INFuture2007

The Future of Information Sciences

Digital Information and Heritage

Odsjek za informacijske znanosti
Filozofskog fakulteta u Zagrebu

1. međunarodna znanstvena konferencija "The Future of Information Sciences: INFuture2007 – Digital Information and Heritage"
Zagreb, 7.-9. studenoga 2007.

Organizatori
Odsjek za informacijske znanosti, Filozofski fakultet, Sveučilište u Zagrebu
Središnji državni ured za *e*-Hrvatsku

Tehnički urednik
Hrvoje Stančić

Dizajn naslovnice
Hrvoje Stančić

Izdavač
Odsjek za informacijske znanosti,
Filozofski fakultet, Zagreb

Za izdavača
Jadranka Lasić-Lazić

Tisak
Vjesnik

Naklada
170 primjeraka

Sve objavljene radove recenzirala je međunarodna recenzentska skupina.

CIP zapis dostupan u računalnom katalogu Nacionalne i
sveučilišne knjižnice u Zagrebu pod brojem 649647.

ISBN 978-953-175-305-0

The Future of Information Sciences

INFuture2007

Digital Information and Heritage

Uredili

Sanja Seljan i Hrvoje Stančić

Zagreb, 2007.

CONTENTS / SADRŽAJ

Predgovor

Budućnost informacijskih znanosti (engl. *The Future of Information Sciences* – **INFuture**) otvara niz konferencija, koje će se održavati svake dvije godine, namijenjenih istraživačima, stručnjacima, poslovnim ljudima i voditeljima projekata iz šireg područja informacijskih znanosti i srodnih struka. Okupljajući stručnjake iz akademskih institucija, vladinih udruga, agencija, kompanija i istraživačkih centara iz Hrvatske, Europe i drugih krajeva svijeta, osnovni je cilj ovih konferencija stvoriti platformu koja bi omogućila razmjenu teorijskih postavki i praktičnih iskustava na području informacijskih znanosti.

S obzirom da Odsjek za informacijske znanosti Filozofskog fakulteta u Zagrebu ove godine slavi trideset godina svoga postojanja, ova konferencija bila je i dobar trenutak preispitivanja postignutih rezultata i vlastitog vrednovanja u relaciji drugih kako bismo procijenili svoj put prema budućnosti (IN Future).

INFuture konferencije istražuju ulogu informacijskih znanosti i srodnih područja kroz tehnološke, obrazovne, istraživačke, organizacijsko- komunikacijske, kulturološke i poslovne aspekte proizašle iz razvoja tehnologije, tržišnih potreba, europskih razvojnih smjernica, znanstveno-nastavnog rada i situacije u Hrvatskoj.

Konferencija 2007. godine pod nazivom **"Digitalne informacije i baština"** (engl. *Digital Information and Heritage*) prva je u nizu INFuture konferencija, ove godine u suorganizaciji Odsjeka za informacijske znanosti Filozofskog fakulteta u Zagrebu i Središnjeg državnog ureda za *e*-Hrvatsku. U knjizi se nalazi više od 50 radova koji kroz različite multidisciplinarne pristupe, relevantne standarde, poslovne sustave i razvojne strategije obrađuju područje primjene informacijske tehnologije u kontekstu očuvanja baštine i kulturnog identiteta.

Knjiga je podijeljena u sedam poglavlja. Nakon prikaza pozvanih izlaganja, obrađene su tematske cjeline vezane uz digitalizaciju, identitet u elektroničkom okruženju i njegovo očuvanje, upravljanje dokumentima u poslovnim aplikacijama i e-vladi, e-baštinu i njezino očuvanje, e-infrastrukturu, te obrazovanje u informacijskim znanostima, koje se međusobno nadopunjuju i isprepliću.

U poglavlju *pozvanih izlaganja* govori se o e-Hrvatskoj i razvojnoj strategiji informacijskog društva, razvoju informacijske znanosti kao društvene znanosti, promjenama u obrazovnom procesu na području informacijskih znanosti, pretraživanju informacija u virtualnom svijetu, potrebi za izgradnjom zajedničkih jezičnih izvora i tehnološke infrastrukture te o potrebi očuvanja digitalnih informacija u Europi.

U poglavlju vezanom uz *digitalizaciju* govori se o arhivskim zapisima kao kulturnom dobru, o nacionalnom projektu digitalizacije u Republici Hrvatskoj, informatizaciji Hrvatskog državnog arhiva, audiovizualnom arhivu na HTV-u, zatim o pouzdanosti podataka na optičkim medijima, digitalizaciji korpusa starohrvatskih tekstova u Institutu za hrvatski jezik, te o repozitoriju u školskim

knjižnicama. Također su prikazane strategije vezane uz izgradnju tematskih zbirki, arhitekture za izradu složene digitalne dokumentacije, uz izgradnju digitalnog leksikona, muzičkih zbirki, zajedničkog označavanja, baštinu i digitalizaciju znanstvenog časopisa te uz digitalni arhiv dječje knjižnice. Prikazana su i nova rješenja vezana uz rekonstrukciju kulturne baštine, oblikovanje baze povijesnih spomenika arhitekture, primjenu digitalnih informacija radi zaštite i rekonstrukcije, nove tehnologije i 3D vizualizaciju, modeliranje, te analizu objekata i nova rješenja u upravljanju kulturnom baštinom.

U poglavlju vezanom uz *identitet u elektroničkom okruženju* prikazani su radovi vezani uz online identitet, kulturološki transfer kroz prevođenje temeljeno na obradi korpusa, uz sravnjivanje dvojezičnih tekstova u postupku izrade prijevodnih memorija, pretraživanje informacija na hrvatskome i rješavanje problema višeznačnosti. Zatim se matematičkim metodama obrađuje korelacija, informacija i kauzalnost te dekompozicija relacijske sheme. Prikazana su i praktična rješenja za digitalizaciju novinske arhive (HINA), obavezni primjerak mrežnih publikacija (NSK) kao i digitalni arhiv mrežnih izvora Hrvatske (HIDRA).

U poglavlju vezanom uz *upravljanje dokumentima u poslovnim aplikacijama i e-vladi* koje je usko povezano s digitalizacijom kulturne baštine, govori se o praksi upravljanja projektima, o upravljanju znanjem u HEP-u, kao i o strateškoj diseminaciji informacija u turizmu.

U poglavlju vezanom uz *e-baštinu i njezino očuvanje* govori se o važnosti razvoja digitalnih knjižnica, ulozi pretraživača za lociranje informacija o kulturnoj baštini, o primjeni softvera u online prikazu kulturne baštine, o virtualnim svjetovima, primjeni softvera s elementima baštine, te o vizualnom identitetu kao obliku komunikacije.

U poglavlju vezanom o *e-infrastrukturi* govori se o biometrijskom identifikacijskom procesu i o značaju skriptnih jezika za administraciju operacijskih sustava.

U poglavlju vezanom uz *obrazovanje u informacijskim znanostima* govori se o statusu informacijsko-komunikacijske tehnologije u obaveznom školovanju, elektroničkim medijskim sadržajima u obrazovanju, multimedijskom softveru i načelima oblikovanja multimedijskog sadržaja, o programima poučavanja informacijske pismenosti i ulozi simulacijskih modela, učenju kroz pružanje usluga, te o novim razvojnim paradigmama.

Vjerujemo da će rezultati rada i radova ove konferencije biti dobar putokaz za promišljanje budućnosti informacijskih znanosti. Vjerujemo da će se rezultati koristiti i za razvoj društva uopće, a onda i za osiguranje pamćenja baštine i našega mjesta u europskom kontekstu. Mi budućnost i rezultate vidimo u napretku i našem doprinosu razvoju i očuvanju baštine kao europskog naroda koji je i do sada svoje intelektualne i materijalne potencijale davao za civilizacijski napredak.

Uredništvo

Preface

The *Future of Information Sciences* (**INFuture**) is a series of conferences that will be held every two years aimed at researchers, professionals, businessmen and project managers from the broad field of information sciences and related professions. The objective of the conference is to provide a platform for discussing theoretical and practical issues in the field of information sciences through networking of researchers and professionals from academy, government institutions, agencies, companies and research centres from Croatia, Europe and other parts of the world.

As the Department of Information Sciences at the Faculty of Humanities and Social Sciences is celebrating its 30th anniversary in the year 2007, the conference was an opportunity for reconsidering of the achieved results and for self-evaluation in comparison with others in order to redefine the way to the future (IN Future).

INFuture explores the role of information sciences and related sciences through technological and educational aspects, research studies, organisational, cultural, communication and business aspects evolved from technological development, market needs, European policies and strategies, educational and research work and the current situation in Croatia.

The conference **"Digital Information and Heritage"** is the first one in the series of INFuture conferences, this year Department of Information Sciences, Faculty of Humanities and Social Sciences in Zagreb co-organising with the Central State Administrative Office for *e*-Croatia. In the book more than 50 papers are presented, elaborating on relevant standards, business applications and development strategies through the multidisciplinary approaches of technology and heritage.

The book is divided in seven chapters. After the introduction, the invited papers are presented followed by topics digitisation, identity in electronic environment and its preservation, document management in business applications and e-government, e-heritage and preservation, e-infrastructure, and information science education.

In the chapter of *invited papers* the following topics have been elaborated: e-Croatia and development of information society, development of information sciences as a social science, curriculum changes in information science education, information search in the virtual world, the need for building of common language resources and technology infrastructure and digital memory preservation challenges in Europe.

In the chapter addressing to *digitisation* the papers related to the following issues have been presented: archival records as cultural goods, national project of digitisation of the Croatian cultural heritage, informatisation of the Croatian archival service, audiovisual archive on Croatian television, data reliability on

optic media, corpus digitisation of old Croatian texts in the Institute of Croatian Language, and repository in school libraries. Strategies related to building of thematic collections, architecture for editing complex digital documents, digital edition of the biographical lexicon, Croatian music heritage and digitisation of sound recordings, collaborative tagging, digitisation of the scientific journal and digital archiving in children’s library. Then, new solutions in cultural heritage reconstruction (3D visualization, modelling etc.) are presented.

In the chapter related to *identity in electronic environment and its preservation* the papers related to the following issues have been presented: online identity, cultural transfer through a corpus-based study, sentence alignment as the basis for translation memory database, retrieving information in Croatian and word sense disambiguation. Further, the issues of mathematical models relating to correlation, information and causality and decomposition of relational scheme are presented. Practical solutions have been presented for digitisation of newspaper-based archive (HINA), depositing system for web publications (NSK) and digital archive of network resources in Croatia (HIDRA).

The chapter *document management in business applications and e-government* is closely related to digitisation of cultural heritage, dealing with improved project management practices, managing knowledge in Croatian power industry, and the issues of strategic dissemination of information in Croatian tourism.

The chapter related to *e-heritage and preservation* deals with importance of usability in development of digital libraries, use of search engines for locating information on cultural heritage, virtual worlds, use of social software of cultural heritage and visual identity as a form of communication.

The chapter on *e-infrastructure* deals with biometrics identification process and the importance of script languages for administration of operating systems.

In the chapter *information science education* the following issues have been considered: the status of information and communication technology in national curriculum for compulsory education, electronic content in education, multimedia software and principles for designing educative multimedia content, programs of teaching information literacy in school libraries, simulation models in information science education, service learning and new paradigm of development.

We believe that the results of this conference and the papers presented here will serve as a sign-post for reconsideration of the future of information sciences. We also hope that the results will be used for the development of the society in general, but also for the preservation of cultural heritage in the European context. We see the future and forthcoming results through our contribution and cooperation in the development of digital preservation in the cultural heritage sector which has been already offering its intellectual and material capital for the progress of our civilisation.

Editors

INVITED PAPERS

e-Croatia 2007 – Fostering the Development of Information Society in Croatia

Diana Šimić
Central State Administrative Office for *e*-Croatia
Gajeva 4, Zagreb, Croatia
diana.simic@vlada.hr

Summary

Programme e-Croatia 2007 is a framework programme for the development of information society in Croatia. Paper presents the main objectives of the Programme, describes the implementation mechanisms, summarises its main results in creating the digital content and discusses future activities necessary to maintain the speed of development achieved so far.

Key words: e-Croatia Programme, implementation mechanisms, information society, ICTs, government, education, research, culture, infrastructure

Introduction

Programme e-Croatia 2007 is a framework programme for promoting the development of information society in Croatia. It was built on the best international practices summarised in the documents of the World Summit on Information Society [ITU 2003a, 2003b], EU programme eEurope 2005 [CEC 2002], and the eSEE Agenda [eSEE 2002] as well as the status of implementation of the Strategy for the Information and Communication Technology [Budin et al, 2002; MST 2003]. Main goals of the Programme were focused on building safe and secure electronic communications infrastructure, creating legal framework for the development of information society, building institutional capacity, fostering public administration reform by decreasing administrative burdens and simplifying business processes in public administration through implementation of ICTs, introducing ICTs as a subject matter and a tool at all levels of education and promoting use of ICTs in public institutions and businesses.

Central State Administrative Office for e-Croatia was created with a mandate to coordinate implementation of the Programme. Its task was to create necessary infrastructure, provide support and coordinate activities and projects applying information and communication technologies carried on by the Croatian national authorities.

In this paper I explain the implementation mechanisms, summarise the main results and outline next steps that need to be taken on Croatia's road to the information society.

Implementation Mechanisms

Implementation of the e-Croatia 2007 Programme involves drafting action plans, coordinating and monitoring their implementation, drafting specific sectoral strategic and implementation documents, preparing implementation reports, commissioning independent benchmarking studies and generally promoting use of ICTs as support for business process change and management.

Since 2004, the Central State Administrative Office for *e*-Croatia has involved more than 30 state administration bodies, agencies and the Croatian Chamber of Commerce in the drafting of three action plans [CSUeC 2004a, 2006, 2007]. Each action plan was an opportunity to include the most current international best practices in the implementation of our national programme. Thus, implementation incorporated the results of the Tunis phase of the World Summit on Information Society [ITU 2005a, 2005b], was harmonized with the EU i2010 Initiative [CEC 2005], and implemented regional broadband [eSEE 2005a] and e-business [eSEE 2005b] initiatives.

Benchmarking of on-line availability of e-government services started in 2004, and continued through 2006, implementing EU methodology to provide comparable results [Capgemini 2005, 2006; T&MC 2006]. Figure 1. shows the increase in availability of on-line information, electronic forms and interactive services since benchmarking started in EU and Croatia.

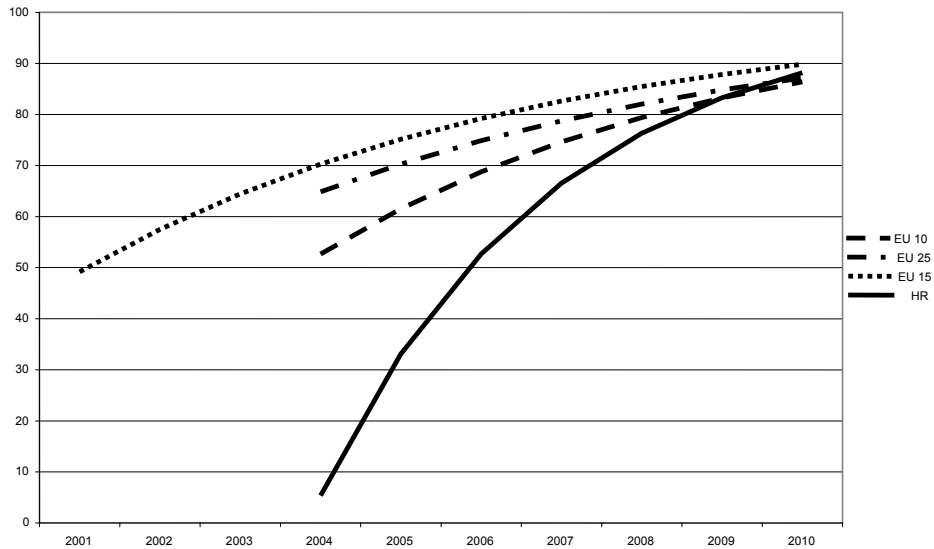


Figure 1. Availability of on-line services in EU member states and Croatia.

While Croatia still lags behind the EU average, the rate of increase of on-line availability was twice as fast in Croatia as in the new member states, and three times as fast as in the old member states.

Study on development of information society was commissioned in 2006. It showed that already in 2005 Croatia was in some aspects better or comparable to EU average. Still, use of ICTs in businesses was at the basic level, and Croatia was specially lagging in broadband availability to the households and the use of advanced internet technologies in the businesses.

Infrastructure

Basic prerequisite for the development of information society is wide availability of safe and secure network infrastructure.

National infrastructure for information security builds upon the National Programme for Information Security [CSAOeC 2005]. Based on this programme, in 2007 a new Law on Information Security was adopted, which sets-up institutional framework for the implementation of an information security policy. CARNet's CERT – Computer Emergency Response Team (http://www.cert.hr) – has thus legally, as well as functionally, become the national centre for responding to on-line treats.

Wide availability and reasonable cost of communication services are directly related to the development of the communications market. Ministry of Sea, Tourism, Transport and Development is responsible for the legal and institutional framework on electronic communications. Building up of the institutional capacity of the Croatian regulatory agency, changes in secondary legislation and adoption and implementation of the broadband strategy [MSTTD 2006] have helped opening up the market, lowering the costs and increasing the broadband penetration in households from less than 1% at the end of 2003 to above 22% in June 2007.

Croatian Academic and Research Network – CARNet has, since its founding, stayed at the forefront of networking developments in international research community. In cooperation with the University Computing Centre of the University of Zagreb (http://www.srce.hr), where CARNet started as a project, today we can say that CARNet leads the way in Europe, and Croatia is among the first countries in the world when it comes to education and research e-infrastructure. All schools, from elementary schools to higher education institutions, and all public research institutes are connected to the common network infrastructure. Number of services – from common authorisation and authentication infrastructure – part of the Eduroam space, to grid infrastructure, e-learning technologies and on-line access to international journals and books are available to CARNet users, and all student dormitories provide broadband CARNet access to all their student users.

With all these results, there are still things that need to be done. In the area of information security there are several by-laws and government decrees that need to be adopted in order to establish a comprehensive and coherent legal framework. CARNet CERT has to increase in staff to be able to continue providing high quality service with an expected increase in the number of internet users.

While the existing Telecommunications Law has enabled development of market and growth in the number of users, it is not harmonized with the new EU framework and in 2008 we expect a new Electronic Communications Law, fully aligned with the EU framework, to be adopted. Finally, Strategy for digital switch-over is being prepared and the introduction of digital TV broadcasting is expected to increase availability and quality of TV signal in many areas of Croatia, while also providing an additional channel for delivery of rich digital content.

Education

Information and communication technologies are essential both as learning infrastructure, and as an area of learning of utmost importance, since knowledge and skills related to the use of ICTs and digital content are becoming more and more a critical factor for employability. Network and equipment infrastructure in Croatian schools is already at a very high level, comparable to the most developed countries. Building on this infrastructure Ministry of Science, Education and Sports (MSES) is funding two portals for elementary and secondary education. CARNet portal for schools is a common web place for exchange of information and digital content for teachers, students and parents. Central portal for distance learning “Nikola Tesla” provides central infrastructure for digital learning content available to all students and teachers in Croatia.

Even though information and communication technology was offered as an elective subject in Croatian schools for a long period, implementation of the Croatian national educational standard provided definition of common content for ICT training. It has also highlighted the need to provide continuous education on ICTs and their use in education to teachers. More than 10 000 teachers went through the ECDL training. Additionally, CARNet’s ELA – e-Learning Academy – provides one-year programs in e-Learning management, e-Learning tutoring and e-Learning design targeted at educators at all levels.

With high quality infrastructure available in Croatian education system, our priorities have to shift to providing high quality training opportunities for teachers, making available high quality digital learning content and gradually introducing information and communication technology as an obligatory subject at earlier stages of elementary education. Teachers should be able to acquire not only basic skills in using computers and the internet, but also the skills to effectively use digital learning content and educational applications, and take an active part in creating new digital educational content. While general educational content of high quality is already available globally and can be localized, it is of the utmost importance to create high quality national educational content (e.g. in language, geography, history, cultural heritage, etc.).

Science

Research in information and communication technology has been the enabling factor for rapid development of information society in the world. While this fact cannot be overstressed, I would like to draw attention to the other facet of research, as a building block of information society that is not as often emphasized. Research is also a producer of vast amounts of digital content in the form of databases, research reports, scientific papers and other publications. The very structure of the world wide web has been developed as a tool for online scientific collaboration.

Croatia has a good track record on creating digital content within the research community. In the World Summit Award competition of 2003 the award in the category e-Science was granted to the Croatian www.blue-world.org, In 2005 the World Summit Award was granted to the www.kopacki-rit.hr in the category e-Science and to the www.fauna.hr in the category e-Learning. This year Croatia was the host of the World Summit Award Grand Jury. At this time the winners are still not known.

The governments of the world agreed in the Geneva Declaration of Principles of the World Summit on Information Society [ITU 2003a] to promote universal access to all scientific knowledge, by inter alia supporting open access initiatives for scientific publications. Croatia is among the few countries that have created an open access portal to scientific journals. Hrčak (hrcak.srce.hr), as it is named, offers access to the journals following the standards adopted through the Open Access Initiative.

Since 2000 Ministry of Science and Technology has funded iProjects – aiming to encourage creation of high quality digital content and services through funding small scale projects in the research institutions and institutions of higher education with a requirement to make results online available. A remarkable number of projects created high quality educational and informative content that remains a part of the Croatian digital landscape.

In order to provide more incentive for creation of digital content within the research community National Council for Science and the Agency for Science and Higher Education should consider including criteria on digital publications among the evaluation criteria for the scientific promotion.

Culture

Croatian cultural heritage is a potentially rich source of digital content. Libraries, museums, galleries and archives hold vast quantities of data and objects that are accessible to only a limited number of visitors or researchers.

Availability of online services in libraries has been among the best since the beginning of national benchmarking, with high number of libraries offering online search of the catalogues and other services. Building on this experience Ministry of Culture initiated the National Programme of Digitisation of Archival, Library and Museum Holdings [MC 2006] aiming to develop and make widely

accessible content in digital form, as a way to protect and evaluate Croatian cultural heritage, present it through national, European and regional networks, preserve cultural diversity and make Croatian cultural contents available for use in education, tourism and other service industries. Implementation of the Programme is carried on through coordination of the Ministry, Croatian State Archives and the Museum Documentation Centre. The first project proposals are already under evaluation. The time-frame for the Programme implementation is 2007-2009.

e-Government and e-Business

Legal framework for information society services in Croatia is in place and its level of harmonization with the Acquis Communautaire is high. It includes Law on Electronic Signature (OG 10/2002), Telecommunications Law (OG 122/2003, 158/2003, 60/2004, 70/2005), Law on Electronic Commerce (OG 173/2003), Law on Personal Data Protection (OG 103/2003), Law on Electronic Document (OG 150/2005), Law on Classified Data (OG 79/2007), and Law on Information Security (OG 79/2007).

Through the implementation of the e-Croatia Programme the Government has introduced a large number of on-line services for citizens and businesses like e-VAT (value added tax), e-Crew (service for yacht charters), e-Land Registry (land ownership), e-Cadastre, e-Customs (online services for shipping agents), e-REGOS (online provision of data for the central registry of retirement insured persons), e-Pension (online provision of data to the Croatian Pension Insurance Administration) as well as online access to other national registries. In addition to providing higher quality of service and increasing its availability, it also aims to stimulate the use of internet by citizens and businesses.

In October this year the Central Portal of Public Administration MojaUprava.hr has been launched as a one stop shop web place for access to online information and services offered by the central public administration. The portal is based on the principle of user centricity. Research on user needs and usability were carried on in order to create more accessible and usable data presentation and portal functionality.

While we have witnessed a rapid increase in availability and use of online services, as well as increase in the use of e-commerce in private sector during the last four years, there are still many assignments to be carried on if we want to continue decreasing the gap between Croatia and the countries with highly developed e-business and e-public administration. Some specific laws and secondary legislation still explicitly require paper document or hand signature. In order to overcome these obstacles, we need to do research on remaining legal barriers to e-business and initiate appropriate change to legislation. Further efforts are needed on raising awareness about benefits of use of ICTs in business and public sector and on existing mechanisms and technologies for ensuring information security and protecting personal data. In the area of basic infrastructure

we need to establish national electronic identity management infrastructure with central gateway for authorisation and authentication for users of on-line services and civil servants. In a longer run we should introduce a national ID on a smart card. We have only started building interoperability framework for public administration information systems. Further work in this area will have to be coordinated with European efforts through the IDABC Programme of the European Communities. Croatia has already joined this programme. Building of the safe virtual private network HITRONet was a necessary prerequisite for back-office integration across different public authorities. Secure transactional middleware is the other such prerequisite that needs to be built.

e-Inclusion

While ICTs open new opportunities, they can also become an additional barrier to access to knowledge, if we do not take special care to remove or at least decrease these barriers from the beginning.

In Croatia, today, many web pages are created implementing latest fads in design and interactivity technologies, without taking into account needs of the people with disabilities. Within the *e*-Croatia Programme we emphasise the need to make digital information available to all groups of people. Thus we promote conformance with W3C WAG AAA (http://www.w3.org/TR/WAI-WEBCONTENT/) for all public web sites. The need to provide resources to all groups of citizens is also addressed in the Broadband Strategy – which aims to make broadband available not only to large cities, where investment into infrastructure is commercially viable, but also to those areas of Croatia where it is not profitable to invest in such infrastructure – such as islands or remote mountain areas. Availability of school computer labs, and broadband connectivity is essential to providing access to digital content and information resources to children in less well-off families. Switching to digital TV broadcasting will also create an additional channel of communication that will be more easily available and also more accessible to all.

Conclusion

Our efforts in building e-Croatia often remind me of the Red queen's reply to Alice: "Now, *here*, you see, it takes all the running *you* can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!" [Carroll, 1988]. We have started later, and we have been running 2-3 times as fast as other European countries. And we need to carry on running. The strategic framework and the implementation mechanisms we introduced have provided results in practice. We need to carry on, assessing the results as we go, and adjusting the objectives and priorities to the current situation. Coordinated action of all stakeholders is necessary to achieve the potential synergistic effect on the growth and competitiveness of the whole society.

References

Budin, Leo; Bajica, Marija; Carić, Antun; Čerić, Vlatko; Glavinić, Vlado ; Lovrek, Ignac; Manger, Robert ; Ursić, Srebrenka. Information and Communication Technology in the Strategy of Development of the Republic of Croatia. (in Croat.) Zagreb: Office for the Strategy of Development of the Republic of Croatia, 2002.

Capgemini: Study on online availability of public services for e-Croatia, Final report. Zagreb: Capgemini d.o.o., January 2005. (http://www.e-croatia.hr)

Capgemini: Study on online availability of public services for e-Croatia (second research). Zagreb: Capgemini doo, February 2006. (http://www.e-croatia.hr)

Carroll, Lewis. Through the Looking Glass and what Alice found there. In: The Complete Works of Lewis Carroll. London: Penguin Books, 1988, pp. 121-250.

CEC: eEurope 2005: An information society for all: An Action Plan to be presented in view of the Sevilla European Council. COM(2002) 263 final. Brussels: Commission of the European Communities, 28.5.2002.

CEC: i2010 – A European Information Society for growth and employment. COM(2005) 229 final. Brussels: Commission of the European Communities, June 2005.

CSAOeC: Operational Plan for the Implementation of the e-Croatia 2007 Programme in 2004. (in Croat.) Zagreb: Government of the Republic of Croatia, Central State Administrative Office for e-Croatia, June 2004a. (http://www.e-croatia.hr)

CSAOeC: Strategy of the One-Stop-Shop Programme (in Croat.) Zagreb: Government of the Republic of Croatia, Central State Administrative Office for e-Croatia, Central State Administrative Office for Public Administration, Financial Agency, December 2004b. (http://www.e-croatia.hr)

CSAOeC: National Programme on Information Security in The Republic of Croatia. (in Croat.) Zagreb: Government of the Republic of Croatia, Central State Administrative Office for e-Croatia, Workgroup for Information Security, March 2005. (http://www.e-croatia.hr)

CSAOeC: Operational Plan for the Implementation of the e-Croatia 2007 Programme in 2006. (in Croat.) Zagreb: Government of the Republic of Croatia, Central State Administrative Office for e-Croatia, April 2006. (http://www.e-croatia.hr)

CSAOeC: Operational Plan for the Implementation of the e-Croatia 2007 Programme with an Overview of Avtivities in 2007. (in Croat.) Zagreb: Government of the Republic of Croatia, Central State Administrative Office for e-Croatia, May 2007. (http://www.e-croatia.hr)

eSEE: eSEEurope Agenda for the Development of the Information Society – a cooperative effort to implement the Information Society in South Eastern Europe. Belgrade: Stability Pact for the South Eastern Europe, eSEE Initiative, October 2002.

eSEE: Memorandum of Understanding on the Development of a Unified Market of Broadband Networks fully Interconnected to the European and Global Networks – Initiative for ‘bSEE’ – Broadband South Eastern Europe. Thessalonici: Stability Pact for the South Eastern Europe, eSEE Initiative, July 2005a.

eSEE: Guidelines for a National Strategy on Building the Framework for e-Business. Thessalonici: Stability Pact for the South Eastern Europe, eSEE Initiative, UNECE, July 2005b.

IDC Adriatic: Study on Information Society Developments in Croatia in 2005. Zagreb: IDC Adriatic, October 2006. (http://www.e-croatia.hr)

ITU: Declaration of Principles: Building the Information Society: a global challenge in the new Millennium. Document WSIS-03/GENEVA/DOC/4-E. Geneva: ITU, December 2003a. (http://www.itu.int/wsis/)

ITU: Plan of Action. Document WSIS-03/GENEVA/DOC/5-E. Geneva: ITU, December 2003b. (http://www.itu.int/wsis/)

ITU: Tunis Commitment. Document WSIS-05/TUNIS/DOC/7-E, Tunis: Second Phase of the WSIS, November 2005a. (http://www.itu.int/wsis/)

ITU: Tunis agenda for the information society. Document WSIS-05/TUNIS/DOC/6(Rev. 1)-E18. Tunis: November 2005b. (http://www.itu.int/wsis/)

MC: National Programme of Digitisation of Archival, Library and Museum Holdings. (in Croat.) Zagreb: Ministry of Culture, November, 2006.

MST: National Report on Strategy Implementation: Information and Communication Technology Croatia in the 21st Century 2002-2003. Zagreb, Ministry of Science and Technology, November 2003.

MSTTD: Strategy of the Development of Broadband Internet Access in the Republic of Croatia by 2008. Zagreb, Government of the Republic of Croatia, Ministry of Sea, Tourism, Transport and Development, September 2006. (http://www.mmtpr.hr)

T&MC: Benchmarking study – online availability of public services (third research). Zagreb: T&MC, October 2006. (http://www.e-croatia.hr)

CLARIN: Common Language Resources and Technology Infrastructure

Steven Krauwer
Utrecht University – Faculty of Humanities
CLARIN/ELSNET/UiL OTS
Trans 10, 3512 JK Utrecht, NL
Phone: +31 30 2536050, fax: +31 30 2536000
steven.krauwer@let.uu.nl
http://www-sk.let.uu.nl/

Summary

This paper gives an overview of the CLARIN project, which aims at the creation of a common language resources and technology infrastructure in Europe, to serve the humanities and social sciences research communities.

Key words: available language resources, technology infrastructure, humanities, social sciences, access.

The mission

The CLARIN [1] mission is to create an infrastructure that makes language resources and technology (LRT) available and readily usable to scholars of all disciplines, in particular the humanities and social sciences (HSS). In our age we are presented by many challenges as we deal with language in electronic formats, in spoken, written, and multimodal forms, as a carrier of information, as an object of study, and otherwise. The volume of texts and recorded spoken texts is enormous, and it is growing exponentially. The sheer size of this material makes the use of computer-aided methods indispensable for many scholars in the humanities and in neighbouring areas who are concerned with language material.

The CLARIN infrastructure is based on the firm belief that the days of pencil-and-paper research are numbered even in the humanities. Computer aided language processing is already used by a wide variety of sub-disciplines in the humanities and social sciences, addressing one or more of the multiple roles language plays, as carrier of cultural content and knowledge, instrument of communication, component of identity and object of study. Current methods and objectives in these disparate fields have a lot in common with each other. However it is evident that to reach the higher levels of analysis of texts that non-linguist scholars are typically interested in, such as their semantic and pragmatic

dimensions, requires an effort of a scale that no single scholar could, or indeed, should afford.
The cost of collecting, digitising and annotating large text or speech corpora, dictionaries or language descriptions is huge in terms of time and money, and the creation of tools to manipulate these language data is very demanding in terms of skills and expertise, especially if one wants to make them accessible to professionals who are not experts in linguistics or language technology. The benefits of computer enhanced language processing become available only when a critical mass of coordinated effort is invested in building an enabling infrastructure, which can then provide services in the form of provision of all the existing tools and resources as well as training and advice across a wide span of domains. Making resources and tools easily accessible and usable is the mission of the CLARIN infrastructure initiative.

The technological challenge

The purpose of the infrastructure is to offer persistent services that are secure and provide easy access to language processing resources. As language, speech and vision technology improve, it should be commonplace to carry out tasks such as: *"summarize Le Monde from 11th March 2007" "list all uses of 'enthusiasm' in 19th century English novels written by women", "find all video clips of Tony Blair on the BBC in 2007"*. But without the proper infrastructure, these technologies to make these tasks possible will only be available to a few specialists. At present one needs to find an appropriate program (to do translation, summarization, or extraction of information, etc.), download the program, make sure it is compatible with the computer that will execute the program, understand the form of input it takes, download the data (e.g. novels, newspapers, corpus, videos), and convert them to the correct format for the programs, and all this before one can get started. For most researchers outside of computer science, at least one of these tasks will be an insurmountable barrier. Our vision is that both the resources for processing language and the data to be processed be made available in usable formats and can be run over a distributed network from the user's desktop. The CLARIN objective is to make this vision a reality: repositories of data with standardized descriptions, language processing tools will be amended to operate on standardized data, legal and access issues will be resolved, and all of this will be available on the internet using Grid architecture.
The nature of the project is therefore primarily to turn existing, fragmented technology and resources into accessible and stable services that any user can share or customize for their own applications. This will be a new underpinning for advanced research in the humanities and social sciences, a "research infrastructure".

Where we stand

To avoid any misunderstandings: the CLARIN infrastructure described here does not yet exist. Even if one finds repositories of language data in most European countries, and even if some of them are technologically quite advanced there has never been an attempt in Europe to link the existing repositories across national frontiers and to interconnect them in such a way that to the user they present themselves as a large scale facility with (at least virtually) one single entry point offering access to a broad variety of data and services.

Unlike in fields such as nuclear physics, climate research, environmental research, energy, space research and many others, where shared infrastructures with transnational access are quite common, nothing of this kind has ever existed for the humanities and social sciences. Only recently the European Commission has taken initiatives towards a long term roadmap for research infrastructures in Europe, explicitly including the infrastructure needs of the social sciences and humanities. This initiative, called ESFRI [2], has recently led to a report describing 35 essential research infrastructures for Europe. This report, called the ESFRI Roadmap, has now been taken up by the EU and by the member states with a view to possible future implementation in a three-stage process: Preparatory Phase, Construction Phase and Exploitation Phase. As a first step the EU has launched a call for proposals whereby the selected infrastructures were invited to come up with a proposal for a Preparatory Phase for each of the envisaged infrastructures.

The CLARIN infrastructure is one of them and the contract negotiations between the participants and the EC are now ongoing. The start-up of the project is expected at the end of this year or early next year. It will have duration of 36 months, after which –if the project is successful- the Construction Phase will start. In the rest of this paper we will provide more information about our approach and our activities in the first three years. The CLARIN consortium is led by Utrecht University, and it has 31 partners from 22 EU and associated countries. In addition there is a wider community of CLARIN members, ca 90 institutions with specific expertise in language resources, spread over 33 countries.

Objectives of the Preparatory Phase

According to the EC call for proposals for the preparatory phase it has to aim at bringing the project to the level of legal, organisational and financial maturity required to implement the project. As the ultimate goal is the construction and operation of a shared distributed infrastructure that aims at making language resources and technology available to the humanities and social sciences research communities at large an approach along various dimensions is required in order to pave the way for implementation. We briefly describe the four main dimensions and the preparatory phase objectives for each of them.

First of all there is the funding and governance dimension. The aim here is to bring together the funding agencies in all (now 22) participating countries and

to work out a ready to sign draft agreement between them about governance, financing, construction and operation of the infrastructure.
Secondly there is the technical dimension. A language resources and technology infrastructure is a novel concept. Even if it will be based on existing and emerging technologies (grid, web services) there are no off-the-shelf blueprints for the architecture of such an infrastructure. The technical objective is to provide a detailed specification of the infrastructure, agreement on data and interoperability standards to be adopted, as well as a validated running prototype based on these specifications. The validation should cover both technical aspects, linguistic aspects (see below) and user aspects (see further below). The construction of the prototype will also help to make realistic cost estimations for the construction and exploitation phases.
The third dimension is the language dimension. For the validation of the specifications of the infrastructure and the proposed standards the running prototype will have to be populated with a selection of language resources and technologies for all participating languages. This population process will normally take place by adaptation and integration of existing resources to the CLARIN requirements although in a number of cases the creation of specific essential resources will be necessary. The objective is to deliver a sufficiently populated and thoroughly tested prototype that demonstrates the adequacy of the approach for all participating languages, a prototype that can be used to bootstrap the construction phase.
The fourth and most important dimension is the user dimension. The intended users are the humanities and social sciences research communities. In order to fully exploit the potential of what language technology has to offer, a number of actions have to be undertaken: (i) an analysis of current practice in the use of language technology in the humanities will help to ensure that the specifications take into account the needs of the humanities, (ii) the execution of a number of typical humanities projects will help validating the prototype and its specifications, (iii) the humanities and social sciences communities have to be made aware of the potential of the use of language resources and technology (LRT) to improve or even innovate their research, (iv) the humanities and language technology communities have to be brought together in order to ensure lasting synergies between the communities. The objective of this cluster of activities is to ensure that the infrastructure has been demonstrated to serve the humanities and social sciences users, and that we create a joint, informed community that is capable of exploiting and further developing the infrastructure.

Technical Infrastructure

The overall objective of the CLARIN research infrastructure project is to turn existing, fragmented technology and resources into accessible and stable services so that users can use them the way they want it. This type of advanced infrastructure enabling eScience scenarios is comparatively new, we can only re-

fer to a few projects, such as for example the Physiome project, that are working on similar goals facing similar degrees of complexity, i.e. before the real construction work can be done prototypical work needs to be carried out to study

- all aspects of applying modern architecture concepts such as Service Oriented Architectures to the field of LRT;
- the level of standardization that is necessary to implement workflow concepts making it more easy for users to use and combine language resources and technology components to solve their advanced problems;
- the various requirements to deploy a distributed service and server landscape in Europe with a broad geographic coverage;
- the problems of integration and interoperability in our domain with a more comprehensive approach.

In building the ingredients of a modern distributed digital infrastructure, we can rely on the activities of many different fields and standardization bodies such as W3C, ISO, OASIS etc. and we can look back on a long formation process in our discipline. Standardization in the area of linguistic resource management has a long tradition (EAGLES, ISLE, MILE, TEI) that recently led to the formation of ISO TC37/SC4, for example. However, the level of unification and the range of acceptance achieved is not yet sufficient to implement a service oriented architecture that would guarantee the required degree of integration and interoperability. Nonetheless, the conformity achieved already now in the field will allow us to take the next steps towards automated workflows. In contrast to earlier attempts CLARIN has the broad coverage of institutes and well-known experts that is needed to overcome the barriers that we are faced with.

Building an infrastructure offering a rich domain of services requires work at different levels. All of them will be based on the fast European network and the goal must be to build a foundation rich and powerful enough to generate new types of complex eScience applications. Grid-like functionality will ensure that communication occurs between trusted servers, that all resources in the domain have unique and persistent identifiers, that authentication and authorization is working seamlessly in distributed scenarios, that users of organizations participating for example in national identity federations are accepted with a single identity etc. We need a network of powerful centres that can offer stable and highly available services of a great variety. These ranges from archiving services allowing others to store data resources with a guarantee of long-term accessibility to advanced ontology services that offer widely accepted and well defined domain concepts. These centres need to offer portals to register all types of resources (data, tools, knowledge components) such that they can be accessed and interpreted by humans and algorithms. Standards need to be worked out that are flexible enough to cope with resource type differences and sub-discipline terminology and that enable automated workflows. Services will be needed that help to overcome the syntactic and semantic differences that we are

faced with when applying a certain tool to a certain resource or when concatenating two tools to a new more complex operation.
For all these different layers different steps are at hand. The middleware technology to establish a federated domain of data repositories has been tested already by projects such as DAM-LR [3]. Here the next step is to test its scalability in a pan-European network with national hubs and to train young people to fully understand and to maintain such basic infrastructures. At other levels such as for achieving structural and semantic interoperability still much analysis and specification work needs to be done to come to a set of generic frameworks, before adaptation, construction and implementation work in large quantities can be carried out. Still for a number of services such as for the ISO TC37/SC4 data category registry for example we have to create prototypes to test out principles such as high availability by distributed and synchronized services and automatic workload distribution.

Landscape of Services and Centres

The CLARIN infrastructure will establish a rich landscape of services and centres. As indicated above, the types of services are very heterogeneous and they will be hosted by certain centres. Centres that want to offer specific services at the European level need to fulfil a number of criteria such as long-term support by funding agencies, competence and cost effectiveness. In addition, centres will be selected based on the criterion of geographical distribution. A set of open criteria will be published together with the specification of the services. Interested parties can apply for taking over these services at the European level and will be selected by means of an evaluation process.

Activities

We will do theory-based analysis work by making use of the results that have been achieved in many standardization projects (EAGLES, ISLE/IMDI, TEI, ISO, W3C, OASIS, OAI/OLAC, OAIS, etc) and related initiatives and projects (OGF, Internet2, D-Grid, HAKA, EGEE, TERENA, DEISA, DAM-LR, etc). We will do design studies with a limited vertical and horizontal coverage to deeply understand the problems that need to be solved. The vertical studies include various layers ranging from the basic grid services to the integration applications, the horizontal studies investigate one layer and the interplay and scalability issues at a pan-European level. Driven by top-down decisions prototyping work needs to be done to test out elements of the infrastructure. We will keep a close and continuous eye on the needs of our customers, i.e. the humanities and social sciences research community. To meet the needs and requirements of our users prototypical infrastructure elements - be it services or centers - will be set up not only to demonstrate the functionality of the design choices, but also to learn from the experiences and to make realistic effort and cost estimates. We will work closely together with IPR experts to specify the types of

agreements that will be necessary from a purely technological point of view on the one hand and to understand the requirements from IPR considerations on the infrastructure on the other hand.

User needs

We will start out analyzing past and existent projects that were carried out in the domain of the humanities such as the Bricks project to get a deeper understanding about the concrete use of LRT in the humanities and about the state of technology in that area. We will also stimulate a few new concrete projects to help pointing to gaps on the one hand and to develop requirements. The projects to be carried out can be extensions of existing or finished projects to achieve compliance with CLARIN or completely new ones tailored to test the emerging infrastructure concepts. To gain a detailed understanding of the requirements for LRT infrastructure in the humanities domain the humanities projects need to be selected very carefully, i.e. criteria need to be set up that in particular include chains of LRT operations, multilingual scope and addressing the increasing interest of young people in multimedia presentations. CLARIN is committed to boosting humanities research in a multicultural and multilingual Europe, by facilitating access to language resources and technology for researchers and scholars across a wide spectrum of domains in the humanities and social sciences (HSS). To reach this goal we will establish an active interaction with the research communities in HSS and to contribute to overcome the traditional gap between the Humanities and the Language Technology communities. We will achieve these goals by gathering information about: potential collaborators in network building, the current impact of LRTs in these fields, and user needs.

LRT Overview

We will collect detailed information about existing language resources and technology, specify and prototypically implement representational standards and strategies for achieving interoperability, and validate the proposed technical standards and service specifications of the CLARIN infrastructure. It is obvious that much effort will be required to integrate the speech/ multimodality resource and technology sub-community which is not yet well-represented in CLARIN and that strong links need to be established with the community working on ontological issues.

The term language resource subsumes the whole range of linguistic data types such as text, speech and multimodal corpora, lexica, treebanks, typological databases, grammars, ontologies, schemas, and the term language technology covers a wide range of processing and annotation components such as taggers, parsers, semantic extractors, manual annotation tools, speech alignment tools, etc. The survey needs to include a detailed analysis of the structural and encoding characteristics of the resources and the interfacing and import/export characteristics of the tools that will serve to design service oriented architecture.

Based on this broad and detailed investigation a comprehensive taxonomy of language resources and tools will be created which will be the basis for the classification and standardization work.
In addition, a Basic Language Resource Kit (BLARK [4]) for the major languages will be specified and gaps for individual languages will be identified by a coordinated action involving the CLARIN members from the various countries. This will include criteria for the quality assessment of resources and tools, since many resources and tools lack basic characteristics to make them effectively available in a service oriented architecture serving the humanities needs. It will be left to the national decision boards whether they will fill the gaps that are identified. Depending on the requirements some new resources and tools may be developed already in the preparatory phase with high priority.
Five activities will be in the focus: (1) investigating all aspects that have to do with the integration of resources and tools into the infrastructure; (2) studying the problems of interoperability in detail; (3) studying usage scenarios including chains of operation in detail; (4) creating missing resources and tools where they are required for the success of the preparatory phase; (5) validate the specifications and the prototypical implementations.

IPR Issues

Despite the clear commitment of CLARIN to open access and open source principles for all resources and code developed in CLARIN, legal and ethical aspects cannot be ignored. A rich LRT domain as intended by CLARIN will be bound to include protected material and therefore we will have to build the necessary legal and ethical agreement patterns into CLARIN. During the preparatory phase we will need to develop a thorough understanding of these problems and need to work out first such patterns to prepare the construction phase. Agreements and licenses are needed for successful cooperation among the various actors and users of CLARIN, and for achieving and maintaining sufficient levels of trust. A network of agreements, licenses and auditing is needed to relate the actors to each other and to avoid or reduce risks incurred in possible violations of intellectual property rights (IPR) or basic ethical rules. CLARIN may also include new business and accounting models. The opportunities and threats of the inclusion of commercial services have to be studied carefully in order to understand the positive and negative effects for the CLARIN infrastructure that will be constructed. We will have to create sufficient model agreements for the operational prototype and test beds that will be built in CLARIN during the preparatory phase and to find out what their coverage and level of acceptance is. The work can build on earlier work such as in the LE-PAROLE project whose licensing scheme consisting of four parties can be adapted to the CLARIN situation. Also for distributed authentication and authorization scenarios CLARIN can build on experience made in a number of national identity federations and in the DAM-LR project where the goal was to

establish a federation of archives. Keys to all these federations are trusted organizations that can offer (1) a reliable identification of users and (2) a reliable certification of the user reading the license and having signed it. Methods for auditing the reliability of trusted organizations must be put in place and provided to the resource collectors, for routine periodic re-assurance of trust. Since CLARIN wants to be open to guest researchers worldwide it has to be analyzed how the strict rules can be applied. Only reliable distributed authentication schemes open the way to providing licenses for large numbers of users to access materials and tools, and to migrating legacy language resources into wide use within CLARIN.
A number of additional issues need to be taken up such as (a) licensing templates for language materials and tool usage, (b) establishing model licenses to be used for new CLARIN materials, (c) migrating legacy licenses into CLARIN, (d) licensing Templates for Language Technology Software and (e) a common code of conduct covering widely agreed ethical rules.

Construction and Exploitation Agreement

Even though a large proportion of the work to be carried out in this project has to do with the specification of the infrastructure, interoperability standards and an IPR framework the main deliverable of this project is a single draft agreement between all participating countries and players to move on to the Construction Phase. This by no means simple, as such an agreement will have to include a significant long term financial commitment: We envisage a construction phase of ca 5 years, followed by an exploitation phase that might easily cover up to 10 years. Other issues to be addressed include governance, management, and coordination of national programmes related to language resources. None of them are hot topics from a scientific point of view, but they are crucial for the success and sustainability of CLARIN.

Concluding remarks

CLARIN has still a long way to go but it offers an exciting opportunity to fully exploit the achievements of especially language (and even speech) technology over the last decade to the benefit of communities that traditionally do not maintain a close relationship with human language technologies. Contrary to many EU programmes the main beneficiaries of this project are not expected to be the big ICT-oriented industries or the bigger language communities in Europe: CLARIN addresses the whole humanities and social sciences research community, and it very explicitly addresses all languages of the EU and associated states, both majority and minority languages, both languages spoken and languages studied in the participating countries.

Acknowledgements and references

This paper is an abridged version of a document written by Peter Wittenburg (Max Planck Institute, Nijmegen), Tamas Varadi (Hungarian Academy of Sciences, Budapest), Martin Wynne (Oxford Text Archive, Oxford), Kimmo Koskenniemi (University of Helsinki) and Steven Krauwer, with contributions by many other members of the CLARIN community.

References

http://www.clarin.eu
http://cordis.europa.eu/esfri/
http://www.mpi.nl/DAM-LR/
http://www.elsnet.org/blark.html

Curriculum 2.0? Changes in Information Science Education for a Web 2.0 World

David Bawden
Department of Information Science, City University London
Northampton Square, London EC1V 0HB, England
dbawden@soi.city.ac.uk

Lyn Robinson
Department of Information Science, City University London
Northampton Square, London EC1V 0HB, England
lyn@soi.city.ac.uk

Theresa Anderson
Faculty of Humanities and Social Sciences, University of Technology
PO Box 123, Broadway NSW 2007, Australia
anderdorfer@gmail.com

Jessica Bates
School of Information and Library Studies, University College Dublin
Belfield, Dublin 4, Ireland
jess.bates@ucd.ie

Ugne Rutkauskiene
Institute for Librarianship and Information Science, Vilnius University
3 Universiteto str., Vilnius LT-01513, Lithuania
ugnerut@gmail.com

Polona Vilar
Department of Library and Information Science and Book Studies, University of Ljubljana
Aškerčeva 2/V, 1000 Ljubljana, Slovenija
polona.vilar@ff.uni-lj.si

Summary

This paper reports an international comparison of changes in library/information curricula, in response to the changing information environment in which graduates of such courses will work. It is based on five case-studies from Australia, Ireland, Lithuania, Slovenia and the United Kingdom. Specifically, it describes responses to changes: an increasing proportion of e-content; and the

impact of the communication and social networking features of Web 2.0. It examines both changes in curriculum content, and in methods of teaching and learning. The latter involves pedagogy adapting and changing in the same way as the professional environment, with a greater emphasis on e-learning, and use of Web 2.0 tools. Students therefore learn about the issues by making use of the tools and systems in their studies.

Key words: LIS curriculum development, Web 2.0, e-learning, information society, collaboration

Introduction

The development of the collection of tools, techniques and approached which are treated together under the general heading of ‘Web 2.0’ is increasingly entering the consciousness of library and information specialists worldwide. Whether these developments threaten traditional library / information services, or whether they provide new opportunities and capabilities, is a matter for debate, but it is certain that the library / information world must be aware of them.
One way in which this will happen is for these topics to impact the library / information studies curriculum; both as subjects to be taught, and as tools for teaching. How this may best be done - given the wide variety of such tools, their rapid development, and the relative lack of expertise in their use among many library / information educators – is far from obvious.
The five contributions to this paper present a ‘snapshot’ of ways in which this issue is being addressed in universities worldwide; from Australia, Ireland, Lithuania, Slovenia and the United Kingdom. The contributors present their views in their own style, without any imposed template or format, so that local concerns and perspectives can clearly emerge. The concluding section brings out some genera issues and findings.

1. Australia

Reflecting on my experience as an LIS educator has meant working at the level of the LIS specialist programs as well as at the level of Faculty wide "core subjects." I will describe briefly two examples to demonstrate my approach. In both instances, students are working with these tools and being asked to critically reflect on the efficacy of the tools in use in various contexts.
A Faculty-wide subject (Communication and Information Environments)
This subject is part of the common first semester curriculum for the approximately 600 first-year undergraduates enrolled in one of the BA in Communication degrees offered in our Faculty. As a Faculty of Humanities and Social Sciences, the specialties include journalism, public communication, media arts production, creative writing, cultural studies, social inquiry as well as information management.

As a foundation course for the undergraduate program, it is essential that students not only study communication and information environments, but that they learn to develop critical capacities involving discussion forums, blogs and collaborative online tools. Thus, as part of my redesign of the subject, I decided to make fuller use of collaborative online tools like small group blogs within our syllabus this past year. Doing so involved shaping class activities, work between classes as well as a major assignment around the use of these blogs.

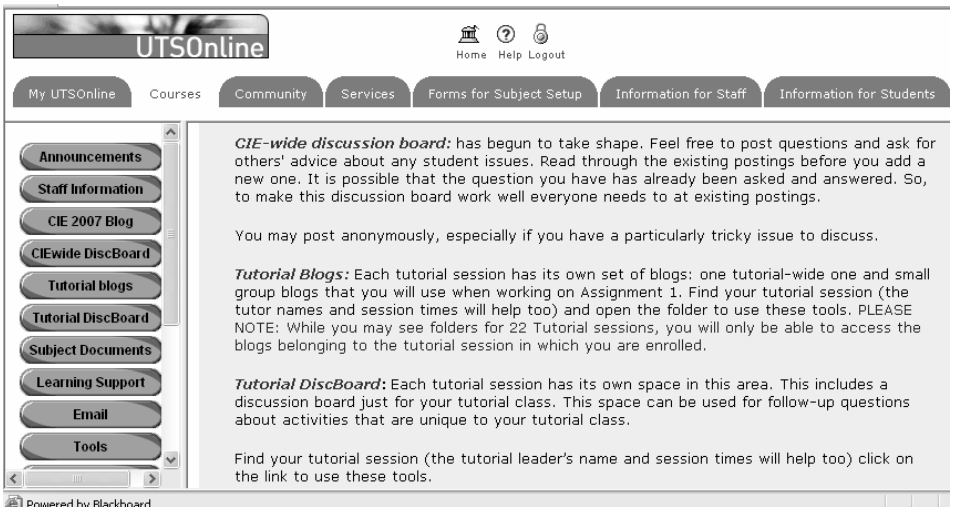


Figure 1 shows the three levels of blogging zones created within the online course site – class-wide, tutorial wide (e.g.: each of the 22 tutorials had a zone) and group wise (e.g.: each tutorial class had four or five small groups who worked weekly in their group blog).

Research has shown that making activity assessable in some way is a great motivator when it comes to online tools such as the ones we would be using in class. Equally, e-learning research also invites caution about the risk of overloading students and crowding curriculum with online elements that do not add value with regards to the learning outcomes and desired graduate attributes. In the case of this particular subject, the link was an easy one to make: learning to think critically about the merits and flaws of various forms of online and face-to-face communication pathways available to these students was explicitly related to the desired attributes of communications graduates. It was also an essential factor to be considered in terms of the actual content of the subject. Through this approach, teaching issues about contemporary communication and information environments including what we now label Web2.0 is a central focus of this subject. Thus, Web2.0 technologies are both sites of learning and tools for learning.

An LIS specialist subject (Social Informatics)

In a second subject (Social Informatics; 60 students in core undergrad for LIS students; elective for postgraduate LIS and rest of Faculty) the focus of weekly class activities centres on discussions about emerging technologies and theoretical frameworks associated with Social Informatics that help us to understand and interrogate the way such e-techs may be used in society. It builds on the first example – taking both the content and the use of e-tools in class further by using blogs, wikis, virtual classroom tools, IM/real time chat and podcasts.

Once again, as in the first example, it is imperative for students looking at these themes theoretically to make active use of such tools. Their first-hand experience can then used to clarify theoretical issues.

In this subject, the students create digital scrapbooks that categorise and analyse reading they are doing on the theme. Many students create blogs (individual) to deliver their scrapbooks and to reflect on the many "texts" (ranging from traditional academic works through to popular literature and music, advertisements and TV programs as well as blogs, rssfeeds and e-zines) they collect as part of their knowledge artefact. They analyse this material in terms of the bidirectional influences of society and information & communication technologies, such as that discussed in class and identified in their reading.

Use of Web2.0 tools goes further in this subject that in the first example – and it does so deliberately as part of a progressive approach to the lesson they need to learn about multimodal collaboration and work. We discuss ways to represent the ideas they wish to communicate and relate their experience of creating digital scrapbooks with the work of creating a catalogue for a museum exhibition. Discussion in class and analysis in the assignment relates this work to issues of knowledge access and organization, helping the students to examine the implications of Web2.0 worlds and uses for enabling all 'users' to effectively become knowledge producers.

Blogs may be used for individual scrapbooks, but the main focus in this second subject is online collaboratories of small groups (4-6 students) who must take responsibility for a 2 week online discussion around their chosen e-technology. For this purpose, wikis are created at two levels:

1. Student teams each get a private work zone assigned in the online content management system. In includes live chat, file sharing and a wiki site where they prepare a wiki that serves as the central site for discussion with their class colleagues around their e-tech.
2. As part of a small team, they have responsibility for facilitating discussion about their e-tech topic and summarizing/weaving the activity in their wiki for the rest of their tutorial class. As part of this Moderating Team, students work together to plan, facilitate, monitor and synthesise the two-week discussion in their tutorial's Blog for that particular emerging technology.

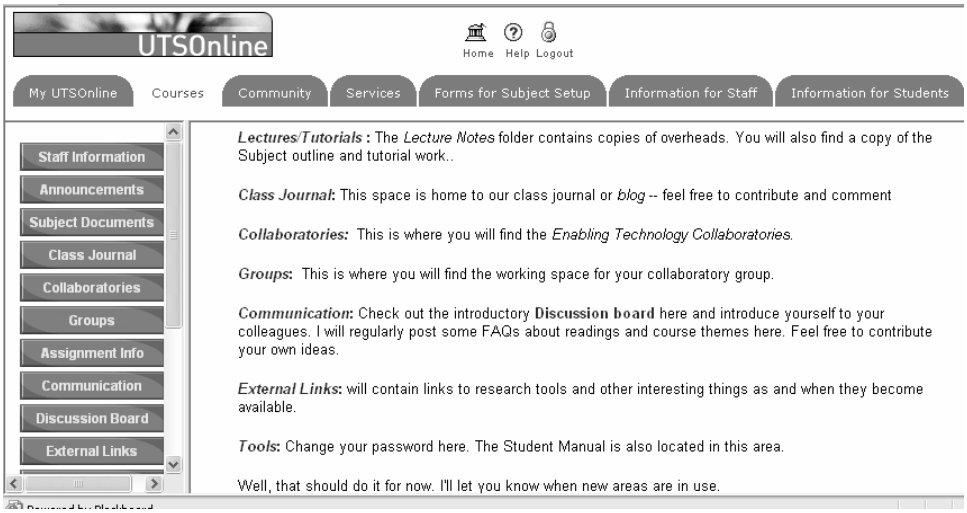


Figure 2 shows the layers in this subject.

Each team has a personal online work zone (e.g.: private discussion space, file exchange, wiki page) that can be used for group-only communication and behind-the-scenes work. Teams are also encouraged to meet in person as required. Students are expected to take active part in each of the emerging technology collaboratories in operation for any given week. They cycle fortnightly in each tutorial so that students are able to focus on one theme at a time.

Students are encouraged to bring their online chats into the classroom (that is, lessons learned, themes covered). A major assessment item involves them preparing a critical reflection of the collaboratory experience (at all levels and in all roles). They look at the benefits and drawbacks of face-to-face and online communication for workplace and social activities. Furthermore, a final assignments (critical evaluation of one of the emerging technologies under discussion) draws on the wikis constructed in the process of working through the student-moderated discussions. The online work also forms the starting point for lecture discussions in second half of semester (along with the class-wide blog that encourages the sharing of ideas on the subject's themes).

2. Ireland

Web 2.0 refers to a second generation of web-based applications and services and in particular the use of the web as a platform for user-generated content and web-based communities. Komito describes the "rubric of Web 2.0" as consisting of "user-generated content, dynamic web publishing and online social groups" (Komito, 2007: 85). Associated technologies include: blogs, social networking sites, wikis, mashups, podacasts, RSS feeds, shared bookmarks and image sites. Web 2.0 is intrinsically linked to the 'semantic web'.

Web 2.0 encompasses a variety of different meanings that include an increased emphasis on user generated content, data and content sharing and collaborative effort, together with the use of various kinds of social software, new ways of interacting with web-based applications, and the use of the web as a platform for generating, re-purposing and consuming content. (Franklin & van Harmelen, 2007: 4)

Around the world LIS curricula increasingly recognise the importance of Web 2.0. There are three main facets to teaching and learning activities that relate to Web 2.0:

- The technological developments
- The social uses / impacts
- Implications for the field and the profession

There are a small number of specific modules that deal with Web 2.0, but more commonly aspects of Web 2.0 are included as a part of existing modules.

Of the modules offered by the UCD School of Information and Library Studies, three in particular include aspects of Web 2.0 in their curricula. Of these two are offered at Level 3 (and taken by mainly students in the third and final year of a primary degree, and also by a small number of postgraduate students on the Graduate Diploma and Masters in LIS) and one at Level 4 (for postgraduate students enrolled in either the Graduate Diploma in LIS or the Masters in LIS).

The two Level 3 modules, IS30010: 'Weaving the Web: The Internet and Society', and IS30070: 'Cybersociety? Technology, Culture, and Communication', examine current developments in Web 2.0 and students taking these modules develop an understanding of the transition from the Internet to Web 2.0. 'Weaving the Web' focuses more on the technological changes that are enabling greater interconnectivity, and the 'Cybersociety' module is more concerned with the social impact of online communities and use of social technologies.

In the Level 4 module, IS40080: 'Information and Society', the emphasis is more towards Web 2.0 from the perspective of library and information professionals. The impacts of Web 2.0 on information provision are examined in a critical way and students are encouraged to explore and debate the implications of Web 2.0 in relation to library and information work.

The use of Web 2.0 tools in teaching is still in its infancy, although it is likely that LIS educators are among early adapters in this respect.

It has been common practice for some years to make use of virtual learning environments (VLE) in university teaching and in the UCD School of Information and Library Studies all taught modules make some use of a VLE, although the extent to which the VLE is utilised, and how it is utilised varies from module to module. The main VLE used in the teaching and learning activities of the UCD School of Information and Library Studies is Blackboard.

"The Blackboard Virtual Learning Environment (VLE) is used to provide elearning services to staff and students. It allows lecturers to share course materials, post announcements for students, and run quizzes and surveys. Blackboard

also includes a Virtual Classroom facility, discussion forums and various other course tools." (http://www.ucd.ie/itservices/staffit/elearning/)
However, VLE's such as Blackboard are not generally true examples of Web 2.0 tools. Innovations in this area would include use of Second Life for teaching and learning activities, and alternative tools such as Moodle. Both of which are currently being considered by the School, as is podcasting.
One other area in which Web 2.0 is anticipated in having an increasing impact in the School's teaching and learning activities is in assessment of student work, for example the use of social software in group projects.
There is also research expertise among staff in the UCD School of Information and Library Studies in relation to Web 2.0 (in terms of technological developments and programming and social dimensions), and at present two postgraduate students are working on doctoral theses in the areas of applying Web 2.0 technologies to collaborative learning environments, and the use of ICTs (including Web 2.0) by a geographical, proximate suburban community in Ireland.

3. Lithuania

The Librarianship and Information Science Institute is a part of the Faculty of Communication of Vilnius University (VUFC). VUFC is the leading Higher Education establishment working for the broad field of information and communication professions in Lithuania. Established in 1991, VUFC is led by an integrated approach towards information and communication, looking at libraries, museums, archives, media, publishing enterprises, information agencies and information businesses as parts of the information infrastructure of the society. Currently, VUFC offers five BA and ten MA study programs, has 40 FTE employees, ca. 1200 BA and MA students and 26 PhD students.
In the LIS Institute of Vilnius University teaching about Web 2.0 is currently integrated into the Internet Communication course for second course LIS students. This course can also be freely taken by any student of other programs of Communication Faculty of Vilnius University.
The format of the course has been developed in the light of the social networking site 43 Things [http://www.43things.com], where users create accounts and then share lists of goals and hopes. The content of the course has been adopted from the Learning 2.0 program, originally created by Helene Blowers at the Public Library of Charlotte & Mecklenberg County [http://plcmc.org] Also the experience of other libraries, running the 23 Things program (Yarra Plenty Regional Library [http://www.yprl.vic.gov.au/Learning_2.0/index.htm], Marylathe public library [http://www.maryland librarieslearning2about.blogspot.com], etc.), has been explored.
The 23 hours length on-line course is based on the Moodle virtual learning environment. It consists of 10 training modules, each exploring the particular Web 2.0 tool. The course begins with the introduction to this new concept of the Wide World Web. After an introduction to Web 2.0 philosophy, students are in-

vited to take part in learning journey through Web 2.0 tools, where they learn about blogs, wikis, Flickr, YouTube, webcasting, podcasting, RSS, Del.icio.us, Library Thing and Library 2.0 concept. The course is a self-discovery program which encourages students to take control of their own learning through exploration and play. Students are encouraged to work together and share with each other their discoveries, techniques and "how to’s" both in-person and through discussion forums.
As the course is a part of formal LIS education, it was not possible to give the full ownership to the learning process to the students. Their learning process is supervised by the tutor, who is monitoring, consulting and assessing learning process of students. The final assessment of each student is generated from the assessment results of each module.
Year 2007/2008 are considered as piloting of the course. After the piloting the program will be reviewed and modified according to the students’ feedback. Also it is planned to transform the distance course into a Wiki web site, to allow free access to the course material to the Lithuanian libraries staff.

4. Slovenia

Since Web 2.0 refers to a perceived second generation of web-based communities and hosted services – such as social-networking sites, wikis and folksonomies – which aim to facilitate collaboration and sharing between users (O’Reilly, 2005), it seems important that LIS students, as future information professionals, are aware of these changes and innovations and that they know more about these than an average user to be able to cope in the tomorrow information world. In this context there are a few questions. One is, how to incorporate these themes into the curriculum. Due to the lack of common understanding as to what Web 2.0 actually is, a certain amount of scepticism and pragmatism is needed.
Another question is, of course, how to use the emerging technologies and means of communications in the teaching process itself. It seems that in this part higher education has been more successful by setting up web learning materials, e-learning environments, etc., on which Web 2.0 facilities may be built. Here we will present some thoughts connected with e-learning, as well as experiences and prospects of e-learning at the Department of LIS at the Faculty of Arts, University of Ljubljana.

E-learning

As argued by many, among them Krevs (2007), e-learning can be seen as use of technology for reaching pedagogical goals. There are a number of benefits we can achieve through e-learning, e.g. better knowledge, better e-literacy, control over study materials, more independent and current student engagement, independence of time/place, easier student and teacher mobility. It offers potential benefits for teachers, students, institution (faculty, university), society. Studies

(e.g. Mullen and Tallent-Runnel, 2006) show that e-learning environment has great influence on student perceptions of the learning process, their teachers and tutors, and in addition also influences their motivation, satisfaction and learning. Donnert (2004) has presented both, positive and negative perceptions of students regarding e-learning. These are summarized in Table 1.

Table 1: Positive and negative student perceptions of on-line learning (Donert, 2004)

Top five negative aspects	**Top five positive aspects**
Technical problems (21%)	Flexibility (89%)
Isolation (18%)	Can interact with peers (27%)
Lack of support (16%)	Access to resources (21%)
Lack of other student interaction (14%)	Effective mode of learning (17%)
Too few face-to-face opportunities (14%)	Can interact with tutors (16%)

The learning is or should be a creative process and key factor is the level of engagement of the learner. With the use of technology this is made very different from the traditional ways of teaching and learning through better accessibility, easier time management, incorporation of individual differences. Lavrič (2006) argues that through e-learning we can achieve integration of the three pedagogical approaches (Figure 3).

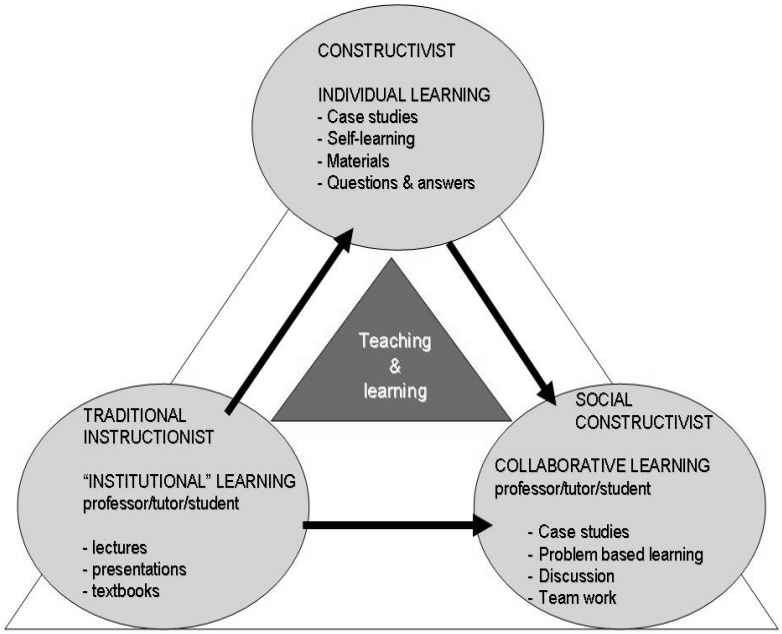


Figure 3: Integration of three pedagogical approaches (Lavrič, 2006)

Krevs (2007) has presented many different aspects of e-learning. The reasons for using e-learning, which can be also seen as possibilities which it offers:

- Better use of teacher's and student's time, greater freedom for planning
- More student activity, greater responsibility of students in the learning process
- More independent and current student engagement
- Better overview over the teaching/learning material
- Different forms of learning, of materials (dictionaries, glossaries, quizzes, AV materials, …)
- Direct student engagement ("the growing classroom")
- Different ways of communication with and among students (contact hours, forums, blogs)

E-learning doesn't guarantee that we will achieve all these, but can effectively help in doing so.

Requirements for the use of e-learning technologies for teachers:

- Knowledge and skills for preparation of e-materials and e-classrooms
- Attention to planning student involvement, their assignments, means of communication, knowledge assessment methods, grading (e.g. portfolio?)

Requirements for the use of e-learning for students:

- Knowledge and skills for use of e-materials and e-classrooms
- Internet access
- Regular visits and steady rhythm of work in e-classrooms
- Strict meeting of deadlines

Faculty of Arts

Previous experience in Slovenia which is rather sketchy, left to enthusiastic individuals (usually coming from the area of Computer Science or Informatics), and scattered across all levels of education, shows that commercial firms usually overestimate potential market, especially in public higher education, but that there are numerous good examples which as a rule stem from good teaching practices (Krevs, 2007). There are actually two approaches to introduction of e-learning: the low-budget, pedagogy driven way (bottom-up) which is less controlled, more spontaneous, and planned, systematic and institutionalized (up-bottom) which has more control and better organization. An ideal scenario should probably take both into account.

The introduction of these technologies at the Faculty of Arts had wide support, but there were also some warnings and concerns. People were mainly concerned with workload, problems with technology and support, lack of expertise, effect on current teaching/learning (Krevs, 2007). Such concerns are likely to appear with greater emphasis with greater introduction of Web 2.0 features.

At Faculty of Arts there have been mainly good experiences with the introduction of e-learning environment http://e-ucenje.ff.uni-lj.si/ based on Moodle

(Krevs, 2007). Moodle is an open source software package which by now has over 200.000 registered users, over 30.000 registered sites in 75 languages in over 192 countries (Moodle statistics, 2007). It is designed to enable teachers create e-courses, offers a variety of interaction possibilities, and can be used both in small classes and in large universities. This allows ready inclusion of Web 2,o facilities.

The process begun by firstly individual teachers using e-learning as guests on outside servers. Soon a special workgroup was formed which quickly set up a server and e-learning environment Moodle and proceeded with courses and workshops to introduce the concept to the faculty. Some pilot projects were set up (pilot e-classrooms, e-courses as support to usual teaching). The final stage were courses for students which were mainly prepared by individual departments. Statistics (Krevs, 2007) shows that by April 2007 there were over 100 e-classrooms (19 from departments of English and Library and Information Science, 12 from department of German, 9 from department of Asian Studies, 1 or 2 from other departments), 51 teachers, 1830 registered participants, 75% of participants visiting at least every 2 days, 60% every day, and around 10.000 per month, in total over 44000 visits and over 262000 lookups with average of 6 pages per visit. This good use is a solid basis for 'building in' Web 2.0 features.

Department of LIS

It has been mentioned that the Department of LIS was one of the first at the faculty to start working with the e-learning environment by setting up a few pilot e-classrooms. The new Bologna study programme which started last year is entirely supported by Moodle. It practice this means that we do not have a lot of experience and can therefore speak mainly of prospects and goals. The faculty workgroup prepared a workshop to introduce Moodle to all the teachers and assistants in the department.

Students in the first year of Bologna study programme were immediately introduced to the environment. No survey has yet been made, but it seems that they did not have great difficulties in using it. They appreciated to have all the materials prepared in advance. They accepted the means of knowledge assessment and grading, e.g. portfolios, mostly respected deadlines for the assignments, and actively participated in other activities, e.g. forums. In the future, Web 2.0 facilities, e.g. blogs, will be included to support communication with the teachers as well as among students themselves.

Among the things which were well accepted both by students and teachers Krevs (2007) mentions appreciation of support to faculty and students, and of environment offering great variety of usage, enhanced student involvement, good possibilities of support to a relatively simple, permanent and qualitative process and last but not least actual added value to the teaching process. There were of course some problems, e.g. teachers were reluctant to become 'technicians' (a problem likely to increase as more 'advanced' Web 2.0 features are

added, high quality user interfaces were not well accepted, the belief that technology itself is already added value needed to be overcome, students complained that some teachers were putting up videorecordings of lectures as teaching materials, sometimes it was difficult to achieve spontaneous co-operation of students.
In perspective we can say that e-learning is an important measure of quality of institutions of higher education. It supports the 'battle' for students as well as for better knowledge of graduates. What we need at this point is integration of various scattered initiatives into a whole in a search for successful and efficient e-learning approaches. Web 2.0 approaches have the potential both to help and to hinder this.

5. United Kingdom

The Department of Information Science at City University London teaches a variety of Masters level courses in various aspects of the library and information sciences. Two courses in the geographic information area may be followed in fully distance mode, with the students relying entirely on e-learning, while the greater majority of students follow a form of flexible learning, blending face-to-face attendance at the university with complementary e-learning.
The advent of Web 2.0 has led to much consideration as to which aspects of this phenomenon should be taught (content) and how they may be themselves used in teaching (methods). This presentation outlines both of these for City University's courses in Information Science and in Library and Information Studies, with the proviso that these are at an early stage, and we are very much feeling our way as to how best to proceed.

Content

While the 'popular culture' instances of Web 2.0 have received a great deal of publicity, their long-term effects on the library/information profession are far from clear. We therefore wished to steer a middle course between avoiding the issues entirely, and making too much of them, when it is far from clear which will be of most long-term significance. We also wished to avoid creating any specific module or course devoted to 'Web 2'; rather we wanted to bring these matters into our existing course structures. Finally, we were aware that our students would typically be more aware of much of Web 2.0 than the academic staff; this calling for a careful way of presentation.
Web 2.0 issues are therefore appearing in many, if not most, modules of the library / information courses. The main issues covered are new forms of communication (blogs, RSS, wikis, podcasting and vidcasting, etc.), social networking (MySpace, YouTube, etc.), media sharing (YouTube, Flickr, etc.), and social tagging and folksonomy. The emphasis is on those aspects of these issues which affect the creation and communication of recorded information, and hence the work of the library / information specialist.

Some specific examples are:

- New communication media affecting the publication chain are covered in detail in a course on 'Libraries and Publishing'
- Social tagging and folksonomies are covered, and compared with more conventional approaches in a course on knowledge organisation
- The advantages and disadvantages of wikis and other socially constructed knowledge resources are covered in a course on digital literacy
- Basic philosophical and societal issues resulting from the development of Web 2.0 are dealt with in courses on the foundations of the library and information sciences

Naturally, given the current interest in these topics, an increasing number of students are keen to undertake dissertation projects on Web 2.0 topics.

Means

We are keen to use Web 2.0 facilities into our teaching for three reasons: they offer real advantages over other methods; they give students an insight into academic and professional use of these facilities; and they increase the credibility of our teaching, in what is sometime seen, wrongly, as an old-fashioned subject.
However, we are aware of the need for care. Merely because it is easy to use Web 2.0 does not mean it is easy to use it well. And, as noted above, many students are very familiar with Web 2.0, and will not be impressed by an amateur or inappropriate approach.
Again, our approach has been to introduce Web 2.0 'organically' into existing approaches. City University uses the WebCT / Blackboard e-learning system, which, though well-tried and robust, have a somewhat old-fashioned interface. We are, for the most part, introducing Web 2.0 facilities incrementally, to augment or replace equivalent but less effective, e-learning measures.
Specific examples include:

- Replacement of discussion board communication by blogs
- Use of wikis to gather student contributions, rather than attached files
- Use of podcasts of audio recordings of lectures, as a complement to text files

We have also adopted blogging as a way of conducting some academic administration, to replace email exchanges.

Conclusions

From the rich and diverse insights given by these five case studies, we may distil some general principles. Such principles must always be qualified by local circumstances, and by the choice of which Web 2.0 tools and features to adopt, from the variety available.
'Integration' is a key concept. Web 2.0 provides both the content of learning, and the tools to promote learning itself. This gives a particularly strong, and de-

sirable, integration of theory and practice. It is also worthwhile to integrate teaching of these topics with related activities of research or consultancy in the teaching department, where such exist.

Students typically have a natural enthusiasm for these topics, and often greater expertise in some aspects than most teaching staff. If this enthusiasm can be harnessed, and students can learn for themselves and from each other, then the effects will be particularly positive. Similarly, the use of Web 2.0 facilities by academic staff themselves, for their own purposes, will enhance their understanding, and hence promote more credible teaching. Where academics are lacking in expertise and confidence, this must be built up gradually and sensitively.

The introduction of Web 2.0 into teaching is best done incrementally, starting with particular courses or topics, and expanding on the basis of knowledge gained. Given the investment of effort into the development of e-learning systems in many academic departments, it seems very sensible to use this as a platform for development of the Web 2.0 curriculum.

References:

Donert, K. (2004). Analysing the use of virtual learning, collaboration and communication tools in Innovation Education. European Teacher Education Network, volume 6. University of Greenwich Press, 101-114

Franklin, T & van Harmelen, M (2007) Web 2.0 for Content for Learning and Teaching in Higher Education, available at: http://www.jisc.ac.uk/media/documents/programmes/digital_repositories/web2-content-learning-and-teaching.pdf (accessed 5 September 2007)

Komito, L (2008, in press) ‘Information Society Policy’, in: Hearn, G & Rooney, D (eds) Knowledge Policy: Challenges for the 21st Century. Cheltenham, UK: Edward Elgar, pp 85-99.

Krevs, M. (2007). Visokošolsko e-izobraževanje: izkušnje, prednosti in nevarnosti. Retrieved September 13 2007 from http://e-ucenje.ff.uni-lj.si/file.php/168/gradiva/krevs-delavnica-4-07.ppt

Lavrič, A. (2006). Osnove visokošolske didaktike: program za izpopolnjevanje. Uporaba informacijske komunikacijske tehnologije v izobraževanju: gradivo za udeležence, 4. del. Ljubljana: Filozofska fakulteta, Center za pedagoško izobraževanje.

Moodle Statistics Retrieved September 10 2007 from http://moodle.org/stats/

Moodle. http://moodle.org (accessed 5 September 2007)

Moodle. Retrieved September 10 2007 from http://moodle.org/

Mullen, G. E., Tallent-Runnels, M. K. (2006). Student outcomes and perceptions of instructors’ demands and support in online and traditional classrooms. Internet and Higher Education 9, 257–266.

O’Reilly, T. (2005).What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software. Posted September 30 2005, retrieved September 13 2007 from http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

Second Life. http://secondlife.com (accessed 5 September 2007)

University College Dublin IT Services – elearning at UCD http://www.ucd.ie/itservices/staffit/elearning/ (accessed 5 September 2007)

Je li informacijska znanost još uvijek društvena znanost?

Miroslav Tuđman
Filozofski fakultet
Ivana Lučića 3, 10000 Zagreb, Hrvatska
mtudman@ffzg.hr

Summary

Autor postavlja pitanje odgovornosti i teorijske opravdanosti postuliranja informacijskog društva kako bi se potvrdio alibi da je informacijska znanost društvena znanost. Tradicionalni predmet informacijske znanosti – javno znanje – nije više svijet objektivnih i istinitih informacija. Novi predmet informacijske znanosti postala je digitalna stvarnost, koja funkcionira kao javno znanje ali više nema zadaću prenositi istinite i objektivne informacije; javno znanje ima funkciju društvene kontrole i nadzora. U digitalnoj stvarnosti informacije postaju objekti bez značenja, a istina irelevantna i/ili zastarjela kategorija. Autor propituje epistemološke odrednice obavijesti, znanja, istine u novom digitalnom svijetu. Smatra da na početku 21. stoljeća sudbina informacijske znanost, kao i ostalih društvenih znanosti, ovisi o ishodu prevrednovanja vrednota na kojima će se graditi budućnost zapadnog svijeta. Sada se informacijska znanost nalazi u procijepu između univerzalnih vrijednosti koje su na odlasku, i ireverzibilnih globalnih vrijednosti koje grade i na kojima se gradi digitalni svijet.

Key words: informacijsko društvo, paradigma informacijske znanosti, digitalna stvarnost, javno znanje, istina, obavijest, znanje

Uvod: informacijsko društvo kao postulat

U medijima, tj. u dnevnom i ideološkom diskursu, agresivno se nameću teze da živimo u informacijskom dobu, da je obavijest ključni potencijal socijalnog, gospodarskog i tehnološkog razvoja, da nema napretka bez primjene informacijskih tehnologija u svim sferama ljudske djelatnosti, da je primjena ICT dovela do globalizacije, da onaj tko ima obavijesti (a ne više nužno sirovine, rad ili energiju) ima moć i može osigurati hegemoniju (ili barem dominaciju u svojem području djelatnosti), te da je konačni cilj socijalnog i kulturnog razvitka stvaranje informacijskog društva, odnosno društva znanja.

Na početku 21. stoljeća većina zemalja u svijetu žele definirati svoju "nacionalnu informacijsku politiku" kako bi ostvarili "društvo znanja", tj. društvo koje svoj razvoj temelji na znanju i po mjeri znanja. Informacijska ekonomija, in-

formacijsko-komunikacijska tehnologija (ICT), odnosno obavijest, postala je globalna pokretačka sila razvoja. Pod konac 20. stoljeća takva gledišta svjedoče konvencije i povelje međunarodnih organizacija, i većine nacionalnih zakona demokratskih zemalja, jer obavijest definiraju kao razvojni potencijal koji mora biti svima dostupan, te se pravo na pristup informacijama smatra jednim od novih temeljnih ljudskih prava u demokratskim društvima.
Ovakve teze prihvaćaju i teoretici i praktičari informacijske znanosti, najčešće kao premise od kojih polaze, a rijetko kao predmet kritičkih rasprava. Razlozi tome mogu biti brojni, no za naše razmatranje važne su implikacije teze da teoretici informacijske znanosti prihvaćaju informacijsko društvo kao zadaću, a praktičari pridonose uspostavi informacijskog društva, odnosno društva znanja. Takvo polazište s jedne strane implicitno osigurava informacijskoj znanosti status društvene znanosti[1], a s druge takvo je stajalište vrlo profitabilno jer informacijskim projektima osigurava prioritete u nacionalnim, regionalnim i globalnim razvojnim planovima.

Pitanje odgovornosti

Pedesetak godina od svog nastanka informacijska znanost ne treba se više brinuti ni za svoj akademski a ni za socijalni status. Ona zato mora postaviti pitanja o svojim mogućnostima i zadaćama i potražiti odgovore koji će joj osigurati vjerodostojnost i budućnost. Naime, osim svojih pozitivnih strana, informatizacija i globalizacija već danas pokazuju i svoje tamne strane. Informacija nije samo čimbenik razvoja i blagostanja; ona je dobila i niz novih neželjenih atributa i primjena. Obavijest je postala oružje (M. Y. Whitehead, 1997); informacijski prostor postao je bojišnica na kojoj se vode informacijski ratovi prema novim doktrinama (Edward Waltz, 1998; Field Manual No. 3-13); korporacije, države i međunarodne organizacije razvijaju informacijske strategije kako bi sebi osigurali hegemoniju i dominaciju u svijetu (Jim Winters, John Giffin, 1997); informacijske mreže postale su sredstvo i prostor potpuno nove forme informacijskog rata: mrežnog rata (*netwar*)[2].
Stvaranjem globalnog informacijskog prostora (*cyberspace-a*) 1990-ih, događaju se radikalne promjene predmeta obavijesti ali i predmeta javnoga znanja. Informacijski rat postaje dominantni oblik za ostvarivanje nacionalnih, gospodarskih, političkih, kulturnih, itd. interesa, s ciljem da se na protivničkoj strani

[1] To slijedi iz definicije društvene znanosti: "Social Science – any discipline or branch of science that deals with human behaviour in its social and cultural aspects" (Encyclopaedia Britannica: Social Science).

[2] "Netwar is an emerging mode of conflict in which the protagonists – ranging from terrorist and criminal organizations on the dark side, to militant social activists on the bright side – use network forms of organization, doctrine, strategy, and technology attuned to the information age" (John Arquilla, David F. Ronfeldt, ed. 2001).

osigura dominacija vlastitih informacija, kako bi se protivnik prisilio da donosi odluke koje nisu u njegovu interesu (M. Tuđman, 2007a).

Već se iz ovih nekoliko naznaka da naslutiti kako se u informacijskom prostoru što ga oblikuju internet i nove informacijske tehnologije, ne oblikuje korpus objektivnog i istinitog znanja; postojeći komunikacijske mreže nemaju zadaću organizacije i prijenosa istinitih informacija, već je njihova zadaća osigurati u javnosti dominaciju određenih informacija; odnosno, u informacijskom prostoru ne postoje mehanizmi kontrole javnoga znanja prema kriteriju istine i objektivnosti (M. Tuđman, 2007b).

Ako su gornje naznake točne, onda su opovrgnute i poništene početne i temeljne postavke informacijske znanosti o istinitom i objektivnom znanju kao predmetu njezina bavljenja[3].

S pravom se možemo zapitati tko je odgovoran za uvjerenja o informacijskom društvu[4] kao panaceji i mitu? Preciznije, može li informacijska znanost nekritički preuzeti takva stajališta i ugraditi ih u polazišta svoje znanstvene paradigme, te samim time preuzeti odgovornost i za sve neželjene i neplanirane posljedice u stvaranju informacijskog društva?

Može li informacijska znanost bez preuzimanja odgovornosti za svoj doprinos istraživanju i svijetlih i tamnih strana razvoja informacijskog društva biti društvena znanost? Ako želi biti društvena znanost, onda informacijska znanost mora smoći snage i odgovornosti upisati i opisati svoj doprinos moralnom, umnom i duhovnom razvoju društvene zajednice, doprinos koji se neće mjeriti tek tehničkim i kvantitativnim pokazateljima (primjene ICT).

Osobno sam uvjerenja da je informacijska znanost društvena znanost, te da je njezin predmet interesa i bavljenja organizacija i razmjena znanja. Međutim, svjedoci smo dekonstrukcije temeljnih postavki paradigme informacijske znanosti, kao i njezina predmeta. Informacije su sve manje istinite i objektivne, a sve češće pogrešne, neistinite i lažne. Korisnici su prema toj činjenici ostali ravnodušni jer su informacije izgubile značenje, a traži se prvenstveno njihov smisao. Korisnicima su na raspolaganju ogromne količine informacija, ali do znanja teško da mogu doći. Znanje nije više cilj po sebi. Ono postoji ako je isplativo, proizvodi se ako je unosno.

[3] Naime, od samog svog početka teoretici informacijske znanosti često su se pozivali na tezu Johna Zimana "znanost je javno znanje"; štoviše ruski su autori (Mihajlov, Giljarevski) definirali informacijsku znanost kao znanost o znanstvenim informacijama. B. Težak vjerovao je da prvenstveno fundamentalne znanosti stvaraju istinite i objektivne informacije koje moraju činiti jezgru javnoga znanja. Istinito i objektivno javno znanje tvrdi B. Težak pretpostavka je slobodnog, javnog i odgovornog djelovanja (vidi M. Tuđman, 2007b).

[4] Sociološke, filozofske i političke postavke su poznate (vidi na primjer F. Webster ili UNESCO, 2003.). Ovdje postavljamo pitanje o teorijskom i epistemološkom utemeljenju informacijskog društva u području informacijske znanosti.

Informacijska je znanost sredinom 20. stoljeća, poput drugih znanosti, težila biti objektivna znanost: utvrditi istinu i zakonitosti organizacije i razmjene znanja u realnom svijetu. Međutim, informacijska znanost u 21. stoljeću ne bavi se više realnim svijetom. Njezino je područje bavljenja *cyberspace*: digitalni svijet u kojem su temeljne kategorije "objektivnog" svijeta iščezle. U digitalnom svijetu eliminirani su kako "vrijeme i prostor", tako i "tijelo i stvari", a ostale su samo znamenke i riječi (R. Capurro). Objektivni svijet odmetnuo se u digitalni svijet (ili prema Baudrillardu u hiperrealnost[5]) a informacijska znanost još uvijek koristi kategorijalni aparat o "objektivnom" svijetu iz sredine 20. stoljeća.

Do sada se informacijska znanost samo pozivala na informacijsko društvo[6] kako bi osigurala sebi status suvremene i lukrativne znanosti. U zbilji ona nije kontrolirala rast i razvoj digitalne stvarnosti. Ona je bila tek alatka u stvaranju informacijskih mreža, *cyberspace-a*, sama bez znanja o tome kuda je krenula i odgovornosti za to kamo će stići.

Ako se želi baviti istraživanjima i organizacijom različitih korpusa znanja (njihovim ustrojem, arhitekturom, organizacijom, razmjenom, itd.), informacijska znanost mora preispitati dosadašnja filozofska, socijalna i teorijska polazišta svoje zastarjele znanstvene paradigme.

Entuzijazam javnog i ideološkog diskursa o ulozi i ciljevima informacijskog društva treba zamijeniti kritički stav o "virtualnoj stvarnosti" kao području informacijske znanosti te o njezinu predmetu kao digitalnom objektu. Iako smo pod pritiskom javnog entuzijazma o svemoći informacija, bilo bi poželjno suočiti se s posljedicama činjenice da se "stvarnost nepovratno strukturira oko informatičke i virtualne mreže", i priznati da je informacijska znanost vrlo malo, ili gotovo ništa, pridonijela utvrđivanju istine o toj "virtualnoj stvarnosti" – iako je u veliko pomogla njezinu oblikovanju.

Nekoliko teza o fragmentaciji i dekonstrukciji znanja

Želimo postaviti tek nekoliko pitanja za koja niti nemamo, niti zagovaramo cjelovite odgovore. Namjera nam je tek s nekoliko fragmenata započeti dijalog o (re)definiranju predmeta i područja informacijske znanosti. Takav zahtjev postavila je nova digitalna stvarnost: ako tim zahtjevima teorija informacijske znanosti ne odgovori, informacijska znanost će zastarjeti, svest će se tek na uslužnu djelatnost, na struku bez teorije i kao znanost će nestati.

[5] "Na djelu je i urušavanje stvarnosti u hiperrealizam, u brižno podvostručavanje stvarnosti…" (Baudrilliard, str. 101), odnosno kako Rade Kalanj piše u predgovoru iste knjige "Novi, a to za Baudrilliarda znači postmoderni način mišljenja moguć je jedino ako shvatimo da se stvarnost nepovratno strukturira oko informatičke i virtualne mreže, oko poretka simulakra koji apsorbira cijelu socijalnu sferu i očituje se kao njezina simulirana hiperrealnost" (str. xxv).

[6] Virtualna stvarnost, *cybercpace*, društvo znanja, itd. samo su sinonimi podjednake nepreciznosti u opisu željenog područja bavljenja informacijske znanost.

Postulat objektivnosti

- Jedini a priori postulat za znanost je objektivnost. Znanost se organizira, i za svoje opravdanje traži "neku stvarnu, 'objektivnu' referencu, u stvarnom procesu" (Baudrillard, str. 84-85). "Objektivna" referenca je ono o čemu znanost izvješćuje, i ta je referenca njoj zadana. Digitalna stvarnost nova je referenca informacijske znanosti.

Kognitivna hijerarhija

- Tradicionalna je postavka da su obavijesti prikaz pojava, podataka ili događaja. Objektivnost i istinitost obavijesti ovisi o preciznosti prikaza sadržaja obavijesti u realnom svijetu. Sadržaj obavijesti "proizvod" je kognitivnih procesa: spoznaje, prosudbe i razumijevanja. Kognitivna hijerarhija (od podatka, obavijesti do znanja) upućuje samo na jednu dimenziju obavijesti: njezin sadržaj, a sadržaj je određen "stvarnošću". Prikaz sadržaja obavijesti određuje logika kognitivne funkcije kako bi se obavijest mogla koristiti kao istinit podatak i objektivna činjenica.

Obavijest bez referencije

- U digitalnom svijetu došlo je do reza između forme obavijesti i sadržaja obavijesti[7]. U semiotičkoj terminologiji: označitelj referira na označeno (sadržaj misli) i na referenta (neki objekt u stvarnosti; referent i označeno zajedno čine sadržaj znaka/obavijesti). U digitalnom svijetu dolazi do redukcije: moguće je povući znak jednakosti između obavijesti i objekta obavijesti[8]. Obavijesti više ne prikazuju sadržaj nekog objekta u realnom svijetu. One same su objekti u digitalnom svijetu; objekti koji nemaju nikakvu referencijalnost, jer nisu više ekvivalent koji predstavlja objekt u nekom drugom svijetu. To ima za posljedicu da su obavijesti izgubile značenje: sadržaj predstavljanja. Zato informacije više ne određuje kognitivna funkcija. Značenje, činjenična vrijednost, istinitost ili lažnost, objektivnost – postali su irelevantni atributi obavijesti. U komunikaciji obavijest ima vrijednost samo ako se može razmijeniti s drugim podacima unutar komunikacijskog sustava, samo ako ima smisao. Komunikacijski sustavi ne vode brigu o činjeničnoj vrijednosti (istini) obavijesti. Zato zadaća komunikacijskih sustava nije razmjena istinitih, već korisnih (relevantnih) obavijesti.

Strukture bez pamćenja

[7] Ili u drugoj terminologiji nestala je razlika između podatka (znaka) i "stvarnog" referenta.

[8] PIN, ili lozinka nisu prikaz nekog objekta u stvarnosti nego su stvarni objekti koji funkcioniraju u "objektivnom" svijetu. Takve obavijesti (znakovi) nisu supstitut stvarnosti – oni postaju i jesu stvarnost.

- Obavijesti su na pola puta između podataka i znanja. Obavijesti su rezultat procesuiranja podataka formalnim jezicima. Što je sustav obrade sofisticiraniji to je privid da se obavijest izjednačila sa znanjem veći. Proizvodi baza podataka, informacijskih sustava, operacijskih sustava, OLAP-a, itd., ostat će tek jedna od vrsta obavijesti. Ma kako složeni procesi računalne obrade "proizvodili" obavijest, ona time neće postati znanje. Može biti samo ono što jest: proizvod formalnih jezika a ne kognitivne funkcije.
- Prirodni jezik kojim se prenose obavijesti već podrazumijeva povijest: različita iskustva što ga pošiljatelj i primatelj imaju u komunikaciji. Upravo zato što se presijecaju "jezični prostori" pošiljatelja i primatelja, postoji komunikacijska napetosti i interes za komunikacijom (J. Lotman, 1998, str. 10).
- U slučaju kada ne bi bilo presijecanja (zbog različitih jezika, ili nepostojanja zajedničkog iskustva, tj. pamćenja) komunikacija bi bila nemoguća. U slučaju potpune identičnosti iskustava pošiljatelja i primatelja komunikacija bi bila bez sadržaja.
- Formalni jezici ne podrazumijevaju povijest, oni su "strukture bez pamćenja". Razlika između kôda i jezika jest u tome što kôd ne podrazumijeva povijest, a jezik je kôd plus povijest (J. Lotman, 1998, str. 9-10)[9].
- Jedan od ključnih razloga zašto arhivistika i muzeologija "kasne" s digitalizacijom svoje građe nije samo "tehničke" prirode. Nerješivi je problem što povijesno, socijalno i kulturno pamćenje arhivalija i muzejskih eksponata postaje predmetom obrade "strukture bez pamćenja". Povijest, socijalno i kulturno pamćenje, su balast, smetnja i neprilika u digitalnom svijetu u kojem su nepoznati i u kojem su eliminirani kako "vrijeme i prostor", tako i "tijelo i stvari" (R. Capurro, isto). Zato ne iznenađuje činjenica što u ne-povijesnom digitalnom prostoru najbolju prođu imaju oni arhivski i muzeološki projekti koji su u funkciji industrije turizma, promidžbe i zabave.

Hiperrealnost/cyberspace nova objektivna stvarnost

- Dominacija digitalne stvarnosti i digitalnih posrednika ima za posljedicu "odvajanje riječi od realnosti" (J. Lotman, 1998., str. 155). Posljedica toga je da je čovjek postao ovisan o informacijama. Čovjek više nema neposredno iskustvo o pojavama i činjenicama, ne može neposredno iskusiti realnost, došlo je do vremenskog i prostornog "odvajanje riječi od realnosti", i ljudi svoj život, odluke i ponašanja organiziraju na temelju

[9] I zato umjetni jezici garantiraju visok stupanj identičnosti. Ako pošiljatelj i primatelj koriste isti kôd, oba bez pamćenja, komunikacija će biti idealna ali bez sadržaja, jer jedan drugome neće imati što reći. Zapravo komunikacija će biti svedena na niz instrukcija.

onoga što čuju i vide u medijima. "Prikaz realnosti zato postaje njihova realnost, koja se doživljava kao točnija i pouzdanija nego realnost sama" (J. J. Britz). Informacijski objekti postaju realna stvarnost. U krajnjem slučaju posjedovanje obavijesti o digitalnom objektu implicira i posjedovanje tog digitalnog objekta. Tamo gdje nije moguće povući razliku između prikaza objekta i samog objekta stvarnost se može mijenjati i manipulirati pomoću podataka/obavijesti[10]. Digitalna stvarnost jest stvarnost: nije fikcija, niti je "virtualna" stvarnost.

- "Na djelu je i urušavanje stvarnosti u hiperrealizam, u brižno podvostručavanje stvarnosti…" veli Baudrillard[11]. "I sama definicija stvarnosti glasi: ono čemu je moguće dati istovjetnu reprodukciju. Ona je suvremenica znanosti, koja tvrdi da je neki proces moguće u cijelosti reproducirati u određenim okolnostima… Na kraju tog procesa reproduktivnosti, stvarno nije samo ono što može biti reproducirano, nego ono što je uvijek već reproducirano. Hiperrealno.[12]"

Javno znanje: društvena kontrola i nadzor

- Korpus javnoga znanja nije se razvijao prema viziji što su je zagovarali utemeljitelji informacijske znanosti: kao korpus istinitog i objektivnog znanstvenog znanja. Korpus javnog znanja u *cyberspaceu* organizira se na novi način: javno je znanje ono što je stvoreno, razmijenjeno, djelatno *samo* u javnosti – samo ako sudjeluje, ako ima učinka u javnoj sferi. Odnosno, javno znanje stvaraju i razmjenjuju ljudi koji se sukobljavaju i bespoštedno bore za svoje interesa.
- Predrasuda je da masmediji ili internet (kao forma javnog znanja) imaju zadaću istinitog i objektivnog informiranja. Njihova je temeljna zadaća društvena kontrola i nadzor, te prisilna socijalizacija kako bi se masa podčinila vladajućim vrijednostima i vladajućim elitama.
- Javno znanje nije objektivno i istinito znanje.

Umjesto zaključka

Utemeljitelji informacijske znanosti polazili su od postavke da njezino područje bavljenja, poput predmeta drugih znanosti, treba biti objektivno i istinito znanje;

[10] J. J. Blitz navodi da je "intelektualno vlasništvo" dobar primjer kako ne postoji razlika između prikaza objekta i samog objekta, odnosno da je tada moguće povući znak jednakosti između posjedovanja obavijesti i objekta obavijesti. Posjedovanje PIN-a, zaporke ili novca može poslužiti kao primjer da je moguće izvršiti neku operaciju ili financijsku transakciju bez da se stvarno transferira sam novac. Digitalizacija bankarstva omogućava globalizaciju svjetskog gospodarstva (J. J. Britz)

[11] J. Baudrillard, 2001., str. 101.

[12] J. Baudrillard, 2001., str. 104.

i to kako sam predmet istraživanja (tj. informacije i znanje) tako i spoznaje do kojih informacijska znanost dolazi. Takvo mišljenje imalo je dugu tradiciju koja se proteže od antičke filozofije, do humanizma i prosvjetiteljstva. Antički su filozofi gledali na istinu kao na normalno stanje svijesti, a "sve drugo moglo je biti samo poremećenost", odstupanje od norme i nešto nastrano[13]. Humanistički je svjetonazor da čovjek može spoznati nove istine, i da samo pomoću istine može spoznati red u kozmosu, stvoriti jedan smisleni svijet u kojem su istina i dobro ključne vrijednosti.

Utemeljitelji informacijske znanosti baštinici su one tradicije mišljenja u kojoj je središnji ideal bila znanost jer se pomoću nje može uspostaviti jedinstvo istine, istine koja je zajednička, razumljiva i dokaziva svima. Oni su vjerovali u znanost i znanstvenu istinu na tragu Hegela koji je smatrao da je istina mjera svih stvari i da istina mora biti suverena i nezavisna.

Konačno, to je ona tradicija mišljenja koja je uvjerenja da u kozmosu i prirodi postoje univerzalni zakoni koje treba otkriti i utvrditi kako bi se kontrolirale prirodne sile; da se i društvo i povijest (subjektivni svijet) kreću i razvijaju po zakonitostima koje spoznajemo onda kada poznajemo njihov uzrok, a razumijemo onda kada poznajemo njihov razlog. U tradiciji je tog uvjerenja da se na tako spoznatoj istini temelji moralni napredak, pravednost institucija i sreća ljudi.

Danas je takvo mišljenje zastarjelo: metafizičko poimanje istine ("bitak je istina" – "verum est ens") doživjelo je slom početkom novoga vijeka kada novovjeki duh postulira maksimu "Verbum quia factum": istinito je i spoznatljivo samo ono što smo sami učinili. Tu maksimu, koja je upućivala čovjeka na povijest kao mjesto istine, od 19. stoljeća sve više potiskuje mišljenje tehnike, kada "dolazi do primata onoga što se može učiniti pred onim što je učinjeno"[14].

Postmoderno mišljenje navodi na novo poimanje istine. Pod konac 20. i početkom 21. stoljeća posebno je za informacijsku znanost važno razumjeti novu prirodu istine koju treba otkriti i razumjeti, jer čovjek u svijetu globalizacije gradi svoju egzistenciju u novom, digitalnom okruženju. Ta je egzistencija neminovnost, ali se gradi na dvojbama i sumnjama, jer nema čvrstog kriterija za novo određenje istine. Ako ne znamo što je istina, kako možemo znati što je dobro i što je pravedno u tom novom digitalnom svijetu?

Za sada odgovore možemo tražiti na tragu izloženih fragmenata i polazeći od negativnih određenja. Digitalni svijet je realnost a ne pitanje izbora. Digitalni svijet nije u vlasti čovjeka, onoga što on čini, kao što je to bilo u slučaju tehnike i tehničkog mišljenja.

Digitalni svijet jest projekt, simulacija i reprodukcija. Kontrolu nad tim svijetom nema čovjek. Čovjek je tek korisnik/potrošač simulakra koji ima svoju autonomnu logiku. Objektivnost u tom svijetu nema referenciju u nekom "realnom" svijetu. Istina je tek sporedni atribut projekta i simulacije. Istina u digital-

[13] Jeremy Campbell, 2005., str. 73.

[14] Vidi raspravu o filozofskom poimanju istine J. Ratzinger, 1972., str. 34-45.

nom svijetu je planska veličina. Ona ne referira na neki sadržaj, te nije više presudni kriterij vrednovanja digitalnog svijeta.

Počinje radikalna vlast digitalnog svijeta, a čovjek u njemu nema uporišta u kojem bi se osjećao sigurnim i zaštićenim. Na djelu je proces prevrednovanja univerzalnih vrednota po kojima se čovjek ravnao u 20. stoljeću: ljudska prava, sloboda, kultura i demokracija. Čovjek je suočen s radikalno novim vrijednostima globalizacije koje određuju njegovu egzistenciju: tehnologija, tržište, turizam i informacija. Suvremeni čovjek danas živi u tom procjepu: čini se da su globalne vrijednosti ireverzibilne, dok su univerzalne na odlasku (J. Baudrillard, 2003.)

I informacijska je znanost zapela u tom procijepu između univerzalnih i globalnih vrijednosti. Ako bude inzistirala na istini kao univerzalnoj vrijednosti, bit će zastarjela znanost, a ako se opredijeli za globalne vrijednosti neće biti društvena znanost relevantna za moralnu i duhovnu sudbinu čovjeka i društvene zajednice.

Literatura

Arquilla, John; Ronfeldt, David F. ed.. Networks and Netwars: The Future of Terror, Crime, and Militancy. National Defence Research Institute, RAND, 2001

Baudrillard, Jean. Simulacija i zbilja. Zagreb: Naklada Jesenski i Turk, 2001.

Baudrillard, Jean. The Violence of the Global. 2003. URL: http://www.ctheory.net/articles.aspx?id=385 (18. rujna 2007).

Britz, J. J. Access to Information: The ethical challenges of the information era . // In: *The ethics of electronic information in the twenty first century*. Ed. Lester Pourciau, str. 9-29.. West Lafayette: Purdue University Press, 1999.

Campbell, Jeremy. Lažljivčeva priča. Historija neistine, Beograd: Knjižara Krug, 2005.

Capurro, Rafael. Beyond the Digital. // Paper presented at the International Festival for Film Video and Media Symposium (VIPER 99), "Cut & Copy," Lucerne (Switzerland), 29-30 listopada 1999. (http://www.capurro.de/home-eng.html), (18. rujna 2007).

Encyclopaedia Britannica. Social Science. // Encyclopaedia Britannica 2004. Deluxe Edition CD.

Field Manual br. 3-13. Information Operations: Doctrine, Tactics, Techniques, and Procedures. Washington DC: Headquarters Department of the Army. 28 studenog 2003. (http://www.us.army.mil)

Lotman, Jurij M. Kultura i eksplozija. Zagreb: Alfa, 1998.

Ratzinger, Joseph. Uvod u kršćanstvo. Zagreb: Kršćanska sadašnjost, 1972.

Tuđman, Miroslav. Information Warfare and Public Knowledge. *ACIPSS – Journal for Intelligence, Propaganda and Security Studies* 1, br 1/2007a, str. 81-94.

Tuđman, Miroslav. Profesor dr. Božo Težak i razvoj informacijske znanosti. // Ur: Đurđica Težak i sur. *Profesor Božo Težak, lučonoša znanosti.* Zagreb: Hrvatska sveučilišna naklada, 2007b, str. 257-271.

UNESCO. Cultural and Lingvistic Diversity in the Information Society. Paris: Unesco, 2003.

Waltz, Edward. Information Warfare. Principles and Operations. Boston – London: Artech House, 1998.

Webster, Frank. Theories of the information society. New York: Roudledge, 1995.

Whitehead, Maj Yulin; USAF. Information as a Weapon. Reality versus Promises. *Airpower Journal***,** jesen 1997, str. 41-54.

Winters, Jim; Giffin, John. Information Dominance vs. Information Superiority, (1. travnja 1997), http://www.iwar.org.uk./iwar/resources/info-dominance/issue-paper.htm (27. travnja 2007).

Pretraživanje i pronalaženje informacija u virtualnom svijetu

Nenad Prelog
Ministarstvo vanjskih poslova i europskih integracija
Trg N. Š. Zrinskog 7-8
Fakultet političkih znanosti Sveučilišta u Zagrebu
Lepušićeva 6
nprelog@mvpei.hr

Sažetak

U radu su analizirani problemi pretraživanja i pronalaženja informacija u uvjetima nastanka sve većeg broja informacija. Kontrola nad novonastalim znanjem jedan je od uvjeta za daljnji razvoj pa i opstanak čovjeka i civilizacije. Danas se procjenjuje da je ukupna godišnja proizvodnja novih informacija jednaka veličini pola milijuna novih Kongresnih knjižnica, a stopa porasta proizvodnje informacija iznosi oko 30% godišnje. Iako informacijska tehnologija omogućava radikalno brže načine pretraživanja i nove pristupe, količina pronađenih informacija na Internetu često obeshrabruje korisnike za detaljniji uvid u dokumente. Uspoređeni su načini komunikacije, pretraživanja i pronalaženja u realnom i virtualnom svijetu kako bi se ukazalo na probleme s kojima se korisnici suočavaju tragajući za pravodobnim i primjerenim informacijama. U virtualnom prostoru mijenjaju se neki temeljni elementi odnosa između izvora i korisnika, pa se kao kriterij pronalaženja informacija sve više koristi istraživanje korisnikovih informacijskih potreba, a manje sam upit za informacijama.

Ključne riječi: pretraživanje i pronalaženje, kontrola znanja, virtualni svijet, informacijske potrebe

> *Ima jedna prastara priča – netko bi sigurno kazao: legenda – koja kaže da je sve, apsolutno sve zabilježeno (a možda će se i ponoviti?). Legenda kaže da je sve što je ikada postojalo, stajalo na mjestu, micalo se po zemlji, letilo zrakom ili plivalo vodom, moralo djelovati na sve ono što ga okružuje. Na druge/na drugo oko sebe. Mirovalo ili kretalo, bacalo je sjenu na kamen ili stablo u blizini, svojom toplinom ili hladnoćom utjecalo je na elektrokemijske i ine procese u nedalekom živom biću, svojim je mirisom u kratkom trenutku prolaska – barem minimalno – promijenilo postojeći miris predmeta, bića ili pejzaža.*

Dakle, ništa više nije isto nakon nas, bilo bi drugačije da se mi nismo dogodili, ali sve bi u nekoj minimalnoj mjeri bilo različito i bez zrnca pijeska na plaži, galeba na pučini, oblaka koji je nestao dok sam pisao ovaj redak, ili trna divlje ruže nikle uz ogradu.

O utjecaju svega na sve i svakoga se po svoj prilici možemo složiti, no s rekonstrukcijom svega što je ikada postojalo ići će nešto teže. Gdje i kako su zabilježene te informacije koje bi bile potrebne da se sve ponovi? Riječ je o golemom, zaista golemom broju informacija koji bi morao biti zabilježen na submolekularnoj ili još vjerojatnije – na subatomskoj razini. A kako nas uvjerava suvremena znanost (čitaj: postojeći stupanj razvoja znanosti), manje čestice od dijelova atoma (proton, neutron, elektron itd.) uglavnom ne postoje, a kamoli procesori ili depozitoriji informacija koje bi bile pretražljive i pronalažljive.

Kako god pogledali, problem je u informacijama. Teorija kaže da je granice između materije i energije vrlo teško precizno odrediti, a slično je i s granicama između trećeg fundamentalnog fenomena bitnog za postojanje i održanje svih sustava – informacije. Gdje prestaje informacija a počinje materija i obratno često je pitanje koje podsjeća na raspravu o kokoši i jajetu, a danas je sve teže postaviti granicu između proizvoda i usluga, između sadržaja i veza (tj. uputnica prema drugim sadržajima), a da ne pričamo o mnogim teorijama i prognozama koje najozbiljnije tvrde da informacije danas predstavljaju najveći izvor svih vrijednosti, jer su u vremenu u kojem živimo već poodavno postale "gorivo koje pokreće svijet" i temelj mnogim nacionalnih i nadnacionalnih ekonomija. Tako malo pomalo dolazimo do problema bilježenja (pohrane) svih informacija koje su ikada nastale, svih poruka koje su ikada prenesene, svih znakova, glasova, oblika, slijeda bitova i sl. koji nisu bili slučajni, koji su imali neko značenje...

Još je davno rečeno, napisano i ponovno naglašavano da je tijekom povijesti čovjek svoje preživljavanje temeljio na sposobnosti prenošenja pamćenja (semantičkog i proceduralnog) s jedne generacije na druge generacije. Kao i sva druga živa bića, čovjek naravno posjeduje sposobnost generičkog pamćenja, postepenog mijenjanja genetskog koda kao reakcija na promjene u okolini. No taj proces, na kojem se osniva koncept evolucije, je vrlo spor i pozicija ljudske vrste na Zemlji bila bi vrlo upitna da se temeljila samo na tom obliku pamćenja. Najnovija istraživanja potvrđuju da je upravo u genomu zapisana cjelokupna evolucijska povijest organizma[1], no budući da je riječ o zbivanjima koja su se se događala kroz duga povijesna razdoblja, na taj način nije moguće utvrđivati utjecaj konkretnih zbivanja i njihove posljedice na živa bića.

[1] Tomislav Domazet Lošo, Split, 4. rujna 2007. 5. svjetska konferencija forenzičke genetike i molekularne antropologije.

Dakle, čovjek je jedina vrsta koja je postepeno usvojila proces bilježenja, prijenosa i pohrane informacija koje u svih drugih vrsta nestaju fizičkom smrću jedinki. Čovjek je "pronalaskom" govora, kasnije pisma, pa onda različitih načina umnažanja zapisanog (od tiska do fotokopiranja) te konačno sredstava masovnog komuniciranja uspio razviti efikasne načine širenja postojećeg i novog znanja u sadašnjosti te njegovu (relativno trajnu) pohranu namijenjenu budućim generacijama. Uz sve poznate probleme očuvanja integriteta fizičkih nositelja informacija (od polaganog raspada papira ili brisanja magnetskih vrpci pa sve do zastarijevanja operacijskih sustava ili aplikativnih programa potrebnih za prikaz informacija pohranjenih na nekonvencionalnim medijima), jedan se problem nametnuo svojom važnošću, problem koji povezuje i sadašnjost (tj. distribuciju informacija) i budućnost (mogućnost sustavnog uvida u pohranjene informacije), a poznat je svima pod imenom **pretraživanja i pronalaženja** informacija.

Bez mogućnosti smislenog povezivanja informacija u cjelinu i provjere da li je to upravo ono najvažnije što je o tome (do sada) rečeno, napisano ili snimljeno nema niti izgradnje znanja, a u uvjetima goleme proizvodnje informacija to se danas sve većem broju ljudi čini kao uzaludan proces.

Mnogo informacija, malo pohrane

O tome ponešto govori i jedna od često spominjanih procjena proizvodnje novih informacija. Istraživanje provedeno u Školi za informacijski menadžment i sustave na Kalifornijskom sveučilištu u Berkeleyu najsustavnije je do sada procijenilo ukupnu količinu informacija koja se tijekom jedne (2002.) godine proizvede u svijetu[2]. Sve informacije zajedno (bez obzira da li su distribuirane/pohranjene na papiru, filmu, magnetskom ili optičkom mediju) imaju oko 5 exabajta (1 exabajt = 1,000,000,000,000,000,000 bajtova odnosno 10^{18} bajtova), ili plastičnije prikazano riječ je o veličini od pola milijuna novih Kongresnih knjižnica. Od toga je 92% bilo pohranjeno na magnetskim medijima, 0,01% na papiru). Ako to raspodijelimo po svakom pojedincu, (uključujući i djecu, nepismene i sl.), proizvodnja informacija promatrane godine je bila 800 megabajta (približno 10 metara knjiga po glavi). No to su tek pohranjene informacije, ako tome još dodamo one distribuirane elektroničkim kanalima (telefon, radio, TV i Internet) tada dobijemo novih 18 exabajta (98% se odnosi na informacije koje su poslane i primljene putem telefona). Stopa porasta proizvodnje informacija iznosila je 30% godišnje.

O "vrijednosti" odnosno "korisnosti" tih informacija u ovom trenutku ne diskutiramo, no nikada se unaprijed ne može utvrditi da nekada, negdje, iz nekog razloga neka informacija nekome neće biti potrebna.

Bez obzira gdje su i kako proizvedene i fizički pohranjene, zapisane informacije (ili barem podaci o njima) u sve većoj mjeri su dostupne putem Interneta, a već

[2] How Much Information? UC Berkeley's School of Information Management and Systems, 2003, Senior Researchers: Peter Lyman and Hal R. Varian.

neko vrijeme vrijedi ona (prilagođena) izreka: čega nema na Webu, odnosno ono što Google neće ili ne može pronaći, kao da i ne postoji, kao da se nije ni dogodilo. Za one koji žele još preciznije određenje mogli bi dodati i: ono što nije prikazano na prvih nekoliko (3-5) stranica s rezultatima pretraživanja kao da i ne postoji[3]. Kada na ekranu vidite zastrašujuću brojku od 50 ili više milijuna stranica, ili ne daj bože milijardu i više[4] jasno je da nitko neće imati volje (snage, vremena itd.) ići na 899. ili 24355. stranicu (što bi još uvijek bilo manje od 1 promila pronađenih dokumenata). No u pravilu rijetko koji pretraživač nudi na pregled više od tisuću dokumenata. A znamo i zašto.

Kada bi samo na čitanje naslova dokumenata potrošili u prosjeku 7 sekundi (što je malo čak i za one s tečajem brzog čitanja), do 100.000. stranice (tj. do milijuntnog dokumenta) bi trebali više od 8 dana neprekidnog pregledavanja (što bi definitivno bio rekord svih rekorda kada bi netko bio u stanju te učiniti).

Dakle, kontrola nad novostvorenim informacijama više ne postoji, a teško da itko i pokušava pohraniti (i učiniti pronalažljivim, odnosno osigurati kasnije korištenje) sve one informacije koje se (trudeći se održavati svoj neometani eksponencijalni porast) pojavljuju na Internetu. Iako je bilo pokušaja arhiviranja "cijelog Weba", vjerojatno najpoznatiji projekt je "Internet Archive" s arhiviranih 85 milijardi stranica,[5] teško je procijeniti da li je riječ o broju dokumenata, broju *mjesta* (site), stranica u značenju kojem mi upotrebljavamo tu riječ (što bi bitno smanjilo veličinu pohranjenog sadržaja itd. Procjenjuje se da sadrži 1,5-2 petabajta podataka (rastući brzinom od 20 terabajta mjesečno), a sve je pohranjeno na 880 računala. Polazeći od radova nekih autora koji tvrde da je u prosjeku životni vijek trajanja neke stranice na Webu 100 dana (!?), iste stranice se pohranjuju svakih 60 dana, bez obzira da li je i u kojoj mjeri njihov sadržaj promijenjen.

U svakom slučaju riječ je o nekom obliku presjeka ("snimka", kako ga sami autori imenuju) kroz povijest Weba, a ne o čuvanju ukupne, ikada objavljene građe. Stranice nisu dinamične, teško je sagledati kontekst u kojem su originalno bile stavljane na Web, linkovi uglavnom više nisu "živi" jer upućuju na mjesta koja više ne postoje ili se drugačije zovu...

Naravno, sve se mijenja, a Web se mijenja brže od svega drugoga.

[3] Istraživanja pokazuju da većina korisnika rijetko odabire dokumente/sadržaje koji se nalaze dalje od treće stranice, stoga se danas svi trude (poznatim a ne uvijek baš i potpuno legalnim sredstvima) biti na vrhu popisa, među onim sadržajima za koje je najvjerojatnije da će biti pregledani.

[4] Ako u neki pretraživač upišemo na primjer riječ "Google" rezultat će biti 1.160.000.000 dokumenata/stranica.

[5] Internet Archive (http://www.archive.org).

Ostali pokušaji pohrane sadržaja s Weba i ne pokušavaju biti sveobuhvatni, pa su neki posvećeni samo određenom tipu sadržaja ili (vremenski ili zemljopisno) ograničenima izvorima.[6]

Tko traži taj i nađe?

Svjesni navedenog ograničenja, činjenice da je samo manji dio nastalih informacija pohranjen i spreman za eventualno korištenje, vrlo brzo se susrećemo s problemom koji stoji na putu ostvarenja naše želje za uvidom u novonastalo znanje: pretraživanjem.

O problemima pretraživanja napisane su tisuće knjiga i stotine tisuća radova i ako se većina autora barem u nečemu slaže, onda im je temeljna teza nerazumijevanje, odnosno nekompatibilnost jednog sustava (stroja ili čovjeka) kojem je zadatak obrada informacija u svrhu pohrane i pronalaženja i drugog sustava koji želi pronaći traženi (potrebni) sadržaj. U tom procesu korisnik je relativno malo pažnje posvećivao načelima funkcioniranja sustava obrade, posebno ako su bili relativno strogo formalizirani.

"Dvoje ljudi nikada neće odabrati iste ključne riječi u opisu nekog dokumenta, prema tome, ključne riječi koriste samo onima koji jako dobro poznaju primjenu nekog sustava označavanja" pisao je Tim Berners Lee, autor Weba i jezika za njegovo pretraživanje.[7] Danas se realni i virtualni svijet sve više razlikuju, o nekima od tih razlika koje se prvenstveno odnose na pretraživanje i pronalaženje informacija biti će riječi na sljedećim stranicama.

Realni svijet (fizički prostor)

Fizički entiteti su temeljni pojam i oni služe za pohranu i prijenos informacija. Dokumenti su najvažniji nositelji informacija a njihov je broj ograničen.

Za **opis sadržaja** (dokumenta) koriste se ključne riječi iz nekog unaprijed definiranog skupa ili po slobodnom izboru obrađivača te svrstavanje u neko (unaprijed definirano) područje ili disciplinu. Sve se temelji na tekstu, odnosno tekstualnom opisu netekstualnih sadržaja (slika, film, video, muzejski izložak itd.).

Uloga tehnologije uglavnom se svodi na **ubrzavanje postupaka** označavanja i traženja ili automatizaciju dijela (odnosno cijelog) procesa (SDI, automatska izrada indeksa itd.).

Do sadržaja se dolazi nakon **selekcije naslova** (recimo putem ključnih riječi ili svrstavanja u neko područje), **čitanja sažetka** ili drugog oblika opisa sadržaja dokumenta, te **saznavanja lokacije** gdje je dokument pohranjen te načina kako se do njega može doći.

---

[6] Slični su ali ograničeni pokušaji Kongresne knjižnice u Washingtonu (samo djela iz SAD), "Svjetske digitalne knjižnice (http://www.worlddigitallibrary.org/project) ili Gutenberg projekta (http://www.gutenberg.org) koji je među prvima počeo digitalizaciju važnijih djela baštine čovječanstva.

[7] Tim Berners-Lee, HyperText and CERN, CERN, svibanj 1990.

Najvažniji problemi proizlaze iz **nedovoljnog razumijevanja procesa obrade i procesa pretraživanja**, između njih gotovo da i nema komunikacije, interakcije, provjere, evaluacije.
Nema interakcije između onih koji pripremaju dokumente za pretraživanje (obrađivača) i onih koji pretražuju (korisnika), odnosno interakcija postoji samo na osobnoj razini (pomoć u pronalaženju, ili u slučaju djelomično obrađenih dokumenata.
Obuhvaćen je samo **manji dio proizvedenih informacija**, dominiraju formalne publikacije koje su prošle kriterije evaluacije.
Tipične institucije su knjižnica, informacijski centar, informacijsko-dokumentacijska služba. **Tradicija i iskustvo** su argumenti koji se navode u korist postojećih institucija.
Dok su **usluge pretraživanja i pronalaženja dokumenata u pravilu besplatne** (osim nekih vrlo specijaliziranih servisa), dokumenti, fizički nositelji informacija imaju svoju cijenu (koja se dodatno pojačava troškovima distribucije, dostave do korisnika).
Do dokumenta se dolazi sporo; nakon što je pronađen podatak o dokumentu (opis) to još ne osigurava dokument pa dostava konvencionalnog dokumenta može biti i izrazito dugotrajan proces. Dominira način distribucije putem konvencionalnih medija. Dokumenti najčešće sadrže konsolidirane informacije (koje su izdržale određene oblike provjere) što nužno utječe na njihovo kašnjenje.
Znanstvene spoznaje su najraširenija vrsta informacija. To podrazumijeva i nedvosmisleno (jasno navedeno) **autorstvo, potpisani tekstovi**, fotografije i drugi prilozi.

Virtualni svijet

Temeljni pojam su **zapisi na nekonvencionalnom** (najčešće računalno čitljivom) **mediju** i postoje uglavnom samo u tom obliku. Budući da je informacija neiscrpan resurs zbog svog nematerijalnog karaktera može ju dijeliti i upotrebljavati više korisnika, a mogućnost korištenja je neiscrpna. Razmjenom i prijenosom se informacije obogaćuju – i onaj koji daje i onaj koji prima na kraju procesa razmjene imaju više.
Za opis sadržaja (odnosno zabilježene informacije koja često ne odgovara do sada prihvaćenim definicijama dokumenta) koristi se (rjeđe) logika podjele na skupine interesa u nekom od kazala; a češće **slobodno pretraživanje svih riječi** od kojih se sastoji neka (tekstualna) informacija. Dodatni elementi pronalaženja informacije su i uputnice (linkovi) s drugih stranica usmjereni ka toj informaciji, te dosadašnje iskustvo u traženju (i nalaženju) informacija (praćenje informacijskih potreba). U razvoju su sustavi za prepoznavanje slike i videa na temelju sadržaja, a ne njegovog tekstualnog opisa.
Uloga tehnologije je dominantna, tehnike pretraživanja i pronalaženja kompletno se temelje na specifičnim programskim pa i strojnim rješenjima, a upravo

pretraživanje informacija predstavlja najdinamičniji (i najprofitabilniji) **smjer razvoja** informacijsko komunikacijske tehnologije i **Interneta**.
U virtualnom svijetu pristup informacijama najčešće počinje od **pregledavanja/prebiranja** ("browsing") **popisa podataka o sadržajima** koji su po različitim načelima prioriteta (korištenost, prisutnost na drugim stranicama itd.) poredani na ekranu kao rezultat korištenja nekog pretraživača. Gotovo se podrazumijeva da je sadržaj od popisa udaljen tek jedan ili dva "klika" miša, te će biti dostupan u nekom od standardnih oblika prikaza. Drugi po važnosti (i učestalosti) je **direktni pristup putem uputnice** (poveznice, linka) od jednog na drugi sadržaj. Postoji bezbroj mogućnosti, daleko više no u realnom svijetu, kako pronaći neku osobu ili instituciju, informaciju o događaju, sliku davno izgubljenog prijatelja ili najnovije sadržaje o rezultatima kliničkog ispitivanja nekog lijeka. Osim informacija kojih je autor tražena osoba, dobit ćemo i sve one informacije koje govore o toj osobi, a skup informacija potrebnih za pretraživanje daleko je manji (može biti i samo dio neke riječi) no što bi bio neophodan u nekom od kataloga ili kazala realnog svijeta.
Najčešći problemi proizlaze iz (pre)velikog broja novoproizvedenih informacija koje ponekad nadilaze kapacitete čak i najvećih pretraživača (Google svakoga dana svoje "farme računala" pojačava sa stotinjak novih jedinica, po potrošnji električne energije je izjednačen s gradom od pedeset tisuća stanovnika); a istovremeno korisniku nude preveliki broj (ne uvijek potrebnih) pronađenih informacija kao odgovor na njegov upit. Stalno upućivanje na nove, dodatne informacije iz teksta, slike ili video isječka (putem linkova) nerijetko dovodi do **prevelikog širenja temeljne teme interesa**, pa čak i zaborava prvobitnog problema. "Nuđenje" različitih vezanih sadržaja ili usluga u pravilu predstavlja samo oblik oglašavanja i vodi prevelikoj komercijalizaciji cijelog sektora pretraživanja, nikada se ne zna što je informacija a što infotainment, a nekad isticana teza da kvantiteta vodi kvaliteti uglavnom ne stoji, jer provjera pronađenog može postati mučan i neučinkovit proces, konačno, rezultat pretraživanja je prvenstveno uputnica koja (kako pokazuju istraživanja) usprkos povremenih provjera ipak u određenom broju slučajeva nikamo ne vodi jer su dokument ili stranica uklonjeni iz različitih razloga s Weba.
Sve se temelji na neprestanoj interakciji, parola "korisnik je kralj" iz odjela za marketing i prodaju se preselila u područje razvoja tehnologije pretraživanja; **prati se** (čak i previše) **što korisnik traži**; iniciraju se rasprave na forumima, objavljuju se (čak i kritička) mišljenja i komentari čitatelja/gledatelja (npr. o knjizi ili putovanju u neku zemlju), potiču se ankete o svemu i svačemu kako bi se na temelju rezultata "krojili" budući pretraživači itd.
Obuhvaćen je golemi broj (novo)proizvedenih informacija, a istovremeno se održava i golemi broj "starih", ranije na Web postavljenih informacija, Daleko **najveći broj pronalažljivih informacija nije stariji od 10 godina** što – za razliku od informacijskih službi u realnom svijetu – bitno ograničava vremenski doseg pretraživanja/pronalaženja. Iako se ponegdje digitaliziraju i starije publi-

kacije (najčešće bez primjerene prilagodbe Webu) to je češće iznimka no pravilo jer ne jenjava nego se neprestano pojačava pritisak novih informacija.Više ne dominiraju formalne publikacije, već najrazličitije informacije (poslovne, obrazovne, zabavne, sportske, servisne itd.) od kojih su neke posebno pripremane za Web dok su druge samo preslikane iz nekog od tradicionalnih medija.

Iako virtualni svijet tek stvara svoje institucije, one se po svemu razlikuju od onih na koje smo navikli u realnom svijetu. Stvaratelji sadržaja (novine, časopisi, radio i TV postaje, različite institucije i organizacije itd) istovremeno uz nove informacije čuvaju i one ranije objavljene, dakle preuzimaju ulogu arhiva i knjižnica; **U središtu informacijskog svijeta caruju pretraživači**, njima je sve podređeno, oni vode razvoj sve novijih tehnologija kojima je cilj što bolje upoznati naše načine korištenja informacija kako bi nam u svakom trenutku mogli ponuditi informaciju primjerenu vremenu i prostoru u kojem se nalazimo, našim navikama ili željama. Najnoviji tip institucija su one koje organiziraju društveno umrežavanje, upoznavanje i druženje u virtualnom prostoru, posebno za mladu publiku, a upravo ta mjesta su najposjećenija i bilježe najbrži razvoj.

U većem dijelu svoje (ne baš duge) povijesti Internet se razvijao pod geslom: Sve je besplatno! To je posebno došlo do izražaja u opadanju cijene prijenosa informacija (taj je iznos postao potpuno zanemariv). No s vremenom je ipak došlo do određenih – u pravilu novih – načina zarade od baratanja s informacijama, no i tu nailazimo na jedan obrat. Za razliku od realnog svijeta, u virtualnom **nije naglasak na cijeni informacije, već se veći dio prihoda stječe pretraživanjem izvora informacija**. Za privilegiju "besplatnog" korištenja pretraživača i kasnijeg uvida u sadržaje sve više plaćamo bilo našim vremenom bilo gubitkom privatnosti. Točnost informacija plaćamo svojom privatnošću, što više sustav zna o nama, to će bolje odrediti naše informacijske potrebe. Davno je još bilo rečeno – "Reci mi što čitaš, reći ću ti kakav si, odnosno koje informacije trebaš!"

Brzina, promptnost, neprestano ažuriranje uvjet su opstanka na Webu. Tekst, slika ili video o nekoj osobi, o događaju ili pojavi trenutačno su dostupni kako onima koji su svjedoci događanja, tako i onima koji su udaljeni desetke tisuća kilometara. Prekid programa zbog emitiranja informacije o nekom važnom događanju ("breaking news"), što stvara teškoće u realizaciji radio ili TV programa (a gotovo je nemoguć u tiskanim medijima) nešto je najnormalnije na Webu gdje se nova vijest uvijek nalazi na vrhu stranice.

Utvrđivanje autorstva ili odgovornosti za pojavu nekog sadržaja na Webu jedan je od većih problema komuniciranja u virtualnom svijetu. Većina priloga je nepotpisana, a često je i bez posebne naznake preuzeta iz drugih izvora. Istovremeno **je teško utvrditi podjelu odgovornosti** između onoga koji je postavio ("proizveo") neki sadržaj, onoga koji ga održava na svojim stranicama, davatelja usluga, distributera ("providera") i konačno mjesta s kojeg je pristupljeno informaciji (posebno ako je riječ o školi, knjižnici ili drugoj javnoj instituciji). Stvar komplicira i nepostojanje primjerene legislative koja se u skladu s dosada-

šnjom poviješću odnosila na neki teritorij, dok se sada autor može nalaziti na drugom kraju svijeta od održavatelja stranica i distributera.

Na kraju

Netko bi nakon svih ovih usporedbi mogao pitati – da li sve to ima veze s pretraživanjem i pronalaženjem informacija? O tome se baš i radi, pronalaženje informacija u virtualnom svijetu danas sve više postaje susretno, ono se manje temelji na pretraživanju a više na otkrivanju naših informacijskih potreba. Netko je već jednom rekao da mi danas postojimo u više različitih oblika. Prvo (i mnogi smatraju jedino) djelujemo kao stvarna (fizička) osoba koja je jedinstvena, čvrsto nogama na zemlji i trošeći materiju, energiju i informacije funkcionira u okvirima poimanja (tradicionalnog) realnog svijeta. Drugo je međutim naš "avatar", kopija koju smo htjeli-ne htjeli stvorili koristeći informacije u virtualnom prostoru. Taj naš drugi ja, koji nema nikakve fizičke dimenzije ostavlja tragove, kreće se od sadržaja do sadržaja putem uputnica, snalazi se bolje od nas u savladavanju informacijske eksplozije i povremeno nas obavještava o nečemu što "on" smatra da mi smatramo važnim. Netko bi još možda dodao da postoji i "treći ja", jer je sve veći broj krađa identiteta i lažnog predstavljanja u virtualnom prostoru, no taj problem (za sada) ostavimo po strani.

Treba naučiti suglasno živjeti u oba svijeta, ostvariti pravu vezu s našim "izaslanikom" i tada će biti manje problema a informacija upravo toliko koliko nam je potrebno.

Nismo li to već nekada čuli?

Addressing Digital Memory Preservation Challenges in Europe

Chiara Cirinnà,
Fondazione Rinascimento Digitale
via Bufalini 6, 50125 Florence, Italy
cirinna@rinascimento-digitale.it

Maurizio Lunghi
Fondazione Rinascimento Digitale
via Bufalini 6, 50125 Florence, Italy
lunghi@culturalheritage.it

Summary

Digital preservation is not just a mechanism for assuring bit sequence persistency, but it involves economic, legal and social aspects. Digital preservation is too big an issue for individual institutions or even sectors to address independently. Concerted actions at both national and international level are required.
Apart some isolated initiatives, the first effort carried out jointly by the European Commission and the Member States was the recommendation issued in July 2002 under the auspices of the Council and in cooperation with the National Representatives Group (NRG). After that some projects funded by the European Commission have developed a knowledge background and experience on digital preservation by cultural institutions.
The coordination action DigitalPreservationEurope (DPE), funded under the Sixth Framework Programme, sets out to improve coordination, cooperation and consistency in current activities to secure effective preservation of digital materials. To this end, the project will facilitate pooling of the complementary expertise that exists across the academic research, cultural, public administration, and industry sectors in Europe. Three important projects funded by the European Commission about digital preservation in the cultural sector, namely DPE, CASPAR, PLANETS, are working strongly together, in order to create a unique framework for the users community at a European level.
Future developments are aimed at enlarging the community of users and directed towards the creation of a network of competence centres on digital preservation related issues. In particular, a study about current practices and architectures implemented in competence centres networks worldwide has been produced and a benchmark model open to experts discussion has been proposed as preparation of a network of excellence in Europe.

Keywords: digital preservation of cultural contents, competence centres

Introduction

Electronic resources are a central part of our cultural and intellectual heritage; but this material is at risk. Digital memory needs constant management, using new techniques and processes, to contain such risks as technological obsolescence. Long-term digital resources preservation implicates the use of management procedures since the digital record creation and continues for as long as the digital object needs to be retained. Digital preservation is not just a mechanism for assuring bit sequence persistency, but it involves economic, legal and social aspects.

Digital preservation is too big an issue for individual institutions or even sectors to address independently. Concerted action at both national and international level is required. To this end the Coordination Activity DigitalPreservationEurope DPE[1], works in order to facilitate pooling of the complementary expertise that exists across the academic research, cultural, public administration, and industry sectors in Europe. It aims at fostering collaboration and synergies among many existing national initiatives across the European Research Area.

DPE, lasting three years, has started the 1st of April 2006 and within the ending date it aims at involving all the European Countries in raising awareness, promoting and carrying out research on technical and organisational issues of digital preservation.

DPE's success will help to secure a shared knowledge base of the processes, synergy of activity, systems and techniques needed for the long term management of digital material.

The DPE Vision

DPE objectives aim at the creation of a network of coordination that helps the diffusion of digital preservation issues through coordination, sharing and dissemination of information, experiences and results. In order to accomplish these objectives, the main project activity is directed to the clustering of a wide user community, so as to produce documentations and recommendations, tutorials, training programmes and factual synergies.

DPE has three main objectives, each of which has more sub-goals:

- *To create a coherent platform for proactive co-operation, collaboration, exchange and dissemination of research results and experience in the preservation of digital objects.* In order to reach this objective, several actions have to be implemented. First we consider important to raise awareness on the issues surrounding the curation and preservation of digital objects across the broad spectrum of national and regional cultural and scientific heritage activity in Europe; second it's a priority the elimination of the duplication of effort of research activities by researchers at

[1] DigitalPreservationEurope – DPE, http://www.digitalpreservationeurope.eu.

different institutions, enabling identification, collection, and sharing of knowledge and expertise; third a conduit between the research community and practitioner community has to be created, in order to foster the collaborative approaches to preservation needs; finally further research on digital preservation in key areas have to be stimulated, encouraging the development of standards and developing research agendas.

- *To increase prevalence preservation services and their viability and accountability*. In order to reach this goal, we support the development of a European-wide approach to the audit and certification of digital repositories as an essential stage in creating content management and delivery services and to repository federation. This means examining core issues that will deliver essential guidelines, methods, and tools to enable the construction of preservation functionality within European. A basic step in this direction is to stimulate ICT companies and software developers to incorporate some of the curation and preservation thinking into newer generations of software and therefore to relate the digital preservation research agenda more directly to the development of exploitable product opportunities and to develop links with the industrial sectors.
- *To improve awareness, skills and available resources.* This action is accomplished by examining core issues that will deliver essential guidelines, methods, and tools and by implementing a suite of training seminars based on best practice, identifying where and what further practitioner training and staff development initiatives might be undertaken.

Delivering these objectives will help institutions, business and individuals in Europe to respond to the increase in e-commerce, e-government, and eContent initiatives as they continue their pervasive penetration of European society.

The main DPE actions

The foremost goal of the proposal is to create some coordination mechanisms across Europe for the digital preservation on-going initiatives. DPE brings together public sector, academic, research, industrial and commercial organisations into a focused initiative on the preservation of digital objects. The project will achieve its objectives through pursuing six core activities.

Advocacy and outreach

A critical weakness in digital preservation work in member states is a widely perceived lack of recognition of the existence of a serious problem amongst the general public and key stakeholders. Awareness raising has not spread beyond a handful of professionals in archives, data centres and libraries. For a cause which relies almost entirely on the collaboration of data creators, intermediaries, and public funds this is a crucial barrier to future progress. Advocacy and outreach to systematically influence the attitudes of others, is therefore a key component of digital preservation action plans in member states. A number of

key stakeholders are common to all national efforts and may be best informed and influenced by coordinated advocacy and outreach at a European level.

Training and Continuing Professional Development

While levels of awareness of digital preservation issues have certainly increased in recent years, a significant obstacle remains for institutions in translating their awareness into practical digital preservation programmes. Both management and operational staff will need training and continuous professional development to be equipped with both the skills and the confidence to embark on a structured but graduated course of action which allows institutions to develop digital preservation programmes to suit their specific requirements.

Digital preservation training is a resource intensive activity which needs to draw on existing expertise and professional methods of delivery and validation. A range of bodies across Europe have begun to develop training primarily addressing awareness raising or single topics in workshops.

Co-ordination of EU Repository Activities

Digital repositories are a new and high-profile area: their number is growing rapidly as institutions, research teams (and even individuals) invest in digital content and networked access to it. The initial focus of activity inevitably has been on the process of establishing and populating repositories. Given the experimental and project-funded nature of many institutional and individual small repositories, it is not surprising that very few repositories can be considered to be "doing digital preservation" but for all digital preservation is becoming a growing and pressing issue. The objectives of this work activity are to add value to repositories and permanent collections by sharing information on strategies for the development of repositories in Member States, encouraging the adoption of best practice and innovative approaches to the preservation of their content, fostering the collaboration between repositories which could lead to sharing of resources, examining the possibility of a common European infrastructure to provide a last line of service for local and national repositories. DPE has developed an on line registry of repositories to store and share information regarding repositories[2].

Audit and Certification Services

Data creators, funders and institutions are increasingly recognising that they need access to trusted repositories. Whether they create these themselves or relying on outsourced services, institutions need mechanisms to validate the trusted status of repositories. OCLC and RLG in their Attributes of a Trusted Digital Repository paper have proposed a high level model for the design, de-

[2] http://www.digitalpreservationeurope.eu/repositories

livery, and maintenance of a digital repository. Subsequently RLG and NARA are progressing towards constructing audit and certification requirements for establishing and selecting reliable digital information repositories. Although there is growing awareness of the certifiable characteristics of repositories (e.g. activities, attributes, functions, processes), the mechanisms for audit and the process by which certificates are issued (and revoked) remain to be agreed.
DPE in collaboration with DCC (Digital Curation Centre) has developed the DCC/DPE Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)[3], the latest development in an ongoing international effort to conceive criteria, means and methodologies for audit and certification of digital repositories. The intention throughout its development has been to build upon, extend and complement existing efforts. A key requirement has been to establish a toolkit that contributes toward a single process for repository assessment.

Research and Practitioner Integration
Digital preservation is a complex inter-related set of technical, organisational, social, and legal challenges. Successful approaches to resolving them benefit heavily from inter- disciplinary and multi-professional contributions and rely on innovative practice developed in a few leading competence centres and industry teams. The potential contribution of research to successful solutions and interaction between practice and research are currently under-developed. Research in digital preservation in Europe is patchy, fragmented, and disconnected. Different research groups do not communicate well or with practitioners.
The objectives of this activity are to add value to research and encourage innovative practice through coordinating research collaboration, building bridges between practitioners and researchers, and developing a database of research activity.
DPE is co-ordinating a digital preservation challenge to increase awareness of the preservation problems and research opportunities among younger researchers. The DPE Digital Preservation Challenge has the goal to raise awareness amongst researchers of the aspects of digital preservation. The challenge invites participants to overcome the barriers hindering access to (sets of) digital objects[4].
DPE offers also exchange of researchers to overcome fragmentation and allow experts to visit a host institution of their choice for up to three months. DPE hopes that these DPE Exchanges will propagate knowledge, capacity and innovation as well as foster better cooperation among research institutions and in-

[3] http://www.repositoryaudit.eu

[4] http://www.digitalpreservationeurope.eu/challenge

dustrial partners working on meeting pressing challenges in digital preservation[5].

Community building

Fondazione Rinascimento Digitale[6] is responsible for the task of community building. A detailed community building strategy has been developed in order to create a solid network for spreading the consciousness of digital preservation issues. The first step is the recognition of different level of awareness on the subject:

- some actors are considerably aware of the importance and criticality of digital preservation issues: the DPE network must develop some strategic alliances with these 'competitors' in the digital preservation arena, like other projects, international or professional organisations, cultural institutions;
- some actors are still not aware of the strategic significance of digital preservation: DPE aims at offering to this user community a common, clear, trustable and open point of view and an access to information about digital preservation related-issues through the exploitation of the tools at disposal.

The ultimate purpose of DPE network is focussed at triggering a "virtuous circle", necessary for the exploitation and capitalisation of different expertise and background, deriving both from 'aware' and 'unaware' actors. The strategy of the coordination action in this direction will be realized through the creation of a large consensus.

DPE has identified distinct types of collaboration with its partners:

1. *Cooperation with other projects and international initiatives*. As Coordination Action would like to avoid duplication of effort by individual organisations and different sectors within the Community, in order to ensure that scarce resources, including financial and human, will be used more effectively and efficiently.

- International organisations and institutional partners have specific tasks and advantages deriving from the collaboration with DPE: some possible terms of cooperation are patronage for events or joint organisation, shared communication channels like newsletters, publications, web-tools and cross-reference;
- Project partners should contribute through the dissemination of good practices, the promotion of recommendations and guidelines, and establishing expert workgroups. They could assist in the transfer of informa-

[5] http://www.digitalpreservationeurope.eu/exchange

[6] http://www.rinascimento-digitale.it

tion and training materials to small organisations interested in digital memory such as libraries, museums, archives and the commercial sector.

2. *DPE and its Associate Partners*. The problem of preserving and providing of continued access to digital information is faced in varying degrees by government and public administrations, cultural heritage, broadcasting, e-health, e-science, commercial organisations, and even private individuals. However digital preservation is too complex an issue for individual organisations or even sectors to address independently. Efforts to meet these challenges must be coordinated if they are to be successful. To achieve this coordination DPE has developed the opportunity for Institutions to become DPE "Associate Partners" (AP). Institutions with experience in activities related to the DPE objectives, and with visibility at national level can act as national representative for the DPE network. AP benefit from the possibility of sharing experiences with other partners, international visibility, having first access to DPE tools and results, the opportunity to work with DPE to disseminate their expertise to the wider DP community, to join expert working groups to help shape international research and policy agenda and to nominate a representative to attend meetings of the DPE governing board. AP should assist to raise awareness Digital preservation issues and DPE within their own country and disseminate DPE publications and recommendations, where necessary translating them in their mother language.
3. *DPE and its User Community.* The main objective of the DPE project is to raise awareness digital preservation issues to the widest community of users and increase the sharing of experience and experience. To achieve these objectives DPE has created the network user community. The user community is composed of various types of individual users, both content holders, creators and end users, that need information and tools to cope with the issue of preserving the digital content. Users are encouraged to register by completing the online Registration Form via the DPE website[7], and have access to the offered benefits.

Preservation Research Roadmap

Over the past five years there has been much discussion of the kinds of research that are needed in the area of digital preservation, but this has not resulted in a concise and well-developed strategy that represents the views of a broad community.

Future research depends upon the availability of such a view. DPE proposes reviewing existing preservation research agendas, and publishing an interim cross-walk and review of them. This review will be accompanied by a question-

[7] http://www.digitalpreservationeurope.eu/mou

naire and opportunity for public comment. The final outcome will be the release of a Research Roadmap in Preservation.
The Roadmap aims at contributing to the planning of our future R&D in Digital Preservation by means of different actions, namely:

- analysing the state of the art in Digital Preservation research and already existing research agendas on a global level;
- analysing the currently available state of computer science, information and communication technology;
- researching and reflecting the needs and demands from the point of view of the Digital Preservation user communities and their leading experts
- researching and reflecting the needs and demands from the point of view of future markets for technology and service providers, to pave the ground for R&D towards such conceptual, technological and application oriented potential building blocks for advanced solutions supporting Digital Preservation applications in a variety of possible scenarios.

The Market and Technology Trends Analysis
This report focuses on market analysis through the systematic consultation of stakeholders on the benefits (access to joint materials and training, licensing of products, etc), and the scale of individual national contribution or individual institutional contribution, to joint collaboration on a European basis. This report has two main objectives:

- Market analysis based on experience and knowledge of all the contributors and the consultation of main stakeholders on their needs and plans so that the outputs of the DPE project meet their present and future demands.
- Technology trends analysis providing main DPE target groups with information on technological solutions available for digital preservation.

The document provides information and knowledge necessary for understanding which are present and future goals and needs of main DPE target groups in the area of digital preservation and how the DPE outputs can meet them in the best way. Moreover the report analyses the technological solutions, both commercial and open source, applicable and already applied for digital preservation in different institutions belonging to the main DPE target group that meet minimal functional requirements[8].

Competence Centres
The European Commission (EC) is actively working to empower its member states to safeguard their digital knowledge for long-term accessibility and reuse. The EC has set a goal for member states to formalise a strategy for the

[8] http://www.digitalpreservationeurope.eu/publications/#reports

long-term preservation of their digital information by mid-2008. The European Commission is aware that a great deal of support and guidance will be necessary to assist in the creation of these strategies and is eager to ensure that competent facilities are readily available to all member states. As a starting point, the EC organised a workshop to explore organisational models, boundaries and priorities for work, and to explore the potential impact and long-term sustainability of digitisation and digital preservation competence centres. It concluded that 'competence centres should generate equal access to excellence from anywhere – but not duplicating excellence unnecessarily'. DigitalPreservationEurope (DPE), as part of its planned activity, has been reviewing the current international landscape with regard to the availability and provision of digital curation and preservation expertise. In particular DPE produced a study about current practices and architectures implemented in competence centres networks worldwide[9]. We expect that the results of this review as well as being of use to European Commission to help inform the development of existing and new competence centres across the EU, it will also be of value to EU Member States in considering the development of new competence centres and the measurement of performance of existing ones.

As a starting point, DPE has considered some basic questions, like:

- What is competence?
- Why is competence significant in the to content creating and memory institutions?
- What is the right mix of expertise?
- How is credibility established?

DPE considered these questions from a range of viewpoints and examined current examples of competence centres in the cultural heritage sector and beyond. These considerations have helped to define DPE's '7C's' benchmarking model. This benchmarking model enables the comparison of competence centres' overall strengths and weaknesses and illustrates areas where improvements might be made.

Conclusions

DPE is confident that by addressing these areas, the European Commission will help to ensure that sustainable support is made accessible to assist content creators, curators and re-users to effectively manage and care for their digital resources over their entire life-cycle.

[9] http://www.digitalpreservationeurope.eu/publications/#reports

References

RLG, Trusted Digital Repositories: Attributes and Responsibilities. An RLG-OCLC Report, 2002, http://www.rlg.org/legacy/longterm/repositories.pdf.

Annex: Priority Actions and Indicative Timetable, Council Conclusions on the Digitisation and Online Accessibility of Cultural Material, and Digital Preservation, Official Journal of the European Union, 7 December 2006, http://eur-lex.europa.eu/LexUriServ/site/en/oj/2006/c_297/c_29720061207en00010005.pdf

Report from the Workshop on Centres of Competence for Digitisation and Digital Preservation held in Luxembourg on 14 November 2006, ftp://ftp.cordis.europa.eu/pub/ist/docs/digicult/competence-centres_en.pdf, page 3.

DIGITISATION

Archival Records as Cultural Goods

Živana Heđbeli
C. Zuzorić 43, 10000 Zagreb, Croatia
zivana.hedbeli@gmail.com

Summary

Regulations on archives and archival records and on protection of cultural monuments apply to archival records. Archive material, records, documents, letters, manuscripts, films, buildings and other facilities that permanently preserve or display cultural goods are cultural goods. The owner of a cultural good is obliged to enable accessibility to the cultural good to the public. Duties connected with the protection of archive material are performed by the archives. What is archival records mainly and mostly depends on records' creator, or on a context of records' creation. Partially that is why is so difficult to identify at first sight if some record is archival. It seems that the main archival records' nature or essence is not as mush its permanent value as its uniqueness, fact that certain act was created only in as much of copies as necessary to perform a task, and is not designed for dissemination. Information technology causes changes in archives' work. Traditional, classical records become digitally accessible by scanning. On-line accessibility of records enables simultaneous usage of the same records by several users. That was not possible with traditional records. Today's demands require that all the users should be allowed to use and interpret records according to theirs own needs. That is a new idea for archivists, and implies acceptance of users' competence in relations to research field and not enforcement of "right" way how to understand the records. IT have lead at the end of the last century, to creation of the first archival standards: ISAD(G) and ISAAR (CPF). IT influence indirectly leads to acceleration of human rights protection. Greater accessibility of electronic records and standardization of finding aids unavoidably leads archival profession to leave its hermetisam and false self-sufficiency. Archival profession has to learn to respect different opinions and to achieve joint point of views by discussion and dialog.

Key words: archival records, cultural goods, users

Introduction

People needs to preserve proofs of theirs relations, activities, properties and possessions are old as human race. In terms of evidences, information and data

protection archives are old as literacy. Archives are bench mark of a society that remembers, and that wants to remember. Records are cultural goods, historical and scientific resource which main function is consultation. Preserved records pass on our cultural heritage from one generation to another.

Review of basic Croatian regulations

The Constitution of the Republic of Croatia[1] promotes people and citizens' economic and cultural progress and social welfare. Respect of human rights is one of the highest values of the constitutional order of the Republic of Croatia and base for interpretation of the Constitution. The Republic stimulates and assists the development of science, culture and the arts, protects scientific, cultural and artistic goods as spiritual national values. Provision 52 determines that things of special cultural, historical… significance which are specified by law to be of interest to the Republic shall enjoy its special protection. Provisions of the Constitution are in detail determined by laws and other rules and regulations.

Archival, as well as regulations on protection of cultural goods are relevant for archival records.[2] Basic archival regulation is the *Law on archival records and archives*[3]. Law determines that archival records are of interest for the Republic of Croatia and enjoy its special protection. Archival and current records are protected regardless of whose property or possession they are in, respectively who owns them, or are they registered or recorded. The regulations on the protection of cultural monuments are applied to archival records too.

Principal act on protection of cultural goods is the *Law on the protection and preservation of cultural goods*[4]. Cultural goods are of interest to the Republic of Croatia and receive its special protection. Cultural good, regardless of ownership, preventative protection or registration, receives protection. The purpose of the protection of cultural goods are protection and preservation of cultural goods in an undamaged and original state; transfer of cultural goods to future generations; and establishing conditions where a cultural good may serve the needs of individuals and general interest according to its purpose and importance. The owner of a cultural good is obliged to enable accessibility to the cultural good to the public.

Cultural goods are, among others, buildings and other facilities that permanently preserve or display cultural goods and documentation concerning them. Archive material, records, documents, letters, manuscripts, a collection of items

[1] Narodne novine (NN) no. 41/2001 (Official Gazette of the Republic of Croatia, free of charge available on: http://www.nn.hr)

[2] All relevant regulations are free of charge available on the Ministry of Culture web site: http://www.min-kulture.hr

[3] NN 105/1997, 64/2000

[4] NN 69/1999, 151/2003, 157/2003

in museums, galleries, libraries and other institutions, as well as in other legal entities, national and administrative bodies, including those with physical persons, and films are moveable cultural goods.
Duties connected with the protection of archive material are performed by archives within the framework of their activities in accordance with the provisions concerning archive materials and archives. The procedure for establishing the characteristics of cultural goods for archive material is undertaken with a decision concerning the establishment of the characteristics of a cultural good brought by the Croatian State Archives, which obliged to submit the decision to the Ministry of Culture so that it may be entered into the Register of Cultural Goods of the Republic of Croatia. The Register[5] is a public book kept by the Ministry's Directorate for the Protection of Cultural Goods. Each entry in the Register is published in the "Narodne Novine" (Official Gazette). [6]
Regulation on archives' registers[7] define that the Croatian State Archives, as a central and parent state archives, is bound to keep, among others, the Register of archival records that are proclaimed cultural goods. Register contains following data: ordinal number and date of registration, basic data on owner/holder of records, title of archival fonds/collection or fonds/collections' entity, date, classification mark and registry number of decree on establishment of characteristics of a cultural goods. [8]

What is archival records

If we take one document, very few persons including professionals will be able to say *at first sight*: "This is archival records!" Let us take example known to everybody: invitation to this conference. Faculty of Humanities and Social Science organizes these meeting. Invitation is the Faculty's permanent record (because it is created by the Faculty's action and documents its activity), so it is archival record.
On the other hand, the very same invitation, received by, for example, the Municipal Library Samobor, that caused no further Library's actions (the Library has not sent its employee as participant or lecturer…) has no permanent value for the Library (do not document any of its activities), so the invitation can be destroyed after several years.
It is even more difficult when records that are not acts are concerned. Proceedings of this conference will be recognized at first sight, by majority of persons, no mater of their education and knowledge, as book, and they will say that pro-

[5] Regulation of Register of cultural goods of the Republic of Croatia, NN 37/2001

[6] NN 151/2002, 8/2005, 5/2006, 124/2006…

[7] NN 90/2002

[8] Register is not available on the Croatian State Archives web-site.

ceedings should be held by some library. But, few will recognize the proceedings as archival records. I suppose that even the Faculty will preserve the proceeding in its library, not its archives.

The Law on archival records and archives determines that archival records are considered records or documents created by legal or physical persons in pursuance of their activities, being of permanent significance for culture, history and sciences regardless of the place and time of their creation, not depending on the form and medium they have been preserved thereon. Archival records are created by selection of current records. Current records make a whole of records or documents created or received through the function and activity of a particular legal or physical person. Records or documents are in particular acts, charters, subsidiary office and business books, card indexes, maps, plans, drawings, placards, blank forms, photographs, moving images (film and video recordings), sound recordings, microforms, machine readable records, databanks, including programs and tools for using thereof.

Recognition of permanent records is somewhat facilitated by regulations in force.

Regulation on appraisal and procedures of selection and disposal of archives[9] defines, among others, appraisal criteria and creators' of archives categorization. Appraisal is procedure for judging records value and defining retention period, and also way of disposal or each type or each unit of records.

Appraisal criteria are criteria for defining obligation, needs and interests, and wide individual or social usefulness for records' preservation until expiration of the retention period. Appraisal criteria are:

- significance of activity and function of creator,
- legal acts, rules and standards which determinate particularly disposal conditions,
- business purposes and purposes for supervising of business procedures,
- protection interests of individual or corporate persons,
- public interest for information and data in archives,
- evidential value of records and possibility of trusted provide insight into activities of creator in transparent and structured form,
- information value of archives,
- meaning of archives for culture, history and other sciences,
- value of archives in sense of cultural heritage.

Categorization is procedure of classification of records' creators in groups depending on meaning and value of the whole records created through their function and activity. There are three categories of creators:

- First category are creators responsible for defining policies, goals and methodologies for performing activities and functions; also the creators

[9] NN 90/2002

which archives are transparent for recognize methodology, wideness and conditions for particularly business in frame of the same activities …

- Second category are creators on field of particular activity which are responsible for implementation of policy and performing current activities, which wideness and methodology aren't clear through the archives of creators of first category …
- Third category are creators which mean and area of activity are mostly documenting with archives of creators in first and second categories …

Regulation[10] *on criterions for establishing value of movables that has cultural, artistic or historical value*[11] gives further criterion for establishing movables' value:

- item's historical provenience,
- item's cultural, artistic or historical significance,
- item's market value,
- others item's characteristics that makes part of its cultural, artistic or historical value (method of its construction, material and so on).

Regulation on establishing cultural items that are considered as a national treasure of members states of the European Union[12] determines that national treasure are archival materials and any parts thereof, older than 50 years, as well as photographs, films and negatives thereof.

What is archival records mainly and mostly depends on records' creator, or on a context of records' creation. Partially that is why is so difficult to identify at first sight if some record is archival. There is a need to evaluate value, significance, needs… so determining which document is archival one depends on knowledge and competence of person – evaluator. It is not rare that the very same act (records of same creator) various archivists evaluate differently; or that the very same act is evaluated by the very same archivist one year as archival one and after one or two years as records of no value that can be destroyed. That is why regulations determine that evaluation should be preformed by a commission.

If a librarian destroys or not purchase particular book, that is not a loss, because some other library has it, or will buy it. However, it is quite different when archival records are concerned; records that are destroyed are destroyed forever and can not be regained. Finality of selection of archival records from currents ones is crucial and impossible to revoke. It seems that the main archival records' nature or essence is not as mush its permanent value as it uniqueness, fact that certain act was

[10] NN 77/2004

[11] Regulation applies to establishing value of movables seized during the Yugoslav communist regime.

[12] NN 38/2004

created only in as much of copies as necessary to perform a task, and is not designed for dissemination.

Archives' breakthrough[13]

Information technology (IT) causes, at least in the world, changes in archives' work. Traditional, classical records, that make major part of archives' holdings, become digitally accessible to users by scanning. Today documents are created and stored electronically as well. By their concept electronic records are not different from classic ones. "Record is recorded evidence on legal act, written in accordance with adequate form, established to quarantines it's authentic and to give it probatory force".[14] IT has not changed basic essence of documents' creation. Documents are output or by-product of administrative procedures. Basic role of archives is not changed as well – to preserve documents as evidence of activities. Basic task of archivist is not changed either: to enable consultation of records and to give infrastructure that make easier to manage them.

IT changes, at first place, transparency of archivists' performances, and their attitude toward users. Till the end of the XVIII century archives were secret and not available to public. Where archives started to open most of the users came from the narrow, academic community, and had to clarify why and for what purpose want to use the records. Even today user of records held by a Croatian state archives, records that are according to the Law available for consultation, has to give written statement for what purpose he/she will use the records, what can be considered very much as intrusion of user's privacy.[15] Users' opinions on this – is not known officially. Last research of users needs in Croatia was conducted in the year 1962.[16] According to the Statistic Annual for the year 2006[17] 258 employees of archives, of archival profession, have assisted 6.324 users. It should be mentioned that archives are open during working days, averagely from 8 a.m. to 4 p.m. So employed citizens can consulate records, for pri-

[13] See: Charter on the Preservation of the Digital Heritage, http://portal.unesco.org/ci/en/ev.php-URL_ID=13367&URL_DO=DO_TOPIC&URL_SECTION=201.html; Preserving Digital Heritage: Principles and Policies, selected papers of the international conference organized by the Netherlands National Commission for UNESCO, edited by Yola de Lusanet, Vincent Wintermans, http://www.unesco.nl/images/preserving_the_digital_heritage.pdf

[14] Definition of record according to Luciana Duranti in Guercio, M. Načela, metode i instrumenti za stvaranje, zaštitu i korištenje arhivskih zapisa u digitalnom okruženju. // Modernizacija hrvatske uprave. Zagreb : Društveno veleučilište u Zagrebu, 2003., pages 247-278

[15] If libraries or museums would ask theirs users to give written statement on purpose of the records consultation, there would be reactions from users, organizations that protect human rights and part of the public.

[16] See Heđbeli, Ž. Korisnici, arhivi, literatura, Zagovor za istraživanje korisnika arhiva u Hrvatskoj, http://www.filos.com/mkrzak/akt.htm#clanci, 29.3. 2002.

[17] http://www.dzs.hr/Hrv_Eng/ljetopis/2006/28-Binder.pdf

vate purpose, only during their annual or sick leave. When, and if, research of users needs in Croatia would be conduct, it is also necessary to research needs of persons that up to now, because of the archives' working hours, are not in position to use records.

Majority of users satisfy their needs with picture of a document, and do not care if the document is held by an archives, library, museum, private collection or some office. Users expect records to be available on their personal computers, any time and any place. On-line accessibility of records enables simultaneous usage of the same records by several users. That is not possible with traditional records, and represents a challenge for archivists that they, on principle, not have met yet. (Libraries are different, several copies of a book can be used, in a same library, simultaneously by several users.)

On-line documents, available on foreign archives' web sites, users all around the world use anonymously. Today's demands require that all the users should be allowed to use and interpret records according to their own needs. For archivists it is new idea and implies acceptance of users' competence in relations to research field, and not enforcement of "right" way how to understand the records.

IT, and possibilities that it gives regarding exchange and dissemination of information, have lead at the end of the last century to creation of the first archival standards: General International Standard Archival Description ISAD(G), and International Standard Archival Authority Record for Corporate Bodies, Persons, and Families ISAAR (CPF). Archival Code of Ethics is introduces as well. Draft of International Standard on Activities/Functions of Corporate Bodies is published recently. Further standardization of archives' performance and output is expected and unavoidable. Standardization of finding aids and on-line available records imply greater submissiveness of archivists' work to public evaluation, which is more or less (un)wanted challenge for majority of archivists, which they just learn to respond.

In the second half of the XX century IT development and acceleration of information exchange possibilities have indirectly lead to acceleration of human rights protection. Article 19 of the Universal Declaration of Human Rights[18] determines: "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers." Article 27, subsection 1, says: "Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits."

[18] Adopted and proclaimed by General Assembly, resolution 217 A (III) of 10 December 1948, http://www.un.org/Overview/rights.html

Council of Europe framework convention on the value of cultural heritage for society[19] defines that rights relating to cultural heritage are inherent in the right to participate in cultural life, as defined in the article 27 of the Universal Declaration of Human Rights. The Convention's parties undertake to develop the use of digital technology to enhance access to cultural heritage and the benefits which derive from it, and to seek to resolve obstacles to access to information relating to cultural heritage.

According to the Statistic Annual Croatian archives hold 90.132 linear meters of records, 63.966 or 71% of them are arranged. Archivists just have to arrange and describe something less then one third of theirs fonds and collections. Unarranged records are not available for consultation, in principle. No Croatian state archives records are available on-line for a time being.[20]

Conclusions

Standardization of finding aids (product of archivists' work) necessary for exchange of relevant data and for greater accessibility of electronic records unavoidably leads archival profession to leave its hermetisam and false self-sufficiency. Archivists, at least Croatian ones, just have to realize that users are in centre of attention, that main purpose of archives' work is fast, easy and simple accessibility of finding-aids and records, and that basic human rights are freedom of information and freedom of participation in the cultural life. That means completely arranged and described records, accurate and reliable data, much more cooperation with related professions and records' creators, archivists' permanent education, and research of users needs. And, what is maybe most difficult for such a small profession as archival profession is, respect of different opinions and achieving joint point of views by discussion and dialog.

References

Atti della Conferenza Internazionale Scelte e Strategie per la Conservazione della memoria. Dobbiaco (BZ) 25-29 Giugno 2002.

Boras, D., Heđbeli, Ž. Položaj i uloga arhiva i pismohrana u suvremenom društvu. Izlaganja s 2. kongresa hrvatskih arhivista : Arhivi i društvo – izazovi suvremenog doba : Dubrovnik, listopad 2005., Zagreb : Hrvatsko arhivističko društvo, 2005., http://www.had-info.hr/dubrovnik2005/Referat%20Boras-Hedbeli.pdf

Drobnjak, V. O nujnosti vsestranskega odpiranja arhivske stroke. 5. zbornik referatov dopolnilnega izobraževanja s področij arhivistike, dokumentalistike in informatike v Radencih od 5. aprila do 7. aprila 2006. Maribor : Pokrajinski arhiv Maribor, 2006., str. 28-35

[19] Republic of Croatia has signed the Convention on 27.10.2005., NN – Međunarodni ugovori 5/07

[20] See Heđbeli, Ž. Korisnici, arhivi, literatura : Zagovor za istraživanje korisnika arhiva u Hrvatskoj – osam godina poslije. 6. zbornik referatov dopolnilnega izobraževanja s področji arhivistike, dokumentalistike i informatike v Radencih 28.-30. marca 2007 : Tehnični in vsebinski problemi klasičnega in elektronskega arhiviranja, Maribor : Pokrajinski arhiv Maribor, 2007. pp. 350-358

Heđbeli, Ž. Korisnici, arhivi, literatura : Zagovor za istraživanje korisnika arhiva u Hrvatskoj, http://www.filos.com/mkrzak/akt.htm#clanci

Heđbeli, Ž. Korisnici, arhivi, literatura : Zagovor za istraživanje korisnika arhiva u Hrvatskoj – osam godina poslije. 6 zbornik referatov dopolnilnega izobraževanja s področji arhivistike, dokumentalistike i informatike v Radencih 28.-30. marca 2007. : Tehnični in vsebinski problemi klasičnega in elektronskega arhiviranja, Maribor : Pokrajinski arhiv Maribor, 2007., pages 350-358

Heđbeli, Ž. Određivanje arhivskoga gradiva kao kulturne baštine. Atlanti, vol. 14, no. 1-2 (2004), pp. 242-255

Modernizacija hrvatske uprave. Zagreb : Društveno veleučilište u Zagrebu, 2003.

Preserving Digital Heritage: Principles and Policies, selected papers of the international conference organized by the Netherlands National Commission for UNESCO, edited by Yola de Lusanet, Vincent Wintermans, http://www.unesco.nl/images/preserving_the_digital_heritage.pdf

Puhovski, Žarko. Kultura laži?. Večernji list, 7.10.2006., str. 60

Cooperative National Project of Digitisation of Archival, Library and Museum Holdings "Croatian Cultural Heritage"

Jozo Ivanović
Hrvatski državni arhiv
Marulićev trg 21, Zagreb, Croatia
jivanov@arhiv.hr

Goran Zlodi
Faculty of Humanities and Social Science
Ivana Lučića 3, Zagreb, Croatia
gzlodi@ffzg.hr

Summary

The cooperative national project of digitisation of archival, library and museum holdings "Croatian cultural heritage" is geared at promoting and assisting the systematic and standardized approach to the digitisation of material in cultural institutions, forming and offering cultural content and services by way of digital copies and use of information technologies in the protection, processing and use of materials.
The development of comprehensive widely accessible content in digital form is one of the key assumptions for the protection and evaluation of cultural heritage, for networking and presentation of Croatian cultural heritage in European, regional and other cultural networks, for the preservation of cultural diversity and use of cultural contents in education, tourism and other service industries.
The project is aimed at formulating the long-term policy of digitisation and promotion of institutional, technological, professional and organisational capacities and infrastructure considered necessary for developing comprehensive, usable and sustainable cultural contents and services in a digital environment.

Key words: digitisation, cultural heritage, Croatia, portal, collection-level description service, cooperation

Introduction

Modern information and communication technologies generate numerous new challenges and possibilities for preserving and representing cultural heritage. They offer powerful tools for processing, protecting and making accessible cultural contents. More important and far reaching is that they open up new “virtual” communication spaces for spending time, acquiring and exchanging experiences, offering, seeking and encountering contents and services from all fields of human activities.

Activities connected with cultural heritage are markedly directed toward information and communication technologies: they depend on the quality and accessibility of information regarding stored and offered contents and on the possibility of having the largest possible number of users access its contents and services in an easy and simple manner. Today they are among the largest producers and suppliers of publicly accessible digital contents. The use of information and communication technologies and the digitisation of services and of the most valuable cultural heritage stored in archives, libraries and museums is the foundation for developing these activities and institutions engaged in them.

The issue of the digitisation of cultural heritage and introduction of services based on contents created in that manner should therefore not be considered a mere “added value” or only one possible form of activity. It is an activity on which the realisation of basic tasks of these activities essentially depends in the near future so that the lack of quality, long-term and sustainable programmes would in time lead to the ever decreasing presence and usability of cultural heritage and values which individuals and communities take over and create in communication with it.

Institutions that offer their contents in this way can expect an increased number of users and greater visibility of what they have to offer, attracting in that way new categories of users who would otherwise not know what kind of contents they are offering. Institutions which do not develop quality digital contents and services can expect to eventually lose their recognisability and presence among the public and (potential) users.

The development of comprehensive widely accessible content in digital form is one of the key assumptions for the protection and evaluation of cultural heritage, for networking and presentation of Croatian cultural heritage in European, regional and other cultural networks, for the preservation of cultural diversity and use of cultural contents in education, tourism and other service industries.

The development of an information society and information services, particularly in the economic and public services fields has recently been in the foci of numerous legislative, normative and strategic initiatives and activities. Undoubtedly they are gradually creating a new environment for cultural activities too, regardless of whether they are directly oriented and to what degree towards these activities.

The national programme of digitisation of archival, library and museums holdings takes into account the principles, policies and experiences gained in the field of digitisation of cultural heritage in the European Union such as *Lund Principles* [2], *Lund Action Plan* [3], *Charter of Parma* [1] and other relevant international standards and guidelines in digitisation. One of the goals of national programme is to develop the capacities of our institutions to participate in an appropriate way in the cultural heritage digitisation programmes in the European Union.

Digital contents are created relatively easily and in principle are readily accessible and usable in a variety of ways for developing various types of services. They are, however, also considerably more exposed to risk, including the risk of complete loss or loss of usability. With the growth of their volume and use (whether digitally born or digitised materials) the requirement of their long-term preservation and accessibility also becomes more expressed.

The long-term preservation of digital contents, namely, the ability to systematically store them and make them available in the near and more distant future is an indispensable precondition for any meaningful digitisation programme.

Objecitves of national programme of digitisation of archival, library and museum holdings

During the 2005 and 2006 working group of 13 representatives from archive, library and museum communities developed “The national programme of digitisation of archival, library and museum holdings” [5]. The national programme is geared at promoting and assisting the systematic and standardized approach to the digitisation of material in cultural institutions, forming and offering cultural content and services by way of digital copies and use of information technologies in the protection, processing and use of materials. The programme is aimed at formulating the long-term policy of digitisation and promotion of institutional, technological, professional and organisational capacities and infrastructure considered necessary for developing comprehensive, usable and sustainable cultural contents and services in a digital environment. The following are its basic objectives.

1. The complete digitisation of material in the long term leads to the creation of recognizable and relevant contents and “cultural space” in the electronic environment.
2. Services based on these contents are readily accessible to all interested parties while selected contents could be accessed in accordance with usability rules through networking.
3. Produced digital contents are adapted, adequately formed and processed according to adopted norms and good practices which enable their inclusion in European, regional and other cultural contents and services networks.

4. Digital collections are comprehensive and usable in the forming and supply of educational, tourist and other similar service contents.
5. Digital collections are part of the organised management system which offers sufficient guarantees for their permanent preservation and availability.
6. Institutions whose resources are not sufficiently developed shall not be deprived of the possibility to digitise their materials.

Activities

The national programme of digitisation of archival, library and museum holdings is not an all inclusive and closed digitisation project the primary goal of which is to produce specific digital contents. It has been envisaged as a framework for encouraging and developing a greater number of individual and adapted programmes and activities on the digitisation of these materials as part of national cultural heritage. Within this programme the following activities have been foreseen.

1. **Cooperative national project of digitisation of archival, library and museum holdings “Croatian cultural heritage”.** The aim of this project is to create and make available series of digital collections of identifiable and nationally relevant contents and through work on the project strengthen institutional capacities of participating institutions. The project is open to all institutions that store valuable archival, library or museum materials and wish to make it available in digital form. The experience acquired through cooperation on the project will help institutions to promote their capacities and become equipped to offer other digital contents and services. The duration of the project is three years.
2. **Elaboration of standards and guidelines for the digitisation of material.** In order to ensure interoperability, exchangeability and long-term sustainability of the contents created through the digitisation of cultural heritage, it is necessary to adopt and apply uniform and harmonised standards and specifications in regard to the procedure and parameters of digitisation, properties of digital objects and collections and functional properties of the digitised contents management and accessibility systems. Given that there are no broadly and formally accepted and recommended standards and procedures in regard to the digitisation and management of digital collections in Croatia, the national digitisation project cannot rely on existing solutions. Therefore, it must contain activities that will generate the required standards, specifications and guidelines.
3. **Establishment of support and monitoring systems for the digitisation projects.** Many cultural institutions are relatively small and cannot count on developing their own resources for the systematic and appropriate digitisation of material. In many cases that would not even be rational. The existing and predictable professional resources in this field are such

that a certain degree of the unification and harmonisation will be required in order to develop an appropriate cultural heritage digitisation support system upon which the institution can really rely. The basic tasks of this system include: strengthening of institutional capacities for the digitisation of material and introducing appropriate services for users, decreasing the gap in the capacities of institutions to digitise their materials, supporting professional training and application of adopted standards as well as providing the continual evaluation and monitoring of the quality of digital collections to be created.

4. **Establishment of cooperative digital archives.** The digitisation of material creates collections of permanent value which need to be adequately stored, preserved and made continuously accessible. The producers of digital archives should have at their disposal quality digital archives either by creating and maintaining them individually or else by using the services of cooperative digital archives or some outside service provider. In addition, the task of permanent storage is one of the basic tasks of the institutions that collect and store material while the lack of capacities to adequately do so with digital materials too could have permanent long-term consequences in carrying out their basic activity. For these reasons it is necessary to develop and make such a system available, both for the storage of digital collections which will be compiled within the framework of the national digitisation project, as well as a model for the organisation of digital archives wherever they are needed given the quantity and significance of digital collections in various institutions. The establishment of cooperative digital archives shall be based on the existing standards and good practices with the maximum use of the existing infrastructure in institutions that will be included in the project.
5. **Portal “Croatian cultural heritage”.** The portal “Croatian cultural heritage” will be the central networking point that will enable accessibility and browsing through digital collections which will be developed within the national project of digitisation of archival, library and museum holdings. It shall contain digitisation guidelines and instructions as well as other information and thereby offer support not only to this project but also to other projects, institutions and individuals who deal with or wish to deal with digitisation of their holdings. The portal will enable new and unique communication channel for delivery of cultural heritage content from institutions to users. To ensure quality of representation and accessibility of communicated content data model is developed as a core infrastructure segment. Data model enables description of digital collections and related information: institutions, projects, services or products and physical collections. Data model is based, with some exceptions and modifications, on MICHAEL Data Model and MICHAEL-EU Dublin

Core Application Profile [4] for use by data providers to the MICHAEL-EU presentation service. This model will support core functions of portal “Croatian cultural heritage” related to collection, processing, publishing and access to digital collections metadata what would result in better visibility and accessibility of digital as well as physical collections.

Implementation of the programme

The operational plan for implementation of the programme establishes a series of activities, measures and tasks which need to be implemented within specific basic fields and activities included in the programme, terms of reference and responsibilities for their implementation including a time limit.

The initiator and founder of the programme is the Ministry of Culture of the Republic of Croatia. The initiator of the programme considers and establishes the justifiability, adequacy and feasibility of the programme and programme activities as they are proposed, establishes the organisational structure of the programme approves the project plan and supervises the implementation of the project.

The holders of the programme are the Ministry of Culture, National and University Library, Croatian State Archives and Museum Documentation Centre as the key institutions in library, archival and museum activities.

Bearing in mind the activities linked to the implementation of the Programme e-Croatia 2007 which established the holders of specific activities of particular significance for this programme too (especially the development of a broadband multi-service network), the holders of the activities in the part of this programme which relates to the provision of the broadband access for key institutions, in addition to the Ministry of Culture include the Central State Office for e-Croatia and the Ministry of the Sea, Tourism, Transport and Development.

References

Charter of Parma, 2003. http://www.minervaeurope.org/structure/nrg/documents/charterparma031119final-hr.htm (2007-07-17)

Lund Principles, 2001 ftp://ftp.cordis.lu/pub/ist/docs/digicult/lund_principles-en.pdf/ (2007-07-01)

Lund Action Plan, 2001 ftp://ftp.cordis.lu/pub/ist/docs/digicult/lund_action_plan-en.pdf (2007-07-11)

MICHAEL-EU Dublin Core Application Profile. http://www.ukoln.ac.uk/metadata/michael/michael-eu/dcap/2006-04-28/ (2007-07-01)

Nacionalni program digitalizacije arhivske, knjižnične i muzejske građe. Zagreb : Ministarstvo kulture Republike Hrvatske, 2006. http://daz.hr/bastina/NacionalniProgramDigitalizacije.pdf (2007-07-02)

Informatization of the Croatian Archival Service – From the Idea to the Realization of ARHiNET Project

Vlatka Lemić
Hrvatski državni arhiv
Marulićev trg 21, 10 000 Zagreb
vlemic@arhiv.hr

Hrvoje Čabrajić
Avicena
Put Supavla 30, 21 00 Split
hrvoje.cabrajic@st.t-com.hr

Summary
Being a central national archival institution, Croatian State Archives (CSA) takes care for planning of archival activities, coordination of archives' professional work and performing information-documentation service about archival records on national level. In the year 2006, CSA started the construction of new archival information-evidence system which should cover all archival functions: keeping, preserving, arranging and using of archival records.
Technical characteristics of new information system include WEB application with MS SQL server as a basis and C# computer language. Advantages of this solution are building of uniform base and uniform system of data protection with minimal costs. Its main characteristics are availability, reliability, flexibility and exstensivity with general and specific defined functionalities. System is created on modular basis which practically means design and implementation of particular modules as separate projects and their continuous connecting in unique system.
ARHiNET system encloses several modules: Arranging of archival material, Archival documentation, Register books, National archival service' central evidences, Service for archival records outside archives and User service. Realization of this development project of archival service will enable establishment of unique national integrated system of data exchange among the institutions that keep archival material as well as standardization and increasing quality of provision and services in archives.

Key words: archival information system, web application, standardization and rationalization of business processes in archives, design and implementation of new services in archives

ARHiNET – the Idea

The tasks of the Croatian State Archives (CSA) as the central archival institution include planning of archival activity, coordination of the professional work of the archives and managing the archival records information-documentation service on national level. The basic information resource is the Register of archival fonds and collections of the Republic of Croatia which contains data on all archival material in the state and of its creators and owners. After many years of working on collecting, processing and presentation of data concerning archival records and records’ creators and owners, the need for standardization of arranging and description of archival material, archival principles and vocabulary, as well as for unifying of archival work, has emerged.

Subsequently, CSA started the construction of new archival information system which should cover all archival functions: keeping, preserving, arranging and using of archival records. Its goal is to, in the unique data base, make possible standardization and control of records, and in the same time rationalization and standardization of business processes in Croatian archives. The project team, consisted of experts from CSA and the Avicena Company from Split, had set the basic guidelines which the information solution would have to support:

1. the use of available and acceptable technologies,
2. creating the technological preconditions for including all of the owners of the archival records into the unique system,
3. modulation, extensivity and simple upgrading,
4. the simplicity of use.

Considering the technologies that were on the work group's disposal and having in mind the guidelines that were defined by the project team, it was decided that the information system must be defined as a web application. This decision is based on following facts: disposability, simplicity of maintaining and financial aspect.

Archives.Net- the Realization

The project team has decided to start the development in small steps, always considering the entirety of the information solution in its own complexity. That, practically, meant that the single solutions were defined in a relatively short period of time, and that they were presented and implemented after the testing phase. After the analysis of all the state archives activity area, complexity and extremely large and functionally different logical business processes, it was decided to approach to the making of the system using the modulation principles. In practice that meant:

a. to define the basic business processes through the separate modules which will, in a specific moment, be able to function as a unique system
b. to ensure that the modules can be realized as an independent separate sub-projects

c. to ensure maximum flexibility of modules in a sense of adding, alterations and supplements of new functionalities
d. to ensure the possibility of mutual integration of different modules

The same business processes are defined and relating to CSA, regional state archives and other owners of archival records. Having in mind specific quality of single users, the project team has divided the functionalities into two segments: common and specific quality functionalities.

Technical characteristics of ARHiNET system include web application with MS SQL 2005 server as a basis, C# computer language and partly JAVA and all on the dot.net technology. The management of digital contents is also one of system functionalities, and it encompass several business processes: digitalization of archival records, processing of digitalized documents, saving master copies for storage, automatically creation of web copies in JPEG format and presentation of digital content within program solution.

ARHiNET –Project Overview

Diagram 1: Initiation phase

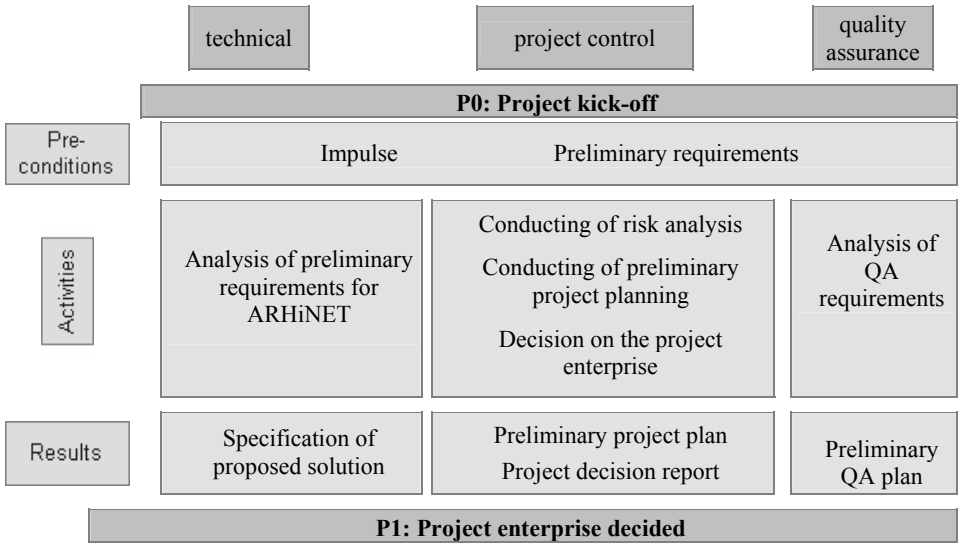


Diagram 2: Definition phase

	technical	project control	quality assurance
Pre-conditions	Preliminary requirements Specification of proposed solution	Project decision report Preliminary project plan Project order	Preliminary QA plan
Activities	**Subphase “Definition of requirements”** Goal: Requirements defined with adequate clarity Definition of the goal Analysis of the domain Acquisition and elaboration of requirements Ordering of the requirements Checking of the requirements Building a domain model Producing a user requirements specification **T21: Requirements defined and checked**	Kick-off activities Risk evaluation Coordinating of user requirements specification with client	Review of user requirements specification
Activities	**Subphase “Definition of the product”** Goal: Product defined with adequate clarity Checking of the requirements Elaboration of product features Building an OOA model Planning of RR Definition of external interfaces Definition of all project processing conditions Making feasibility studies Creating prototype(s) Producing a software requirements specification **T22: Product defined and checked** **Subphase “Drawing up a tender”** Goal: Tender submitted	Drawing up the project agreement Planning the project Planning the CM Setting up the basis CM system Setting up the project infrastructure Coordinating the software requirements specification with the client	Review of the OOA model Review of the software requirements specification Review of project plan Planning the QA measures Review of the QA plan Review of the feasibility studies

Activities	Checking of the requirements Selecting and defining the contractual framework Defining the services Estimating the effort Coordinating price/performance Producing the tender document	Submitting the tender Project checks and control	Review of the tender
	T23: Tender defined, checked and submitted	**P2: Project plan drawn up and checked**	**Q2: QA plan drawn up and checked**
Results	User req. spec. Domain model SW req. spec. Feasibility studies Tender Prototype(s) RR plan OOA model	Project agreement Project plan CM plan Basis CM system Estimation report Open Source SW decision report	QA plan Review reports

Diagram 3: Prototyping phase

	technical	project control	quality assurance
Pre-conditions	SW req. spec. Feasibility studies Tender Prototype(s) RR plan OOA model Domain model Open Source SW decision report	Project order Project plan CM plan Basis CM system	QA plan
Activities	Subphase "Design" Designing the architecture Selecting/defining a prototyping development environment Elaborating/defining design principles and style guides Deciding on reuse of patterns/sample solutions/component ware Evaluation of Open Source components Phased adoption of patterns/component ware Phased creation of an OOD model	Project checks and control Setting up the complete CM system	Revision of the QA plan Drawing up the test plan
Activities	Subphase "Implementation" Phased implementation of a user interface Phased implementation of functions and sequences Phased adoption of design patterns/sample solutions/componentware	Saving checked states of the product in the CM system Defining the preparatory measures for deployment Setting up the test	Design of test cases Review of the test plan Review of solution documentation Ongoing validation of development

Activities	Producing documentation of solution Producing product documentation Performing the system test	infrastructure Providing for acceptance	states Producing the test reports
	T43: System test completed	P4: Product ready for acceptance	
	Subphase “Preparation of operations” Goal: System ready for use		
	Finalizing the product documentation Drawing up the introduction plan Elaborating and conducting user training courses Performing process integration	Performing the acceptance procedure	Review of product documentation Producing the acceptance report
	T44: Preparation of operations completed	P5: Product accepted	Q4: Product checked
Results	Approved product of ARHiNET Documentation of solution for ARHiNET Product documentation for ARHiNET Introduction plan for ARHiNET GUI Styleguide Evaluation report	Project plan CM plan Complete CM system Release note	QA plan Test plan Review reports Test reports Acceptance report

Diagram 4: Design phase

	technical	project control	quality assurance
Pre-conditions	SW req.spec Feasibility studies Tender Prototype(s) RR plan OOA model Domain model Open Source SW decision report	Project order Project plan CM plan Basis CM system	QA plan
Activities	**Subphase “Architectural design”** Goal: Architecture defined		
	Checking (and complementing) the external interfaces Designing the architecture Defining the architecture and product components Building the preliminary OOD model Building the data model Creating prototype(s) Producing the architectural design specification	Project checks and control Planning and commissioning external services Augmenting the CM planning Setting up the complete CM system Defining the HW/SW development tools and programming languages	Definition of design rules and processes Planning and organizing the test procedure Review of the architectural design specification

	T31: Architecture specified and checked		
	Subphase "Detailed design" Goal: Components defined		
Activities	Defining global methods and components Refining the OOD model Designing the individual components Producing the detailed design specification(s)	Planning the product integration Defining the preparatory measures for use	Design of test cases Review of the OOD model Review of the detailed design specification(s) Review of the test plan
	T32: Internal structure specified and checked		
	Subphase "Design when using existing software" Goal: Adaptations specified	Setting up the test infrastructure	Revision of the QA plan
Activities	Evaluation/testing of existing SW Evaluation of Open Source SW components Determining the need for adaptations Producing the adaptations specification(s)	Decision on the use of existing software	Drawing up the evaluation plan Review of the adaptations specification(s)
	T33: Use of existing software specified	**P3: Complete CM system set up and checked**	**Q3: Test plan drawn up and checked**
Results	Architectural design specification Detailed design specification(s) OOD model RR plan Prototype(s) Evaluation report Adaptations specification(s)	Project plan CM plan Complete CM system	QA plan Test plan Review reports Evaluation plan

Diagram 5: Implementation phase

	technical	project control	quality assurance
Pre-conditions	Software requirements specification Architectural design specification Detailed design specification(s) Adaptations specification(s) RR plan Feasibility study (studies) Prototype(s)	Project order Project plan CM plan Complete CM system	QA plan Test plan

	OOD model Open Source SW decision report		
	Subphase “Producing the code” Goal: Developed components implemented	Project checks and control Detailed planning and organization of integration	Revising the QA plan
Activities	Coding the software Performing stand-alone tests Producing the product documentation		Conducting code reviews Finalizing test planning
	T41: Code produced and checked		
	Subphase “Adapting existing SW” Goal: Purchased/RR components adapted	Management of components in the CM system	
Activities	Adaptation of SW and interfaces Stand-alone test of adaptations Adaptation of product documentation		
	T42: Code adapted and checked		
	Subphase “Integration and test” Goal: System tested	Providing for acceptance	Producing the test reports
Activities	Creating tools for integration, testing and installation Phased integration and testing of the system Performing the system test		
	T43: System test completed	**P4: Product ready for acceptance**	
	Subphase “Preparation of operations” Goal: System ready for operations	Performing the acceptance procedure	Review of product documentation Producing the acceptance report
Activities	Completing the product documentation Drawing up the introduction plan Elaborating and conducting user training courses Performing process integration		
	T44: Preparation of operations completed	**P5: Product accepted**	**Q4: Product checked**

Results	Approved product Product documentation Introduction plan	Project plan CM plan Release note RfA note	QA plan Test plan Review reports Test reports Acceptance report

Diagram 6: Operations phase

	technical	project control	quality assurance
Pre-conditions	Software requirements specification Released product Preparation for performing the operations	Project order Project plan Deployed CM system	QA plan Test plan
Activities	**Subphase "Pilot operation"** Goal: Ready for productive operations Preparation of pilot operation Installation of the product	Project checks and control Performing the acceptance procedure	Planning of phase-specific QA measures Recording of metrics data
	T51: Start of pilot operations		Producing the acceptance report
	Support of pilot users		
Activities	**Subphase "Productive operations"** Goal: Stable productive operation Installing and commissioning the product	Releasing maintenance releases	Revising the test plan/ complementing the test data
	T52: Start of productive operations		
Activities	Support of product deployment Analyzing problem reports and change requests Eliminating errors Making approved changes	Deciding on problem reports and change requests **P5: Product accepted**	Performing of regression tests

Results	Deployed product	Release note Project plan	QA plan Quality-related evaluations, metrics data Acceptance report Test plan & Test reports

Diagram 7: Termination phase

	technical	project control	quality assurance
Pre-conditions	Order terminated, completed or negative project decision		
Activities	Archiving of relevant documents and data	Producing a final report	Holding a review of the project Collecting reusable project results Consideration of deviations
Results	Archived documents and data Reusable project results	Project file Final report	Project experience report
P6: Project terminated			

List of abbreviations: CM = Configuration management, HW = Hardware; GUI = General User Interface; SW = Software; OOA = Object-oriented Analysis; OOD = Object-oriented Design; QA = Quality Assurance; Rfa = Ready for acceptance; RR = Round Robin (back up procedure)

Conclusion

The information system of state archives is a dynamic structure which is in a phase of a continued growth and development. The defining of new functionalities and making meaningful the additional modules make this project interesting, dynamic and challenging. ARHiNET enables inclusion of all owners of archival records in Croatia into a unique system which represents a great turning point in the work of state archives and owners of archival records, as well as a long-term developing interest of archival service.

New archival information system contains several modules: Arranging of archival material, Archival documentation, Register books, National archival service’ central evidences, Service for archival records outside archives and User service. Realization of this development project of archival service will enable establishment of unique national integrated system of data exchange among the institutions that keep archival material as well as standardization and increasing quality of provision and services in archives.

Strategies and Approaches to Building Thematic Collections in WebArchiv

Libor Coufal
National Library of the Czech Republic
Klementinum 190, Prague 1, the Czech Republic
libor.coufal@nkp.cz

Petr Žabička
Moravian Library
Kounicova 65a, Brno, the Czech Republic
zabak@mzk.cz

Summary
WebArchiv is a joint web-archiving project run by the National Library of the Czech Republic in cooperation with Moravian Library and the Institute of Computer Science of Masaryk University since the early 2000s. First documents were harvested in September 2001. We attempted five large-scale harvests of the whole national domain since then. The archive contains nearly 6TB of data with over 138.5 million documents to date.
The project deploys three strategies of web archiving: automated large-scale crawls of the Czech national web domain; selective harvesting of high-quality websites selected by curators according to a set of criteria; and thematic collections capturing events of national importance. This paper deals with thematic collections only.
Web resources for inclusion in a thematic collection are currently hand-picked by library staff which is a time-consuming and labour-extensive process. When analysing previous thematic collections we noticed that a large part of selected resources comes from a handful of news sites. This paper discusses some possible strategies and approaches for automating the process of building thematic collections including using RSS feeds, alert services and semi-automated harvesting of selected (news) servers. We also present some preliminary findings from a pilot project which tested some of these approaches.

Key words: web archiving, thematic collections, WebArchiv

Introduction

WebArchiv is a joint web-archiving project run by the National Library of the Czech Republic in cooperation with Moravian Library in Brno and the Institute

of Computer Science of Masaryk University in Brno since the early 2000s. First documents were harvested in September 2001. We attempted five large-scale harvests of the whole national .cz domain since then with various, but progressively improving, degree of success. The archive contains nearly 6TB of data with over 138.5 million documents to date.

The project deploys a combination of three different strategies of web archiving: (1) automated large-scale crawls of the Czech national web domain; (2) selective harvesting of high-quality websites selected by curators according to a set of criteria; and (3) thematic collections capturing events of national importance such as elections, natural disasters, cultural events etc. This paper is concerned with thematic collections only.

Depending on the topic, some collections are once-off while others are continuous or long-term by their nature. So far we have either completed or started seven thematic collections, three of them recently (in 2007). Web resources for inclusion in a thematic collection are currently hand-picked by library staff which is a time-consuming and labour-extensive process. It usually starts with using search engines to discover web pages or whole websites related to the topic of the collection. The curator may also try to directly identify websites covering the given topic. The next step involves following links to other related materials. The whole process is then repeated as necessary throughout the duration of the collection.

This cumbersome task obviously raised a question whether it could be made easier. When analysing previous thematic collections we noticed that a large part of selected resources comes from a handful of news sites. By coincidence, around that time we found out that our colleagues from the Danish netarchive.dk project had reached very similar conclusion and came up with a solution that allowed automating the collection process, at least to a degree. However, as it had some potentially serious drawbacks it was not perfect for what we needed. It was an inspiration, nevertheless. We started to look into the matter and identified some possible solutions to the problem:

1. The netarchive.dk approach – regular, repeated automated harvesting of several selected news servers a few levels down from the homepage in short intervals during the duration of the collection; the number of levels along with the duration of the collection need to be determined. The positive of this approach is that it is fully automated; in fact it is a modification of large-scale crawls. As the servers to be collected and the depth of collection (number of levels) can be determined beforehand, collection can be initiated with a minimum delay as the need arises. A big downside is that a large volume of junk is collected during the process, leaving open a question how to provide access to relevant documents only. Also, relevant resources from other serves not included in the collection set are omitted. This drawback can be partially remedied by including topic-specific servers.

2. Cooperation with publishers – all the news servers we regularly use in thematic collections publish RSS feeds that allow us to subscribe to automatically receive new posts from the whole server or its sections. However, these generic feeds are not suitable for our purposes as they still generate a lot of noise and would require curators to sift through the posts to find collection-related content. It would be much better to be able include search terms (keywords) in the feeds that would limit their scope. This approach requires active participation of publishers as specific feeds would have to be created for each collection. The feeds could be created upon a request by publishers based on specifications provided by Web Archiv or directly by WebArchiv and then sent to the publishers. It would slow down the reaction time at the start of collection but the delay in initiating collection can be minimized. The approach would solve the problem of irrelevant content and how to provide access to relevant sources only but it cannot solve the problem of omitting relevant sources from other servers.
3. Using web-based "alerting" services – this approach uses existing free or fee-based web services that allow users to create alerts or RSS feeds for searches. In a way, it is a modification or extension of the previous approach. Rather than having to create individual feeds for each server, it allows users to create only one search feed or alert and use it for repeated searches over a number of servers at once. Some services actually let users specify a set of servers to be covered. Once the searches are created they are periodically run by the service at predefined intervals and the results are delivered either via email or an RSS feed. Examples of such services are Google Alerts or the "Search the future" feature of Bloglines. We will briefly describe and discuss some of these services in the next section.
4. Building an own application – if none of the existing services is suitable it is possible to build an own application. Luckily, it is not necessary to build it whole from a scratch. Many existing services allow users to build "mash ups" by providing application programming interfaces (APIs), for example Google Code or Yahoo Pipes. However, as some programming may be involved it is advisable that the staff working on the project have at least some basic programming skills. Obviously, the more technically apt the staff is, the more complex applications can be built.

A brief description of some web-based "alerting" services

In this section we will briefly introduce two of the services that we call "alerting" – Bloglines and Google Alerts. We have started to experiment with these two services in our thematic collections but we plan to test some other as well in the near future. There is a number of services of this type available that differ in

the features they offer. Many of them are free but some of them operate on commercial basis and charge a fee.

Bloglines

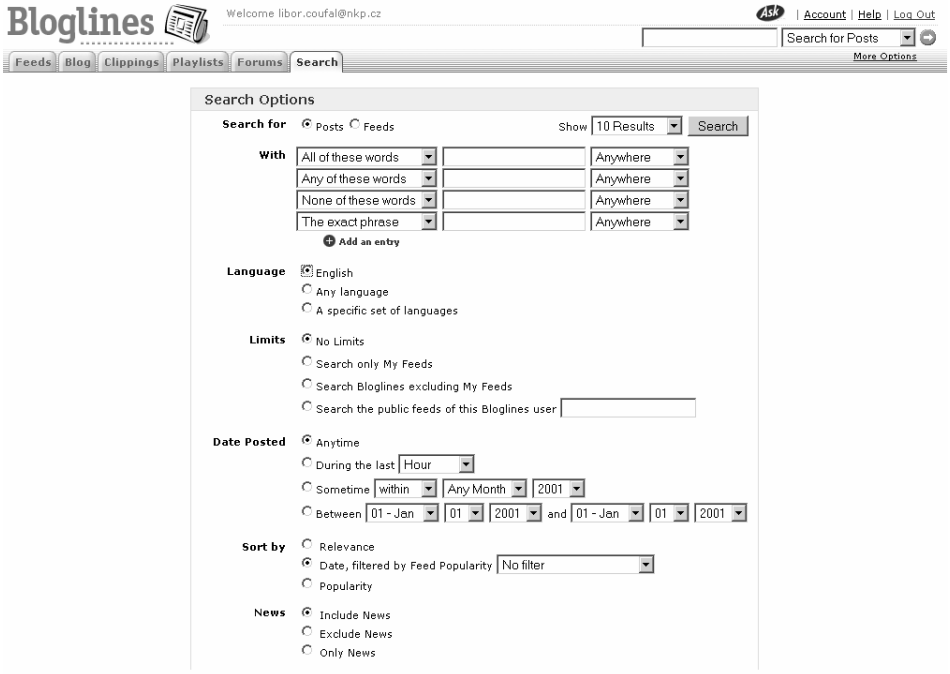
This web-based feed aggregator offers a feature called "Search the future" that lets users to do a search within indexed blogs and then subscribe to an RSS feed that includes these search terms. Anytime new posts appear in the indexed blogs that include these search term, the user gets instantly notified. The feed can be subscribed to in Bloglines or any other RSS aggregator. The search can be further limited to a set of specific subscriptions.

The initial search screen offers only simple "Google style" search box where you can enter keywords with defaulted Boolean AND. However, from the advanced search screen a number of settings can be changed to obtain the best result, including:

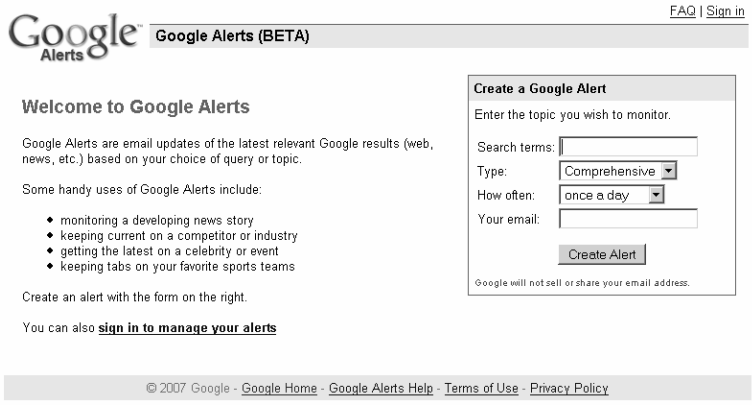
- Searching for posts or feeds
- Boolean logic (AND, OR, NOT); phrase search
- Where search terms should appear (in title, author, URL, body, subject, citation, or anywhere)
- Limiting by language and date posted
- Search in all Bloglines or particular only subscriptions
- Results can be sorted by relevance, date or popularity and limited further by time or inclusion/exclusion of news

Google Alerts

Google lets users to subscribe to "email updates of the latest relevant Google results." Alerts are then send to a chosen email account – which doesn't need to be Gmail – whenever there are new Google results for the given search terms. The users can specify whether they want results from News, Web, Blogs, Groups or all of them (Comprehensive). What exactly "relevant Google results" means varies depending on where the results come from. For example, for a 'Web' alert it means "within the top twenty results of a Google Web search", while for a 'News' alert it is "the top ten results of a Google News search." The periodicity of the search can be set to once a day, once a week or continuously. However, alerts are emailed only when new results are discovered. A user can have up to 1,000 alerts at a time. Alerts can be created without the need to sign up but to manage all alerts conveniently from one place it is necessary to have a Google account. The familiar Google search syntax including "phrase search" and all advanced Google search features can be used. The best way to build a complex search query is to use the general Advanced Search page to generate it and then copy and paste it into the Google Alerts search box.



Picture no. 1 – Bloglines "Search the future" advanced search screen



Picture no. 2 – Google Alerts page

Conclusion

Our work on making thematic collections in WebArchiv less labour extensive is just at the beginning. We started to gather ideas for possible solutions and experiment with some of the approaches and services described in this paper. The results we have got so far are only preliminary and they are influenced by the advanced stages of our currently running thematic collections in which not many new resources appear. We will have to wait for a new thematic collection to do a proper test that will give us a better picture. We plan to try out more “alerting” services, as each of them has a slightly different coverage and it can be therefore expected that the best results will be obtained by using a combination of a few of them. We also plan to test the fully-automated approach of netarchive.dk. When we initiate a new collection we will test a few different strategies and compare them against each other as well as with the results of our curators’ work. It will be also interesting to see if some of these tools could be integrated with existing web archiving tools such as Web Curator Tool or Netarchive Suite.

Architecture for Editing Complex Digital Documents

Tomaž Erjavec
Department of Knowledge Technologies
Jožef Stefan Institute
Jamova cesta 39, Ljubljana, Slovenia
tomaz.erjavec@ijs.si

Summary

In several on-going projects we were faced with the dilemma of how to reconcile our goal of delivering standardly encoded digital editions, yet have the actual editing and annotation performed by researchers and students who had no knowledge of XML and the Text Encoding Initiative Guidelines (TEI), and, for the most part, no great interest in learning them. The developed solution consists of allowing the annotators use familiar and flexible editors, such as Microsoft Word (for structural annotation of documents) and Excel (for word-level linguistic annotation) and automatically converting these into XML/TEI. Given the unconstrained nature of such editors this sounds like a recipe for disaster. But the solution crucially depends on a dedicated Web service, to which the annotators can up-load their files; the service then, via XSLT scripts, converts them to the canonical encoding in XML/TEI, and from it back to a visual format, either HTML or Excel XML. These files are returned to the annotators, giving them immediate feedback about the quality of their encoding in the source, and can thus correct errors before they accumulate. The paper describes the web service and details its use in a project compiling a digital library and lemmatised corpus of XIXth century Slovene books.

Key words: production of digital editions, collaborative work, standards for text encoding, XML, Text Encoding Initiative, Web services

Introduction

A valuable part of digital heritage constitute historically important texts, so called text critical editions, typically comprising the facsimile, and one or more transcriptions, along with commentary. Furthermore, such text can be linguistically annotated, say lemmatised or part-of-speech tagged, for easier access and retrieval. They are thus multimedia editions, with a rich and complex structure, comprising metadata, editorial interventions, extensive cross-linking, and fine-grained annotations.

It is today generally assumed that such scholarly annotated digital texts must be stored in XML to ensure longevity as well as platform and software independence, ensuring the preservation of the extensive intellectual effort that went into their production. It is also accepted by many of the producers of such editions that the XML vocabulary for annotating the editions should be based on the TEI Guidelines (Sperberg-McQueen and Burnard, 2002), which have become the de-facto standard for constructing the XML schemas for a broad range of scholarly digital texts.

But while XML is well-suited for machine processing, it is less than ideal for authorial or editorial interventions into the text, esp. when used with the complex TEI-derived schemas. It is of course possible to edit XML documents directly in a plain text editor or, better yet, in specialised XML editors that support on-the-fly validation against a schema and schema-dependent drop-down menus. But, depending on the text type and required manual interventions, these generic editors are too clumsy for extensive work, and do not enable complex constraints on the allowed content and changes in the annotations. An additional problem with using XML editors for editorial work is the fact that many humanities scholars or students who are most likely to be doing this work have no knowledge of XML or TEI or any experience in editing it. This problem becomes all the more relevant in collaborative projects in which, say, a large number of students are hired to annotate a certain text or text collection. The effort required to first teach them how to use an XML editor and the underlying concepts might be prohibitive, and saving time necessary to perform each editorial intervention is essential.

Such problems of manual intervention can be resolved as they had been before the advent of XML: by developing specialised editing programs for the task at hand which store the data in the required format, and are optimised to perform validity checking and enable fast and easy editing of particular texts types or annotations. But such development is expensive in programming time, (human) editors need to be taught the specifics of the system and last but not least, the program might need to be installed on many different computers or on a Web server for which an uninterrupted internet connection is required.

On the other hand, standard desktop office editors are very versatile, can be easily configured for particular tasks, and most computer users are literate in their usage and already have them installed. In the context of scholarly text editing and annotation, two editors are especially relevant:

- Microsoft Word, the most used text editor, allows easy text authoring (e.g. spell-checking) and editing (e.g. hot keys), definition of complex document structures (sections, tables, notes), has good support for Unicode, allows the inclusion of graphics, etc.; and
- Microsoft Excel, a widely used spreadsheet editor, allows sorting, content-dependent formatting, cell protection, drop-down menus, multiple sheets, etc.

Both editors have applications in scholarly digital texts production: Word is appropriate for text authoring and editing and simple alignments (e.g. between the transcription and its facsimile), while Excel can be usefully employed for linguistic annotation, especially lemmatisation and part-of-speech tagging.
The missing link is, of course, the transformation from the file format of the respective editor to the format defined by a specific (TEI) XML schema. The implementation of such a transformation brings about two problems:

- how to parse complex input files the formats of which are under the control of a software company, hence without the necessarily accessible specifications and with no guarantees on modifications from one version of the software to another; and
- how to reconcile the very free ("visual") structure allowed by the editors to the much stricter ("semantic") XML schema controlled output.

This paper explains how we overcame these two hurdles in the otherwise appealing scenario: humanities editors are free to use tools they are familiar with (Word, Excel), no time investment into project specific (computer) editors is required, yet the final digital edition is stored in a standardised, well-documented and processable format, the TEI XML.

The conversion architecture

The conversion architecture is centred on a Web service which takes as input (possibly a combination of) Word and Excel documents and returns XML and HTML documents. The conversion consists of:

1. parsing the input documents;
2. converting (and merging) them to a TEI document;
3. validating the TEI document;
4. converting the TEI document to HTML;
5. (converting the TEI document to Excel);

Parsing the input

The first hurdle mentioned above, parsing – and making sense of – the vendor-encoded input documents, would have been much more difficult (if not impossible) to overcome some years ago, as we would need to develop or obtain software to parse native editor formats (RTF, Excel) or fall back on the lowest common denominator (plain text, tab separated file). We would also have to deal with significant problems of character set encodings, especially problematic with historical documents. Now, however, many applications offer XML formats for their data. Both Word and Excel (at least in the Office Professional edition) support saving and opening such documents as XML. Encoding problems with well-formed XML do not exits, as its native character set encoding is Unicode. Having XML as the input format allows for formally validating each input file but also allows – even in the absence of public and well-defined formal specifications, i.e. XML schemas – a window into the source format, as

XML documents are easier to reverse-engineer than fully-proprietary encodings. Possible changes in the format between versions are hence also relatively easy to accommodate.

Up-conversion to TEI and validation
The conversion from the proprietary XML to the TEI-compliant XML typically consists of a pipeline of XSLT transformations. The source XML is first converted into a simple TEI and from there into the project-required TEI encoding. The Extensible Stylesheet Language Transformations (XSLT) is an XML-based language used for the transformation of XML documents into other XML or "human-readable" documents, such as HTML.
The formal, “syntactic” validation of the produced TEI document is straightforward, as the TEI document can be checked by any validating XML parser against the project specific TEI schema.

Down-conversion to HTML
The canonical TEI document is then converted, again with XSLT transforms, into a “readable” version of the document. On the on hand, this has to be done for end-users (of digital libraries), but it is also of crucial importance in overcoming the second hurdle mentioned above, i.e. how to reconcile the unconstrained and presentation-oriented nature of documents, especially the ones created in Word, with the strict and interpretative TEI schemas. Namely, the XSLT generated HTML format, esp. with its table of content, various indexes and other cross-references and use of HTML formatting features, such as colour, gives the humanities editors the feedback they need in order to validate whether their source documents are indeed well-formed; only if the structure and annotations are correct in the produced HTML, are they valid in the source. While, as will be discussed below, good guidelines are still needed, the proposed approach also enables self-correction and self-teaching of editors who ultimately produce an exact TEI document.

Down-conversion to Excel
In certain scenarios we also generate Excel (XML) documents which then serve as input to the editing process, and are uploaded to the server after they had been corrected. Such automatically generated Excel documents can be quite sophisticated using a simple trick: a template Excel document is created by hand, with certain cells containing “hooks”, and stored as XML. The conversion then takes this template and replaces the hooks with actual data from the TEI document, duplicating the rows as necessary.

The Web Service
The implemented web service runs under Linux/Apache, using CGI/Perl. The Perl script:

1. takes the uploaded file, possibly compressed, with the archive containing multiple files;
2. calls various transformations with user-selected parameters;
3. returns the result, either directly via HTTP or as an archive file; and
4. logs each transaction, possibly archiving the input and output files.

The presented architecture has thus the following characteristics:

- it enables the editors to work with familiar and powerful tools yet produces TEI conformant output;
- it allows for a gradual learning process and step-wise refinement of the target documents;
- the data is standards-compliant (TEI, XML, (X)HTML, XSLT);
- the software components are Open Source, (Linux. Apache, Perl, libxml); and
- it is not very difficult to modify for new projects.

The AHLib project

So far, the presented web service has been used in several projects / editions. In the initial attempts the Web interface supported only uploading of a Word file and then displayed the derived TEI and HTML files. In this setting, Word is used primarily as an authoring environment. Experience has shown that quite exact guidelines are needed to enable the production of sufficiently constrained Word documents to allow for further processing. This is why editors in later projects have been given short courses, as well as written, quite specific guidelines about what and how to annotate the source document, together with a Word dot file containing the styles used in the project.

In this section we present the Web interface used to compile the AHLib digital library / corpus of XIXth century Slovene books. This is still work in progress, although the Web interface has already been implemented, as well as a "debug" version of the TEI to HTML conversion.

In AHLib each book is represented by the facsimile and a structured diplomatic transcription, hand-corrected from OCR. The text is automatically lemmatised, using the methods described in Erjavec et al. (2005), and then corrected manually. The AHLib Web interface is thus used for correcting two types of errors. The first are errors in the text itself for which each book must be proof-read from the OCR original. At this stage text structure is annotated as well, e.g. headings (divisions), footnotes, figures, page breaks (for alignment with the facsimile), foreign language passages, critical corrections (in case of typos in the original), etc. The second type of errors concerns linguistic annotation. Each word token in the text is first automatically lemmatised and then this lemmatisation is corrected by hand.

The set-up and the intended workflow in this application are rather complicated, mainly due to the fact that there is no simple way of splitting these two annotation types into two separate stages. In particular, checking the lemmatisation often reveals overlooked OCR errors in the text which can only be corrected by going back to the RTF. A further problem is that the automatic lemmatiser (a machine learning program, coupled with a trainable tagger) has been trained on contemporary Slovene which differs considerably from the (non-standardised) language of a century ago. Therefore the lemmatiser consistently makes errors with certain archaic words.

We solved these problems by splitting the process into three stages, allowing for multiple file input and output, and up- or downloading partially corrected files:

1. The user uploads the RTF file and receives either the TEI or HTML; an example of the produced HTML is given in Figure 1. This stage is appropriate for structuring the document and initial proof-reading.
2. The user uploads the RTF or TEI file which is automatically lemmatised and the (structured and) lemmatised version returned as TEI or HTML; an example lemmatised HTML file is given in Figure 1. The lemmas are also furthermore checked against a large contemporary dictionary of Slovene. The unknown lemmas are returned with word-forms and context from the corpus as an Excel table, as illustrated in Figure 2. This table is then manually checked: in case a word is a typo, it is corrected in the RTF file and deleted from the Excel table; if a word is lemmatised incorrectly, its lemmatisation is corrected, and the corrected Excel table is uploaded to the service where it serves as a gold-standard lexicon for the lemmatisation of further texts. It is also possible to perform this process cyclically by submitting the RTF/TEI files and the (partial) Excel table of unknown words together.
3. The user uploads the RTF or TEI documents, and the complete lemmatised text is returned as an Excel table, as illustrated in Figure 3. The user has to check / correct the lemmatisation of each token in this table, and finally submit the RTF/TEI together with the corrected Excel in order to arrive at the final structured and linguistically annotated TEI document. Again, cyclic improvement is possible by submitting the RTF/TEI together with partial Excel. This step is slightly more complicated as Excel imposes an upper limit of 64,000 rows per table, while a book can have more than that number of words. We therefore support the download of multiple Excel files, each containing a portion of the book. Furthermore, the user has the option of retrieving the Excel in the text order or sorted alphabetically.

Conclusions

The paper has presented an environment for manual interventions into XML-based scholarly editions. The basic assumption is that it easier for au-

thors/editors/annotators to use generic and readily available editors than to edit the XML, as well as faster for computer linguists to write or modify XSLT scripts than to develop specialised editors for particular projects. The architecture relies on a Web service that transforms input documents into a standardised format, validates them syntactically, and returns them for semantic validation. The approach was illustrated in a setting in which Word and Excel are used for authoring or correcting the base text and word-level linguistic annotation respectively.

Our experience with the presented approach shows that it is important to give annotators a tutorial and detailed guidelines, and that the approach is mostly appropriate for shallow encodings. For example, trying to unambiguously represent nested annotations in Word (e.g. a correction consisting of a deletion and addition) or cross-references is very difficult, and complex XSLT transformations are needed to transform this information into TEI. We therefore see the usefulness of this approach esp. in collaborative projects with each annotator investing minimal time in training and annotation. Such an annotation scenario is becoming increasingly popular in the HLT community, and wider. So, for example, Mihalcea and Chklovski (2004a, 2004b) describe their "Open Mind Word Expert" site where e.g. student contributors are presented with a set of natural language (e.g., English) sentences that include an instance of the ambiguous word and are asked to indicate the most appropriate meaning with respect to the definitions provided, thus building a word-sense disambiguated corpus. Similarly, Good et al. (2006) report on an experiment where volunteers untrained in knowledge engineering developed a partial ontology via a Web interface. In a non-HLT context, a distributed approach to annotating is used by the Mechanical Turk (http://www.mturk.com/) by Amazon (also dubbed "Artificial Artificial Intelligence") where humans are paid to classify instances of e.g. pictures or texts into predetermined categories. Such tasks are posted to the Turk by companies that need large annotated datasets, typically for training machine learning systems. Both examples above are much more sophisticated than ours in terms of the number of users they allow, but much simpler in the kinds of annotations they envisage – the main difference is that they directly employ a web interface, while our architecture assumes off-line editing in Word or Excel.

As could be noticed, our approach is built around the notion of open standards and software, so a legitimate question is why we have opted to support Microsoft Word and Excel, rather than their open source Open Office (http://www.openoffice.org/) equivalents, OO Writer and OO Calc. The reason is simple: most of the editors and annotators that we worked with have Microsoft Office already installed on their computers, and are reluctant to install the OO suite and learn to use it. Also, in our experience, OO tools still lag behind Microsoft in terms of usability. However, the use of the XML-based OpenDocument standard as the native format for OO applications is a significant ad-

vantage, so we might reconsider our decision in future version of the Web service.

In our further work we also plan to address the question of version control, which is currently lacking in our system, to enable multiple editors to correct a set of documents without the danger of conflicts.

References

Erjavec, T., Ogrin, M. (2005) Digitalisation of literary heritage using open standards. In Paul Cunningham, Miriam Cunningham (eds.). Innovation and knowledge economy: issues, applications, case studies, (Information and communication technologies and the knowledge economy). Amsterdam [etc.]: IOS Press, 999-1006.

Erjavec, T., Ignat, C., Pouliquen, B., Steinberger, R. (2005) Massive multilingual corpus compilation: ACQUIS Communautaire and totale. In Proc. of the Second Language Technology Conference. April 2004, Poznan.

Good, B. M., E. M. Tranfield, P. C. Tan, M. Shehata, G. K. Singhera, J. Gosselink, E. B. Okon, and M. D. Wilkinson. (2006) Fast, Cheap and Out of Control: A Zero Curation Model for Ontology Development Pacific Symposium on Biocomputing 11. http://psb.stanford.edu/psb-online/proceedings/psb06/good.pdf

Mihalcea, R. and T. Chklovski, Teaching Computers: Building Multilingual Linguistic Resources with Volunteer Contributions over the Web, in The LISA Newsletter - Globalization Insider, September 2004. http://www.lisa.org/archive_domain/newsletters/2004/3.3/mihalceaChklovski.html

Mihalcea, R. and T. Chklovski, Building Sense Tagged Corpora with Volunteer Contributions over the Web, book chapter in Current Issues in Linguistic Theory: Recent Advances in Natural Language Processing, Editors Nicolas Nicolov and Ruslan Mitkov, John Benjamins Publishers, 2004.

Sperberg-McQueen, C. M. and L. Burnard, (eds.) (2002). Guidelines for Electronic Text Encoding and Interchange, the XML Version of the TEI Guidelines. The TEI Consortium, http://www.tei-c.org/

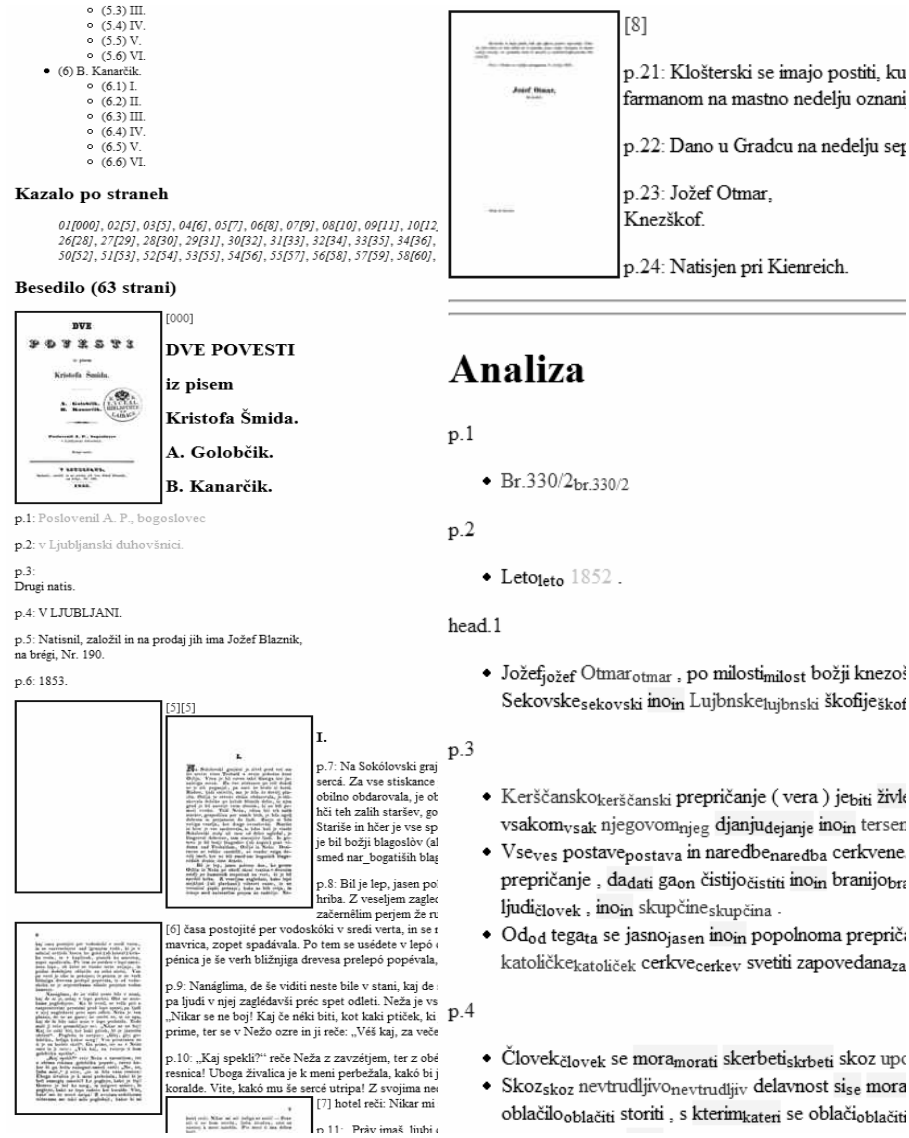


Figure 1. Two screenshots of the HTML view, first one from the beginning of the book, giving the end of table of contents, index by pages, and start of text containing facsimile and the transcription. The second one gives the end of the text, and the start of the linguistic analysis, i.e. the lemmatisation – only lemmas different from the word-form are shown. The colour indicates the status of the correction status of the lemma.

Figure 2. Excel spreadsheet for first round of corrections, giving only the words with out-of-vocabulary lemmas. Column A gives the lexical sort order of the table, B the text sort order, C the word-form, D the lemma (this is the column that annotators correct), E the status of the lemma (here only unknown), and F,G,H the concordance of the word.

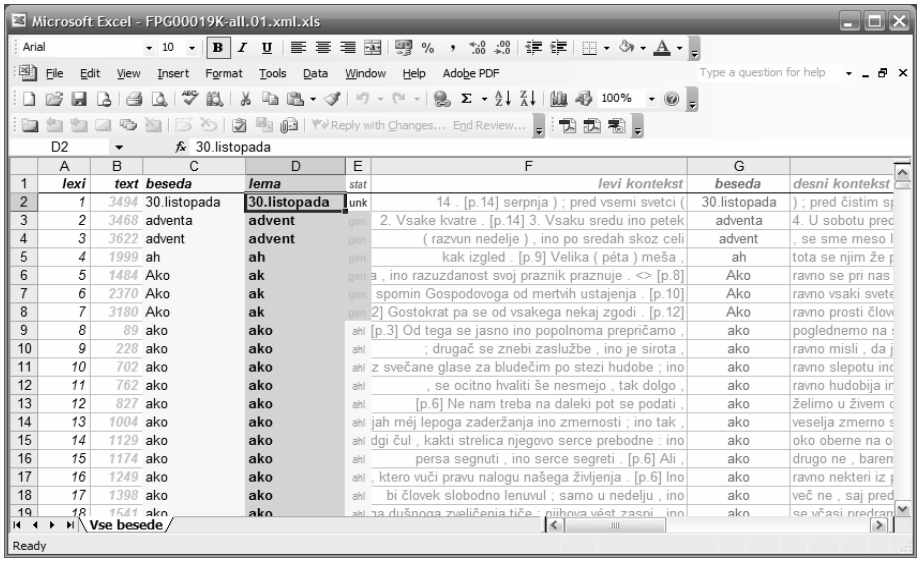


Figure 3. Excel spreadsheet for final corrections, giving the lemmatisation of all word tokens in the text. Here the status (E) of the word can be, in addition to unknown lemmas, also that of general lemmas, lemmas added to the AHLib specific lexicon, and ambiguous lemmas (not shown in the example).

Towards a Digital Edition of the Slovenian Biographical Lexicon

Petra Vide Ogrin
Slovenian Academy of Sciences and Arts, Library
Novi trg 3, Ljubljana, Slovenia
petra.vide@zrc-sazu.si

Tomaž Erjavec
Department of Knowledge Technologies
Jožef Stefan Institute
Jamova cesta 39, Ljubljana, Slovenia
tomaz.erjavec@ijs.si

Summary

The paper presents work-in-progress in the project of the digitization of the Slovenian Biographical Lexicon (SBL). SBL summarizes the lives and work of notable figures from Slovenia's history and is an important reference work for research in the Slovenian humanities and social sciences as well as in the history of the natural sciences. SBL uses the bio-bibliographical method and gives a synthetic assessment of their work and significance, based on the extensive use of primary sources, thus offering support to other encyclopaedic and reference editions. Today especially the first few volumes are almost impossible to obtain while, at the same time, requests for copies of SBL are frequent not only in Slovenia but also from abroad. In order to widen the availability of SBL, the Slovenian Academy of Sciences and Arts (SASA) and the Scientific Research Centre of the SASA have started a project aimed at digitising the SBL and making it freely available on-line. The methodology hinges on the use of open standards and software, in particular the XML-based Text Encoding Initiative Guidelines, which provide a wide, formally specified and well-documented annotation vocabulary for texts and analytical mechanisms. The paper presents the general methodology and steps of the SBL digitization work, concentrating on the TEI elements to be used in the digital edition, in particular: structural mark-up, identification of names and other biographical data, expansion of abbreviations, and cross-linking. The paper gives motivation for our choices and exemplifications from the SBL, and concludes with plans for further work.

Keywords: digital library, encyclopaedias, TEI-XML mark-up

Introduction

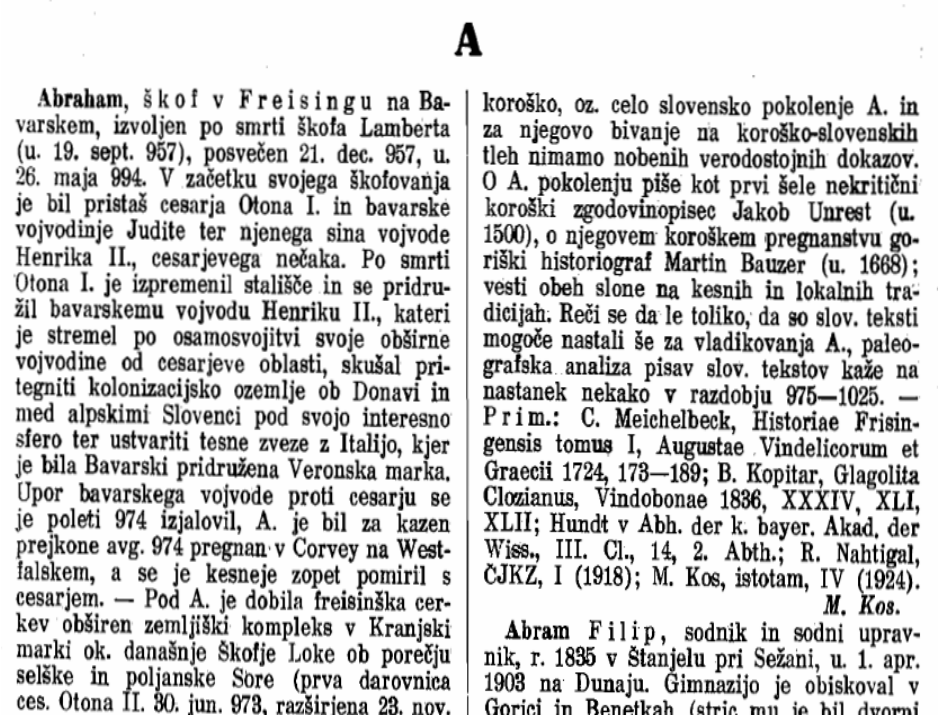
Publication of the Slovenian Bibliographical Lexicon (SBL, 1925-1991) started in 1925 with the appearance of the first of 15 volumes. The introduction to the first volume explains the motivation for such a publication: to give an accurate picture of Slovenia's cultural life, from its beginnings up to the contemporary time by including everybody who participated in the cultural development, is of Slovenian nationality or born on Slovenian ground, and was active in the homeland or abroad, as well as persons of foreign origin who with their work among the Slovenians influenced the Slovenian cultural life. SBL aims to cover not only a person's biography but also the important literature depicting the person's life and work, information where to find their unpublished works or photographs, in short, the SBL aims to be a reliable signpost in orientation and a helpful tool in scientific research.

The first editorial board proposed a list of persons to be included in SBL, which comprised 2,335 names who mostly covered the areas of humanities and social sciences, although the awareness of the necessity to include also figures from other areas was increasingly present. In its rather long history the publication of SBL saw various changes regarding the list of persons to be included. The new list proposed after the WW2 was supposed to "reflect the shifts in the society and a broader concept of national culture that should consider the increasing development of natural sciences, modern technologies and their applications, as part of the spiritual superstructure" (SBL, vol. 15, 1991). In the end, and despite the fact that a few dozens of the names from the original list were eliminated as their significance was examined anew, this led to the enlargement of the original number. So with the last volume having been published in 1991 SBL comprises 5,031 biographical entries and, since some entries are family names, SBL covers more than 5,100 persons, from the beginnings of the Slovenian culture up to the time after WW2.

The primary aim of the SBL was to be informative as well as exhaustive, so a balance had to be found between the length of an article and substantial information. The data in articles are checked against relevant historical materials, e.g. biographical and other dates are always compared to dates in registers and other primary documents, literary citations with originals, etc. SBL also includes an index of all person names that appear in the articles and a list of abbreviations.

In order to widen the availability of SBL, the Slovenian Academy of Sciences and Arts (SASA) and the Scientific Research Centre of the SASA have started a project aimed at digitising the SBL and making it freely available on-line. This paper explains the first steps in this direction, concentrating on the underlying methodology, and the information we plan to explicitly mark in the digital edition. The rest of this paper is structured as follows: Section 2 gives account of the TEI P5 Guidelines and outlines the changes they bring in markup practice, Section 3 gives account of the methodology used in the digitization process of

SBL, and Section 4 sketches the possibilities of the future work on the project of the digital SBL.



A

Abraham, škof v Freisingu na Bavarskem, izvoljen po smrti škofa Lamberta (u. 19. sept. 957), posvečen 21. dec. 957, u. 26. maja 994. V začetku svojega škofovanja je bil pristaš cesarja Otona I. in bavarske vojvodinje Judite ter njenega sina vojvode Henrika II., cesarjevega nečaka. Po smrti Otona I. je izpremenil stališče in se pridružil bavarskemu vojvodu Henriku II., kateri je stremel po osamosvojitvi svoje obširne vojvodine od cesarjeve oblasti, skušal pritegniti kolonizacijsko ozemlje ob Donavi in med alpskimi Slovenci pod svojo interesno sfero ter ustvariti tesne zveze z Italijo, kjer je bila Bavarski pridružena Veronska marka. Upor bavarskega vojvode proti cesarju se je poleti 974 izjalovil, A. je bil za kazen prejkone avg. 974 pregnan v Corvey na Westfalskem, a se je kesneje zopet pomiril s cesarjem. — Pod A. je dobila freisinška cerkev obširen zemljiški kompleks v Kranjski marki ok. današnje Škofje Loke ob porečju selške in poljanske Sore (prva darovnica ces. Otona II. 30. jun. 973, razširiena 23. nov. koroško, oz. celo slovensko pokolenje A. in za njegovo bivanje na koroško-slovenskih tleh nimamo nobenih verodostojnih dokazov. O A. pokolenju piše kot prvi šele nekritični koroški zgodovinopisec Jakob Unrest (u. 1500), o njegovem koroškem pregnanstvu goriški historiograf Martin Bauzer (u. 1668); vesti obeh slone na kesnih in lokalnih tradicijah. Reči se da le toliko, da so slov. teksti mogoče nastali še za vladikovanja A., paleografska analiza pisav slov. tekstov kaže na nastanek nekako v razdobju 975—1025. — Prim.: C. Meichelbeck, Historiae Frisingensis tomus I, Augustae Vindelicorum et Graecii 1724, 173—189; B. Kopitar, Glagolita Clozianus, Vindobonae 1836, XXXIV, XLI, XLII; Hundt v Abh. der k. bayer. Akad. der Wiss., III. Cl., 14, 2. Abth.; R. Nahtigal, ČJKZ, I (1918); M. Kos, istotam, IV (1924).

M. Kos.

Abram Filip, sodnik in sodni upravnik, r. 1835 v Štanjelu pri Sežani, u. 1. apr. 1903 na Dunaju. Gimnazijo je obiskoval v Gorici in Benetkah (stric mu je bil dvorni

Figure 1: An SBL page excerpt

Text Encoding Initiative P5

In 2007, it has been exactly twenty years since the beginning of Text Encoding Initiative (TEI), whose main remit was to produce recommendations or guidelines for the creation and processing of electronic texts for better interchange and integration of scholarly textual data in all languages and from all periods (Burnard, 1988). The first public version of the TEI Guidelines, so called TEI P3 was published in 1994 (Sperberg-McQueen, Burnard, 1994) and was based on the ISO standard SGML. TEI P4, published in 2002 (Sperberg-McQueen, Burnard, 2002) converted the underlying annotation scheme to XML, although it still maintained backward compatibility with P3, and hence SGML. TEI P5 Guidelines, which are to be finalized in 2007, do not maintain compatibility with P4, and are thus free to explore new solutions to text encoding.

For an XML document to be valid (not merely well-formed), its structure must be checked against a XML schema. For a valid TEI document, this schema must be a conformant TEI schema, which is a project-specific combination of TEI modules. Each module contains a set of related elements, typically for use to describe a particular text type or analytical approach.. Typically in building a

TEI schema we decide for a combination of different modules, according to our needs. The TEI infrastructure module, however, is a required component of any TEI schema. It provides declarations for all datatypes, and for the attribute classes, model classes, and macros used by other modules in the TEI scheme.[1] To define a schema in P5, i.e. the required combination of TEI modules and their customizations, the so called ODD language is used.[2].

Apart from the extension and generalization of a modular system one of the main goals of TEI P5 Guidelines is interoperability with current standards, in particular those of ISO and W3C. TEI encoders, for example, are mandated to use Unicode characters. Similarly, the old global attribute lang has been replaced by the W3C-defined attribute xml:lang, which takes as its value the ISO language code. P5 also brings some new elements which are especially relevant for the digital edition of SBL, in particular a structured description of real world entities such as persons and places independent of textual references to them. In this way, "the scope of TEI encoding scheme expands beyond the simple representation of textual structure to include the representation of the knowledge inferred from or implicit in those textual structures." (Burnard, 2007).

Methodology

The methodology of encoding SBL hinges on the use of open standards and software, in particular the adoption of the TEI P5 Guidelines. The preparation of the materials revolves around the up-conversion of the original digital document into TEI-XML, and the down-conversion of this storage format into the XHTML Web presentation language, along with the implementation of digital library open source software. This software should be able to process users' search demands regarding full-text search and allow for complex search and display tasks combining different criteria. In the following sections we detail these aspects of the technology used in preparing the electronic edition of SBL**.**

Preparation of the text

An SBL article typically starts with a name, which gives the variant of the name, used by the person him/herself towards the end of his life or that was generally used, whereas other variants, if they exist, are put in brackets. What follows is a chronological summary of a person's life and activity: information about birth, death, residence, occupation etc. An article is mostly concluded with a brief bibliography that depicts the person's life and work or with locations of other materials relevant to the person's life, e.g. photographs. An article may consist of one or more paragraphs, depending on the exhaustiveness of the

[1] http://www.tei-c.org/release/doc/tei-p5-doc/html/ST.html#STIN

[2] One Document Does it all; see http://www.mulberrytech.com/Extreme/Proceedings/html/2004/Burnard01/EML2004Burnard01.html

article. As is characteristic for encyclopaedic texts, the SBL articles are written in a dense language, containing many abbreviations. Figure 1 shows the opening page of SBL with part of its first article.

In the process of encoding we have to pay special attention to two types of abbreviations: those that are contained in already existing lists, and those which represent the initial letter of a name entry. There are three existing lists of abbreviations with their expansions: a list of bibliographical abbreviations, a list of abbreviated names of SBL contributors and a list of general abbreviations. The lists changed with the publication of each volume in accordance with its content, so, for the purpose of our project, all lists from every volume are combined.

The text is first prepared in a text-editor with the prospect of further conversion into XML. The styles of the original text are retained in order to preserve some basic encoding in the up-conversion, such as for articles and <p> for paragraphs. The text-editor we use is OpenOffice,[3] an open source office tool, which is able to import and export arbitrary XML, if provided with XSLT stylesheets to describe the transformation. TEI provides a simple set of stylesheets and templates to let TEI-XML and OpenOffice work together, in the TEI OO[4] package. So, after correcting the mistakes that still remain after OCR, text is converted into the basic TEI-XML format. At this point we consider which information we want to encode or which "intelligence must be embedded in the text in such a way that the program can derive information from it", (Hockey, 2000, p. 24) The structure of a TEI-XML document takes into account the encyclopaedic nature of the text and with the prospect of what information we want the user of our final database to be able to retrieve and in what way we want it to be presented.

TEI mark-up

The structure of SBL TEI-XML document follows the latest version of TEI P5 Guidelines[5] and especially the module on biographical and prosopographical[6] data. The module defines some special purpose elements which can be used to encode biographical, historical, and prosopographical data. The guidelines propose also a possibility of a more structured entry, for which the information contained in the text must be first extracted and then encoded separately using specific elements. The principal idea is that each article is treated individually

[3] http://www.openoffice.org/

[4] http://www.tei-c.org/Software/teioo/

[5] http://www.tei-c.org/release/doc/tei-p5-doc/html/ND.html#NDPERS

[6] Prosopography: a study that identifies and draws relationships between various characters or people within a specific historical, social, or literary context (Gk. *prosopon* = person, face*; graphein* = (to) draw, write)

as a unit. Essential information about the subject of an article contained in the text is extracted and encoded in a separate structural block at the beginning of the article, in which the <person> element contains detailed further information. Following TEI P5 Guidelines, information about people may comprise a series of statements or assertions that mainly fall into three categories: traits, states and/or events. Traits do not, by and large, change over time, and are encoded in elements such as <sex> or <trait>; states hold true only at a specific time and are encoded in elements such as <occupation> or <floruit>; events or incidents lead to a change of state or, less frequently, trait and are encoded as <event>, <birth> etc. The choice of TEI elements in SBL articles depends on the information contained in a particular article, so the number and the types of elements are to some extent different for each individual article. Nonetheless, we decided for a selection of typical elements that will be important for IR later on by the users, e.g. marriage, ordination, exile, further education, occupation, residence, active period, e.g. as a writer, doctor, judge, bishop etc.

Figure 2 shows the elements used for the first article of SBL. Within the <person> element there are elements indicating sex, date of birth and death (in this case date of birth is not certain), nationality, faith, residence, occupation, active period, occasion of ordination and the role of bishop. It should be noted that the values of attributes are ISO or W3C standard wherever possible (sex, dates etc.). The elements in the <listPerson> are to be (semi-)automatically extracted from the annotation in the article itself. So the basic principle underlying the encoding process of the SBL is first to encode the information in the article with the appropriate elements and then capture it within the <listPerson> element. More on the reason for such doubling of metadata in the section on information retrieval.

The elements also contain linking attributes that join elements conveying the same content or in some other way corresponding to each other. Cross-linking may connect a biographical entry with the occurrences of the name or name variants elsewhere in the entire SBL text and to the entry in the index of personal names. The elements of <occupation> and <floruit> are connected by virtue of adjecency and give information about which occupation a person had in a certain period or give indication of the field of activity in that period. Explicating such information will ideally allow a user to form search demands such as: when did so and so practice law or what kind of interests someone exhibited between 1850 and 1880.

```
<div>
  <listPerson>
    <person xml:id="A.001">
      <persName>Abraham <roleName type="eccl">škof</roleName></persName>
      <sex value="1"/>
      <birth notAfter="0937"/>
      <death when="0994-05-26"/>
      <nationality key="si"/>
      <faith>krščanska</faith>
      <residence notAfter="0974">
        <placeName>
          <settlement>Freising</settlement>
          <region>Bavarska</region>
        </placeName>
      </residence>
      <residence notBefore="0974">
        <placeName>
          <settlement>Otok ob Vrbskem jezeru</settlement>
        </placeName>
      </residence>
      <occupation>duhovnik</occupation>
      <floruit from="0957-12-21" to="0994-05-26"/>
      <event type="ord" when="0957-12-21">
        <label>škof</label>
      </event>
    </person>
    <person>
      <persName>Lambert</persName>
    </person>
  </listPerson>
  <p><persName corresp="#A.001">Abraham</persName>, škof v <placeName>
```

Figure 2: TEI-XML structure of extracted information in <listPerson

Special attention is paid to handling the problem of abbreviations. As mentioned above, there are three types of abbreviations and the comprehensive existing lists give the expanded form or explanation for each abbreviation. We wrote Perl programs to automatically tag the abbreviations in the text with suitable elements, as shown in Figure 3. A program is also used to identify and encode the initials of the subject of an article that occur further in the text of an article contained in <p> element(s). However, it should be noticed that the expansion of the abbreviations is, as yet, preliminary, as they would still need to be correctly inflected.

```
      <expan>slovenski</expan>
    </choice> teksti mogoče nastali še za vladikovanja <persName>
      <choice>
        <abbr corresp="#A.001">A.</abbr>
        <expan>Abraham</expan>
      </choice>
    </persName>, paleografska analiza pisav <choice>
      <abbr>slov.</abbr>
      <expan>slovenski</expan>
    </choice> tekstov kaže na nastanek nekako v razdobju 975-1025. — Prim.: <listBibl>
      <bibl>C. Meichelbeck, Historiae Frisingensis tomus I, Augustae Vindelicorum et Graecii
        1724, 173-189;</bibl>
      <bibl>B. Kopitar, Glagolita Cloesianus, Vindobonae 1836, XXXIV, XLI, XLII;</bibl>
      <bibl>Hundt v Abh. der k. bayer. Akad. der Wiss., III. Cl., 14, 2. Abth.;</bibl>
      <bibl>R. Nahtigal, <choice>
          <abbr>ČJKZ</abbr>
          <expan>Časopis za slovenski jezik, književnost in zgodovino</expan>
        </choice>, I (1918); </bibl>
      <bibl>M. Kos, istotam, IV (1924)</bibl>
    </listBibl>.</p>
  <docAuthor>
    <choice>
      <abbr>M. Kos.</abbr>
      <expan>Dr. Milko Kos, univ. prof. v Lj.</expan>
    </choice>
  </docAuthor>
</div>
```

Figure 3: Annotation of abbreviations

Retrieval of biographical information

The idea of placing essential information in a special structural block within the <listPerson> element preceding the text of an entry organizes the information and so makes it accessible in a well defined way. Information captured in this sort of unified metadata structure is analogous to a TEI Header or a bibliographic record, where data are placed into a variety of predefined fields and then made searchable by those fields. Such a system addresses demanding users, having specifically defined research needs. In the case of SBL a user will be able to find for example female writers, who lived and worked in the period between 1830-1860 in Ljubljana, or priests, who were also philosophers and born in Maribor etc.

Future plans

The project of the digital edition of SBL is still at a very early stage. Our future work will concentrate on various aspects of the SBL digitization. First we will work on the implementation of a system that will allow for the kind of advanced searching described above. One possibility considered is PhiloLogic[7], a full-text search, analysis, and retrieval tool developed by the ARTFL Project[8] and the Digital Library Development Center at the University of Chicago. The system supports boolean and proximity searches and provides a number of reporting functions, including KWICs, and achieves quick display of results by pre-indexing each word and storing byte offsets in a flat file. PhiloLogic handles document structure, currently extending down to seven levels of depth, abstracting a structure by assigning numbers to each nested object-level of each document. It uses an abstract representation of document structure, shredding the XML into sets of related database tables. This means that PhiloLogic can process a TEI-XML encoding scheme by extracting structural information from available text tagging. Storing object-level data in SQL tables, PhiloLogic can search document structure and refine word searching by using the shredded XML. (Cooney et al., 2007)

Apart from working on the implementation and adaptation of the retrieval system as a whole, we will concentrate on semi-automatic tagging of the text. Dedicated filters written in the Perl programming language and XSLT will be able to produce further mark-up in the up-conversion from OpenOffice into TEI. XSLT, which is the XML transformation language, also a recommendation of W3C and hence a standardized specification, is ideal for defining XML structure conversions. It is less suitable though for cases where certain string patterns should give rise to XML structures and in such cases filters are written in Perl (Erjavec, Ogrin, 2005).

Semi-automatic annotation also means exploring the field of language technologies, especially in the extraction of various biographical data with specific consideration of Slovenian linguistic particularities, e.g. inflected forms that need to be normalised, or, as in the case of abbreviations, lemmas that need to be inflected. Regarding the character of data the emphasis will first be probably on the development of a Named Entity Recognition (NER) tool, (Jackson, Moulinier, 2002; Bekavac, 2002) for Slovenian, which will undoubtedly speed up the encoding process.

[7] http://philologic.uchicago.edu/

[8] http://humanities.uchicago.edu/orgs/ARTFL/

Conclusions

The motivation for digitization of the SBL is for the research and other community to be able to access the valuable reference information captured in the lexicon. The digital medium offers searching and retrieval possibilities unthinkable to a classical reader of printed text, especially regarding a text of such encyclopaedic nature as is SBL. Thus, the main idea is to produce a digital edition of SBL with an implementation of a satisfactory retrieval system that will be able to process intelligent searching tasks. To this purpose we have decided for a more complex TEI-XML structure of a biographical entry. That often means (semi-)manual extraction of data that are not explicitly expressed in the text itself. Adding such metadata may mean, in some way and to some extent, creating “new” data, or making implicit information explicit, which represents further editorial interventions. We expect future development of linguistic tools for data extraction to prove helpful in being able to do automatic data extraction, which will speed up the process of encoding.

We believe the implementation of a suitable open source digital library software together with the TEI P5 encoding as outlined in this paper will answer the needs of a demanding user.

References

Bekavac, Božo. Strojno obilježavanje hrvatskih tekstova – stanje i perspektive // *Suvremena lingvistika.* 53-54 (2002), 173-182. http://nl.ijs.si/et/tmp/matija/BB_disertacija_verT9.pdf (2007-09-14)

Burnard, Lou. Encoding standards for the electronic edition // Znanstvene izdaje in elektronski medij: razprave / Ogrin, M. (ed.). Ljubljana: Založba ZRC, ZRC SAZU, 2005, 25-42

Burnard, Lou. Report of Workshop on Text Encoding Guidelines // *Literary & Linguistic Computing*. 3 (1988), 131-133

Cooney, Charles M. et al.. Extending PhiloLogic. April 2007.
http://www.digitalhumanities.org/dh2007/abstracts/xhtml.xq?id=175 (2007-09-09)

Erjavec, Tomaž; Ogrin, Matija. Digital Critical Editions of Slovenian Literature: an Application of Collaborative Work Using Open Standards. // *From Author to Reader: Challenges for the Digital Content Chain: proceedings of the 9th ICCC International Conference on Electronic* Publishing, Arenberg Castle / Dobreva, M.; Engelen, J. (eds.). Leuven: Peeters, 2005, 151-156

Hockey, Susan. Electronic Texts in the Humanities: Principles and Practice. Oxford: University Press, 2000

Jackson, Peter; Moulinier, Isabelle. Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization. Amsterdam: John Benjamins, 2002, 180-185

Cankar, Izidor et al. (eds). Slovenski biografski leksikon. Ljubljana: Slovenska akademija znanosti in umetnosti, 1925-1991

Sperberg-McQueen, C. Michael; Burnard, Lou (eds.). TEI P4: Guidelines for Electronic Text Encoding and Interchange (TEI P4). Text Encoding Initiative Consortium. XML Version. Oxford, 2002. http://www.tei-c.org/Guidelines2/index.xml.ID=P4 (2007-09-09)

Condition of Croatian Music Heritage. Croatian Art Music Sound Recordings

Željka Radovinović
Academy of Music, Library
Berislavićeva 16, Zagreb, Croatia
zradovinovic@muza.hr

Summary

Just over 130 years ago Thomas Edison invented the first practical machine for recording a sound. That event has changed the whole historical and sociological picture of sound-expressing arts – first of all music. The media for sound recording and reproduction have been developing rapidly – the older ones are going to ruin physically, and/or vanish in the flow of novelties on the market. Serious information institutions all over the planet make great efforts to preserve the recorded echoes of the history and keep them in a safe and wide accessible form.
Croatia came up early with the beginnings of sound industry. The changes in the cultural politics in the decades of state administration led it to a retardation in care for both material and intellectual value of recorded sound heritage – especially that of art music of Croatian composers and interpreters. The recordings have been recently kept on a few locations: Croatian Film Archive, sound archives of Croatian Radio *in Zagreb and seven local stations; rich but closed for the general public are the* archives of publishing companies. *Some sound recordings, mostly gramophone records and CD-s are kept in the music departments of* The National and University Library, public libraries *and* Academy of Music Library in Zagreb.
Until now, nobody has taken care of Croatian sound heritage systematically. Initiative is reduced to persistent private efforts of a group of enthusiasts, without official support of information experts and funds. The purpose of this article is to initiate activities in collecting information, listing, determining priorities in digitisation for preservation, organising funds, resolving the copyright-problems and establishing a digital repository opened for scientific and educational purpose.

Key Words: Croatian music heritage; preserving musical sound recordings; digitisation of sound; sound archive; digital repository

Introduction – Short history of sound recording

Humans always tried to keep and interpret complexes of sound they liked – imitating, orally, writing down, using mechanical devices and finally recording. The first appliance that could record the sound mechanically, but not to reproduce it, was "phonoautograf", invented 1857 by Edouard-Leon Scott. The first practical device for sound recording and reproduction was invented by Thomas Edison 130 years ago. His mechanical phonograph with cylinders from 1877 spreaded fast over the planet, initiating the inception of the new industry. In 1887 Emile Berliner issued a patent of the zinc disc coated with fatty film, and 1888 promoted a machine for the reproduction – gramophone. With many reconstructions and improving, in accordance with the development of electronic recording and reproduction of the sound, such as the new plastic materials, the gramophone remained the most popular media of sound industry until well into the 20th century. In the fifties of the last century began the fast development of magnetic tape as recording media, enabling the enrichment of the broadcast-production conceived in the 20-ies. The changes in media are getting faster and faster – in the 80-ies the sound was recorded in electronic machine readable zooming patterns of digits, formats were created and became obsolete, immense amount of information accumulated and interwave following the imperative of "development" for the purpose of increasing the consummation … and where is here the history here, as the solid point of support that will not allow us to drown in oblivion?

Sound recordings are preserved from the beginning in the countries that systematically take care of the whole national cultural heritage. The first sound archive – *Phonogrammarchiv* of The Academy of Science in Vienna was funded 1899. Among scientific, very rich and important are the publisher's archives and recordings collections of the broadcasting institutions. Libraries started to collect the sound recordings after the 2.nd World War.

Recorded Sound in Croatia

Croatia is a country with proportionally rich music culture. The beginning of the sound recordings production on its territory had already begun in 1908, when the engineer Slavoljub Penkala cut the first gramophone records. In that time Croatian musicians were recorded by the foreign companies like *Berliner Grammophon, Columbia Co., Victor, Oesterechische Grammophon Gesellschaft* etc.

M.(Mavro) Drucker (Ilica str. 39) had some kind of the publishing house of his own, for the Croatian music. In the catalogue were the recordings of Croatian interpreters: his own with the symbol of hawk, as well as the *Odeon*, *Writing Angel* and *His Master's Voice*.

The first big discography house in Croatia was founded in 1927, named *Edison Bell Penkala* – as a branch of the English company *Edison Bell International*. The market in discography, without any copyright at all, was very convenient

for the development of such industry, and a production was huge. It covered the market of contemporary Yugoslavia, Austria, Poland, Czechoslovakia of that time, Rumania, Albania, Greece, Little Asia, Syria and Palestine. Beside their own records, *EBP* sold the products of a dozen other companies. In the time of its production 1927-1933, *Edison Bell Penkala* generated a big number of records. The participation of Croatian interpreters were about 5%.
A year after the closing of *Edison Bell Penkala*, the new factory named *Elektroton*, was founded. After the II. World War the company was nationalised and was named *Jugoton*. Her authentic and licence production covered the territory of the whole Yugoslavian federation. From the 60-ies, particularly in the independent Republic of Croatia, the manufacturing of the commercial sound recordings had increased among growing number of publishing houses.
In 1926 in Zagreb the radio broadcasting station began to work. Allready in 1934 The Croatian Radio had started the proper production of records, using the *Philips* technology. Today it has seven regional radio-stations, three national and seven regional radio programmes.[1]

The state of sound heritage in Croatia – recordings of the Croatian composers and interpreters of art music: initial research

The music heritage – the segment of the recordings of Croatian composers and famous interpreters of art music – until now it hasn't been a subject of systematic research and care, although the Croatian Law on archival material and archives[2] as well as The Law on protection and preserving cultural goods[3] obliges every institution and individuals who owns them to protect them and permit the access for the research and education purpose. Depending on the investment amount and decisions of every single owner, the recordings are kept in different physical conditions, and the access to the information about their existence is more or less possible.
My research began with the list of the music libraries and collections in Zagreb.[4] The figures are completed in contacts with colleagues from some of the collections. According to this list, the following institutions are keeping the sound materials in Zagreb:
The Croatian State Archive unfortunately is not in function as a state music sound archive.[5] The efforts of the retired director of the Archive, dr Josip

1 IZLOŽBA fonografija u Hrvatskoj 1927-1997. (1997-1998; Zagreb).

2 ZAKON o arhivskom gradivu i arhivima. Zagreb, 29. rujna 1997.

3 ZAKON o zaštiti i očuvanju kulturnih dobara. Zagreb,25. lipnja 1999.

4 Juričić, Vedrana. Vodič kroz glazbene knjižnice i zbirke Zagreba.

5 Vidačković, Zlatko. Nova zgrada nakon trideset godina čekanja // Vijenac, Broj 301, 29. rujna 2005.

Kolanović, were continued by the new management in a very modest way.[6] In Croatian State Archive, within the competence and in the space of *The Croatian Film Archive*, without any qualified person employed, with very modest annual funds from *The Croatian Ministry of Culture*, the researches of the Croatian sound heritage in the institutions and private collections in Croatia and abroad are being done. There is some internal computational list with 1168 gramophone records (shellac and vinyl) looking like an ordinary table in word-format (with columns named order-number; title; interpret; producer; note). The interested person can get the printout of the list. The records are property of *The Croatian State Archive, The Croatia Records* and *The Croatian Radio*. The records that are not in possession of *The State Archive* were given to that institution for the restoration, which has been done partially.

The restoration is conducted by the outside collaborator who repairs them physically, than making an archival magnetic tape copy from the original historical gramophone device on his own, and finally cleaning the sound of any non-musical noises. The tapes are for now possible to listen in the Croatian Film Archive, on the tape-recorders borrowed from *The Croatia Records*. In a short time, when the devices will be returned, it would not be possible to reproduce the materials preserved in a such way. Talking to Ms Carmen Lhotka, the person in authority for the sound recordings in *The Film Archive*, I was told that they were planning to ask for the funds to buy some ordinary audio-CD-burning-device and a mini hi-fi appliance for the reproduction. Their idea is to digitise the archival tapes by themselves in so primitive way, and giving the copies to the disposal for listening on the mini-hi-fi within the Archive. They are not planning to create either a computerized data-base with the inventory list, or to fullfill the description of items, or to put the recordings on the Internet.

The Croatian Radio Sound Archive in Zagreb is the richest Croatian sound archive. Although it is a part of the public institution, until 1994 owned by the state, and even though it contains a great number of items in the category of heritage, it is closed for the outside users. It was founded after the foundation of the radio-station, for the program needs. The initial holdings was consisted only of gramophone records (they still keep the collection of so called hard – PVC records – some of them are lent to *The Croatian Film Archive*, for the restoration). In the fifties, the new technology of copying on the magnetic tapes had been accepted. The important parts of the collection of sound documents are documentary recordings (i.e. concerts) – some of them have a special value for the Croatian cultural heritage. Recent fund of the radio-archive includes gramophone records – PVC, shellac, vinyl, magnetic tapes on the opened and closed reels, DAT-tapes and CD-s. They are kept in metal compact archival closets, in the rooms with inappropriate climate. The records are listed in the paper-cata-

[6] Vidačković, Zlatko. Treba nam arhiv glazbe i govorne riječi // Vijenac, Broj 340, 15, ožujka 2007

logues, and mainly registered in the unique computational data base. The base was formed out by the international accepted standards. The data input is frequently inconsistent, with the wrong subject references, taken in by the group of ad hoc gathered outside co-operators.

The Croatian Radio has been trying fore years to save and to bring up to date the fund of its sound recordings, in their own studios for digitisation; the results are being kept on the few hard-discs. The selection of the materials for the digitisation is adequate to the needs of that commercial institution (everyday program) – so called serious music is at the end of every list of priorities.

The Croatian Radio is a member of The European Broadcasting Union. The people engaged in its sound archive are also striving to synchronize their work on digitisation with the standards accepted by the EBU members. The technique of the treatment for the old records is submitted to the needs of broadcasting (with cleaning of the non-musical-sounds, lowering the loudness etc.), in consistence with the NOA standard for the radio broadcasting.

The sound archives of the local radio-stations of the Croatian Radio: in Dubrovnik, Knin, Osijek, Pula, Rijeka, Split and Zadar, also contain a bigger amount of the sound recordings of Croatian composers and interpreters. Their lists were for this occasion not available to me.

The Music Collection of the National and University Library in Zagreb contains ca 25,000 gramophone records, ca 12,000 audio cassettes and some 9,000 audio CD-s. By a simple search it is not possible to now how many recordings contain the art music of Croatian authors. The items of the fond are mainly obtained as obligatory copies, from the year 1965 when the obligation had been legalized. They are listed in the library catalogue, according to the standards ISBD(NBM) and UNIMARC, and mainly available to the public in the very well equipped listening room. The older records (speed 78) are protected and not publicly available (there is no reproducing device either), so they are the candidates for the priority in the selection for the possible digitisation. The concrete plan of the digitisation of the sound-recordings is not included in the project of the digitisation of heritage within the *National and University Library*.

Some ten music collections of the **public libraries in Zagreb** and other Croatian towns also contain a number of older gramophone records and audio CD-s. The oldest of them – *The Zagreb City Library* collection (founded in 1962), is the biggest. The fund is described and listed in the computational data base, by the library standards and available for the research in the Internet. It is mainly mechanically damaged. It is planned to digitise some rare titles for the purpose of protection of the endangered media (gramophone records) because of the inaccessibility in any other format.

The collection of **The Academy of Music Library in Zagreb** contains a number of important gramophone records and commercially accessible audio CD-s. Some of the gramophone records are rare. They are invented and described in

the library catalogues, and accessible for the listening in the reading room. For now, there are no plans for digitisation.

The **archives of publishing houses** should keep the master copies of their issues. As a former director of *The State Archive*, dr Josip Kolanović told me, the result of his efforts was that, the company named *Hrvatska naklada zvuka i slike d.d. Croatia Records* (until 1991. under the name *Jugoton*), excluded the archive from its means of the privatisation, preserving it from the probable destruction. It is formally under the jurisdiction of *The State Archive*, but still in the premises of the company. The archive contains the collection of gramophone records and magnetic tapes, audio cassettes, video cassettes and CD-s (in 1997 about 40,000 titles)[7]. The collection contains their own publications (including the licence recordings of Croatian artists published on the gramophone records), as well as the issues made for the other publishers. It is listed for the internal use – since 1987 the list has been run in the computational data base, with no ambitions to obey any standards of the information branch. They keep one copy of every issue. Some recordings of 78rpm are preserved and wait for the restoration in *The Croatian Film Archive*. *The Croatia Records* digitise just the items for the re-publishing on the audio CD-s, and for that purpose the recordings are cleaned from the unwanted sounds. The selected titles from the gramophone records, under the serial title *Glazbeni spomenar* are prepared for the publishing in the last years of the vinyl era. A few music-lovers within the company, led by mr Veljko Lipovšćak, have prepared some 30 records. Only the small part of that uncommercial series is really published.

The single items of the first phonograph and gramophone issues are kept in the **museums of bigger Croatian cities** (Zagreb, Rijeka).

In the project named *Croatian Heritage* by the **Croatian Ministry of Culture** the sound recordings are not even mentioned.

In the daily press it is from time to time possible to find the stories about the collections of some old-sound-media lovers – especially gramophone records. The doctor of the technology science and composer from the town Split, Damir Tomčić, possesses some 20,000 gramophone records and 14 antique gramophone devices. Beside his collection, some private expert-group also mention the other collections, like Čapka, Kraker and Mirnik. Some of them are prepared to entrust their heritage treasure for the preservation and use to some serious institution, but there is no such one in Croatia[8].

Interesting and totally realistic is the initiative of three stubborn lovers – by already mentioned Damir Tončić, another two retired persons – Veljko Lipovšćak and Ivan Stamać from Zagreb. They, mostly research by themselves and at their

[7] Juričić, Vedrana. Vodič kroz glazbene knjižnice i zbirke Zagreba.

[8] Čelan, Joško. Hrvatska je gluha za svoju zvučnu baštinu // *Nedjeljna Slobodna Dalmacija*, 04/02/2007., str. 34-35.

expense the older history of the Croatian music sound heritage, consulting the recent press materials and publishers' catalogues, as well as the catalogues of foreign sound archives. From time to time they present the results of the research by organising the popular exhibitions in museums, or writing for the newspapers and serials.

They have, for some ten years asked for help from the state institutions – from *The National and University Library* to *The Croatian Ministry of Culture* – unfortunately until now without any success. With the support of respectable institutions (*Croatian Composers Society, Croatian Section of The Audio Engineering Society, Section for the literature of The Croatian Academy of Arts and Science,* as well as *Oesterreichisches Phonogrammarchiv and Oesterreichiches Mediathek*), this year they have asked for the funds from the pre-accession fund of the European Union, for establishing the Croatian sound archive. They are planning to find and equip their own plase where the rare materials would be kept in ideal conditions, the whole Croatian sound heritage restored and digitised, and founded the digital repository[9].

For the complete rewiew of the state of Croatian art musical sound heritage it is necessary to consult the lists of the items of all institutions already mentioned, and interwiev other possible owners. It is necessary to contact the similar institutions abroad – in the countries where the Croatian artists stayed and/or made any sound recordings.

Criteria for the selection of materials for the digitisation

Inevitable systematically digitisation of sound recordings that are worthy preserving of the national music heritage is complex, expensive and time-consuming process. Therefore it is necessary also in Croatia to determine priorities for the conversion of materials – in the segment of the art music of Croatian composers and interpreters, taking into consideration the following criteria:

1. the condition of sound carriers (by the recommendations of IASA[10])
2. cultural, scientific and academic meaning of the content
3. rarity of the title

ad 2) Music is the kind of art in continued recreation, where interpreters take a part as well as the composer. By the selection it would be necessary to give the priority to the valuable interpretations of the artist that took the part in the Croatian history

ad 3) The recordings of some compositions survived just in documentary form (i.e. the live concert recording), on the magnetic tape, when others are published on many media and/or editions. Some compositions – especially of the recent

[9] Disopra, Tina: Prvi fonoarhiv // *Globus*, 29/06/2007, str. 66-67.

[10] Task force to establish selection criteria of analogue and digital audio contents for transfer to data formats for preservation purposes. [s.l.], IASA, 2004.

authors, are not written down at all, or they contain the element of improvisation, that makes the performance recordings unique.
How long is the life of the digital archive? – ask the sceptics[11]. Is it possible to loose the history trusted to the machines in constant changes? – is "five years or eternity" enough guaranties for the preservation of the digitised documents?
By selecting priorities for the preserving of the recordings by digitalisation it is necessary to take in consideration the following parameters:

4. the physical durability of sound carriers
5. obsolesce of the reproduction devices
6. disappearance of experts

ad 4) The physical durability of sound carriers
The IASA document mentioned above gives the list of the physical durability of sound carriers. About the **magnetic tapes**: "… Generally, only standard play tapes open reel tapes (SP, 52 μm total thickness) should be trusted to be mechanically stable… The lesser the mechanical stability, the greater the chance that the tape suffers from inadequate winding, which is one of the most underrated risks for magnetic tapes…prolonged storage of badly wound tapes causes irreversible deformations, which may lead to severe replay problems, specifically with thin tapes and high density recordings, e.g. R-Dat."
The same IASA document describes the new, **optical carriers.** "The data integrity of CD-s, like all other digital media, is objectively measurable by special CD players and suitable software … According to digital archival principles (cf. IASA-TC 03, § 11), every CD must be free of uncorrectable errors." The replicated audio CD-s may contain the interpolations that are not original signal. It is advisable to test them for full error correction repeated in regular intervals.
ad 5) Reproduction equipment
All audio carriers are machine readable. Following the IASA *Task force*, by choosing the priorities for the migration in the new (digital) media it is necessary to take into consideration the availability of professional equipment of high quality. There are some new inventions in the sound reproduction technology, where the gramophone record in the absence of the historical gramophone, or for the preservation of mechanical damage, could be "read" by a laser ray[12]. In the American *Library of Congress* a machine for the scanning of gramophone discs, even broken ones is being improved, and reproduced by a computer.[13]

[11] Warnke, Martin. Speicher, Archiv, Gedaechnis. Das Paradox der digitalen Archive. // Musik-Sammlungen – Speicher interkultureller Prozesse : Teilband A, pp. 95-111.

[12] http://www.laserturntable.com/ (01/09/2007)

[13] Boyce Nell. You Can Play the Record, but don't Touch

Techniques of music digitisation – standards for the archival custody of digitised music recordings

Digitisation of sound in the information institutions all over the world started in the 90-ies in last century. Leading were the rich countries with highly developed tradition in archives and libraries (Australia, the USA, the United Kingdom, the Netherlands etc.). From their practice in the short time the standards have developed – first on the national level and by the time on the level of the international community of information branch. As a subject of the new scientific and professional research, it has generated a list of relevant studies, helpful for the new beginners to find the proper solutions.

The study entitled *Moving Images and Sound archiving Study AHDS*[14] recommends the following technical standard for the archival custody of digitised sound recordings:

- 96kHz and 24 bits for the archival copy, that enables the dynamic range of 110 decibel, without compression, and at least stereo
- for the users copy rather AAC compression than MPEG(MP3)

The data migrated in the digital media, like born digital, due to obsolescence need to be copied using the new technical devices – to emulate.

Metadata

To organise the collection of digitised recordings and to ensure the possibility to access of the information, it is necessary to describe it. The description contains three types of data: technical, administrative and data for research and use. Some institutions developed the standards for the description (Dublin Core, CD Terms, PB Core, METS, PREMIS, AudioMD etc.). The automation of the metadata extraction is increasing.

It is recommended to use existing sets of metadata – the creation of the new set is expensive and not worthy the trouble.

Organising repository

Audio-visual collections are deposited and available for the use in various organisations, better or poorly functional equipped. Making efforts to improve the accessibility to the information, it is necessary to make a decision whether the depository will be associated with the institution, or to become a part of the collaborative service of the group of other repositories, or to give it for the administration to the external agency or some national service. The good practice has proved that the most profitable is to develop separate but networked collections, with the use of network technology and tools like the Storage Resource Broke (SRB).

[14] Moving Images and Sound Archiving Study // Arts and humanities data service.

Examples from the neighbourhood

National and University Library in Ljubljana

With the funds from abroad, the employees of the *National and University Library in Ljubljana* (3 computer specialists and 1 musicologist-librarian) carried out the project with a simple title *Digitisation of old sound recordings*. They have selected a delimited collection of physically endangered and historical interesting shellac gramophone records with the recordings of approximately 100 Slovenian compositions. They have followed the good practice of the European project *Minerva*, and recommendations of information experts. As they didn't have the original historical gramophone for the old 78rpm recordings, they played them at 33rpm and the speed adjusted by means of a computer. The sound recording is accurately transcribed from the analogue to digital medium with all scars, noises and cracks and by means of the program Adobe *Audition* digital recording is cleansed from all non-musical compositions. Digital copy is stored in WAV (standard Windows audio format in big files) and in user, quality controlled MP3 files. For the preservation of the visual image of the original source, both sides of a record are scanned. A metadata scheme is made according to the Dublin Core standard, accepted by the European projects TEL-MEMORE and TEL, as well as OAI protocol. To fit in Slovenian OPAC named COBISS, based on the UNIMARC standard, a computer interface was designed which obtained data from COBISS(UNIMARC), adjusted to fit Dublin Core and transferred them into another database. Searching the database is possible according to the general parameters of the bibliographic description.

Users can export the electronic source, by a telephone or mail order. Because the recordings are no longer under Copyright Law, reproduction is legal. The next step in digitisation process will be digitisation of catalogues and documentary materials of Slovenian musicians and institutions, followed by the digitisation of endangered music periodicals and their archives[15].

National Library of Serbia

According to the text from the beginning of August this year[16] the *National Library of Serbia* is planning to establish the national sound archive, to keep digitised sound recordings of all available sound recordings from all over the Serbia – on the model of the national archive of Great Britain.

The four years work on the digitisation of the complete library collection of 78rpm gramophone record is coming to an end (about 600 of about 1000 records). Their plan is to consolidate all Serbian institutions with the similar fund.

[15] Moličnik Šivic, Simona: From presentation to reality, from preserved to hard. Experience in the digitization of analogue sound recordings in the Music Collection of the National and University Library in Ljubljana // Pregled NCD 8 (2006), 75-79.

[16] Tanjug, 08.08.2007.: NSB planira osnivanje Nacionalnog zvučnog arhiva // *Blic Online* I Kultura.

How to preserve Croatian music sound heritage – proposal instead of conclusion

Based on the limited picture of the condition of materials in institutions and collections, as well as on good practices from the developed countries, I'm willing to first of all propose the strategy to save the important segment of Croatian history – sound recordings of art music in the first place – composers and interpreters.

At the moment, the whole responsibility for the preserving the materials is given to the employees initiative. According to the Law, authorized institutions – among all the *Ministry of Culture* and *The Croatian State Archive* should take over the responsibility and to determine the standards for the preservation. Before all it is necessary to make a detailed study, or a comprehensive project in following steps:

1. making a detailed list of owners of sound materials (according to the survey directed to the institutions and private persons)
2. making a detailed standardised list of materials – with the specification of the media format, age, deposit and physical conditions
3. based on the list to select the priorities for the preservation, considering the physical endangered condition, intellectual value of materials and rarity of items
4. to determine the standards for the migration in digital media, according to the above mentioned recommendations; to instruct the institutions that are already working on the digitisation to coordinate with standards
5. to educate the employees in the institutions that take care of the sound materials
6. according to above mentioned recommendations to make at least two copies of the recording – archival and compressed, for the public access – it would perhaps be of greatest worth to equip the unique studio for the digitisation (by the Croatian State Archive or some future Croatian sound archive?)
7. to determine the unique standard for the description (metadata) of the digitised recordings
8. to establish a unique repository for the sound recordings, by some institution – *Croatian State Archive* or some future Croatian sound archive?
9. to harmonize the variations of the explanations of the Copyright Law – enabling the free access to the heritage materials for the educational and scientific purpose
10. the project should be financed by the Croatian State and, in future perspective, European institutions that are helping to the approaching countries

References

Aleksandrović, Vesna. Digitalizacija građe posebnih fondova Narodne biblioteke Srbije // *Glasnik Narodne biblioteke Srbije* (1), pp. 107-120 http://eprints.rclis.org/archive/00002166/ (18/01/2006)

Boyce Nell. You Can Play the Record, but Don't Touch . http://www.npr.org/templates/story/story.php?storyId=11851842 (01/09/2007)

The British Library. The world’s knowledge. http://www.bl.uk/collections/sound-Archive/nsaabout.html (23/08/2007)

Brygfjeld, Svein Arne ; Solbakk, Svein Arne. Norvegian Digital Radio Archive Initiative http.//www.rlg.org/preserv/diginews/diginews2-6.html (01/09/2007)

Commision of the European Communities. On the digitisation and online accessibility of the cultural material and digital preservation. Brussels, 24.8.2006. http://ec.europa.eu/information_society/newsroom/cf/document.cfm?action=display&doc_id=160 (01/09/2007)

Čelan, Joško. Hrvatska je gluha za svoju zvučnu baštinu // *Nedjeljna Slobodna Dalmacija*, 04/02/.2007, pp.34-35.

Digitalizirana baština. http://www.nsk.hr/Heritage.aspx?id=25 (01/09/2007)

Digitisation of Records. Better practices Checklist. Version 1, 2004 http://www.agimo.gov.au/checklists (25/05/2007)

Disopra, Tina: Prvi fonoarhiv // *Globus*, 29/06/2007, pp. 66-67.

Dubrović, Ervin. Čarobna igla // *Sušačla revija* no. 46/47. http://www.klun-suscana.hr/revija/clanak.asp?Num=46-47&C=6 (21/08/2007)

Europe’s cultural and scientific heritage at a click of a mouse. http://ec.europa.eu/information_society/activities/digital_libraries/index_en.htm (01/09/2007)

Fleischhauer, Carl. The Library of Congress Digital Audio Preservation Prototyping Project. http://www.arl.org/preserve/sound_savings_proceedings/fleischhauer.html (1 of 9). December 15, 2003 (20/05/2007)

Going Digital: Strategies for Access, Preservation, and Conversion of Collections to a Digital Format. New York, London : The Haworth Press, Inc., [s.a.].

Handbook for Digital Projects: A Management Tool for Preservation and Access. December 4, 2004.Andover, Massachusetts : Northeast Document Conservator Center, 2000. http://www.nedcc.org/oldnedccsite/digital/dman.pdf (01/09/2007)

Harvey, Ross. Preserving Digital Materials. Muenchen : K.G. Saur, 2005.

HRTweb. http:// www.hrt.hr (01/09/2007)

Hrvatska kulturna baština. 20.03.2007. http://daz.hr/bastina/ (01/09/2007)

Hunter, Gregory S. Preserving Digital Information. New York, London : Neal-Schuman Publishers, Inc. [s.a.].

IASA Tehnical Committee. The safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy. Version 3, December 2005. http.//www.iasa-web.org (01/09/2007)

IFLA Guidelines for Digitisation Projects. March 2002. http://www.ifla.org/VII(s19/sconsv.htm (01/09/2007)

IFLA. Smjernice za audiovizualnu i multimedijsku građu u knjižnicama i drugim ustanovama. Zagreb : Hrvatsko knjižničarsko društvo, 2005.

IZLOŽBA fonografija u Hrvatskoj 1927-1997. (1997-1998; Zagreb). Zagreb : Tehnički muzej, 1997.

Juričić, Vedrana. Vodič kroz glazbene knjižnice i zbirke Zagreba. Zagreb : Hrvatsko muzikološko društvo … [et al.], 1997.

Laser Turntable. http://www.laserturntable.com/ (01/09/2007)

Lipovšćak, Veljko. Zaštita audiozapisa // *Arhivski vjesnik*,.43 (2000), pp.109-136.

Moličnik Šivic, Simona: From presentation to reality, from preserved to hard. Experience in the digitization of analogue sound recordings in the Music Collection of the National and University Library in Ljubljana // *Pregled NCD* 8 (2006), 75-79. http://www.ncd.matf.bg.ac.yu/casopis/08/NCD08075.pdf (01/09/2007)

Moving Images and Sound Archiving Study // Arts and humanities data service, http://ahds.ac.uk/about/projects/archiving-studies/moving-images-sound-archiving-final.pdf. (01/09/2007)

NOB Cross Media Facilites. Digital Facility for Playout and Archiving, Netherlands. http://www.broadcast-technology.com/project_prinatble.asp?ProjectID=2978 (22/08/2007)

Sapunar, Marko. Radiodifuzni sustavi. Zagreb : Alinea, 1994.

Stančić, Hrvoje. Proces digitalizacije // *Informacijske znanosti u procesu promjena*. Zagreb, Filozofski fakultet, Zavod za informacijske studije Odsjeka za informacijske znanosti, 2005., pp. 45-83.

Stančić, Hrvoje. Teorijski model postojanog očuvanja autentičnosti elektroničkih informacijskih objekata. Doktorska disertacija. Zagreb : Filozofski fakultet, 2005.

Tanjug, 08.08.2007.: NSB planira osnivanje Nacionalnog zvučnog arhiva // *Blic Online*, Kultura http://www.blic.co.yu/kultura.php?id=10154. (21/08/2007)

Task force to establish selection criteria of analogue and digital audio contents for transfer to data formats for preservation purposes. [s.l.], IASA, 2004.

Vidačković, Zlatko. Nova zgrada nakon trideset godina čekanja // *Vijenac*, Nr 301, 29/09/2005. http://www.matica.hr/Vijenac/vijenac301.nsf/AllWebDocs/Razgovor,(01/09/.2007)

Vidačković, Zlatko. Treba nam arhiv glazbe i govorne riječi // Vijenac, Broj 340, 15/03/2007 http://www.matica.hr/Vijenac/vijenac340.nsf/AllWebDocs/Treba_nam_arhiv_glazbe_i_govorne_rijeci (01/09/2007)

Warnke, Martin. Speicher, Archiv, Gedaechnis. Das Paradox der digitalen Archive. // Musik-.Sammlungen – Speicher interkultureller Prozesse : Teilband A, Stuttgart : Franz Steiner Verlag, 2007, pp. 95-111.

ZAKON o arhivskom gradivu i arhivima. Zagreb, 29/09/1997 http://www.nn.hr/clanci/sluzbeno/1997/1617.htm (01/09/2007)

ZAKON o zaštiti i očuvanju kulturnih dobara. Zagreb, 25/06/1999 http.//www.nn.hr/clanci/sluzbeno/1999/1284.htm (01/09/2007)

Digitization of Sound Recordings as an Example for Preservation of Oral and Music Folklore Heritage

Basic principles and two examples from practice

Aleksandra Mežnarić Karafin
Academy of music, Library
Berislavićeva 16, Zagreb, Croatia
alexm@muza.hr

Summary

Sound recordings are the result of almost every field research conducted in Zagreb Institute for ethnology and folklore research (IEF), and Vienna Phonogrammarchiv of Austrian Academy of sciences. Documenting orally transmitted traditions, sound recordings are becoming specific witnesses of time, and they cannot be adequately replaced by any transcription. They have archival value, and their protection and preservation is of utmost importance. These recordings are mostly analogue, on magnetic tapes as carrier, which opens a set of questions regarding their preservation. Preservation professionals find that digitization may provide an answer to this problem. This paper analyzes the importance of field recordings in folklore research, and internationally accepted standards for their digitization. It also presents magnetic tapes digitization praxis in earlier named institutions, comparing approaches and results.

Keywords: folklore heritage preservation, preservation of sound recordings, digitization of sound recordings

Introduction

Since the end of the 19th century, sound recordings have been the witnesses of almost every important segment of the human practice. Speech recordings of important historical figures and recorded interpretations of great musicians are the only way of preserving for the future unique parts of history. Documentary recordings resulting form field research are essential in a study of folklore as communication process embodied in orally transmitted forms of cultural expression of all segments of society.[1] Original recordings of authentic speakers and different forms of musical expression contain spontaneous, unrepeatable elements in a way written notes could not, which then documents segments of

[1] About problems regarding definition of *folklore, see:* Lozica, Ivan. Izvan teatra (…), p. 18- 31

past reality. For more then a century they have been deposited in archives and sound recordings collections of the research institutes, the oldest sound archive – *Phonogrammarchiv* of Austrian academy of science is founded in Vienna as far back as 1899.

Our ability to hear recorded sound depends on physical integrity of the carrier, sound record format, and availability and functionality of reproduction equipment. Sound carriers, analogue and digital, generally have shorter life expectance then quality traditional printed materials, and they are more subject to damage caused by inadequate handling, storage and replay equipment. The concern about these recordings becoming inaccessible is thus objectively founded. Problem of their preservation became a subject of increasing number of discussions among expert community in last two decades. Digitization of analogue sound recordings, according to internationally accepted standards, in this moment appears to be the best solution for preservation of these endangered materials and for insurance of their accessibility. Lifetime of digital carriers, problem of selection and eventual obsolescence of digital format, independent from the media, raises then an all-new set of questions.

International expert associations, libraries and archives are recognizing the value of these materials and set of problems regarding their preservation. They are working on detailed standards, recommendations and projects for protection and preservation by digitization of deteriorating sound recordings. In United States their protection is under regulation of U. S. Congress National Recording Preservation Act (2000), and from it consequent expert bodies.[2] There are more and more projects (TAPE, PRESTO, UNESCO Memory of the world programme, etc.) oriented on preserving audiovisual materials and heritage. In the same time this problem, although indicated in the *Law about archive materials and archives,* remains unsolved, staying at the margin of interest of expert community. There are few possible reasons for that: **a)** in principle, value of these recordings as significant archival materials of national interest is not recognized, **b)** there is no systematic care about its archiving, protection and preservation **c)** lack of educated personnel engaged in its care and preservation.

This paper was initiated by those reasons, and its purpose is to point out the importance and problem of protection and long-term preservation of field recordings on analogue magnetic tapes. Large quantities of these tapes are the core of sound recordings collections of the *Institute of ethnology and folklore research* (IEF) in Zagreb, and *Phonogrammarchiv* in Vienna. Both collections are as a whole (Zagreb), or partially (Vienna) acknowledged as protected entities of a great value and importance. Documentation of IEF, from year 1991, is included into the Registry of mobile culture monuments of the Regional

[2] National Recording Registry, National Recording Preservation Foundation, National Recording Preservation Board

institute for the protection of cultural monuments in Zagreb, and so as a whole obtained the status of the monument of the highest ("0" i "I") category. Historical collections of the *Phonogrammarchiv*, with materials from the period of 1899-1950, are included in the registry of UNESCO *Memory of the world Programme in year* 1999.

Influence of development of the recording technologies on the field research

Sound recording technology development, which dramatically begun with Edison's invention of the phonograph in 1877, had a direct impact on the scope, focus and methods of the scientific research of traditional culture by the beginning of 20th century. After Edison placed the new, improved version of the phonograph, with which first field recordings on the wax cylinders were made, on the market in 1880, phonograph recording quickly found the place in the scientific circles. Researchers in the filed of anthropology and linguistics were the first ones who used the advantages of new sound recording technology. Due to this technology they could record and preserve the sounds, speech and musical performances, which could only be partially documented in the written notes. In 1890 ethnologist Jesse Walter Fawkes recorded the first filed recordings of Passamaquoddy Indians songs with phonograph.[3] Linguist Milan Rešetar recorded the first phonograph recordings in Croatia in 1901, as a part of his research of Slavic languages for Vienna *Phonogrammarchiv*.[4] At the beginning of 20th century, Bartók and Kodály recorded traditional music of Hungary and Rumania with phonograph. In the twenties and thirties of the 20th century Božidar Širola and Milovan Gavazzi made first phonograph recordings while working in the Department for the folk music, founded in 1921 in Ethnographic Museum in Zagreb. Vienna *Phonogrammarchiv* donated the first phonograph to the Department as its branch, and Department was obliged by the contract to turn over all original cylinders to the Phonogrammarchiv, their copies were deposited in Zagreb.[5]

None of the future recording methods found its way in the filed research practice as deeply as magnetic recording technique. The U.S. engineer Oberlin Smith set basic principles of this technique as far as in 1878; it was perfected in Germany and in the U.S. during and after the World War II. Magnetic sound recording was then widely spread in the fifties. Sound carrier evolved from the steel wire and steel band, into the first magnetic tape, as we know it today. Recording equipment became more available and affordable, and practical transportable recording units that used batteries for the power supply made fieldwork

[3] Brady, Erika. A spiral way (…)

[4] Lechleitner, G. Ethnomusicology and the historical collections (…)

[5] Bezić, Jerko. Etnomuzikološka i etnokoreološka djelatnost Instituta (…), p. 21

much easier. Vienna Phonogrammarchiv introduced magnetic tape technique in 1951; from 1958 portable tape recorders were used in the field research. In 1954, Vinko Žganec made for IEF first magnetic tape recordings on his field trip in Međimurje. These recordings are held in the sound recordings collection of IEF.
Magnetic tape technique was used as a primary field recording technique by the eighties, when digital recording technology was perfected, and came into common use. Results of the wide use of the magnetic tape recording are large collections of sound recordings that were made all over the world. They bare witness to the focus of the scientific research in home institutions, but more importantly, of significant cultural changes in this turbulent period.

Preservation of sound recordings

Problems

Only constants in the development of sound recording and sound reproduction technology are continuous and ever more rapid alternations of innovation and obsolescence in every aspect of recording, reproduction and archiving of sound recordings. Our ability to hear recorded sound depends on several elements: sound carrier, system for its reproduction, and lately, on digital format and operation system. In case problem occurs in just one of these elements, sound recording may become inaccessible. In attempting to preserve analogue sound recordings, priority is mostly given to preservation of original sound content, more then on the preservation of the original sound carrier.

Approaches

The only way to preserve a sound recording stored on the carrier that is, even with best storage conditions and protection techniques, endangered by inevitable deterioration, is the transfer of the recording on to the new media, that is, its conversion to a new format. Still in the last decade of 20th century, some expert associations as Audio Engineering Society, National Academy of Recording Arts and Association of Recorded Sound Collections expressed the concern about use of digital technology and digital storage media, whose stability and durability were not yet confirmed in the practice in their discussions about long-term preservation of sound recordings. For this reason, as a solution for preservation and storage of sound recordings, they prefer re-recording of endangered materials on analogue magnetic tapes.[6] Ch. A. Patton (1998) also recommends the use of analogue recording technology, and magnetic tape as a storage media in his study that covers all aspects of sound recordings preservation. Although,

[6] “[...] because analogue tape has been proved to last, [...], and because the shelf life of digital tape is unknown, recordings should be stored and backed up, at least in the analogue tape format.” Capturing Analogue Sound for Digital Preservation , p. 2

in some aspects, he acknowledges few advantages of digital technology,[7] he finds more arguments against its use in the field of sound recordings preservation and archiving. These are: **a)** rapid change and improvement of the technology – results in rapid obsolescence of hardware, digital format and storage media, **b)** lack of consensus in the expert community regarding sample rate, bit depth and record format for sound archiving, **c)** questionable stability and durability of the storage media.

Yet in time when digitization of conventional library materials was perfected and became common practice, Abby Smith refers to digitization in 1999 primarily as to a method of providing access to rare, endangered, or distance materials, and not as to a permanent solution for preservation[8]. However, statement of E. Cohen "distribution is preservation" relays exactly on possibilities that lays in digital domain and networked environment: creating unlimited number of identical copies without the loss of original information, simple distribution of the documents, web access, permanent, media-independent preservation of digital content. Urgent conversion of analogue recordings into a digital domain is an imperative and to hesitate means to compromise preservation.[9] Lately this approach is considered the best solution for preservation of original sound content stored on deteriorating analogue sound carriers. In the same time, preservation of the original carrier and original replay equipment is of same importance: future development of sound reproduction technology could enable the transfer of yet unavailable subtle sound content.

The main obstacle for the adequate preparation and realization of the digitization projects is the lack of financial support, but even more, uneducated personal in charge of the collection. In this sense, there is a visible difference between the archives in the narrow sense, and sound recording collections in research institutions. Institutions whose basic purpose is archiving and preservation usually have established preservation policy and ways for financing preservation and digitization projects. Collections that are part of research institutes usually have a different role – they serve as a source of materials for current research, mostly without defined preservation and digitization policy, and with no finances planed for this purpose.

[7] "[They are] easier to edit, recording systems are inherently less noisy than analog systems. In theory […] should be possible to "clone" digital recordings, permitting the creation of many generations of identical copies". Patton, Ch. A. Preservation re-recording of audio recordings (…), p .207

[8] Smith, Abby. Why Digitize? , p. 5

[9] Cohen, Elizabeth. Preservation of Audio, p. 1

Analogue sound recording digitization: international standards and recommendations

Documents as IFLA Guidelines for digitisation projects (2002), IASA-Technical Committee Standards IASA-TC 03(2005), IASA Task force to establish selection criteria of analogue and digital audio contents for transfer to data formats for preservation purposes (2003) are setting the standards and recommended procedures for transfer and digitization of analogue sound recordings, for storage, permanent preservation, and for providing of access to digital sound recordings. They are defining reasons for digitization, selection criteria, and basic phases in the digitization process.

Reasons for digitization

Considering analogue magnetic tapes, primary reasons for digitization are: **a)** saving original sound content from endangered carrier / protection of the original carrier – digital copy preserves sound content, in the same time preventing the wear of the original carrier; **b)** easy access – replay equipment for analogue magnetic tapes became obsolete, it is complicated for handling, incompetent user can cause a permanent damage to the original tape; **c)** providing new services for the users.

Selection criteria

Selection criteria will be determined with primary function and activity of the home institution. According to *IFLA Guidelines for digitization projects* basic criteria for selection materials for digitization are:

- *Content:* priorities are determined by evaluation of intellectual value of materials, their historical, scientific and cultural significance. Priority must be given to unique materials /sources.
- *Demand:* priority should be given to the materials in constant demand. Transfer of the content into the digital form will satisfy the needs of current users, at the same time attracting the new ones.
- *Condition:* fragile and damaged unique materials have a priority in transfer into a digital domain; this minimizes the possibility of a permanent damage caused by the frequent use. Digitization process alone can be of a high-risk for the damaged materials. They must be handled with great care, some restoration procedures may be needed before the transfer. Priorities in digitization projects of analogue sound recordings are set by IASA-TC 03 § 16: **a)** documents at immediate risk / recordings on endangered media **b)** documents who are a part of an obsolete or commercially unsupported system **c)** documents in regular demand.

Standards for transfer of analogue recordings

To achieve optimal transfer of original analogue recording, it is necessary to insure maximum fidelity during the reproduction of original. Expert knowledge of original format and all of its characteristics is required, as well as availability and optimal adjustment of well maintained, and completely functional replay equipment. Lack of professional knowledge and adequate replay equipment can result in inadequate transfer of analogue sound signal and irreversible damage of original carrier. Necessary steps in reproduction and transfer of analogue recordings on magnetic tapes should be:

- Checking, preparation and cleaning of original tape. Procedure includes identification of the tape (tape base – paper, acetate, PVC, polyester), establishing tape condition and eventual damage repair (damaged splices, mould, deterioration of the tape base, oxide or binder, acetate tapes are considered to be most unstable)
- Configuration and calibration of replay equipment: **a)** setting the tape speed (most common are 4,75 cm/s, 9,5 cm/s, 19 cm/s; in field recordings often more recordings recorded at the different speed and formats can be found on the same tape **b)** defining the recording format (mono, stereo), recording and replay equalization standards (CCIR, NAB), azimuth adjustment (adjustment of the reproduction head in the same angle as it was during the recording of original)

All data about original tapes, all restoration procedures, replay and transfer parameters must be documented in details, and available for the future reference.

Sound recording digitization standards

In process of digitization, sound signal that is analogue by nature, in computer Sampling/quantization processing is divided into a specific number of vertical segments – samples. Number of those segments is defined by sampling frequency (kHz). Quantization then divides every vertical segment, which is defined by number of bits and often called "word length". Sampling frequency and bit rate define the quantity of sound information transferred from the original source. The quality, fidelity and dynamic range of digitized sound recording are depending on them. Sampling frequency and bit rate adequate for sound archiving is the subject of many discussions among preservation experts and audio engineers. Some archivists are suggesting high sampling frequencies of 192 kHz, with arguments "just in case", "there are hard to hear harmonics that should be kept".[10] IASA-TC 03 is suggesting 96 kHz sampling frequency, with 24-bit word length, as archival sound record standard. From archival copy user copy can be made, with downsampling to CD-Audio quality (44, 1 kHz/16 bit), and with data reduction (MP3) when needed for network transfer, etc.

[10] Fleischhauer, Carl. The Library of Congress Digital Audio Preservation (…), p. 5

Unmodified transfer
Transfer of analogue sound signal to new archive format should be carried out without any modifications or de-noising. From archival point of view, secondary, "unintended" sound artefacts (clicks, noise) are considered to be part of the sound document, and are equally important as primary, "intended" sound content.[11] By the use of signal processing from unmodified, archival copy, modified sound document can be created for commercial presentation or to answer the users' needs.

Sound record format
To minimize the possibility that digital sound format becomes obsolete too quickly in creating digital archival sound record, widely accepted formats should be chosen. Format should support high-resolution audio records, it should be transparent (with simple coding schemes – Pulse Code Modulation is recommended, without data reduction), and should have enclosed metadata about data extraction procedures.
IASA-TC 03 at this moment recommends following formats for digital sound archiving: WAVE (.wav), Broadcast Wave Format (.bwf), Audio Interchange File Format (.aiff), CD Digital Audio (CD-DA), with PCM or LPCM coding schemes. Formats with data reduction (MP3 is most popular) can be acceptable for access copies or transfer thru the web (web streaming), but they are not appropriate for long-term preservation of sound documents.

Storage of digital sound record
IASA TC03 suggests following principles of digital sound archiving:

- Each digital copy created in archival purposes must be checked against the imported file (verified), must be free of uncorrectable errors, with lowest possible number of correctable errors
- Each carrier containing digital recordings must be regularly checked for data integrity
- Digital content must be copied to a new carrier each time when number of error increases significantly, before uncorrectable errors occur (refreshment)
- Digital content must be copied before digital carriers, formats and/or hardware becomes obsolete (migration)
- It is essential to keep at least two digital preservation copies, and to
- use additional copies for access. The preservation copies should be kept in different location whenever possible.

[11] IASA-TC 03, § 8: "It is important to understand that the intended signal is only part of a given document. The unintended and undesirable artefacts (noise, clicks, and distortions) are also part of the sound document. Both have to be preserved with utmost accuracy."

Like analogue media, digital magnetic and optical storage media also have expected lifetime and stability that depends on conditions of their use and storage and they are subjected to format and hardware obsolescence. To date, in digital storage practice following storage media have been used: R-DAT, CD-R, DLT and LTO c7omputer magnetic tapes. Currently, R-DAT is already considered obsolete and concern rises also about the use of CD-R and DVD-s as digital target formats for archives.[12] Because of the rapid changes and development of digital storage media, formats and hardware, automated, media-independent approach appears to be the solution for long-term preservation of digital sound record, its migration before format obsolescence, or deterioration of built-in error correction. Storage of digital data in repositories based on OAIS model[13] or in DMSS (*Digital Mass Storage System)* is an ideal solution. These automated systems for storage, management, maintaining and integrity check of digital data, their preservation and distribution of digital objects with embedded metadata, are closest to solution of permanent preservation of digital objects problem. Currently, digital mass storage systems are installed in some major sound archives, but because of large financial investments required for this purpose, they are out of reach to smaller institutions and archives.

Metadata

Metadata in digital environment are extension of basic catalogue description, and they are indispensable for finding, control and usage of digital documents. Beside descriptional, administrative and structural metadata, set of preservation metadata is mandatory at assessment of technical parameters of the recording. It should contain details about the original carrier, format and preservation; equipment for reproduction and their parameters; digital resolution, digital format; operators included into a process, digital signature for document Most used basic schemes of metadata are MARC bibliographic record and Dublin Core Metadata Element Set. They can be stored as standalone documents, or as integral part of digital document, where Standard Generalized Markup Language (SGML) is most used authentication; details about secondary sources of information.[14]

Examples from praxis

Institute of ethnology and folklore research

Sound recording collection of IEF is an integral part of its Documentation collection, which collects, catalogues, and preserves documents which are result of scientific research activities of Institute's scientists and collaborators from year

[12] Bradley, Kevin. Risks associated with the use of recordable CDs (…)

[13] Fleischhauer, Carl. The Library of Congress Digital Audio Preservation (…), p. 8

[14] IASA-TC 03, § 15

1948, when *Institute* was founded,. This material in unique way bears witness to Croatian traditional culture of 20th century, to other nationalities in Croatia, and to Croats abroad. As mentioned before, IEF documentation has obtained status of cultural monument "0" and "I" category. Sound recording collection of IEF stores 3,300 magnetic tapes –around 4,000 hours of recorded material.

During last war in Croatia magnetic tapes were, for the purpose of protection from war danger, stored in the only available place at this time – wet cellar of an old building in Zvonimir Street, where IEF was located. Tapes were protected by polyvinyl bags.

At year 1995 ethnomusicologist Grozdana Marošević has discovered first tapes with signs of mould, and after consultations with audio engineer of Croatian radio and founder of Samofix d.o.o company, Vito Gospodnetić, digitization of sound recording collection was initiated. It was decided to transfer magnetic tape recordings to audio CD-s. Reason for that was, in the first place, an attempt to save sound material on magnetic tapes that was showing obvious signs of decay, and then preservation of original tapes. However, plan for digitization, with defined priorities, standards for converting and storing of digital sound record based on expert archival criteria, did not exist. Overall plan, damaged tapes restoration procedures, conversion of original audio recording, digital resolution and choice of storage media were defined mostly by project proposal by Vito Gospodnetić. Project objectives in this proposal were: **a)** to make material available for larger number of users**, b)** easier browsing thru CD content, **c)** technical "make up", which would increase quality of recordings, and make them clearer.[15] Choice of CD-R as a storage media is based on its wide acceptance and practical use and, at that time (1996), unconfirmed assumptions about its durability and life expectancy.[16] There were no priorities set in digitization procedure; recordings were transferred in natural order, beginning from the oldest ones. Sometimes some "newer" recordings were digitized, to meet the needs of current research projects. During the period of nine years, which is how long it took to carry out digitization, recordings from 521 tapes were transferred to 659 CD-s.

Tapes were PVC based; their state of preservation was dependent on recording date, conditions of their use and storage. Older tapes were brittle because of the base PVC layer decay. This required great care in tape handling, and adjustment of tape tension thru reproduction. Some tapes had signs of mould and they needed to be cleaned, some had damaged splices that were repaired. State of preservation of original tapes, and restoration procedures carried out, were not documented. In the transfer of sound recording, according to Mr. Gospodnetić, there was procedure of "cleaning of sound content", "removal of useless situa-

[15] Gospodnetić, Vito. Digitalizacija i restauracija fonoteke Instituta (…), p. 5

[16] Ibid., p. 3

tions" – someone entering the room, coughing, etc. These interventions were also not documented.
Digital record is made in resolution of 44, 1 kHz/16 bit; it is stored on audio CD, in .cda digital sound format. Two identical copies in resolution, format, and type of carrier were made – one archival, and one access copy. They are both stored in the same place. All recordings are catalogued in a computer database made in File Maker Pro application. Description contains basic data about author of recording, location were recording was made, year, and title that broadly defines the content, and detailed list of recorded content. Information about original tape, recording equipment and transfer procedures is not mentioned. Description is not based on any accepted and recommended metadata scheme.
Digitization of sound collection of IEF was never formally defined as an actual project. It was carried out with difficulties in financing; it was partially financed by Institute itself, and for some time with the support from Ministry of culture, that ended in 2003. At this moment, IEF finances on its own digitization of tapes when they are needed in current projects. It should be noted, that IEF initiative of sound recording digitization was one of the first in this part of Europe. Although some solutions may be problematic in the long-term, this was, to a degree, a ground-braking project of small unprofitable cultural institution that is worth mentioning. If project continues, it would be necessary to revise its current basic principles, and solutions according to internationally accepted standards should be defined.

Phongrammarchiv of Austrian Academy of Sciences
Sound collection of Vienna Phonogrammarchiv has in its possession 43074 recordings on acetate, PVC and polyester magnetic tapes of different speeds. According to Dr. Gerda Lechleitner, a curator of historical collection, due to good conditions and full functionality of tapes and reproduction equipment, priorities were not defined, and there is no-long term plan for digitization of collections stored in the archive. Nevertheless, archive begun with digitization in 1995, when some amount of magnetic tapes has been converted to R-DAT. Systematic digitization started at year 2000. Archive digitizes parts of its one collection, and collections of other archives and institutions.[17] From the beginning, digital resolution of 96 kHz/24 bit was used. Archive uses WAVE format, and digitization is performed according to IASA TC04 standards.

[17] This year, with support of EU Interreg Programme, archive completed extensive project of digitization of endangered collection of Alfred Quellmalz, containing folk music recordings from South Tyrol, from period 1940-1942. For realization of this project, prestigious Jikji Prize is awarded to Phonogrammarchiv. Prize is given every two years to promote objectives of UNESCO Memory of the World Programme: preserving and digitizing humanity's documentary heritage. http://www.pha.oeaw.ac.at/phawww/news_e.htm

According to standards, original archive sound document is not modified in any case, and it is real copy of original. There were no signal enhancements or denoising procedures applied, except physical restoration of original sound carrier. Archival sound records and complete accompanying documentation is stored in two identical copies on LTO3 magnetic tapes. For storing, access and manipulation, combination of servers, LTO Juke Box and manual manipulation of LTO tapes is used. User copies are available online in MP3 format, 192Kb/s, and stored on server.
For metadata, Dublin Core Metadata Element Set is used, supplemented by IASA Cataloguing Rules, extending basic DC scheme.

Conclusion

Preserving field sound recordings of folklore music, customs and telling, means to preserve valuable testimonies about history of specific nation, and its traditional culture. International organizations as IFLA, IASA (International Association of Sound and Audiovisual Archives), AES (Audio Engineering Society), and numerous others projects and programmes whose objective is preservation of audiovisual materials and heritage, have defined standards and recommended procedures in this field. This paper presents the analogue magnetic tapes digitization practice in two institutions that have two things in common: scientific research of folklore, and significant collections of field sound recordings on magnetic tapes. However, they are different by their primary function (archive / research institute), staff profile, and most of all, financial possibilities.
Project of the *Institute for ethnology and folklore research* was at that point in Croatia, in year 1995 when it was initiated, in a way pioneering attempt to save unique and valuable sound collection, an already by decay endangered tapes. Project was not conducted by staff educated in field of sound recordings preservation and archiving, but by audio engineer, highly experienced in sound processing for radio, film and television. Digitization procedure was not in all segments based on expert standards for analogue sound materials digitization, but digitized recordings stored on CD-s reduced the possibility of damage to original carrier, and provided easier access and use of these recordings. Because of undocumented modifications in transfer of analogue sound signal, fidelity of digital record to original is questionable. Primary goal – preserving original sound information and all of its segments was thus not entirely achieved. Question is, will the state of preservation of original tapes permit, in some financially more favourable times, repeating of digitization procedure according to expert digitization and sound preservation standards.
On the other side, magnetic tapes digitization procedure conducted in *Phonogrammarchiv* in all aspects follows the standards of expert organizations mentioned before. Archive, by its definition, has different priorities then a research institute; it is supported by his home institution – Austrian Academy of Sciences, and has established sources of finances. In this sense significant is in-

volvement in international and local projects (EU projects TAPE, DELOS, AES Standards Committee, and Subcommittee on Audio Preservation). Archive has highly trained experts for every aspect of its work, from field research, to preservation and reproduction of obsolete sound carriers, international standards, technical and IT support. These are all necessary conditions for successful conducting of complex projects.

Field recordings on magnetic tapes are here taken as an example of material with specific function and documentary value. However, they are only a small part of comprehensive sound heritage stored on different formats, in depositories of libraries and collections that collect sound materials – National and University Library, Croatian State Archive, Archive of Croatian Radio and unknown number of private collections. In Croatia, only recently certain interest is shown regarding the preservation of sound heritage, primary by interested enthusiasts. Necessary steps that should be taken by expert societies, to change current practice of marginalizing the problem of sound heritage preservation, are:

- To encourage persons in highly responsible positions in institutions that are collecting sound materials, to recognize its significance as a part of cultural heritage.
- Education of experts in the field of preservation of sound materials
- Defining analogue sound materials digitization projects respecting international standards and recommendations
- Ensuring financial support from responsible state administrative bodies and international projects and programmes

In case that significant changes in current practice of sound recordings "preservation" thus not occur, there is a serious possibility that, because lack of interest and knowledge, a significant part of cultural heritage will be irreversibly lost.

References

Bezić, Jerko. Etnomuzikološka i etnokoreološka djelatnost Instituta od kasnih četrdesetih do osamdesetih godina. // Narodna umjetnost. 35/2 (1998) ; p. 21-48 http://www.ief.hr/page.php?id=364&lang=hr

Boston, George. Survey of endangered audiovisual carriers. Paris : UNESCO , 2003. http://portal.unesco.org/ci/en/file_download.php/dfb2ad0ec5e386a5040cf35fc58f029bSurvey+Report.pdf

Bradley, Kevin. Risks associated with the use of recordable CDs and DVDs as reliable storage media in archival collections – strategies and alternatives. Paris : UNESCO , 2006. http://unesdoc.unesco.org/images/0014/001477/147782E.pdf (28/08/2007)

Brady, Erika. A spiral way : how the phonograph changed ethnography. Jackson, MS : University Press of Mississippi , 1999.

Capturing Analog Sound for Digital Preservation : Report of a Roundtable Discussion of Best Practices for Transferring Analog Discs and Tapes. Washington D. C. : Council on Library and Information Resources, Library of Congress, 2006. http://www.clir.org/pubs/reports/pub137/pub137.pdf (25/07/2007)

Cohen, Elizabeth. Preservation of audio. // *Folk Heritage Collection in Crisis.* Washington D. C. : Council on Library and Information Resources, 2001. http://www.clir.org/pubs/reports/pub96/pub96.pdf (25/07/2007)

Fleischhauer, Carl. The Library of Congress Digital Audio Preservation Prototyping Project. 15.12.2003. http://www.arl.org/preserv/sound_savings_proceedings/Digital_audio.shtml (20/08/20007)

Gospodnetić, Vito. Digitalizacija i restauracija fonoteke Instituta za etnologiju i folkloristiku : projektni zadatak. Radna verzija teksta. 07.11.1996.

IASA-TC 03 The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy, Version 3, December 2005.
http://www.iasa-web.org/IASA_TC03/TC03_English.pdf (01/09/2007)

IASA Task force to establish selection criteria of analogue and digital audio contents for transfer to data formats for preservation purposes. International Association of Sound and Audiovisual Archives , 2004. http://www.iasa-web.org/taskforce/taskforce_00.htm (10/08/2007)

IFLA Guidelines for digitisation projects. March 2002. http://www.ifla.org/VII/s19/pubs/digit-guide.pdf (25/08/2007)

Knight, Gareth, McHugh John. Preservation handbook : digital audio. 25.07.2005. http://www.ahds.uk/ (20/07/2007)

Lechleitner, Gerda. Ethnomusicology and the historical collections of the Phonogrammarchiv.
14.02.2004. http://ethnomusicologie.free.fr/jetu-docs-pdf/jetu03-lechleitner.pdf (01/09/2007)

Lechleitner, Gerda. Magnetic tape digitisation. Privatna poruka. (30.08.2007.)

Liebl, Christian. Magnetic tape digitisation. Privatna poruka. (03.09.2007.)

Lozica, Ivan. Izvan teatra : teatrabilni oblici folklora u Hrvatskoj. Zagreb : Hrvatsko društvo kazališnih kritičara i teatrologa, 1990.

Ord-Hume, Arthur W. J. G. [et al.]. Recorded Sound.. // *Grove's Dictionary of Music and Musicians.* Oxford : Oxford University Press , 2001. Vol. 6., p. 7-37

Paton, Christopher Ann. Preservation re-recording of audio recordings in archives: problems, priorities, technologies, and recommendations // *The American Archivist.* 61(1998), p. 188-219

Pegg, Carolle [et al.]. Ethnomusicology. *Grove's Dictionary of Music and Musicians.* Oxford : Oxford University Press , 2001. Vol. 8., p. 367- 403

Smith, Abby. Why Digitize? 02.1999. http://www.clir.org/PUBS/reports/pub80-smith/pub80.html (01/08/2007)

Stančić, Hrvoje. Proces digitalizacije. // *Informacijske znanosti u procesu promjena.* Zagreb : Filozofski fakultet, Zavod za za informacijske studije Odsjeka za informacijske znanosti, 2005. p. 45-83

Wallaszkovits, Nadja. Magnetic tape digitisation. Privatna poruka (07.09.2007.)

Zakon o arhivskom gradivu i arhivima. Zagreb : Narodne novine , 105(1997) http://www.nn.hr/clanci/sluzbeno/1997/1617.htm

Digitalni audiovizualni arhiv u teoriji i praksi u informativnom programu HTV-a

Snježana Mihačić
HRT
Prisavlje 3, Zagreb, Hrvatska
smihacic@hrt.hr

Sažetak

Digitalizacijom informativnog programa HTV-a došlo je do promjene i u dosadašnjem načinu arhiviranja audiovizualnog materijala. Pristup poslužitelju (serveru) doveo je do neovisnijeg rada arhivista kao i neposredniji pristup uradcima novinara – sirovinama (snimljenim i još nemontiranim materijalima) i polufinalima (prilozima bez potpisa i sa svim tonovima). To je omogućilo i kreativniji pristup arhiviranju kao neizostavnom dijelu produkcije koje se nalazi na njenom kraju ali često i na početku.

Ključne riječi: digitalizacija, arhiv, informativni program, poslužitelj

Uvod

Iza blještavila produkcije i ekskluzivnih javljanja novinara, uloga audiovizualnog arhiva često je potisnuta u drugi plan, a njegova uloga podcijenjena. Do sada se to najbolje očitovalo u nedovoljnoj suradnji novinara i arhivista te nepovjerenju da će materijali koje predaju arhivu biti sačuvani i ponovno dostupni. Tako su novinari i realizatori Informativnog programa stvarali svoje "privatne" arhive na štetu "zajedničkog" arhiva kojega su na taj način ostavljali bez vrijednih materijala, najčešće sirovina. Suvremeni život nametnuo je brži ritam u svim čovjekovim aktivnostima te je krilatica "vrijeme je novac" često korištena kao pravilo. Utrka TV kuća za bržim i kvalitetnijim plasiranjem informacija znači moć, gledatelje i novac. Malo tko misli što će biti sutra i što će biti s informacijom kada prođe njena vrijednost koju joj je donijela aktualnost. Arhiviranje, odnosno primjerena obradba i čuvanje audiovizualnog (AV) materijala uvijek je ovisilo i pratilo tehnologije stvaranja i emitiranja programa. Tako je i njihova digitalizacija donijela promjene te nametnula i unaprijedila rad AV arhiva pa tako i arhiva Informativnog programa HTV-a.

Digitalizacija informativnog programa HTV-a

Sve medijske kuće današnjice ovise o automatizaciji. Ni Hrvatska radiotelevizija nije izuzetak, stoga je u izradi dnevnih informativnih emisija počela s kori-

štenjem digitalnog sustava za prikupljanje, izradu i emitiranje vijesti. Radi se o iNEWS NRCS (News Room Computer Systems) sustavu, produktu tvrtke Avid, koji pokriva sve potrebe modernog redakcijskog sustava izrade informativnog programa. iNEWS NRCS u sebi objedinjuje sve radne procese izrade vijesti; od prikupljanja najnovijih agencijskih vijesti, njihovog kategoriziranja i pohranjivanja u centralnu bazu podataka, pisanja i stvaranja priče, organizacije programa vijesti, do kreiranja rundown liste i kontrole emitiranja. Nakon testiranja i prilagođavanja sustava radnim procesima uredničkih redakcija HRT-a, CS-ovi inženjeri proveli su plansko školovanje korisnika iNEWS NRCS sustava ovisno o djelatnom profilu svakog korisnika. Program školovanja bazirali su na osnovu predavanja koja su pohađali u Avid iNEWS centru Engleskoj.

Promjene koje digitalizacija donosi arhivu

Osnovna i najveća promjena arhiva u digitalnom okruženju je nestanak uvriježenog koncepta arhiva kao skladišta vrpci i filmova. Nije na odmet podsjetiti koliko bogatstvo predstavlja arhiv. Tu se prije svega misli na neprocjenjivu socijalnu, kulturnu i povijesnu vrijednost nacionalnog audiovizualnog naslijeđa zemlje kao i na osiguravanje izvora koji će moći obogaćivati programe svojom objektivnom, dokumentarnom i subjektivnom, nostalgičnom vrijednošću. Arhiv će sve više postajati poveznica produkcije i distribucije i preuzimati aktivniju ulogu od one koja mu je bila namijenjena kao spremištu materijala čije je vrijeme prošlo. U digitalizaciji arhiva trebamo imati na umu kako će još jedno izvjesno vrijeme arhivi živjeti u hibridnom svijetu između analogne i digitalne tehnologije – u istom trenutku imamo materijale koji su nastali korištenjem nove, digitalne tehnologije i analogne zapise koji se čuvaju na medijima koji zahtijevaju čitače čije vrijeme prolazi. Tu se nameće i otvara pitanje migracije s analognih na digitalne medije koja je dugotrajan , organizacijski složen i skup proces. Bit digitalizacije je čuvati sadržaj a ne vrpcu, a "vječni" file se može proizvesti samo putem digitalizacije. Ovdje se izuzimaju filmski materijali za koje vrijedi pravilo "nikad ne uništavaj original". Digitalizacija omogućava pristup podacima većem broju korisnika istovremeno a također osigurava svako kopiranje "master" (izvornik) izdanja koje se neće izlagati promjeni i brisanju korisnika.

Digitalni svijet gotovo je izbrisao granicu između produkcije i arhiviranja. Tradicionalno arhiviranje smatrano je pomoćnom djelatnošću, prva postaja na koju dolaze pretraživati potrebni materijal i posljednje odredište završenih i često zaboravljenih materijala – sirovina, emisija… U idealnim uvjetima dolazi do integracije produkcije i arhiviranja, u stvarnom svijetu treba učiniti sve kako bi se produkcija i arhiviranje što više približili. Tako u lancu digitalne proizvodnje arhiv nije više na kraju lanca već se integrira u proizvodnji programa lakšim i bržim pristupom materijalu. Međutim, to podrazumijeva ne samo tehnološku prilagodbu nego i mijenjanje svijesti o aktivnoj kooperaciji između arhivista, produkcijskog osoblja i tehničke potpore. Arhivisti prestaju biti neprimjetni,

manje važni djelatnici , oni postaju "media asset managari" koji slijedom svoje profesije rukovode informacijama (pregled, odabir, obrada i trajna pohrana).

AVID i MERIDIO

Nabavom aplikacija u obliku AVID softverskih rješenja i tehničkih uređaja iste kompanije te obučavanjem svoje tehničke službe i djelatnika informativnog programa na AVID Profesional Services: Workflow Workshop – Workflow treninzima, Hrvatska radio televizija po opremljenosti i načinu obrade materijala stala je uz bok svjetskih televizijskih kuća. Akvizicija AV sadržaja prvi je i početni korak u radnom procesu digitalne produkcije informativnog programa. Snimanje video materijala koji pristižu od stranih agencija, međunarodnom i nacionalnom razmjenom, pokreće se sa stolnog računala korisnika kroz Capture Manager – Feed Ingest koji je modul aplikacija Media Browsa te se kroz Avid AirSpeed u visokoj i IPV nSpectra SDI u niskoj rezoluciji pohranjuju na središnjem prostoru za pohranu AV sadržaja. Usnimavanje materijala s terena ili arhivskog materijala putem odgovarajućih magnetoskopa, pokreće se i kontrolira korisničkim računalom kroz Capture Manager – Feed Ingest koji je također modul aplikacije Media Browsa te se na isti način putem Air Speeda i IPV encodera pohranjuje na središnji poslužitelj. Kao prvi ulazni stupanj akvizicije sadržaja nalaze se samostojeći Avis Air Speed uređaji sa svim potrebnim AV priključcima te s GigE mrežnim sučeljem prema postojećem DNR (Digital News Room) sučelju. Oni omogućuju veliki broj ulaznih kanala za snimanje (digitalizaciju) AV sadržaja u IMX50 formatu visoke rezolucije izravno na središnji Unity sustav za pohranu, a svaki od Air Speed uređaja spojen je na središnji i dijeljeni prostor za pohranu (Avid Unity Media Network) preko Port Servera PRO dok su međusobno svi uređaji spojeni na isti Gigabit Ethernet swich Cisco Catalyst 2970. Avid Airspeed uređaji su kontrolirani Avis Capture Manager aplikacijom koja omogućuje trenutno i unaprijed planirano snimanje.

Sukladno tome, danas je posao (još uvijek!) AV arhivista možda ponajmanje puko arhiviranje. Svakodnevnim pregledavanjem pristiglih materijala u projektima sirovina vijesti i dnevnika, EBU feeda (za unilateralne priloge koje dopisnici HTV-a šalju iz inozemstva), sirovina ili polufinala tjednih i dnevnih emisija (Hrvatska danas, Život u živo, Euromagazin), priloga iz dopisništava HTV-a kao i dnevnog spremišta – bina te materijala koji možda neće ni biti emitiran, odabiru se materijali koji su po zahtjevima arhivske struke vrijedni za pohranu. Potrebno je naglasiti da se ipak radi o specifičnim uvjetima rada i strukturi i potrebama krajnjih korisnika, novinara kojima treba omogućiti da iskoriste potencijale arhiva znajući točno kako i kada i znajući što je sve u domeni arhiva (upravljanje informacijama o materijalu, omogućavanje dostupnosti onoga što korisniku treba, kada treba i u obliku koji mu je najpodesniji za korištenje). Kriteriji za odabir materijala su slijedeći: sirovine, materijali od političkoj značaja što znači svi materijali koji se odnose na aktivnosti predsjednika RH i predsjednika vlade, hrvatske vojske i policije, materijali od povijesnog značaja za

Republiku Hrvatsku bilo na gospodarskom, društvenom ili kulturnom planu Međutim, uz sve što ima dokumentarističku vrijednost svjedočenja vremena i djelovanja, tu su svakodnevne potrebe i zahtjevi korisnika koji se odnose na takozvano "pokrivanje" – nadopunu novinarske priče. Tako kadrovi državnih, kulturnih, gospodarskih ustanova, kadrovi prolaznika, poplava, gužvi na prometnicama, kadrovi gradova, prirode, turističkih odredišta…nemaju arhivski ali i te kako imaju uporabni značaj u svakodnevnom kreiranju i stvaranju programa. Tu se otvara i najveće polje djelovanja arhivista u novom, digitalnom okruženju. Aktivna uloga ogleda se u suradnji s korisnicima te na uzajamnoj, obostranoj koristi. Njihovim upoznavanjem novog načina rada arhiva i našeg osluškivanja potreba korisnika stvara se suradnja koja bi trebala znatno poboljšati kvalitetu produkcije koja se odnosi na "arhivske snimke" koje se nalaze u području autorstva emitiranih priloga.

Nova tehnologija koja se primjenjuje u arhivu Informativnog programa dala je novu dimenziju ulozi AV arhivista. Oni više nisu na kraju stvaralačkog procesa, već su aktivno uključeni u stvaranje svoje baze podataka, svoga arhiva. Pristup informacijama postao je neposredniji, a odgovornost za priloge koji svakodnevno pristižu na poslužitelj (server) veća. Upravo je na arhivistu da, vođen iskustvom, znanjem i dobrom informiranošću, odabire materijale za trajno čuvanje ali i one koji neće biti emitirani, ali čija vrijednost ili važnost nisu ništa manje od onih emitiranih. Nakon odabira materijala, u Media Composeru materijali se (u formatima master clipova) nelinearnom montažom montiraju u sekvencu koja se presnimava na digitalne, IMX vrpce. Nastavak rada na obradi, indeksiranju i katalogiziranju materijala (sada već dokumenta) još uvijek nije doživio promjene jer se radi u programu MERIDIO aplikacija za unos i pretraživanje, kao i prije digitalizacije. Kao slijedeći korak predviđa se integracija AVID i MERIDIO programa što će imati doista revolucionarne posljedice za AV arhiviranje.

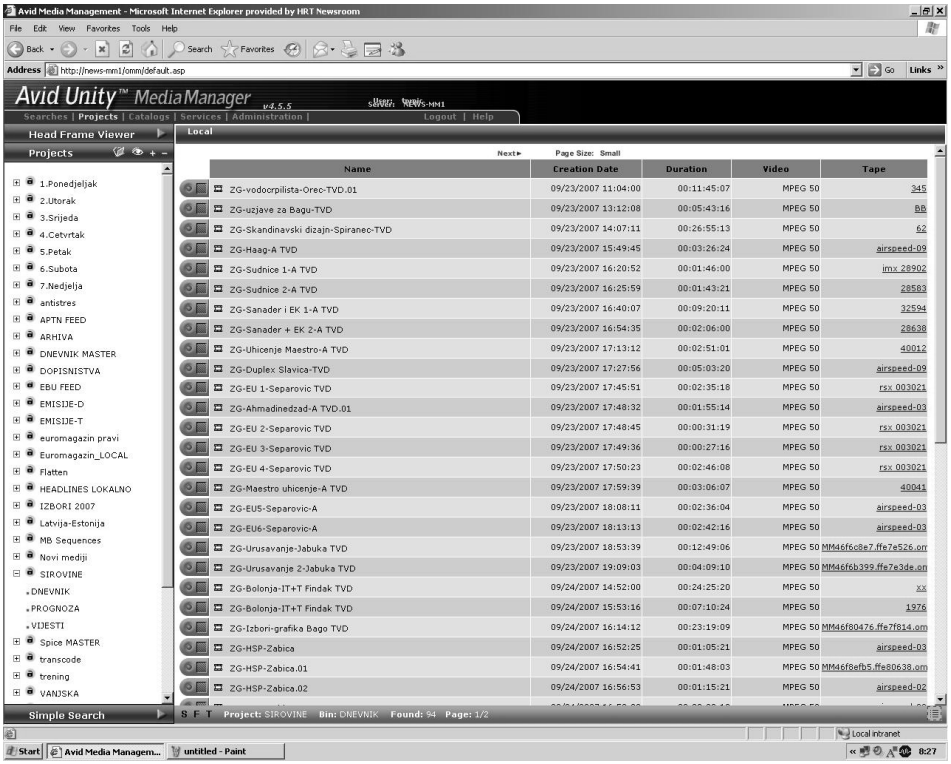
Bolje prepoznavanje uloge arhiva te bolja suradnja s ostalim čimbenicima proizvodnje programa poboljšali bi kvalitetu kako arhiva tako i same produkcije, odnosno njezina dijela koji se odnosi na uporabu arhivskih materijala. Značaj i uloga arhiva je često podcjenjenja te se zaboravlja kako bogat i dobro uređen arhiv uz stručne djelatnike predstavlja vrijedan izvor informacija i danas se smatra da je 30% tv produkcije upravo arhivskog porijekla.

On line arhiva

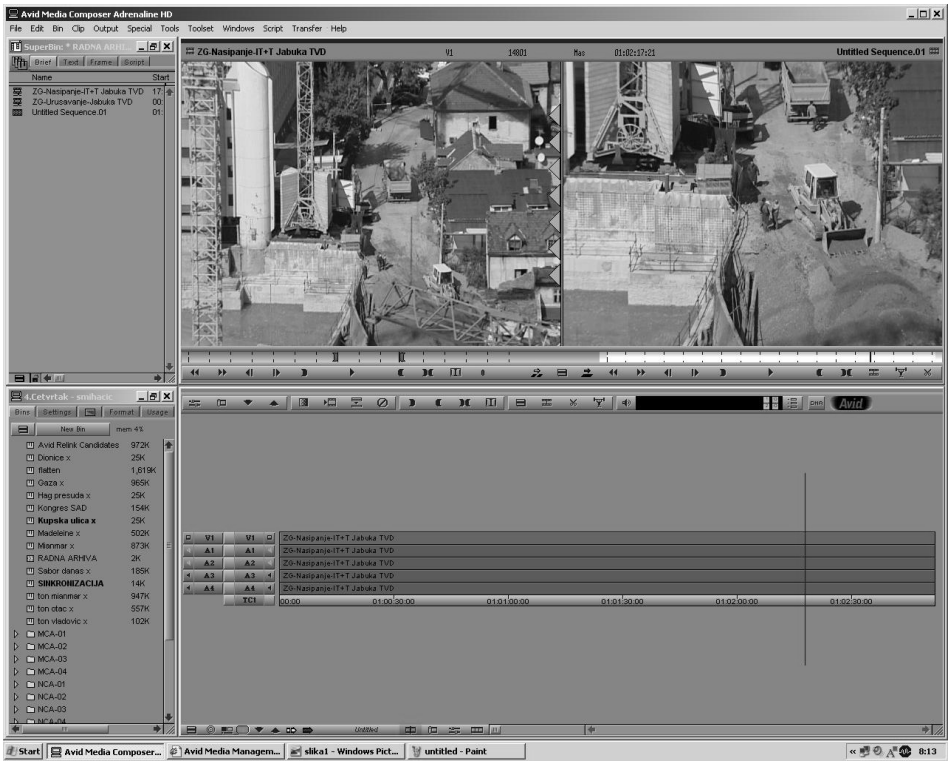
Na Unity Workspace Arhivi , koja predstavlja on line arhivu – sigurni jocker u rukavu stalno užurbanih novinara, u trajanju od 360 minuta nalaze se tematska spremišta binovi koji se ažuriraju i prilagođavaju korisnicima. Te, uvjetno rečeno mape su organizirane u suradnji s novinarima i realizatorima informativnog programa sukladno njihovim sugestijama i potrebama. Tako postoje tematski binovi:

USTANOVE (zgrade državnih institucija)
PRAVOSUĐE (sudovi)
ZDRAVSTVO (bolnice, ljekarne)
PROMET (ceste, zračne luke, kolodvori)
GOSPODARSTVO (poduzeća)
VANJSKA POLITIKA (institucije EU i UN)

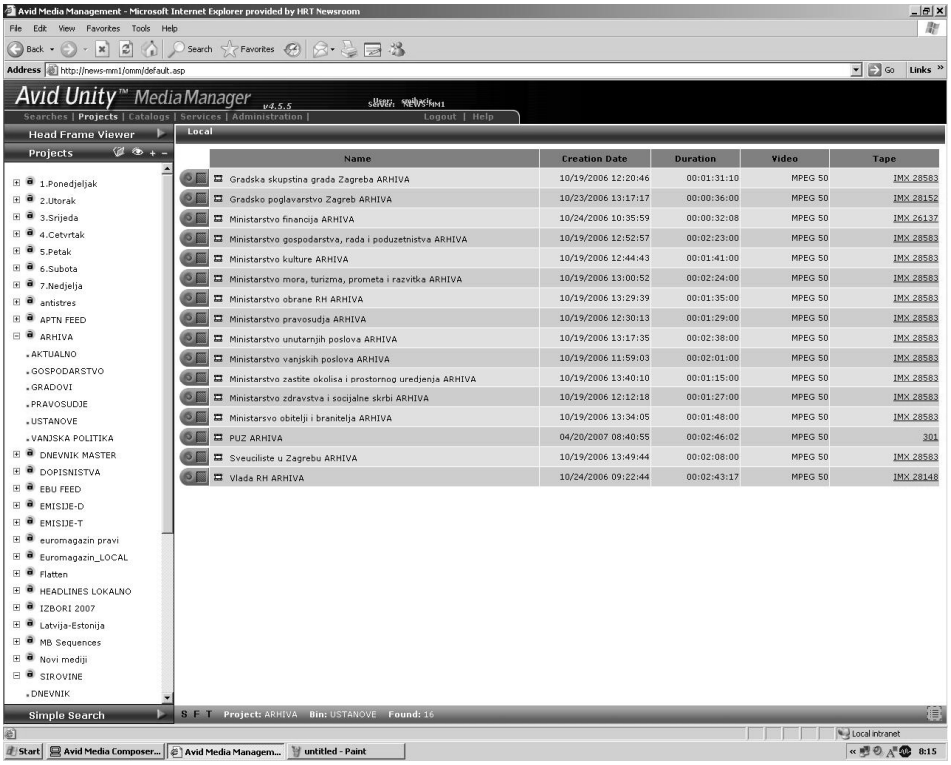
a u tijeku je formiranje mapa u kojima bi bile kulturne institucije kao i ostali materijali koji se svakodnevno ili vrlo često koriste što će osjetno ubrzati proces proizvodnje i produkcije.
Bitno je naglasiti kako se desio značajan organizacijski pomak u radu arhiva pa su tako djelatnice organizirale i vrlo uspješno provele dvodnevno snimanje državno – političkih institucija u gradu Zagrebu te inicirale snimanje i slanje materijala iz gradova u kojima su dopisništva HTV-a s kadrovima gradova, njihovih političkih, gospodarskih i kulturnih institucija, a dodatno će se voditi računa i da kadrovi odgovaraju trenutnim godišnjim dobima. Obzirom na novostečena znanja, materijali koji se nalaze u radnom prostoru arhive uvršteni su (ingestirani) s arhivskih vrpci i sa servera iz jednog, dnevnog, projekta u drugi projekt – arhivu. Kako se svi materijali koji su u Media Manageru nalaze u visokoj rezoluciji, pokretanjem "fletanja" prebacuje se u nisku rezoluciju kako bi korisnici u Media Browsu imali pristup materijalima i kako bi ga mogli oblikovati i prilagođavati. Jednako tako, sirovine koje novinari donesu na svojim radnim vrpcama kako bi ih pohranili u arhivi, a kako se ne bi dodatno opterećivao ingest – mjesto na kojemu se materijali stavljaju na server bilo s vrpci ili iz dopisništava HTV-a odnosno montaže (u slučaju da se presnimava na arhivsku vrpcu) arhivisti digitalnog arhiva IP- a u stanju su lokalno na server (u dnevni projekt) spremiti sirovinu, ukoliko je potrebno montirati je te snimiti na arhivsku vrpcu odnosno spremiti u Workspace arhivu. Od studenog 2005. godine do kolovoza 2006. godine ukupno je odabrano i arhivirano 6500 priloga sa servera (sirovine, polufinali, unilaterali) što iznosi oko 20000 minuta.



Slika 1. Zaslon servera s prikazom otvorenog projekta Sirovine/ Dnevnik pristiglim materijalima u periodu od tri dana



Slika 2. Zaslon servera s prikazom nelinearne montaže odabranog materijala iz master clipova u sekvencu



Slika 3. Zaslon servera s prikazom otvorenog poglavlja Ustanova online Arhive

Zaključak

Nova tehnologija sveprisutna, moćna digitalizacija došla je i do nas obećavši mnogo dobra onome tko je savlada, usvoji i iskoristi. Vidjeli smo dokud se stiglo u digitalizaciji arhiva Informativnog programa – što je novoga donijela i što je ostalo isto. Ukoliko bi se suradnja arhivista, novinara, realizatora nastavila u nagovještenom pravcu, značaj koji bi se ostvario omogućio bi kvalitetnije i brže kreiranje i realizaciju cjelokupnog programa te bi se i početničke greške i nedostaci (koji će se nastojati ukloniti) mogli brzo zaboraviti. Korak koji je Arhiv IP-a učinio zahvaljujući svojim "ljudskim resursima" bez kojih nijedna tehnologija ne bi bila moguća značajan je jednako za nas kao i za Informativni program čija je dobrobit i kvaliteta naš konačni cilj.

Literatura

Digital newsroom: thematic visit; Visit to DR News, Copenhagen, Denmark, 14-15. november 2005.

Acquilian Stefab, Ching Yee Lai, Van Dessel Dalida: Reports om speeches of the media manegament seminar "Changing scenaries, changing roles", Amsterdam 2004.

Dooley, Brid: Digital Archive "Mision posible", EBU, Geneva, December 2005.

Potpuno digitalizirana produkcija vijesti na HRT-u; CS završio instalaciju digitalnog redakcijskog sustava vijesti (NRCS) na HRT-u, http://www.cs.hr/o_nama/iNEWS_HRT.asp, 25.4.2007.

Archives in digital Broadcasting, izvještaj interdisciplilnarne arhivske grupe EBU 2002. godine

Digitalizacija informativnog programa, http://www.tportal.hr/tehnologija/mixer/page/2005/09/05/0394006.html, 12.1.2007.

Alicia Conesa, TVC – *Digition. Digital archive in Television de Catalunya*. September 2003 Conference, http://www.fiatifta.org/restricted/Antalya2002/dejong.doc,18.12.2006.

Napomena

Ovaj rad je dio stručnog rada "Audiovizualni arhiv Informativnog programa HTV-a u digitalnom okruženju" koji je prihvaćen za objavljivanje u Arhivskom vjesniku za 2007. godinu.

Pouzdanost podataka na optičkim medijima

Mladen Tušek
Filozofski fakultet, Zagreb
Ivana Lučića 3, Zagreb, Hrvatska
mladentusek@yahoo.com

Sažetak

U ovom radu biti će riječi o sve većoj količini informacija koje okružuju pojedinca te rastuće potrebe arhiviranja istih. Da li je moguće uspješno sačuvati informacije koje nas preplavljuju eksponencijalnim rastom? Problematika čuvanja informacija na određenom mediju koji smatramo superiornim u određenom trenutku aktualna je kroz povijest civilizacije pa sve do danas. Nasumce izabran primjer crteža sa zida špilje Altamire u Španjolskoj pokazuje da je informaciju moguće sačuvati tisućama godina. Navedeni primjer svjedoči o uspješnom očuvanju informacija. Promjene medija na koji bilježimo određene informacije uvijek ispočetka stavlja pred nas izazov kako uspješno sačuvati zapis tijekom dugog vremenskog perioda. U današnjem informatičkom dobu usredotočiti ćemo se na problem očuvanja digitalne informacije na digitalnom mediju. Optički medij postaje dominantan za korištenje i čuvanje arhivskog gradiva, no da li možemo gradivo na optičkom mediju sačuvati 50, 100 ili više godina? Koji su to ključni problemi i prepreke koje je potrebno savladati da spriječimo gubitak arhivskog gradiva ili informacije uopće na optičkom mediju? Kakvi su nam alati i metode dostupni da unaprijedimo očuvanje informacija na CD ROM, DVD ROM ili HD DVD ROM disku? Ovo će biti jedan pokušaj da se opiše uspješno čuvanje informacija i znanja na optičkom digitalnom mediju.

Ključne riječi: Arhiva, informacija, optički medij, pouzdanost zapisa, CD, DVD

Uvod

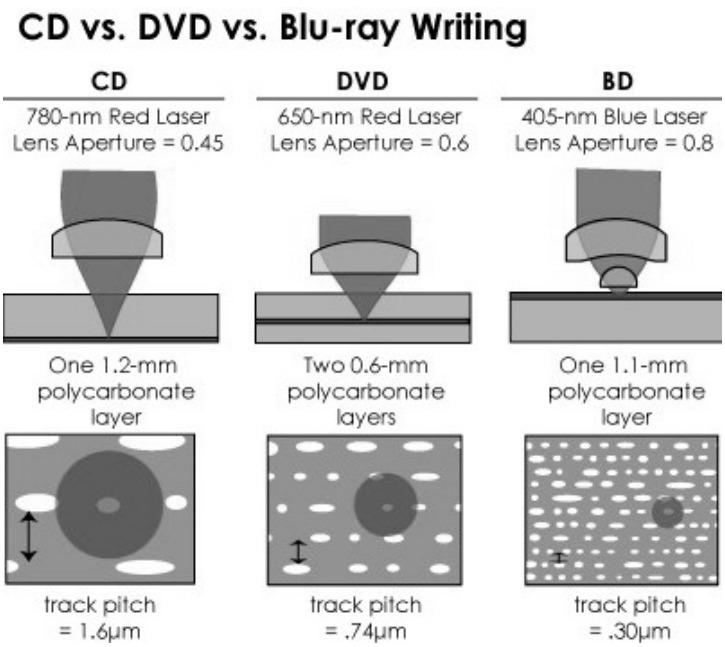
Optički mediji svoju javnu promociju doživljavaju kroz audio CD (compact disc) promjera 12 cm. Nekoliko godina nakon toga (u vrijeme IBM AT386 računala), na tržište dolazi računalni CD koji donosi 650 MB kapaciteta. U upotrebi su bili ili jesu i drugi promjeri i kapaciteti, no danas je optički disk 12 cm dominantan, te se i studije koje se bave projekcijama razvoja optičkog medija u budućnosti mahom baziraju na istoj vanjskoj dimenziji od 12 cm.
Sam optički medij predstavlja idealno sredstvo za dugoročnu pohranu i arhiviranje podataka. Vrlo niska cijena kako medija tako i uređaja za zapisivanje i či-

tanje te fizička otpornost na vanjske utjecaje daju mu prednost u odnosu na dosad navedene medije. Trajnost u upotrebi mu je teoretski beskonačna, budući da nema mehaničkog habanja tijekom rada – cijeli proces čitanja i pisanja odvija se optički putem laserske zrake određene valne duljine. Postoji dosta velika mehanička otpornost i primjerena elastičnost samog diska što olakšava korištenje i snižava troškove skladištenja. Klimatski uvjeti koji se moraju osigurati za čuvanje lako su ostvarivi bilo u arhivima ili okolini gdje se koriste podaci s medija. Postoje dva polja koja je potrebno unaprijediti kod ovog tipa medija za masovnu pohranu podataka: kapacitet koji je moguće ostvariti na takvom 12 cm mediju te kemijska stabilnost površine u koju se upisuju podaci na disku.
Razvoj CD medija od početnih 650 (700) MB nastavio se prema DVD disku. DVD disk također promjera 12 cm posjeduje kapacitet od 4.7 GB do 9 GB po mediju. Uređaji koji čitaju i pišu na DVD medije potpuno su kompatibilni sa CD medijima. Zbog veće gustoće zapisa na DVD mediju, laserska zraka u uređajima je manje valne duljine. Kod DVD medija pojavljuje se jedna tehnička karakteristika koja dodatno omogućuje pohranu gotovo dvostruke količine podataka na medij koji standardno posjeduje mogućnost zapisa definirane gustoće podataka. Radi se o zapisivanju digitalnih podataka u slojevima, u konkretnom slučaju o dva sloja. Gornji sloj je polu proziran što omogućuje optici laserskog pisača odnosno čitača da promjenom fokusa piše ili čita podatke sa odgovarajućeg sloja, čime dobivamo privid dva optička diska. Potrebno je napomenuti da u slučaju zapisa važnih podataka moramo biti oprezni pri izboru ovog načina rada. Dobitak u količini podataka je znatan, no potrebno je imati na umu da je pouzdanost zapisa ipak nešto niža.
Razvojem optičkih medija dolazimo do HD DVD i BluRay diskova. Trenutno predstavljaju međusobno nekompatibilne standarde, no postoje naznake da bismo ih uskoro mogli koristiti u jedinstvenim uređajima koji bi pisali i čitali oba standarda. Kapaciteti im se kreću za HD DVD 15 i 30 GB, te za BluRay 25 i 50 GB. Iz navedenih kapaciteta se vidi da je i kod njih dozvoljeno spremanje podataka u dva sloja (postoje informacije da HD DVD postoji u inačici od tri sloja što čini ukupni kapacitet od 45 GB). Kroz određeno vrijeme potrebno da tehnologija sazrije i cijena postane prihvatljiva predstavljat će vrlo dobru opciju za pohranu velikih količina podataka. Na slici 1. zorno je predočen razvoj optičkih tehnologija za pohranu podataka.

Razrada

Problematika dugoročnog očuvanja podataka na optičkim medijima

Kada govorimo o problematici dugotrajnog čuvanja zapisa na optičkom mediju, možemo reći da su tu prisutni problemi fizikalnog i (ili) kemijskog oštećivanja.
Fizikalno oštećivanje može nastati uslijed loma diska, ogrebotina gornje ili donje strane diska, rupture diska te oštećenja ili uništenja diskova izazvanih visokim temperaturama ili vatrom, te posebno izlaganjem intenzivnijem sunčevom zračenju.



Slika 1.

Kemijsko oštećenje može biti izazvano prolijevanjem kemijski agresivnih supstanci po samoj površini diska sa vanjske strane. Kemijski sastav plastike od koje se sastoji veći dio optičkog diska je vrlo stabilan te je otporan na vodu i veći broj blažih kemikalija što olakšava čuvanje i korištenje samih medija.

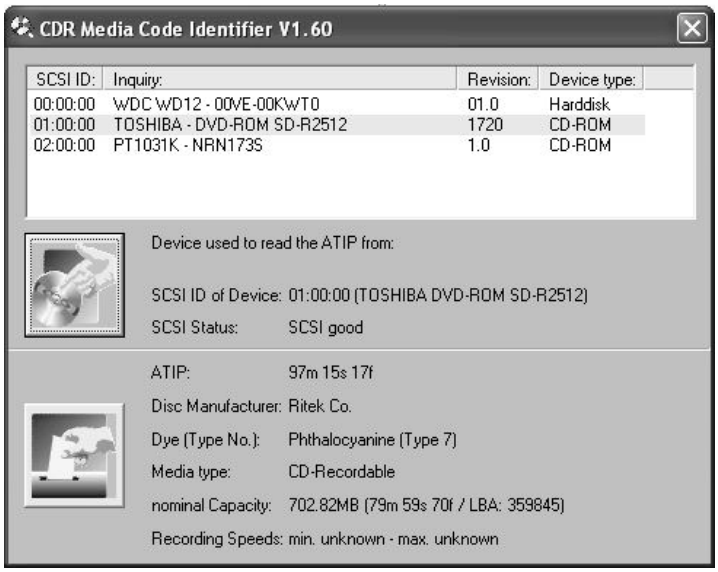
Najveći problem kod dugoročnog čuvanja podataka na optičkim medijima predstavlja upravo kemijska nestabilnost sloja diska u koji se laserom upisuju podaci. To predstavlja najveći izazov proizvođačima optičkih diskova kao i ustanovama koje su zadužene za dugoročno čuvanje elektroničke građe.

Poznato je da proizvođači samih optičkih diskova koriste različite kemijske sastave koji služe za laserski upis podataka na sam disk. U svijetu postoji velik broj proizvođača optičkih diskova, no malen je broj proizvođača kemijske podloge za upis podataka. Tako dolazimo do situacije da kemijska postojanost diska ovisi o kemijskom sastavu sloja koji je proizvela neka tvrtka te o kvaliteti implementacije tog sloja u sam disk. Isto tako, više tvrtki proizvođača diskova koristi istu kemijsku podlogu, kao što i jedna tvrtka proizvođač diskova može za različite modele optičkih diskova koristiti različite kemijske podloge. Važno je napomenuti da su kemijske podloge koje se koriste za CD ROM diskove u odnosu na DVD ROM odnosno HD DVD ili BluRay diskove različite. Kao najstabilnija kemijska podloga pokazao se "Phthalocyanine". Slijedeća važna karika za postojanost zapisa na optičkim diskovima je reflektivna podloga koja se nalazi ispod kemijske podloge, te osigurava pravilno očitavanje zapisanih podataka sa površine optičkog diska. Najstabilnija je zlatna podloga, dok srebrna

podloga daje nešto slabije rezultate. Prilikom arhiviranja velikih količina podataka biti će nam interesantno da znamo koju kemijsku podlogu koristi optički disk na koji upisujemo podatke. Da bismo to saznali moramo koristiti neki od programskih alata koji će biti u stanju pročitati koja se kemijska podloga koristi u dotičnom optičkom disku.

Implementacija programskih alata za identifikaciju i analizu

Koristan program koji daje informacije o određenom mediju je “CDR Media Code Identifier”. Pomoću njega ćemo brzo doznati tko je stvarni proizvođač nekog optičkog medija, te koju kemijsku podlogu zapravo nalazimo u disku.

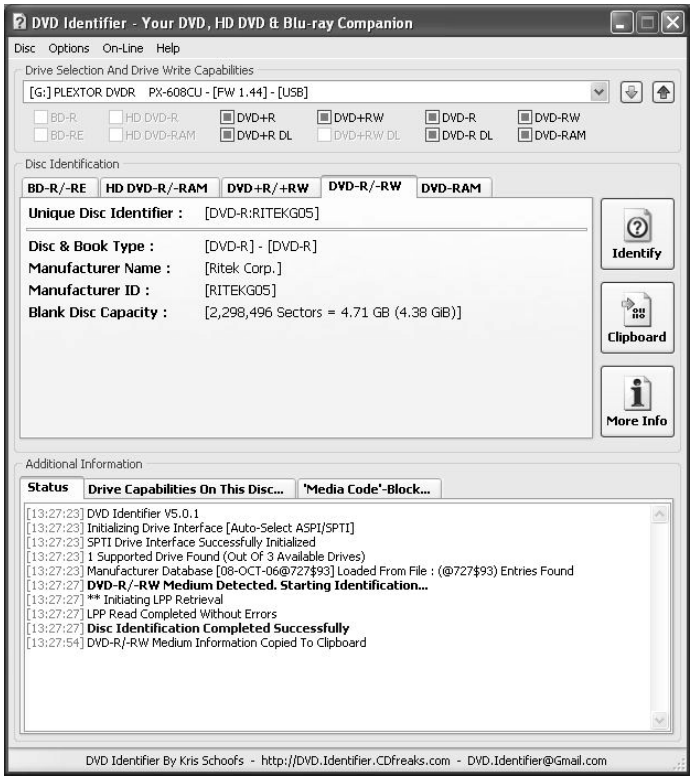


Slika 2.

Na slici 2. vidi se rezultat ispitivanja optičkog medija. U optički pogon umetnut je disk tvrtke “Traxdata”, no na slici se vidi da je proizvođač zapravo “Ritek Co.”. Isto tako možemo vidjeti naziv kemijske podloge na disku: “Phthalocyanine (Type 7)”. Prilikom odluke na koji medij ćemo upisivati podatke ove informacije mogu i trebaju biti od presudnog značaja.

Kod dugoročnog skladištenja optičkih medija dolazi do starenja kemijske podloge, te će to sigurno dovesti do gubitka podataka što može biti i nepovratno. Pitanje na koje nitko ne zna odgovor je ono najvažnije – kada će se to dogoditi. Postoje različite prognoze proizvođača medija koje se kreću u rasponu od 10 pa do 100 godina. Naravno da se u takve informacije ne možemo osloniti u procesu arhiviranja elektroničkog materijala, te nam preostaje da sami poduzimamo određene mjere koje će podići nivo sigurnosti pri arhiviranju podataka na optičkim medijima.

Slijedeći koristan program koji možemo koristiti je “DVD Identifier”.

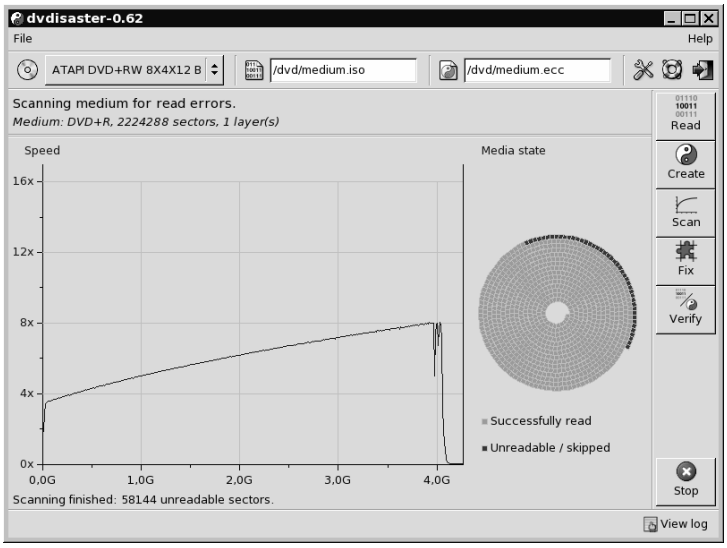


Slika 3.

Kako se vidi na slici 3., pomoću ovog programa možemo analizirati BluRay, HD DVD kao i obične DVD medije. Program nam pruža više korisnih informacija o optičkom mediju koji želimo koristiti. Korištenje ovog i sličnih programa predstavlja vitalni korak prilikom izbora optičkih medija koje planiramo koristiti za dugoročno čuvanje podataka. Trgovački nazivi, reklamne poruke, pa čak i informacije na internetskim stranicama proizvođača optičkih medija redovito ne daju važne podatke za donošenje odluke o izboru medija.

Implementacija programskih alata za provjeru i povrat podataka

U slučaju da dođemo u situaciju da nismo u mogućnosti standardnim postupcima pročitati dio sadržaja optičkog medija, koristit ćemo neki od programa koji su specijalizirani za analizu i oporavak podataka sa optičkih medija. Postoji više programa te namjene, a ja bih ovdje izdvojio “DVDisaster”.



Slika 4.

Slika 4. prikazuje analizu optičkog diska. U desnom dijelu je grafički predočena pozicija problematičnih sektora (ne nužno i nepovratno izgubljenih). U lijevom dijelu slike predočena je brzina čitanja. Upravo je brzina čitanja medija izvanredno koristan podatak pri procjeni stanja podataka na optičkom mediju. Kako se vidi na slici, krivulja brzine kreće od neke početne, zatim približno linearno raste kako se laserska glava sa optikom približava rubu diska u optičkom uređaju, no pri kraju se javljaju nepravilnosti koje opisuju drastičan pad brzine čitanja medija. Takav očiti diskontinuitet nam daje do znanja da je došlo do problema u čitljivosti podataka na disku. Rezultat može biti trojak:

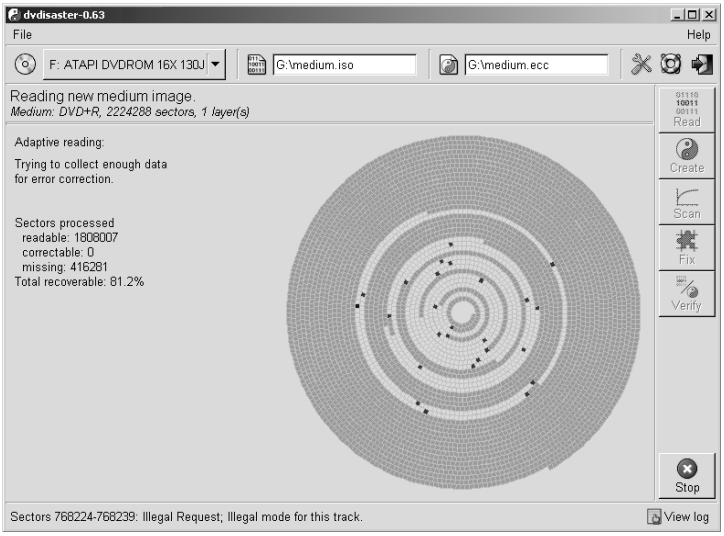
- brzina čitanja je opala na nekim dijelovima diska, no ne postoje loši sektori – potrebno je žurno kopirati podatke na drugi medij
- brzina čitanja je opala na nekim dijelovima diska, te postoje loši sektori – potrebno je žurno pristupiti procesu oporavka integriteta podataka te kopirati podatke na drugi medij
- brzina čitanja je opala na nekim dijelovima diska, te postoje loši sektori – potrebno je žurno pristupiti procesu oporavka integriteta podataka, no dijelovi medija su toliko oštećeni da je dio podataka nepovratno izgubljen

Brzina čitanja kod ispravnih medija je približno linearna ili se mijenja u pravilnim koracima (obično tri). Uzrok promjena ovisi o uređaju koji čita podatke sa optičkog diska, uređaj može čitati konstantnom obodnom brzinom ili konstantnom kutnom brzinom. U slučaju da uređaj nije u stanju ispravno pročitati neki dio diska, smanjiti će brzinu te pokušati ponovo izvršiti proces čitanja ali sada manjom brzinom. Takve promjene brzine uzrokuju nepravilan oblik krivulje što

nam je pouzdana informacija da su se počeli pojavljivati problemi sa podacima na disku uzrokovani iz dva razloga:

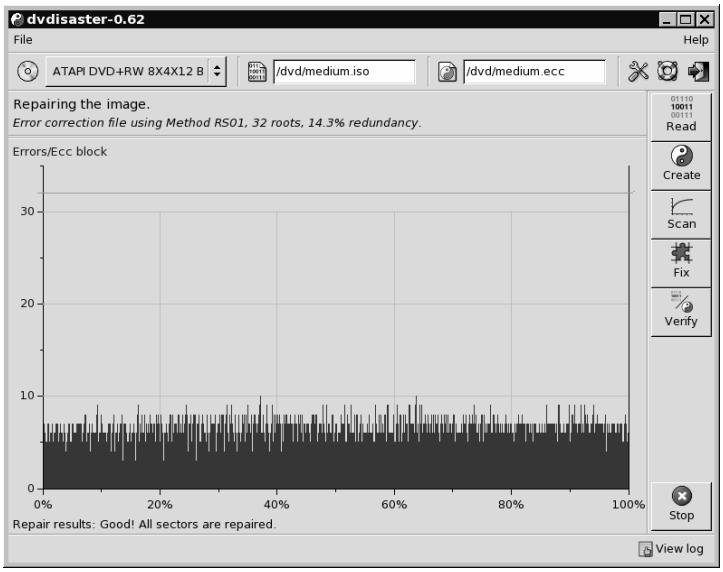
- kemijsku podlogu je zahvatio poodmakli proces starenja, za što možemo pretpostaviti da je pokazatelj stanja i ostalih optičkih medija istog proizvođača i tipa korištenih u arhivi,
- kemijska podloga ima po površini nejednoliku kvalitetu sloja što može indicirati loš proizvodni proces određenog tipa medija te da u arhivi možemo imati medije koji su nečitljivi kao i one u dobrom stanju, a sve od istog proizvođača i istog tipa.

Ovaj posljednji slučaj može se u većoj mjeri izbjeći prilikom samog procesa zapisivanja podataka na medij koristeći se metodom verifikacije i provjere čitljivosti u drugom optičkom uređaju.



Slika 5.

Na slici 5. Vidimo proces čitanja podataka sa defektnog diska. Postoje i drugi programski alati koji omogućuju sličnu analizu stanja zapisa na disku. Smisao tih analiza je da možemo na nedvojben način, dokumentirano, utvrditi proces starenja kemijske podloge koja sadrži podatke, te na vrijeme reagirati. Rezultat će biti mogućnost očuvanja digitalnih zapisa na optičkim medijima sa visokim stupnjem sigurnosti. Korištenjem ove metodologije i programskih alata nam se ne može dogoditi da kad utvrdimo nečitljivost medija bude prekasno. U slučaju dokumentiranja i uspoređivanja stanja kemijske podloge optičkih medija potrebno je obratiti pozornost na različite karakteristike uređaja za čitanje. U slučaju zamjene odnosno korištenja drugog optičkog uređaja podaci mogu varirati iz razloga različite kvalitete optike kod različitih proizvođača uređaja te implementacije različitih algoritama korekcije grešaka unutar samih uređaja.



Slika 6.

Slika 6. Pokazuje tijek procesa spašavanja podataka sa optičkog medija. U slučaju da graf pri dnu koji je u opsegu vrijednosti 3 do 10 nadvisi liniju vrijednosti 32 u ovom slučaju, oporavak podataka neće biti moguć ovim algoritmom.
Ovako smo opisali ključne postupka s optičkim medijima u okruženju kako malih tako i velikih količina optičkih diskova.

Zaključak

Promatrajući problematiku očuvanja elektroničke građe na optičkim medijima, teško se oteti dojmu da se susrećemo sa sličnim poteškoćama kao u počecima čuvanja papirnate građe.
Za pretpostaviti je da će nam ispravni optički mediji uspješno sačuvati podatke nekih 5 do 10 godina, no da li se možemo pouzdati u to? Da bismo bili sigurniji u uspjeh očuvanja elektroničke građe, potrebno bi bilo provesti određeni broj koraka kako prilikom pisanja podataka tako i tijekom čuvanja građe. Također je moguće da imamo kvalitetne medije sa stabilnom kemijskom podlogom pa ćemo podatke sačuvati i tridesetak godina na istom mediju. No tu ćemo se sigurno susresti sa pitanjem migracije zbog ubrzanog razvoja tehnologije. Već danas možemo usporediti medije koji su u upotrebi, i to CD ROM i dvoslojni BluRay. Odnos kapaciteta na fizički jednako velikom mediju je 1:77.

Prijedlog procesa očuvanja:

- birati kvalitetne medije i to po dva medija različitih proizvođača i različite kemijske podloge,
- zapisivanje podataka potrebno je vršiti paralelno na dva računala i sa dva različita optička uređaja (različiti proizvođači: Plextor i Pioneer, npr.),
- koristiti manje brzine zapisivanja (4x) i obavezno verifikaciju zapisanog sadržaja,
- nakon zapisivanja zamijeniti medije u optičkim uređajima te provjeriti čitljivost na cijelom mediju (brzina čitanja),
- po završetku spremanja jedne cjeline a prije brisanja izvornog materijala provjeriti čitljivost po metodi slučajnog uzorka,
- tijekom arhiviranja u pravilnim vremenskim razmacima vršiti provjeru čitljivosti (svakih 6 mjeseci) po metodi slučajnog uzorka te u slučaju problema korištenjem specijaliziranih programa osigurati kopiranje sadržaja prije nepopravljivog uništenja podataka.

Procesi starenja kemijske podloge nisu brzi, te povremena ispitivanja stanja mogu uspješno spriječiti gubitak podataka na optičkim medijima.

Podrazumijeva se da je prilikom arhiviranja i korištenja elektroničke građe potrebno poduzeti i sve ostale mjere potrebne radi očuvanja arhivske građe.

Izvori

Backup Critic. 2007. http://www.backupcritic.com/backup-media/dvd/durability.html (31.08. 2007)

Blanck, Artur G. Archival Media. 2005. http://www.delkin.com/pdf/user_guides/archivalgold_whitepaper.pdf (31.08.2007.)

CDVD info 2001. http://www.afterdawn.com/software/video_software/dvd_tools/cdvdinfo.cfm (31.08.2007.)

Digital Atlantic Corp. Ltd. 2007. http://www.cdroller.com/ (31.08.2007.)

DVD Identifier. 2006. http://dvdidentifier.cdfreaks.com/ (31.08.2007.)

Gnörlich, Carsten Dvdisaster. 2007. http://dvdisaster.sourceforge.net/en/ (31.08.2007.)

How To Choose CD/DVD Archival Media. 2006. http://adterrasperaspera.com/blog/2006/10/30/how-to-choose-cddvd-archival-media/ (31.08.2007.)

Jufsoft. 2007. http://www.jufsoft.com/badcopy/ (31.08.2007.)

Perko, Mitja. 2007. http://www.kvipu.com/CDCheck/ (31.08.2007.)

Smart Projects. 2007. http://www.smart-projects.net/cdrecovery.php (31.08.2007.)

Softpedia. 2007. http://www.softpedia.com/get/CD-DVD-Tools/CD-DVD-Rip-Other-Tools/CDR-Media-Code-Identifier.shtml (31.08.2007.)

Solidlabs Technology. 2007. http://www.soft32.com/download_77336.html (31.08.2007.)

Stellar Phoenix. 2007. http://www.stellarinfo.com/cd-data-recovery.htm (31.08.2007.)

Digitalizacija korpusa starohrvatskih tekstova i kritika teksta

Amir Kapetanović
Institut za hrvatski jezik i jezikoslovlje
Ulica Republike Austrije 16, Zagreb, Hrvatska
akapetan@ihjj.hr

Summary
Digitalization of Old Croatian texts was initiated within the framework of the project "The Old Croatian Dictionary". The purpose of the initiative was to make a dictionary that would lexicographically describe the lexis of the oldest periods of Croatian literacy in the Croatian language (from first records to the end of the 15th century). Digitalization of Old Croatian texts is more complex than digitalization of contemporary texts, not only owing to predominantly non-authored corpus in manuscript form, but also owing to the fact that the collected Old-Čakavian / Old-Štokavian corpus, originally written in three scripts, first needs to be critically analyzed. The Croatian medieval texts have been published since the 19th century in different forms (photograph, transcript, transliteration, transcription), and the publications have varying degree of quality. In addition, some texts that may become integral parts of the corpus have not been published yet. The paper will present the corpus structure, and major issues and principles in corpus analysis and design.

Key words: digitalization, corpus, the Old-Croatian language, Middle Ages, lexicography, textual criticism

Uvod

Danas, u vremenu sveopće digitalizacije, oko nas nastaju različiti tekstovni arhivi (digitalne biblioteke) i korpusi različite kvalitete. Pri dohvaćanju neke aproksimativno određene milijunske brojke u utrci za kolikoćom pojavnica određenoga korpusa kadšto se zaboravi na kakvoću primarnih korpusnih podataka. Jamačno kvaliteta korpusa, koji se sastavlja s ciljem njegova iskorištavanja u lingvističke svrhe, ovisi ponajprije o valjanoj digitalizaciji odabrane građe. U toj prvoj fazi rada u izradi korpusa moraju se primarni podatci (tekstovi) kvalitetno obraditi i pohraniti, a tek potom može se korpus nadograđivati metapodatcima.

Put do označenoga (anotiranoga) korpusa ipak je kraći i lakši ako stvaramo *korpus suvremenoga jezika* nego *jezičnopovijesni korpus*. Prilikom izrade korpusa suvremenoga jezika mora se oprezno pristupiti izboru reprezentativnih inačica

tekstova (ako je poznato nekoliko autoriziranih izdanja teksta) i vjerno iz izvornika prenijeti primarne podatke u elektronički oblik (ako je riječ o tekstovima iz neelektroničkih medija). Ako kanimo sastaviti jezičnopovijesni korpus, priprema je za digitalizaciju veoma zahtjevna. Primarni podatci u takvu korpusu, za razliku od korpusa suvremenoga jezika, nisu *izvorni*, nego *posredni* (prijepis ili transliteracija ili transkripcija izvornih tekstova). Primjerice, digitalne slike izvornih srednjovjekovnih rukopisnih vrela hrvatske pisane baštine mogu poslužiti samo za dokumentaciju. OCR-programi do danas nisu tako usavršeni da bi mogli izvršiti prepoznavanje rukom pisanih znakova, osobito u hrvatskim srednjovjekovnim tekstovima izvorno pisanim glagoljicom, ćirilicom i starom latinicom. Nijedan stroj danas ne može osim *prepoznavanja* znakova izvršiti i filološku *interpretaciju* znakova, a bez toga se ne može izgraditi korpus baštinskih tekstova pisanih trima pismima, nenormiranim pravopisom i slovopisom. Posredni primarni podatci mogu se naći u dosadašnjim izdanjima baštinskih tekstova (preslovljeni izvori), ali mnoga izdanja u manjoj ili većoj mjeri nisu pouzdana. Zato je potrebno vratiti se opet originalima i provjeravati valjanost preslovljenih izvora. Ako se jezičnopovijesni korpus izgrađuje sa zadatkom da omogući izradu kakva povijesnoga rječnika, treba predvidjeti koje informacije za leksikografsku obradu u konačnici mora pružati korpus.

Povijest projekta

U Institutu za hrvatski jezik i jezikoslovlje (IHJJ) u Zagrebu izgrađuje se Korpus starohrvatskih tekstova (KST) koji će u prvom redu biti upotrijebljen za izradu *Starohrvatskoga rječnika*. Hrvatska filologija nema ni pokusni računalno čitljiv korpus starohrvatskih tekstova ni na suvremenim leksikografskim načelima izrađen rječnik koji bi nudio informacije o hrvatskom leksiku starijih razdoblja. Mažuranićevi *Prinosi za hrvatski pravno-povjestni rječnik* obuhvatili su samo pravna vrela, a za *Akademijin rječnik* (AR) iskorišteni su kao izvori samo neki starohrvatski tekstovi iz kojih nije ekscerpiran sav leksik. Danas jedino AR u hrvatskoj filologiji ima ulogu opsežnoga papirnoga korpusa (obilno su navođene potvrde), premda je to vrlo varljiva baza podataka, osobito ako se u njemu traže podatci o prvim ili jedinim potvrdama određenoga leksema. Tako je, primjerice, prema *Akademijinu rječniku* topomim *Evropa/Europa* potvrđen u hrvatskim tekstovima tek od 17. stoljeća[1], a prividni hapakslegomenon *taržan* rezultat je nedovoljno proučene grafije Zoranićevih *Planina*[2].

Izradu rječnika najstarijih tekstova otpočeli su neki slavenski narodi u 20. stoljeću (npr. nedovršeni dvotomni Gebauerov *Slovník staročeský*, nedavno dovršeni višetomni *Staropolski słownik*), a zamisao o starohrvatskom rječniku po-

[1] Uvid u starohrvatske tekstove, kao što su različiti hrvatski prijevodi *Lucidara*, pokazuju da se potvrđenost toga toponima može pomaknuti barem u 15. stoljeće; o tome v. Kapetanović 2005: 3.

[2] O tome v. Kapetanović 2006: 23.

tječe iz ranih 90-ih godina 20. stoljeća, kada se u tadašnjem Zavodu za hrvatski jezik razmišljalo o izradi *Rječnika hrvatskoga srednjovjekovlja* (Malić 1998: 61). Godine 1998. dr. Dragica Malić ponudila je u jednom članku skicu za izradu rječnika, a 2001. i nacrt s popisom izvora i pravilima leksikografske obrade u knjižici *Nacrt za Hrvatski rječnik do Marulića i njegovih suvremenika.* Od te godine do 2005. obavljena su u IHJJ-u preliminarna istraživanja pod vodstvom dr. D. Malić, koja su ponajviše bila usmjerena na prikupljanje i provjeravanje predviđenih izvora za rječnik kako bi se moglo otpočeti s izgradnjom strojno čitljiva korpusa. U tim su se istraživanjima mnogi predviđeni izvori (transkripcije i transliteracije starih tekstova) pokazali nepouzdanima i moralo se ulagati mnogo truda, znanja i vremena u provjeravanje nekih izvora prema originalima. Taj opsežni dio posla nije dovršen. Unatoč tim spoznajama i rezultatima intenzivnoga trogodišnjega istraživanja malobrojnih suradnika, projekt je kritiziran na temelju polaznoga popisa izvora: "*Hrvatski rječnik do Marulića i njegovih suvremenika*, izrađen na korpusu samo objavljene građe, ne bi mogao biti temeljno djelo hrvatske filologije. Jer, kao što je već istaknuto, objavljen je tek manji dio građe. K tomu objavljena je građa − ponavljam − dobrim dijelom nepouzdana i nužno je usporediti je s originalima. Pa kad je tako, uputno bi bilo građu ekscerpirati iz originala ili njihovih preslika, kako su činili priređivači Akademijina povijesnoga *Rječnika*. (Nazor 2005: 453)"

Drugim riječima, time je predloženo da se na početku 21. stoljeća leksik iz originala ili preslika ekscerpira kao što se nekoć kadšto činilo pri izradi AR-a i uvijek pri izradi *Rječnika crkvenoslavenskoga jezika hrvatske redakcije*[3]. Na taj način stvarao bi se *papirni korpus* (ili u modernijoj verziji: *računalna zbirka potvrda*), a ne strojno čitljiv korpus starohrvatskih tekstova. Takav korpus ne bi mogao biti podlogom iscrpne leksikografske obrade, otežano bi bilo provjeravanje i povezivanje podataka prilikom leksikografske obrade i teško bi se taj korpus mogao u budućnosti iskorištavati za druga deficitarna jezičnopovijesna istraživanja hrvatskoga jezika u IHJJ-u i izvan njega. I u skicama i nacrtima za rječnik njegova se izrada povezivala s izradom računalno čitljiva korpusa: "Svi će se izvori konkordirati, što znači da će sve potvrđene riječi ući u obradu. Time će se izbjeći samovoljan odabir riječi, kojim se npr. uvelike odlikuje *Akademijin rječnik*." (Malić 1998: 68); "Treba nabaviti/napraviti računalni program za konkordiranje izvora i za optimalno iskorištavanje građe, odnosno za stvaranje korpusa izvora kao osnove za leksikografsku obradu." (Malić 2002: 152)

[3] O ekscerpiranju za *Rječnik crkvenoslavenskoga jezika hrvatske redakcije* osnovnom i komparativnom metodom v. u uvodu toga rječnika na str. III.

Planovi i opis korpusa

Rad na izradi korpusa nastavljen je 2007. u IHJJ-u u sklopu projekta *Starohrvatski rječnik* (voditelj A. Kapetanović), koji je dio institutskoga programa *Hrvatska jezična riznica* (voditelj Dunja Brozović Rončević), uz potporu Ministarstva znanosti, obrazovanja i športa Republike Hrvatske. Polazni popis vrela koji je izradila dr. Malić reduciran je s obzirom na gornju granicu starohrvatskoga jezika (u korpus ulaze izvori do 1500, uz neke iznimke mlađih prijepisa). Sustavno se pristupilo izradi kritičkih izdanja hrvatskih srednjovjekovnih tekstova i njihovoj digitalizaciji. Odlučeno je da se pri izgradnji KST-a primijene isti ali korpusu prilagođeni računalni standardi i sustavi kao i u izradi ostalih korpusa u okviru današnje *Hrvatske jezične riznice* IHJJ-a, kako bi u budućnosti KST mogao postati potkorpusom krovnoga hrvatskoga jezičnoga korpusa u IHJJ-u. Do tada će KST funkcionirati kao samostalan korpus, razdijeljen na sastavnice prema vrstama tekstova, u čemu se ogleda i struktura građe budućega rječnika:

SASTAVNICE	IME	VRSTE TEKSTOVA
SASTAVNICA 1	*Pjesništvo i drama*	pjesme, plačevi, prikazanja
SASTAVNICA 2	*Proza I*	djelomični prijevodi Biblije, molitvenici, lekcionari
SASTAVNICA 3	*Proza II*	apokrifi, vizije, čudesa Marijina, svetačke legende, romani, priče, poučna proza...
SASTAVNICA 4	*Pravni tekstovi*	isprave, statuti, zakoni, urbari, regule
SASTAVNICA 5	*Povijesni tekstovi*	ljetopisi, kronike
SASTAVNICA 6	*Kratki napisi*	natpisi, grafiti, različiti zapisi i bilješke
SASTAVNICA 7	*Pabirci*	glose, izbor tipičnoga hrvatskoga općega leksika iz crkvenoslavenskih vrela i starohrvatskih imena i titula iz hrvatskih latinskih isprava

Raznolikost tekstova razvrstanih u sedam sastavnica osigurava jednu drugu ne manje važnu kvalitetu korpusa (uzorkovanje različitih tipova i vrsta pisanih vrela). Korpus će u konačnici biti tako označen da se može pretraživati čitav korpus ili samo dijelovi korpusa (po određenim sastavnicama, po stoljećima: 11-15, po starohrvatskim književnim idiomima: čakavski/štokavski). U pretposljednjoj navedenoj sastavnici bit će dosta zapisa koji nisu veći od sintagme i jedne rečenice. Budući da su sintagma i jedna rečenica premaleni korpusni uzorci, morat će se odlučiti kako će takva građa, nastala u rasponu od pet stoljeća, biti organizirana. Posljednja navedena sastavnica razlikovat će se od svih ostalih jer neće biti sastavljena od cjelovitih tekstova. To će biti pomoćna sastavnica koja će se iskorištavati u leksikografskoj obradi ako ne bude (dovoljno) potvrda iz hrvatskih vrela. Istina je da u hrvatskom srednjovjekovlju nije postojao "čvrst zid" između hrvatskih narodnih idioma i crkvenoslavenskoga jezika, a to se odnosi ponajviše na proznu beletristiku. Tekstovi koji su pisani pretežno crkvenoslavenskim jezikom, neće se uzimati kao vrela, a crkvenoslavizmi

u tekstovima pisanim pretežno hrvatskim obrađivat će se u *Starohrvatskom rječniku* ravnopravno kao i drugi *izmi* (talijanizmi, germanizmi, bohemizmi…). Tako će u korpusu i u rječniku biti zastupljen neizbrisiv trag utjecaja crkvenoslavenskoga jezika na hrvatske književne idiome srednjovjekovlja. Zapisi starohrvatskih imena i titula u hrvatskim latinskim ispravama navodit će se u rječniku kao rekonstrukcije s izvornim zapisom, npr. **Budəc* (Budeç, Budicius, Budiço, Budiz, Budizo).

Korpus starohrvatskih tekstova mogao bi se ovako definirati: računalno čitljiv, označen, jednojezični, nespecijalni statični povijesni korpus pisanoga jezika sastavljen od cjelovitih tekstova (osim pomoćne sastavnice br. 7). Što se raspona korpusa tiče (od prvih zapisa do 1500), i tu postoji problem s gornjom vremenskom odrednicom. Bit će iznimaka u vezi s nekim tekstovima, pa će tako ući u korpus neki srednjovjekovni tekstovi sačuvani samo u prijepisima iz 16. stoljeća. Nezamislivo je, primjerice, da *Vinodolski zakon* sastavljen 1288. (sačuvan u najstarijem prijepisu iz 16. stoljeća) ne bude dio korpusa jer on predstavlja važno vrelo starohrvatske pravne terminologije. Opseg korpusa (broj pojavnica) teško je za sada predvidjeti, ali se prema grubim procjenama može pretpostaviti da bi *Starohrvatski rječnik* mogao sadržati oko 50 000 natuknica.

Prijenos informacija i izbor načina transponiranja primarnih podataka

Andrea Hofmeister analizirala je u jednom teorijski usmjerenom tekstološkom prilogu (2005) odnos "nalaza" (*Befund*) i "tumačenja" (*Deutung*) s obzirom na polja informacija i stupnjeve (vrste) izdanja. Na primjer, ona posve točno zaključuje kada kaže da bazična transliteracija kumulativno sadrži informacije o pisanom tekstu svih polja (od paleografskoga do semantičkoga), ali upravo zbog toga prijeti nepreglednost i manja pristupačnost korisniku (*benutzerfreundlich*).[4] Ipak se ne možemo složiti s njezinom tvrdnjom da originalni sadržaj informacija može ponuditi jedino rukopis sam ili visokokvalitetni faksimil rukopisa jer faksimil ne može za razliku od originala pružiti paleografima podatke o vodenim žigovima, što nije nevažna informacija istraživaču koji želi datirati rukopis. A. Hofmeister u svojem radu nudi shemu odnosa polja informacija i stupnjeva (vrsta) izdanja, a mi ćemo ovdje ponuditi sličnu, prilagođenu našim potrebama i terminologiji (umjesto polja informacija navest ćemo izbor nekih konkretnih informacija koje se posreduju različitim načinima transponiranja teksta):

4 V. A. Hofmeister 2005: 5-6.

POSREDOVANJE INFORMACIJA U RAZLIČITO TRANSPONIRANOM TEKSTU					
	original	*snimka*	*prijepis*	*transliteracija*	*transkripcija*
materijal za pisanje	+	–	–	–	–
oblik slova i duktus	+	+	–	–	–
izvantekstni znaci	+	+	–	–	–
kontinuirano pisanje	+	+	+/–	–	–
slovopis	+	+	+	+	–
jezik	+	+	+	+	+

Takva editorička pitanja nisu nevažna za korpusnu lingvistiku. Osiguravanje zalihosnih informacija ne samo da je neekonomično u ciljanim istraživanjima (nepotreban utrošak vremena i novca) nego može pomutiti i transparentnost podataka te tako smanjiti razinu primanja važnih informacija. Stoga, nije nevažno hoćemo li primarne podatke u korpus prenositi transliteracijom, fonološkom transkripcijom, prijepisom[5].

	snimka	*prijepis*	*lat. transliteracija*	*lat. transkripcija*
glagoljica	ⰕⰂⰑⰡ ⰕⰟⰂⰑⰡ ⰕⰂⰑⰊⰡ ⰕⰂⰑⰼⰡ	ⱅⰲⱁⱑ ⱅⱏⰲⱁⱑ ⱅⰲⱁⰺⱑ ⱅⰲⱁⰼⱑ	tvoě t'voa tvoia tvoĵa	tvoja[6]
ćirilica	ТВѠꙖ̂ ТВОѢ ТВѠА ТВѠꙉА	твѡꙗ̂ твоѣ твѡа твѡꙉа	tvoja tvoě tvoa tvoĵa	
latinica	TVOIA tuoa tuoija tuoya	*TVOIA* *tuoa* *tuoija* *tuoya*		

[5] *Transliteracija* (lat. *translitteratio*) označuje postupak prenošenja znakova jednoga pisma adekvatnim znakovima drugoga pisma, npr. *краль* > *kral'*. *Fonološka transkripcija* (lat. *transcriptio*) označuje postupak kojim se grafemi za određene foneme iz jednoga slovopisa jednoga pisma prenose u slovopis istoga ili drugoga pisma, npr. iz 13. stoljeća *краль* > *kralj*. *Prijepis* označuje vjerno i dosljedno prenošenje znakova originalnoga slovopisa jednoga pisma bez konverzija i interpretacije znakova, npr. *краль* > *краль*.

[6] Komentar tablice: Vodoravno su ispisani načini prenošenja izvornoga teksta, okomito su poredana hrvatska pisma. Fingira se zapis riječi *tvoja* različitim pismima (uglavnom "mlađi" načini zapisa glagoljicom i ćirilicom), kao što "snimka" "originala" pokazuje. Potom se zapisi prepisuju

Odluka o tome mora se donijeti prije otpočinjanja digitalizacije jer su posljedice te odluke dalekosežne. Mora nam biti jasno koje informacije želimo prenijeti i kojem krugu korisnika (užem/širem). Budući da se KST izrađuje s ciljem istraživanja *jezika* (najprije leksika, potom gramatike), a ne različitih slovopisa hrvatskih triju pisama, sve potrebne informacije može ponuditi fonološka transkripcija. Osmišljena su unutar projekta pravila transkripcije za izradu KST-a jer ih za srednjovjekovne tekstove nijedan hrvatski tekstološki priručnik do danas nije ponudio.[7] Odbacivanjem zalihosnih informacija olakšava se računalna obrada podataka i ne ugrožava se transparentnost jezičnih podataka u korpusu i u rječniku. Na primjer, zbog višenamjenskoga *s* u staroj hrvatskoj latinici zapis *kosa* mogao bi se ovisno o kontekstu transkribirati kao *koža* / *koza* / *kosa* (posljednja mogućnost može odražavati tri homonima). Ako se primjer ne bi transkribirao, korisnik korpusa i rječnika morao bi se samostalno upustiti u filološku interpretaciju potvrda *kosa* bez uvida u grafijski sustav čitava teksta iz kojega je zapis. Promjena originalnoga slovopisa i pisma nije krivotvorenje jezika. Obično se u tradicionalnoj filologiji transliteracija povezuje sa znanstvenim, a transkripcija s popularnim izdanjima tekstova[8], ali se pritom zaboravlja da su za znanstvena istraživanja neiskoristive one transliteracije koje sadrže puno pogrešaka i krivih interpretacija znakova, a one transkripcije koje maksimalno čuvaju izvorni jezik teksta i koje prati kritički aparat, mogu biti korisne za znanstvena istraživanja, kao npr. neka izdanja u *Starim piscima hrvatskim*. Loša su ona izdanja starih tekstova u kojima se miješaju dva postupka: transliteracije s primjenom nekih transkripcijskih načela (npr. transliterira se *kral* > *kral*, ali *Maria* > *Marija*) i transkripcije s nekim transliteracijskim načelima (npr. transkribiranje: mo*j*i očči = posebno markiranje grafema s obzirom na to kako je fonem zapisan u izvorniku i čuvanje ortografskih geminata).

("prijepis") i pritom tip pisma ne mora biti isti kao na "snimci". Nadalje se izvodi latinička transliteracija svih zapisa osim latiničkih, jer "transliteracija iz latinice u latinicu nije moguća, tj. ona bi bila čisti preslik originala u grafemskom smislu" (Damjanović 1998: 61; Vončina 1999: 25). Teorijski bi bilo moguće ćirilicom i latinicom transliterirati glagoljički tekst i ćirilički tekst glagoljicom i latinicom, ali danas to više nije uobičajeno (o tome v. Bratulić 1981). Različito zapisana riječ *tvoja* (ovdje na 12 načina, i to nisu svi mogući) svedena je latiničkom transkripcijom na samo jedan zapis, i to tako da jezik ni u jednoj pojedinosti nije promijenjen.

[7] U jedinom hrvatskom tekstološkom priručniku (Vončina 1999) ne daju se pravila transkripcije srednjovjekovnih tekstova, premda primjeri iz nekoliko rukopisa pokazuju što bi mogla biti uzorna transkripcija. D. Kapetanić ne bavi se u svojem članku (1970) editoričkim problemima srednjovjekovnih rukopisa. Pomoć se ne može naći ni u drugim južnoslavenskim priručnicima, na primjer, knjiga srpskoga tekstologa D. Ivanića (2001) nema pravila transkripcije i ne bavi se starosrpskim tekstovima.

[8] Na primjer: "Latiničku transkripciju ili transliteraciju prihvaćali su kasnije i drugi izdavači [poslije I. Črnčića, op. A. K.], a danas je već postalo redovitom praksom da se hrvatskoglagoljski tekstovi izdaju u latiničkoj transliteraciji ili transkripciji, ovisno o tome da li se radi o znanstvenom ili popularnom izdanju. " (Grabar 1978: 40)

Leme u korpusu i u rječniku također će biti transkribirane (normalizira se grafija, a ne jezik) jer je zbog različitosti idioma, slovopisa i pisama nemoguće stvoriti grafijski i jezično idealnu[9] natuknicu (npr. dvojna natuknica *meja*/*međa* predstavljat će zapise *meiam*, *meyu*, *medji*, *medyom*…). Osim interpretacije rad na korpusu predviđa i rad na dokumentaciji, i to će biti dodatne informacije u računalno pretraživu korpusu. Planira se izrada opisa svih vrela i povezivanje digitalnih slika izvornika s transkripcijama. Na taj način nadoknadila bi se redukcija informacija za filologe koji žele provjeriti valjanost transkribiranoga zapisa. Osim toga, korpus će u konačnici imati i statističke podatke o čestotnosti.

Srednjovjekovni tekst i nova kritika teksta

Izrada KST-a i *Starohrvatskoga rječnika* treba se suočiti ne samo s primjenom novih računalnih tehnologija[10] nego i s novim odnosom prema srednjovjekovnom tekstu i novom kritikom teksta koju je razvila tekstologija/editorika 20. stoljeća.

KST temelji se uglavnom na rukopisnoj srednjovjekovnoj baštini (iznimka su inkunabule). Srednjovjekovni tekst nema autora (s biografijom) i sve su njegove inačice ravnopravne.[11] Autorstvo u modernom smislu nije starije od konca 18. stoljeća, stoga tekstovi korpusa nisu autorizirani i nema autografa. U srednjovjekovnim anonimnim tekstovima autorska bi se svijest doduše mogla tražiti u određenim slojevima tekstova[12], ali, dakako, ne autorstvo kako ga shvaćamo od početka 19. stoljeća. Ne može se ipak reći da su srednjovjekovni ljudi zazirali od isticanja vlastitih imena; pokazuju to brojni hrvatski srednjovjekovni glagoljički natpisi i grafiti od Istre do Dubrovnika (v. Fučić 1982).

Srednjovjekovni je tekst, pisan narodnim jezicima, nestabilan i otvoren, javlja se u brojnim varijantama i verzijama jer se prevodi, kompilira, prepisuje, prerađuje i širi bez kontrole prvotnoga sastavljača. Egzistira u svojim mijenama sve dok se širi prepisivanjem ili usmenim prenošenjem. Jedina iznimka je biblijski tekst. U srednjem vijeku važnije je bilo što više umnožiti tekst prepisivanjem, manje se u to doba pazilo na kritiku teksta, kao u antici ili u renesansi. Otkriće

[9] U *Rječniku crkvenoslavenskoga jezika hrvatske redakcije* natuknice se navode "u normaliziranom ('idealnom') liku crkvenoslavenskoga jezika hrvatske redakcije". V. prvi svezak rječnika (2000), str. VIII. i dalje.

[10] Novija hrvatska stručna (teorijska) literatura vezana uz probleme izrade korpusa usmjerena je na izradu korpusa suvremenoga hrvatskoga jezika, npr. Tadić 2003. Mora se reći da je nedovoljno hrvatskih priručnika s postupnim uvođenjem u probleme korpusne lingvistike, kao što su npr. dva novija njemačka: Scherer 2006; Lemnitzer−Zinsmeister 2006. U posljednjem navedenom priručniku postoji popis i osnovni opis različitih njemačkih korpusa i arhiva, v. Lemnitzer−Zinsmeister 2006: 113-126.

[11] Diskusiju o tome v. npr. u Stackmann 1994: 412 i dalje.

[12] Kao što to pokazuje R. Schnell (1998) kritikom neofiloloških stajališta o *autoru* i o *djelu* u njemačkom srednjovjekovlju.

tiska uvelike je promijenilo odnos prema tekstu. Riječima B. Cerquiglinia: "Izum knjigotiska je puno više mentalna nego tehnička revolucija." (Cerquiglini 2005: 125; prev. A. K.)

Varijantni srednjovjekovni tekstovi nekoć su se promatrali kao i novovjekovni autorski tekstovi. Danas je realno prikazivanje teksta (izvora) zamijenilo nekadašnje *rekonstruiranje arhetipa* teksta (na temelju svih rukopisa sastavljao se jedan koji bi mogao biti najbliži originalu) ili konstruiranje *idealnoga* teksta. U novijoj tekstologiji smatra se da su svi sačuvani oblici nekoga teksta (od njegova nacrta do tiskanoga izdanja) ravnopravne inačice jer svaka pokazuje određeni stupanj razvoja teksta. Zato će različite varijante i verzije srednjovjekovnih tekstova ravnopravno ući kao izvori u KST.

Mnoga dosadašnja izdanja srednjovjekovnih tekstova nisu loši izvori za KST samo zato što su dosadašnji njihovi "čitači" činili omaške i prezentirali ih u svojim izdanjima bez ujednačenih tekstoloških načela nego i zato što su i neka pouzdana izdanja izrađena "starom" kritikom teksta. Riječ je o izdanjima u kojima se, primjerice, rukopis iz 15. stoljeća ispravljao prema prijepisu iz 17. stoljeća ili su se lakune srednjovjekovnih rukopisa popunjavale prema vrlo mladim prijepisima.

U KST-u će dakle ravnopravno biti zastupljene sve poznate inačice jednoga hrvatskoga srednjovjekovnoga teksta nastale do konca 15. (ili početka 16. stoljeća). Na primjer, srednjovjekovna pjesma *Bog se rodi v Vitliomi* bit će u korpusu predstavljena ravnopravno dvjema inačicama (jedna iz *Pariške pjesmarice*, druga iz *Zbornika duhovnoga štiva* IV a 92)[13]: neće se od dvije inačice izrađivati jedan kontaminiran tekst s namjerom da se rekonstruira arhetip te pjesme i neće se pristupiti "popravljanju" tih dviju najstarijih inačica na temelju njihovih mlađih prijepisa iz 17. i 18. stoljeća. Ako se budu činile nužne emendacije i konjekture, svaka od njih morat će biti dobro argumentirana i dokumentirana.

Sve nam to govori da će kvaliteta KST-a i *Starohrvatskoga rječnika* ovisiti o koordiniranoj suradnji računalne lingvistike, tekstologije/editorike i leksikografije.

Zaključak

U članku je usmjerena pozornost na kritiku teksta prilikom izrade korpusa, i to u prvoj fazi rada: prikupljanje primarnih podataka za korpus. U izradi KST-a to je otežano zbog niza faktora (dosadašnja nedovoljno dobra izdanja tekstova, primjena stare kritike teksta, neujednačenost tekstoloških postupaka). Stoga je potrebno vratiti se izvornicima i kritički ih transkripcijom prenijeti u elektronički oblik. Struktura građe omogućuje drugu kvalitetu korpusa (raznovrsnost). KST treba biti računalno čitljiv, označen, jednojezični, nespecijalni statični povijesni korpus pisanoga starohrvatskoga jezika.

[13] O izvorima i čitanju te pjesme v. Ivšić 1939.

Literatura

AR = Rječnik hrvatskoga ili srpskoga jezika. I-XXIII. Zagreb: JAZU, 1880-1976.

Bratulić, Josip. Ediciona praksa hrvatskih istraživača i izdavača srednjovjekovnih tekstova u XIX i XX stoljeću (Historijski prikaz). // *Međunarodni naučni skup Tekstologija srednjovekovnih južnoslovenskih književnosti.* / Bogdanović, Dimitrije (ur.). Beograd: SANU, 1981, 137-147.

Cerquiglini, Bernard. Textuäre Modernität. // *Texte zur Theorie des Textes.* / Kammer, Stephan, Lüdeke, Roger (priredili i komentirali). Stuttgart: Philipp Reclam jun., 2005, 116-131. [Fragmentarni njemački prijevod francuskoga izvornika *Éloge de la variante*, Paris, 1989.]

Damjanović, Stjepan. Što (ne) smijemo mijenjati u piščevu tekstu? // *Radovi Zavoda za slavensku filologiju*. 32 (1998); 57-65.

Fučić, Branko. *Glagoljski natpisi*. // *Djela JAZU*. 57 Zagreb: JAZU, 1982.

Grabar, Biserka. O nekim problemima kritičkog izdavanja hrvatskoglagoljskih tekstova. // *VIII međunarodni slavistički kongres: Prilozi*. // Kuzmanović, Mladen; Stamać, Ante; Šojat, Antun (ur.). Zagreb: Hrvatsko filološko društvo, 39-44.

Hofmeister, Andrea. Textkritik als Erkenntnisprozeß: sehen − verstehen − deuten. // *Editio*. 19 (2005); 1-9.

Ivanić, Dušan. Osnovi tekstologije. Beograd: Narodna knjiga−Alfa, 2001.

Ivšić, Stjepan. O tobožnoj "najstarijoj sačuvanoj hrvatskoj pjesmi prije god. 1920”. // Građa za povijest književnosti hrvatske. XIV (1939); 1-10.

Kapetanić, Davor. Kako pripremati izdanja djela novijih hrvatskih pisaca. // *Croatica*. 1 (1970), 1; 237-259.

Kapetanović, Amir. Toponimi u hrvatskim “Lucidarima”. // *Folia onomastica Croatica*. 14 (2005); 1-22.

Kapetanović, Amir. Leksik Marulićeve *Suzane*. // *Colloquia Maruliana*. 15 (2006); 15-25.

Lemnitzer, Lothar; Zinsmeister, Heike. Korpuslinguistik: Eine Einführung. Tübingen: Gunter Narr Verlag, 2006.

Malić, Dragica. Skica za *Hrvatski rječnik do Marulića i njegovih suvremenika*. // *Filologija*. 30-31 (1998); 61-70.

Malić, Dragica. Nacrt za *Hrvatski rječnik do Marulića i njegovih suvremenika*. Zagreb: Institut za hrvatski jezik i jezikoslovlje, 2002.

Mažuranić, Vladimir. *Prinosi za hrvatski pravno-povjestni rječnik. I-II*. Zagreb: JAZU, 1902-1922.

Nazor, Anica. Dragica Malić i glagoljski tekstovi. // *Rasprave Instituta za hrvatski jezik i jezikoslovlje*. 31 (2005); 449-454.

Rječnik crkvenoslavenskoga jezika hrvatske redakcije. I. svezak. Zagreb: Staroslavenski institut, 2000.

Scherer, Carmen. Korpuslinguistik. Heidelberg: Universitätsverlag Winter, 2006.

Schnell, Rüdiger. ‘Autor’ und 'Werk' im deutschen Mittelalter: Forschungskritik und Forschungsperspektiven. // *Neue Wege Mittelalter-Philologie*. / Heinzle, Joachim; Johnson, L. Peter; Vollmann-Profe, Gisela (ur.). Berlin: 1998, 12-73.

Stackmann, Karl. Neue Philologie? //*Modernes Mittelalter. Neue Bilder einer populären Epoche*. / Heinzle, Joachim (ur.). Frankfurt a. M., Leipzig: Insel-Verl., 1994, 398-427.

Tadić, Marko. Jezične tehnologije i hrvatski jezik. Zagreb: Ex Libris, 2003.

Vončina, Josip. Tekstološka načela za pisanu baštinu hrvatskoga jezičnog izraza: Posebni prilog Stoljećima hrvatske književnosti. Zagreb: Matica hrvatska, 1999. (^{2}2006)

Building Repository of School-librarian Works on Croatian Network of School Librarians

Josip Rihtarić
2. osnovna škola Varaždin
Augusta Cesarca 10, Varaždin
info@josiprihtaric.iz.hr

Summary

This article discusses incentives and preparation actions connected with starting repository of school-librarian works, criteria for selecting the contents management system, solutions for indexing contents, which are the file and the article on the repository. This article discusses problems that emerged while using program MediaWiki, portalization of network center and plans for future.

Key words: repository, school library, portal, wiki, network

Informational technologies and school librarians

Recently development of informational technologies motivated many changes in libraries and librarian profession. School libraries in Croatia are at the back of those changes – there is still great number of school libraries without the computer with an internet access. Because of chronic lack of money for equipment, book collections and library staff, many school librarians have no conditions to build more quality library collections or to implement quality educational activities. Still, despite of those limitations, school librarians have embraced informational and computer technologies and are using them more and more in their jobs of school librarians. During last years they have started documenting more and more of their examples of good practice using computers and periphery computer units. Those their works, in the first place examples of good practice, they have started exchanging frequently during last few years without compensations, using CDs and memory sticks. Works in question are of different sizes, kinds and qualities, but in their contextual, pedagogic, methodic and technical form lot of intellectual engagement, creativity, knowledge and time is frequently invested in all the phases, from planning to realization. Since school librarians are in most cases the only employees in their school libraries, it would be uneconomic when such examples of good practice couldn't be used by colleagues as well, it is best to use them on more occasions. Exchanging and using works, school librarians are helping each other mutually to achieve more quality work with their customers.

Need for creation of professional repository

After smaller or greater collections of such works, that were dispersed on computers and amongst CDs, were collected on computers of school librarians, it was noted that such "private" collections are not practical or economic, it was noted that it was getting harder to find the desired work and the solution is not in the increase of exchange but in unifying and clarity. It was necessary to use advantages of librarian profession and of modern age we are living in: to unify works and information about works in one network place from which all school librarians will be able to take them and to process them using certain (librarian) classification systems (schemes) in order for one to be able to find the desired work quickly and simply in multitude of works, so that overview of works by queries – according to different criteria, would be available.

What was on computers of individual school librarians should have been transferred to common, remote network server, so that all works would become united and available to everyone at any time from any place that has access to the internet.

Preconditions for repository

While planning building of network repository of school-librarian works, several opposed wishes had to be considered.

On the one side, colleagues who know how to store the file on a memory stick do not have to know how to use HTML nor to please FTP. For them, transferring files on remote network server has to be as simple as to choose partition and folder in which files are stored, in order to determine, in fewest steps possible (screen changes), place on the internet for file storage. During transfer of file one must not waist too much time nor fill in tiresome, exhaustive forms, especially such where data is lost if network connection is broken.

On the other side, from the position of visitor, files have to be described quality and in – detail, so one could reach desired choice asking complex queries like limiting selection by years, authors, types of files, contents, keywords… It is necessary to have multiple summaries like newest, best graded, most visited, most downloaded files, overview by content (subject), keywords…

For associates and visitors it is important for repository to be simple and clear, without even one undesired link that would distract attention. But also, both associates and visitors need to get in fewest steps possible to any page on the networks site following simple and logical organization of pages.

These requirements are very hard to fulfill – to have simultaneously ascetic simple interface and enable reaching any desired page in simple, short and precise way and offer detailed overview of all existing files and articles.

Transferring files is not very demanding job, it is simplest to determine one folder on the network server in which files would be sorted and in one step transfer files from your computer, but accumulating files (which often have nontransparent names) on common server is not functional, it turns visitors and

associates away if files are not adequately processed. Generating alphabetic lists of united files is like an alphabetic catalogues that would only contain titles of books, without even one other information, and that is too less for quality information about the book or file. Visitors who would have an opportunity download works based solely on the name of the file would quickly conclude that, because of time spent on their download and study, it is not worthwhile to waste time, that it is very hard to find the desired work in multitude of works. No one likes or have time to download on his computer every single file to see what it is about only after opening it. It is, anyway, in opposition with librarian work. Gathering and finding information is just one part of the job, and value of a library is in good organization and not in simple amassing of information carriers.

Less time we spend indexing and describing files, less information we will have about them, and more detailed "archiving form" is files will be indexed with more elements, so visitor will be able to find them easier. However, in that case network page will be filled with that information. We are in the gap: on the one side, if too high demands are posed upon associates, they will refuse to cooperate and will prefer to exchange files in usual way. On the other side, big and disorganized series of files turns visitors away because it brings them in the situation to be forced to manage on their own finding desired information, creating bad image of librarians.

We have chosen the name *Croatian Network Of School Librarians* for the whole project, name *UDK 02* for repository, and http://www.knjizinicari.hr for URL – which mean *librarians* on English. Thought we started using word *repository*, we kept in mind that our repository will be different from repositories like institutional repositories which grown up since Budapest Open Access Initiative.[1] Lots of institutional repositories have been built, and according to The Directory of Open Access Repositories[2] and Registry of Open Access Repositories[3] there are more then 850 such repositories around the world. Looking into Croatian institutional repositories, like repository of Medical faculty[4], Faculty of Humanities and Social Sciences[5] and Faculty of Mechanical Engineering and Naval Architecture[6] of University in Zagreb and vocational repositories like E-lis[7], we noticed that they have selected the solutions in which authors cannot

[1] http://www.soros.org/openaccess

[2] http://www.opendoar.org

[3] http://roar.eprints.org

[4] http://medlib.mef.hr/

[5] http://darhiv.ffzg.hr

[6] http://www.fsb.hr/library/search.php

[7] http://eprints.relis.org/

self-archive their works or they can, but in that case they are not available as long as an authorized person does not give the allowance. Based on the experience of leaders of named Croatian faculty repositories it is noticeable that authors, generally, are not particularly interested to self-archive their works, e.g. they love to leave that work to leaders, even when "archiving form" is quite short and simple one.

Since we imagined the repository of school-librarian works foremost as simple continuation of exchange of works that we exchange anyway on CDs or memory sticks, just in this case we would not sore them on outside media but on one network place, this concept is not acceptable for us. Besides, since in our case we have professional repository at hand and we are mutually in equal and not hierarchic relation, it would be inappropriate to give individuals among us editorial authorizations or to make that part of work professional, because there are no more competent persons than ourselves who could realize our desires and ideas better, and we do not have money either to pay that kind of professionals.

We have decided for approach without hierarchy, in which all parts of repository are accessible to all associates, and every associate can change every part, but, also, every associate or group of associates can from whole base make their "editorial bibliographies", "feature articles", clear pages and similar.

Regarding procedure of transferring, indexing and describing files, we have decided for variant in which every associate can completely self-archive his own (or someone else's) work in all its parts successively or partially, and that work can be done by several associates, through mutual cooperation. Work is done in the way that everything that is missing or is not done correctly can be fixed by any other associate, which ensures speed and efficiency, and if someone does not want to give the part of the job to other associates, he can always reserve that part of work for himself, putting appropriate note or locking page. At this point we have not yet provided to offer procedure of archiving using complex form because of two reasons. First, there is great animosity in community towards filling in the pools and forms. And second, we are still working on all elements of describing and indexing files and articles, so that kind of form would require modifications from time to time.

With this kind of approach – building repository based on equality and cooperation instead of declaring editorial boards and setting hierarchy of privileges and protecting copyrights using licenses GNU, GPL[8] and CC[9] , not only that we achieve principles of democracy and ensure access to the information for all members of the community, but we are also on the track of most successful trends in network cooperation.

[8] http://www.gnu.org/copyleft/gpl.html

[9] http://creativecommons.org

Indexing files

Though all school librarians are obliged to use UDC in their school libraries, though UDC (generally) is used in all other libraries in Croatia, though it is widespread in the world and though its attribute is universality and applicability to all sources of the information, and explicitly for indexing files, not only books[10] - with wondering and disbelief we can note that we can almost find no application of UDC in some network site, besides network catalogues of classical (paper) libraries.

With introduction of computers into business, by global acceptance of only few operation systems (DOS and Windows) and unifying graphic interfaces of different programs (for example with usual appearance and positioning of menus), school librarians have accepted as well positioning of files in the folders or subfolders named by arbitrary names according to momentary needs, and number of which as well as the number of levels also increases gradually. Multiple indexing is generally not used, because of nature of programs (Windows Explorer and similar), and in case two librarians would get in their personal collections few hundred files with task to arrange them in groups and subgroups in order for these files to be as easy as possible to find – it is evident that each of them would have his own solution in naming folders, in number of folders and subfolders and in organization of their hierarchy (programs like Windows Explorer do not offer creation of mutual relations).

With appearance of many general and specialized network portals that offer clear lists and descriptions of network sites it is also notable that librarian classification schemes were not used, but network sites' administrators were finding their own solutions, creating groups and subgroups of information that are not in accordance with librarian schemes.

We have predicted possibility of files' (and articles') classification using UDC, but we have not yet realized this, not only because of stated reasons but because of need of resolving copyrights for usage of UDC on http://www.knjiznicari.hr. Besides, because of often changes of UDC during last years, national libraries in Croatia, with which librarians closely cooperate, apply UDC unequally in their funds. That is why it seemed that classifying and often changing of marks would take too much time and energy, and based on experience of other repositories and portals, it is questionable how many visitors and associates even use UDC. Because of everything that's been said feasibility of using UDC is questionable. All files and articles are marked with subject words, that is keywords, and because of characteristics of MediaWiki[11] program that enables simple way for establishing mutual relations using references and creating hierarchy, it seems better to use those keywords in building thesaurus of school libraries.

[10] http://www.udcc.org/

[11] http://www.mediawiki.org/wiki/MediaWiki

Problems with usage of MediaWiki

Thinking about technical side, from the beginning we were planning to run network site on such a platform that would enable gradual growth and expansion without program limitations, system for contents' management based on database. We generally agreed to choose system for contents' management from some of offered "free" software, e.g. open source software. Thought there are wide area of free software specialized for institutional repositories[12], and some of repositories has its own wiki, we have choose different approach. We determined criteria that potential contents' management system must have and according to that criteria we have narrowed down the selection of programs that could represent program platform for librarian works' repository, and later on for portal of school library science. The primary criteria:

- User interface should be as easy as possible to enable beginner to upload different kind of files and to write texts
- Interface must be in Croatian language
- Admin's side of software should be as simple as possible so average school librarian could use it

The secondary criteria:

- Users should be able to change files and text after they have been uploaded and written
- Visitors should be able to post comments
- Besides text, users should be able to add picture, videos and audios
- There should be automatically generated pages with statistics
- There should be automatically generated authors index pages
- There should be automatically generated pages with recently changes
- There should be integrated searching pages

Comparing characteristics of some twenty programs, in closest selection were Joomla[13], Moodle[14], PHP Nuke[15] and MediaWiki[16]. We have selected MediaWiki because it met most of criteria (besides grading / points), and mostly because there is smallest difference in the level of computer pre – knowledge between "regular" associate and "admins". We supposed that repository will contain multitude of different documents, and librarians' job is to organize information. This system enables for every librarian to manage simply and on his own with the contents, every school librarian can via network interface publish

[12] http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf

[13] http://www.joomla.org

[14] http://moodle.org/

[15] http://phpnuke.org/

[16] http://www.mediawiki.org

and classify his own and works of other people, and he does not have to learn HTML, PHP or satisfy FTP. Everything can be done via interface of his network browser. Still, it is necessary to learn some twenty new techniques, but multitude of associates on Wikipedias[17] is proof that those skills can be quickly and easily mastered. MediaWiki derived form the idea of GNU encyclopedia (which is immanent to librarians), it is in constant growth and development because it is program platform of Wikipedia – the most famous world network encyclopedia – which guarantees popularization of software, and it is constant trend that, besides classic network pages, many firms and institutions start their own wikies, as places for exchange of ideas and information and of together, cooperative work on projects. With this we open opportunity for school librarians to gather, exchange ideas and works on the best quality program platform with aim of achieving common goal – affirmation of vocation, improvement of work conditions, offering more quality service to users, which finally lead to fulfillment of librarian mission – building better world.

Though Wiki is world trend and world phenomenon, we have expected most problems in managing repository architecture, that is, of whole network site, first experiences indicate that problems are not of that nature, but they are in using Wiki as such. Namely, greatest problems for many associates are managing through the menus (which were typified), using "edit" card, fear to spoil or ruin something and similar. Regardless of that if those problems are because of different levels of computer pre-knowledge, that is, because of not being used to Wiki, or because Wiki is still not what is should be – fast and easy to master and use, we conclude that to use even this simple and quickly learnable program best to use workshops in informatics classrooms and directly teach participants in way of cooperation, especially because it showed that associates do not like to use help pages, that they prefer to skip them and they prefer to learn on computer using method of tries and fails, and during workshops they can always directly ask leader for clarification of any vagueness so learning is much faster and more efficient.

Repository and Portal

We have started thinking about repository of digital and digitalized school – librarian works (and called that project "UDC 02"), but from the beginning we have predicted possibility of portalization of network site, e.g. for http://www.knjiznicari.hr to become starting point for all school librarians for different contents, and not access point only to the repository. Since we have chosen MediaWiki program and with its usage of same techniques and procedures possible to add file to the repository and to write an article, there was no need to limit from the beginning contribution of associates to "filling up" the

[17] http://www.wikipedia.org/

repository with files, and since one of agreed principles was to enable modification of network site according to stated desires and needs of users, from the very beginning parallel are being built portal of school library science and librarians’ works repository.
In the repository of files and on the portal contents from area of class preparations and presentations of educational work are being gathered and presented; literary critics and annotations, displays of exhibitions and other examples of good practice; presentations of work of county expert councils and mother offices in all phases, from announcement of invitation for meetings, display of work of certain councils, to final reports about work; we arrange lists, overviews and comments of laws and other prescriptions that refer to school librarians; we follow and give overviews of works of associations, societies and institutions that are engaged in school library science; we model lists and clear pages of school libraries and librarians; we announce, follow and note events and gatherings, happenings and ongoings – creating thus a kind of almanac, and recently we have introduced rubrics which change on weekly or daily bases on starting page.

Future plans

Since there are many school librarians who have yet not heard of existence of network site on http://www.knjiznicari.hr, they should be introduced and included in the cooperation.
Many school librarians readily visit network site, but they do not dare to cooperate, they should be motivated for cooperation.
In network environment boundaries between states become irrelevant, and language barriers manageable, and it is our desire to achieve cooperation with colleagues from other states, from other language areas, respecting language, cultural and other specificities, but with desire to, with getting to know with examples of good practice, mutually influence on quality of work, affirmation of vocation of school librarian and, finally, for benefit of our users and community in which we live and work.

Conclusion

Though we have not encountered during building school – librarian works’ repository in wiki – environment examples of repositories that would be built by school librarians, amount of published works and written articles in less than a year from starting the network site indicates good perspective of building repository as well as of portal of school library science, solving the appearing problems because of use of new technology and new way of network cooperation as we go.

References

Bubaš, Goran. Objekti učenja: zašto i kako. 30.ožujka.2007. http://www.skole.hr/nastavni-materijali/o-materijalima?news_hk=5372&news_id=213&mshow=509 (31. kolovoza 2007.)

Chawner, Brenda; Lewis, Paul H. WikiWikiWebs: New Ways to Communicate in a Web Environment. // *Information technology and libraries*. 25 (2006), 1; 33.-43.

Frumkin, Jeremy. The Wiki and the digital library : digital libraries: modern pracdtices, future vision. // *OCLC Systems & Services: International Digital Library Perspectives.* 21 (2005.), 1; 18.-22.

Lasić-Lazić, Jadranka; Banek Zorica, Mihaela; Špiranec, Sonja. Repozitoriji digitalnog obrazovnog materijala kao sastavnica kvalitete suvremenih koncepta obrazovanja. // *Edupoint : časopis o primjeni informacijskih tehnologija u obrazovanju*. 33 (2005.), 33; 6.-12.

Lasić-Lazić, Jadranka; Slavić, Aida; Banek Zorica, Mihaela. Bibliotečna klasifikacija kao pomagalo u organizaciji znanja. // *Odabrana poglavlja iz organizacije znanja* / Lasić-Lazić, Jadranka (ur.). Zagreb : Zavod za informacijske studije, 2004., 10.-32.

Collaborative Tagging: Providing User Created Organizational Structure for Web 2.0

Mihaela Banek Zorica
Department of Information Sciences
Faculty of Humanities and Social Sciences
Ivana Lučića 3, 10000 Zagreb Croatia
mbanek@ffzg.hr

Sonja Špiranec
Department of Information Sciences
Faculty of Humanities and Social Sciences
Ivana Lučića 3, 10000 Zagreb Croatia
sspiran@ffzg.hr

Krešimir Zauder
Department of Information Sciences
Faculty of Humanities and Social Sciences
Ivana Lučića 3, 10000 Zagreb Croatia
kzauder@ffzg.hr

Summary

The Web 2.0, having the user both creating and organizing content, has changed much of how one approaches to and uses the Web. While the concept of user-submitted content is by no means new, user created organizational structure is. The article gives an overview of the organizational means and processes that enable it. To provide the general framework the article gives a short overview of Web 2.0. It then centres on the collaborative tagging process as a central organizational process and means for the Web 2.0 and provides definitions for the Web 2.0 terminology used. After describing the general process, its strengths and weaknesses and pointing out that, while useful, it cannot replace professional indexing tools and library and information science professionals the article goes on to describe collaborative tagging and its specific features in general. Some of the more common services to use collaborative tagging are then described.

Keywords: Collaborative tagging, Web 2.0, resource discovery, IR, social bookmarking

Introduction

Organization of the World Wide Web resources in the traditional sense has been a problem since its explosion. The early attempts to categorize the Web in subject directories have failed in comprehensiveness and have a significant lag in adding new resources due to the speed of Web growth and the sheer amount of resources present. Metadata, which could have enabled automatic gathering and organization, has failed for the Web as a whole due to lack of use and widely accepted standards and even due to misuse. This has left us with the search engine as the main access point to Web resource since any subject or author based access requires organizational structure that is simply not present. There was no task force large enough to provide the organization for the World Wide Web and the automatic means for providing subject based access still have to be developed. However, one of the new approaches to organization of online resources, central to the so-called Web 2.0, has managed to enlist the help of the common user in organizing vast quantities of these resources.

We can characterize the phenomenon known as Web 2.0 by its technological and design aspects but they are just a support for its conceptual nature which is what distinguishes it the most from the “old Web”. Web 2.0 is dependent on and built from user supplied content and organization in a collaborative environment. We could call this aspect of Web 2.0 “social Web” or “collaborative Web” but “Web 2.0” is far more common. It is this collaborative principle which gives it power and it has come a long way from merely linking to other pages.

Web 2.0 services constitute the backbone of collaboration. Technologically, Web 2.0 treats the Web as a platform with its own applications accessible through the browser. These applications make the services which provide the needed collaborative environment and enable one of the key Web 2.0 principles: the service automatically becomes better the more users it has [7]. In using many of these services users collaboratively create content (as is the case with Wikis) or, more frequently, create collections of either their own resources (e.g. Flickr) or publicly available ones (e.g. del.icio.us, LibraryThing).

User collections are thus one of the central concepts for Web 2.0 IR and resource discovery. One of the essential features of a collection, and the one which distinguishes it from a heap, is its organization, a topic which has plagued the web from its beginning. The collection organization, if done right, enables IR and resource discovery by providing different access points to organized resources without just resorting to full text search combined with statistical methods and ranking algorithms. While search engines are great for known item retrieval they fail somewhat on subject related searches and casual resource discovery (comparable to shelf browsing in libraries).

The central principle and means of organization, and thus the creation of access points to resources, for Web 2.0 is tagging which is most frequently implemented in its collaborative form.

Tags, tagging and folksonomies

We can define tagging as the organizational method and the very process of assigning ad hoc user-created natural language keywords to information resources thus organizing them into user specific collections. It should be noted that indexing term is here used in its broadest sense since indexing terms are traditionally used to denote the subject of the resource while the scope of tags is broader. The ad hoc created index terms used for tagging are most frequently called tags although one can also find other terms such as topics or labels. From the definition we can clearly see the three needed components in a tagging system: users, resources and tags. Indeed, we can call a basic information object of these systems a *post* and say it consists of a (user, resource, {tag}) triple [1]. In other words, to have a post we need one user, one resource and zero or more tags. Depending on a service, a resource may be an URL (or bookmark), an image, a Web clip, a bibliographic description of an academic paper, metadata about a real world object (a book or just about anything else), etc. As we can see the resource can be either just the metadata about an information object or it can include the information object a user wants to post. Tags are separated from the resource as a distinct feature supporting several functions within a service and thus a special case of metadata.

Although seemingly simple, there are various ways one can think of a tag. Tag has the features of metadata (more specifically, an indexing term or a keyword but is broader in scope), a category name and a navigational tool. A tag as an ad hoc created category name is an especially important notion since modern cognitive science has clearly shown categorization is central to our thinking [5]. The ability of the user to use the categories and category types that first come to mind and are easiest to use while not having to cross-reference a controlled vocabulary of some kind contributes to the cognitive ease of use of the tagging process. Also, this kind of approach to knowledge organization does not have to be learned beforehand. The problems that arise from tagging, the skill in tagging and some advanced approaches (e.g. the faceted approach) to tagging all may be learned afterwards. This ease of use might be one of the prime reasons for the current popularity of tagging. It may also be one of the pitfalls of tagging: the perceived ease of use might backfire later when a user's collection organized in this manner grows too big to be easily navigated through and used in general. This might have been prevented by planning the tagging process beforehand but the user doing the tagging did not have the needed knowledge and experience to plan it and quite probably did not even perceive the need to do so. It remains to be seen how many users will re-organize their collection when this happens since there is quite serious amount of work involved. This might also be just a short term solution which will lead to increased awareness of the problems present and thus to ease of use of more complicated tools and processes by the average user. This in turn might help in realization of the Semantic Web understood as an extension of the current Web in which data is semanti-

cally described and automatically connected across resources thus facilitating machine to machine communication, greater reuse of data and a general shift from document to data.

We can analyse a tag in terms of its specificity, objectivity/subjectivity and in terms of the properties it implicitly presupposes. If we consider the hierarchy of categories we employ in our thinking a tag may be more or less specific (or more or less general). For example, where one user, who is especially interested in knowledge organization, might use the tag "collaborative_tagging" others might use "tagging" or "metadata". Objectivity/subjectivity refers to what a tag actually describes: the resource or the user's relation to the resource. For example a tag might be "cars" which quite obviously relates to the resource and is useful for IR and resource discovery or "to_read" which is useful primarily to the user who tagged it. Objective tags thus have a much greater value for information retrieval and discovery of other users while subjective ones primarily have value for individuals who used them and sometimes for other users of their specific collections but not for users of the system in general.

In terms of the property a tag implicitly presupposes, a tag might deal with the subject, content, author, page type, task, frequency of use etc. In the traditional approach to metadata the properties are explicitly defined and they take certain values (e.g. <meta name="DC.creator" content="John Smith" />). Tags can be likened to values of these properties but the properties themselves are left implicit.

Tagging problems

Although tagging is the most prominent knowledge organization method for Web 2.0, and, due to the user acceptance, the only one that works at this level, it is not without its problems. We can summarize them in these five crucial ones:

- lack of vocabulary control
- lack of defined relationships between tags
- lack of explicitly defined properties for tags
- lack of user education and educational material
- users tag primarily for own use

Tagging is not backed by a controlled vocabulary so many problems these tools solved for databases and traditional institutions are present. There is no synonym control or homonym discrimination which presents a barrier to IR and resource discovery. For example which jaguar does a user want? An animal or the car? If a user searches for "folksonomy" it is quite probable he also wants "tagging" and "collaborative_tagging" to be included in the search. In addition, there are no conventions for the form of word to use (e.g. singular or plural?) and for the creation of compound tags. Tags are most frequently separated by spaces so "Web 2.0" would produce two tags, "Web" and "2.0". Since there are no conventions for compounding tags, this gives rise to many possibilities of the

semantically same tag: “Web_2.0”, “Web-2.0”, “Web2.0”. Some services allow usage of spaces in tags by separating the tags with another character (e.g. “;”) which alleviates the problem somewhat but does not solve it completely.

Lack of defined relationships presents a problem related to specificity of tags. Searching for a tag will not find tags its hyponyms i.e. the semantically subordinate tags. For example looking for “furniture” will not find posts tagged with ”chairs”, “tables” and similar. This is another problem that was traditionally solved with various knowledge organization tools. Perhaps the main problem is that the use of these tools is a skilled process normally undertaken by highly trained information professionals [6] so one currently cannot implement them without losing users without which tagging has no significance since its greatest quality is that “common folk” are doing it in massive numbers. This fact also clearly shows that tagging is just another organizational layer for a specific purpose which cannot replace existing knowledge organization tools and methods but can add to them. One might, for example, design a Website of a large library with traditionally organized resources but allow users to build personal collections organized by tagging. If enough users did this it would provide a new layer of access points to library collection.

Lack of clearly defined properties of which tags are instances is a more serious problem than is immediately apparent. Traditional metadata implementations had explicitly defined properties which were then filled with instances that were quite often taken from controlled vocabularies. In tagging, however, these properties are present only implicitly. This means that while human users might sometimes recognise them (i.e. when a bookmark is tagged with “cars” one will naturally suppose it is the subject of the resource) they cannot be used automatically by the machine. Even human users will sometimes be confused: if a bookmark is tagged with “images” and “John_Smith” one cannot know if the resource contains images of John Smith, images taken by John Smith or even text about images (i.e. resolution and colours) written by John Smith. So the property of which “John_Smith” is an instance might be “author” or “subject” and “images” might be the instance of “content” or “subject”.

As in traditional environment, tagging is also a matter of knowledge and skill. The “common user” still does not possess that qualities since these subjects have yet to enter the curricula and most of the services using the tagging approach to organization still do not offer tagging tutorials, FAQs, and other documentation which would describe collection building problems and give extensive tagging advice.

It is important to keep in mind that users of services employing tagging for users’ collections organization primarily tag for themselves. This means that they will frequently use tags that make sense only to them and generally be sloppy in their tagging. They might also employ different means to overcome tagging problems which might make sense to them and even be quality solutions for their collection organization but which do not contribute to the system as a

whole. A user might, for example use prefixes to define properties (such as "subject.images") or use different tags to connote different properties (such as "audio" for content and "music" for subject or vice versa).

Tagging benefits

The fact that tagging is an organizational method and process that common users are actually employing to organize massive quantities of very heterogeneous resources themselves is its most important benefit. This is the first time this has happened on such a massive scale and the first time we have the infrastructure to make it happen. Given the number and the growth of information resources and given that automatic methods still have many difficult problems to solve, this constitutes a very important aspect of Web organization. It is this fact that makes tagging worthwhile in spite of its many problems. Besides that, it is quite interesting that many things that constitute the problems of tagging are also its benefits. The lack of vocabulary control also means, as already mentioned, that users can start using it without advance preparation and that it is easy to use every time since there is no need for cross-referencing terms. Also since terms are created ad hoc there is no delay until the terms enter the vocabulary and it is quite difficult to imagine a creation of a controlled vocabulary comprehensive enough for general purpose tagging.

The fact that the users are tagging for themselves and the free nature of the tagging makes sure that the organization of every collection will make sense to its user. In other words, every user is able to use objective tags which make sense to him or her on the level of specificity which suits him or her best and subjective tags to facilitate an organizational layer tailored according to tasks, opinions, frequency of use and other personal parameters.

It should also be mentioned that this kind of approach, where users are included in the resource organization, is an important step to raising the users' organizational skills and their awareness of organizational problems. This could in turn lead to increased educational efforts and materials (both professional and available through the Web services in forms of tutorials, FAQs and other documentation) which is necessary if the users will get to the next step, i.e. the Semantic Web. It is quite probable, given the scope of the problem, that only a smaller part of the Web will ever become "semantic" if the users are not involved.

The tags taggers use might help in gathering data for controlled vocabulary creation for which the terms users use are an important input. The idea of combining the results of tagging with controlled knowledge organization systems is well perceived as a possible building block for the Semantic Web. According to different research activities [7] Web 2.0 ideas and applications can contribute to the creation of ontologies for a multitude of domains, which is essential for the development of the Semantic Web. The model of ontologies defines precise but non exhaustive semantic relationships between terms while the model of tagging associates terms into exhaustive contexts with no specific relationships.

The optimistic combination of these models would result in an exhaustive and precise model of knowledge organization.
Besides ontology creation, users might also play a significant role in actual deployment of this model. Use of ontologies and semantic description in general for Web organization has similar problems as many other approaches: the sheer amount of heterogonous data involved and the need for human intellectual input in organizing a large part of it. However, users might be brought to help overcome this problem in much the same way they are tagging, only in connection with ontology and on data, rather than document, level.

Collaborative tagging

The most frequent implementation of tagging is in its collaborative form. Collaborative tagging is tagging of resources in such a way that it is possible for different users of the same service to tag the same resource. This can manifest itself in two ways. The most frequent way (as in del.icio.us, citeUlike, etc...) is that the resources are publicly available to all users and all users may add it to their collections and tag it separately. The same resource may exist in collections of different users where it may (and frequently is) tagged with a different set of tags. The types of resources in this kind of services include metadata about: Web pages (i.e. bookmarks), academic papers, music artists, albums and tracks, books and anything else that is not unique to the individual user. They may also include the whole information object as the resource part of the post but the usual resource of these systems is constructed of just metadata. In this case just a resource is not enough to identify a post: it is identified by both resource and a user. A resource as input can provide a list of users who have that specific resource in their collections which can be useful for identification of users with similar interests and thus support casual resource discovery.
The other and much less frequent form of collaborative tagging happens when a resource is specific for a user, is located just in his or her collection and some or all of the other users are allowed to change the tags of the resource. This is the case with Flickr, for example. If other users were not allowed to add or change tags (e.g. YouTube) it wouldn't be *collaborative* tagging. For this type of collaborative tagging system it is characteristic that the user is the owner or creator of the information object and that the whole information object with the accompanying metadata constitutes a resource. The user may or may not give the permission to other users or groups of users to change the tags of his posts. Sometimes this permission is automatically given. For example, when the post is included in a meta-collection (e.g. a collection of posts based on a subject or a group) other users who have their posts in the meta-collection can change the tags of all the posts in that collection, not just their own.
Collaborative tagging is a form of tagging most prominent because by its very nature it alleviates (but doesn't solve!) some of the problems present in tagging and displays some social organizational aspects and thus supports knowledge

discovery. The central phenomenon is that a same resource in most services is usually tagged by several users. Indeed, the number of users who have a same resource tagged in various ways in their collections is frequently very big and the totality of the tags used to describe the resource clearly shows which tags are most prominent among users for the description of that specific resource. This totality of tags is highly significant for information retrieval and is important for resource evaluation. It helps with the problems of objectivity/ subjectivity and specificity as well as sloppy tagging by users and some of the language related problems. If one user tagged a resource just with “to_do” others quite probably collaboratively “corrected” this by using more objective tags (e.g. subject or content related). The same thing happens with more or less specific tags (e.g “metadata” and “tagging”), synonyms (e.g. ”movie” and “film”), compound tags (e.g. “Web_2.0” and “Web2.0”), sloppy tagging (e.g. “metdata” and “metadata”) all of which increases the recall of the system and is not possible in a system without the collaborative approach. Although the approach cannot distinguish homonyms, when searching with a tag it can provide related tags (i.e. the tags users frequently used in conjunction with the tag used as query) to further specify the search.

This however provides much meta-noise since some semantically same words constitute different tags. Examples are singular and plural forms, tags compounded in different ways, shorter forms of the same word, abbreviations, etc. So a del.icio.us tag cloud for flickr will include “photo”, “photos” and “photography” all as highly popular and thus highly ranked tags for this resource. Another problem is for a service to attain the critical mass of users that tag frequently and thus support other users’ tagging and resource discovery. In the future, there will probably be few main collaborative tagging services for each type of resource and the highest challenge for new services will be attaining enough “taggers” for the system to reach its full potential. It remains to be seen just how serious are the problems derived from too many users. As things stand now a large group of diverse users seems more of a benefit than a hindrance but as new approaches are implemented and as community based Web develops both technically and socially, benefits may also be gained from tailoring a service for a highly specific community.

An interesting benefit of these services is an instant feedback. While tagging, a user can usually see the tags other users frequently used when tagging. Immediately after tagging the user can see how many other users have the resource in their collections, and look at the tag cloud of the resource and start finding other users with similar interests.

Resource popularity plays a highly significant role in this kind of services since it is used for ranking and recommendation. Also, a popular resource will be easier to find since its folksonomy will be rich and tagging a popular resource will be somewhat easier since it will be easier for the system to recommend tags. The resource popularity is not, however, gained from the collaborative

tagging method and process but from the inclusion in user collections and would thus be present in the same form if some other organizational approach was used.

Folksonomies

Another term which is frequently used in conjunction with collaborative tagging is folksonomy. As is the case with other terms in this area, folksonomy is frequently ill defined. It is often used either to denote a whole variety of phenomena or defined to broadly to be of any use.

We define a folksonomy as the totality of tags emerging from the process of (collaborative) tagging. According to this we can liken a folksonomy to an "open vocabulary" which currently does not include any relationships but does include other data such as the number of times a tag has been used in the whole system, for a specific resource and by a specific user. Since some services have started implementing relationships between tags (e.g. Bibsonomy), soon we might see folksonomies with some relationships included.

From this we may distinguish three types of folksonomy: a service folksonomy, a user folksonomy and a resource folksonomy. Each of these types of folksonomies may be visualised by a tag cloud. A tag cloud is a primarily navigational device that visually shows the most popular tags where the frequency of tag usage is denoted by font size: the larger a tag's font the more often it was used. It may or may not show all the tags present in a folksonomy depending on the number of tags. A tag cloud representing the folksonomy of a service shows the popularity of tags which in turn shows "hot" topics among the users of a service. A tag cloud representing all the tags of a single user shows his interests within the resources current service allows to collect, while a tag cloud representing a folksonomy of a single resource shows how the resource was tagged by all the users who have it in their collection in one service.

Another way to distinguish folksonomies is to "broad" and "narrow" types [8]. Broad folksonomies are the ones emerging from services in which all users are able to (which doesn't mean they do) tag all resources that can be included in a service. Narrow folksonomies emerge from the services in which a resource unique to one user's collection is tagged by just that user and possibly other users which were given permission to do so by the "owner" of the resource.

Conclusion and further directions

There is no doubt that the Web 2.0 is a propulsive and already widespread phenomenon. A plethora of Web 2.0 services are freely available to the common user that very frequently use collaborative tagging to facilitate organization, ranging from social bookmarking (del.icio.us), academic paper sharing (CiteULike, Connotea), media sharing (flickr, last.fm), collections of real world items (LibraryThing) etc., or a combination of those. These and other services offer new models, methods, and technologies that can be adapted to improve

different domains with a strong focus on formal knowledge management and IR, like the corporate sector, information sector or education.
With the application of the above described Web 2.0 services, traditional educational institutions or information agencies like libraries could transform themselves from a place of passive information consumption to a dynamic, participative and creative knowledge production space. Besides using traditional organizational and navigational tools, users could produce and consume knowledge, create new information architectures and change the information landscape by using collaborative, social and personalized means.
An important facet of the transformational function of the Web 2.0 ideas is their potential contribution to solutions to some of the recognized existing problems within IR, although it is believed that the real strength of these services lies in their combination with more controlled knowledge organization systems.
Finally, Web 2.0 concepts, particularly collaborative tagging, could significantly enhance and facilitate the further development of the Semantic web idea, by using emerging folksonomies for the development of ontologies and thereby acknowledging the idea that even this machine oriented concept can’t be realized without a strong social dimension.

References

[1] Cattuto, C., Loreto, V. and Pietronero, L. Collaborative Tagging and Semiotic Dynamics. PNAS, 104 (5), January 2007. http://arxiv.org/abs/cs/0605015v1 [15/05/2007]
[2] Golder, S. and Huberman, B.A. (2006). Usage Patterns of Collaborative Tagging Systems. Journal of Information Science, 32(2): 198-208.
[3] Guy, M. and Tonkin, E. (2006). Folksonomies: tidying up tags? D-Lib Magazine, 12 (1), January 2006. http://www.dlib.org/dlib/january06/guy/01guy/html [04/05/2007]
[4] Hammond, T. et al. (2005). Social Bookmarking Tools (I): A General Review. D-Lib Magazine, 11 (4), April 2005. http://www.dlib.org/dlib/april05/hammond/04hammond.html [03/26/2007]
[5] Lakoff, G. Women, fire and dangerous things: What categories reveal about the mind. Chicago and London: The University of Chicago Press, 1987.
[6] Macgregor G. and McCulloch, E. Collaborative tagging as a knowledge organisation and resource discovery tool. Library Review, 55 (5), 2006. http://www.emeraldinsight.com/10.1108/00242530610667558 [05/07/2007]
[7] Mesnage C. and Jazayeri, M. Towards global collaborative tagging. 2006. http://cedricmesnage.org/resources/pdf/mesnage06aSubmitted.pdf.
[8] O’Reilly, T. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. O’Reilly Media, Inc; 2005. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html [02/27/2007]
[9] Vander Wal, T. Explaning and Showing Broad and Narrow Folksonomies, http://www.personalinfocloud.com/2005/02/explaining and .html [04/05/2007]

Heritage and Digitization as in the Example of the Scientific Journal "Glasnik matematički"

Jasna Dravec Braun
PMF-Matematički odjel Središnja matematička knjižnica
Bijenička c. 30, Zagreb, Croatia
jdravec@math.hr

Summary

Based on the example of journal "Glasnik matematički" it is possible to present what influence modern technology has in spreading of scientific information and on its preservation. Simple digitizing by scanning the journal's articles contents resulted in material that can be quality managed thanks to technical and software solutions. According to that, access to the journal is modernized increasing its accessibility - "visibility" and the number of potential authors too.
At the moment the journal is available online in three ways [http://web.math.hr/glasnik/PastVol.html]. As bibliographic data for the period 1966-1997 year, as full-text in free access for the period 1998-2003 year, from 2004-2007 combining as a bibliographic base and with free access from subscribing institutions only. Its back volumes will be successively scanned and digitized. It could be said that in this part of processing the journal contents, preservation of scientific information for future generations is in the function of heritage. Through open access it will be available to any user, any scholar who is interested, without limitations. Digitizing and archiving in Google Scholar, for now, will process preservation.
Electronic and print journal issues parallel appearance has the practical value of preserving print version, partially. Time will show at which point, if any at all, there will be a need for an online article and simultaneously the abandoning of a printed issue. If this occurred, care about the preservation management of a printed issue would be conveyed to preservation of the e-journal. With experience and accomplishments in publishing and storing till then, it would not be a more serious problem then the one advancement of information technology is putting before electronic contents.

Key words: digitizing, digital archiving

Introduction

Journal "Glasnik matematički" has been published since 1966.g. as a 3rd series of its precedents: "Glasnik Hrvatskog naravoslovnog društva" (Series I), founded in 1886, "Glasnik matematičko-fizičko astronomski" (Series II), founded in 1946. Its long tradition and quality give it a value of national knowledge heritage. Volumes of all three series are part of serials library holding in the Central Mathematical Library at the Department of Mathematics, University of Zagreb. It is also reviewed by two most famous mathematical data bases: MatSciNet maintained by the American Mathematical Society and Zentralblat-MATH, issued by Heidelberg Akademie der Wissenschaften Fachinformationzentrum Karlshrue, as well as in Russian mathematical database Referativnyj žurnal. Articles are written in English, German and seldom in French and Russian, with abstracts in Croatian until 1990, when this changed to abstracts in English only. The journal is published semiannually by the Croatian Mathematical Society and the Department of Mathematics, University of Zagreb. Financially it is supported by the Ministry of Science, Education and Sports. It also has long term exchange policy with 280 international journals all over the world. It publishes original research papers from all fields of pure and applied mathematics.

The journal is indexed in SCOPUS database and in Croatian portal HRČAK-Portal of Scientific Journals of Croatia.

Network accessibility is a real challenge in organizing particular contents, and web sites. On one hand it enables a comprehensive way of linkage through mathematical databases, sites with thematically arranged journals and similar and on the other hand hierarchical restriction of access in order to ensure economic stability of print issue production**.** According to this "Glasnik matematički" is included in two thematical journal lists: *Topology-Related Journals*, and *Journals containing articles related to number theory*. It is also included in more general journal lists: *Printed Journals with Mathematics and Internet sites; AMS Digital Mathematics Registry; The Index of all Journals worldwide for Mathematic, Statistics, Econometrics, Actuarial Sciences, Quantitative Finance, OR; Mathematics Journals on the Web.*

Subscription to the journal enables access to its online content while anonymous users have restricted access to four previous years.

For preserving written heritage it is necessary to answer some basic questions: how, in which extent, by whom it will be done and from where the financial support is to be ensured. What was done in that direction for "Glasnik matematički" until today is based mainly on the enthusiastic work of managing editor PhD. Andrej Dujella, and software editor PhD. Josip Tambača, professors at the Department of Mathematics, University of Zagreb.

Preservation of the journal

"Glasnik matematički", as many other journals in the contemporary information era, has different possibilities solving preservation needs. One classic solution is to deposit print versions in appropriate places. But publishing of the print version has a parallel electronic issue for the last ten volumes and reasonable as much as modern consideration of preserving goes into the direction of digital electronic archiving. Volumes from the period 1998-2003 have articles in full-text and free access accessible from its main web site as well as last four volumes, which have free access from subscribing institutions only. Digitizing was done by scanning deposited volumes and creating Pdf files of each article for volumes 33 and 34. For other eight volumes Pdf files of the original articles that were prepared for printing were extracted from the *glasnik* document class for LaTeX2e. This whole procedure was without expenses what is a challenge for the preservation policy.

Two decisions were brought regarding the journal "Glasnik matematički" preservation, depending on accessibility of its articles at the present time. One is successive digitizing of back volumes starting from volume 31 backwards. It is an extensive job to do that cannot be left to enthusiastic workers. The Editorial committee accepted Google's offer for taking over this project. The journal will be digitized using destructive scanning**.** There will be possibility of retreiving digitized articles from Google Scholar. The publisher maintains copyright and ownership of their contents.

The second decision is connected with online issues with restricted access, accessible from the journal's main web site and Google Scholar, and the portal Hrčak, as well. The ease with which modern technology allows contents to be presented on the Internet sometimes may lead to redundancy. To avoid this, serious efforts were made by associations of libraries, at different levels, not to duplicate work and records during preservation on different media (de Lusenet (ed.), 1997). But when a whole process moves towards e-environment, everything is different because the focus of today's preservation is not the preservation of particular storage media, but rather on long- term access to the intellectual content of the resource (Gould, Varlamoff, 2000). The question of redundancy emerges due to numerous digital collections, which offer the same range of information. For journals due to their web sites with the whole or partial content independent of collection or archive they are included or indexed in. However, existence of institutional archives did not cut subscriptions of scientific journals, so many of them became hybrid (Silobrčić, 2007) and so did "Glasnik matematički". It has the moving wall model of balancing accessibility from restricted to open access. Editorial committee defined the moving wall of four years. This could affect exchange with international journals. Because of that the possibility of exchanging electronic articles should be taken in consideration. When once digitized, documents are in the same process as born-digital. If in future the print version of the journal will be abandoned, its preserva-

tion will change toward preservation of born-digital documents. The only present and future doubt could be existence of online archives if access is ensured. What is preserved from damage are print versions, less or not copied at all because e-versions exist providing better printed texts than the copy itself.

Conclusion

In the example of the journal “Glasnik matematički” decision on long term storage and preservation is to be brought by the publisher as well as the initial creation of the digital documents without expenditure. A strategy for digital preservation is based on existing print versions of the journal and its electronic issue for the last ten volumes. As far as quantity of work on digitization rice the project is taken over by Google Scholar. At this point the aims to have wider public and to be preserved are interlaced.

Hybrid in regard to its accessibility it balances archiving and publishing by four years moving wall.

References

Choosing to preserve: towards a cooperative strategy for long-term access to the intellectual heritage. Papers of the international conference organized by the European Commission on Preservation and Access and Deutsche Bibliothek, Leipzig/Frankfurt am Main, March 29-30, 1996 / Edited by Yola de Lusenet. European Commission on Preservation and Access Amsterdam, 1997

Gould, Sara; Varlamont, Marie-Thérèse. The preservation of digitized collections: recent progress and persistent challenges world-wide. // A reader in preservation and conservation. (IFLA Publications; 91). München: K.G.Saur, 2000

Silobrčić, Vlatko. Slobodan pristup ocijenjenim znanstvenim informacijama: moguća budućnost informiranja znanstvenika. // Vjesnik bibliotekara Hrvatske, 50 ½ (2007); 51-61

Digital Archiving in Children's Library

Dunja Holcer
Public Library and Reading Room Sisak, Children's Department
Kranjčevićeva 8, 44000 Sisak, Croatia
dunjaholcer@gmail.com

Summary

Digital technology has given the libraries a way to preserve their materials and documentation in a secure way which doesn't demand as much storage space as paper does. Digital materials and documentation are easily accessible with numerous ways for later use. Children's Department of Public Library and Reading Room Sisak has recognized possibilities digital technology gives and started a project called "Digitization Project of Exhibition Activities". The goal of that project is creation of digital archive of educational and promotional materials and official documentation.

Key words: digital archiving, digitization, children's libraries

Introduction

Libraries are dedicated to archiving promotional and exhibition materials, photos, newspaper articles about library activities as well as annual reports in order to use them if the need ever arises. That leaves a big paper trail to archive and preserve which can be a problem since public libraries do not have enough storage space intended for keeping paper documents and, usually, those storage spaces are too small to keep everything librarians consider important.

Digital technology, digitization and digital carriers can solve that problem by reducing the amount of paper stored inappropriately in a secure way that doesn't need a lot of space at the same time archiving and preserving all the documents for later use.

Children's libraries are living spaces which grow and develop with their users. Important to the librarians and their users are sources of information that are accessible, simple to use, relevant and accurate. Digital archiving provides a way to store and preserve valuable information and educational contents, a mode for all relevant information on certain subject to be accessible in one place and a method to make promotional and archival materials important for every library.

This work will present all the stages in archiving materials in Children's Department of Public Library and Reading Room Sisak: from archiving paper documents and creating first digital documents to complete digital archiving

which provides the way for documents and information to be easily accessible and used as additional source of information for the users as well as promotional material and the material for annual reports, retrospectives etc. for the librarians. The use of this kind of digital archive is unlimited.

The beginning – archiving paper documents

For a very long time Children's Department of Public Library and Reading Room Sisak is dedicated to promotion of education and literacy. The exhibitions are the most important and most suitable way for Children's Department to achieve a connection and cooperation between itself and its users, children, teachers, students…

All the activities that are done for and during the exhibition involve creating materials that are valuable source of information for children as well as for librarians and they are worth preserving for possible later use.

Some years ago, before photocopy machines and computer technology all the work was done by hand. Texts for the exhibition posters were hand written, then cut out, glued to the cardboards or thicker papers and displayed in that way. After the exhibition was closed, handwritten texts were taken down from the cardboards and put away, in pieces, gathered in envelopes and archived in the storage space.

As technology developed, photocopy machines, printers and computers became available for creation of exhibition posters. Texts were printed and pictures photocopied, first in black-and-white and later in color, cut out and glued to cardboards but, after the exhibition, they were still archived in the same way as before, in envelopes, in paper pieces.

What is the end result of this kind of archiving? After a couple of years the oldest exhibition materials started to deteriorate, piles of paper become a problem for storage and, in the end, nothing could be used again since all exhibition materials were just small paper pieces which nobody could sort out.

First digital documents – promotional materials

Interest for the new technology among librarians of Children's Department grew as they realized that it could give them a new and modern way to improve and promote library work so the first step towards the digital archiving happened after two exceptionally successful exhibitions which were held during October and November 2003. By that time Children's Department was equipped with better computers, color printers, scanners, and digital photo cameras were owned by librarians personally, so they were also available for taking photos during exhibition.

During October 2003 an exhibition "Community Policing" in cooperation with Police Department of Sisak and Moslavina County took place in Children's Department. After the exhibition a PowerPoint presentation was made and presented at Interliber – Zagreb Book Fair and then in Osijek at The Conference

for Public Libraries in April 2004 where CD-ROMs containing presentation were given to other librarians as a promotional material.
Second CD-ROM with PowerPoint presentation was made in March of 2004. It was called "Blue World of Dolphins" and it was made to accompany the paper given at annual Spring School for the School Librarians. "Blue World of Dolphins" was an exhibition done in cooperation with the Blue World Institute for Research and Conservation of the Sea from Veli Lošinj and held in Children's Department during Month of Croatian Book in 2003. That CD-ROM was also given to interested librarians as a promotional material.
Both PowerPoint presentations contained a series of photographs taken during the exhibitions, schedule of events, all the reasons for making of the exhibitions, names of people involved, results, reactions of the visitors, newspaper articles, etc. They were made to present a very successful cooperation between different public and scientific institutions, library, schools and kindergartens.
After those two promotional CD-ROMs, librarians of Children's Department made more of them on different subjects presenting all aspects of activities including Preschool Playroom, Art Workshops, Storytelling Corner, Puppeteer Troupe "Together", etc. and most of them were translated to English language for promotional purposes. It, also, became a standard that every bigger exhibition was documented in that way.

Exhibition materials in a digital form – multimedia educative CD-ROM with digital and digitized content

Progress in making previously mentioned CD-ROMs was fast and soon it was not satisfying. In September of 2004 a very extensive ecological exhibition "Woods of Croatia" was prepared in Children's Department and for the first time Adobe Photoshop program was used for designing of complete exhibition posters in digital form which were later printed on A3 paper format and needed no cutting or gluing. That was the first step towards the creation of multimedia educative CD-ROMs with digitized content.
The whole process begins with the arrangements for an exhibition including decision about the topic of the exhibition after which experts are contacted for the purpose of arranging cooperation. Cooperation with various experts, scientists, authors of scientific articles, ambassadors etc. gives an opportunity to collect relevant materials, ephemerals, books, photographs, music, newspaper articles and scientific articles needed for the exhibition. That way it is made sure that materials gathered for the exhibition are accurate and relevant and that permission for the usage of those materials is given.
The next step is to write texts for the exhibition posters. Very important is a selection of pictures and photographs which will accompany and illustrate the texts. Chosen pictures and photographs are scanned usually in 300 dpi color format to ensure the quality of the reproduction and than edited. All scanned

pictures are saved in .jpg format in thematic folders after which additional editing of the pictures in Adobe Photoshop program follows.
The designing of exhibition posters is the process in two steps. First, written texts are copied from MS Word documents to Adobe Photoshop and scanned pictures are added. Next follows the final designing of the exhibition posters. Finished posters are saved in .jpg format but they are, also, saved in .psd format in case additional changes are needed.
Posters are then printed on A3 paper format with no cutting or gluing. Since they exist in digital form there is no real need to archive them in paper form so they can be lended or given to another library to use them.
After the exhibition is over, all digital documents, including exhibition posters used for the exhibition, are collected in one place. Photographs taken during the exhibition are added, as are digitized newspaper articles about exhibition. That is the beginning of making of a CD-ROM which has a multiple purpose: it is educative, fun, promotional, archival. The content of each CD-ROM is different depending on the activities done during exhibition duration time.
This kind of designing and storing of exhibition activities products is an archival record of library activities with a long period of durability, it is simple to use and is a multimedia educative product that has all the relevant information about a certain topic in one place and is accompanied by illustrative music, quizzes of knowledge which can be used for testing knowledge on the subject, etc. At the same time, this kind of product is a very good presentation and promotion of the work that is being done in Children's Department.

All-inclusive digital archiving – plans for the future

Previously described method for digital archiving is now called "Digitization Project of Exhibition Activities in Children's Department of Public Library And Reading Room Sisak". Plan is to develop it further in the way which will include indexing of all the documents concerning exhibitions which will be set in the future and attach it to the search software incorporated on the auto run DVD-ROM. That new approach will enable the librarians and users to quickly find what is needed without opening a number of folders and files but by simply entering key words.

It will include all the posters made for the exhibition which contain all the information about a certain subject (like, for example, history, literature, climate, geographical characteristics, etc. of Canada) in one place, digitized newspaper articles, photographs, data about authors of the exhibition, all the official documentation such as a correspondence between librarians and institutions, schools, experts, etc., interactive quizzes of knowledge made for the exhibition, statistical information about visitors, leaflets, promotional materials, etc.
Exhibitions which are and which will be digitally archived in this way will exclusively interpret topics for which it is known that are of real use, interest and benefit to the users, which promote knowledge, increase possibilities of educa-

tion and develop an interest in preservation of cultural, historical and natural heritage. That is, at the same time, the most important criteria for choosing topics which are and will be interpreted and archived in this way.

Conclusion

Digitization, a modern way of preserving rare and valuable materials that increases an access to visual records of the past and gives various new possibilities in education and promotion of knowledge and heritage is a complex process which includes different types of materials. It is done in several steps and each step considers specifics of a material that is being digitized which requires different expensive hardware and software solutions.

However, to be in a category of digital, documents can be made from scratch in a digital form without expensive equipment. Documents made in that way give a wide range of possibilities for management and use. Materials for exhibitions, research or various activities made in digital form are more suitable for different applications and they are valuable sources of information for librarians as well as users and, therefore, worth archiving and preserving.

Children's Department of Public Library and Reading Room Sisak do not posses valuable and rare materials. It is dedicated to education and promotion of reading and literacy so digitization is set on a different basis from which comes a different methodology and process of making the end product – digital archive of educational and promotional materials and official documentation.

References

Holcer, Dunja. Digitalizacija u dječjoj knjižnici // Godišnjak Knjižničarskog društva Sisačko-moslavačke županije. 1(2007.) ; 12-13

Stančić, Hrvoje. Proces digitalizacije // Informacijske znanosti u procesu promjena / Lasić-Lazić J. (ed.). Zagreb : Zavod za informacijske studije, 2005., 45-84

Vrana, Radovan. Izgradnja digitalnih zbirki: određivanje i izbor kriterija za digitalizaciju knjižnične građe // 7. seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne infrastrukture: zbornik radova / Katić T. (ed.). Zagreb : Hrvatsko knjižničarsko društvo, 2004., 79-85

Application of Digital Information in Prevention and Reconstruction of Monuments, Architecture and Urbanization – Specific Application of BIS Machine

Maria Styblińska
Faculty of Informatics and of Materials Sciences
University of Silesia
ul. Będzińska 39, 41 200 Sosnowiec, Poland
maria.styblinska@us.edu.pl

Małgorzata Gajos
Faculty of Informatics and of Materials Sciences
University of Silesia
ul. Będzińska 39, 41 200 Sosnowiec, Poland
malgorzata.gajos@us.edu.pl>

Krešimir Buntak
CSU d.o.o., Zagreb, Croatia
csu@zg.t-com.hr

Summary

The new information technologies imply new understanding of access to cultural and scientific heritage materials. The preservation and reuse of these digital assets forms both the cornerstone of future civilization growth and development.

The cultural and scientific heritage consists of treasured values, accumulated during the centuries. They are the foundation for the future of human being memory. These can remain beyond time only if we take now the responsibility for their preservation or reconstruction using digital methods. It is necessary to prepare before hand physical copies of the most valuable products of human intellect, in order to save them for the future generations. Increasing amounts of cultural, natural and scientific heritage are being created or represented in digital forms, as we are living in the digital ages.

These digital materials represent our heritage in order to monuments, architecture and urbanization and are our future intellectual capital. The fast pace of change in the technological landscape makes ensuring long-term access to these materials as a challenge.

The paper presents some specific theoretical and practical aspects of preservation of cultural assets in digital form in long-term aspects and practical issues in information digitisation through the practical example of some part of BIS Machine related to Apoksiomen, as a connection of digitisation and production of monuments.

Key words: monuments, architecture, threats, digital information, digital preservation, reconstruction

Introduction

The cultural and scientific heritage consists of treasured values, accumulated during the centuries. They are the foundation for the future of human being memory. These can remain beyond time only if we take now the responsibility for their preservation or reconstruction using digital methods. It is necessary to prepare before hand physical copies of the most valuable products of human intellect, in order to save them for the future generations.

The digital world is still relatively young and immature in relation to the larger information universe, parts of which have been under development for centuries. Therefore nowadays information is being generated in greater quantities and at higher frequency than at any time in human history. Such information is penetrating and transforming nearly every aspect of our culture and life. Increasing amounts cultural, natural and scientific heritage is being created or represented in digital form.

In recent years, rapid advances in digital technologies, from 3D graphics, CAD systems, etc. to multimedia, and virtual reality, have given heritage new hope: from archaeology to architecture, emerging digital tools offer promise in documenting, analysing, and disseminating culture. From the now obliterated treasures as Buddhist statues at Bamiyan, Afghanistan, to Minoru Yamasaki’s, Twin Towers’ in New York, Croatian sculptures and monuments laying on the deep end of the Adriatic, and so on, we owe it to future generations to harness our digital tools to preserve and protect the world’s cultural legacy.

Cultural Transformations and Digital Dependence.

The increasing dependence upon digital information is having several dramatic effects. It is changing the way in which our culture is recorded and our culture itself is being transformed. If we think of physical products of culture as artefacts, we should also be thinking of digital and electronic products as digital-facts (d-facts) or e-facts. However digital-facts are very fragile, cause digital technology is very fragile also. Digital information is inherently more fragile than traditional technologies – more easily corrupted or altered, without recognition. Digital storage media have shorter life spans and require access technologies that are changing at an ever-increasing pace. If we think of technologi-

cal advances, need to consider the time frame in which we archiving become much shorter.

The cultural heritage reflects the way of life, the history and identity, provides the link between the past, the present and the future, and contributes also the economic sustainability and welfare. People, a community cannot function effectively in the larger societal context if deprived of its identity through the loss of its cultural heritage. "The digital heritage is inherently unlimited by time, geography, culture or format. It is culture specific, but potentially accessible to every person in the world. Minorities may speak to majorities, the individual to a global audience. The digital heritage of all regions, countries and communities should be preserved and made accessible, so as to assure over time representation of all peoples, nations, cultures and languages" [19].

Definition of the architectural heritage

For the purposes of Convention for the Protection of the Architectural Heritage of Europe [7] (Granada, 3.X.1985), the expression "architectural heritage" shall be considered to comprise the following permanent properties (art. 1):

1. Monuments: all buildings and structures of conspicuous historical, archaeological, artistic, scientific, social or technical interest, including their fixtures and fittings;
2. Groups of buildings: homogeneous groups of urban or rural buildings conspicuous for their historical, archaeological, artistic, scientific, social or technical interest which are sufficiently coherent to form topographically definable units;
3. Sites: the combined works of man and nature, being areas, which are partially built upon and sufficiently distinctive and homogeneous to be topographically definable and are of conspicuous historical, archaeological, artistic, scientific, social or technical interest.

The architectural heritage constitutes an irreplaceable expression of the richness and diversity of cultural heritage, bears inestimable witness to the past and is a common heritage.

Treats and responsibility

Our ability and commitment as a society to preserve cultural memory, and natural and science heritage are far from secure. It is so easy to damaged and lost all these in dramatic immediately, only if remind hurricanes (lately Katherine), tsunami, terrorist's attacks (11 September 2001), fires or floods

Threat – be called out phenomenon the working the strengths of nature man, which cause that the feeling of safety, are diminishes completely it disappears. We distribute the threats on natural (e.g. the impulsive defeats) and connected from activity of man (distribute these on: treats of the civilisation, e.g. the mass events of, disease; destructive treats, e.g. the terrorism, crime, sabotage; economic treats, e.g. the dirt of environment, defective constructions). Treats can

be divided from regard on sizes (territory) on which it sets – global treats, regional and local [2].
Examples of threats depending of monumental, sculptural and architectural heritage [12], besides mentioned before:

- conditions of settled;
- the working the moisture and water;
- the atmospheric factors;
- the chemical aggression of environment;
- the biological processes;
- the strain corrosion of materials;
- the influence of constructional defects and execution;
- the influence of underground excavations;
- kinematics of geological movements;
- decay and inappropriate using of relics, their thefts, disappearances; illegal transportation for border;
- lack or inappropriate devolution of protective heritage legislation;
- lack of awareness of their heritage values;
- lack of expertise for specific materials repairs;
- pressure from urban growth and redevelopment;
- pressure from application of uniform building codes.

The Limits of Digital Technology and the Technological Obsolescence

However digital technology poses also new threats and problems as well as new opportunities. Its functionality comes with complexity. Reading and understanding information in digital form requires equipment and software, which is changing constantly and may not be available within a decade of its introduction. Digital information today is produced in highly varying degrees of dependence on particular hardware and software.
The digital world is still too new for us to describe fully the life cycle of the information objects that do now or will in the future reside there, but what surely unites the community of actors in their various information-based activities is their common purpose in support of the pursuit of knowledge.
Knowledge cannot advance without consistent and reliable access to information sources, past and present. The application of computer hardware and software has also generated other new kinds of information objects, including the products of simulation, remote sensing, computer-aided design (CAD) and information systems (IS). These objects come into being and exist as creatures of the digital environment; if nurtured well, digital technologies will certainly get still other kinds of information objects, which we can now only anticipate.

Preservation requirements

In the digital world, to succeed in the preservation of digital objects, preservation measures must ensure that as many of these aspects as possible persist over time. In preserving a digital object, we aim to:

- Fix the object as a discrete whole – the boundaries of digital objects are less clear, especially if they are compound objects created by assembling different media or by linking to resources from around a network.
- Preserve the physical presence – preserving the physical file does not mean that the object will remain accessible.
- Preserve content – refers to maintaining the ability to access the content at its lowest level, such as ASCII text, without the embellishments of font variations and layout features.
- Preserve the presentation – content is typically rendered in some presentation, format or layout. To retain the original look of a document, these layout specifications must also be preserved, especially when they contribute significantly to the understanding and interpretation of the content.
- Preserve functionality – digital objects can contain multimedia components (i.e. text, graphics, audio, and video), exist in hypertext format, contain dynamic content generated automatically from data stores, or have navigation functions, such as toolbars, keyword search, or interactive tables of contents. Special efforts must be made to preserve the functionality.
- Preserve authenticity – activities to guard authenticity include securing digital objects against unauthorized changes and monitoring digital objects through multiple "copying" cycles to ensure that each copy is an acceptable rendition of the original.

Locate and refer to the digital object over time – objects can be readily altered, copied or moved. An individual must be able to match a citation to a digital object, and to distinguish it from other versions or editions.

- Preserve provenance – provenance is an archival concept that asserts the origin and chain of custody of an object and contributes to defining it as a whole. Imprint statements and bookplates. Establishing an object's origin and history help confirm that the work is authentic and its content is intact.
- Preserve context – digital objects are partly defined by their hardware and software dependencies, their mode of distribution and linkages to other digital objects. Preserving digital objects may mean weaning them from some technical dependencies, changing the mode of distribution, and deactivating links. In these circumstances, preserving context is a particular challenge.

Digital Preservation

The term "digital preservation" refers to both preservation of materials that are created originally in digital form and never exist in print or analogue form (also called "born-digital" and "electronic records") and the use of imaging and recording technologies to create digital surrogates of analogue materials for access and preservation purposes.

Preservation must allow future users to retrieve, access, decipher, view, interpret, understand, appreciate, and experience informational entities in meaningful and valid ways. It is required specifically exclude the potential use of digital technology to preserve the original artefacts through digitisation.

Digital preservation strategies are:

I. Technology preservation – means of overcoming technological obsolescence by retaining the hardware and software used to access the digital resource.

II. Migration of Digital Information – refreshing digital information by copying it from medium to another medium and the possibility of maintaining a complex set of emulators describe two distinct points on a continuum of approaches to preserving digital information. Data migration includes: medium refreshing, medium conversion, format conversion, version upgrade and migration of technical environment.

III. Change Media – transfer digital materials from less stable to more stable media. Copying from one medium to another has the distinct advantage of being universally available and easy to implement.

IV. Change Format – the migration strategy for digital archives with large, complex, and diverse collections of digital materials is to migrate digital objects from the great multiplicity of formats used to create digital materials to a smaller, more manageable number of standard formats that can still encode the complexity of structure and form of the original.

V. Emulation – refers to the process of mimicking, in software, a piece of hardware or software so that other processes think the original equipment/function is still available in its original form. The essential idea behind emulation is to be able to access or run original data/software on a new/current platform by running software on the new/current platform that emulates the original platform. Emulation means technological obsolescence of hardware and software by developing techniques for imitating obsolete systems on future generations of computers. Three options are: emulate applications, emulate operating systems, and emulate hardware platforms.

VI. Adherence to standards – adhering to stable and widely adopted open standards when creating and archiving digital resources.

VII. Backwards compatibility – being able to retain accessibility to a digital resource following upgrade to new software and/or operating systems.

VIII. Encapsulation – grouping together a digital resource and whatever is necessary to maintain access to it. This can include metadata, software viewers, and discrete files forming the digital resource.
IX. Converting to stable analogue format – converting certain valuable digital resources to a stable analogue medium.
X. Digital archaeology – rescuing digital resources, which have become inaccessible as a result of technological obsolescence and/or media degradation. Museums of hardware and software where ancient versions of hardware, operating systems, and applications software would be lovingly preserved so that people could read old data. This idea born some obstacles, but is sounding very nostalgic.

Implementations of New Technologies

The new technology has brought some new possibilities enabled to give effect through original copies production and the databases for valuable cultural heritage objects. One of a known method and methodology of an implementation of new technologies (IT, mechanical and other) is a "BIS Machine" – authors Kresimir Buntak and Davorin Kerekovic, which have been presented in Warsaw at 2004 year.
"BIS Machine" system could be described in flow diagram of work process. The diagram shows step by step every work process and check points necessary to evaluate all production steps during replica production.

Phase I

1. Object selection
2. Object imaging harmonized with requirements and possibilities (boundaries) by measuring, D scanning, photo-grammatically, digitally analogically or using combination of the aforesaid data entry methods
3. Processing of the imaged (quantified) data processing
4. Developing of imaging (quantifying) reports
5. Formation of the virtual object i.e. digital file etalon pattern
6. Data direction and harmonization

Phase II

7. CAD preparation
8. CAD model design
9. Validation and -pattern (etalon) calibration
10. Re-design – if necessary
11. Prototyped model design (rapid prototyping – 3D printing and the like)

Phase III

12. Material selection
13. Selection of machine and tool
14. CAM program development
15. Computer development simulation

16. Machine or device-developed production
17. Validation of the developed object
18. CAM program redesign – if needed
19. Sorting and saving all data systematized for data base
20. Object delivery and fitting

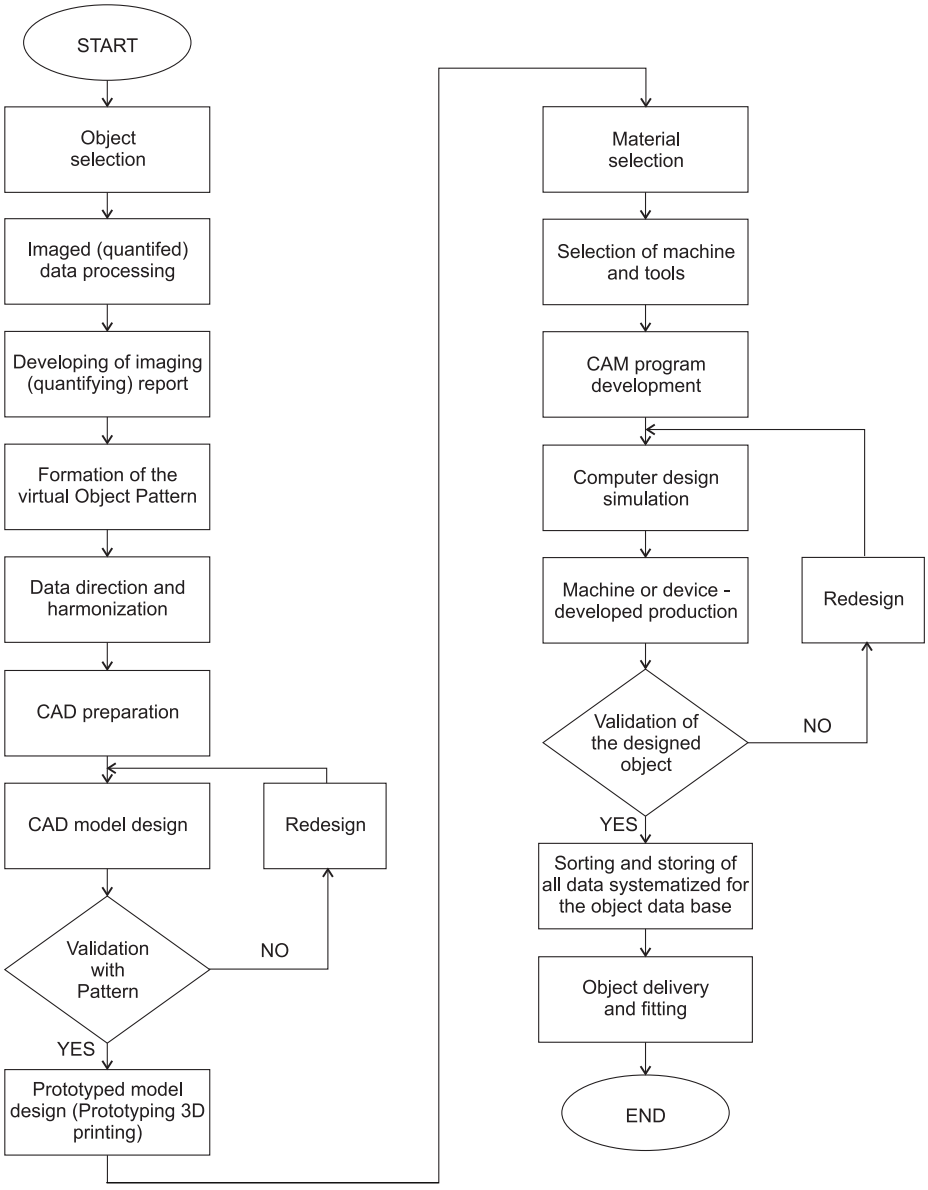


Figure1. Diagram of process flow of the BIS Machine’s method.

"BIS machine" is the method composed of three phases enables of the documenting and makes virtual " the solid "the CAD of model over selections of materials and technologies

Main part of "BIS machine method" is

1. Stereophotogrametric survey of object
2. 3D modelling of received graphic data
3. Production of replica with one of several solutions able to give result

This process flow is prepared according with ISO standards and demands and represents step foreword in the sense of methodology and quality in project preparation and realisation.

Beside replica application of BIS Machine method gives database of targeted object as a very important aim of project. "BIS machine" it is the method composed of three phases enables of the documenting and makes virtual " the solid " the CAD of model (the virtual standard), over selections of materials and technologies, to the creation of replication or the original copy of very large precision (to 1 μm) (Figure 2).

Figure 2. The Replica of the head

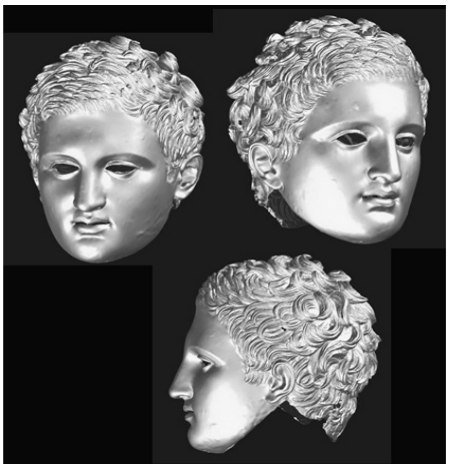


Figure 3. The 3D virtual model of Apoxiomenos

Antique sculpture entitled "Apoxyomenos, il Bronzo della Croazia" excavated from the sea in 1999 near Losinj island, in the northern Adriatic. The sculpture of 192 cm height originated from the period between Classical Greece and Hellenism in the 4th century BC and represents one of the masterpieces of world cultural heritage.

Applying that method it is possible to displace and safe valuables sculpture, parts of rare architecture masterpiece or other object in a way that the original copy will occupy the place of original (Figure 3).

Conclusion

Increasing amounts of cultural, natural and scientific heritage are being created or represented in digital forms, as we are living in the digital ages. Human being will need to develop a range of strategies to ensure the preservation of and access to various categories of digital objects. Custodial and non-custodial arrangements will need to be considered both from preservation and an access perspective. One of the methods of preserving real world to digital one it is a BIS Machine. One of the final goals of this method it is the creation of virtual museums.

References

1. 20th Century Heritage; ICOMOS. International Council of Monuments and sites. http://www.international.icomos.org/20th_heritage/index.html [access 30.09.2007]
2. A Digital Dark Ages? Challenges in the Preservation of Electronic Information. Paper delivered at a workshop of the Audiovisual & Multimedia, 63rd Annual IFLA General Meeting, Copenhagen, Denmark. September 4, 1997. Published in *International Preservation News 17*, May 1998
3. Article: Zagrożenie (Threat); http://pl.wikipedia.org/wiki/ [access 30.09.2007]
4. Burke, Sheridan. Trends, Threats and Risks, http://www.international.icomos.org/risk/2001/synthesis.htm [access 30.09.2007]
5. Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocol No. 11, http://conventions.coe.int/Treaty/en/Treaties/Html/005.htm, [access 30.09. 2007]
6. Council of Europe Treaties, http://conventions.coe.int/general/ [access 30.09.2007]
7. Digital Preservation Coalition; http://www.dpconline.org/graphics/orgact/storage.html [access 08.2007]
8. Digital Preservation; http://www.digitalpreservation.gov/ndiippch.html [access 08. 2007]
9. European Treaty Series – No. 121; http://conventions.coe.int/Treaty/Commun/ [access 10.09. 2007]
10. Hodge, Gail; Carroll Bonnie C. Digital Electronic Archiving, The State of The Art and The State of The Practice, *A Report: International Council for Scientific and Technical Information Policy Committee*, CENDI, Oak Ridge, TN, 1999,
11. Kerekovic, Davorin; Buntak, Kresimir, Dautovic, B. ENCORE machine. Authors solution, Cultural of heritage yen the mugs of threats the yen war and peace thereby, The international conference organized he the occasion of 50th anniversary of signing of Hague convention the forward pass the protection of cultural will crucify the yen the of armed conflict, Warsaw, 13-15 May 2004
12. Kerekvic, Davorin. The modern technologies in cultural heritage management and reconstruction. BIS Machine. Informatyka a ochrona dóbr kultury. Materiały seminarium naukowego. M. Gajos, J. Gawlica (eds.) Sosnowiec: Uniwersytet Śląski, Instytut Informatyki, 2005
13. Research Challenges in Digital Archiving and Long-term Preservation, Report on the NSF Workshop on Research Challenges in Digital Archiving. Towards a National Infrastructure for Long-Term Preservation of Digital Information, 2002, Library of Congress National Digital Information Infrastructure and Preservation Program, USA

14. Seamus, Ross. Changing Trains at Wigan. Digital Preservation and the Future of Scholarship, National Preservation Office
15. Styblińska, M; Long-Term Preservation of Digital Information vs. User Expectations and Needs – Some Specific Aspects; International Multiconference on Computer Science and Information Technology, Wisła 2006, Poland; CD-edition
16. Syllabus. http://www.syllabus.agh.edu.pl/GR/C-GiHMC-MJ/main7.html [access 10. 09.2007]
17. UNESCO Charter on the Preservation of the Digital Heritage, Article 9 – Preserving cultural heritage; http://portal.unesco.org/ [access 10.09.2007]

IDENTITY IN ELECTRONIC ENVIRONMENT AND ITS PRESERVATION

Mediji i demokracija: baza podataka, dubinsko pretraživanje, izravni pristup – digitalizacija novinskog arhiva i organizacija znanja u bazu podataka

Željko Hodonj
Hrvatska izvještajna novinska agencija HINA
HINA informacijska baza / HIB
Informacijsko dokumentacijski centar / IDC
Marulićev trg 16, Zagreb, Hrvatska
zeljko.hodonj@hina.hr

Lana Žaja
Hrvatska izvještajna novinska agencija HINA
HINA informacijska baza / HIB
Informacijsko dokumentacijski centar / IDC
Marulićev trg 16, Zagreb, Hrvatska
lana.zaja@hina.hr

Sažetak

HINA informacijska baza jedan je od alata kojim se služi HINA u vlastitom organiziranju, HIB Informacijsko dokumentacijski centar povezuje interes korisnika sadržaja što ih HINA generira kao svoj proizvod osiguravajući dostupnost javnog znanja. HIB/IDC suvremenim načinom osigurava zaštitu i javnu upotrebu hrvatske kulturne baštine zabilježene u tiskanim i AV medijima. HIB IDC podržava umreženje s bazama u okruženju. Godišnje će obraditi milijun dokumenata, tijekom 2008. godine razvit će automatsku kategorizaciju sustava.

Ključne riječi: Hina, baština, digitalizacija, Hrvatski sabor, baza podataka

Uvod

Hrvatsku izvještajnu novinsku agenciju osnovala je Republika Hrvatska s ciljem da prikupi i pošalje informacije utvrđene na činjenicama o zbivanjima u RH i u svijetu za potrebe medija i drugih sudionika javnog života. HINA djeluje po načelima neovisnog, nepristranog i profesionalnog novinsko-agencijskog izvještavanja. Svakodnevno Hina priprema dvadesetak općih i specijaliziranih servisa vijesti u kojima se dnevno na hrvatskom i engleskom jeziku objavljuje oko 400 tekstova s više od 150.000 riječi. Hina pruža i multimedijske usluge, ima vlastiti

foto (FaH) i audio servis te elektroničku bazu podataka (EVA). Tijekom 24 sata Hinine vijesti prima oko 500 korisnika u zemlji i svijetu. U agenciji je ukupno 180 stalno zaposlenih, od toga 130 novinara i urednika. Povremeno za Hina-u izvještava još šezdesetak vanjskih suradnika-dopisnika iz zemlje i svijeta. Hina je od 1993. članica Europske udruge novinskih agencija EANA čije su članice su nacionalne novinske agencije iz 30 europskih zemalja.

Hrvatska izvještajna novinska agencija – HINA, od osnutka 1990. godine, opredijelila se da će djelovati pomoću opće pristupnog interaktivnog sustava baze podataka. U samom početku to je značilo:

1. Stručno, sa stajališta agencijskog novinarsko uređivačkog posla, trebalo je svoju svakodnevnu aktivnost podvrgnuti normama obilježavanja / atribucije svakog novonastalog sadržaja: tekstualnog, vizualnog ili auditivnog.
2. Stvorena je EVA Hinina baza podataka s današnjih 1,3 milijuna dokumenata – pokazana je dimenzija zadatka pohrane dokumenata već na razini tekstualnih formi, njihove organizacije i pretraživanja. Trebalo je odgovoriti na pitanja dostupnosti sadržaja za krajnje korisnike. Bio je to veliki izazov novinarima, urednicima, IT timu.

   Bilo je potrebno primijeniti opće prihvaćene standarde za taj posao – IPTC kategorizaciju[1].

   EVA je postala sa stajališta zaštite baštine, neprocjenjiv izvor podataka / dokumenata o stvaranju samostalne hrvatske države, o događajima i osobama koji u svakodnevnom agencijskom djelovanju predstavljaju sadržajne silnice novinarskih tekstova.

U ovom ćemo tekstu predočiti zahtjeve koji su utjecali da se na jednom mjestu dogodi sinergija odgovora na pitanja o zaštiti baštine, podržavanju demokracije i dostupnosti organiziranog javnog znanja.

Na prijedlog Vlade RH 2006. godine, Hrvatski sabor je potvrdio zaštitu najveće novinske baze podataka u papirnatom i elektronskom formatu – Vjesnikovu novinsku dokumentaciju / VND online. Tako su stvoreni preduvjeti da se VND

[1] IPTC International Press Telecommunications Council – IPTC međunarodna organizacija novinskih agencija koja je osnovana sredinom prošlog stoljeća s ciljem unapređenja i profesionalizacije međusobne razmjene (tehničkih i novinskih) podataka. U tom kontekstu započelo je korištenje IPTC dokumenata koji su pomagali jednostavnijoj klasifikaciji i organizaciji. Prvenstveno je taj oblik metapodataka bio rješenje za upravljanje slikovnim dokumentima. Kao što je Eurovoc organiziran prema potrebama rukovanja spisima Europske unije, i to mu je odredilo strukturu i metodu označavanja, tako je i IPTC imao svoju prvotnu namjenu. Takav oblik označavanja metapodatcima zahtijevao je točnost i jednoznačnost područja. Među pojmovima međuodnos je bio određen širi/uži pojam. No, gledajući prema osnovnoj namjeni IPTC kategorizacija i njezina osnovna "područja" bliža su klasifikaciji HIB/IDC (odnosi se na razdoblje prije povezivanja s HINA-om kada je to još bila baza VND online) nego klasifikaciji Eurovoca jer je polazišna građa za koju se klasifikacija organizira sadržaj javnih glasila, novinski tekst. (Danijela Miklec)

proglasi razvojnim projektom HINA-e; te da se financira zaštita i razvoj baze podataka sredstvima državnog proračuna.[2]
Tako je projekt digitalizacije zaštićenog novinskog arhiva Vjesnikove novinske dokumentacije s dvanaest milijuna dokumenata i njegova organizacija u javno dostupnu bazu znanja postao je 2006./2007. godine nacionalni projekt i započela je njegova integracija u sustav Hrvatske izvještajne novinske agencije.
Postupajući u skladu s takvim opredjeljenjima Hrvatska izvještajna novinska agencija odmah je uputila Ministarstvu kulture RH projektnu ideju o povezivanju Hinine informacijske baze EVA i Vjesnikove novinske dokumentacije VND.
Osnova za pokretanje projekta bila je kako sljedeća:
Članovi Odbora za medije Hrvatskog sabora inicirali su u lipnju 2006. objedinjavanje Hinine baze podataka Eva i Vjesnikove novinske dokumentacije kako bi ta vrijedna arhiva mogla biti na korist državi i drugim korisnicima.
U saborskoj raspravi povodom Izvješća o radu Upravnog vijeća Hine za 2005/2006. godinu zatraženo je da Vlada razmotri mogućnost povezivanja HINA baza EVA i VND.
Mišljenje je Hrvatskog državnog arhiva iz siječnja 2006. g. da se VND, budući da Vjesnik d.d. to nije u mogućnosti, povjeri na skrb i korištenje institucijama poput HINA-e, HIDRA-e ili druge srodne ustanove.
Kao ciljevi projekta određeni su očuvanje primarne, recentne dokumentacijsko-informacijske vrijednosti gradiva Hinine baze EVA i VND, kao baze podataka i kao povijesno-kulturnog dobra; potpuna digitalizacija baze podataka; stvaranje i razvoj novih, samostalnih i međusobno povezanih informacijskih usluga za različite skupine korisnika (media monitoring, on line novinska dokumentacija i posebni projekti); komercijalno iskorištavanje gradiva kako bi se budući razvoj i održavanje u što većoj mjeri samofinancirali.
Takav kontekst djelatne zaštite kulturnog dobra predstavlja izuzetno kvalitetan i odlučan odgovor na pitanja o modelima zaštite baštine.[3] Suvremene okolnosti zajedničkog života svojim su posebnostima tražile da zajedno s drugima odgovorimo na izazove vremena u kojem živimo.

[2] Krajem 2006. godine u Zagrebu je boravio prof. emeritus Robert Hayes i s nama je proveo desetak sati svog dragocjenog vremena. Potpuno besplatno je vrednovao naše rezultate dajući nam snažnu podršku za nastavak rada i integraciju u sustav HINA-e. O tome je iscrpno razgovarao s Fjodorom Polojcem, pomoćnikom ravnatelja HINA-e, koji je bio zadužen za provođenje projekta povezivanja i s Darkom Gulijom, voditeljem IT tima HINA-e. Vjerujemo da ćemo se prof. Robertu Hayesu uskoro moći dolično zahvaliti.

[3] Državnim proračunom RH za 2007. godinu osigurana su sredstava za razvoj projekta ali je zbog administrativne procedure tijekom ustrojavanja uprave HINA-e, postupak provedbe projekta HIB/IDC započeo tek 1. lipnja 2007. umjesto 1. siječnja 2007. godine.

Historijski je trend[4] da se dominantne funkcije i procesi u informacijskom dobu sve više organiziraju oko mreža. Kultura se dubinski transformira; nakon tisućljeća dominacije prirode nad kulturom i stoljeća dominacije kulture nad prirodom, ulazimo u doba dominacije kulture nad kulturom, gdje priroda postaje kulturni (re) konstrukt. Teoretičari poput Manuela Castellsa, sugeriraju da su čimbenici zaduženi za upravljanje našom tranzicijom u informacijsko doba državne institucije te da nam je u toj tranziciji potrebno političko zastupanje, demokracija u kojoj svi sudjeluju, procesi iznalaženja konsenzusa…
Trebamo imati na umu i to da je razvoj tehnologije rezultat triju društvenih procesa: potrebe ekonomije za većom fleksibilnošću, koja je u konačnici dovela i do uspona elektroničkog poslovanja; potrebe društva u kojem vladaju vrijednosti individualnih sloboda koje su omogućile trijumf umreženog individualizma – stvaranje mrežnih zajednica prema individualnim afinitetima; a (tek tada) i napretka u računarstvu i telekomunikacijama koji se dogodio krajem milenija.

Znanje – novac i moć

U takvom kontekstu trebalo je izabrati način pomoću kojeg će nam vlastiti dokumenti biti dostupni, organizirani u baze znanja, dostupni u svakom trenutku nama i drugima. Osobito nam je bilo važno omogućiti uvid u sadržaje koji su "prekriveni" prvom razinom čitanja.
Rečeni navod potkrijepit ćemo primjerom svoje prakse organizacije baze podataka. Primjer je krajnje jednostavan: pomoću javno dostupnih izvora / baza podataka možemo otkrivati podatke i sadržaje, odnose i procese koji su izvan standardne pozornosti. Primjer pokazuje potencijalni smjer za provjeru kolika je dostupnost i organiziranost baštine, odnosno koliko su nam dostupni kulturni artefakti vremena i to oni koji su potrebni znanju, sudjelovanju i odlučivanju.
Izabrali smo skup podataka koji upozoravaju na informacijski paradoks što se u opasnoj mjeri može preobraziti u grotesku. Riječ je o projektu koji je od VND naručio Ured glasnogovornice Ministarstva pravosuđa RH 2004. godine s namjerom da dobije sliku o recepciji i percepciji reforme pravosuđe u Hrvatskoj.
Od tada do danas ta posebno projektirana baza podataka broji više od šezdeset tisuća dokumenata/članaka iz novina. Početkom 2006. godine projekt je predefiniran u *custom made product,* a trebalo je provjeriti koliko i kako hrvatska tiskana glasila pišu o korupciji, sukobu interesa, odnosno kako tiskana glasila prate i interpretiraju nastojanja hrvatske izvršne vlasti da obračuna s korupcijom na svim razinama hrvatskog društva.
Od 1. siječnja 2006. do 31. prosinca 2006. godine u bazi podataka VND online, područje Pravosuđe, sudstvo i kriminal; obrađen je 22.791 dokument/tekst iz 16 hrvatskih novina (Jutarnji list, Novi list, Večernji list, Slobodna Dalmacija, Glas

[4] Manuel Castells, Informacijsko doba, Golden Marketing, Zagreb 2000/2002, Internet galaksija, Jesenski i Turk, Zagreb; 2003.

Slavonije, Vjesnik, Poslovni dnevnik, Globus, Feral, Nacional, Arena, Sportske novosti, Banka, Privredni vjesnik, Glorija, Vijenac).
Obrađeni su svi članci koji su imali relaciju prema temama: mito i korupcija, zloporaba položaja i ovlasti, sukob interesa.
Usred opće kampanje za borbu protiv korupcije, za tzv. nultu točku tolerantnosti na korupciju, obrada izabranih glasila pokazala je da od ukupnog broja objavljenih tekstova koji su imali sadržajne veze s područjem Pravosuđe, sudstvo i kriminal, samo je 7 posto, odnosno u apsolutnom broju 1633 članka, imalo dodirnih sadržaja s pojmom ili konotaciju prema korupciji. Usporedbe radi, broj tekstova o korupciji od početka devedesetih do 2006. godine, s razine stotinjak napisa godišnje, porastao je do broja koji sada spominjemo. Tekstovi koji utječu na formiranje stavova su: kolumne (33), komentari (44), tzv. pisma (17); naglašavaju moralne principe i opisuju desetak istih događaja ili osoba. U toj godini novinari istraživači nisu objavili toliko tekstova koliko kolumnisti i komentatori. Možda je lakše pisati i objavljivati tekstove o principima nego istraživati stvarnost. Izabrana tiskana glasila zauzimaju približno 90 posto novinskih naklada. Informacije o korupciji najčešće su iz "druge ruke".
U to je doba Hrvatska dobila ozbiljna upozorenja Europske komisije o korupciji u hrvatskom društvu, a poruke su bile potkrijepljene: anketama građana, istraživanja stavova građana i tome slično.
Za sada su bez odgovora pitanja o tome kako je moguće da ocjene o općoj korumpiranosti društva, što izravno dovodi u pitanje demokraciju *per se*, nemaju uporište u sadržajima hrvatskih tiskanih glasila. Hipotetički, možda je riječ o tome da – korupcija nema takvog maha i utjecaja koji joj se pripisuje?
Ili – mediji su korumpirani i zato su takve teme izvan njihova interesa?
Ili nešto treće?
Odgovor na takva pitanja moguć je temeljitim analitičkim i vrijednosnim uvidom u baze podataka, dubinskim pretraživanjem brojnih relacija.
Sliku jednog društva s njegovim kulturnim artefaktima, među kojima je i korupcija, treba promotriti što cjelovitije.
Klasične teorije demokracije pretpostavljaju da je društvena aktivnost građanska dužnost, a biračko tijelo ne bi trebalo biti samo "informirano", nego bi trebalo kontinuirano procjenjivati i tako kreirati pučki interes, stvarati svojevrsnu samoinicijaciju za participaciju u društvenim poslovima. Biračko bi tijelo, s te točke gledišta, trebalo sudjelovati promišljeno i tolerantno prema različitim gledištima držeći se norme da je javni interes povrh pojedinačnog.
Praksa je pokazala da je takva omnikompetencija iznimka, a ne pravilo. Većinom je građanska aktivnost posljedica incidentnih poticaja ili ih – mobiliziraju mediji.
Javna glasila, mediji koji se bave političkim / javnim temama, ne samo da ocrtavaju / opisuju političko okruženje; oni u sve većoj mjeri sami upravo predstavljaju / jesu – političko okruženje.

Primijenjeno na pojedinačnog građanina – rastući dio javnog djelovanja zbiva se pod utjecajem medija.
U tom je kontekstu moćni utjecaj novih medijskih formata: Interneta i World Wide Weba; radija (govorni) i televizijskih kontakt programa, ali i raznovrsnih zabavnih programa koji pozivaju (političke i nepolitičke) goste u zabavnu raspravu o političkim pitanjima (poput zagrebačke PIRAMIDE, by Merlić & Ogresta)
Kroz te elektronske sustave građani su izloženi političkim informacijama, politički im se lideri čine dohvatnima, uključuje ih se u javne rasprave o važnim pitanjima, a političari su izloženi izravnom javnom propitivanju.
Sumnjičavci upozoravaju da novi mediji tako bespovratno utječu na žrtvovanje izravne, osobne komunikacije. Promatrano s te točke gledišta – tako se stvaraju uvjeti za manipulaciju i podržava se društvena izoliranosti. Ta skepsa nije neuvjerljiva ali uvjerljivo je i to da interaktivni mediji u najmanju ruku mogu potaknuti pojedinca / građanina na samoaktivnost.
Zato je HINA u svom strateškom opredjeljenju odlučila da sve svoje sadržaje i sadržaje drugih; i u dogovoru s njima; generira kao baze podataka "podržavajući umreženje svakog pojedinog građanina Hrvatske. Glavni je cilj osigurati dostupnost podataka iz javnih glasila pomažući osviještenu obaviještenost o izabranim razinama javnog znanja u hrvatskom društvu".
Zahvaljujući dosljednoj podršci Vlade RH prema poštivanju takvih kriterija djelovanja HINA-e, stvoreni su uvjeti za razvoj HINA informacijske baze / HIB Informacijsko dokumentacijskog centra.
Stoga je važno znati što je to motiviralo odluke Vlade RH i Hrvatskog sabora da 2006. godine donesu odluku kojom je najveća baza novinskih podataka u Hrvatskoj poznata kao Vjesnikova novinska dokumentacija, VND on line, zaštićena kao prvorazredno kulturno dobro.
Citiramo iz pismohrane Hrvatskog sabora[5]**:**
"Vjesnikova novinska dokumentacija – VND online" novinski je dokumentacijski centar organiziran u računalnom mediju i jedini je u Hrvatskoj. Nastao je na temeljima Vjesnikove novinske dokumentacije koja je ustrojena 1964. godine.[6]
Od tada do 2005. godine kreirano je više od 6.000 zbirki dokumenata / izrezaka, članaka / tekstova iz novina s približno 12 milijuna pojedinačnih zapisa, a

[5] Interpelacija o odgovornosti Vlade RH za stanje društva Vjesnik d.d., 20. sjednica Hrvatskog sabora 1. lipnja 2006. godine.

[6] Osnutak novinskog arhiva nekadašnje Vjesnikove kuće izravno se vezuje uz ime Bože Novaka, glavnog direktora NIŠPRO "Vjesnik" 1964. godine. Od tada pa sve do osnutka HINA informacijske baze / HIB Informacijsko dokumentacijski centar, bazu su održali dokumentalisti: Anđelka Grubišić Čabo, Vlado Kvitek, Deša Kolar, Jasminka Cesar Miličević, Božica Jurčić, Luca Jelić; i dosljedno inzistiranje Hrvatskog državnog arhiva da građa VND bude zaštićena kao baština iznimne spomeničke vrijednosti.

uza sve to je i posebna zbirka članaka za približno 100.000 domaćih i inozemnih javnih osoba. Baza podataka u izrescima članaka, mjereno smještajem na policama, iznosi približno 480 dužnih metara. Baza podataka u računalnom mediju[7] *broji 600.000 dokumenata.*
VND je projektirana za pohranu i organizaciju poveznica starijih i novih podataka..
Tako su ostvarene, kao jedine u Hrvatskoj, posebne eksportne baze podataka tiskanih glasila na DVD-nosačima.
Baza podataka VND online u svom je papirnatom dijelu zaštićena normama/nadzorom Državnog arhiva. (...) Baze podataka su dio organizacije sustava i iznimno su važne u nacionalnoj mreži kulture. U svom općem servisu digitalne baze podataka razvijaju i poslovnu djelatnost povezivanjem agencijskih sadržaj i media monitoringa, a ustrojavanjem posebnih baza podataka izravna su podrška razvoju i postojanju novih medija.
Vlada Republike Hrvatske izrazila je Nadzornom odboru i Upravi Društva izričito protivljenje da se građa i djelatnosti Vjesnikove novinske dokumentacije (dalje VND) prodju (privatiziraju) jer je riječ o javnom arhivskom gradivu. Istog mišljenja je Hrvatski državni arhiv koji je 17. siječnja 2006. godine uputio pismo sličnog sadržaja Ministarstvu kulture.
Istodobno, ako Vjesnik d.d. ne može dalje skrbiti o djelatnosti i građi VND, primjereno je da o navedenom skrbe i da ga koriste institucije koje se primarno bave informacijsko-dokumentacijskim poslovima. Primjerice: Hrvatska izvještajna novinska agencija (HINA)… (...)
Vjesnikova novinska dokumentacija od 26. studenog 1991. godine u režimu je zaštite Državnog arhiva u Zagrebu (Rješenje Historijski arhiv u Zagrebu, Opatička 29. od 26. studenog 1991.): "Građa je od izuzetne vrijednosti u prvom redu za praćenje i analizu događaja kako za period RAT U HRVATSKOJ tako i za sva zbivanja od 1945. do danas u našoj Republici i šire."
Povijesni arhiv u Zagrebu svojim rješenjem 22. srpnja 1994. godine na temelju Zakona o zaštiti arhivske građe donio je rješenje kojim se nalaže: "Sačuvati cjelovitu dokumentaciju te se pojedini dijelovi ne smiju izdvajati."
Istodobno se ističe: "Zbog izuzetne vrijednosti građe potrebno je što hitnije riješiti status stalnog financiranja ove službe." Državni arhiv u Zagrebu svojim Zapisnikom od 11. lipnja 2001. i 8. svibnja 2002. ističe: "Budući da ova služba

[7] Željko Hodonj je autor prijedloga koncepta organizacije baze podataka i njegovo prihvaćanje od strane Uprave Vjesnik d.d. bio je preduvjet za dolazak (Ž. H.) na mjesto zamjenika glavnog urednika dnevnog lista "Vjesnik" od 1. siječnja 2001. godine. Nakon ostavke na mjesto zamjenika glavnog urednika, Željko Hodonj je Upravi "Vjesnik" d.d. 10. listopada 2002. godine predložio "Projekt pokretanja poduzeća – Dokumentacija Vjesnik; poduzeća za proizvodnju i trgovinu informacijama, dokumentaciju i analizu, stvaranje baza podataka za javnu i posebnu upotrebu". Projekt je prihvaćen u travnju 2003., 1. svibnja 2003. Projekt je formaliziran, a nakon parlamentarnih izbora 2003. godine Uprava Vjesnik d.d. donijela odluku da se nastavi s razvojem Projekta.

vrši vrlo značajan posao, nužno bi bilo da se organizacija i tehnološka podrška modernizira."
"Odgovarajući na zahtjeve Državnog arhiva, Vjesnikova novinska dokumentacija/VND provela je naloge iz Rješenja i Zapisnika. Sadašnji ustroj i organizacija obrade građe nastali su tijekom 2004. i 2005. godine. (...) Baza je kreirana za pohranu, organizaciju i pretraživanje 360.000 članaka godišnje. Sustav je ustrojen kao pouzdana informacijska paradigma za pretraživanje i razvrstavanje dokumenata na zahtjev korisnika dokumentacijske građe. (...) U 2006. Uprava Društva ukinula je sva investiranja u daljnji razvoj te djelatnosti i trenutačno provodi postupak izrade popisa građe prema zahtjevu Vlade Republike Hrvatske i Hrvatskog državnog arhiva. Treba napomenuti da spomenuta građa postoji u Nacionalnoj i sveučilišnoj biblioteci, ali samo u kompletima novina, a ne razvrstana prema interesnim područjima. Nitko u Hrvatskoj trenutačno ne razvija takvu interaktivnu bazu podataka tiskanih glasila."

Početak projekta

Sve je počelo naoko nedužno.
Početkom 2001. godine, tijekom 2002. i 2003. godine kao i nekoliko desetljeća ranije mala skupina dokumentalista u Vjesnikovoj novinskoj dokumentaciji škarama je izrezivala izabrane novinske članke.
Zatečeno je približno šest tisuća registratora u kojima su dokumenti grupirani u područja.
Početkom 2004. godine odredili smo prvu razinu našega zadatka: stvoriti pretraživu arhivu u elektronskom formatu s pretraživim sadržajem tekstova i atributima, dostupnu u svako vrijeme i na svakom mjestu, koju je moguće kreirati u što kraćem vremenu uz utrošak što manje ljudi, opreme i novaca.
Pri tom je trebalo imati na umu da obrađujemo devet (9) dnevnih novina, petnaest (15) tjednika/revija – novina. Trebalo je dnevno obraditi približno 500 stranica novina, ali i 700 članaka/dokumenata. Pošli smo od uvjerenja kako stvaramo novinski dokumentacijski centar koji organizira poslove prikupljanja, selekcije, obrade i pohrane građe u elektronsku bazu podataka. Bilo je potrebno izabrati postupke za pretraživanje i razvrstavanje dokumenata kao odgovor na posebne korisničke zahtjeve posebnih sadržaja. Dokumenti su grupirani prema područjima znanja.
Zajedničkim dogovorom definirani su osnovni deskriptivni elementi teksta novinskih članaka i standardizirane su ključne riječi koje su bile potrebne da se linearni tekst pohrani kao strukturirani **hipertekst** u elektroničku bazu podataka.
Hipertekst / tekst nema ograničenja linearnosti, nego je tekst koji u sebi sadrži veze na druge tekstove (članke) ili druge dijelove istog teksta. Takvi dijelovi novinskih članaka su čvorovi ili nodes. Čvor obično sadrži jednu ideju, tj. pojam, a može sadržavati i podatak o vrsti informacije (npr. "sukob interesa" je standardizirana ključna riječ i predstavlja jedan čvor, odnosno vrstu informacije u području Unutarnja politika). Skup takvih čvorova čini novinsku elektronsku

bazu podataka, a čvorovi i njihove veze, unutarnje i vanjske, mrežnu strukturu između novinskih članaka. Unutarnje veze članka se nalaze unutar istog teksta, a vanjske veze članka su poveznice na neki drugi članak.
Ključne riječi usuglašavane su radom na posebnim projektima. Svaki dokumentalist je unutar svog područja odredio ključne riječi od relevantnog značenja za zadano područje ovisno o sadržaju novinskog članka. Teme unutar stručnih područja nisu razvrstane hijerarhijski, nego abecednim slijedom zbog bržeg snalaženja i pretrage novinskih članaka koja je potrebna pri korisničkim zahtjevima i tematskim upitima novinara kojima potrebne informacije uvijek moraju biti pripremljene i predane u vrlo kratkom roku.
Strukturiranje tekstova u bazi podataka uspostavljeno je korelacijskim odnosima pomoću tezaurusa, odnosa ekvivalancije, homonima, sinonima, hijerarhijskih i asocijativnih odnosa.
Struktura tezaurusa novinskih članaka se sastoji od odnosa među standardiziranim pojmovima ključnih riječi koji su ujedno pomoćno sredstvo za indeksiranje. Tezaurusom, tj. kontroliranim rječnikom za označavanje teksta, olakšava se pretraživanje relevantnih novinskih članaka. Prednosti tezaurusa u VND on-line bazi su što on obilježava područja znanja koja želimo pretražiti, omogućava generičku povezanost pojmova preko dokumentalističke atribucije i rješava mnoge semantičke probleme pobližim određivanjem specifičnih riječi. Tako je omogućena višekratna upotrebljivost teksta i neovisnost o programskoj podršci, platformi ili aplikaciji te je osigurana je precizna definicija i efikasnija obrada samog teksta.
Indeksiranje novinskih članaka za ulazak u elektronsku bazu podataka radi se prema tematski definiranim kriterijima u VND pregledniku.
Proces indeksiranja novina je određivanje bitnih karakteristika novinskih članaka, te je svaki dokumentalist unutar svog područja samostalno izradio stručne teme koje su relevantne za određeno područje.
Organizacija informacija u novinskoj elektronskoj bazi podataka (opis novinskog članka) podijeljena je na dva velika dijela (pripremni i stručni) i sastoji se od sljedećih elemenata:

1. Pripremna indeksacija VND on-line za unos tekstova u elektroničku bazu podataka: Autor, Datum, Izvor, Naslov, Opis, Stranica
   OCR-irani tekst je programska podrška koja omogućava pretragu teksta po svim riječima ili full text search:
   Optical Character Recognition je programska podrška koja omogućava prepoznavanje teksta na bazi slike dokumenta. Za OCR podsustav se koristi Abby FineReader 7.0 s podrškom za sve jezike. Klijentska aplikacija sadrži odabir slike, OCR postupak nad slikom, korekciju pogrešaka, opis dokumenta meta podatcima, te export dokumenta. Serverski dio je aplikacija izgrađena na temeljima Microsoft SharePoint portal Servera koja koristi WebStore podatkovno spremište za pohranu podataka (u XML obliku) i Microsoft SQL Server 2000 za pohranu meta podataka. Aplikacije za pre-

gled objavljenih podataka / meta podataka je prilagođeni portal koji omogućava svu potrebnu funkcionalnost u odabiru podataka i / ili pretraživanja podataka. Postoje tri tipa pretraživanja: jednostavno (po sadržaju), napredno (po sadržaju i dijelu meta podataka) i kombinirano pretraživanje (po sadržaju i svim meta podacima uz mogućnost izgradnje upita). OCR tekst vrlo često ima mnogo grešaka što će se vidjeti u primjerima iz prakse koji slijede. Ljudski faktor u pripremnoj indeksaciji ima važnu ulogu u programskom ispravljanju tekstualnih grešaka, a dokumentalistička obrada također onemogućava greške unosom stručnih termina u polje ključnih riječi.
Područje znanja: unosi priprema, a obilježeno je unaprijed od strane dokumentalista;
Tip sadržaja: može biti članak ili stranica;
Status obrade: daje podatak da li je članak samo OCR-iran ili je i dokumentalistički obrađen;
Kreiranje članka: pripremna indeksacija unosi i kreira članka u za to zadano polje, upućuje na ime i prezime osobe koja je novinski članak kreirala, OCR-irala, ispravila pogreške i unijela potrebne parametre;
Modificiranje članka: upućuje na ime i prezime dokumentalista koji je članak preuzeo i stručno obradio.

2. Atribucija je VND-ova stručna dokumentalistička obrada i sadrži:
Ključna osoba: u novinskom članku može biti nekoliko ključnih osoba, zato je programski omogućen unos i pretraga više osoba istovremeno;
Ključne riječi: međusobno su povezane unutarnjim i vanjskim čvorovima pomoću tezaurusa;
Medijska analiza: Negativno, Neutralno, Pozitivno
Novinska vrsta: Ankete i istraživanja, Feljton, Intervju, Izjava, Izvještaj, Kolumna, Komentar, Pisma, Portret osobe, Reportaža
Ostale osobe: su osobe koje se nalaze u tekstu i vezane su uz ključne osobe, slučajeve, afere ili događaje te ih je potrebno izdvojiti u posebno polje;
Područje znanja: moguće je istovremeno označiti nekoliko navedenih područja jer pojedini novinski članci u sebi sadrže problematiku iz više područja: Društvo / Zdravstvo / Socijalna skrb, Ekologija, Ekonomija / Gospodarstvo / Privreda, Financije, Komunikacije / Mediji, Kultura i umjetnost, Odgoj i obrazovanje, Promet, Religija, Sport, Sudstvo / Kriminal, Svijet / Međunarodni odnosi / Vanjska politika, Unutarnja politika, Znanost;
Tema: može biti samo jedna, odnosi se na područje znanja u kojem se nalazi te omogućava sužavanje pretrage;
Institucija: podrazumijeva poduzeća, državne ustanove, Europska tijela, političke stranke itd. Za sve institucije programska podrška podržava i pretraživanje preko ključnih riječi, a ne samo preko zadanog polja Institucija. Zemljopisna odrednica: sve zemlje i gradovi.

Naša minimalistička norma morala je zadovoljiti sljedeće relacije i upite:
osoba
osoba tema
osoba tema zemljopisna odrednica
osoba tema zemlj. odr. vremenska odrednica
osoba tema zemlj. odr. vrem. odr. ključna riječ
osoba tema zemlj. odr. vrem. odr. ključna riječ institucija
osoba tema zemlj. odr. vrem. odr. ključna riječ institucija izvor
osoba tema zemlj. odr. vrem. odr. ključna riječ institucija izvor autor
osoba tema zemlj. odr. vrem. odr. ključna riječ institucija izvor autor novinska vrsta
osoba tema zemlj. odr. vrem. odr. ključna riječ institucija izvor autor vrsta, full text search
osoba tema zemlj. odr vrem. odr. ključna riječ institucija izvor autor vrsta, full text search medijska analiza

Prve recenzije: veljača 2005. godine

Istodobno smo u suradnji s našim razvojnim timom[8] riješili potrebne definicije za eksport baza podataka na DVD nosačima. Naše su tadašnje strateške postavke zadatka glasile: na izabrani način prikupljene i organizirane skupine informacija osposobljenom analitičaru omogućavaju pronalaženje neuočljivih sadržaja/obavijesti i trendova. Bili smo dosljedni u tvrdnji da bazu podataka moramo organizirati kao javno dostupno znanje zato što uz primjerenu organizaciju *informacija* postaje i ostaje dugotrajni izvor ne samo novih obavijesti i tema, nego važan doprinos posebnim znanjima. Imali smo na umu i etičnost svog postupka jer smo znali da moramo voditi računa o vlastitim sposobnostima upotrebljavanja *informacij**e*** tog novog/starog izvora moći.
Tada su kreirane naše prve e-zbirke: "Oluja, 5. kolovoza 1995 – vojnoredarstvena akcija" te "RAT U BIH – Sukobi Hrvata i Bošnjaka u Bosni i Hercegovini".
Molbu za recenziju postignutoga poslali četvorici relevantnih stručnjaka: prof. dr. Damiru Borasu, prof. dr. Miroslavu Tuđmanu, prof. dr. Ivi Josipoviću i dr. Inoslavu Beškeru.
Ukratko ćemo citirati njihove navode:

Prof. dr. sc. **Damir Boras** prodekan za znanost i međunarodnu suradnju Filozofskog fakulteta Sveučilišta u Zagrebu:
"Bilo mi je veliko zadovoljstvo pregledati građu Vjesnikove novinske dokumentacije koju ste mi stavili na uvid na dva programska CD-a koja ste pripra-

[8] Za razvoj VND online iznimno je zaslužan Branimir Grilec, tada voditelj razvoja projekata u tvrtki eVision, a sada je vlasnik tvrtke Conscious d.o.o. Bile su korisne iscrpljujuće svađe s tadašnjim šefom IT tima Vjesnik d.d., Robertom Mekinićem.

vili, 'Oluja, 5. kolovoza 1995 – vojnoredarstvena akcija' te 'RAT U BIH – Sukobi Hrvata i Bošnjaka u Bosni i Hercegovini', tekstova objavljeni u hrvatskim tiskanim glasilima.
Prema viđenim podacima, vaša je baza najkvalitetnija, i moglo bi se reći elitna baza dokumentacije koja omogućuje dostup do svih objavljenih podataka u glasilima koje pratite.
Neobično me veseli da se konačno u Hrvatskoj pojavila konzistentno i sveobuhvatno, te na standardan način organizirana građa novinskih tekstova koja predstavlja izvor ne samo za povijest i kulturu nego i za politiku, odnosno izvor podataka iz raznih izvora koji međusobno ukršteni omogućuju dobivanje kritičke i vjerodostojne slike o događajima koje opisuju. Velika je prednost to što se mogu vidjeti razni novinski izvori, razni autori tekstova, pogledi raznih država itd. U svakom slučaju organizacija građe u potpunost, pa čak i više od toga, odgovara međunarodnim standardima novinske i druge dokumentacije.
Posebno moram naglasiti da se kroz Vašu dokumentaciju afirmira i naš studij informacijskih znanosti jer one upravo to predstavljaju: organiziranje znanja. Očito se to vidi i u činjenici da zapošljavate mnogo naših studenata.
Dozvolite mi da Vam predložim suradnju u edukaciji naših studenata, budući da im je prema novom ECTS sustavu potrebna praksa baš u ustanovama ovakvog tipa, a smatram da bi u Vašoj dokumentaciji mogli s lakoćom postići vještine i kompetencije potrebne danas svim informacijskim stručnjacima, i to ne samo dokumentalistima, nego i arhivistima, bibliotekarima i ostalima."

Prof. dr. sc. **Miroslav Tuđman**, Odsjek za informacijske znanosti, Katedra za organizaciju znanja, Filozofski fakultet Sveučilišta u Zagrebu:
"Sustav organizacije i pretraživanja baze dokumenata "Oluja" i "Rat u BiH. Sukobi Hrvata i Bošnjaka u Bosni i Hercegovini' s dokumentalističkog stajališta zaslužuje najbolju ocjenu: prema kriteriju obrade sadržaja dokumenata, prema kriteriju pretraživanja, i prema vremenu pristupa pretraženim dokumentima.
Manji prigovori (naveli smo ih nekoliko) više su 'tehničke' naravi koje se odnose na prikaz sustava, tj. kako koristiti VND on-line baze dokumenata.
Osim što organizacija i pretraživanje ovih baze dokumenata zaslužuju najbolju ocjenu, ovo su, koliko je meni poznato, jedine hemeroteke koje se nude korisnicima na DVD nosačima. Zato je potrebno ocijeniti i procijeniti njihovu vrijednost i sa stajališta sadržaja dokumentima.
Vjesnikova novinska dokumentacija (VND) prva i jedina u Hrvatskoj kreira baze tekstova hrvatskih tiskanih glasila prema temama koje su ključne za interpretaciju suvremene hrvatske povijesti. Izbor tema na dva ponuđena DVD nosača primjer su važnosti organiziranja građe koja jest već sada relevantna u Haagu, ali koja će vrlo brzo biti relevantna i na procesima koje će ICTY preseliti u nadležnost hrvatskom pravosuđu. Ovako organizirane hemeroteke, dostupne na DVD nosačima velikom broju korisnika – od novinara, znanstvenika,

odvjetnika, do javnih i kulturnih djelatnika – na najbolji način pomažu utvrđivanju podataka i dolaženju do istine o tim zbivanjima.
Projekt 'VND on-line. Baze dokumenata' u tehničkom smislu može se širiti u barem tri nove dimenzije: a) digitalizaciji članaka (tj. dostupnosti izvornih dokumenata u digitalnoj formi), b) on-line dostupnoj bazi podataka, c) obradi svih područja, tj. dnevnoj obradi ukupne nacionalne novinske produkcije.
Takav jedan projekt po svom je karakteru nužno nacionalni projekt. Kako po svojim ambicijama, obuhvatu, značenju, potrebama korisnika, ali i po cijeni koštanja."

Prof. dr. sc. **Ivo Josipović,** Pravni fakultet Sveučilišta u Zagrebu:
"Koristim priliku da Vam čestitam na kvalitetnom projektu koji ste pokrenuli. Mislim da je ovakav projekt posebno važan za profesionalce, novinare, državnu upravu, političare, stručnjake, znanstvenike i studente, te medijske kuće, znanstvene i stručne organizacije.
Mislim da je organizacija građe primjerena. Alati za pretragu su dobro postavljeni. Preporučam da se alati za pretragu prije stvaranja baze podataka ustroje konzultacijom stručnjaka i potencijalnih korisnika. Očekujem da će alati biti specifični s obzirom na temu i potencijalne korisnike. Sustav je funkcionalan i jednostavan za rukovanje.
Projekt je vrlo perspektivan i s obzirom na njegove potencijale, mislim da ga treba razvijati kao javni servis. Sastavni dio projekta treba biti povezivanje i s drugim bazama podataka. Uz to, trebalo bi ispitati potrebe Vlade i državne uprave. Daljnji potencijal projekta sadržan je u mogućnosti suradnje s gospodarstvom.
Suvremeno gospodarsko poslovanje traži stalnu dostupnost informacijama radi donošenja odgovarajućih poslovnih odluka. O važnosti projekta za znanstveni i stručni rad gotovo da je suvišno govoriti. Posebno, ako bi se uspostavila suradnja i razmjena podataka s odgovarajućim stranim bazama podataka.
Sumirajući, smatram da je riječ o projektu koji je iznimno vrijedan i koji zavrijeđuje apsolutnu podršku. Smatram da bi u nastavku projekta trebalo inzistirati na njegovom autonomnom ustroju i profesion-alizaciji. Uz to, bilo bi dobro naći strateške partnere u državnoj upravi, znanosti i gospodarstvu."

Dr. sc. **Inoslav Bešker**, sveučilišni predavač i novinar:
"Recenzirane dionice pokazuju razvedenu i za novinara veoma pogodnu organizaciju građe. Svaka pretraga se može organizirati po jednome od 11 kriterija ili po njihovoj kombinaciji, što je u praksi sasvim dovoljno za brzo snalaženje, te omogućava preciznost.
Organizacija po dionicama, kakva je ponuđena na recenziranim nosačima, izrazito je korisna i korisnicima izvan novinarstva, koji se usredotočuju samo na pojedine teme. Smijemo pretpostaviti da može bitno skratiti vrijeme istraživanja i proširiti tematski okvir kako studentima prava ili politologije, tako i djelatni-

cima u pravosuđu i politici – kao što bi druge teme našle druge korisnike na istim razinama bilo obrazovanja, bilo uprave.
Alati ponuđeni u tu svrhu su jednostavni, pa ih može rabiti svaki novinar s nižom srednjom razinom informatičke uvježbanosti. Pri smještanju cjelokupnog projekta na Internet koristan bi bio 'tutorial'.
Funkcionalnost sustava, kakav je demonstriran u recenziranim dionicama zavidno je visoka. Ako se uzmu u obzir ograničenja zadana tipom nosača informacija – ograničenja koja bi bila uklonjena dostupnošću istih informacija sa servera – funkcionalnost se može ocijeniti potpunom, kada je riječ o interesima istraživačkog novinara, koji u veoma kratkom vremenu može dobiti cjelovit uvid, a u razumnu vremenu i detaljne podatke.
Budućnost tog projekta jamačno bi morala biti u njegovu širenju na cjelokupnu bazu podataka i na omogućavanju da ta građa bude dostupna online, 24/24 sata.
Puna djelotvornost te baze mogla bi se postići, i za novinare i za ine korisnike, kada bi se ona mogla relaciono umrežiti s drugim takvim bazama, kao što su Narodne novine, Državni arhiv, te slične ustanove, i kada bi se samo jednom potragom mogle obuhvatiti različite baze (kada bi, na primjer, potraga po ključnim riječima 'Gotovina Haag zločin' ponudila i međunarodne ugovore, i nacionalne zakone, i javnosti dostupne arhivske dokumente npr. iz Ureda Predsjednika).
Dionice projekta 'VND on-line' su uzorak javne baze podataka, informatički dostupne najširem krugu korisnika, ponajprije u javnom informiranju, te u stručnom obrazovanju i javnoj upravi.
Preporučuje se razvijanje projekta u javnu bazu podataka dostupno online, kao javni servis na nacionalnoj razini i od nacionalnog interesa, te njezino umrežavanje s drugim nacionalnim bazama podataka."

Postignuti i novi ciljevi

Od tada iz baze podataka eksportirano je nekoliko desetina posebnih baza podataka. Među njima vrijedi izdvojiti prvi *tailor made product* – za Ured glasnogovornice Ministarstva pravosuđa RH kojim je do kraja 2006. godine provjeravan medijski odjek procesa i rezultata provedbe reforme pravosuđa. Danas ta posebna baza podataka broji više od 60 tisuća dokumenata.
Posebna je cjelina projekt "100 javnih ličnosti hrvatske političke scene" nastao uoči parlamentarnih izbora 2003. godine sa 4.279 intervjua političkih lidera iz svih hrvatskih političkih stranaka objavljenih u razdoblju od 1999. do izbora za Hrvatski sabor 2003. godine.
Uz bok tom projektu je građa na DVD nosaču "Predsjednički izbori 2004. godine". Od tog projekta započela je naša suradnja sa Studijem informacijskih znanosti Filozofskog fakulteta Sveučilišta u Zagrebu. Studenti Odsjeka postali su naši stalni gosti provodeći obveznu studentsku praksu radom na poslovima digitalizacije, unosa teksta u bazu podataka i – pripremajući svoje seminarske i

diplomske radove tako da sve što je napisano bude u službi razvoja baze podataka na različitima razinama zahtjeva. Jedan od njihovih radova je DVD nosač sa sadržajima intervjua hrvatskih ministara kulture. Taj skup dokumenata omogućava kontekstualnu provjeru kulturoloških i političkih, pragmatičnih i strateških razina promišljanja i odlučivanja osoba koje su odlučujuće utjecale na događaje što obilježavaju kulturu novije hrvatske povijesti.
Tijekom ljeta 2007. godine tom se procesu pridružio prijedlog dr. Jasminke Dobše, Fakultet organizacije i informatike, Varaždin, Sveučilište u Zagrebu da se u našoj bazi podataka testiraju metode i programske podrške. To je logičan izbor jer je na našoj bazi podataka moguće stvoriti vrlo kompleksan i relevantan *training set* potrebnih dokumenata. Predložila je suradnju HIB/IDC i Fakulteta organizacije i informatike kao podršku dodiplomskim radovima apsolvenata FOI.
Tako stvaramo HINA informacijsku bazu kao jedan od alata kojim se služi HINA u vlastitom organiziranju i HIB Informacijsko dokumentacijski centar koji povezuje interes korisnika sadržaja što ih HINA generira kao svoj proizvod osiguravajući pri tom dostupnost javnog znanja. Postupcima kojima se služi i poveznicama koje ostvaruje, HIB/IDC suvremenim načinom osigurava zaštitu i javnu upotrebu hrvatske kulturne baštine zabilježene u tiskanim i AV medijima. HIB/IDC svojim sustavom podržava kreiranje Informacijskog centra na Internetu pomoću umreženja s bazama u okruženju.
Na godišnjoj programskoj razini HIB/IDC obradit će milijun dokumenata, a prema posebnom programu obradit će bazu isječaka iz novina. Tijekom 2008. godine s našim suradnicima razvit ćemo SW za automatsku kategorizaciju sustava. Naše su pristupne odrednice općepoznate: utemeljenost na lematizaciji, svrstavanje tekstova u predefinirane kategorije (pronalaženje kategorija prema unaprijed naučenim pravilima, sposobnost sustava da "uči" pravila prema već napravljenoj kategorizaciji, sustav temeljem pravila samostalno određuje koje od predefiniranih kategorija odgovaraju pojedinom novom tekstu, strojno dobiveni rezultati postaju prijedlozi za humanog kategorizatora), pronalaženje imenovanih entiteta (u obradi sadržaja niz je entiteta koji karakteriziraju tekst i nužno ih je prepoznati, izdvojiti za dalju obradu: osobe, institucije, toponimi…). Rezultat će omogućiti približavanje cilju – bazi podataka i alatima, sustavima i sredstvima koji omogućavaju dubinsko pretraživanja podataka od interesa za korisnika tko god on bio i s kojeg god mjesta želi znati.

Zaključak

Tako će HINA informacijska baza podataka / Informacijsko dokumentacijski centar osigurati da u Hrvatskoj djeluje konzistentna i sveobuhvatna te na standardan način organizirana građa javnih glasila što je izvor ne samo za povijest i kulturu, nego i za svakog pojedinog hrvatskog građanina, gospodarstvo i politiku. Stvarajući takav javni servis pokazujemo ne samo da smo zreli za informacijski povezanu i organiziranu Europu, nego i prema nama samima u kulturnoj i

u političkoj sferi te da kao stabilno društvo organizirano djelujemo u području zaštite kulturnog identiteta, organizacije javnog znanja i dostupnosti znanja.

Literatura

Castells, Manuel, Informacijsko doba, Golden Marketing, Zagreb 2000/2002, Internet galaksija, Jesenski i Turk, Zagreb; 2003.

Hodonj, Željko; Prijedlog : Projekt pokretanja poduzeća – Dokumentacija Vjesnik, Vjesnik interno izdanje, 2002.

Hrvatski državni arhiv; Pismo ministru kulture RH mr. Boži Biškupiću, 17. siječnja 2006.

Hrvatski sabor, Državni proračun Republike Hrvatske, 2007.g., http://www.nn.hr/clanci/sluzbeno/2006/3098.htm,

Hrvatski sabor, Odbor za informiranje, informatizaciju i medije; Izvješće o radu Upravnog vijeća HINA-e 2005/2006 g. http://www.sabor.hr/Default.aspx?art=8080

Interpelacija o odgovornosti Vlade Republike Hrvatske za stanje Društva Vjesnik d.d. 2006. http://www.sabor.hr/Default.aspx?art=8503

prof. dr. sc. Boras, Damir; prof. dr. sc. Tuđman, Miroslav; prof. dr. sc. Josipović, Ivo; dr. sc. Bešker, Inoslav ; recenzije *e-zbirki: “Oluja, 5. kolovoza 1995 – vojnoredarstvena akcija” te “RAT U BIH – Sukobi Hrvata i Bošnjaka u Bosni i Hercegovini”,* veljača 2005.

Vjesnik on line i Novinska dokumentacija, Vjesnik, interno izdanje, travanj 2003.

Going Offline: How Online Initiatives Revive Offline Civic Engagement

Domagoj Bebić[1]
University of Zagreb, Faculty of Political Science
Lepušićeva 6, Zagreb, Croatia
domagoj.bebic@fpzg.hr

Marijana Grbeša
University of Zagreb, Faculty of Political Science
Lepušićeva 6, Zagreb, Croatia
grbesa@fpzg.hr

Summary

Departing from the concept of the new identity and acknowledging a growing concern about disengaged citizenry, soundly expressed in the accounts of eroding social capital and the crisis of public communication thesis, this paper will develop two assumptions. First, it will argue that online interactions in virtual communities have the potential to create group identity hence providing a source of content that has the capacity to transform virtual into physical communities. Second, it will assume that these virtually created and physical consumed communities have the capacity to induce public action and positively contribute to civic engagement.
In order to explore aspects of virtual communities as local e-engagement spaces, this paper will present two case studies – the MoveOn and the Meetup Initiatives.

Key words: collective identity, online communication, social capital, crisis of public communication, civic engagement

Introduction

The process of "modernization", characterized by "increasing social complexity" (Swanson and Mancini, 1996: 9) led to a break with the traditional social ties. The political order that was once organized around social institutions – po-

[1] Domagoj Bebić and Marijana Grbeša are research and teaching assistants at the Faculty of Political Sciences in Zagreb. This paper was written within the project *Public, elites, media and communication strategy of Croatia's accession to the EU.*

litical parties, trade unions and church in particular – and rooted in ideological commitments and group loyalties, has now been replaced by a more fragmented and individualistic society. Along with some other developments (trends in economy, emergence of television etc.) this has led to dissolution of traditional communities and consequently to erosion of social capital. Contemporary trends in political communication have contributed to deterioration of public sphere and further alienated citizens from politics. Internet emerges as a potential solution to reinvigorate citizens' engagement which is considered to be a "building block of successful democracy" (Rensohn, 2000: 200).

Building on the assumption that identity is a source of cohesion in any community and that online interactions are often characterized by elaborate identity construction, proponents of e-engagement argue that it is possible for Internet to provide space where new collective identities are built. These new identities have the capacity to affect collective action in both instrumental and symbolic terms.

In the first chapter we explain the notions of individual and collective identity. We argue that collective identities developed online evolve group cohesion which may encourage collective action. In the second chapter we look at the two concepts concerned with perceived decline of civic engagement – erosion of social capital and the crisis of public communication. In the third chapter we argue that Internet may have the potential to encourage public participation thus providing solution to rebuild broken social ties and reconnect citizens with politics. In the fourth chapter we present some of the most salient concerns addressing the idea of Internet as a new space of citizens' engagement. Finally, we briefly present two virtually created communities that managed to induce offline public action and apparently positively contribute to both Putnam's and Habermasian notion of civic engagement.

New collective identities

The discussion about *identity* has two perspectives: the "I" and the "we" perspective. According to Tanno and Gonzales (1998, in Zhong, 2000: 38), "I" is concerned with an identity of an individual while "we" is concerned with collective identity. Researches of individual identity focus on psychological constructs and the influences of social interaction (Erikson 1956; Mead, 1934; Slugoski and Ginsburg 1989, in Zhong, 2000: 38). Although this perspective strongly penetrates Internet studies, the issue of collective identity still seems to be more salient (Zhong, 2000).

Turkle (1996, in Kennedy, 2006: 860) argues that the key features of individual virtual identities are that they are anonymous, fluid and fragmented. Turkle bases her arguments on the findings from a research conducted on a group of students who participated in a MUD (Multi User Domain) games and witnessed fragmentation of their own identities in an anonymous setting: "part of me, a very important part of me, only exists inside PernMUD" (ibid: 862).

Haraway (1998, in Kennedy, 2006: 863) argues that this fragmentation is a necessary precondition to understand the other side: "the split and contradictory self is the one who can interrogate positioning and be accountable, the one who can construct and join rational conversation and fantastic imaginings that change history". However, other authors argue that online identities are often continuous with offline selves, not reconfigured versions of subjectivities in real life (Kennedy, 2006: 863). Kendall (1999, in Kennedy, 2006:863) likewise argues that members of the Internet community "continually work to reincorporate their experience of themselves and of others' selves into integrated, consistent wholes". Consequently, presumption of an offline identity which continues to live online is a necessary precondition to take our discussion one step further: a person moves back and forth from online to real world to pursue his/her interests and causes which continue to exist both in virtual and physical worlds.

Individuals engage in online communication thus creating virtual communities which Rheingold (1994 in Vedel 2006: 229) defines as "the social aggregations that emerge from the Internet when enough people carry on public discussions long enough and with sufficient human feeling to form webs of personal relationships in cyberspace". Members of the same virtual community share the same collective identity which fosters group cohesion. Smith and Kollock (1999, in Chadwick, 2006: 27**)** argue that identity is a source of cohesion in any community. Some researches have demonstrated that online interactions are often characterized by elaborate identity construction: rituals, rule writing, and reinforcement.

According to Mellucci (1989, in Wall, 2007: 261) collective identity is a "shared definition produced by several interacting individuals who are concerned with the orientations of their actions as well as the field of opportunities and constraints in which their actions take place". Collective identity "takes place via three similar dimensions: (1) the production of cognitive definitions that establish movement goals; (2) the establishment of network of relationships (particularly evident in forms of organization, technologies of communication, etc.) among actors to communicate and negotiate; and (3) creating emotional investments in which movement members feel as if they belong to the movement, allowing them to recognize themselves in each other" (ibid: 262)**.**

Although nature, structure and dynamics of collective identities in online communities still seem to be underinvestigated, Rheingold (1993, in Wall 2007:263) contends that new communication technologies "appear to have opened up new spaces for public and private participation as well as broadened public participation in political matters". Denning (2001, in Wall 2007:262) identifies five general models of Internet communication within social movement activism: "(1) collection of information (2) publication of information (3) dialogue (4) coordinating action and (5) lobbying decision makers. So, all these accounts assert that online communication may affect collective action in both instrumental

and symbolic terms, by improving the effectiveness of communication and creating collective identities and solidarity.
In the next chapter we look how these developments may encourage civic engagement and alleviate consequences of the perceived erosion of social capital and the crisis of public communication.

The troubles of disengaged citizenry

Putnam's (2000) highly influential "bowling alone" thesis contends that developed states have witnessed a decline in social capital during the last thirty years. Social capital theory locates the foundations of democracy not primarily in citizens' beliefs nor in their institutions, but in *relationships* of each to the other. Putnam argues that the decline of social capital in America is visible in the decline in membership of social groups and voluntary associations, and in many forms of collective political participation such as attending public town meetings or working for political parties (Putnam, 1995). Such development is to be blamed on several factors: trends in the structure of the US economy, changes in the family, growth of the welfare state and, perhaps decisively, emergence of television which, according to Putnam, alienates people from each other and the political process and diminishes their sense of citizenship (Norris, 1996: 474).
The argument about increasing public alienation from political process has been equally soundly expressed in the accounts of the proponents of the *crisis of public communication thesis* (Blumler and Gurevitch, 1995; Rosen, 1996, 1999). Their arguments, to rephrase Pippa Norris (2000: 2), basically come down to one: common practices in political communications as deployed by the news media and by party campaigns hinder civic engagement, meaning learning about public affairs, trust in government and political activism. Media with its predominant focus on candidates instead on issues, strategic election game, tabloid scandals and down-market sensationalism, trivialize politics and turn political communication into just another branch of show business (Street, 2003: 86). Party spin doctors and campaign strategists, on the other hand, blur real political substance, praise image over issues and "packaged" personalities over programs thereby contributing to growing public distrust and cynicism. Blumler argues (1997: 396) that if we agree that there has been clear deterioration in the capacity of political communication to serve citizens more than politicians and journalists; offer meaningful choices between governing teams and agendas; promote a broad sense of participation in government; satisfy our symbolic commitment to the notion of democracy, than we must also agree that we are facing a crisis.
So underlying both concepts is the assumption that citizens are increasingly reluctant to engage in any form of public action. If we agree that "it is the engagement of citizens that provides the building blocks of successful democracy" (Rensohn, 2000: 200), then the concern so soundly expressed by the authors like Putnam, Blumler or Rosen is hardly surprising.

Citizens reconnect

The growth of the Internet and its rapid expansion led to extensive researches on possible (positive) implications it might have for democracy. The bulk of literature has been addressing interactivity as the key element to change the nature of citizens' participation in politics and public life in general. Proponents of "electronic democracy" (for instance, Coleman, 2004; Street, 2001) believe that Internet has the potential to restore deteriorating *public sphere*[2] providing a forum in which citizens debate issues of public concern, hold those in power accountable and improve the existing form of democracy or – in a more fundamental form – revive the ancient form of a direct democracy. Summarizing arguments in favour of electronic democracy, Street (2001: 217) argues that Internet may offer solutions for problems that have been obstructing political participation – "time, size, knowledge and access". Internet has overcome boundaries of time and space and it is no longer necessary for citizens to be physically present to contribute to discussion. Limited political knowledge of ordinary citizens and unequal distribution of resources, which has been hampering their capacity to enrol in the process of deliberation may no longer be a problem (ibid.: 217).

Curran refers to Negroponte (1996, in Curran, 2000: 137) who thinks of cyberspace as generating a new world order based on international communication and popular empowerment. Keane (2000: 67) suggests that 'Internet stimulates the growth of macro public spheres' since one segment of the world population uses Internet to 'generate controversies' (ibid: 67) about matters of common concern with other members of the virtual community. It is the forum for discussion or interaction between the members of special interest groups, *ad hoc* pressure groups or cyber protesters. Websites provide infrastructure for deliberation, which may end up in real actions. In that, sense, Internet opens solutions for development of the international civil society. All these arguments basically lead to one conclusion: reinvent community in cyberspace and political participation will follow (Chadwick, 2006: 26)

In short, arguments advocating the role of Internet in reconnecting fellow citizens and invigorating civic engagement are two-folded: first, Internet is seen to have the potential to engage people into public discussion about matters of common concern thus bringing politics back to the people and restoring public sphere; second, Internet is believed to have the capacity to restore broken social ties. In Chadwick's words it emerges as "a medication for the perceived ills of modern society: isolation, fragmentation, competitive individualism, the erosion

[2] *Public sphere* is in the work of Jürgen Habermas (1974; 1989) conceived as a neutral social space for critical debate among private persons who gather to discuss matters of common concern in a free, rational and in principle disinterested way. Criticized as a working model, the concept of the public sphere triggered many controversies. However, it remains widely praised as a normative ideal, especially by the advocates of participatory democracy.

of local identities, the decline of traditional religious and family structures, and the downplaying of emotional forms of attachment and communication" (ibid: 26).

The criticism

These enthusiastic conceptions of the Internet face several problems. "Perhaps the biggest one", argues Chadwick (ibid: 26), "is that life online exhibits many of the social pathologies communitarians wish to transcend". It has been argued that the ties that bind members of virtual community are not as strong as the old ties of family, locality, religion, or even political structures like local party and lobby group associations. The Internet, in this view, takes the impersonality of modern society to a new level, substituting a diluted form of community and social capital for the real things (Doheny-Farina, 1996, in Chadwick, 2006: 27). The argument basically comes down to Putnam (1994, 1995): the only functional community is the one based on a face-to-face communication; the more we connect with other people on a face-to-face basis, the more we trust them. Besides, face-to-face interaction usually imposes the well-known demands of basic civility. Chadwick (2006: 27) argues that "removal of such discipline from the online environment makes it much easier to express all manner of other prejudices flourish online, where individuals can hide behind the cloak of anonymity or pseudonym, both widely accepted practises in cyberspace". Face to face connections remain important. Failing to focus on human contacts in favour of technology may mean diverting precious resources from other areas (Carlsson, 1995; Danitz and Strobel, 1999, in Wall 2007:264).

The potential of the Internet to create free public spheres of political deliberation – a kind of a "civic commons" in cyberspace – celebrated by many (for instance, Blumler and Coleman, 2001) has simultaneously been criticized for the poor quality of interaction between individuals, as well as their tendency to produce a plurality of deeply segmented political associations. "Democracy involves deliberation and dialogue in the formation of collective goals, rather than the aggregation of individual preferences" argues Street (2001: 219). And Internet is all about registering *individual* preferences. This concern has been probably best articulated in Berry Wellman's (2003, in Chadwick, 2006:27) conception of the Internet as a "hybrid form of networked individualism".

Besides classical digital-divide objection, the bulk of literature addressing the civic potential of the Internet has been concerned with problem of power distribution. Mc Chesney (1999, in Chadwick, 2006) argues that the patterns of Internet control suggest that the dominance by a handful of companies over much of the world's communication system is merely replicated on the Internet. Communication on the Internet only reinforces set balance of power and reliance on the Internet may mean privileging certain groups to the exclusion of others. Curran (2000: 137) similarly argues that the fastest-growing branch of

the Internet is e-commerce which only reveals that relations of power shape new technologies and not the other way around.
Although scepticism expressed in these accounts may be justified, it doesn't entirely annul the hypothesis that Internet may provide a way around the practical problems posed by modern democracies, at least to a certain degree. In the next chapter we support this assumption by sketching two online initiatives which mobilised "real life" civic action.

See You Offline

Chadwick (2006: 115) differentiates between three types of e-mobilisation: the first type encompasses traditional groups who went online to argument their offline strategies; in the second type e-mobilization takes a transnational form; and in the third type, groups' and movements' online activities sometimes take a form of direct action.
In this chapter we briefly present two online initiatives which both fall into Chadwick's third category. MoveOn and Meetup both started as a form of online activism and soon after transformed into civic mobilisation movements which continued to love in both virtual and physical worlds. For the reasons of space we won't examine specific features of these movements. Instead, we'll treat both examples as illustrations to support the arguments we have laid so far.

Move on

MoveOn.org Civic Action was started by Joan Blades and Wes Boyd, two Silicon Valley entrepreneurs who launched the initiative out of deep frustration with the partisan warfare in Washington D.C. and the endless attention media were devoting to Clinton-Lewinski tittle-tattle which almost led to Clinton's impeachment. In 1998 they first launched an online petition (at a total cost of $89) to "Censure President Clinton and Move On to Pressing Issues Facing the Nation." Within days they had hundreds of thousands of individuals signed up.
In about a month time, the number of volunteers supporting the initiative mounted to 2000 while by the end of the year the number of petitioners reached 4.150,000 (Clausing, 1999 in Chadwick 2006: 122). Later in 1998 the founders launched MoveOn.org Political Action "so that like-minded, concerned citizens could influence the outcome of congressional elections, and in turn, the balance of power in Washington".[3] The movement focused on combining Internet fundraising, online mobilisation and "real life" activities to support Democratic candidates, often with stunning results. For instance, upon the announcement of its support for Kerry, MoveOn.org sent out e-mails to its 2.1 million supporters asking for donations. Simultaneously, MoveOn actions took place in the "real" world. For instance, mass bake sale (known as the "Bake Back the White

[3] http://www.moveon.org/about.html

House") saw half a million Americans raise over $750 000 on a single Saturday in May 2004 (Chadwick 2006:123). Besides canvassing and fund-rising, some of their most prominent "real life" actions include a massive visit to the opening of Michael Moore's Fahrenheit 9/11 (which was preceded by an online pledge launched and circulated by 100 000 members) and series of concerts called "Vote for Change Tour" which brought together some of the world famous rock starts such as Bruce Springsteen or Dixie Chicks.[4]

The "original" MoveOn initiative, MoveOn.org Civic Action, uses the same pattern. For, instance in 2002 and 2003 it launched online campaign to stop the war in Iraq which were later translated into "real life" demonstrations, such as a huge marches in hundreds of cities in February 2003 (Hickey, 2004; Kahn and Kellner 2004, in Chadwick 2006:123).

Thus, all MoveOn.org activities basically seek to combine online mobilisation campaigns with the "real life" action. It is hardly surprising that scholars have been facing difficulties trying to classify it into any of the "traditional" organisation categories. Chadwick (2006:124) for instance calls it a "hybrid institution" because it doesn't follow any known interest group organisation pattern; it is neither a social movement nor a progressive wing of the Democratic Party. Besides, it employs only a small number of people: until 2003 it had only four officially registered employees (Von Drehle, 2003 in Chadwick 2006:123). It is interesting that just until recently, some authors insisted on a difference between interest groups, which deliberately work within established political institutions and social movements which mobilize for collective action remote from policy elites. Yet, the utility of this distinction has declined (Chadwick 2006:115-116) in terms of groups' and movements' goals, constituencies of support, tactics and policy impact.

The founders of the organization call themselves "a service". They argue they provide "a way for busy but concerned citizens to find their political voice in a system dominated by big money and big media."[5] Their main goal is to bring real people back into democratic process, engaging them either into some sort of political or civic action which will eventually affects the decision making process or the course of the country politics.

Meetup

Contrary to MoveOn which has a clear civic and political mission, Meetup's main concern is to revive local community life in the USA. It is a non-partisan, private initiative started in 2002 by Scott Heiferman who read Putnam's "Bowling Alone," and vowed to use technology to revive heydays of American community. The main goal of the Meetup.com is to match people who share

[4] http://www.moveon.org/about.html

[5] http://www.moveon.org/about.html

same "interest or cause, and form lasting, influential, local community groups that regularly meet face-to-face".[6] So the underlying idea of the site is to encourage people to go online in order to *meet* offline. Heiferman says that Meetup is all "about offline"[7].

The founder believes that people spend too much time staring in their television and computer screens so his initiative was set to encourage face-to-face interaction based on shared interests and causes. Heiferman believes that people are actually not interested in creating online contents but in solving real life problems. In this context, Internet provides just a successful vehicle.

Meetings are being organized in bars, parks, bowling places, restaurants, even living rooms. The project has transcendent the borders of America and today is present in 55 countries of the world[8].

Their programmatic *Bill of Rights*[9] stipulates The Right to Meet, The Right to Privacy, The Right to Not Get Annoying Ads, The Right to Meet About Almost Anything, The Right to Choose Where to Meet.

The site is backed by investors such as eBay, Omidyar Network, Draper Fisher Jurvetson, Esther Dyson, Allen & Company, and Senator Bill Bradley.[10]

Conclusion

Departing from the concept of the *new collective identity* and growing concern about disengaged citizenry, soundly expressed in the accounts of social capitalists and the proponents of the *crisis of public communication* thesis, this paper developed two assumptions. First, we argued that online interactions in virtual communities have the potential to create group identity hence providing a source of content that has the capacity to transform *virtual* into *physical* communities. Second, we assumed that these virtually created and physical consumed communities have the capacity to induce public action and positively contribute to civic engagement.

Internet's biggest asset is interactivity which enables it to engage its audiences into *mutual communication,* which is a prerequisite for construction of group identities. This potential is recognized by many as a possible remedy to the perceived ills of contemporary democracies: growing social fragmentation, dissolution of traditional communities and public alienation from politics which all are all mirrored increasingly disengaged citizenry.

[6] http://www.meetup.com/

[7] Newsweek, May 29, 2006

[8] http://www.meetup.com/about/ (20/08/2007)

[9] http://www.meetup.com/about/ (20/08/2007)

[10] http://press.meetup.com/pdfs/onesheet.pdf (18/08/2007)

Despite loudly expressed scepticism, there are more and more initiatives which prove Internet's potential to reengage citizens into public life. To illustrate our case we have briefly presented two initiatives: MoveOn and Meetup. Both initiatives have clearly demonstrated capacity to a) provide a platform to elaborate the cause people can identify with; b) build group cohesion strong enough to encourage c) "real action". For the reason of space, we had to restrict ourselves to basic facts. Yet, a fruitful avenue of future research would be to closely examine the nature and the dynamics of these virtually created and physically consumed communities (and their respective collective identities) and compare them to traditional "real life" groups and communities. This might help isolate their unique propositions which may prove enriching in terms of civic engagement and democratic ideals.

References

Alexander C. and Pal L. (eds) (1998) Digital democracy: Policy and Politics in the Wired World. New York Oxford: Oxford University Press

Blumler, J. i Gurevitch, M. (1995) The Crisis of Public Communication. London and NY: Routledge.

Blumler, J. (1997) 'Origins of the Crisis of Communication for Citizenship', Political Communication 14: 395 – 404.

Carlsson, C. (1995) 'The Shape of Truth to Come: New Media and Knowledge', in
G. Gerbner, H.Mowlana and H. Schiller (eds) *Invisible Crises:What Conglomerate Control*
of Media Means for America and the World, pp. 235–44. Boulder, CO:Westview Press.

Chadwick A. (2006) Internet Politics: States, Citizens, and New Communication Technologies. New York Oxford: Oxford University Press

Clausing, J (1999) 'Anti-impeachment Website Tallies Millions in Pledges' New York Times January 8 at http://partners.nytimes.com/library/tech/99/01/cyber/articles/08move.html (accessed August 27, 2007)

Coleman, S. (2004) Re-inventing a space for public communication In Tambini D. and J. Cowling (eds) Public Service Communications London: IPPR)

Curran, J. (2000) 'Rethinking Media and Democracy' pp. 120-154 in J. Curran and M. Gurevitch (eds) Mass Media and Society. London: Arnold.

Danitz,T. and W.P. Strobel (1999) 'The Internet's Impact on Activism:The Case of
Burma', *Studies in Conflict and Terrorism* 22(3): 257–69.

Denning, D.E. (2001) 'Activism, Hactivism, and Cyberterrorism:The Internet as a Tool for
Influencing Foreign Policy', in J.Arquilla and D. Ronfeldt (eds) *Networks and Netwars*,
pp. 239–88. Santa Monica, CA: RAND National Defense Research Institute.

Doheny-Farina, S. (1996) The Wired Neighborhood. Yeale University Press; New Haven.

Erikson, E. H. (1956) 'The problem of ego identity', Journal of the American Psychoanalytic Association 4: 56-121

Habermas, J. (1974) 'The Public Sphere: An Encyclopedia Article', New German Critique 1(3): 49-55

Habermas, J. (1989) The Structural Transformation of the Public Sphere. Cambridge: Polity.

Haraway, D. (1998) 'The Persistence of Vision', in N. Mirzoeff (ed.) *The Visual Culture*
Reader, pp. 191–8. London: Routledge.

Hickey, R. (2004) 'Onward Deaniacs' at http://primarybuzz.com/index.php?articleID=2873&sectionID=39

Jenkins H., Thornburn D. (eds) (2003) Democracy and New Media Cambridge, Mass: MIT Press.

Kahn, R. and Kellner, D. (2004) 'New Media and Internet Activism: From the Battle of Seattle to Blogging' New Media Society 6 (1) pp 87-95

Keane, J. (2000) 'Structural Transformation of the Public Sphere' pp. 53-74 in M. Scammell and H. Semetko (eds) The Media, Journalism and Democracy. Ashgate.

Kendall, L. (1999) 'Recontextualising "Cyberspace": Methodological Considerations for Online Research', in S. Jones (ed.) *Doing Internet Research: Critical Issues and Methods for Examining the Net*, pp. 57-74. London: Sage.

Kennedy H. (2006.) 'Beyond anonymity, or future directions for Internet identity research' New Media Society 2006:8: p. 859-876

Lister M., Dovey J., Giddings S., Grant I., Kelly K. (2003) New Media: A Critical Introduction. London New York: Routledge.

McChesney, R. W. (1999) Rich Media, Poor Democracy: Communication Politics in Dubious Times. University of Illinois Press: Chicago

Mead G.H. (1934) Mind, self, and society. Chicago: University of Chicago Press

Melucci, A. (1989) Nomads of the Present: Social Movements and Individual Needs in Contemporary Society. Philadelphia, PA:Temple University Press.

Norris, P. (1996) 'Does television erode social capital? A reply to Putnam'. PS: Political Science & Politics, 29: 474-480.

Putnam, R. (1994) Making Democracy Work. Princeton, NJ: Princeton University Press.

Putnam, R. (1995) 'Tuning In, Tuning Out: The Strange Disappearance of Social Capital in America'. PS: Political Science and Politics, 27(4): 664-683

Putnam R. (2000.) Bowling Alone: The collapse and Revival of American Community New York: Simon & Schuster

Renshon, S. A. (2000) "Political Leadership as a Social Capital: Governing in a Divided National Culture", Political Psychology, 21(1): 199-226

Rheingold, H. (1993) The Virtual Community: Homesteading on the Electronic Frontier. Reading, MA: Addison-Wesley.

Rheingold H. (1994) The Virtual Community: Finding Connection in a Computerized World London: Secker Warburg.

Rosen, J. (1996) Disconnect: The Origins of Public Journalism http://www.cpn.org/topics/communication/wichita.html#disconnect

Rosen, J. (1999) 'The Action of the Idea: Public Journalism in a Built Form' u Theodore Glasser (ur.) The Idea of Public Journalism. New York and London: Guildford Press.

Scott J. and Johnson T. (2005) 'Bowling Alone but Online Together: Social Capital in E-Communities', Community Development 38 (1)

Smith, M.A., and Kollock, P. (Eds.) (1999) Communities in Cyberspace: Mapping the Social Structure of the Usenet, Routledge: London

Slugoski, B. R., Ginsburg, G.P. (1989) 'Ego identity and explanatory speech' in W.G. Austin & S. Worchel (Eds) pp 36-55. London: Sage.

Street, J. (2001) Mass Media, Politics and Democracy. Palgrave.

Swanson, D. and Mancini P. (eds) (1996) Politics, Media and Modern Democracy. Praeger Series.

Tanno, D. V.,& Gonzales, A. (1998). 'Sites of identity in communication and culture' pp 3-7. In V. T. Tanno & Gonzales (eds.) Communication and identity across cultures. Thousand Oaks, CA: SAGE

Turkle S. (1996) Life on the Screen: Identity in the Age of the Internet London Weidenfeld and Nicholson

Vedel T. (2006) 'The idea of Electronic Democracy: Origins, Visions and Questions', Parliamentary Affairs, 59 (2): 226-235

Von Drehle, D. (2003) 'From Screen Savers to Progressive Saviour" Washington Post, June 5 at http://www.commondreams.org/headlines03/0605-03.htm (accessed August 12, 2007)

Wall M. (2007) ‘Social movements and email: expressions of online identity in the globalizations protests’, New Media Society 9 (2): 258-277

Webster F. (ed) (2004) Culture and Politics in the Information Age: A new politics? London New York: Routledge

Wellman, B. Quan-Haase, A., Boase J., Chen, W. Hampton, K. de Diaz, I. I. and Miyata, K. (2003) ‘The Social Affordances of the Internet for Networked Individualism’ Journal of Computer-Mediated Communication 8 (3) at http://jcmc.indiana.edu/vol8/issue3/wellman.html (accessed August 25, 2007)

Zhong, M. (2000) “Dialectics of Identity in the Global Society”, pp. 35-47 in Chen G., Straosta W. (eds) Communication and global society. New York: Peter Lang Publishing

Obvezni primjerak mrežnih publikacija: hrvatska legislativa i metoda arhiviranja

Tomica Vrbanc
Nacionalna i sveučilišna knjižnica, Zagreb
Hrvatske bratske zajednice 4, Zagreb, Hrvatska
tvrbanc@nsk.hr

Sažetak

Rad obrađuje područje koje čini temelj aktivnosti na arhiviranju mrežnih publikacija, a to je zakonski okvir i odabir metode arhiviranja. Zakonski propisi i ove temeljne odluke utječu na svaki daljnji korak u arhiviranju, zbog čega je potrebno da se temelje na promišljenim i utemeljenim odlukama. Početak rada na arhiviranju mrežnih publikacija obilježilo je nedostatak definicije mrežne publikacije kao vrste publikacija, odnos prema obveznom primjerku, kao i nedostatak šire rasprave o odabiru metode arhiviranja. Budući da su za odabranu selektivnu metodu arhiviranja vrlo bitni kriteriji odabira građe, u radu će biti analiziran dokument kojim se oni definiraju i predlaže njegova dorada. Također se daje i prilog raspravi o prijedlogu novog Zakona o knjižnicama u ovom segmentu.

Ključne riječi: arhiviranje mrežnih publikacija, obvezni primjerak, metode arhiviranja, kriteriji odabira, DAMP (Digitalni arhiv mrežnih publikacija)

Početak obrade i arhiviranja hrvatskih mrežnih publikacija

Nacionalna i sveučilišna knjižnica (NSK) kao centralna knjižnica u Republici Hrvatskoj ima zadatak prikupljanja, opisivanja, čuvanja i davanja na korištenje svih vrsta publikacija objavljenih na njenom području, ali i onih izvan nje značajnih za nacionalnu kulturu i baštinu. Uz tiskane, audio-vizualne i mjesno dostupne elektroničke publikacije, mrežne publikacije smatraju se također dijelom ove baštine koju je potrebno sačuvati za budućnost i omogućiti njeno korištenje. U Hrvatskoj je, za razliku od nekih iskustava u svijetu[1], značaj mrežnih publikacija prepoznat dosta rano, pa su već u Zakonu o knjižnicama iz 1997. godine "on-line" publikacije uključene u propise o obveznom primjerku. Iako spomenuti zakon nije definirao posebnosti s obzirom na način dostave i pohrane mre-

[1] Prema podacima sa stranice projekta PADI 2004. godine samo je šest zemalja imalo propise o obveznom primjerku koji uključuje i mrežnu građu. Vidi o tome: Legal deposit / PADI. Dostupno na: http://www.nla.gov.au/padi/topics67.html.

žnih publikacija, NSK je prepoznala svoju obvezu da barem počne sa identifikacijom i katalogizacijom hrvatskih mrežnih publikacija, i tako im omogući pristup preko računalnog kataloga. Bibliografska obrada mrežnih publikacija počinje u NSK 1998. godine, i do kraja 2003. godine obrađene su 783 jedinice mrežne građe[2]. Međutim, katalogizirane publikacije nisu bile arhivirane, već im je pristup bio moguć samo na njihovu izvoru, tj. izvornoj adresi nakladnika/proizvođača. Sa uspostavom sustava za arhiviranje počinje se 2003. u suradnji sa Sveučilišnim računskim centrom Sveučilišta u Zagrebu (SRCE). Kao rezultat ove suradnje nastala je u studenom 2004. godine prva inačica sustava Digitalni arhiv mrežnih publikacija (DAMP)[3] i puštena u rad. Potrebno je ovdje naznačiti da je početak rada na uspostavi sustava DAMP bio obilježen prije svega potrebom da se arhiviraju već bibliografski identificirane publikacije, što je djelomice odredilo i metodu arhiviranju hrvatskih mrežnih publikacija, tj. selektivni pristup. Najšire gledano, odabir metode prikupljanja mrežnih publikacija "zavisi od vrste ustanove, raspoloživih sredstava, utvrđenih kriterija odabira građe … a osobito su određeni odgovornostima i pravima koje određenim ustanovama osigurava zakonski okvir"[4]. Zbog toga ćemo ovdje najprije razmotriti kakav je u Hrvatskoj zakonski okvir koji daje osnovu za prikupljanje mrežnih publikacija.

Obvezni primjerak mrežnih publikacija

UNESCO-ve *Smjernice za očuvanje digitalne baštine*[5] ističu kako bi odluke o kriterijima odabira elektroničkih baštinskih materijala trebalo temeljiti na "vrijednosti materijala u podupiranju misije same organizacije koja preuzima odgovornost za očuvanje"[6]. Nacionalna i sveučilišna knjižnica prema Zakonu o

[2] Willer, Mirna; Milinović, Miroslav. Prema trećoj generaciji knjižnično-informacijskih sustava : hibridna knjižnica za hibridne usluge // 8. seminar Arhivi, knjižnice, muzeji : mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova. Zagreb : Hrvatsko knjižničarsko društvo, 2005., str. 47.

[3] Milinović, Miroslav; Topolšćak, Siniša. DAMP II : Digitalni arhiv mrežnih publikacija : nova funkcionalnost, novi planovi // 9. seminar Arhivi, knjižnice, muzeji : mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova. Zagreb : Hrvatsko knjižničarsko društvo, 2006, str. 41.

[4] Klarin, Sofija. Predmet, motivi i metode arhiviranja sadržaja weba // 8. seminar Arhivi, knjižnice, muzeji : mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova. Zagreb : Hrvatsko knjižničarsko društvo, 2005, str. 30.

[5] Guidelines for the preservation of digital heritage / Prepared by the National Library of Australia. United Nations Educational, Scientific and Cultural Organization, 2003., str. 71-72. Dostupno na: http://unesdoc.unesco.org/images/0013/001300/130071e.pdf.

[6] Cit. prema Stančić, Hrvoje. Teorijski model postojanog očuvanja autentičnosti elektroničkih informacijskih objekata : doktorska disertacija. Zagreb, 2005, str. 30.

knjižnicama[7] iz 1997. godine ima odgovornost izgrađivanja i organiziranja hrvatske nacionalne zbirke knjižnične građe i njenog dugoročnog očuvanja, a prema propisima o obveznom primjerku ta građa uključuje i "on-line publikacije". Ove odredbe daju zakonski okvir koji nedvojbeno utvrđuje odgovornost NSK za očuvanje ove vrste građe i misiju kojoj bi očuvanje trebalo služiti. Međutim, već na ovoj razini nedostaje nekoliko bitnih pojašnjenja s obzirom na mrežne publikacije. Prvi se tiče nedostatka definicije i preciznijeg određenja pojma "on-line publikacija", za koji se možemo pitati da li obuhvaća jednostavno sve sadržaje objavljene "on-line", ili samo one sadržaje objavljene na mreži koji su ekvivalentni tradicionalnim publikacijama (tj. mrežne knjige, časopisi…). Ukoliko se, prema prvom tumačenju, objavljenim smatra samo ono što je javno dostupno na mreži,[8] znači li to da se publikacije na mreži kojima je zaštićen pristup (uvjet je besplatna registracija, ili uz plaćanje) ne mogu smatrati obveznim primjerkom za koje NSK ima odgovornost prikupljanja i dugoročnog očuvanja?[9] U vrijeme kad su se knjižnice počele sučeljavati s problemima vezanim uz građu objavljenu na internetu, bilo je i mišljenja "da netipično knjižnični, tj. novi sadržaj weba koje se ne može odrediti kao tradicionalna *publikacija* ne predstavlja vrijednu građu za knjižnice"[10]. Takav je stav bio karakterističan za vrijeme kad je, prema riječima Johna Byruma bila prisutna "određena nemoć knjižničara da u dinamičnom svijetu interneta dostignu obuhvat, kvalitetu i pouzdanost bibliografskog nadzora[11]" koji je bio dosegnut u svijetu tradicionalnog nakladništva. Kako se vidi iz navedenog Byrumovog iskaza, knjižničarsko razmišljanje o tome što je to *publikacija* u mrežnom okruženju upravo je bilo potaknuto pitanjem mogućnostima obuhvata ove vrste građe, koje se naročito zaoštrava s obzirom na ispunjavanje odgovornosti i zadaća nacionalnih knjižnica. Naime, one imaju, u zemljama u kojima je tako zakonski regulirano, odgovornost za očuvanje cjelokupne objavljene građe na različitim medijima. Ovaj zadatak sveobuhvatnosti teško je primjenjiv na mrežne publikacije, koje, osim njihove brojnosti, također karakteriziraju osobine dinamičnosti, nepredvidivosti i nestalnosti.

[7] Zakon o knjižnicama. Dostupno na: http://www.nn.hr/clanci/sluzbeno/1997/1616.htm.

[8] Selection guidelines for archiving and preservation by the National Library of Australia. (cit. prema Buzina, Tanja. Digitalni arhiv mrežnih publikacija : način prikupljanja i kriteriji odabira građe // 8. seminar Arhivi, knjižnice, muzeji : mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova. Zagreb : Hrvatsko knjižničarsko društvo, 2005., str. 56.

[9] Nejasnoće su vezane i uz pojam "objavljivanje" koji u okruženju interneta ima drugačije značenje od onog u tradicionalnom nakladništvu.

[10] Klarin, Sofija, nav. dj., str. 24.

[11] Cit. prema Klarin, Sofija, isto.

Sljedeće potrebno pojašnjenje u odnosu na "on-line publikacije" kao obvezni primjerak vezano je uz pojam "dostava". Zakon u članku 37 utvrđuje obavezu svih pravnih ili fizičkih osoba koje izdaju ili proizvode tiskovine, audiovizualnu građu i elektroničke publikacije (izrijekom nabrojanu u čl. 38) besplatnog dostavljanja NSK devet primjeraka od te građe. I dok se ova odredba može primijeniti na svu građu koja se objavljuje na materijalnim medijima, pitanje je kako se može primijeniti na "on-line publikacije" kao jednu vrstu elektroničkih publikacija. Način dostavljanja ovih publikacija nije u zakonu posebno opisan, što bi bilo potrebno s obzirom da se one ne izdaju i ne raspačavaju u materijalnom obliku.

Zakonom je također definirana obaveza dostave svakog novog izdanja određene publikacije. Ovdje se postavlja pitanje kako se na mrežne publikacije može primijeniti pojam "novo izdanje" kako je definirano u članku 37, stavak 4., prema kojem se "novim izdanjem … smatra izdanje koje se stavlja u promet s promijenjenom naslovnom stranicom, novim omotom, dopunjenim ili promijenjenim tekstom, novim slikama, tablicama i sličnim promjenama." Brojnost mrežnih publikacija čiji se sadržaj nadopunjava često i u nepravilnim vremenskim razdobljima nameće zaključak da bi se svaka nadopuna novim tekstom, slikama ili drugim sadržajima mogla smatrati novim izdanjem koje je potrebno "dostaviti".

Bez ulaženja u daljnje analize, možemo zaključiti da je Zakon o knjižnicama iz 1997. godine uključio "on-line publikacije" u propise o obveznom primjerku ne uzimajući u obzir njihove posebnosti koje ih razlikuju u znatnoj mjeri od svih drugih tradicionalnih oblika publikacija, kao i od elektroničkih publikacija u materijalnom obliku. Izgleda da je namjera zakonodavca bila da ukaže na važnost novog medija za objavljivanje i širenje informacija i drugih sadržaja, ali da u tom trenutku stručna i šira društvena zajednica još nije bila dovoljno zrela da se pozabavi svim aspektima problema koje to uključuje.

Nova rješenja u prijedlogu Zakona o knjižnicama

U vrijeme pripreme ovog rada u tijeku je javna rasprava o prijedlogu novog Zakona o knjižnicama i knjižničnoj djelatnosti.[12] Očito je da je stručna knjižničarska i šira društvena zajednica prepoznala da je vrijeme za ponovno definiranje ove djelatnosti, također i s obzirom na činjenicu da nove tehnologije svakim danom donose nove vrste publikacija, nove načine objavljivanja i distribucije, što utječe također i na knjižničnu djelatnost. Tekst prijedloga Zakona kakav je stavljen na javnu raspravu bit će potrebno još znatno doraditi, a u mom prilogu ovoj raspravi ograničit ću se na dijelove koji nadopunjuju odredbe sadašnjeg zakona koje sam komentirao u prethodnom dijelu rada.

[12] Prijedlog Zakona objavljen je na stranicama Hrvatskog knjižničarskog društva. Dostupno na: http://www.hkdrustvo.hr/datoteke/280.

Već u uvodnom dijelu prijedloga Zakona nastoji se ispraviti manjak definicije i višestruko upitan izbor termina "on-line publikacije" iz još važećeg zakona. Publikacija se ovdje definira kao "djelo koje je tiskano, umnoženo nekom drugom tehnikom ili priopćeno javnosti putem žica ili bez žica te stavljeno na raspolaganje javnosti". U ovoj definiciji kao *terminus generalis* koristi se zastarjeli termin "djelo", što ponovno daje osnova tumačenju prema kojem knjižničarska zajednica ostaje vjerna tradicionalnim tipovima publikacija, a ostale nove sadržaje ne smatra vrijednom građom. Zbog toga bi bilo bolje definirati publikaciju kao "dokument", već široko prihvaćen termin, tim više što se on koristi u istom dijelu prijedloga zakona, u definiciji knjižnične građe. Definiciju mrežne građe donosi članak 40, stavak 1: "Obveznim primjerkom mrežnih publikacija obuhvaćene su publikacije i integrirajuća građa objavljena u domenama Republike Hrvatske ili u drugim domenama namijenjenim hrvatskim korisnicima". Ovdje se ponovno zamagljuje smisao termina "publikacija" time što se navodi kao *jedna od vrsta* mrežnih publikacija, uz integrirajuću građu. Znači li to da se integrirajuća građa ne smatra publikacijama? Ova definicija zahtijeva potpunu preradu, također jer uvodi novi termin (integrirajuću građu) kao *različit* od publikacije, a sam termin se nigdje posebno ne definira.

Pravu novost ovog prijedloga zakona sadrži stavak 2 istog članka: "Tehnički nezaštićene mrežne publikacije Nacionalna i sveučilišna knjižnica prikuplja pobiranjem, obrađuje i dugoročno pohranjuje sukladno Pravilniku o obveznom primjerku i Kriterijima odabira obveznog primjerka mrežne građe." Ovom rečenicom, kao i rasporedom odredbi o obveznom primjerku koje su podijeljene na propise o publikacijama u materijalnom obliku (članak 39) i odredbe koje se odnose na "prikupljanje obveznog primjerka mrežnih publikacija", konačno se priznaje razlika u načinu prikupljanja ovih dvaju vrsta publikacija. Naime, u slučaju mrežnih publikacija ne može se raditi o "dostavi" (iako je i taj način moguć) i ozakonjuje se posao koji se zadnje tri godine radi u Nacionalnoj i sveučilišnoj knjižnici, a to je "prikupljanje pobiranjem" i obrada mrežnih publikacija. Isti stavak upućuje na Pravilnik o obveznom primjerku, koji još nije donesen, a u kojem bi trebalo biti detaljnije definirano pobiranje, obrada i dugoročna pohrana. Druga značajna novost koju donosi ovaj stavak je da se prikupljanje odvija sukladno Kriterijima odabira obveznog primjerka mrežne građe. Spominjanjem bilo kakvih kriterija odabira može se posredno zaključiti kako se "prikupljanje pobiranjem" ne vrši za sve one publikacije koje su definirane kao obvezni primjerak mrežnih publikacija, nego se između njih vrši odabir, dakle selekcija. Ovako značajnu odluku Zakon bi morao posebice istaknuti i barem ukratko obrazložiti razloge takvog rješenja. Ovaj dokument o kriterijima odabira, za razliku od Pravilnika o obveznom primjerku, već je donesen i objavljen[13], i on će biti predmet analize u nastavku rada.

[13] Kriteriji odabira obveznog primjerka mrežne građe za obradu i arhiviranje. Dostupno na: http://www.nsk.hr/DigitalLib.aspx?id=83.

S obzirom na to da ovaj stavak govori o pobiranju "tehnički nezaštićenih mrežnih publikacija", ostaje otvorenim pitanje zaštićenih publikacija , ukoliko se iz sljedećeg stavka o tome da NSK "ima pravo prikupljanja ... i korištenja mrežnih publikacija" ne bi moglo zaključiti da pokriva i te publikacije.
Značajni nedostatak prijedloga zakona u ovom dijelu čini pitanje koja je obveza nakladnika s obzirom na prikupljanje mrežnih publikacija? Budući da je obveza nakladnika publikacija u materijalnom obliku da dostavi NSK određen broj primjeraka publikacija, ne bi li, sukladno tome, bilo potrebno definirati obvezu nakladnika mrežnih publikacija da obavijesti NSK o objavljivanju svoje mrežne publikacije? Tim više što je ovaj način već uveden u okviru (zasad zakonski nereguriranom) suradnje s nakladnicima u okviru izgradnje Digitalnog arhiva mrežnih publikacija[14]. Ovaj način obavještavanja o objavljivanju mrežnih publikacija ne bi bio u sukobu sa (vjerojatnom) odlukom o selektivnom pobiranju, jer uvjet za dobru selekciju bila bi što potpunija informacija o objavljenim publikacijama.
I, na kraju, na pitanje što predstavlja novo izdanje elektroničkih publikacija pokušalo se odgovoriti u članku 39, stavak 13, koji svojim smještajem u odjeljak o dostavi publikacija u materijalnom obliku upućuje da se odnosi samo na tu vrstu elektroničkih publikacija. Ako se to i zanemari, sama definicija da je "novo izdanje elektroničke publikacije određeno promjenom formata računalne datoteke i/ili izmjenama sadržaja pod kojima se ne podrazumijeva novi oblik ambalaže i korektura teksta" prilično je nejasna i teško primjenjiva na mrežne publikacije. Možemo zaključiti kako ovaj prijedlog zakona ne daje zadovoljavajuća rješenja za većinu otvorenih problema, iako pokušava definirati ključne termine i daje osnovne naznake o načinu prikupljanja mrežnih publikacija.

Metode prikupljanja i arhiviranja mrežnih publikacija

Kako je spomenuto u uvodu ovog rada, u počecima rada na izgrađivanju sustava DAMP odlučeno je da će se primijeniti selektivna metoda u prikupljanju i arhiviranju hrvatskih mrežnih publikacija. Da bi se razumjelo razloge ovakve odluke, potrebno je prikazati koje su osnovne metode koje se koriste za prikupljanje mrežnih stranica, kao i njihove prednosti i nedostatke.
Sveobuhvatno ili neselektivno arhiviranje je metoda širokog sveobuhvatnog snimanja *(harvesting, crawling)* cjelovitog sadržaja određene domene ili više njih koje se provodi u određenim vremenskim razmacima. Ovom metodom dobivaju se sveobuhvatni uzorci (*snapshot*) sadržaja weba[15]. Prednost je ove metode što osigurava širi kontekst prikupljenih dokumenata, a to je vrlo važno s

[14] Digitalni arhiv hrvatskih mrežnih publikacija. Dostupno na: http://www.nsk.hr/DigitalLib.aspx?id=80

[15] O metodama arhiviranja vidi Klarin, Sofija. Predmet, motivi i metode arhiviranja sadržaja weba // 8. seminar Arhivi, knjižnice, muzeji : mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova. Zagreb : Hrvatsko knjižničarsko društvo, 2005., str. 30-32.

obzirom da je kontekst, naročito u okruženju weba, ključan za "interpretaciju značenja informacije".[16] Druga prednost ove metode je ekonomska isplativost, jer vremenski kraće traje i ne uključuje veći i intenzivniji udio ljudskog rada. Nedostatak metode je nemogućnost kvalitetne bibliografske obrade, što se može nadomjestiti primjenom programa za indeksiranje sadržaja i njihovo pretraživanje.

Metodom selektivnog arhiviranja pobiru se samo određene publikacije odabrane prema unaprijed utvrđenim kriterijima, u prethodno utvrđenim intervalima pobiranjima za svaku jedinicu. Ključnu važnost u primjeni ove metode ima određivanje jasnih i konzistentnih načela za utvrđivanje kriterija selekcije. Ukoliko su kriteriji postavljeni na ovaj način, prednost je primjene ove metode mogućnost kvalitetne bibliografske obrade, svaka se arhivska kopija može pojedinačno pregledavati, što je pogodno i za unapređenje sustava arhiviranja i postavljanja zahtjeva za dugoročnom zaštitom.

Nedostatak ove metode naličje je prve navedene prednosti, a odnosi se na teškoće utvrđivanja i primjene kriterija selekcije. Više autora vidi glavnu poteškoću u pitanju da li će onu građu koju danas smatramo vrijednom korisnici u budućnosti također smatrati takvom[17]. Postoji čak mišljenje da će istraživačima budućnosti neformalni sadržaji kao forumi i blogovi biti važniji nego sva građa koja danas ima prioritet u projektima arhiviranja (znanstvene publikacije, službene stranice institucija…). Drugi nedostatak ove metode je upravo nedostatak konteksta čija je važnost već spomenuta, a treći je vezan uz zahtjevan i dugotrajan ljudski rad, što je vezano također uz veće troškove.

Treća metoda koja se koristi u manjoj mjeri od prethodne dvije je **dostava** mrežne građe, a može služiti i kao dopunska metoda već opisanima. Ovom metodom građu izravno dostavlja nakladnik na nekom fizičkom mediju, ili drugim načinom komunikacije (FTP protokol), u skladu s pojedinačnim ugovorima knjižnice i nakladnika. Koristi se posebno za slučajeve kad knjižnica neku građu iz određenih razloga ne može arhivirati (npr. publikacije sa zaštićenim pristupom).

Kriteriji odabira obveznog primjerka mrežne građe

U ovom kontekstu lakše se može razumjeti pristup i način rješavanja pitanja vezanih uz obradu i prikupljanje mrežnih publikacija u NSK. Osim nedefiniranosti postupanja s obveznim primjerkom mrežnih publikacija u Zakonu o knjižnicama iz 1997., treba imati na umu kako na međunarodnom planu u to doba nisu još bile donesene nikakve smjernice niti standardi s obzirom na arhiviranje

[16] Ross, Seamus. Cyberculture, cultural asset management, and ethnohistory – preserving the process and understanding the past, str. 8. Dostupno na: http://www.deflink.dk/upload/doc_filer/doc_alle/740_sross_cyberculture_rev2.doc#_Toc526580743. Naročito se ističe mišljenje da je "sačuvati informacijske objekte bez konteksta besmisleno".

[17] Stančić, Hrvoje, nav. dj., str. 130 i Ross, Seamus, nav. dj., str. 15.

elektroničke građe. Temeljni standard za elektroničke arhive, Referentni model za otvoreni arhivski informacijski sustav (OAIS RM)[18] donesen je 1999. godine, a prvi model koji ga primjenjuje za jedan knjižnični elektronički arhiv razvijen je u okviru projekta NEDLIB 2000. godine[19]. Nisu još bile donesene niti smjernice sa općenitim pristupom problemu očuvanja svjetske digitalne baštine i kriterijima odabira građe za očuvanje, što je učinjeno UNESCO-vim *Smjernicama za očuvanje digitalne baštine* iz 2003. Zbog toga je razumljivo zašto se kod nas nije počelo izgradnjom teoretskog okvira za aktivnosti vezane uz očuvanje mrežnih publikacija, a također i zašto od početka nisu definirani kriteriji odabira građe koja je bila samo identificirana i bibliografski obrađivana. Izradi ovakvog dokumenta pristupilo se kasnije, tijekom uspostave sustava za preuzimanje i arhiviranje obveznog primjerka hrvatskih mrežnih publikacija, i objavljen je na mrežnim stranicama Digitalnog arhiva pod nazivom *Kriteriji odabira obveznog primjerka mrežne građe za obradu i arhiviranje*. Analiza kriterija kako su definirani u tom dokumentu bit će predmet nastavka ovog rada.

Temeljno načelo od kojeg treba početi razmatranje kriterija odabira je "vrijednost materijala u podupiranju misije same organizacije" iz već citirane rečenice UNESCO-vih *Smjernica*. Uvodna rečenica kojom počinje tekst dokumenta o kriterijima odabira može se smatrati upravo iskazom koji govori o misiji organizacije: "Nacionalna i sveučilišna knjižnica ima zadatak sustavno prikupljati, opisivati, pohranjivati i davati na korištenje i mrežne publikacije kao sastavni dio kulturne baštine".

Kriteriji se u ovom dokumentu dijele na opće i posebne. Opći kriteriji su "isti opći kriteriji kao i za tiskanu građu", naime kriteriji koji se primjenjuju kod odabira građe koja se objavljuje u tekućoj nacionalnoj bibliografiji knjiga[20]. Sva četiri navedena kriterija (djela hrvatskih autora, tema djela odnosi se na Hrvatsku, djela na hrvatskom jeziku i djela objavljena u Hrvatskoj) možemo gledati kao najopćenitije, najšire određenje građe koja pripada hrvatskoj kulturnoj baštini, za čije očuvanje je NSK odgovorna ustanova. Ovi kriteriji sami po sebi nisu dovoljni, već predstavljaju samo prvi stupanj selekcije. Ovaj "dvostupanjski" način selekcije razlikuje se od postupka s tiskanom građom, budući da se za nju postavlja cilj što potpunijeg obuhvata nacionalne građe, pri čemu je dovoljno primijeniti kriterij pripadnosti nacionalnom korpusu[21].

[18] Rerefence Model for an Open Archival Information System (OAIS). Consultative Committee for Space Data Systems, NASA, Washington, D.C., 2002. Dostupno na: http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf.

[19] Werf, Titia van der. The deposit system for electronic publications : a process model. Dostupno na: http://www.kb.nl/coop/nedlib/results/DSEPprocessmodel.pdf.

[20] Vidi predgovor u: Hrvatska bibliografija. Niz A, knjige. Dostupno na: http://www.nsk.hr/DigitalLib.aspx?id=278.

[21] Osim u slučajevima odabira tzv. "sitnog tiska" i ostalih rubnih primjeraka publikacija, u slučaju kojih se svejedno radi odabir.

Prvi od šest definiranih posebnih kriterija odnosi se na sadržaj publikacije, prema kojem se prednost pri odabiru daje građi "čiji je sadržaj koherentan, neovisan, potpun, ima trajnu intelektualnu ili umjetničku vrijednost i bavi se temama od općeg kulturološkog značenja", a u nastavku se nabrajaju teme kao umjetnost i kultura, obrazovanje, zdravstvo itd. U definiranju ovog kriterija koji se u kontekstu dokumenta može iščitati kao najvažniji, uvodi se pojam (intelektualne i umjetničke) vrijednosti. Budući da je ovo ključni pojam koji igra važnu ulogu u postupcima selekcije za bilo koju knjižničnu zbirku, valja se na njemu zadržati. UNESCO-ve *Smjernice* predlažu nekoliko pitanja koja nam mogu pomoći kod određivanja vrijednosti materijala koji namjeravamo očuvati. Potrebno se zapitati s čim je povezana vrijednost materijala: dokazom, informacijom, umjetničkim ili estetskim kriterijima, značajnom inovacijom, povijesnim ili kulturnim faktorima, time što omogućava korisniku na nešto učini s materijalom ili karakteristikama značajnim za kulturu?[22]. Iz citirane rečenice dokumenta možemo zaključiti da se smatralo kako je vrijednost materijala povezana sa informacijom, umjetničkim i estetskim kriterijima, kao i povijesnim i kulturnim faktorom. Međutim, u primjeni ovog kriterija osnovna teškoća nalazi se u već navedenom pitanju da li će onu građu koju danas smatramo vrijednom takvom smatrati i korisnici budućnosti. Osim toga, sigurno su za očuvanje mrežnih publikacija kao dijela kulturne baštine bitne i ostale ponuđene vrijednosti u UNESCO-vim *Smjernicama*. Spomenut ću samo vrijednost materijala kao inovacija (inovativno korištenje novih tehnologija u izradi) i vrijednost koju daje "mogućnost da korisnik učini nešto" (što se ogleda u funkcionalnosti koju pružaju neke mrežne publikacije).

Drugi kriterij NSK za odabir obveznog primjerka mrežne građe vezan je uz prethodni, a to je "ugled i pouzdanost nakladnika/autora". Budući da je vrijednost i korist za buduće korisnike teško odrediti, mogu se koristiti provjerene metode u selekciji i izgradnji knjižničnih zbirki tradicionalne građe, a jedna od njih je vrednovanje nakladnika. U tradicionalnom nakladništvu bitna je uloga nakladnika, od biranja rukopisa koje će objaviti, do različitih postupaka uređivanja prihvaćenog teksta i instituta recenzije[23]. Nakladnik svim navedenim postupcima jamči kvalitetu svojih izdanja. Možemo također reći da nakladnik već vrši određen postupak selekcije između rukopisa koji mu se nude za objavljivanje, i u tom kontekstu proces selekcije za knjižnične zbirke dio je "lanca aktivnosti" koji počinje još od nakladnika[24]. Zbog toga se može reći da ugled nakladnika i dalje zadržava svoje značenje u mrežnom okruženju.

[22] Guidelines for the preservation of digital heritage, str. 75 (cit. prema Stančić, Hrvoje, nav. dj., str. 131).

[23] Vidi Živković, Daniela. Elektronička knjiga. Zagreb : Multigraf, 2001., str. 109: "U objavljivanju elektroničke knjige ostala je sačuvana ključna i tradicionalna uloga nakladnika kao što su izbor sadržaja i uređivanje rukopisa."

[24] Vidi Ross, Seamus, nav. dj., str. 11.

Budući da mnogi autori sami izdaju svoje publikacije na internetu, *Kriteriji odabira* navode dodatno obrazloženje kako je "kod osoba pri odabiru važan i njihov stručni i društveni autoritet". Primjena ovako formuliranog kriterija za osobe je upitna, naročito u smislu vidljivosti procesa selekcije iz UNESCO-vih *Smjernica*[25], koja rezultira jasno i eksplicitno izraženim stajalištem zašto je jedna publikacija prošla selekciju, a druga nije.

Za razliku od prva dva kriterija koji se tiču kvalitete sadržaja publikacije, sljedeća dva mogli bismo nazvati formalnim kriterijima. Kriterij strukture publikacije odnosi se na postojanje podatka o naslovu, podataka o odgovornosti za sadržaj i izradu publikacije, preglednost rasporeda podataka na publikaciji, redovitosti osuvremenjivanja, te broju i pouzdanosti poveznica. Iz ovog kriterija proizlazi da prednost kod odabira ima građa koja je oblikovana na način da su prisutni svi navedeni podaci, što je prednost za lakšu i kvalitetniju bibliografsku obradu. Ali, budući da postojanje ovih podataka ne predstavlja najbitniji čimbenik koji utječe na vrijednost publikacije, iz odabira se ne bi smjelo isključiti publikacije samo zbog toga što ne udovoljavaju ovom uvjetu. Kriterij domene publikacije po kojem prednost kod odabira imaju ima građa "koja je objavljena izvorno u domeni .hr" također je opravdan, budući da opći kriteriji navode uvjet da su "djela … objavljena u Hrvatskoj" . Dodaje se kako može biti odabrana i građa koja je objavljena u kojoj drugoj domeni (.com, .net, .org), "ukoliko udovoljava općim kriterijima za odabir građe". Publikacije objavljene na drugim domenama ne bi trebale biti isključene iz odabira, pod uvjetom da udovoljavaju posebnim kriterijima odabira, od kojih je najvažniji intelektualna i umjetnička vrijednost publikacije.

Zadnji kriterij odabira odnosi se na arhiviranje također i onih publikacija koje su objavljene i na nekom materijalnom mediju (knjiga, CD-ROM), što je opravdana odluka, iz dva razloga. Prvi je da je mrežno izdanje ima drugu funkcionalnu vrijednost od tiskanog izdanja, a drugi je u tome što se odabirom i mrežnog izdanja za arhiviranje omogućava šira dostupnost publikacije korisnicima.

Ako, kao zaključak, usporedimo analizirani dokument sa preporukama i načelima iznesenim u UNESCO-vim *Smjernicama*, možemo reći da se kriteriji odabira u ovom dokumentu temelje na izraženoj misiji organizacije koja ima odgovornost za očuvanje hrvatske kulturne baštine. U sastavljanju kriterija pokušalo se odgovoriti na najsloženije pitanje vrijednosti građe koja se želi sačuvati, ali nije dovoljno naglašena važnost ovog kriterija, kao i načini utvrđivanja vrijednosti građe. Uzete su u obzir i neke formalne i tehničke specifičnosti koje utječu na odabir građe, ali bi neke odredbe trebalo preispitati i preraditi.

[25] Guidelines for the preservation of digital heritage, str. 71.

Zaključak

Možemo se zapitati na kojoj se razini organizacije i provedbe postupka prikupljanja i pohrane mrežnih publikacija nalazi Nacionalna i sveučilišna knjižnica kao institucija koja ima odgovornost i provodi ove aktivnosti?. Prema kategorizaciji Anne R. Kenney[26] vidljivo je da je ova ustanova prešla razinu uočavanja problema i početka djelovanja, tj. inicirala je i provodi projekt na očuvanju elektroničkog gradiva u okviru izgradnje Digitalnog arhiva mrežnih publikacija. Uočavaju se već određeni elementi razine konsolidacije, koju karakterizira prelazak s pojedinih projekata na programe, uviđanje da postoje standardi za određene postupke koji su razvijani izvan institucije i uvažavanje pristupa i rješenja drugih sličnih institucija. U arhiviranju hrvatskih mrežnih publikacija ovi pomaci se uočavaju u približavanju sličnim projektima drugih institucija, kao što je to projekt Hrvatske informacijsko dokumentacijske referalne agencije DAMIR (Digitalni arhiv mrežnih izvora republike Hrvatske). Tek nam predstoji faza institucionalizacije koju odlikuje "razvijena jasna okosnica propisanih politika i propisa". Ovim radom nastojao sam pokazati kako ova okosnica još nije stvorena, ali i nužnost njenog izgrađivanja. Postizanje ovako značajnog cilja iziskuje nemalu volju i napor šire stručne i društvene zajednice, pod uvjetom da je uočena važnost ove djelatnosti za očuvanje jednog dijela nacionalne kulturne baštine.

Literatura

Buzina, Tanja. Digitalni arhiv mrežnih publikacija : način prikupljanja i kriteriji odabira građe // 8. seminar Arhivi, knjižnice, muzeji : mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova. Zagreb : Hrvatsko knjižničarsko društvo, 2005.

Digitalni arhiv hrvatskih mrežnih publikacija. Dostupno na: http://www.nsk.hr/DigitalLib.aspx?id=80

Guidelines for the preservation of digital heritage / Prepared by the National Library of Australia. United Nations Educational, Scientific and Cultural Organization, 2003. Dostupno na: http://unesdoc.unesco.org/images/0013/001300/130071e.pdf

Hrvatska bibliografija. Niz A, knjige. Dostupno na: http://www.nsk.hr/DigitalLib.aspx?id=83

Klarin, Sofija. Predmet, motivi i metode arhiviranja sadržaja weba // 8. seminar Arhivi, knjižnice, muzeji : mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova. Zagreb : Hrvatsko knjižničarsko društvo, 2005.

Koerbin, Paul. Report on the crawl and harvest of the whole australian web domain undertaken during June and July 2005. Dostupno na: http://pandora.nla.gov.au/documents/domain_harvest_report_public.pdf

Kriteriji odabira obveznog primjerka mrežne građe za obradu i arhiviranje. Dostupno na: http://www.nsk.hr/DigitalLib.aspx?id=83

Legal deposit / PADI. Dostupno na: http://www.nla.gov.au/padi/topics67.html

Milinović, Miroslav; Topolšćak, Siniša. DAMP II : Digitalni arhiv mrežnih publikacija : nova funkcionalnost, novi planovi // 9. seminar Arhivi, knjižnice, muzeji : mogućnosti suradnje u

[26] Vidi Stančić, Hrvoje, nav. dj., str. 123-129; prema Kenney, Anne R., Digital Preservation in Digital Libraries: Issues, Obstacles, and Possibilities, slajdovi s predavanja na DELOS International Summer School 2005 – Digital Preservation for Digital Libraries, INRIA, Sophia Antipolis, Francuska.

okruženju globalne informacijske infrastrukture : zbornik radova. Zagreb : Hrvatsko knjižničarsko društvo, 2006.

Pigac, Sonja. Digitalni arhiv mrežnih publikacija : rad u sustavu za preuzimanje i arhiviranje obveznog primjerka hrvatskih mrežnih publikacija // // 8. seminar Arhivi, knjižnice, muzeji : mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova. Zagreb : Hrvatsko knjižničarsko društvo, 2005.

Prijedlog zakona o knjižnicama i knjižničnoj djelatnosti. Dostupno na: http://www.hkdrustvo.hr/datoteke/280

Rerefence Model for an Open Archival Information System (OAIS). Consultative Committee for Space Data Systems, NASA, Washington, D.C., 2002. Dostupno na: http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf

Ross, Seamus. Cyberculture, cultural asset management, and ethnohistory – preserving the process and understanding the past. Dostupno na: http://www.deflink.dk/upload/doc_filer/doc_alle/740_sross_cyberculture_rev2.doc#_Toc526580743

Stančić, Hrvoje. Teorijski model postojanog očuvanja autentičnosti elektroničkih informacijskih objekata : doktorska disertacija. Zagreb, 2005.

Werf, Titia van der. The deposit system for electronic publications : a process model. Dostupno na: http://www.kb.nl/coop/nedlib/results/DSEPprocessmodel.pdf

Willer, Mirna; Milinović, Miroslav. Prema trećoj generaciji knjižnično-informacijskih sustava : hibridna knjižnica za hibridne usluge // 8. seminar Arhivi, knjižnice, muzeji : mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova. Zagreb : Hrvatsko knjižničarsko društvo, 2005.

Zakon o knjižnicama. Dostupno na: http://www.nn.hr/clanci/sluzbeno/1997/1616.htm

Živković, Daniela. Elektronička knjiga. Zagreb : Multigraf, 2001.

DAMIR

Ana Garvas Delić
Hrvatska informacijsko-dokumentacijska referalna agencija – HIDRA
Trg maršala Tita 3, Zagreb, Hrvatska
Ana.Garvas@hidra.hr

Tamara Horvat
Hrvatska informacijsko-dokumentacijska referalna agencija – HIDRA
Trg maršala Tita 3, Zagreb, Hrvatska
Tamara.Horvat@hidra.hr

Sažetak

Tijela javne vlasti Republike Hrvatske objavljuju službenu dokumentaciju, publikacije i informacije na preko 900 web sjedišta. Teško snalaženje u velikoj i rasutoj količini sadržaja, često samo privremeno dostupnog na Internetu, pokušava se olakšati automatskim prikupljanjem, obradom, pohranom, čuvanjem i organizacijom pristupa u Digitalnom Arhivu Mrežnih IzvoRa Republike Hrvatske – DAMIR.
U radu se prikazuje razvoj sustava za pobiranje i potreba njegova stalnog dopunjavanja i razvoja, postupci koji se primjenjuju u radu s prikupljenim sadržajima, važnost kriterija za odabir dokumenata i nužnost njihove sustavne dopune te razvoj pristupa sadržaju DAMIR-a.
Analizira se sadržaj arhiva prema oblicima dokumenata, prema količini objavljenih dokumenata pojedinih tijela javne vlasti, te odnos ukupno pobranih i prihvaćenih dokumenata. Odvojeno se razmatra profil pobiranja webova kao cjelina.
Pobiranje, u uvjetima nepoštivanja preporuka za korištenje identifikacijskih podataka pri izradi dokumenata te nepoštivanja tehničkih standarda pri izradi mrežnih izvora daje rezultate koji samo djelomično ostvaruju postavljene ciljeve i upućuje na potrebu primjene standarda i razvoja organiziranog informacijskog prostora tijela javne vlasti u Republici Hrvatskoj.

Ključne riječi: tijela javne vlasti, arhiv mrežnih dokumenata, pobirač (robot), kriterij odabira dokumenata

Uvod

HIDRA sustavno i redovito prikuplja javne službene dokumente i publikacije sa službenih internetskih stranica tijela javne vlasti i političkih stranaka Republike Hrvatske od 2004. godine. U suradnji sa Sveučilišnim računskim centrom Sve-

učilišta u Zagrebu – ***Srcem*** razvijen je sustav za arhiviranje mrežnih izvora i periodički se sadržaji pohranjuju u **D**igitalni **A**rhiv **M**režnih **I**zvo**R**a Republike Hrvatske – **DAMIR**.
Automatizacijom procesa pobiranja sadržaja s odabranih mrežnih izvora postignuta je nova kvaliteta u prikupljanju službene dokumentacije RH. Postavljeni su osnovni kriteriji odabira sadržaja primjereno novim tehnologijama rada, multipliciranom objavljivanju istih dokumenata na različitim web sjedištima, različitim formatima objave dokumenata, strukture dokumenata i pratećih elemenata za identifikaciju dokumenata. Osiguran je trajan pristup obrađenim sadržajima s jednog mjesta na internetskim stranicama HIDRA-e.
Ostaju trajni ciljevi:

- razvijati sustav trajne pohrane sadržaja u DAMIR
- obraditi formalno i sadržajno prikupljene i odabrane sadržaje i učiniti ih dostupnima javnosti u što kraćem vremenu
- razvijati mogućnosti pristupa DAMIR-u
- osigurati tehničke uvjete za brzi i nesmetani rad pobirača i pohranu sadržaja sve većeg broja web sjedišta
- ukazivati na niz pitanja koja su otvorena primjenom automatskog prikupljanja službene dokumentacije, a koja se trebaju rješavati suradnjom tijela javne vlasti uvođenjem normizacije prikaza sadržaja na internetskim stranicama.

Razvoj sustava

Okosnicu sustava DAMIR čini **A**rhiv **M**režnih **D**okumenata – **AMD**, modularan, proširiv alat jednostavan za uporabu, utemeljen na programskoj podršci otvorenog koda. Razvijen je na temelju iskustava ***Srca*** u području istraživanja weba, a prema potrebama HIDRA-e. Najvažnije komponente su podsustav za pobiranje, skladište podataka i web sučelje za upravljanje pobiranjem i odabirom sadržaja.
Izvor podataka za AMD sustav su podaci o web sjedištima institucija javne vlasti prikupljeni u Hidrinim bazama podataka. Razvijena je procedura obnove podataka odabranih web sjedišta tijela javne vlasti i političkih stranaka u skladištu podataka DAMIR i obavlja se jednom mjesečno. Procedura podrazumijeva promjenu postojećih podataka o već aktivnim web sjedištima i dodavanje podataka o novim web sjedištima.
Osnovna namjena DAMIR-a rješava potrebu automatskog pobiranja pojedinih dokumenata objavljenih na web sjedištima tijela javne vlasti. Za svako web sjedište moguće je podešavanje parametara vremena i učestalosti pobiranja, parametara pobiranja određenih formata dokumenata, dubine pobiranja i sl. Nakon pobiranja pojedinog web sjedišta dodatno se periodički podešavaju parametri kako bi se postigli što bolji rezultati. Parametri se naročito prilagođavaju pri svakom redizajnu pojedinog web sjedišta.

Ovisno o učestalosti objavljivanja dokumenata na pojedinom web sjedištu modeliraju se i podaci o rasporedu i učestalosti pobiranja. Putem web sučelja za svako pobiranje pojedinog web sjedišta moguć je pristup rezultatima pobiranja, pregledavanje arhiviranih dokumenata i odabir dokumenata za daljnju obradu.
Automatsko pobiranje sadržaja s web sjedišta ponekad je otežano ili nemoguće, ovisno o načinu na koji se primjenjuju pojedine tehnologije prilikom izrade sjedišta. Tijekom izgradnje DAMIR-a i stjecanja iskustava objavljene su tehničke preporuke sukladne s preporukama W3C organizacije, krovnog tijela koje brine o standardizaciji web tehnologija s ciljem bolje dostupnosti internetskih stranica. Za potrebe pohrane važnih dokumenata u DAMIR, a koji nisu mogli biti pobrani automatski iz tehničkih razloga, nadograđen je modul za ručni unos dokumenta.
Isto tako dio dokumenata dostupan na web sjedištima pohranjen je u bazama dokumenata i moguć im je pristup odabirom ili pretraživanjem, ali ne i pobiračem. U bibliografskim bazama HIDRA-e dio je takvih sadržaja povezan i čini virtualnu bazu dokumenata. Zbog opasnosti promjene pristupa ili naprosto brisanja dokumenata, razvijen je modul za unos dokumenata prema listi dokumenata.
Neovisno o prvobitnoj namjeni, razvijen je i profil za pobiranje i arhiviranje web stranica svih web sjedišta. Profil ima zajedničke osnovne podatke naziva i adresa za pobiranje s profilom dokumenata, ali zasebne parametre za pobiranje sadržaja, poput dubine i učestalosti pobiranja.
Učestalost pobiranja webova je jednom godišnje i dosad je obavljeno dva puta. Za sva web sjedišta postavljena je dubina pobiranja 3. Time se pobire gornja razina web sjedišta, dobiva osnovni pregled sadržaja pojedinog sjedišta, pregled glavnih tema i pregled primjene novih tehnologija. Moguće je pratiti razvoj uvođenja web tehnologije u primjenu i razvoj komunikacije s korisnicima sadržaja. Obrada arhiviranih web stranica odgovara načinu obrade periodičkih publikacija.
Arhiv se izgrađuje samo dopunjavanjem sadržaja bez brisanja, ni pojedinih dokumenata, ni institucija u cjelini. Moguće su samo selekcije za izdvajanje dijelova za različite potrebe. Pojava institucija koje prestaju djelovati rješava se podešavanjem parametra za učestalost pobiranja i time postojeći dokumenti ostaju u arhivi, ali pobirač više ne obilazi web sjedište takve institucije.
Prihvaćeni i obrađeni (formalno i sadržajno) dokumenti postaju dio zbirke Službene dokumentacije Republike Hrvatske ili zbirke Političkih stranaka (u bibliografskoj obradi svi nose oznaku medija **online e-arhiv**), te postaju dostupni:

- na razini bibliografskih jedinica u okviru tiskanog godišnjeg biltena prinova: *Katalog službenih publikacija i baza podataka Republike Hrvatske*
- na razini punog teksta ukupno prikupljenih dokumenata predstavljenih bibliografski po abecedno svrstanim tijelima javne vlasti kao korporativnim autorima ili izdavačima na DVD-u: *E-arhiv službenih publikacija i dokumenata tijela javne vlasti Republike Hrvatske*

- na razini punog teksta kroz pretraživanje cjelovitog fonda zbirke *Službena dokumentacija Republike Hrvatske* putem e-Kataloga (http://www.hidra.hr/dok/1katalog.htm)
- u dokumentaciji *Političke stranke u Republici Hrvatskoj* na http://www.hidra.hr.

Kriteriji odabira dokumenata

Prvi kriterij odabira građe obavlja se izborom adresa web sjedišta s kojih se građa pobire. Sljedeći važan korak je odabir formata dokumenata pri modeliranju sustava za pobiranje. Formati se zadaju ovisno o načinu i vrsti prikaza sadržaja na pojedinim web sjedištima i važnosti samih dokumenata.

Prikupljena građa s web sjedišta tijela javne vlasti i političkih stranaka RH odabire se prema HIDRA-inim *Kriterijima odabira mrežnih izvora službene dokumentacije Republike Hrvatske za postupak arhiviranja* koji se odnose na:

- kategoriju službenosti sadržaja dokumenata, što podrazumijeva prihvaćanje službene dokumentacije tijela javne vlasti koje je autor i/ili izdavač tijelo s čijeg se web sjedišta prikuplja dokumentacija; prihvaća se dokumentacija tijela državne vlasti neovisno o sadržaju, a za ostale skupine (ustanove javnog sektora, poduzeća od javnog interesa, diplomatsko-konzularna predstavništva RH, lokalna i područna samouprava i političke stranke) se prihvaća službena dokumentacija koja se izravno odnosi na rad tijela i zakonski djelokrug poslova
- identifikaciju odgovornog tijela kao tijela javne vlasti RH ili registrirane političke stranke u RH; temelji se na godišnjem *Popisu tijela javne vlasti* koji se objavljuje u Narodnim novinama i *Registru političkih stranaka*
- strukturu/cjelovitost dokumenta.

Kriterijima se želi postići formiranje zbirke službenih dokumenata Republike Hrvatske koji se mogu odrediti prema vremenu i izvoru nastanka nakon izdvajanja iz okoline koju imaju kao dokumenti na web sjedištu. Za provođenje što kvalitetnijeg arhiviranja, HIDRA je objavila *Arhiviranje mrežnih izvora informacija tijela javne vlasti Republike Hrvatske: preporuke autorima i nakladnicima*. Preporuke se odnose na dodjelu identifikacijskih podataka mrežnim izvorima službene dokumentacije RH i na tehničke preporuke za izradu mrežnih izvora.

Rezultat primjene kriterija pokazao je velike razlike u količini prikupljenih i prihvaćenih dokumenata. Naime, dokument ne može biti prihvaćen ako ne postoje identifikacijski podaci unutar samog dokumenta (naslov i intelektualna odgovornost; podaci o objavljivanju; materijalni opis i identifikacijske oznake). Osim toga, na web sjedištima se objavljuju i dokumenti koji svojim sadržajem i autorstvom ne odgovaraju definiciji službenog dokumenta, isto kao što je čest slučaj da se isti dokument objavljuje na službenim web sjedištima raznih tijela javne vlasti, no prihvaća ga se samo s onog koje odgovara autoru/izdavaču do-

kumenta. Isto tako, AMD sustav prepoznaje promjene u formatu dokumenta, veličini i datumu stvaranja. Ukoliko se isti dokument objavi u nekoliko različitih formata, sustav prikuplja sve zadane formate koje razlikuje po ekstenziji u nazivu dokumenta. Promjena veličine dokumenta i datuma stvaranja su podaci koji određuju ponovo pobiranje dokumenta iako se ponekad ne radi o novoj verziji sadržaja dokumenta, nego intervenciji u identifikacijske podatke, formalni prikaz ili samo prolaz dokumenta kroz neki sustav za rukovanje dokumentima.

Sadržaj arhiva

Profil pobiranja dokumenata

Broj web sjedišta s kojih se automatski prikupljaju dokumenti u kolovozu 2007. godine iznosi 946. Od toga s 340 web sjedišta nije prikupljen ni jedan dokument. Uglavnom se radi o web sjedištima s osnovnim informacijama o djelatnosti tijela u html formatu i bez službenih dokumenata. S obzirom da se format html prikuplja samo u izuzetnim slučajevima, takva sjedišta nemaju prihvaćenih dokumenata (Tablica 1.).

Tablica 1. Osnovni pokazatelji sadržaja DAMIR-a

Broj web sjedišta	946
Broj web sjedišta bez pobranih dokumenata	340 (36%)
Broj pobranih dokumenata	125.947
Broj prihvaćenih dokumenata	23.843 (19%)
Veličina arhiva	37 GB

Do sada je ukupno pobrano 125.947 dokumenata i to pretežno u PDF i DOC formatu (Tablica 2.).

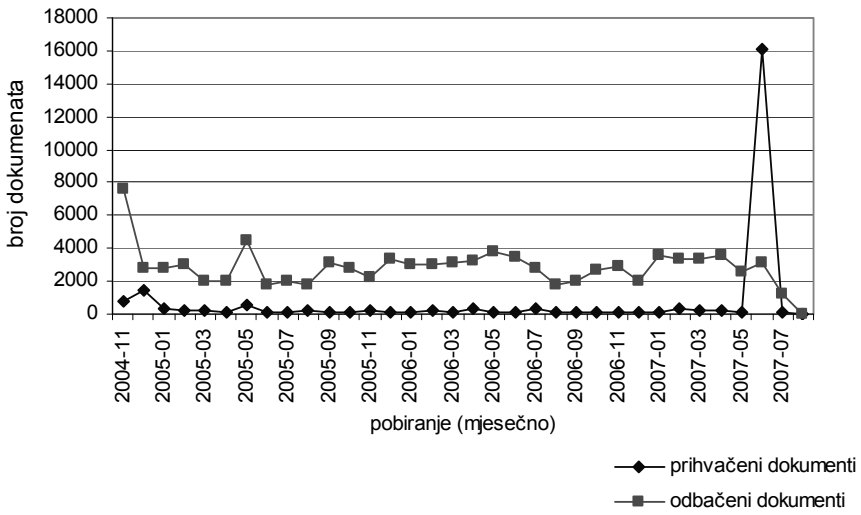
Tablica 2. Broj pobranih dokumenata prema formatu

pdf	71.814
doc	25.744
html	17.446
xls	7.829
ppt	1.757
zip	686
rtf	552
ostalo	119

Odabrano je i katalogizirano 23.843 dokumenata (oko 19%) prema navedenim kriterijima odabira. Prema Tablici 3. *Odnos prihvaćenih i odbačenih dokumenata mjesečno* vidi se da je omjer podjednak u cijelom razdoblju prikupljanja dokumenata uz iznimke u prosincu 2004. godine na samom početku rada s pobiračem, u svibnju 2005. godine kada je instalirana poboljšana verzija pobirača kojom je bilo moguće i poboljšano pobiranje te u lipnju 2007. godine kada je

uključeno pobiranje po listi. Svi dokumenti s liste prethodno su formalno i sadržajno obrađeni i dobivaju oznaku prihvaćen.
Odnos prikupljenih i prihvaćenih dokumenata zabrinjavajuće je mali, no uz sve navedene i primijenjene kriterije odabira dokumenata on je približno stalan.

Tablica 3. Odnos broja prihvaćenih i odbačenih dokumenata mjesečno



Analiza odnosa prema državnom ustroju prema Tablici 4. pokazuje da postoje skupine koje se postupno više pridržavaju standarda i potpunije opremaju objavljene dokumente pa oni ulaze u skupinu prihvaćenih.
Velik dio prikupljenih dokumenata tijela državne vlasti odnosi se na radne materijale sjednica Hrvatskog sabora i Vlade RH, te govore Predsjednika Republike Hrvatske. Iako je prihvaćeno oko 50% tih dokumenata, velikom broju dokumenata nedostaju osnovni identifikacijski podaci i vrlo je teško pratiti zakonodavni postupak od nacrta prijedloga zakona sa sjednica Vlade, promjene kod rasprava na Saborskim sjednicama do konačnog prijedloga zakona.
Slijede ministarstva, a najviše prikupljenih dokumenata imaju Ministarstvo mora, turizma, prometa i razvitka, Ministarstvo kulture, Ministarstvo poljoprivrede, šumarstva i vodnog gospodarstva, Ministarstvo zaštite okoliša, prostornog uređenja i graditeljstva, no to nužno ne povlači i najveći broj prihvaćenih dokumenata. Npr. Ministarstvo kulture s 2.298 prikupljenih dokumenata ima samo 4 prihvaćena, jer se većina objavljenih dokumenata odnosi na izvatke iz tiska.
S web sjedišta ministarstava prihvaćaju se razna izvješća, planovi, strategije, ali i mnogobrojna izdanja namijenjena informiranju građana, u čemu prednjači Ministarstvo vanjskih poslova i europskih integracija koje izdaje mnoge publikacije na temu približavanje EU građanima.

Zanimljivo je da jedino ministarstvo koje nema niti jedan prikupljeni dokument jest Ministarstvo obitelji, branitelja i međugeneracijske solidarnosti, a razlozi se mogu tražiti u načinu izrade web stranica i nemogućnosti pobiranja sadržaja. Prema Tablici 5. *Web sjedišta institucija s najvećim brojem objavljenih dokumenata* Splitsko-dalmatinska županija ima prikupljenih 2.475 dokumenta, ali niti jedan prihvaćen, jer svaki prikupljeni dokument predstavlja jednu stranicu službenog glasnika. To bi značilo da bi svaki broj glasnika u bio sastavljen od 50-tak i više linkova.

Tablica 4. Odnos prikupljenih i prihvaćenih dokumenata u DAMIR-u RH prema državnom ustroju

Institucije	Broj web sjedišta	Ukupno dokumenata	Broj prihvaćenih dokumenata	% prihvaćenih
Tijela državne vlasti RH	81	27.363	4.342	15,87
Ustanove javnog sektora	202	52.134	1.498	2,87
Poduzeća od javnog interesa	18	20.753	15.489	74,63
Jedinice lokalne i područne (regionalne) samouprave	218	23.038	2.182	9,47
Diplomatsko-konzularna predstavništva RH	15	384	12	0,26
Diplomatsko-konzularna predstavništva u RH	19	496	0	0.00
Političke stranke	53	1779	320	17,98
Ukupno	**606**	**125.947**	**23.843**	**18,93**

Kod ustanova javnog sektora, od 52.134 prikupljena dokumenta prihvaćeno je njih 1.498. Omjer prikupljenih i prihvaćenih dokumenata manji je od 3%. Velikom broju prikupljenih dokumenata na ovoj razini doprinose sveučilišta i fakulteti. Te institucije produciraju velik broj dokumenata, no većina ih ne zadovoljava kriterije službenog dokumenta nego su to tekuće obavijesti za studente, pa to bitno utječe na vrlo veliku razliku između prikupljenih i prihvaćenih dokumenata kod ustanova javnog sektora. Ovo potvrđuje i podatak iz Tablice 5., gdje se Fakultet strojarstva i brodogradnje našao na 5. mjestu po broju objavljenih dokumenata, a od toga je prihvaćen samo jedan.

Poduzeća od javnog interesa odnose se na 18 tijela, ali je s tih web sjedišta prikupljeno preko 20.753 dokumenta, a gotovo 75% dokumenata je prihvaćeno. Tome najviše doprinosi pobiranje po listi sadržaja Narodnih novina. Većinu neprihvaćenih dokumenta čine izvadci iz novina, razna priopćenja za medije i slično, no prihvaćeni dokumenti mogu se ocijeniti vrlo visokom ocjenom.

Naime, većina ustanova javnog sektora vrlo je redovita i točna u objavi službenih dokumenata, a problem često predstavlja objava istih dokumenata s promijenjenim metapodatcima, pa pobirač više puta prikuplja isti dokument.

Jedinice lokalne i regionalne samouprave trenutno od 23.038 prikupljenih imaju 2.182 prihvaćenih dokumenata, odnosno oko 10%. Većina prihvaćenih dokumenata su Službeni glasnici (trenutno 50 naslova), Regionalni operativni planovi, Programi razvoja, Izvješća o stanju okoliša te razni priručnici namijenjeni informiranju građana o zdravlju, poduzetništvu, poljoprivredi, mladima i slično te se zahvaljujući sve bržoj informatizaciji lokalne i regionalne samouprave očekuje daljnji porast dokumentacije.
Mnoge jedinice lokalne i područne samouprave na web sjedištima objavljuju radne materijale sa sjednica gradskih vlasti, što je vrlo značajno za transparentnost rada tih tijela, ali prema kriterijima odabira Hidra radne materijale sa sjednica ne prihvaća. Primjer takvog nerazmjera u prikupljenim i prihvaćenim dokumentima je grad Rijeka, gdje je od 2.300 dokumenata prihvaćeno njih 10-tak, jer ostali predstavljaju radne materijale sa sjednica.
Općenito diplomatsko-konzularna predstavništva imaju mali broj web sjedišta, a postojeći nisu izvori obilne dokumentacije.
Oko polovica registriranih političkih stranaka Republike Hrvatske ima web sjedište, a oko 18% dokumenata je prihvaćeno i obrađeno.

Tablica 5. Web sjedišta institucija s najvećim brojem objavljenih dokumenata

Institucija	Ukupno dokumenata	Broj prihvaćenih dokumenata	% prihvaćenih
Narodne novine d. d.	15.456	15.308	99,04
Hrvatska narodna banka	5.104	359	7,03
Hrvatski sabor	2.941	1.438	48,89
Ministarstvo mora, turizma, prometa i razvitka RH	2.639	28	1,06
Fakultet strojarstva i brodogradnje Sveučilišta u Zagrebu	2.637	1	0,04
Vlada Republike Hrvatske	2.496	1159	46,43
Splitsko-dalmatinska županija	2.475	0	0,00
Grad Rijeka	2.397	11	0,46
Državni ured za reviziju	2.390	7	0,29
Ministarstvo kulture RH	2.298	4	0,17

Profil pobiranja webova

Za pojedino tijelo javne vlasti službeni dokumenti se pobiru s jednog web sjedišta. Međutim, s pobiranjem webova to nije slučaj. Odabir webova je kompleksniji, prikupljanje podataka opsežnije i zahtijeva daljnji rad na izradi kriterija za prikupljanje. Takvo se stanje jasno vidi pri pregledu web sjedišta koje otvara npr. Ministarstvo obrazovanja i športa RH kao potporu međunarodnim konferencijama, nacionalnim projektima i sustavima razvojnog značaja (*Mind and Brain*, *Nuclear and Particle Physics*, *Croatian Medical Journal*, *International*

Conference on Telecommunications, *Nacionalni informacijski sustav knjižnica RH*, *Sustav znanstvenih informacija RH* itd.).
Rezultati arhiviranja web sjedišta pokazuju obogaćivanje sadržaja i razvoj komunikacije s građanima – korisnicima. Tako je npr. velika razlika od suhoparnog sadržaja na stranicama Hrvatskog sabora u prvom prikupljanju u odnosu na novo stanje koje je bogatije novim dizajnom, primjenom novih tehničkih mogućnosti uz prijenose sjednica u vrijeme kada Sabor zasjeda.
Naročite promjene vidljive su od travnja 2007. i na stranicama Vlade RH. Sadržaj stranica je proširen, nove grupe sadržaja su Aktualne teme, Multimedija, Kutak za medije, Projekti ministarstava, Audio zapisi sjednica Vlade, pa je preko profila pobiranja webova zanimljivo pratiti kako se web sjedišta prilagođavaju potrebama korisnika usporedo s razvojem tehnologije.
Rezultati arhiviranja webova za tijela javne vlast još se istražuju i nakon katalogizacije, te formalne i sadržajne obrade bit će dostupne javnosti sve pohranjene verzije.

Zaključak

DAMIR po svojoj funkcionalnosti i razvoju u ovoj fazi treba shvatiti kao način prikupljanja sadržaja u razdoblju neuređenog korištenja web tehnologije, nerazumijevanja posljedica nepridržavanja standardima i želje da se bar dio objavljene građe što bolje koristi i trajno sačuva. Borba s neuređenim sustavom masovne objave informacija i dokumentacije doprinosi djelomičnom uređivanju sustava, razvoju klasifikacija i primjeni normi za organizaciju pristupa građi. Na temelju stečenih iskustava DAMIR može poslužiti i kao smjernica na putu stvaranja zajedničkog informacijskog prostora tijela javne vlasti i njihovih digitalnih sadržaja i usluga.
Odluke da u DAMIR-u ostaje pohranjen sav prikupljeni sadržaj ostavlja mogućnost i kasnijim procjenama za potrebu revizija kriterija odabira i značaja određenih sadržaja.
Rezultati izgradnje DAMIR-a su očiti. Prikupljeno je dosta sadržaja koji bi izostali prikupljanjem klasičnim metodama. Sadržaj je formalno i sadržajno obrađen te trajno dostupan, a ovo je prvi pokušaj da se na jednom mjestu objedinjeno obradi ukupno izdavaštvo tijela javne vlasti i stavi na uporabu javnosti.
Neposredni planovi razvoja DAMIR-a u suglasju su sa željom za što boljim prikupljanjem građe i promocijom zahtjeva na uređivanje i objavljivanje dokumenata te razvoj pristupa DAMIR-u. Razvoj sustava ovisi o više vanjskih faktora: pridržavanju osnovnih pravila pri objavi dokumentacije tijela javne vlasti, primjeni standarda pri izradi web sjedišta i razvoju alata za pobiranje.

Literatura

Arhiviranje mrežnih stranica tijela javne vlasti Republike Hrvatske: Preporuke autorima i nakladnicima // HIDRA (2006.). http://www.hidra.hr/hidra/hidra-preporuke-v01.pdf (16.08.2007.)

Baršić-Schneider Ž. Pretpostavke i kriteriji za prikupljanje, odabir i pohranu dokumenata. 2004. http://www.hidra.hr/hidra/10godina/hidra-srce101204.pdf (16.08.2007.)

Bjelić M., Plavac J., Popović D.: The Archive of Network Resources Collection, selection, processing and storage of documents acquired from the official websites of the public authority bodies of the Republic of Croatia. 2005. http://www.hidra.hr/hidra/10godina/lida-hidra05-30.ppt (16.08.2007.)

Cabinet Office e-Government Unit. Annex L. Archiving Websites. 2002 – [UPDATED MARCH 2004]. http://www.cabinetoffice.gov.uk/e-government/resources/handbook/html /annexl.asp (16.08.2007.).

Garvas Delić A. Arhiva Mrežnih Dokumenata – Prikupljanje, Odabir, Obrada i Pohrana Dokumenata sa službenih internetskih stranica državnih tijela RH. 2004. http://www.hidra.hr/hidra/10godina/amd-poop.pdf (16.08.2007.)

Garvas Delić A., Milinović M.: Prikupljanje, odabir, obrada i pohrana dokumenata sa službenih mrežnih stranica tijela javne vlasti Republike Hrvatske, Arh. Vjesn., god 49 (2006), str 91-106

Gatenby P. Collecting and managing web resources for long –term access : web harvesting and guidelines to support preservation (ICABS Actions 3.3 and 3.4). 2004. http://www.ifla.org /IV/ifla70/papers/026e-Gatenby.pdf . (16.08.2007.).

Information Management Office-Australian Government. GovernmentOnline: the Commonwealth Government’s strategy. 2002. http://www.agimo.gov.au/publications/2000/04/gov online. (16.08.2007.).

Milinović M. Arhiv mrežnih dokumenata. 2004. http://www.hidra.hr/hidra/10godina/amd-10122004.pdf (16.08.2007.)

Operativni plan provedbe Programa e-Hrvatska 2007. za 2006. godinu // Središnji državni ured za e-Hrvatsku (2006.). http://www.e-hrvatska.hr/repozitorij/dokumenti/downloads/plan_ provedbe_e-Hrvatska_za_2006.pdf (16.08.2007.)

Popis tijela javne vlasti za 2007. godinu // Narodne Novine (2007) 11. http://www.nn.hr/clanci/ sluzbeno/2007/0456.htm (16.08.2007.).

The National Archives. Records of the UK government from Domesday to the present. http://www.nationalarchives.gov.uk/ (16.08.2007.).

Zakon o knjižnicama // Narodne Novine (1997) 105. http://www.nn.hr/clanci/sluzbeno/1997/ 0767.htm (16.08.2007.).

E-Identity: Responsibility or Commodity?

Sanja Seljan
Hrvoje Stančić
Dina Crnec
Anita Salopek
Faculty of Humanities and Social Sciences
Ivana Lučića 3, HR-10000 Zagreb, Croatia
sanja.seljan@ffzg.hr
hrvoje.stancic@zg.t-com.hr
dcrnec@ffzg.hr
asalopek@ffzg.hr

Summary

The aim of this paper is to analyze importance of e-identity, related issues (e.g. trustworthiness of information, pseudonimity, privacy, fraud, property, etc.) and role of virtual life in the real one. As Internet users are in great proportion service-oriented and interested in interaction, intermediation and social network communication, the reputation of e-identity has become increasingly important. In this paper two aspects of digital identities are analyzed: encrypted identity (e.g. passwords) and personal virtual identities, used in forums, chats, blogs, etc. which are created through explicit statements (true or false) of users and by implicit ways of interaction and behaviour.
After the analysis of the main types of identification and web services, results of the pilot study, conducted among students of the Faculty of Humanities and Social Sciences, regarding reliability of information of e-identity has been presented. The aim of the paper is not to list detailed issues but to raise the awareness related to the potential problems in the sphere of e-identities and suggest the appropriate actions.

Key words: e-identity, real identity, virtual identity, information, reliability, reputation, protection

Introduction

Online identity vs. real life identity?

In most dictionaries identity is defined as the condition of being the same with something described or asserted. While humans feel unique and have unique physical identity, digital identity is a set of characteristics asserted by one digital subject. According to Nabeth (2005a) the quality of online identities (how they represent themselves and how they are perceived) has direct implications

on the quality of the space and interactions (e.g. evaluation of the product, relevance of information, raising interest for certain topics). In its simplest way digital identity is represented through the user name and the password, which is just one type of authentication.
Using e-identities in different areas of life (finances, education, fun, etc.) in some instances people can freely redefine themselves on the Internet as they wish and in many cases it involves deception: most users prefer to identify themselves by means of pseudonyms, which allow them to hide their real identities while online, but still let them build a reputation with their usernames. Also, often users have more than one e-identity (e.g. eBay for bids or simply for escaping the consequences of negative reputation). One of the greatest fears of e-identity is having one's identity stolen or abused, and in most services these issues are being dealt with by law protection.
Regarding e-identities in the academic area, users can ask their provider to hide their identity but then they lose the possibility of using certain services in the system. Here, as well as in e-banking, the question of trusting the "guaranteed privacy, without fear of unauthorized access" is being posed. Also, major concerns regarding virtual identity revolves around the areas of misrepresentation, the effects between on and offline existence, as well as the validity of online relations. Whether we refer to online communities, e-classes, e-banks or blogs, these include the same issues, being either positive or negative and building up a sense of e-identity and its preservation.
The notion of human identity is an intuitively understood concept, but hard to formalize. Identity is created out of attributes that differentiate each individual from other individuals.
According to Wiki, "digital identity refers to the aspect of digital technology that is concerned with the mediation of people's experience of their own identity and identity of others" or it is defined as the representation of a "human identity used in a distributed network interaction with other machines or people with purpose to ease human communication".
Digital identity is created through a set of claims made by one digital subject about itself or another digital subject, therefore it is a subset of the characteristics that identify a person. Closely related notion is digital identity management described as an "interaction between user and his identity information stored and maintained digitally" (Gautirez, 2006).
The saying *in cyberspace nobody knows you're a dog* reflects part of identity problems in the electronic environment.
E-identity raises many issues, such as trustworthiness or reliability of data, pseudonimity, privacy, fraud, property, propagation, etc.

Virtual life vs. real life

Every Internet user is dealing with this doubt: should there be difference between my virtual life and my real life? In every sphere of e-life (finance, educa-

tion, fun, etc.) exists request for e-identity. When will users disclose or hide themselves? What is the main reason of hiding or consequence of disclosing? Considering student's opinion we will try to answer some of these questions and elaborate on e-identity in the areas of finance, education and fun showing positive and negative sides.
The idea of a virtual world, such as that found in Second life, is all about wish fulfilment. People are not showing who they are but who they would like to be. Here is misrepresentation in charge. What can virtual life offer? For some users, it is easier to express their opinions and feelings incognito on chats, blogs or forums, or to use services on Internet (e.g. e-banking), or to check a lesson using e-learning system. One blogger said that virtual life is a great opportunity to escape from problems of real life. That's his way of dealing with problems. But if virtual life is so perfect, why are users then afraid to use e-banking or online buying services? There is also threat of stealing usernames, passwords and disturbance of privacy. So virtual life besides the lack of face-to-face communication could also bring problems that users can hardly avoid.

e-Banking

Users of e-banking services require e-identity. Users are given a token (password protected device) that gives access to the control bills, of buying and selling ect. The token is used for protection of electronic identity.
One of the key aspects is an authentication, providing codified assurance of the identity. Authentication methods include in most cases passwords, bank credit card or encryption methodologies, but also some new techniques such as handprinting or voiceprinting. According to Tsui (2006) today there are two main problems on the Internet: safety (security and privacy) and convenience.
The study undertaken by Arcot (2007) claims that "the identity theft and fraud is the nation's fastest growing crime (…) and a big business too" thus showing the necessity for protection of e-identities.

E-identity in education

Security and identity management at universities are ranked as critical issues because of their strategic importance. Applying basic functions for access management (identification, authentication, authorisation accountability – IAAA) the system is required to:

- request or assign the username (identification),
- validate if the token belongs to the user (authentication) which can be password, identity card or personal attribute (fingerprint, voiceprint, retina pattern),
- ensure the user has allowed level of access to the services and data (authorization),

- ensure that only authorized persons can assume authorized active sessions, providing the appropriate identification, authentication and authorization (accountability).

AAI@EduHr is an infrastructure enabling a unique authenticity and authorization check created with the purpose to ensure simple and efficient use of all resources of the system, including CARNet (Croatian Academic Research Network) services, Edu Roam (using the same computer during the visit in European universities), CMU (CARNet Modem Access) etc. The system is cheap and can be used only by students, professors and members of CARNet. CARNet's e-identity is a virtual identity that is proving student's status and enables him or her Internet access, e-mail address, CARNet online courses ect. E-identity is determined by username (login, user id) and password. User also gets account on public server. Anonymous or group accounts are not allowed. User can ask other extra services (ftp, web site). Personal information about a user is public and can be changed by the user. On user's demand information could be removed. But in that case user is losing the possibility of using services of AAI@EduHr system.

E-identities and web services

Reputation of e-identity has become increasingly important, since Internet is more and more used for interaction (online chats, forums, multi-player online games, virtual community system, Wiki, CMS, e-mail, MUDs-Multi-User Dimensions) or intermediation (digital marketplaces, online social networks) with other people creating social life of the virtual community. According to Nabeth (2005b) a considerable part of population is less centred on information and more on the social process.

E-identities can reveal various amounts of personally identifiable information. Online users at different web services (including forums, instant messaging, massively multiplayer online games etc.) can freely redefine themselves as they wish and in most cases, it involves deception. Most users prefer to identify themselves by means of pseudonyms, which allow them to hide their identities online, but still let them build a reputation with their username. Users can hide behind their computers saying whatever they want to say with little consequences and without fear of persecution.

Although these "online masks" do not reveal the actual identity there seems to be certain connection between virtual and real identities. When it comes to online communities such as MySpace and Facebook, people create their own social network and an important part of it is popularity. The attention one gets as a result of their page visits makes signing in onto the networks on the daily basis even more addicting. In MMORPGs (Massively Multiplayer Online Role Playing Games) such as World of Warcraft, online identities that people choose are "less ordinary" than their real selves. Virtual environments can be a way of

expressing a different side of personalities or escaping the social constraints of the real life, and they often mirror human behaviour in the real world.
An online reputation is often achieved while in a pseudonymous state. A digital reputation can be useful in situations where credibility must be established, but exposure of one's real life identity is undesirable or possibly hazardous.
An online identity that has acquired an excellent reputation is a valuable one. A person invested a great deal of time and effort to build the identity's reputation while other users decide whether it is sufficiently trustworthy. Some of those online identities have been put up for sale at online auction sites, but there are conflicts over the ownership of online identities – whether they belong to the person who created them, or to the company that owns the software used to create those identities.
Because of the malleability of online identities, different sites, such as the trading ones, have developed certain forms of reputation management (e.g. eBay's feedback system that records transactions and provides the technical means by which users can rate each others' trustworthiness). Here, when pseudonymous identities propose to enter into an online transaction, the deal can succeed only if the parties are willing to trust each other, but they have no rational basis for doing so.
One of the situations is selling online identities with good reputation. According to the information on wiki, a good reputation player of the online game Everquest, owned by Sony Online Entertainment, attempted to sell his identity on eBay. Sony protested pointing out that this identity is their intellectual property. Therefore, the question arises: who is the owner of identities on commercial websites?

Stealing & abuse

One of the greatest fears concerning e-identities is having one's identity stolen or abused. In most online services these issues are being covered by law regulations. The law forces often state that online anonymity gives access to criminal actions willing to mask their identities and call to start with identity management infrastructure that would relate online identity with person's legal identity.
Identity thieves can be divided in groups concerning the type of identity theft committed: financial (obtaining goods and services), criminal (including e.g. terrorism and espionage), business (using other's business name to obtain credits) and identity cloning (using other's information to assume his or her identity in the real life).
One of the examples regarding the role of the reputation and selling e-identities (Nabeth, 2005a) is when a committer has succeeded to already take $200,000 before being caught and his real identity revealed. After creating an excellent reputation on eBay, by selling computers in order to create vision of authenticity, he used multiple user IDs and every time when he was discovered he had already assumed a new identity.

Another example was in online gaming when certain users with unauthorized powers made other players impossible to win or survive (Becker, 2002). Another illustration could be an example of false identity created by social phishing (Nabeth, 2005a). The story says that after being disfigured in the car accident, caused by a drunk driver, the girl remained mute and in wheelchair and her boyfriend was dead. Using computer and participating in a BBS (Bulletin Board System) the girl made many friends and showed shining personality. Afterwards it was revealed that the girl was not at all a disabled person and that "she" was actually "he".
According to the BBC Alerts, the Javelin Identity Fraud Survey Report and an update of the Federal Trade Commission's Identity Theft Survey Report "despite the growing fears about identity theft and online fraud these crimes are more frequently committed offline than online. Internet-related fraud problems are actually less severe, less costly and not as widespread as previously thought". The Javelin study also says that computer crimes are included with 11.6 per cent in overall identity frauds in 2004 (often by opening pop-ups). The most frequently committed fraud was done by stealing a wallet or a checkbook. Another question relating to the fraud committer, reported that half of all identity fraud is done by a person who knows the victim. The report also indicates that the majority of crimes are self-detected by the victims which show the importance of monitoring electronic transactions. Based on the previous findings, some advices for consumers are given regarding regular updating, being suspicious to sending account numbers, taking care of important documents, keeping passwords hidden, ignoring suspicious Internet links, checking credit cards' status and resolving fraud promptly.

Who would you like to be?

According to Chan, a photographer, analyzed relationship between gamers' real and online identities by comparing images of the two persons – the real one and the online version, aiming to show whether digital representations in the role-playing environments were echoes of true personalities. He indicates that in the beginning the users enjoys the anonymity, but then this character becomes a bigger part in the life. Online identity is more flexible, can change attributes (gender, age, occupation, etc.), and generally were more interesting (good-looking, more power, better, faster) and when trying to find a fat one, he did not succeed.
A study at the Stanford University (Yee/Bailenson, 2007) has shown some measurable trends in the character creation, e.g. introvert persons would generally choose their avatars as ideal version of themselves. Study suggests that virtual environment can also be an escape from the real life, but when having complete freedom, users often bring themselves into appropriate norms and do regular things (like shopping) thus falling into a trap of complete freedom. One

of the conclusions is that avatars change depending on the way of our communications with others (Yee/Bailenson, 2007).
According to Wiki, although a person is hiding behind a mask, it does however reveal certain interests. To prove that fact and research further into the opinions of the students on the questions concerning e-identities we have decided to conduct a study.

Research at the Faculty of Humanities and Social Sciences

The study was made among the students of the Faculty of Humanities and Social Sciences in Zagreb regarding reliability of e-identity representations. The questionnaire that was given to 30 randomly chosen students had eleven questions. Some questions allowed multiple answers. Here we present only the most important results.

Chart 1: What kind of person do you think you are in real life?

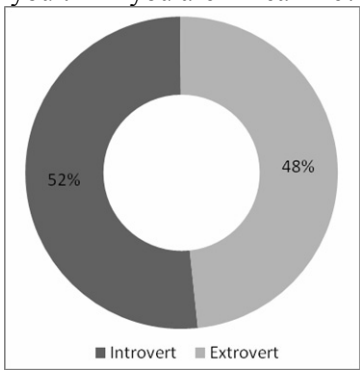


Chart 2: Have you ever created virtual identity of yourself using your own data?

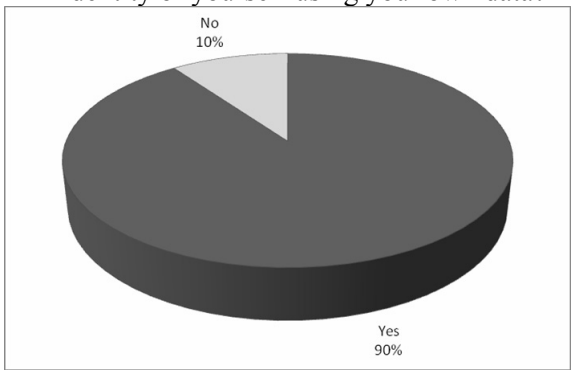


Chart 1 interestingly shows that roughly half of the surveyed students consider themselves as extrovert and the other half as introvert person. In our opinion this result makes further results personality independent. Chart 2 unexpectedly showed that 90% of the surveyed students have created a virtual identity by themselves. This does not include formal virtual identities created by, e.g. a bank or the Faculty. Such a high percentage shows great involvement of the students with the electronic environment. This result is an important one for gaining a better insight in the functioning of the young population. Results from Chart 3 show that the students are mostly involved in the discussions on Internet forums and chats followed by the virtual identities created for the purpose of education. Chart 4 shows that they use their virtual identities fairly regularly although some of them have obviously created a virtual identity, probably for some one-time reason or out of the curiosity, and never used it afterwards.

Chart 3: Where did you use your virtual identity?

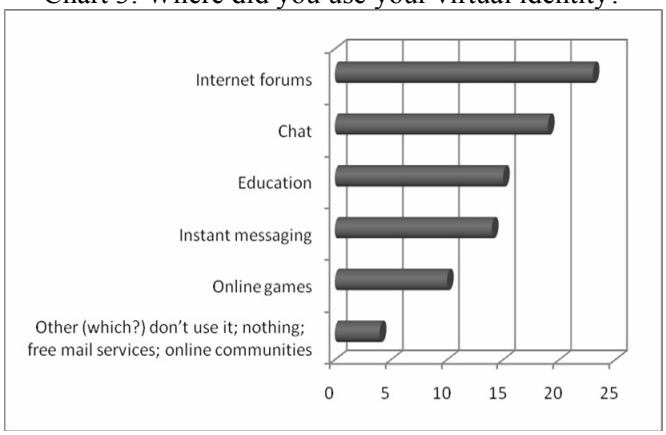


Chart 4: How often do you use your virtual identity?

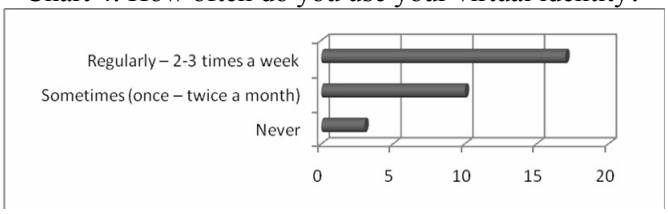
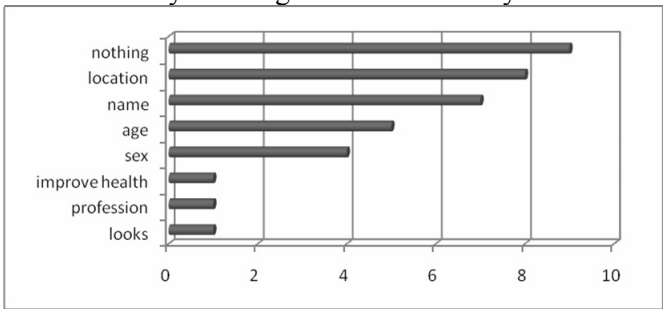


Chart 5: What would you change from real life in your virtual identity?



It is worth pointing out that the surveyed students claimed that their virtual and real identity would not differ, as seen from charts 5 and 6. Usually, it is believed that people are creating a different personality in the virtual environment. It seems that their experience goes in that direction too, as seen from chart 7. However, Chart 5 shows that they would change the information about their location and name which enables them to achieve anonymity. Other results, not graphically presented here, showed exactly their opinion that the main advantage of virtual identity is anonymity which allows them to speak freely (56.3%). Yet they claim the main disadvantage of virtual identity is its difference from

the real identity. It seems that although the students liked the anonymity and freedom of speech enabled by the creation of a virtual identity, they in most cases either did not use the opportunity of an anonymous questionnaire in the paper form to state that they would probably change something in their virtual identities or the surveyed pool of students on overall did not fit in the general picture.

Chart 6: If created, what would be the main difference between your virtual identity and your real identity?

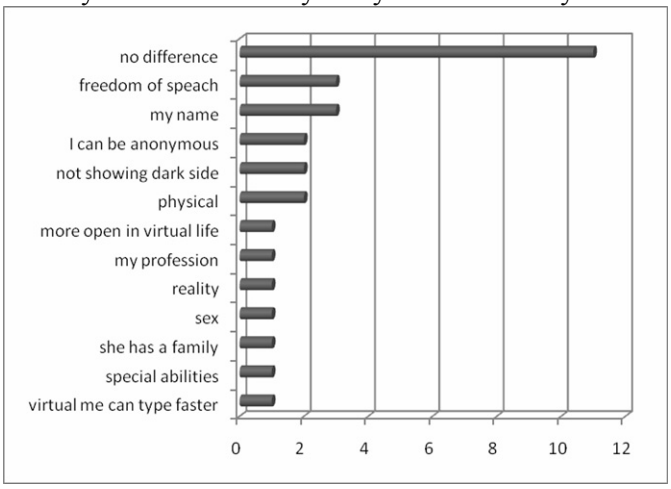


Chart 7: In your opinion what is the main disadvantage of virtual identity?

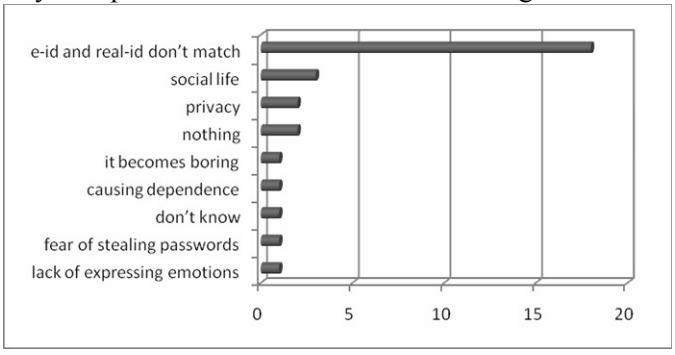
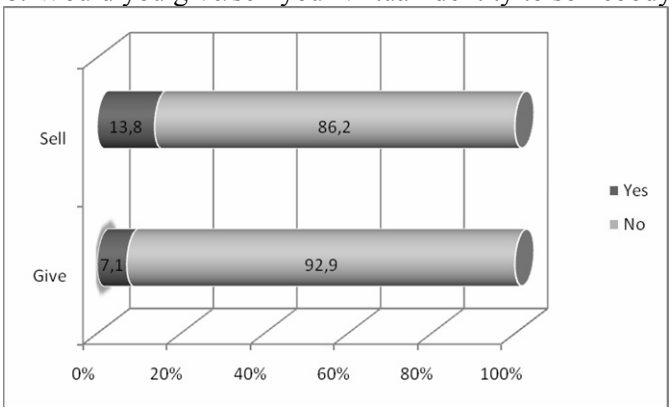


Chart 8 shows the dilemma whether give or sell private virtual identity to somebody else or not. The results showed that 7% of students would give their virtual identity to other person, but when asked whether they would sell it – the percentage doubled to 14%. It is clear that the dangers of giving or selling an identity in the electronic environment should be explained more thoroughly to

their users prior the exchange or transaction occurs. These results show what ideas should be stressed out in the process of education.

Chart 8: Would you give/sell your virtual identity to somebody else?



The idea of the ability of every user to self portrays his or her virtual identity has resulted in much discussion about the validity of online relations. The connection between online and offline lives poses the question what are real experiences and how a virtual one may affect one's offline emotions.
It was proved that an e-identity is actually an echo of people's true selves (e.g. real portraits and virtual images of gamers and people's digital representations in role-playing environments). According to several researches, most people enjoy the anonymity at first, but the character they created then becomes a bigger part of their life. An online identity has given people the opportunity to feel comfortable in wide ranging roles, some of which may be underlying aspects of the user's life. In a virtual world, an online identity is potentially much more flexible than the real identity, allowing easy changes of race, gender, age, socio-economic background and even species. It more freely offers self-definition, including multiple identities and shared identity, where there are no behaviour guidelines or prescribed etiquette.

Conclusion

The concept of this paper relies more on an informal perspective of e-identities than on the formal, financial perspective or security and technical issues. Users are engaged, through a number of web services and mechanisms (blogs, e-mails, wikis, reputation systems, forums, Massively Multiplayer Online Role Playing Games), in the creation of e-identities using either their own views of themselves or their virtual identities form according to the views of others. Social digital networks bring benefits in quick and easier communication, interaction, mediation, but also raise the important questions regarding privacy, reli-

ability, trust, establishment of social relations, identity management in e-education, e-banking, etc.

Digital identity is one of the key issues in economy, research and ICT education, addressed by different EU initiatives and projects, conferences, workshops, forums and education courses.

As the border between the real and the virtual world has become interwoven and is constantly changing, there is certain complementarity between virtual and real lives. This was proved in the survey of the students where 90% of them claimed that they have created at least one virtual identity. That finding showed the amount of possible impact of the mutual influences of virtual and real identities. It is very important to acknowledge that fact especially knowing that, according to the survey, 14% of the students, i.e. future young professionals, would sell their virtual identities. This fact calls for a wider action of raising the awareness of the importance of the responsible management of e-identities. These ideas should became embedded throughout the education system and thoroughly studied and analyzed in the information sciences' curriculum.

References

Arcot – The Authentication Authority. Taking eFraud out of eBanking: Preventing Identity Fraud in Financial Services, 2007. http://www.cmp-customevents.com/arcot/, September 5, 2007.

Becker, David, Online Gaming's Cheating Heart. Cnet news, June 7, 2002.

Bruhn, M.; Gettes, M.; West, A. Identity and Access Management and Security in Higher Education, EDUCAUSE, 2003.

Chan, Michelle Jana. Identity in a Virtual World. http://edition.cnn.com/2007/TECH/06/07/virtual_identity/index.html, September 5, 2007.

Crawford, Susan P. Who's in Charge of Who I Am? Identity and Law Online. New York Law School Law Review, 49, 2004.

e-Banking, http://www.zaba.hr/info/pb/pc/ezaba_gradani.htm

e-Identity in education, http://www.aaiedu.hr

Gutierrez, Alvaro. Towards Better Digital Identity Management, 2006.

Javelin Strategy Research. One Surprising Finding: In Half of the Cases Where the Perpetrator is Known, Identity Fraud is committed by Someone Close to the Victim, 2003. http://www.bbb.org/ALERTS/article.asp?ID=565, September 5, 2007.

Nabeth, Thierry. Privacy in the Context of Digital Social Environments: A Cyber-Sociological Perspective. INSEAD CALT-FIDIS working paper, 2005a.

Nabeth, Thierry. Understanding the Identity Concept in the Context of Digital Social Environments. INSEAD CALT-FIDIS working paper, 2005b.

Tsui, Will. Digital Identity Management on the Internet. 2006.

Wiki. What is Digital Identity? http://en.wikipedia.org/wiki/Digital_identity, September 1 2007.

Yee, N., Bailenson, J. N. The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. Human Communication Research, 33, 2007.

Cultural and Scientific Transfer through Translation – a Corpus-Based Study of Term Formation in the Period 1848-1919

Špela Vintar
Department of Translation, Faculty of Arts
Aškerčeva 2, SI – 1000 Ljubljana
spela.vintar@guest.arnes.si

Summary

Contacts between cultures are a driving force of technological, scientific and linguistic development, where a culturally or economically more advanced region "feeds" its neighbouring regions. The Austro-Hungarian Empire was a multi-cultural environment where this transfer can be observed through – among other processes – translation. The study focuses on the development of Slovene technical and scientific terminology under the influence of German and is based upon a sample from a recently built digital library of translations from German into Slovene from the period 1848-1919, which was created within two nationally-funded projects in Austria and Slovenia.
The period between 1848 and 1919 was particularly crucial for the Slovene linguistic development, and German was at the time an influential language from which many concepts, expressions, phrases and names were borrowed. Apart from the level of the lexicon, significant changes can also be observed on other linguistic levels, from orthography, morphology, phonology, to syntax and pragmatics.
This study analyses several term formation processes such as borrowings, loan translations, term variation etc.
The corpus-based method allows us quantify the observed linguistic phenomena and compare some generic traits of texts from various domains, such as the type/token ratio or the ratio of borrowings compared to native vocabulary.

Key words: Term formation, Digital library, Corpus-based study

1. Introduction

Much of the world's knowledge is already available in electronic format, however the same is not true for cultural heritage, of which texts in various formats are a significant part. From the point of view of intercultural contacts, texts are

also powerful vehicles of cultural influence and thus also carriers of technical and scientific progress.

The field of digital libraries and new access modes to scientific and cultural heritage is developing rapidly on the EU level, with several large initiatives under way, e.g. **DELOS**[1] **"Network of Excellence on Digital Libraries"**, which conducts a joint program of activities aimed at integrating and coordinating the ongoing research activities of the major European teams working in Digital Library-related areas with the goal of developing the next generation Digital Library technologies, **ERPANET (Electronic Resource Preservation and Network)** (IST-2001-32706) which is establishing an expandable European Consortium, which will make viable and visible information, best practice and skills development in the area of digital preservation of cultural heritage and scientific objects. Both of these are also continuing the work of IST SM **DigiCULT**[2] **"Digital Culture"** (2002-2004), which established a regular technology watch for cultural and scientific heritage.

The aim of our study is to show that cultural and technical developments go hand in hand with linguistic development, whereby translation is in effect the bridge between the "feeder" and the "receiver" cultures. The Austro-Hungarian monarchy was extremely diverse in terms of national and linguistic structure, yet the processes of cultural transfer between Austria and Slovenia in the 19th century were by no means always fluent. Especially as far as literary production was concerned, the general opinion among the Slovene literary elite was that only original production was worthwhile and valuable, while translations – especially from German – would contaminate the language and jeopardize the Slovene cultural identity (Hladnik 1992, Orel 2005). This attitude was not completely mirrored in translations of technical and scientific texts from German into Slovene, particularly in the domains that were considered ideologically unproblematic (agriculture, housekeeping, medicine, natural sciences etc.).

As an attempt to create a digital resource that would enable both quantitative and qualitative research of phenomena related to translation, two parallel national projects were launched, namely: the Austrian Government funded **2004-2006** project "**German-Slovene/Croatian translation 1848-1918 (FWF P17465)**", with Graz Academy of Sciences as the coordinator, and the Slovene Government funded 2004-2007 project "**Slovene Translations of German Texts in the period 1848-1919 – Cultural and Linguistic Impacts (J6-6078)**", coordinated by University of Maribor (Teržan-Kopecky 2004). The result of the joint efforts of all project partners is the AHlib digital library consisting of the following parts:

[1] http://www.delos.info/

[2] http://www.digicult.info/

- TraDok – a comprehensive bibliography and database of Slovene, Croatian and other translations from German from the period 1848-1919, with their German counterparts, containing over 6,000 bibliographical units and equipped with a multi-function search interface[3],
- digitised and processed texts constituting the AHlib digital library, where each text has undergone scanning, OCR, manual correction, semi-automatic linguistic annotation (part-of-speech tagging and lemmatization), analysis of historical wordforms, and finally conversion into TEI (cf. Erjavec/Ogrin 2005).

Although part-of-speech tagging and lemmatization of contemporary Slovene are procedures for which reliable tools have been developed, linguistic annotation of archaic texts is an entirely different story. Many words are not recognized either because they are no longer in use or because they were spellt differently. Furthermore, orthographic variation was a common phenomenon both with general language words (eg. rujavo, rjavo, rjujavo [brown]) and specialized terms (operacija, operacia [operation]). The number of lemma types compared to word form types in a text can help us estimate variation in a text (see Figure x). Lemmatization of the AHlib digital library is being performed as an iterative process where each round of automatic lemmatization is followed by a manual correction phase. After each manual correction of unknown lemmata the lexicon for automatic annotation is expanded and the results consequently better.

Final versions of the above processing steps (in PDF, RTF and XML formats) are being uploaded into TraDok, while an online interface for the above conversion is available at the Jozef Stefan Institute[4].

Since the processing of the texts – especially the time-consuming manual correction phase – is still underway, the study presented in this paper is based upon a small sample from the target digital library, namely on 7 technical or scientific books from different domains and time periods. Another small corpus of literary works from the same period was compiled for the purposes of comparison. The listing of all books used can be found below.

2. Methods of corpus-based analysis

For the quantitative and qualitative analysis of the historical text collection we are using Wordsmith Tools (Scott 1998), a suite of programs and utilities for statistical text processing and concordancing.

As a measure of lexical density, which can give insight into the size of vocabulary used, the standardised type/token ratio (std. TTR) is used. This measure compares the number of different words (types) found in a corpus or text to the number of running words (tokens). The ratio between the two typically de-

[3] https://buedo22.uni-graz.at/pub/tradok/

[4] http://nl.ijs.si/ahlib

creases with corpus size, therefore comparison between different texts is only possible if samples of equal size are drawn from each text (standardised TTR). In all below experiments we use sample size of 5,000 words.
As an indicator of the complexity of vocabulary we can observe word length, more specifically the amount of extremely long words in a text – usually these words are found to be terms. An indicator of syntactical complexity can be average sentence length, although this may be also related to the text type.
By constructing word lists ordered alphabetically, by frequency or backwords we can analyse orthographic shifts, term variation and term formation from a single root. The example below shows a terminological nest around **daljnogled** (binocular), with the term variant **daljnovid**.

DALJNOGLED	19	1	14,29
DALJNOGLEDA	4	1	14,29
DALJNOGLEDE	9	1	14,29
DALJNOGLEDI	1	1	14,29
DALJNOGLEDNA	1	1	14,29
DALJNOGLEDOM	1	1	14,29
DALJNOVID	1	1	14,29
DALJNOVIDE	1	1	14,29

Comparing different word lists with the Keywords function helps us identify words that occur in a specific text with a higher relative freqnency than in a larger reference corpus. These words most often represent the terminological or technical inventory of the text, however they may include other domain- or author-specific words.

3. Text analysis

3.1 General corpus characteristics

As described above, two subcorpora were compiled for the purposes of this study, one consisting of 7 technical or scientific works spanning the years 1847-1908 (Tech), and the other containing 5 literary works from the same period (Lit). All books were translated into Slovene from German. The sizes of both subcorpora and some general statistics are given in Table 1.

	Tech	Lit
Size in bytes	1,251,569	737,815
Tokens (running words)	212,814	128,669
Types (different words)	25,759	19,385
St. type/token ratio	30,73	37,28
Av. sentence length	16	17
Av. word length	5,2	4,7

Table 1: General corpus characteristics

There appears to be a significant difference in TTR, with the corpus of literary texts exhibiting a higher lexical density than the corpus of technical texts. This could be an indicator of the fact that literary texts indeed use a richer vocabulary than technical texts, and indeed one must acknowledge that literary texts were normally translated by language professionals, either writers or dedicated translators. In contrast, technical texts were often translated by domain experts or clerical people.

Another variable that may influence TTR is the amount of orthographic variation. If the same word is spellt in several different ways, each form is counted as a separate word type. We do not however believe that this had a major impact on the observed difference between the corpora.

The slightly higher average word length in the Tech corpus is upon nearer examination the result of term usage.

3.2 Lexical density over time

In the rest of the study we are focusing mainly on lexical properties of non-literary texts in our sample. An intuitive hypothesis was that older texts would exhibit lower TTRs than newer ones because the term inventory of a language expands over time. There is however no empirical evidence of that, on the contrary – there is a notable tendency of decreasing TTR with texts produced at a later date (see Figure 1). Figure 2 shows the TTRs of literary texts for comparison.

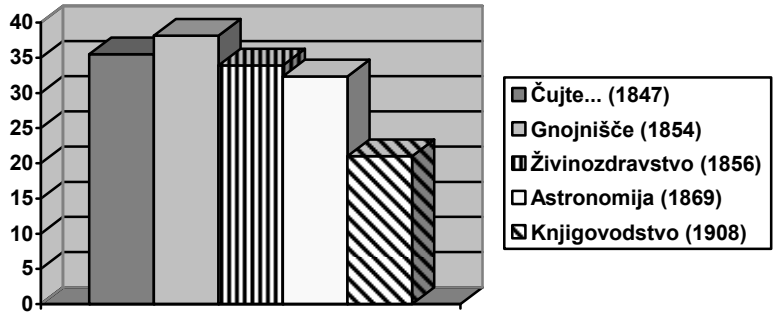


Figure 1: Type-token ratios of technical texts

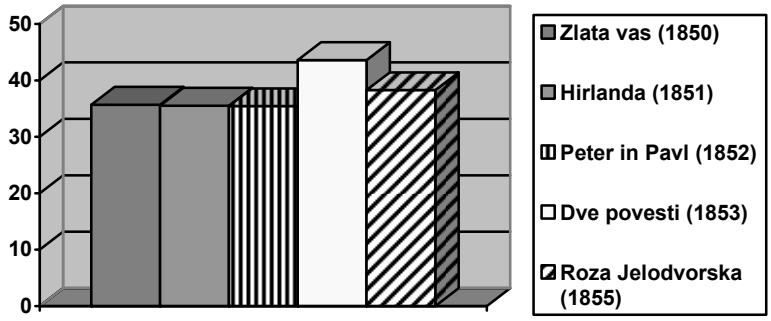


Figure 2: Type-token ratios of literary texts

3.3 Foreign citation as a sign of term pre-formation

A common practice still at work today is to introduce a new term in language accompanied by an original term in brackets. The citation of the original term serves as reference for the readers familiar with the source language, and on the other hand eases the translators dilemma in introducing a new expression. There are numerous examples of this practice in our text collection:

> *13. Pasja kuga ali mor* ***(Staupe)*** *207 14. Poglavje. Ploščnatna glista* ***(Bandwurm)*** *208 15. Poglavje. Gliste v želodcu in črevah 209 (Živinozdravstvo)*
> *Tako n. pr. zaznamavamo narazje nebeskih teles z zvezdno daljavo, z zemeljskimi poloméri; zemeljsko površje merimo z miljo, s protom* ***(Ruthe)****, s sežnjem, z metrom, in reči manjše raztege s čevljem, s palcem in črto. (Astronomija)*
> *Tudi ne sme konj imeti kraka* ***(Spath)****, ne podplatnih otisk* ***(Steingallen)****, ne nadkosti na zadnji strani skoknega člena* ***(Hasenhacke)****, ne pipe* ***(Piphacke)****, ne lupine* ***(Schale)****. Obedva jajca se morata dobro viditi. (Živinozdravstvo)*

It is clear from the above examples that the citations are used as a reference to another concept system from which the translated work is drawing knowledge, but we can also see the translators' efforts of using and using Slovene terminology whenever possible. Although citations occasionally reveal instability in term usage (see example below from Živinozdravstvo, where *Bandwurm* is translated first as *ploščnata glista* and second as *trakulja*) and are generally taken to be signs of an extremely passive attitude to term formation, the analysed works show quite the opposite.

...14. Poglavje. Ploščnatna glista ***(Bandwurm)****....*
....ampak le gosenica trakulje ***(Bandwurm).*** *.....*

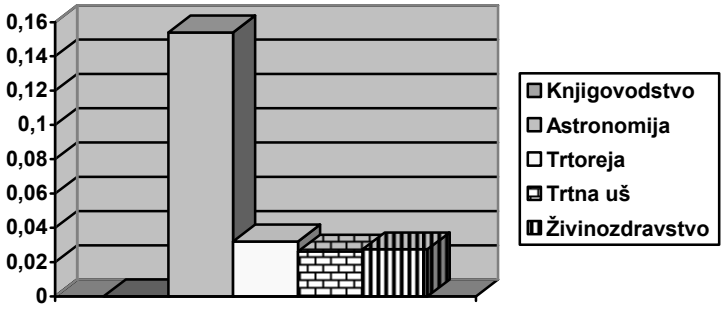


Figure 3: Percentage of German cited wordforms

Figure 3 shows the ratio of German cited words compared to the number of all types.

3.4 Term variation

As mentioned above, there is considerable variation in spelling in our text collection, and this naturally applies to term spelling too. Several variations are systematic and indicate an overall orthographic shift in Slovene, such as *serce - srce, smert – smrt; rudeč – rdeč, rujav – rjav; kteri – kateri, kedar – kadar*. Other variations typically occur with borrowed words (from German or classical languages) at phoneme boundary I and A: *salmiak – salmijak – salmjak, operacija – operacia – operacja*.

Cases of terminological inconsistency where several expressions are used to refer to the same concept are difficult to detect automatically, moreover we do not regard them as variations but as alternative representations of the same concept (e.g. *polutnik – ravnik – aequator – ekvator*; all synonyms from Astronomija). Yet another set of variations occur on the level of grammar, such as the ending –i in the locative case: *na obrežji* – [contemporary Slovene] *na obrežju*.

Since the process of lemmatization usually reduces these variants and assigns them a common lemma, we thought it would be interesting to compare the ratios of lemma types and word form types. The lemma/wordform ratio is thus the number of different lemmata found in a text divided by the number of different wordforms (see Figure 4).

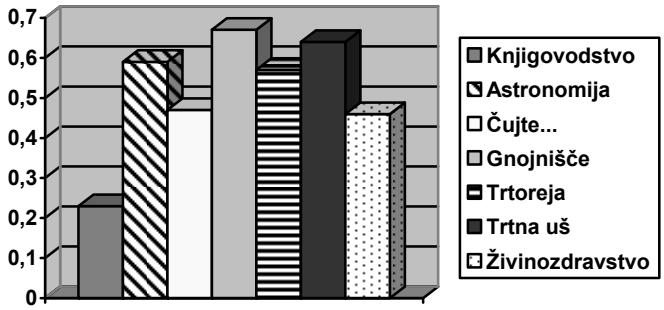


Figure 4: Lemma/wordform ratios

The underlying hypothesis is that texts with a higher lemma/wordform ratio actually contain more lexical diversity and less orthographic variation. Perhaps Knjigovodstvo should be excluded from this comparison because it contains many numbers that had been counted as different wordforms but all have the same lemma. In other texts the hypothesis might be confirmed, especially if we see that the lowest lemma/wordform ratios (apart from Knjigovodstvo) are the ones of the oldest books in the sample.

3.5 Loan translations

A highly productive term formation strategy is direct or loan translation, meaning that the original term is transferred into the target language by translating the original bits (Cabré 1998: 94). There are numerous examples of loan translations in our text collection, especially from Astronomija:

Brennpunkt – gorišče, ognjišče
Hundstage – pasji dnevi
Kegelschnitt – kegeljosečnica
Nachtbogen – ponočni lok

Upon detailed examination of the texts however we find that a more frequent strategy of translators was not to translate directly but to create genuinely Slovene expressions according to word formation principles. Whether neologisms were at that time recieved as reluctantly as today, we can only speculate, but as the general linguistic climate was very much against borrowings, especially from German, we believe that Slovene neologisms were definitely more welcome than cited foreign words.

4. Conclusions

The study explored some aspects of term formation in Slovene in the period 1848-1919. Some intuitive hypotheses have not been confirmed by corpus evidence, for example that the type-token ratio would grow with time as a reflec-

tion of the expanding vocabulary. It seems that the vocabulary expansion that had certainly taken place through technological and scientific development was on the other hand balanced by more orthographic standardization and less variation. A comparison of technical and literary texts showed that lexical density was slightly higher in the latter.

Several volumes we explored showed a high percentage of cited German expressions. These were however rarely used as borrowings without any attempt of introducing a Slovene term, rather they were used as additional reference points for the informed reader who may not yet be familiar with Slovene terminology. On the whole, translation was in the selected time period still regarded as a creative process of restructuring and adapting the text to the target audience, while the principle of equivalence only gained importance from 20th century onwards.

A more extensive and detailed study of the above phenomena will be possible when the AHlib digital library is finished, and several studies using AHlib materials are already underway. However, to gain a thorough insight into linguistic and cultural processes at work within the chosed time frame, one should have available a comparable digital library of Slovene original production as well as German originals of the analysed works.

To conclude in a visionary tone, digital libraries of the future will offer entirely new access modes to the knowledge and cultural achievements of past generations, and we may look forward to the applications still to be developed in this field.

References

Cabre, T. (1998). Terminology: Theory, methods and applications. Amsterdam/Philadelphia, PA:John Benjamins.

Erjavec, Tomaž and Matija Ogrin (2005). Digitalisation of literary heritage using open standards. In Paul Cunningham, Miriam Cunningham (eds.). Innovation and knowledge economy: issues, applications, case studies, (Information and communication technologies and the knowledge economy). Amsterdam [etc.]: IOS Press, 2005, str. 999-1006.

Honzak-Jahić, Jasna (2002) O vplivu češkega in nemškega jezika na razvoj slovenskega strokovnega izrazja. In: Slavica comparativa, Prague: Euroslavica, 2002, pp. [33]-39.

Hladnik, Miran (1992) "Vloga prevoda v slovensko-nemški literarni tekmi", in: Hladnik, Miran / Pečaj-Rus, Darinka eds. XXVIII. seminar slovenskega jezika, literature in kulture. Zbornik predavanj. Ljubljana: Center za slovenscino kot drugi tuji jezik pri Oddelku za slovanske jezike in knjizevnosti Filozofske fakultete, 109-119 [http://www.ijs.si/lit/prevodsl.html-l2].

Orel, Irena (ed.) (2005) Razvoj slovenskega strokovnega jezika. Proceedings of an international symposium, Ljubljana, 17-19 November 2005, Ljubljana: Center za slovenščino.

Scott, Mike (1998) WordSmith Tools Version 3. Oxford University Press, Oxford, England.

Teržan-Kopecky, Karmen.(2004) Slovenski prevodi nemških besedil v obdobju 1848-1918 : jezikovni in kulturni vplivi : vloga za (so)financiranje raziskovalnega projekta za leto 2004 = Slovene translations of German texts in the period 1848-1918 : linguistic and cultural impact : (research proposal). Maribor: Pedagoška fakulteta, 2004.

Bibliography of the sample corpus

Gasteiner, Josip (1908) *Knjigovodstvo za dvorazredne trgovske šole* [übers. v. Volc Ivan]. Ljubljana: Slov. trgovsko društvo “Merkur”.

Goethe, Arminij (1881) *Trtna uš. Poljudno razlaganje o tem kakšne lastnosti ima in kako živi ovi najnevarnejši sovražnik vinske trte in kaj moramo storiti v obrambo zoper ta mrčes*. V Gradci: Štajersko društvo za omikanje ljudstva.

Goethe, Hermann (1891) *Iz biologičnega trtorejskega poskušališča od Hermana Goethe-ja v Badenu pri Dunaju* [übers. v. C. kr. kmet. društvo v Gorici]. Gorica: Tisk. Paternolli.

N.N. (1852) *Peter in Pavl, ali Bóg ubózih sirót najboljši oče. Povest za otroke in mladénče, za odrašene, kakor tudi za starše in učitelje*. V Celovcu: Pri založniku Joanu Leonu, bukvarju.

Rohlwes, Johann Nikolaus (1856) *Domače živinozdravstvo v boleznih konj, govedja, ovac, prešičev, koz in psov, ali nauk, kako mora kmetovavec svojo živino rediti, ji streči, jo kermiti in ozdravljati* [übers. v. Robida Karl]. V Celovcu: Janez Leon.

Schödler, Friedrich Karl Ludwig (1869) “Astronomija” [übers. v. Ogrinec Vil.[jem]], in: *Knjiga prirode. I. del. Fizika, Astronomija in Kemija*. Band: 1, v Ljubljani: Matica Slovenska, 229-339.

Schmid, Christoph (^{2}1853) *Dve povesti iz pisem Kristofa Šmida. A. Golobčik. B. Kanarčik* [übers. v. P.[intar] A.[nton]]. V Ljubljani: Jožef Blaznik.

Schmid, Christoph (1855) *Roza Jelodvorska. Lepa podučivna in kratkočasna pripovest za mlade in odrašene ljudi* [übers. v. Pijelik Dragoslav]. V Ljubljani: Henrik Ničman.

Šmid, Krištof (1851) *Hirlanda bretanjska vojvodnja ali zmaga čednosti in nedolžnosti. Nauka polna povest za starost in mladost*. Ljubljana: J. Giontini.

Wimmer, Anton (1854) *Gnojnišče kmetovavca zlati rudnik* [übers. v. Bleiweis Janez]. Ljubljana: c.k. kmetijska družba.

Zschokke, Heinrich (1847) *Čujte, čujte, kaj žganje dela! Prigodba, žalostna ino vesela za Slovence* [übers. v. Globočnik Felicijan]. V Celovci: Joan Leon.

Zschokke, Heinrich (1848) *Zlata Vas. Podučna in kratkočasna povest* [übers. v. Malavašič Fran]. Ljubljana: Natisnil Jožef Blaznik.

Sentence Alignment as the Basis for Translation Memory Database

Sanja Seljan
Department of Information Sciences, Faculty of Humanities and Social Sciences
Ivana Lučića 3, 10 000 Zagreb, Croatia
sanja.seljan@ffzg.hr

Angelina Gašpar
SOA Centre Split
Katalinića prilaz 2, 21000 Split
ginasplit@yahoo.com

Damir Pavuna
Integra d.o.o.
A. Stipančića 18, 10000 Zagreb
damir.pavuna@integra.hr

Summary

Sentence alignment represents the basis for computer-assisted translation (CAT), terminology management, term extraction, word alignment and cross-linguistic information retrieval. Created out of the sentence alignment process, translation memory (TM) represents the basis for further research in translation equivalencies. Automatic sentence alignment, based on parallel texts, faces two types of problems: robustness and discrepancies between source and target texts in layout and omissions which have an influence on the accuracy of the alignment process.
The aim of the paper is to present research on the sentence alignment process carried out on the Croatian-English parallel texts (laws, regulations, acts and decisions) and implemented by the alignment tool WinAlign 7.5.0 by SDL Trados 2006 Professional.
The alignment process and its impact on the creation of translation memories is presented through comparison of translation memories that differ regarding the levels of expert intervention in the set up of the alignment program and preparation of the source text for the segmentation. Recommendations for further development using statistical analysis, automatic learning techniques and language knowledge are suggested.

Key words: sentence, alignment, translation memory, computer-assisted translation (CAT), tool, segmentation, set up

Introduction

The need for fast translation of a large number of pages in several languages, use of specialized and consistent terminology, sharing of common resources, time-saving, cooperation in larger translation projects and cost-saving, have caused a growing use of translation memories.

Sentence alignment represents the basis for computer-assisted translation (CAT), terminology management, term extraction, word alignment, cross-linguistic information retrieval, etc. Created out of sentence alignment process, translation memory (TM) represents the basis for further research in translation equivalencies.

Witnessing the importance of the sentence alignment process, different international projects have been undertaken in order to develop its evaluation metrics: the two-year project ARCADE (1995-96) aiming to produce a bilingual French-English corpus suited for the alignment task and its evaluation; MULTEXT-East Project (Multilingual Text Tools and Corpora for Eastern and Central European Languages) where each of the six translations of the novel 1984 by G. Orwell were sentence aligned with the English original, and the alignments hand validated; the Egypt Statistical Machine Translation Toolkit (1999) and the following GIZA++ for training statistical translation models.

In this paper, the sentence alignment problem is elaborated through several aspects. The main reasons are presented for the use and development of translation memories created out of the alignment process and their integration into translator's workbench.

As translation memories (TMs) work best on the voluminous and highly repetitive types of texts (e.g. new versions of software or products, regulations/laws, decisions, catalogues, manuals) the research was done on Croatian-English parallel legislative texts. Therefore, the research was done on a highly structured type of text (legislation) whose main characteristics are presented (structure, enacting formulas, specific terms and expressions) that can influence the alignment process.

The results of the alignment processes are presented through bitexts, created out of Croatian-English parallel texts, which are imported into translation memories. As various approaches in the alignment process are used, different types of TMs are elaborated and tested. In the conclusion, recommendations for further development are suggested.

The research presented here is an outcome of the research project "Information Technology in Computer-Assisted Translation of Croatian and in e-Language Learning" (130-1300646-0909) undertaken with the support of the Ministry of Science, Education and Sport of the Republic of Croatia.

When to use TMs?

Need for fast and consistent translations are obvious in the EU and for candidate member states when a large number of legislative documents are to be trans-

lated, but also in multilingual societies, multinational companies, in government institutions and agencies, or when translating simultaneously multilingual documentation for new versions of products and services in several languages. Translation memories are used to speed up the translation process, enable the sharing of resources, consistent terminology and cost-reduction.
In the process of creation of TMs, there are two possible ways:

- When the translator is giving input through the source text which has to be translated, the program scans the text trying to find matching pairs as full match or fuzzy matches, which are then subject for the translator's review. The new segments are then stored in the database, i.e. translation memory and can be used in future work.
- Another way is the building of TMs out of already translated material through the alignment process, as presented in this case study.

Translation memories are usually integrated with other CAT (Computer-Assisted Translation) tools (e.g. word processing programs, multilingual dictionaries, thesauri, terminology management system, machine translation software) into translator's workbench. Local translation memories can be integrated into the central TM stored on the central server, which is usually part of the global management system.
Translation memories are mostly used when a translator has a feeling of "having already translated something like this". This is where the TM has the best effect: it offers the same translation (100% match) or similar translations (fuzzy matches) using already translated units. Besides this, TMs are a valuable resource for concordance search to determine the appropriate use of the term or as a terminology management source providing specialized terminology.

Corpus used

Translation memories are mostly used for translation of highly structured and voluminous documents. The use of controlled language, specific forms and structures can augment results obtained using CAT tools.
As the research is done on a highly structured type of text (legislation) the main characteristics are presented regarding its structure, enacting formulas, specific terms and expressions that can influence the alignment process. The results of the alignment processes are presented through created bitexts of Croatian-English texts that are imported into translation memories.
The alignment is carried out on parallel texts, consisting of corresponding texts between Croatian and English. The analysis is based on Croatian legislative acts: laws, regulations, decisions, and ordinances (NN122/03, NN51/04, NN49/03, NN30/97, NN10/02, NN164/04, and NN 173/03) related to competition acts, crafts act, trademarks, electronic signature, agreement of minor importance, of relevant market, on concentrations and bylaws. The bilingual corpus was examined regarding its structural and lexical levels since various acts have their own standard presentation and standard formulas. They are set out in

the Style Guide for Croatian Legislation inspired by the English Style Guide from EC DGT (2005) and the Interinstitutional Style Guide (2005). The percentage ratio for word count in English translations is 33.15%, which is due to the fact that the English language is an analytic type of language, contrary to the Croatian language, which uses a highly flective system. This is presented in Table 1, which compares the number of words in Croatian and English texts.

Table 1: Size of parallel texts

	Words	**Characters** (with spaces)	**Pages** (1,800 char./page)
Croatian	**39,956.00**	274,198.00	**152.32**
English	**53,203.00**	330,513.00	**176.29**
Total	93,159.00	604,711.00	**328.61**

Standard structure and formulas

The drafting style takes account of the type of act for the sake of uniformity and terminological consistency. However, every act is specific for its own repetitive legal terms, phrases and sentences. For instance, the following main components of a regulation are: the title, preamble (citations, recitals, enacting formulas which can be obligations, permissions, admissions and statements), enacting terms, addressee, place, date and signature.

Acts with a simple structure comprise articles and subdivisions of articles. The arrangement of enacting terms in Croatian acts is the following: Part (Division, Title, Chapter, Section, Subsection, Heading, Subheading) and Article (Paragraph, Subparagraph, Point/Item, Indent, Annex, Appendix; Schedule). Also, the textual components of the enacting terms comply with relatively strict rules of presentation: the subject matter and scope, the definitions, the provisions conferring implementing power, provisions concerning penalties or legal remedies, transitional and final provisions. Each article contains a single provision or rule laid down in an act. Furthermore, the standard form prescribes the layout on the page, including spacing, paragraphing, punctuation and even typographic characteristics (capitalisation, typeface, boldface and italics).

Use of verbs in enacting terms

The enacting terms of binding Croatian legislation can be divided into imperative and declarative terms. While declarative terms refer to definitions or amendments, in order to express commands and prohibitions, the Croatian language often uses the present tense whereas 'shall' is used in English translations. For instance, /Sudionici koncentracije obvezni su podnijeti…/ is translated into English as / The parties to the concentration shall be obliged to submit…/or /Zabranjeni su svi sporazumi…/ /There shall be prohibited all agreements/. Also, modals in Croatian legislation such as 'morati' or 'trebati' are also translated into English as 'shall' or 'must' meaning 'is required to'; for instance, /Informacijski sustav …treba biti oblikovan tako..'/ The information system

shall be organized in such way...'/ or /'davatelj usluga certificiranja….mora čuvati svu dokumentaciju'/ /'Certification authorities…. must safeguard all documentation'/. Prohibition is expressed in Croatian legislation by terms 'ne može se' or 'nije dopušteno', which have an English equivalent in the term 'may not'. Also, permission in the Croatian language is expressed by the terms 'može se', 'smije se' or 'dopušteno je', which have the English equivalents 'may' or 'it is admissible'. The terms such as 'it is permitted' or 'it is allowed' are not used in English legislation. For instance, / 'Agencija može…odobriti produženje roka'/ The Agency may ...extend the period...'/ or /'Iznimno od odredbe…dopušteno je članstvo...'/ 'Without prejudice to the provision…, it is allowed to be a member...'/; /'Dopuštena je svaka gospodarska djelatnost...'/ 'Every economic activity ...is permitted'/.

Authorisation is expressed by two terms in Croatian: 'ovlašten je' or 'može se', whereas English legislation uses several terms such as :'may', 'is authorised to' , 'is empowered to' , 'has the power to' or 'shall', in case the authorisation is binding or comprises certain activity. For example, /'Predsjednik i članovi Vijeća mogu pisati...'/ /'The President and the members of the Council are authorised to write...'/.

As far as expressing the rights is concerned, Croatian legislation uses indicative expression 'imati pravo' which corresponds to the English expressions 'has the right to', 'is entitled to' or 'may'. Definitions of legal terms are binding and are expressed with the present tense in Croatian but still translated as 'shall' in the EU, although the Anglo-American legislative has recently started using present. /'Pojedini izrazi koji se rabe u ovom zakonu imaju sljedeće značenje:'/ /'Individual terms in this Act shall have the following definitions:'/

Verbs expressing descriptive functions or statements such as 'biti', 'postojati', 'nalaziti se' and 'imati' are used in the present tense in Croatian, which correspond to English 'There shall be' or rather 'There is hereby established'. For instance, /'Upisnik je knjiga koja sadrži podatke i isprave...'/ /'The Register is a book containing the data and documents...'/.

Capital letters

In Croatian language only the first noun in names of institutions, administrative bodies, laws, geographical names, agreements etc. is capitalised whereas in English all nouns and adjectives are written in capitals. For instance, /Ministarstvo obrazovanja i sporta/ Ministry of Education and Sports, / Zakon o zaštiti tržišnog natjecanja/ The Competition Act, / Madridski sporazum/ the Madrid Agreement. All titles in both source texts and their translations are capitalised.

Hyphens and compound words

The hyphens used in English compound words, for instance, 'a five-year-term' / 'razdoblje od pet godina' or 'local self-government' / 'jedinica lokalne samo-

uprave', are omitted in the source text. It is to stress that some compound words in the Croatian language do not have the same word order as in English translation, e.g. / 'jedinice lokalne i regionalne samouprave' / 'bodies of regional government and local self-government'/.

Punctuation

The punctuation is almost the same in both the source texts and the translations. A semicolon is mostly used instead of a linking conjunction and to separate intends. The comma is used to divide adjectives in series but also before 'and' and in parenthetic and introductory phrases. For instance, / 'Usmena rasprava je, u pravilu, javna.'/ 'The oral hearing is, as a rule, public.' /. Round brackets or parentheses are used when citing numbered paragraphs in English, for instance, /'Article 11 paragraph (1) item1 '/, but in source texts they are omitted, /'…članak 11.stavka1.točka 1.

Numbers

Numbers up to 10, but also larger numbers, are written either in words or in figures in both the source texts and the translations, for instance, /....'najmanje jednu milijardu kuna'/ corresponds to English /'1 billion Kuna/ or /'..najmanje 100.000.000,00 kuna'/ ...at least 100,000,000.00 Kuna. Although in plural and lowercased, the symbol for Croatian money 'kuna' is capitalised and written in singular, 'Kuna', in translations.

Dates

Dates use figures for days and words for months both in the source texts and the translations. However, the usage is different: / '21.srpnja 2003.'/ '21 July 2003' / or '7.21.03' in the American dating system, or '2003-07-21' in the international one.

Foreign words, expressions and synonyms

Some foreign Latin words and expressions can be found in both the source texts and the translations, and also English words and expressions in the source texts (e.g. *ex officio, know- how, joint venture, world-wide, franchising*). Synonyms used in the source texts are mostly avoided in the translations, for instance, /'pripajanjem ili spajanjem poduzetnika'/ 'merger association of undertakings'/ or /'Stjecanje dionica ili udjela'.../Acquisition of Shares'.../ /'grafički prikaz ili dijagram' / 'the graphic presentation (diagram)'/.

Both the source texts and the translations have consistent terminology. Defined terms are used in a uniform manner in order to facilitate comprehension and interpretation of legislative acts. Gender-neutral language is preferable. While the Croatian language uses the active voice more frequently, the English language makes more use of passive. Sentences in the active voice are generally, though not always, clearer and more concise than those in the passive voice because

fewer words are required to express action. Unlike English, Croatian language has a very rich case system whose nouns, pronouns and adjectives are inflected by the case. It is due to this variety that the cases bear the main burden in marking the syntactic functions of a noun phrase and that word order is relatively free.

Research

The alignment was carried out on Croatian legislative acts: laws, regulations, decisions, and ordinances and their respective English translations (a total of 328.61 pages). After comparing the number of words and pages, automatic alignment was carried out with the total number of translation units. The comparison between different types of alignment models is presented together with their impact on the creation of translation memories. The results between four types of translation memories are elaborated distinguishing the level of expert intervention in the setting up of segmentation parameters and source document segmentation.

Tools used

In the research we used SDL Trados 2006 Professional, part of which is the alignment tool WinAlign 7.5.0. For document structure analysis we used a very common tool for such purposes, AnyCount 4.0 (version 405). Bitexts in the .txt form are exported out of the WinAlign tool and imported into SDL TRADOS Translator's Workbench 7.5.0. Translation memory is then saved in SDL Trados native .TMX format, which is standard and convertible to almost all recommended format.

Activities

In the field of language technologies, the term bitext is used, denoting a merged document consisting of the source and target texts, generated by the alignment tool. A collection of *bitexts* is called bitext database or bilingual corpus. The main difference between bitext and translation memory is that matched segments are stored in the way that is unrelated to the original, with lost sentence order, while the bitext holds up the original sentence order.

As translators often have at their disposal a considerable amount of translated material, it can be aligned and converted into a TM database, although certain preparatory activities should be taken:

- comparison of the source and target texts (whether all text is translated)
- defining set up of end and skip rules (delimiters, creating abbreviation user list)
- preparation of the source text for better segmentation (spelling, automatic bullets and numbering, deleting of soft returns, hyphens, certain punctuation, tables created with tabs and revision marks)

- modification of set up rules
- verification of the alignment (especially 1:2 and 2:1 pairs and commitment of pairs)
- creation of translation memory and verification.

Automatic alignment

WinAlign has language independent algorithms that count:

- the quality of translation units which can have tree levels (low, medium, high)
- translation units aligning 1:2 or 2:1 pairs
- unconnected target segments.

In the case study, nine legislative documents in parallel Croatian and English languages were aligned. Statistics from Table 2 represent the results of the automatic alignment using language independents algorithms.

Table 2. Automatic alignment

Target File Name:	1	2	3	4	5	6	7	8	9	**Total**	**Percent**
N. of source segments:	530	525	262	82	107	218	249	132	484	2,589	104.73
N. of target segments:	504	512	290	89	107	224	246	129	482	2,583	104.49
N. of aligned units:	503	504	255	77	104	211	230	124	464	**2,472**	
N. of com-mitted units:	0	0	0	0	0	0	0	0	0	0	0.00
N. of high quality units:	259	193	71	25	34	51	57	43	226	959	**38.79**
No. of medi-um quality u.:	217	297	152	45	67	149	153	73	222	1,375	**55.62**
N. of low quality units:	27	14	32	7	3	11	20	8	16	138	**5.58**
N. unconnec-ted source s.:	8	0	1	0	0	0	0	0	2	11	**0.44**
N. unconnec-ted target seg:	0	0	18	2	0	0	1	0	0	21	**0.85**
N. of 1:2 and 2:1 units:	20	29	23	15	6	20	34	13	36	196	**7.93**
										2,472	**Aligned**

152.32 pages of the Croatian text were aligned with 176.29 pages of the English text, creating automatically all together 2,472.00 translation units (alignment performed 104.73%) and 11 units that were not aligned (0,44%). Out of the total number of aligned units, 38,79% (959) were marked as high quality units, 55,62% (1.375) as medium quality units, and 5,58% (138) as low quality units. 7,93% (196) were marked as 1:2 and 2:1 aligned units.

From the figures presented, it can be seen that every alignment should be verified, manually corrected and the whole process supervised. Part of the problems relate to different layout of texts, omissions, inversions, different structure orders and paragraph numbering. Therefore, expert intervention in the set up of the alignment program and pre-editing activities of the source text for better segmentation should be included. That way, improper segmentation would be reduced, since automatic marks would be hidden, alignment would be carried

out relating to the text and the number of high quality translation units augmented.

Automatic and manual alignment

The significant difference between alignment processes made with and without expert interference is presented in Table 3. The first column shows automatic alignment without a language expert. As WinAlign uses language independent algorithms, it is estimated that the automatic alignment process found 5.58% of low quality units, 55.62% of medium quality units and 38.79% of high quality units.

Table 3. Alignment: automatic vs. manual

	Automatic alignment		**Difference**	**Manual alignment**	
Text 8 (Bylaws)	No.	%		No.	%
N. of source segments:	132	104.73	**?**	120	100.00
N. of target segments:	129	104.49	**?**	125	104.17
N. of aligned units:	124		**?**	120	
N. of committed units:	0	0.00	**High**	120	**100.00**
High quality units:	43	**38.79**	**High**	0	0.00
Medium quality units:	73	**55.62**	**High**	0	0.00
Low quality units:	8	**5.58**	**High**	0	0.00
Unconnected source s.:	0	0.44	**OK**	0	0.00
Unconnected target s.:	0	0.85	**OK**	0	0.00
N. of 1:2 and 2:1 units:	**13**	7.93	**High**	**5**	4.17
Aligned:				120	

After the manual alignment, all sentence pairs are marked as committed units (as presented in 'Manual alignment' columns), out of which the TM base can be created. Another problem are 1:2 and 2:1 units. As WinAlign does not have any language algorithm and does not care which are the source and target languages, in this study there are 13 cases (7.93%) marked as 1:2 or 2:1 pairs. Out of the 13 suggested cases suggested in automatic alignment, 8 were wrong, and the total number of such cases is 5, as stated in manual alignment.

Comparison of TMs

Table 4 presents four types of translation memories:

- TM created out of the automatic alignment using SDL Trados language independent engine with null expert intervention (Raw TM)
- TM created out of the alignment but without any expert intervention in the set up of the alignment program and without intervention on the text segmentation, stating only that the source text corresponds to translated target segment (Aligned TM)

- TM created out of the automatic alignment, with expert intervention in the set up of the alignment program (Aligned TM + Set up rules, e.g. segment and skip rules, abbreviation user list)
- TM created out of the manually confirmed alignment, including setting up of segmentation rules in the alignment program and expert intervention on the segmentation of the source text (e.g. changes of soft returns, check of colon segmentation)

The presented translation memories differ regarding expert intervention in the WinAlign set up and in the segmentation of the source text. When translating the same text, out of which the TM has been created, it is to be expected that this automatic translation would completely match the created TM, and that machine translation would match 100%. But this is not the case, especially when TMs are created automatically using a language independent engine and without expert intervention in the setting up of segmentation rules and revision of the segments in the source text.

For the purpose of this study the following changes were made in the set up of the alignment program: ":" is not considered as delimiter, "br." is not considered as delimiter but moved to the abbreviation user list, "I." is also not considered as delimiter but moved to the abbreviation user list.

In the process of segmentation of the source text, soft return was eliminated and ":" deleted.

Therefore, the presented evaluation was made on the automatic translation of the same text, out of which different types of TMs had already been made with the difference regarding expert intervention in the process of creation of TMs.

Results

The automatic alignment presented in the first column was carried out by the language independent engine and with null expert intervention (Raw TM). Although the results seem very good (the same text translated with 91.67%) thanks to highly structured texts, alignments are very imprecise and wrong, often without sense and linguistically incorrect. The TM created out of this alignment contains 61.2% of medium and low quality units (see Table 3), suggesting that more than every other segment is not properly aligned. Therefore, the generated translation would be unclear and useful only for experimental purposes.

Although the results using aligned TM without any expert intervention (Aligned TM) seem much worse (80.30%), all translated text is linguistically correct. The first column, in spite of results, can not be compared with other three columns since their translated segments are linguistically correct and correspond to each other, which is not the case with Raw TM presented in the first column.

With the setting up of segmentation parameters (e.g. segment end and skip rules, creating abbreviation user list), the result is much better (88.89% text translated).

Table 4. Alignment: automatic, manual

	Raw TM	**Aligned TM**	**Aligned TM**	**Aligned TM**
			Set up rules	**Set up rules**
Segments/				**Segmented source docum.**
Context TM	0	0	0	0
Repetitions	0	0	0	0
100%	121	106	112	**120**
95% - 99%	0	0	0	0
85% - 94%	2	5	0	0
75% - 84%	2	2	1	0
50% - 74%	1	1	2	0
No Match	6	18	11	0
Total	132	132	126	120
Percent	**91.67%**	**80.30%**	**88.89%**	**100.00%**

Ultimately the 100% translated text could not be produced without the setting up of segmentation parameters and without the preparation of the source text for segmentation, as in the last case when an expert used the tool to see hidden characters and made final changes (in practice it is advisable to do it at the beginning of the whole alignment process) in the source text and according to the experience in the alignment process (e.g. the changing of soft returns, checking colon segmentation).

Conclusion

Sentence alignment is prerequisite for further corpus processing and research in the fields of computer-assisted translation (CAT), terminology management, term extraction, word alignment and cross-linguistic information retrieval. In the process of sentence alignment two main types of problems are considered: robustness, differences in layout (omissions, inversion, 1:2 or 2:1 alignments) between the source and target texts, and the segmentation of the source text in order to achieve better accuracy and to create a translation memory of good quality.

The standard and uniform manner of legislative texts, prescribing the layout (space, paragraphs, punctuation and capitalisation) and relatively strict rules should facilitate sentence alignment, although expert intervention is necessary, as presented in comparison of different translation memories.

The translation memories created in this study out of different types of the alignment processes give different results regarding the quality of the translated material. The results show necessary interventions of an expert when defining the set up rules, in preparation activities for the source text segmentation and in the verification of suggested translation units.

A good quality translation memory created out of such an alignment process then becomes a valuable source for further research in translation equivalencies, terminology extraction, terminology management, word alignment or cross-linguistic information retrieval. Although the results are augmented, they are static

and could be improved with the integration of language knowledge, rule-based algorithms and further research in statistical alignment. Integrated with other translation tools, TMs can be a useful tool to increase the speed and augment terminology consistency in the translation process, but with a human as the main supervisor.

References

Arcade: Evaluation of parallel text alignment system. (http://www.up.univ-mrs.fr/veronis/arcade/arcade1/index-en.html, 01st September 2007)

Barzilay, Regina ; Elhadad, Noemie. Sentence alignment for Monolingual Corpora. Proceedings of the Conference on Empirical Methods in NLP - EMNLP, 2003, Japan.

Ceauşu, Alexandru; Ştefănescu, Dan; Tufiş, Dan. Acquis Communautaire Sentence Alignment using Support Vector Machines, LREC 2006, Italy.

Corpus Markup. COP Project 106 MULTEXT-East: Work Package WP2 - Task 2.3, 1997. (http://nl.ijs.si/ME/CD/docs/mte-d23f/node2.html, 25 August 2007.)

del Pino, Santiago. Using Translation Memory Software (TMS): An Organisational Checklist. Terminologie et Traduction, pp. 132-139.

DGT of the EC. Translation Tools and Workflow, 2005. http://europa.eu.int/comm/dgs/ translation/bookshelf,/toolsandworkflowen.pdf

EC DGT. English Style Guide: A handbook for authors and translators in the European Commision, 2005 (http://ec.europa.eu/translation/writing/style_guides/english/style_guide_en.pdf)

Egypt Statistical Machine Translation Toolkit, 1999. (http://www.clsp.jhu.edu/ws99/projects/mt/toolkit/, 25 September 2007)

GIZA++: Training of Statistical Translation Models. http://www-i6.informatik.rwth-aachen.de/Colleagues/och/software/GIZA++.html, 25 August 2007)

Krstič, Adriana. Translation of EU legislation in Slovenia. Proceedings of the 5th EAMT Workshop, 2000, Slovenia.

Knauf, Ansgar. TR16: Development, Use and Profitability of Transaltion memory Systems. Translation Issues, 4-99. December 1999.

LISA: 2004 Translation Memory Survey (http://www.lisa.org/products/surveys/tm04survey.html)

LISA: The Localization Industry Standards Association. (http://www.lisa.org)

Marrafa, Palmira ; Ribiero, António. Quantitative Evaluation of Machine Translation Systems : Sentence Level. Workshop on Example-Based Machine Translation, Proceedings of the MT Summit VIII conference, 2001, Santiago de Compostela.

Och, Franz Josef; Ney, Hermann. Statistical Machine Translation. Proceedings of the 5th EAMT Workshop, Ljubljana, 2000.

Office for Official Publications of the European Communities:Interinstitutional Style Guide, 2004 (http://eur-lex.europa.eu/en/techleg/pdf/en.pdf)

Planas, Emmanuel. Extending Translation Memories. Proceedings of the 5th EAMT Workshop, Ljubljana, 2000.

Priručnik za prevođenje pravnih propisa RH na engleski jezik, MVPEI, Zagreb, 2006. (http://www.mvpei.hr/ei/download/2007/02/26/prirucnik_za_prevodenje_pravnih_propisa_RH.pdf)

Seljan, Sanja; Pavuna, Damir. Translation Memory Database in the Translation Process // Proceedings of the 17th International Conference on Information and Intelligent Systems IIS 2006, Varaždin : FOI.

Simard, Michel; Plamondon, Pierre. Bilingual Sentence Alignment: Balancing Robustness and Accuracy, 1998.

Steinberger, Ralf ; Pouliquen, Bruno ; Widiger, Anna ; Ignat, Camelia ; Erjavec, Tomaž; Tufiş, Dan; Varga, Dániel. The JRC-Acquis : A Multilingual Aligned Parallel Corpus with 20+

Languages. Proceedings of the 5th International Conference on Language Resources and Evaluation LREC, Genoa, Italy, 24-26 May 2006.

Wiki http://en.wikipedia.org/wiki/Translation_memory (2nd September 2007)

Zetzsche, Jost. Translation Memories: The Discovery of Assets: Recognizing opportunities and overcoming obstacles to TM sharing. MultiLingual Computing & Technology 72, vol. 16 (4), 2005.

Retrieving Information in Croatian: Building a Simple and Effcient Rule-based Stemmer

Nikola Ljubešić
Faculty of Humanities and Social Sciences,
Department of Information Sciences
Ivana Lučića 3, 10 000 Zagreb
nljubesi@ffzg.hr

Damir Boras
Faculty of Humanities and Social Sciences,
Department of Information Sciences
Ivana Lučića 3, 10 000 Zagreb
dboras@ffzg.hr

Ozren Kubelka
Vern' Polytechnic
Trg bana Josipa Jelačića 3
ozren.kubelka@vern.hr

Summary

Since Croatian is a highly flective language there is a need for morphological normalization of natural language information so that information could become retrievable in a more efficient way. Although this topic has been researched for more than two decades in Croatia, the vast majority of information systems that store information written in Croatian still do not have this problem solved. The primary cause for this situation is the high price of existing systems. The aim of this paper is to analyze the current situation in the industry regarding this problem and to build a rule-based stemmer which would consist of a minimal set of rules for expanding queries to the whole possible paradigm. Such a system could make expensive morphological databases in information retrieval obsolete.
We used a corpus sample, a morphological lexicon and a query sample of 1.000 most frequent nouns in base form to build a rule-based stemmer optimized through the steepest ascent hill climbing algorithm. Using this method we built a stemmer which performs almost equally good as the noun lexicon with F1 measures of 97.82% without the rules for adjectives and 97.64% with them.

Key words: Information retrieval, Croatian language, rule-based stemming, hill climbing optimization, industry awareness

Introduction

The aim of this paper is twofold – first, to analyze the practice in the IT industry in Croatia concerning the problem of Croatian morphology in information retrieval, and second, to construct a simple and efficient stemmer which could be easily used in the industry.

As far as we know, up to this point there were two stemmers built for retrieving information in Croatian. The first one was built by Tomislava Lauc [1]. She wrote rules for noun and adjective paradigms as well as for their morphonological alternations in Croatian. The reported precision of her stemmer was 90.26%. As input, nouns and adjectives in all forms were used. The stemmer was not tested on a corpus, but on a lexicon which completely neglects the frequency of specific forms. The other stemmer was built by Dobrica Pavlinušić [2]. The main purpose of this stemmer was to help in retrieving texts of the official gazette of Republic of Croatia (Narodne novine) [3]. The retrieval system using the stemmer can be found on [4]. His stemmer was not quantitatively tested and is the only one proven to be used in information retrieval of publicly available information. The only other system in public usage which probably uses a stemmer as well as a lexicon is the search engine pogodak.hr [5], but this information is being considered a business secret and couldn't be confirmed.

In the first part of the paper we discuss the results obtained through a online questionnaire taken by 16 web administrators of top web sites in Croatian web space. In the second part we develop a simple and efficient stemmer which could be used widely regardless of the size of the information system. The developed stemmer is rule-based and is optimized by the steepest ascent hill climbing algorithm. Therefore we used a corpus sample, a query sample and a gold standard. As our corpus sample we used a portion of the Vjesnik online newspaper corpus (72M, 4.5M used). We used a morphological lexicon of Croatian nouns for building gold standards – morphological indices (connect tokens that belong to the same paradigm). As our query sample we used 1,000 most frequent nouns in the corpus in their base form.

Current situation in the industry

The aim of the first part of our research was to acquire an insight in retrieval capabilities and the technology used in top 30 Croatian web sites regarding to [6], concerning the problem of morphology in Croatian. For that reason we built an on-line questionnaire and invited the 30 web administrators to fill out the questionnaire. We received 17 answers.

The first and basic question was: "Is your web site search engine sensitive to morphological changes in Croatian language? (e.g. for question 'banka' is it capable of finding documents that contain word forms such as 'banaka', 'bankama', 'banci' etc.):".

Regarding the first answer we divided our questions into two groups:

1. If the answer was "YES", the offered questions were:
   1.1. Which search method do you use?
      - stemmer
      - lexicon
      - both
   1.2. Which engine?
   1.3. Do you find morphological sensitive search techniques enhancing the work of your search engine?
2. If the answer was "NO", the offered questions were:
   2.1. Do you consider morphological sensitive search capable of enhancing the work of your search engine?
   2.2. Are you informed about capabilities of morphological sensitive search?
   2.3. Would you be ready to incorporate morphological sensitive search into your web site in near future?
      - yes – open source engine
      - yes – commercial engine
      - yes – both
      - no – I am not interested

The results showed that at this moment there are hardly any morphologically sensitive search engines used at all. That is to say, just one of the received answers to the first question was positive. Search engines mostly use engines adjusted to English (e.g. Google) or are constructed with simple search techniques like the sameness with the query (=), or similarity with the query (like).

Still, 82% of the users that gave an answer to the question 2.1 believe that introducing a morphologically sensitive search engine would be useful, and 65% of them base their answer on basic knowledge of such search methods (question 2.2). A majority of the users interested in upgrading their search engine didn't care whether the engine would be commercial or open source (76%). Twelve percent was in favor of open source, 6% would prefer a commercial product, and the remaining 6% were not interested.

Building a simple and efficient stemmer

As stated in the introduction, stemmers for Croatian built up to this day were often too complex, mostly built with no regard to specificities of information retrieval and none of them was properly evaluated. In the second part of our research our goal is to build a simple and efficient stemmer which would be empirically tested.

As our dataset, we used two portions of the Vjesnik on-line newspaper corpus [7]. Some specifications regarding that corpus are shown in table 1. Both portions used as the validation set and the test set consist of 10,000 randomly chosen articles that don't overlap. The validation set has 4.515,651 tokens includ-

ing punctuation while the test set has 4.459.519 tokens. The validation set was used for finding the most frequent nouns, for building the set of suffix rules and for finding the optimal set of these rules. The test set was used to test the chosen set of suffix rules. In both cases, the gold standard was built with help of the morphological lexicon.

The lexicon we used consists of 21,003 nouns. It was built with the help of the xfst tool [8]. For purposes of this research we transformed the noun part of the lexicon into two hash tables. The first hash table lists all possible forms regarding the lemma {'kava' : ('kava', kave', 'kavi', 'kavu', 'kavo', 'kavom', 'kavama')}.

Table 1: Basic data about the Vjesnik on-line newspaper corpus (PM = puncuation marks)

Number of articles	187,323
Number of tokens (with PM)	82.862,497
Number of tokens (without PM)	71.935,880
Number of punctuation marks	10.926,617
Number of articles with subtitles	76,982
Number of tokens in article bodies (with PM)	78.245,043
Number of tokens in article bodies (without PM)	67.741,145
Number of sentences in article bodies	3.105,495
Average number of tokens per sentence (with PM)	25.20
Average number of tokens per sentence (without PM)	21.81

The second hash table – an inverse index – gives all possible basic forms regarding the form given {'borom' : ('bor', 'bora')}

With the help of those hash tables we formed a frequency list of nouns in the validation set regardless of the homonymy problem. The list consisted of 13,261 nouns. We checked manually the First 1,033 nouns and excluded 33 of them which got such a high ranking because of the homonymy clash with a very frequent lexeme (such examples are bit, bilo, oko, toga, kada, dok etc.). We used the remaining 1,000 nouns in base form as our query sample. This decision is based on the intuition that in most cases queries in information retrieval are nouns and adjectives in their base forms.

As the gold standard we considered morphological indices built with the help of our noun lexicon. As our query sample, we assumed the 1,000 most frequent nouns in base form from the validation set and built a list of hashes of hashes. Every outer hash consists of the key – the query (one of the 1,000 most frequent noun lemmata) and the value – the inner hash – its forms found in the corpus – the keys – with their number of occurrence – the values. An example of a list item would be {'kapacitet': {'kapacitetima': 8, 'kapacitetu': 2, 'kapacitet': 24, 'kapaciteti': 31, 'kapacitetom': 3, 'kapacitete': 38, 'kapaciteta': 112}}. We built a list of 32 suffix rules by hand through the previously described list of 1.000 most frequent nouns in the validation set. An example of a rule would be ('C',

‘a’, ‘e’, ‘i’, ‘u’, ‘om’, ‘ama’) where the first element of the tuple defines whether the last character of the invariable part of a lexeme should be a consonant or a vocal.

We also put together a set of plain suffixes (140 of them) to examine the simplest approach – just a set of suffixes where there are no specific restrictions.

Using the validation set, our goal was to find the set of rules which would maximize the F1 measure regarding our gold standard. Therefore we used the steepest ascent hill climbing algorithm which adds a rule to the rule set that, in combination with rules already in the set, maximizes the F1 measure, iterating the procedure as long as the F1 measure increases. While searching for the optimal rule set we were also interested in the form of the rule that gives highest F1 measure. Therefore, we put together five forms of our rules:

1. there is a constraint regarding the last character of the invariable part of the lexeme (if consonant or vocal) (‘C’, ‘a’, ‘e’, ‘i’, ‘u’, ‘om’, ‘ama’)
2. there is a constraint regarding the first, entry suffix meaning that the query must _t the entry suffix for the rule to be applied to it (the entry suffix marked with square brackets) ([‘a’], ‘e’, ‘i’, ‘u’, ‘om’, ‘ama’)
3. there are both constraints from the first and the second form of rules (‘C’, [‘a’], ‘e’, ‘i’, ‘u’, ‘om’, ‘ama’)
4. there are none of the constraints described above (‘a’, ‘e’, ‘i’, ‘u’, ‘om’, ‘ama’)
5. there is no set of rules, but just a set of suffixes without any entry restrictions (‘’, ‘a’, ‘u’, ‘om’, ‘i’, ‘ima’, ‘e’, ‘ama’, ‘em’, ...)

Regarding the different forms of rules, the final F1 measure, precision, recall and the number of chosen rules are given in table 2. The maximal F1 measure was achieved with the second form (97.81%) with a small advantage towards other forms. The second form on the other hand requires the biggest amount of rules (25). The least success proved to give the fifth form with a F1 measure of only 94.17%. Since the second form, disregarding the fifth one, is the simplest to apply (there is no necessity for checking the character preceding the potential suffix, and only one suffix – namely the first one (the entry suffix) – has to be checked when searching for rules that can be applied to a specific query token) and has a slightly higher F1 than the others, we have decided to use to second form of rules with the 25 chosen rules.

Table 2: F1, precision and recall measures and number of rules on the validation set regarding the five forms of rules

rule form	F1	precision	recall	# of rules
1	0.9730	0.9775	0.9686	23
2	0.9781	0.9843	0.9719	25
3	0.9747	0.9828	0.9667	24
4	0.9754	0.9774	0.9734	23
5	0.9417	0.9698	0.9151	19

The optimal rule set consisting of 25 rules in order how the rules were added to the set is: ((‘’, ‘a’, ‘u’, ‘om’, ‘i’, ‘ima’, ‘e’), (‘a’, ‘e’, ‘i’, ‘u’, ‘om’, ‘ama’), (‘e’, ‘a’, ‘u’, ‘em’, ‘ima’), (‘o’, ‘a’, ‘u’, ‘om’, ‘ima’), (‘’, ‘a’, ‘u’, ‘om’, ‘ovi’, ‘ova’, ‘ovima’, ‘ove’), (‘ak’, ‘ka’, ‘ku’, ‘kom’, ‘ci’, ‘aka’, ‘cima’, ‘ke’), (‘k’, ‘ka’, ‘ku’, ‘kom’, ‘ci’, ‘cima’, ‘ke’), (‘ac’, ‘ca’, ‘cu’, ‘cem’, ‘ci’, ‘aca’, ‘cima’, ‘ce’), (‘anj’, ‘nja’, ‘nju’, ‘njem’, ‘njom’, ‘nji’, ‘anja’, ‘njima’, ‘nje’), (‘’, ‘ka’, ‘ke’, ‘ci’, ‘ki’, ‘ku’, ‘kom’, ‘aka’, ‘kama’), (‘ar’, ‘ra’, ‘ru’, ‘rom’, ‘ri’, ‘ara’, ‘rima’, ‘re’), (‘ao’, ‘la’, ‘lom’, ‘lu’, ‘lovi’, ‘lova’, ‘lovima’, ‘love’), (‘’, ‘a’, ‘u’, ‘om’, ‘em’, ‘evi’, ‘eva’, ‘evima’, ‘eve’), (‘an’, ‘na’, ‘nu’, ‘nom’, ‘ni’, ‘ana’, ‘nima’, ‘ne’), (‘in’, ‘ina’, ‘inu’, ‘inom’, ‘i’, ‘a’, ‘ima’, ‘e’), (‘am’, ‘ma’, ‘mu’, ‘mom’, ‘movi’, ‘mova’, ‘movima’, ‘move’), (‘t’, ‘ta’, ‘tu’, ‘tom’, ‘ti’, ‘ata’, ‘tima’, ‘te’), (‘zak’, ‘ska’, ‘sku’, ‘skom’, ‘sci’, ‘zaka’, ‘scima’, ‘ske’), (‘tak’, ‘tka’, ‘tku’, ‘tkom’, ‘tci’, ‘ci’, ‘taka’, ‘tcima’, ‘cima’, ‘tke’), (‘dac’, ‘ca’, ‘cu’, ‘cem’, ‘ci’, ‘daca’, ‘cima’, ‘ce’), (‘ga’, ‘ge’, ‘zi’, ‘gi’, ‘gu’, ‘gom’, ‘gama’), (‘st’, ‘sti’, ‘šću’, ‘stima’), (‘g’, ‘ga’, ‘gu’, ‘gom’, ‘zi’, ‘zima’, ‘ge’), (‘sao’, ‘sli’, ‘šlju’, ‘slima’), (‘t’, ‘ti’, ‘ću’, ‘tima’)). Every rule can be accessed through the first item in every tuple – the entry suffix. The growth of the F1 measure and the respective recall and precision regarding the number of rules added to the set is shown graphically in figure 1.

When applying the rules chosen on the validation set on the test set, we got a F1 measure of 97.82%, a precision of 98.40% and a recall of 97.24%.

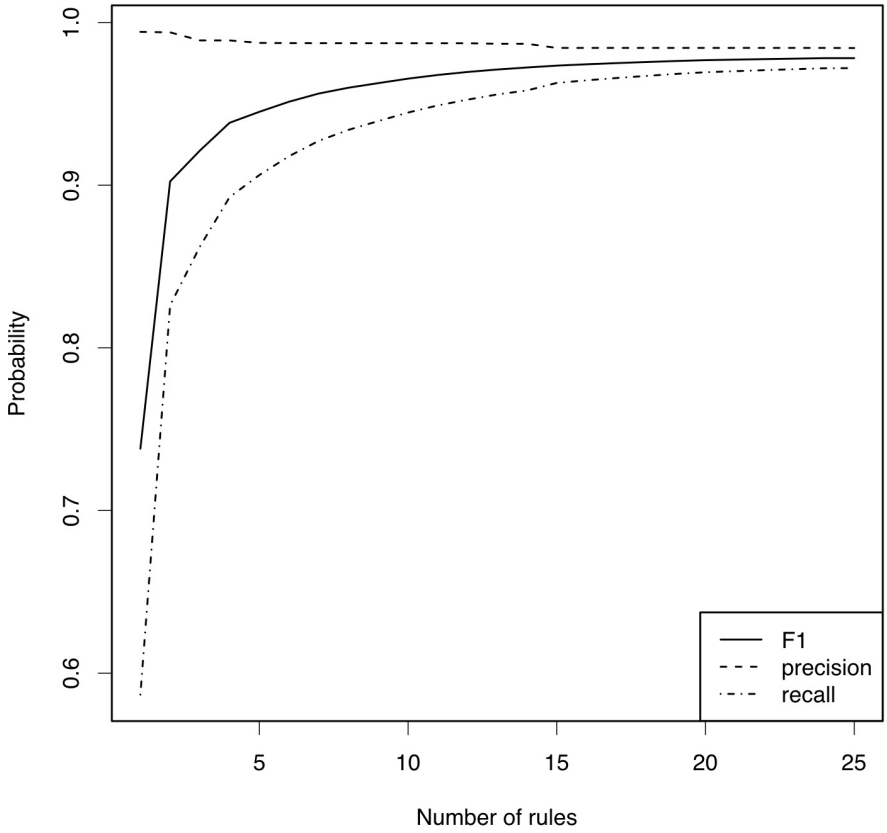
Adding rules for adjectives

Although the rules concerning adjectives are not a primary part of this phase of research, we were interested in how much the most general rules for adjectives, which cover most of them, would harm the task of finding all forms of noun queries in our corpus.

We built six most general rules for adjectives and implemented them on top of the 25 previously selected rules for nouns. We designed the rules for adjectives in the same manner as we decided to design those for nouns. The F1, precision and recall measures when applying these six rules – in order of harmlessness – are shown in table 3.

On the test set, when applying all the 31 rules, the final F1 measure was 97.64%, precision was 98.00% while recall was 97.29%. Thereby we experi-

enced a rise in the recall measure. We believe it is so because of the similarity of paradigms for nouns and adjectives. On the other hand, we lost on precision since some noun queries hopped over to their related adjective forms. Therefore we experienced a small decline in the overall F1 measure.



Picture 1: Growth of the F1 measure with respective precision and recall on the validation set as new rules in second form are added to the rule set

Table 3: F1 measure with precision and recall on the validation set as rules for adjectives are added to the rule set

# of rules	F1	precision	recall
	0.9781	0.9843	0.9719
1	0.9781	0.9843	0.9719
2	0.9779	0.9836	0.9722
3	0.9776	0.9831	0.9722
4	0.9773	0.9822	0.9725
5	0.9769	0.9814	0.9725
6	0.9763	0.9803	0.9725

Conclusion

In this paper we first analyzed the situation in the industry regarding the problem of morphology in retrieving information in Croatian. Our conclusion was that almost none of the most visited on-line portals have this problem solved (just one), although the majority of web administrators (95%) is interested in using a web service which could deal with such a problem.

In the second part of the paper we developed a simple and efficient rule-based stemmer which we optimized through the steepest ascent hill climbing algorithm on a corpus and query sample regarding the lexicon as our gold standard. By optimizing the stemmer we searched for a set of rules which would maximize the F1 measure. We experimented with five forms of rules and accepted the simplest one in form of a set of rules that also got the highest F1 score. Out of 32 rules we accepted 25 of them. In the last step we also measured the decline of measures while adding general rules for adjectives. On the test set, we got a similar F1 measure to our lexicon-based gold standard – 97.82% without the rules for adjectives and 97.64% with them.

We also built a web service which does query expansion regarding the chosen rules and a web application which uses the web service in searching the 4.5M portion of the Vjesnik on-line newspaper corpus. Instructions how to use the web service as well as the example web application can be found on http://faust.ffzg.hr/stemming/.

Our future research will include analyzing query samples received from different information systems such as [9] to back up our assumption that most queries are nouns and adjectives in their base forms. We plan to use different optimization algorithms, enhance our quality measures regarding the number of over generated forms and use a real-world query sample while optimizing the stemmer.

References

[1] Lauc, Tomislava. Problemi obrade prirodnoga jezika u sustavima za pretraživanje obavijesti putem pretraživanja punoga teksta na hrvatskome književnom jeziku. PhD thesis. Zagreb: Faculty of Philosophy, 2000.

[2] Pavlinušić, Dobrica. NN: pretraživač za hrvatski; Information wants to be free. 04-10-20. http://www.rot13.org/ dpavlin/nn.html (07-07-07)

[3] Narodne novine. 07-07-07. http://www.nn.hr/sluzbeni-list/sluzbeni/pregled.asp (07-07-07)

[4] Pavlinušić, Dobrica. Narodne novine pretraživanje. http://nn.rot13.org/ (07-07-07)

[5] Pogodak! 07-07-07. http://www.pogodak.hr (07-07-07)

[6] geminusAudience, pattern, n=16.531, may 2007, Croatia, VALICON i GEMINUS S.A

[7] Vjesnik on-line. 07-07-01. http://www.vjesnik.hr. (07-07-01)

[8] Beesley, Kenneth R.; Karttunen, Lauri. Finite-state morphology. Stanford: CSLI Publications, 2003.

[9] http://www.hr; hrvatski homepage od 1994. 07-07-20. http://www.hr (07-07-20)

What makes sense? Searching for strong WSD predictors in Croatian

Nikola Bakarić
Department of Information Sciences,
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
nbakaric@ffzg.hr

Jasmina Njavro
Department of Information Sciences,
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
jnjavro@ffzg.hr

Nikola Ljubešić
Department of Information Sciences,
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
nljubesi@ffzg.hr

Summary

The goal of this research was to investigate and determine position of strong predictors for word sense disambiguation of Croatian nouns. Research was conducted using supervised learning methods and a corpus of around 70 million words. We have concluded that words in the immediate vicinity of an observed lexeme (1-5 words left and right) have the highest discriminative power. We have also measured the applicability and accuracy of the one-sense-per-discourse method and found it to be very successful as well as the impact of sentence boundaries which proved not to be a good criterion for selecting strong predictors.

Key words: word sense disambiguation, Croatian language, strong predictors

Introduction

Multi-sensed words have presented a problem in computer processing of natural languages since its beginnings. These words carry more than one sense and therefore present a problem in many high-level NLP tasks like information retrieval, automated indexing and machine translation. There are two general approaches in dealing with such problems: stochastic and deterministic [10].

In recent decades the stochastic approach has gained popularity due to increase of processing power and its high efficiency. Again, there are two approaches in stochastic methods – supervised which uses labeled data to build statistical models and unsupervised which uses clustering algorithms without having any labelled data on hand. Supervised algorithms, as expected, achieve much better results than unsupervised [1]. In this research we use a supervised method.
Determining sense of a word is often a complex task, even for humans. The inter-annotator agreement between annotators that prepare data for the SENSEVAL competition is around 60% [3]. In the SENSEVAL competition annotators focus on fine-grained sense distinctions. This research deals with lexemes that have related, but distinctly different senses (strong polysemy) because we believe that there is no point in trying to distinguish fine nuances of meaning which are often unclear to human evaluators. The approach we use is gaining popularity in the NLP community [4]. When trying to determine the sense of a particular lexeme, humans rely on the information given through the context [9]. This research focuses solely on the context of observed lexemes as we try to determine the relationship between position of a word regarding the observed lexeme and its discriminative power in WSD.

Preparing the data

The corpus on which the research was conducted consists of on-line articles of Vjesnik daily paper from May 30th 1999 to December 31st 2006 and it is not POS tagged or lemmatized [8]. The main identifier of the article is the URL and the structure is as following: title, subtitle, text.

- Two separate lists were put together. Each list consisted of articles extracted from the corpus in which lexemes "miš" ("mouse" – the first list) and "stanica" ("cell" – the second list) appear. The lists were then randomly divided into ten sets used in 10-fold cross-validation. They were verticalised and sentence boundaries were marked.
- Next step was to determine possible word senses present in the lists and then to manually annotate the sense of very occurrence of the observed lexemes. Around 1000 occurrences were evaluated. The occurrences in the first 60 percent of the lists were annotated by both annotators together to determine the sense inventory. The remaining 40 percent was annotated separately so as to determine the inter-annotator agreement. The lexeme "miš" was annotated with eight different senses while "stanica" was annotated with six different senses.

Naïve Bayes classifier

Naive Bayes is the simplest probabilistic learning method of all supervised corpus-based methods for word sense disambiguation [7]. The main idea of this classifier is that it calculates in the training corpus the conditional probability of an event (in our case a token in a specific window around the observed lexeme)

regarding a specific sense of the lexeme. Each token contributes potentially useful information about the sense of the ambiguous word present. The classifier does no feature selection – all types are features – it uses all tokens as bag-of-words around the observed lexeme [6]. It is possible to use some feature selection method as the chi-square or mutual information [5], but at this point our primary interest lies in the relative difference in accuracy concerning the size of the window and its distance from the observed lexeme. The greatest disadvantage of such simple classifier is the fact it assumes that the variables given are independent. In spite of this naïve design and apparently over-simplified assumptions, naïve Bayes classifier often works better than some other, more complex classifiers. Due to its simplicity, this classifier is robust enough not to be affected by the curse of dimensionality. Like all probabilistic classifiers under the maximum a posteriori decision rule, it arrives at the correct classification as long as the correct class is more probable than any other class; hence class probabilities do not have to be estimated very well [7].

Results

The one-sense-per-discourse hypothesis assumes that in one discourse a polysemous lexeme is used in only one sense. Yarowsky measures that phenomenon as applicable to English and uses it effectively in his semi-supervised approach to WSD [11]. Since we annotated all occurrences of chosen lexemes in selected documents, it was possible to measure the applicability and accuracy of this hypothesis in our corpus. The method has proven to be applicable in almost one third of cases as well as quite accurate as can be seen in Table 1.

Table 1. One-sense-per-discourse applicability and accuracy percentage

	Applicability	Accuracy
“miš”	28.92%	88.98%
“stanica”	26.31%	97.10%

We trained and tested the Naive Bayes classifier by using 10-fold cross-validation. Since there was no additional parameter estimation, we did not need a validation set. The experiments were performed with the varying window size and the varying window distance with window size one. Results are shown in figures 1 to 4.

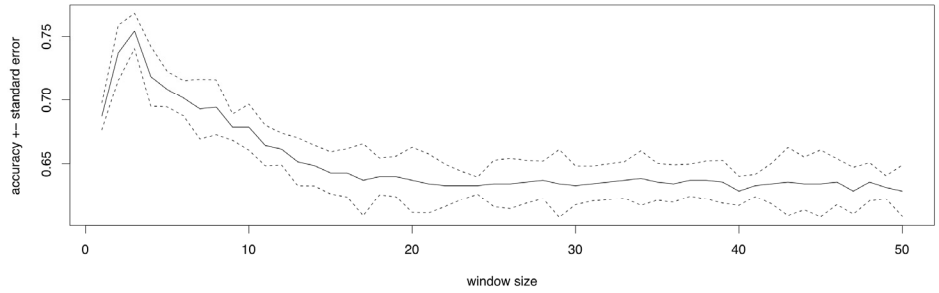


Figure 1. Window size/accuracy ratio for “miš”

The results of varying window size show that accuracy decreases as the window size increases. In the case of the lexeme “stanica”, it decreases constantly while the highest accuracy for the lexeme “miš” is obtained with window size 3. The results for varying window distance show that in case of both lexemes best sense predictors lie in the first five positions from the observed lexeme and that the discriminative power of more distant tokens is quite constant.

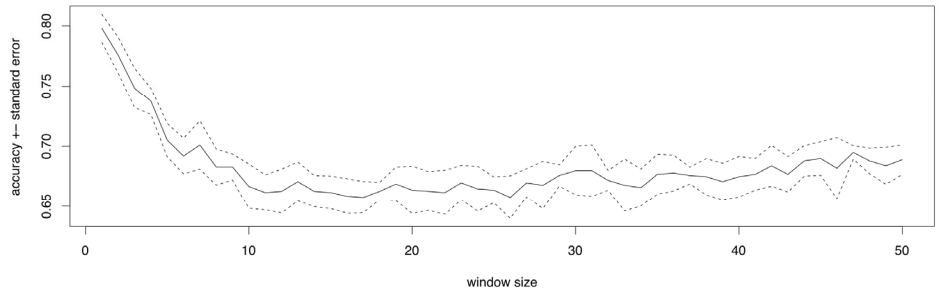


Figure 2. Window size/accuracy ratio for “stanica”

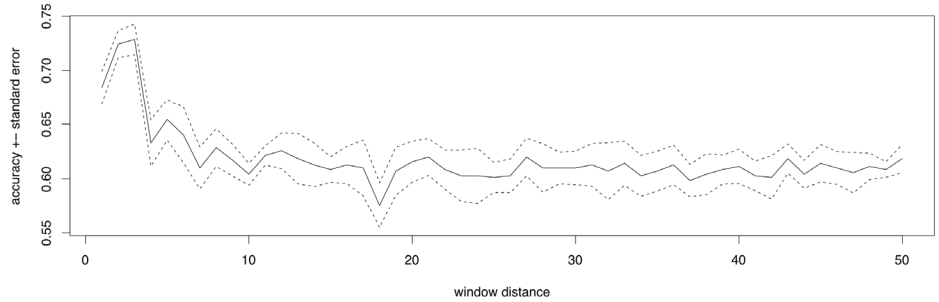


Figure 3. Window distance/accuracy ratio for “miš”

The difference between the lexemes “miš” and “stanica” lies in the fact that the lexeme “stanica” mostly makes strong NP collocation (“matične stanice”, “autobusna stanica”). That is in our belief the reason why the lexeme “stanica” has its accuracy peak with a window distance and size of one. The lexeme “miš” needs, as stated before, a window distance or size of three to achieve peak accuracy.

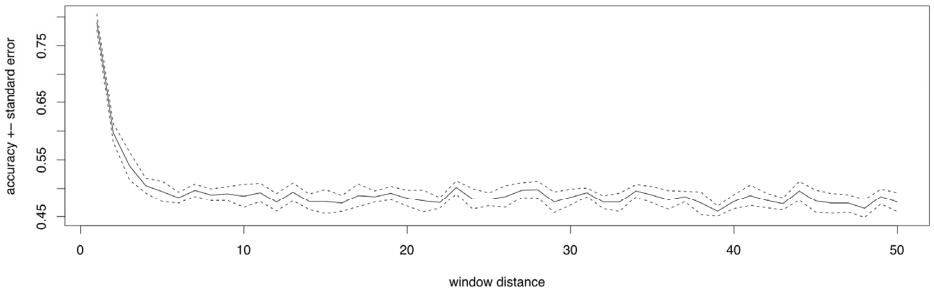


Figure 4. Window distance/accuracy ratio for “stanica”

Furthermore, we experimented with the importance of sentence endings for WSD. We trained one classifier with three first and last tokens in the sentence in which the lexeme occurs and one classifier with three last and three first tokens in neighbouring sentences. The accuracy difference between these two classifiers is shown in Table 2. While tokens in the sentence of the observed lexeme are better sense predictors, the difference is rather small and it remains unclear to what extent it is the result of the smaller distance from the observed lexeme in comparison to its possibly bigger discriminative power.

Table 2. Accuracy with standard error in relation to 3 tokens before/after observed lexeme sentence boundary

	Before sentence boundary	After sentence boundary
“miš”	68,00%±1,52% (SE)	64,14%±2,09%
“stanica”	57,37%±1,75%	57,27%±1,18%

Conclusion

Applications of sense disambiguation systems are many. Apart from machine translation; information retrieval, information extraction and text mining could also benefit from a working word sense disambiguation system as well as lexicography [2]. The main goal of the research was to examine the connection between the distance of a token to the ambiguous lexeme and its’ discriminative power for WSD. Our main conclusion is that best predictors of a sense of the observed lexemes are situated near that lexeme, usually from 1 to 5 places to the left or right. The one-sense-per-discourse is proven to be applicable in a third of cases and is quite accurate. Since this method is applicable only when the

lexeme is mentioned more than once in a discourse, its possible application in a WSD system is limited, but it can still strongly affect the final results, especially in unsupervised and semi-supervised approaches. The sentence limits have not proven to be any significant border of strong WSD predictors.
Since we are not aware of any research of WSD for Croatian, we believe that conclusions drawn in this paper represent a stepping stone for further research in WSD and natural language processing of Croatian language.

References

[1] Agirre, E.; Edmonds, P., editors. Word Sense Disambiguation: Algorithms and Applications. // Text, Speech and language Technology. Vol. 33 (2006), page 7
[2] Agirre, E.; Edmonds, P., editors. Word Sense Disambiguation: Algorithms and Applications. // Text, Speech and language Technology. Vol. 33 (2006), pages 10-11
[3] Edmonds, P. SENSEVAL: The evaluation of word sense disambiguation systems. // ELRA Newsletter. Vol. 7 (2002), No. 3
[4] Ide, N.; Wilks, Y. Making Sense About Sense // Word Sense Disambiguation: Algorithms and Applications. // Text, Speech and language Technology. Vol. 33 (2006), pages 48-49
[5] Manning, C.D.; Schütze, H. Foundations of Statistical Natural Language Processing. London: MIT Press, 1999, pages 169-172
[6] Manning, C.D.; Schütze, H. Foundations of Statistical Natural Language Processing. London: MIT Press, 1999, pages 237-239
[7] Marquez, L.; Escudero G.; Martinez D.; Rigau G. Supervised Corpus-Based Methods for WSD // Word Sense Disambiguation: Algorithms and Applications. // Text, Speech and language Technology. Vol. 33 (2006), pages 185-186
[8] Vjesnik d.d., On-line edition from 30 May 1999 to 31 Dec 2006. http://www.vjesnik.hr (25 Mar 2007)
[9] Wikipedia, Distributional Hypothesis. http://en.wikipedia.org/wiki/Distributional_hypothesis (25 Mar 2007)
[10] Wikipedia, Natural Language Processing. http://en.wikipedia.org/wiki/Natural_language_ processing (25 Mar 2007)
[11] Yarowski, D. Unsupervised word sense disambiguation rivalling supervised methods. // Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL), Cambridge, USA, 1995, pages 189-196

Korelacija, informacija i kauzalnost

Božidar Tepeš
Jacinta Grbavac
Odsjek za informacijske znanosti, Filozofski fakultet
Ivana Lučića 3, Zagreb, Hrvatska
btepes@ffzg.hr

Sažetak

Statistička povezanost zove se korelacija, a stupanj statističke povezanosti mjeri se koeficijentom korelacije. U prirodi jedan od glavnih zakonitosti je kauzalnost, a kauzalnost ne vidimo iz nezavisnosti i povezanosti. Kauzalnost pretpostavlja relaciju između slučajne varijable koja je uzrok i slučajne varijable koja je posljedica. U slučaju nezavisnosti relacija između slučajnih varijabli je komutativna ili govorimo o skupu nezavisnih slučajnih varijabli, a ne o nizovima slučajnih varijabli kakve imamo kod kauzalnosti. Kauzalnost se prikazuje acikličkim usmjerenim grafom koji predstavlja kauzalnu povezanost slučajnih varijabli. Informacija u radu je mjera stupnja sličnosti između statističkog modela odnosa slučajnih varijabli i stvarnih odnosa varijabli mjerenja dobivenih uzorkom. U radu će se pojmovi korelativnosti, informacije i kauzalnosti ilustrirati Markovljevim i kauzalnim modelom predikata hrvatskog jezika

Ključne riječi: Korelacija, informacija, kauzalnost, Markovljev model, kauzalni model, predikat hrvatskog jezika.

1. Korelacija

Koeficijent korelacije [1] govori o nezavisnosti pojava. Nezavisnost je pojam vezan za vjerojatnost i slučajne varijable, a statistika koristi uzorak koji je niz slučajnih varijabli koje imaju svoje razdiobe. Ako nezavisnost slučajne varijable X i slučajne varijable Y označimo $X \perp Y$ tada za njihove funkcije razdiobe vrijedi:

$$f(x,y) = f(x) \cdot f(y) \tag{1.1}$$

Koeficijent korelacije slučajnih varijabli X i Y označimo s r_{XY} . . Slučajne varijable mogu biti nezavisne i tada je koeficijent korelacije jednak nuli. Obrat ne vrijedi ili ako je koeficijent korelacije jednak nuli slučajne varijable ne moraju biti nezavisne. Tu činjenicu možemo zapisati relacijom (1.2).

(1.2) $$(X \perp Y) \Rightarrow (r_{XY} = 0)$$

U slučaju niza slučajnih varijabli $X_1, X_2, ..., X_n$ možemo govoriti o uvjetnoj nezavisnosti dvije slučajne varijable X_i i X_j iz niza slučajnih varijabli $X_1, X_2, ..., X_n$ uz uvjet ostalih slučajnih varijabli iz niza koje se razlikuju od slučajnih varijabli X_i i X_j. Uvjetnu nezavisnost označimo $X_i \perp X_j \mid X_1,...,X_{i-1},X_{i+1},...,X_{j-1},...,X_{j+1},...,X_n$ i za uvjetnu nezavisnost vrijedi jednadžba (1.3).

(1.3)
$$f(x_1,x_2,...,x_n) = f(x_i \mid x_1,...,x_{i-1},x_{i+1},...,x_n) \cdot f(x_j \mid x_1,...,x_{j-1},x_{j+1},...,x_n)$$

Uvjetna nezavisnost mjeri se koeficijentom parcijalne korelacije $\rho_{X_iX_j \cdot Z}$ gdje smo sa Z označili skup slučajnih varijabli koje se razlikuju od X_i i X_j ili $Z = \{X_1,...,X_{i-1},X_{i+1},...,X_{j-1},...,X_{j+1},...,X_n\}$. Također vrijedi relacija (1.4) između vrijednosti koeficijenta parcijalne korelacije i uvjetne nezavisnosti.

(1.4) $$(X_i \perp X_j \mid Z) \Rightarrow (\rho_{X_iX_j \cdot Z} = 0)$$

Obrat relacija (1.2) i (1.4) vrijedi u slučaju normalne razdiobe slučajnih varijabli. To znači ako su koeficijent korelacije i koeficijent parcijalne korelacije jednaki nuli onda su slučajne varijable nezavisne i uvjetno nezavisne. To znači da možemo govoriti o nezavisnosti za male vrijednosti koeficijenata korelacije.

2. Informacija

Pojam informacije temelji se na pojmu entropije ili neodređenosti. Možemo reći informacija je razlika entropija ili smanjenje neodređenosti. U slučaju jedne diskretne slučajne varijable X koja ima vrijednosti x s vjerojatnostima $f(x)$ za koje vrijedi $\sum_x f(x) = 1$ entropija $H(X)$ je određena:

(2.1) $$H(X) = -\sum_x f(x) lb(f(x))$$

gdje je $lb(f(x)) = \log_2(f(x))$ logaritam po bazi dva. Kažemo slučajna varijabla X ima entropiju $H(X)$ bita.

Statističkim modelima na temelju uzorka ocjenjujemo razdiobu slučajne varijable X. Ako na temelju modela M ocijenimo vjerojatnosti slučajne varijable X s $m(x)$. Temeljno pitanje je koliko naš model odgovara stvarnim vrijednostima ili koja je razlika između $f(x)$ i $m(x)$. Ocjenu neodređenosti modela daje nam ukrštena entropija $H(X,M)$:

$$H(X,M) = -\sum_x f(x) lb(m(x)) \tag{2.2}$$

Informacija o točnosti modela je razlika između ukrštene entropije $H(X,M)$ i entropije $H(X)$. Ta informacija zove se relativna entropija ili Kullback-Leiblerova razlika [2] $I(f \parallel m)$:

$$\begin{aligned} I(f \parallel m) &= H(X,M) - H(X) \\ &= \sum_x f(x) lb \tfrac{f(x)}{m(x)} \end{aligned} \tag{2.3}$$

Može se pokazati nenegativnost Kullback-Leiblerove razlike ili $I(f \parallel m) \geq 0$. KL razlika je informacija dobivena uvidom u stvarne vrijednost slučajne varijable ili neodređenost modela koja se smanjila uvidom u stvarne podatke o slučajnoj varijabli. Informacija kao razlika entropija također se mjeri u bitima. Minimalizacijom Kullback-Leiblerove razlike možemo ocijeniti model koji se najmanje razlikuje od stvarnih podataka.

Markovljev model M je stohastički model koji preko Markovljevog lanca opisuje razdiobu $m(x)$ slučajne varijable X. Markovljev lanac prvog reda određuje razdiobu $m(x_i)$ stanja i na temelju razdiobe $m(x_{i-1})$ stanja $i-1$ koje prethodi stanju i. Povezanost je određena vjerojatnostima prijelaza $p_{i-1,i}$:

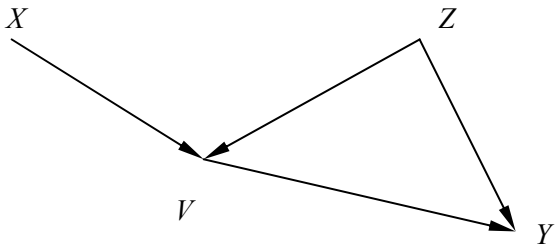
$$m(x_i) = \sum_{i-1} p_{i-1,i} m(x_{i-1}) \tag{2.4}$$

Markovljev lanac može biti i višeg reda r govorimo o $r-$ gramskim modelima u kojima su vjerojatnosti prijelaza iz r prethodnih stanja $i-1, i-2, ..., i-r$ u $i-$to stanje koje slijedi iza prethodnih stanja.

3. Kauzalnost

Kauzalna povezanost varijabli mjerenja određena je smjerom od varijable koja je uzrok prema varijabli koja je posljedica. Tako se grafički kauzalne poveza-

nosti prikazuju acikličkim usmjerenim grafom koji se naziva kauzalna struktura. Kauzalna povezanost može biti direktna ako postoji usmjereni put između dvije varijable ili zbunjujuća kauzalna povezanost ako postoji put koji nije usmjeren između dvije varijable. Primjere direktne i zbunjujuće kauzalnosti prikazane su na kauzalnoj strukturi na slici 1. gdje je direktna kauzalnost između uzroka X i posljedice Y preko varijable V i zbunjujuća kauzalnost između uzroka X i posljedice Y preko varijable Z.



Slika 1. Kauzalna struktura

Pomoću IC* algoritma induktivne kauzalnosti [3], [4] određuje se kauzalna struktura ili aciklički usmjereni graf koji pokazuje kauzalnu povezanost varijabli mjerenja. Intenzitet kauzalne zavisnosti mjeri se informacijom o međuzavisnosti [5]:

$$I(X \perp Y \mid Z) = -0{,}5\, lb\left(1 - \rho^2_{XY \cdot Z}\right) \tag{3.1}$$

gdje je lb logaritam po bazi 2, a Z je podskup skupa varijabli mjerenja u kojem nisu varijable X i Y.
Kauzalni model ili kauzalna mreža je kauzalna struktura na kojoj varijablama mjerenja ili vrhovima v su pridružene funkcije f_v oblika:

$$v = f_v(R, u_v) \tag{3.2}$$

gdje su R roditelji vrha v, a u_v su nezavisne slučajne varijable. Funkcije f_v mogu biti razdiobe slučajne varijable v ili $p(v)$koje se mogu odrediti iz uvjetnih razdioba $p(v \mid R)$ za koje vrijedi:

$$p(v) = p(v \mid R) \cdot p(R) \tag{3.3}$$

Jednadžba se rekurzivno primjenjuje na cijeloj kauzalnoj strukturi i određuje razdiobu kauzalnog modela $p(km)$:

$$p(km) = \prod_{v,R} p(v \mid R) \tag{3.4}$$

Informacija o točnosti kauzalnog modela je informacija o točnosti statističkoga modela određena Kullback-Leiblerova razlika $I(f \parallel m)$ određena izrazom (2.3) gdje je f stvarna razdioba $p(v)$ varijabli mjerenja, a m je razdioba $p(km)$ određena kauzalnim modelom.

4. Model predikata hrvatskog jezika

Pojam korelacije, informacije i kauzalnosti ćemo ilustrirati na primjeru predikata hrvatskog jezika. Iz baze označenih rečenica hrvatskog jezika [6] uzet je uzorak od stotinu rečenica. U tim rečenicama označeni su dijelovi predikata: pomoćni glagol *I*, glavni glagol *V*, odrednica (zamjenica) *D*, pridjev *A* i imenica *N*. Koeficijenti korelacije r_{ij} su prikazani matricom R:

$$R = \begin{bmatrix} 1 & 0{,}99 & 0{,}97 & 0{,}96 & 0{,}94 \\ 0{,}99 & 1 & 0{,}97 & 0{,}96 & 0{,}94 \\ 0{,}97 & 0{,}97 & 1 & 0{,}94 & 0{,}95 \\ 0{,}96 & 0{,}96 & 0{,}94 & 1 & 0{,}95 \\ 0{,}94 & 0{,}94 & 0{,}95 & 0{,}95 & 1 \end{bmatrix} \tag{4.1}$$

Iz matrice koeficijenata korelacije R vidimo jaku povezanost pojedinih dijelova predikata. Parcijalni koeficijenti korelacije $\rho_{ij \cdot Z}$ između dijelova predikata i i j uz uvjet skupa Z u kojem nisu i i j prikazani su matricom ρ:

$$\rho = \begin{bmatrix} - & 0{,}85 & -0{,}54 & 0{,}16 & 0{,}17 \\ 0{,}85 & - & 0{,}44 & 0{,}13 & -0{,}21 \\ -0{,}54 & 0{,}44 & - & -0{,}16 & 0{,}44 \\ 0{,}16 & 0{,}13 & -0{,}16 & - & 0{,}49 \\ 0{,}17 & -0{,}21 & 0{,}44 & 0{,}49 & - \end{bmatrix} \tag{4.2}$$

Parcijalne koeficijente korelacije možemo podijeliti na dvije skupine. U prvoj skupini su koeficijenti veći od 0,4, a u drugoj su skupini koeficijenti manji od 0,4. Prva skupina koeficijenata pokazuje povezanost, a druga skupina nezavisnost dijelova predikata.
Iz uzorka od stotinu rečenica hrvatskog jezika izračunata je razdioba dijelova predikata $f(x)$:

x	I	V	D	A	N
$f(x)$	0,30	0,49	0,08	0,07	0,06

Tabela 4.1 Razdioba dijelova predikata na uzorku

Uz pretpostavku uniformne razdiobe dijelova predikata entropija je 2,81 bit, a entropija uzorka je 2,07 bita. Informacija o razdiobi uzorka je razlika entropija i iznosi 0,74 bita.
Sada pristupimo modeliranju predikata Markovljevim modelom (MM) i kauzalnim modelom (KM).
U Markovljevom modelu predikata bitna je matrica prijelaza P s elementima p_{ij} koji pokazuju vjerojatnosti nalaženja dijela predikata j iza dijela predikata i:

(4.3)
$$P = \begin{bmatrix} 0{,}021 & 0{,}575 & 0{,}043 & 0{,}255 & 0{,}106 \\ 0{,}568 & 0{,}189 & 0{,}216 & 0{,}027 & 0{,}000 \\ 0{,}000 & 0{,}778 & 0{,}000 & 0{,}111 & 0{,}111 \\ 0{,}100 & 0{,}200 & 0{,}000 & 0{,}200 & 0{,}500 \\ 0{,}200 & 0{,}600 & 0{,}200 & 0{,}000 & 0{,}000 \end{bmatrix}$$

Uz pretpostavku početnog stanja uniformne razdiobe dijelova predikata i sustavnog množenja tog stanja s matricom prijelaza P dobivamo:

(4.4)
$$\begin{bmatrix} 0{,}2 & 0{,}2 & 0{,}2 & 0{,}2 & 0{,}2 \end{bmatrix} \cdot \begin{bmatrix} 0{,}021 & 0{,}575 & 0{,}043 & 0{,}255 & 0{,}106 \\ 0{,}568 & 0{,}189 & 0{,}216 & 0{,}027 & 0{,}000 \\ 0{,}000 & 0{,}778 & 0{,}000 & 0{,}111 & 0{,}111 \\ 0{,}100 & 0{,}200 & 0{,}000 & 0{,}200 & 0{,}500 \\ 0{,}200 & 0{,}600 & 0{,}200 & 0{,}000 & 0{,}000 \end{bmatrix}^{n}$$
$$\underset{n\to\infty}{\longrightarrow} \begin{bmatrix} 0{,}27 & 0{,}40 & 0{,}12 & 0{,}11 & 0{,}10 \end{bmatrix}$$

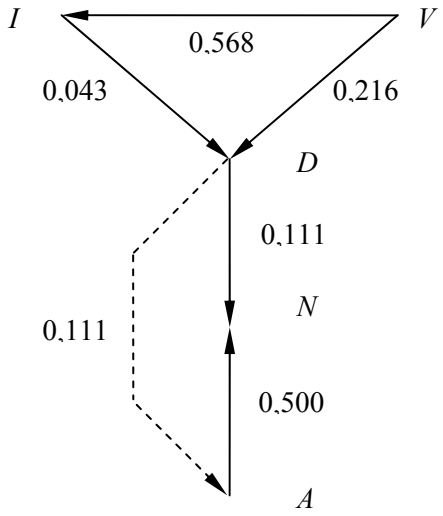
Iz toga slijedi razdioba dijelova predikata na temelju Markovljevog modela:

$$m_{MM} = \left[0{,}27 \quad 0{,}40 \quad 0{,}12 \quad 0{,}11 \quad 0{,}10\right] \tag{4.5}$$

Kullback-Leiblerova razlika je:

$$\begin{aligned} I\left(f \,||\, m_{MM}\right) &= 0{,}30 lb \tfrac{0{,}30}{0{,}27} + 0{,}49 lb \tfrac{0{,}49}{0{,}40} + 0{,}08 lb \tfrac{0{,}08}{0{,}12} + 0{,}07 lb \tfrac{0{,}07}{0{,}11} \\ &+0{,}06 lb \tfrac{0{,}06}{0{,}10} = 0{,}052 bita \end{aligned} \tag{4.6}$$

i pokazuje razliku između Markovljevog modela i uzorka predikata hrvatskog jezika.
Kod kauzalnog modela koristimo kauzalnu strukturu dobivenu iz parcijalnih koeficijenata korelacije:



Slika 4.1 Kauzalna struktura predikata

Iz kauzalne strukture je vidljiva dominantna uloga glagola V koji je uzrok nalaženja svih ostalih dijelova predikata jer glagol otvara u predikatu mjesta svim ostalim dijelovima predikata V, D, A i N. To je u suglasnosti s gramatičkim ustrojstvom rečenica hrvatskog jezika gdje je glagol temeljna riječ predikata [7].
Iz direktnih kauzalnih i zbunjujućih kauzalnosti dobivamo razdiobu dijelova predikata kako slijedi:

$$m_{KM}(I) = p_{21} = 0{,}568 \approx 0{,}57$$
$$m_{KM}(D) = p_{23} + p_{21}p_{23} = 0{,}216 + 0{,}568 \cdot 0{,}043 = 0{,}240 \approx 0{,}24$$
$$m_{KM}(A) = p_{34} = 0{,}111 \approx 0{,}11$$
$$m_{KM}(N) = (p_{23} + p_{21}p_{23})p_{35} + p_{34}p_{45} = 0{,}240 \cdot 0{,}111 + 0{,}111 \cdot 0{,}500$$
$$= 0{,}082 \approx 0{,}08$$

Razdioba dijelova predikata I, D, A i N izvedena iz razdiobe svih dijelova predikata I, V, D, A i N na uzorku od stotinu rečenica hrvatskog jezika je:

x	I	D	A	N
$f_1(x)$	0,58	0,17	0,14	0,11

Tabela 4.2 Razdioba dijelova predikata iz uzorka bez V

Kullback-Leiblerova razlika je:

$$I(f_1 \,||\, m_{KM}) = 0{,}58 lb \tfrac{0{,}58}{0{,}57} + 0{,}17 lb \tfrac{0{,}17}{0{,}24} + 0{,}14 lb \tfrac{0{,}14}{0{,}11} + 0{,}11 lb \tfrac{0{,}11}{0{,}08} = 0{,}029\, bita \tag{4.7}$$

i pokazuje razliku između Markovljevog modela i uzorka predikata hrvatskog jezika. Dobiveni rezultat pokazuje prednost kauzalnog modela pred Markovljevim jer je Kullback-Leiblerova razlika kauzalnog modela manja jer je razlika modela od uzorka manja. Cijeli model bi trebalo testirati na cijeloj Baza morfološki i sintaktički označenih rečenica hrvatskog jezika i proširiti na ostale dijelove rečenica hrvatskog jezika i tako definirati kauzalnu strukturu cijelih rečenica.

Literatura

[1] I. Pavlić, Statistička teorija i primjena (1988), Tehnička knjiga, Zagreb
[2] V. Hatzivassiloglou (2006), *Cross Entropy and Relative Entropy,* University of Texas,Dallas, http://www.hlt.utdallas.edu/~vh/Courses/Spring06/StatisticalNLP.html
[3] J. Pearl (2001), *Causality: Models, Reasoning, and Inference,* Cambridge University Press, Cambridge UK
[4] B. Tepeš (2007), *Statistički modeli na grafovima,* (skripte) Filozofski fakultet, Zagreb, http://www.ffzg.hr
[5] J. Whittaker (1990), *Graphical Models in Applied Multivariate Statistics,* John Wiley & Sons, Chichester UK
[6] Baza morfološki i sintaktički označenih rečenica hrvatskog jezika (2006), znanstveni projekt Označavanje i prepoznavanje riječi hrvatskog jezika, voditelj B. Tepeš, http://infoz.ffzg.hr/tepes/
[7] R. Katičić (1991), *Sintaksa hrvatskog književnog jezika*, HAZU, Globus, Zagreb

Dekompozicija relacijske sheme bez gubitka informacija

Vladimir Mateljan
Odsjek za informacijske znanosti, Filozofski fakultet
I. Lučića 3, Zagreb, Hrvatska
vmatelja@ffzg.hr

Goran Đambić
Croatia Airlines
Zagreb, Hrvatska
goran.djambic@croatiaairlines.hr

Sažetak

Postupkom normalizacije podataka relacijska shema dekomponira se na više manjih i pravilnijih podshema. Pomoću Rissanenovog kriterija nezavisnosti komponenata moguće je dokazati je li polazni skup informacija sačuvan, tj. da je dekompozicija početne relacijske sheme reverzibilna. Zahvaljujući tome Rissanenov test ima izvanrednu važnost u postupku oblikovanja baze podataka. Za Rissanenov test potrebno je imati komplet relacijskih shema koje se dobiju dekompozicijom, tj. potrebno je najprije izvršiti normalizaciju baze na relacijske sheme, a zatim izvršiti test. U radu se predlaže jedan novi test pomoću kojeg se može provjeriti reverzibilnost dekompozicije relacijske sheme bez da se prethodno izvršio postupak normalizacije.

Ključne riječi: funkcijska zavisnost, relacijska shema, normalizacija, Rissanenenov test, dekompozicija, reverzibilnost, gubitak informacija

1. Uvod

Normalizacija je metoda oblikovanja baze podataka. Osnovni ciljevi ove metode su eliminacija anomalija održavanja baze i svođenje redundancije u bazi na kontroliranu redundanciju. Normalizacija sintezom polazi od skupa funkcijskih zavisnosti zadanog na skupu atributa i direktno konstruira skup relacijskih shema u trećoj normalnoj formi. Konačan model baze podataka proizlazi iz skupa funkcijskih zavisnost koji je zadan na skupu atributa. Ako u takvim modelima, koji proizlaze iz skupa funkcijskih zavisnosti, vrijede Rissanenovi uvjeti reverzibilnosti, onda je dekompozicija početne relacijske sheme reverzibilna, odnosno tokom dekompozicije nije došlo do gubitka informacija. Za normaliza-

ciju baze do treće normalne forme, u radu ćemo koristiti Bernsteinov algoritam za normalizaciju sintezom.

2. Bernsteinov algoritam za normalizaciju sintezom

Algoritam za vertikalnu normalizaciju sintezom definirao je Bernstein (1976) i glasi:

Ulaz : Skup funkcijskih zavisnosti (**FZ**) **F**.
Izlaz : Komplet relacijskih shema u trećoj normalnoj formi (**3NF**).

Postupak:

(1) Nađi atribut **Z** tako da vrijedi $\mathbf{Z} \in \mathbf{R}$.
(2) Konstruiraj funkcijsku zavisnost $\mathbf{R} \rightarrow \mathbf{Z}$ i dodaj je u **F**.
(3) Nađi reducirani neredundantni prstenasti pokrivač **G** za prošireni **F**.
(4) Za svaku sastavljenu funkcijsku zavisnost (**SFZ**) $(\mathbf{X_1},\mathbf{X_2},...,\mathbf{X_n}) \rightarrow \mathbf{Y}$ u **G** konstruiraj relacijsku shemu $\mathbf{R_i}(\mathbf{X_1X_2}...\mathbf{X_nY})$ s ključevima $\mathbf{X_1},\mathbf{X_2},...,\mathbf{X_n}$.
(5) Izbaci atribut **Z** iz relacijske sheme $\mathbf{R_i}$.
(6) Ispiši skup relacijskih shema u **3NF**.

3. Reverzibilnost dekompozicije relacijske sheme

Normalizacijom podataka, relacija se dekomponira na više projekcija. Pri tome ne smije doći do gubitka informacija, odnosno mora biti moguće operacijom prirodnog pridruživanja nad projekcijama uspostaviti polaznu relaciju. Za ovakvu dekompoziciju kažemo da je reverzibilna. Skup projekcija koje zadovoljavaju ovaj zahtjev Rissanen naziva nezavisnim komponentama relacije.

3.1. Rissanenov uvjet nezavisnosti komponenata

Prema Rissanenu (Rissanen, 1977), nezavisne komponente relacije moraju ispunjavati sljedeće uvjete reverzibilnosti, tj. moraju biti ispunjeni kriteriji:

(1) $\mathbf{R}=\mathbf{R_1} \cup \mathbf{R_2}$
(2) $(\mathbf{R_1} \cap \mathbf{R_2} \rightarrow \mathbf{R_1}) \vee (\mathbf{R_1} \cap \mathbf{R_2} \rightarrow \mathbf{R_2})$

Riječima ove kriterije možemo izraziti na sljedeći način: Relacijske sheme $\mathbf{R_1}$ i $\mathbf{R_2}$, koje su dobivene dekompozicijom relacijske sheme **R,** su nezavisne ako je unija atributa u projekcijama relacije jednaka relacijskoj shemi te presjek $\mathbf{R_1}$ i $\mathbf{R_2}$ sadrži ključ bar jedne od njih.
Ako želimo ustanoviti jesu li zadovoljeni Rissannenovi uvjeti, moramo normalizirati bazu i za dobivene relacijske sheme u **3NF** provjeriti prethodna dva uvjeta.

3.2. Provjera reverzibilnosti normalizirane relacijske sheme

Dekompozicija je reverzibilna ako se skup relacijskih shema **R** sastoji od relacijskih shema $\mathbf{R_1},...,\mathbf{R_n}$ koje su dobivene postupkom vertikalne normalizacije sintezom i ako su ispunjeni Rissanenovi uvjeti reverzibilnosti ($\forall \mathbf{R_i} \in \mathbf{R}$) ($\exists \mathbf{R_j} \in \mathbf{R}, \mathbf{j} \neq \mathbf{i}$) ($\mathbf{R_i} \cap \mathbf{R_j}$ sadrži ključ od $\mathbf{R_i}$ ili od $\mathbf{R_j}$).

4. Testiranje reverzibilnosti na skupu funkcijskih zavisnosti

Pokazat ćemo da se provjera reverzibilnosti dekompozicije relacijske sheme može izvesti na skupu funkcijskih zavisnosti, bez da se prije baza normalizirala. Razmatrat ćemo skup funkcijskih zavisnosti $\mathbf{F}=\{\mathbf{X_1}\rightarrow\mathbf{Y_1},\mathbf{X_2}\rightarrow\mathbf{Y_2},...,\mathbf{X_n}\rightarrow\mathbf{Y_n}\}$ u kojem niti jedna funkcijska zavisnost $\mathbf{X_i}\rightarrow\mathbf{Y_i}$ neće sadržavati isti atribut na svojoj lijevoj i desnoj strani. Naime, skup funkcijskih zavisnosti **F** uvijek se može zamijeniti skupom $\mathbf{F}=\{\mathbf{X_1}\rightarrow\mathbf{Y_1}\setminus\mathbf{X_1},\mathbf{X_2}\rightarrow\mathbf{Y_2}\setminus\mathbf{X_2},...,\mathbf{X_n}\rightarrow\mathbf{Y_n}\setminus\mathbf{X_n}\}$.

Tvrdnja:

Neka je $\mathbf{F}=\{\mathbf{X_1}\rightarrow\mathbf{Y_1},\mathbf{X_2}\rightarrow\mathbf{Y_2},...,\mathbf{X_n}\rightarrow\mathbf{Y_n}\}$ skup funkcijskih zavisnosti nad skupom atributa **R**. Ako **F** sadrži **FZ** $\mathbf{X_r}\rightarrow\mathbf{Y_r}$ i $\mathbf{X_s}\rightarrow\mathbf{Y_s}$, $\mathbf{r}\neq\mathbf{s}$ za koje vrijedi:

(a) $\exists \mathbf{W}\ (\mathbf{W}\subset \mathbf{X_r}\cap\mathbf{Y_s})\ \&\ (\mathbf{W}\not\subset \bigcup_{i\neq r\vee i\neq s} \mathbf{X_i}\ \mathbf{Y_i})$

(b) $\exists\ \mathbf{X_{r0}}\ (\mathbf{X_{r0}}\subset\mathbf{X_r}\setminus\mathbf{W})\ \&\ (\mathbf{X_{r0}}\not\subset \bigcup_{i\neq r} \mathbf{X_i}\ \mathbf{Y_i})$

(c) $\exists \mathbf{Y_{r0}}\ (\mathbf{Y_{r0}}\subset\mathbf{Y_r})\ \&\ (\mathbf{Y_{r0}}\not\subset \bigcup_{i\neq r} \mathbf{X_i}\ \mathbf{Y_i})$

(d) $\exists\ \mathbf{X_{s0}}\ (\mathbf{X_{s0}}\subset\mathbf{X_s})\ \&\ (\mathbf{X_{s0}}\not\subset \bigcup_{i\neq s} \mathbf{X_i}\ \mathbf{Y_i})$

onda se Bernsteinovim algoritmom za normalizaciju sintezom dobiva skup relacijskih shema koji sadrži relacijsku shemu $\mathbf{R_r}$ koja nije povezana niti s jednom od preostalih relacijskih shema dobivenih u postupku normalizacije.

Drugim riječima:

Ako skup funkcijskih zavisnosti **F** nad skupom atributa **R** zadovoljava (a), (b),(c) i (d) onda se Bernsteinovim algoritmom za normalizaciju sintezom dobiva skup relacijskih shema koje ne zadovoljavaju Rissanenov test reverzibilnosti.

Dokaz:

Primijenimo li Bernsteinov algoritam za normalizaciju sintezom na skup funkcijskih zavisnosti **F** koji zadovoljava uvjete (a), (b),(c) i (d) tada imamo:

(1) Neka vrijedi $\mathbf{Z} \not\subset \mathbf{R}$
(2) Zamijenimo skup **F** sa skupom $\mathbf{F}' = \mathbf{F} \cup \{\mathbf{R} \rightarrow \mathbf{Z}\}$
(3) Reducirani neredundantni pokrivač:

Neredundantni pokrivač:

Funkcijska zavisnost $\mathbf{X_r} \rightarrow \mathbf{Y_r}$ u postupku traženja neredundantnog reduciranog pokrivača neće biti izbačena iz skupa funkcijskih zavisnosti. Naime, kako njezina desna strana sadrži atribut $\mathbf{Y_{r0}}$ koji, prema (b), nije niti u jednoj drugoj preostaloj funkcijskoj zavisnosti u $\mathbf{F}'$, zaključujemo da ne može biti izvedena iz $\mathbf{F}' \setminus \{\mathbf{X_r} \rightarrow \mathbf{Y_r}\}$.

Lijevo reduciranje:

Skup $\mathbf{X_{r0}} \subset \mathbf{X_r} \setminus \mathbf{W}$ neće biti izbačen u postupku lijevog reduciranja **FZ** $\mathbf{X_r} \rightarrow \mathbf{Y_r}$ jer, prema (b) i (c), vrijedi $\mathbf{Y_{r0}} \not\subset (\mathbf{X_r} \setminus \mathbf{A})^+ \mid_{\mathbf{F}}$, niti za jedan $\mathbf{A} \in \mathbf{X_{r0}}$.

Skup $\mathbf{X_{s0}} \subset \mathbf{X_s}$ neće biti izbačen u postupku lijevog reduciranja **FZ** $\mathbf{X_s} \rightarrow \mathbf{Y_s}$ jer, prema (c) i (d), vrijedi $\mathbf{Y_{r0}} \not\subset (\mathbf{X_r} \setminus \mathbf{A})^+ \mid_{\mathbf{F}}$, niti za jedan $\mathbf{A} \in \mathbf{X_{r0}}$.

Dakle, lijevim reduciranjem **FZ** $\mathbf{X_s} \rightarrow \mathbf{Y_s}$ preostat će **FZ** $\mathbf{X'_s} \rightarrow \mathbf{Y_s}$ i vrijedit će $\mathbf{X_{s0}} \subset \mathbf{X'_s}$.

Skup $\mathbf{W} \subset \mathbf{X_r}$ neće biti izbačen u postupku lijevog reduciranja **FZ** $\mathbf{X_r} \rightarrow \mathbf{Y_r}$ jer, prema (a), (b) i (c), vrijedi $\mathbf{Y_{r0}} \not\subset (\mathbf{X_r} \setminus \mathbf{B})^+ \mid_{\mathbf{F}}$, niti za jedan $\mathbf{B} \in \mathbf{W}$.

Niti jedan atribut iz skupa $\mathbf{Y_{s0}}$ neće biti izbačen iz lijeve strane **FZ** $\mathbf{R} \rightarrow \mathbf{Z}$ jer, prema (b), osim u $\mathbf{X_r} \rightarrow \mathbf{Y_r}$, skup $\mathbf{X_{r0}}$ se ne nalazi niti u jednoj drugoj **FZ** $\mathbf{X_i} \rightarrow \mathbf{Y_i}$ ($\mathbf{i} \neq \mathbf{r}$).

Pokažimo da u postupku lijevog reduciranja neće biti izbačen skup atributa $\mathbf{X_{s0}}$ sa lijeve strane **FZ** $\mathbf{R} \rightarrow \mathbf{Z}$.

Ako bi bio izbačen neki atribut $\mathbf{A} \in \mathbf{X_{s0}}$, značilo bi da postoji $\mathbf{X} \subset \mathbf{R}$ za koji bi vrijedila **FZ** $\mathbf{X} \rightarrow \mathbf{A}$, što je nemoguće, jer se, prema (d), atributi iz $\mathbf{X_{s0}}$ ne nalaze na desnoj strani niti jedne **FZ** iz skupa **F**.

Pokažimo da će u postupku lijevog reduciranja biti izbačen skup atributa **W** sa lijeve strane **FZ** $\mathbf{R} \rightarrow \mathbf{Z}$.

Funkcijsku zavisnost **FZ R→Z** možemo pisati:

$\mathbf{X_1Y_1X_2Y_2...(X_r\backslash W)Y_r...\ X'_s\ Y_s...X_nY_nW\rightarrow Z}$

Prema (a) i (d), iz $\mathbf{X'_s\rightarrow Y_s}$ i $\mathbf{W\subset Y_s}$ slijedi $\mathbf{X'_s\rightarrow W}$. Prema tome, skup atributa **W** biti će u postupku lijevog reduciranja izbačen iz **R→Z** jer je **W** određen skupom atributa $\mathbf{X'_s}$. Dakle, lijevim reduciranjem **FZ R→Z** dobivamo **FZ** $\mathbf{R'\rightarrow Z}$, gdje skup atributa $\mathbf{R'}$ ne sadrži **W**.

Desno reduciranje:

U postupku desnog reduciranja, prema (b), skup $\mathbf{Y_{r0}}$ ne može biti izbačen iz skupa atributa sa desne strane **FZ** $\mathbf{X_r\rightarrow Y_r}$ jer se $\mathbf{Y_{r0}}$ nalazi samo u **FZ** $\mathbf{X_r\rightarrow Y_r}$ i niti u jednoj drugoj **FZ** $\mathbf{X_i\rightarrow Y_i}$ (**i≠j**).

U postupku desnog reduciranja niti jedan atribut $\mathbf{A\in W}$ ne može biti izbačen sa desne strane **FZ** $\mathbf{X_s\rightarrow Y_s}$ jer se **A** ne nalazi na desnoj strani niti jedne preostale funkcijske zavisnosti.

(4) i (5):

Zatvarač $\mathbf{X_r^+\,|_F}$, neće biti ekvivalentan niti jednom zatvaraču $\mathbf{X_i^+\,|_F}$, (i≠r) jer $\mathbf{X_r}$ sadrži skup $\mathbf{X_{r0}}$ koji nije u lijevoj strani niti jedne **FZ** $\mathbf{X_i\rightarrow Y_i}$ (**i≠r**). Iz toga slijedi **SFZ** $\mathbf{(X'_r)\rightarrow Y'_r}$ kojoj odgovara relacijska shema $\mathbf{R_r(X'_rY'_r)}$ s ključem $\mathbf{X'_r}$. Nadalje, pokazat ćemo da $\mathbf{R_r}$ nije povezana niti s jednom od preostalih relacijskih shema koje su dobivene u postupku normalizacije.

Najprije ćemo pokazati da relacijska shema $\mathbf{R_z}$, koja je generirana funkcijskom zavisnošću **R→Z**, nije povezana s relacijskom shemom $\mathbf{R_r}$. Naime, relacijska shema $\mathbf{R_z}$ sastoji se od svih atributa koji su preostali nakon lijevog reduciranja funkcijske zavisnosti **R→Z**. Ti atributi sačinjavaju ključ relacijske sheme $\mathbf{R_z}$. Kako $\mathbf{R_z}$ sadrži skup $\mathbf{X_{s0}}$ koji nije u $\mathbf{R_r}$, slijedi da $\mathbf{R_r}$ ne može sadržavati strani ključ koji pokazuje na $\mathbf{R_z}$. S druge strane, ključ od $\mathbf{R_r}$ sadrži atribut **W** koji nije u $\mathbf{R_z}$. Prema tome, $\mathbf{R_z}$ ne pokazuje preko stranog ključa na $\mathbf{R_r}$. Dakle, relacijske sheme $\mathbf{R_r}$ i $\mathbf{R_Z}$ nisu međusobno povezane.

Sada ćemo pokazati da niti jedna druga od preostalih relacijskih shema dobivenih u postupku normalizacije nije povezana preko stranog ključa s relacijskom shemom $\mathbf{R_r}$. Naime, niti jedna relacijska shema, osim relacijske sheme $\mathbf{R_r}$, ne sadrži atribute $\mathbf{X_{r0}}$ koji su dio ključa od $\mathbf{R_r}$ pa prema tome niti jedna relacijska shema ne može pokazivati sa svojim stranim ključem na $\mathbf{R_r}$. Vrijedi i obratno, tj. $\mathbf{R_r}$ ne pokazuje sa svojim stranim ključem niti na jednu relacijsku shemu $\mathbf{R_i}$

(**i≠r**) koja je dobivena u postupku normalizacije. Naime, ako bi $\mathbf{R_r}$ pokazivala na neku relacijsku shemu $\mathbf{R_i}$ (**i≠r**), to bi značilo da $\mathbf{R_r}$ sadrži ključ od $\mathbf{R_i}$. To je nemoguće jer je jedini skup atributa koji je zajednički nekoj relacijskoj shemi $\mathbf{R_i}$ (**i≠r**), skup atributa **W** koji je desni dio **FZ** $\mathbf{X_s \rightarrow Y_s}$, a atributi desne strane **FZ** ne mogu biti ključni atributi.

Primjer:

Neka je zadan skup **FZ F**={**CE→DF**, **AFG→BH**,**EI→CD**,**D→C**,**CD→EI**} na relacijskoj shemi **R**=(**ABCDEFGHI**).

Uvedimo supstitucije:

$\mathbf{X_r}$**=AFG**, $\mathbf{Y_r}$**=BH**, $\mathbf{X_s}$**=CE**, $\mathbf{Y_s}$**=DF**, **W=F**, $\mathbf{X_{r0}}$**=AG**, $\mathbf{Y_{r0}}$**=BH,** $\mathbf{X_{s0}}$**= CE**

Za $\mathbf{X_r}$, $\mathbf{Y_r}$, $\mathbf{X_s}$, $\mathbf{Y_s}$, **W,** $\mathbf{X_{r0}}$, $\mathbf{Y_{r0}}$ i $\mathbf{X_{s0}}$ vrijede uvjeti iz prethodno dokazane tvrdnje. Prema tome zaključujemo da će se normalizacijom dobiti skup relacijskih shema koji će sadržavati relacijsku shemu $\mathbf{R_r}$(**X'Y'**) te da $\mathbf{R_r}$ neće biti povezana preko stranih ključeva niti s jednom od preostalih relacijskih shema koje će biti dobivene normalizacijom. Dakle, zaključujemo da će se normalizacijom relacijska shema **R** dekomponirati na skup relacijskih shema uz gubitak informacija.

Isti zaključak mogli smo dobiti normaliziranjem relacijske sheme **R** te provjerom Rissanenovih uvjeta reverzibilnosti. Normalizacijom se dobiva slijedeći skup relacijskih shema u **3NF**:

$\mathbf{R_1}$(**CDEFI**) s ključevima **CE**, **EI**, **D**
$\mathbf{R_2}$(**ABFGH**) s ključem **AFG**
$\mathbf{R_3}$(**AEGI**) s ključem **AEGI**

Međusobni presjeci relacijskih shema su:
1. $\mathbf{R_1 \cap R_2}$**=F** ne sadrži ključ niti iz $\mathbf{R_1}$ niti iz $\mathbf{R_2}$
2. $\mathbf{R_1 \cap R_3}$**=EI** sadrži ključ iz $\mathbf{R_1}$
3. $\mathbf{R_2 \cap R_3}$**=AG** ne sadrži ključ niti iz $\mathbf{R_2}$ niti iz $\mathbf{R_3}$

Iz **1.** i **3.** zaključujemo da relacijska shema $\mathbf{R_2}$ nije povezana niti s $\mathbf{R_1}$ niti s $\mathbf{R_3}$. Iz **2.** zaključujemo da je relacijska shema $\mathbf{R_3}$ povezana s $\mathbf{R_1}$. Prema Rissanenovim uvjetima slijedi da dekompozicija početne relacijske sheme **R** nije reverzibilna, odnosno da je tokom dekompozicije došlo do gubitka informacija. Može se primijetiti da relacija $\mathbf{R_2}$ u ovom primjeru odgovara relaciji $\mathbf{R_r}$ u tvrdnji. Prednost ispitivanja reverzibilnosti na skupu funkcijskih zavisnosti je u tome što se ne treba izvršiti postupak normalizacije relacijske sheme, tj. prije postupka

normalizacije možemo zaključiti hoćemo li izgubiti informacije ako relacijsku shemu normaliziramo.

Zaključak

Pomoću Rissanenovog kriterija nezavisnosti komponenata moguće je pokazati je li pri normalizaciji relacijske sheme došlo do gubitka informacija. Da bi se izvršio Rissanenov test, potrebno je imati komplet relacijskih shema koje se dobiju normalizacijom početne relacijske sheme. U radu je pokazano da je moguće provjeriti je li dekompozicija relacijske sheme ostvarena bez gubitka informacija, a da se prethodno nije morala izvršiti normalizacija, čime se dobiva na efikasnosti testiranja reverzibilnosti jer postupak normalizacije može biti dugotrajan.

Literatura

Bernstein, P. A. Synthesizing third normal form relations from functional dependencies. //ACM Transitions on Database Systems, 1, 1976, pp. 277-298

Date, S. J. An introduction to Database Systems, Add.-Wesley Pub. Co, 1995, New York.

Maier, D. The theory of Relational databases, Comp. sciences Press, Rockville, MA.

Mateljan, V. The Possibility of Applaying the Calculus of Functional Dependences to a Knowledge Base. // Journal of information and organizational sciences. 4 (2000) , 1; 69-81

Rissanen, J. Independent Components of Relations // ACM Transitions on Database Syste ms, Vol. 2, No. 4, December 1977, pp. 317-325

Tkalac, S.; Mateljan, V. Algorithm for Nodes Arrangement in Graphic Reprezentation of Functional Dependencies Set. // Informatologija. 24 (1992), 3-4; 101-108.

Tkalac, S.; Mateljan, V. Proposition for modification of Bernstein's Algorithm for Vertical Normalization by Synthesis. // Informatologija. 24 (1992) , 1; 1-11.

DOCUMENT MANAGEMENT IN BUSINESS APPLICATIONS AND E-GOVERNMENT

Improved Project Management Practices as a Key to the Successful Information System Implementation

Željka Požgaj
Faculty of Economics and Business, University of Zagreb
J.F.Kennedy Square 6, 10000 Zagreb, Croatia
E-mail: zpozgaj@efzg.hr

Hrvoje Sertić
Research & Development Center, Ericsson Nikola Tesla
Krapinska 45, 10000 Zagreb, Croatia
E-mail: hrvoje.sertic@ericsson.com,

Alen Škatar
Sales department, T-HT d.d. - Region 3 west
Ciottina 17a, 51000 Rijeka, Croatia
E-mail: alen.skatar@zg.t-com.hr

Summary

Project organization for the information system development is a highly demanding task. To enable efficient implementation, development project team must be able to capably perform development tasks and quickly respond to all sorts of change requests. Traditional project organization methods are no longer suitable to answer all challenges related to the information system implementation caused by the customer demands.
This work analyzes typical project organization practices and proposes method to organize information system development projects in a way to enable efficient use of available competences and provide a way to implement changes during the project execution. Proposed method is applied and evaluated on the actual development project to present all advantages of proposed approach.

Key words: software development, project management, information systems

Introduction

Companies today require highly complex information systems (IS) which are used as a main business foundation and primary connection with customers and partners. Development of such systems has changed significantly over the last several years: from small development organizations and just a few developers,

IS development currently is performed by complex development organizations with large number of developers, who are structured into complex organizations containing large projects split into several sub-projects executed by different development companies. Conditions on the global market force developers to implement IS in tremendously short time frames by utilizing most advanced technologies [Berkun, 2005]. The role of project manager has also changed significantly: from main technical developer, project manager has become primary driver of the project execution. This means that the project manager must be able to organize all aspects of project execution, from work on requirement collection, through implementation and testing to IS introduction on customer site. Ability to rapidly respond to customer's change requests has become one of the key project managers' skills and precondition for efficient IS implementation.

Amount of knowledge required to implement successful IS has also increased significantly: beside high amount of knowledge required to use advanced software technologies, software developers must be able to understand complete business environment in which customers operate in order to implement usable IS [McConnel, 1996]. Customers nowadays demand highly customized development according to their particular requirements; mass-produced IS can no longer be used to gain competitive advantage on today’s global market.

Typical development practices, in which IS requirements are frozen on the project start are no longer suitable, since environment in which IS will be used changes so rapidly that it is no longer to completely define required IS functionality in advance. Customers can no longer wait for the IS to be implemented to try its services and verify its behavior; instead they must be able to constantly verify implemented parts of IS during implementation and make corrections necessary to implement IS according to their particular needs. This has significant impact on project organization: to answer all customer requests, complete development project must be organized to deliver usable parts of IS as soon as possible to customers for verification, all change requests made by customers must be quickly accepted and implemented, and complete available competences must be efficiently utilized [Schwalbe, 2007].

This work analyzes typical development practices and project manager role in today’s complex project organization and planning, knowledge utilization and role assignments, requirement definition and change management, with particular focus on development improvements. Method for project organization according to the available competences aimed to enable rapid implementation of customer change requests is proposed and evaluated on actual development project to present all benefits which improved project management practices can bring to the IS development.

This paper is organized as follows: after the introduction, typical development practices and project manager role in these practices are analyzed. Third section presents method for project organization according to available competences,

while fourth section presents results of method application on actual development project. Conclusions are given at the end.

Project manager's role in development organization

Typical development organizations usually have some sort of formal organization, which defines high-level roles for most of development team. This means that typical development teams usually have predefined basic roles, so that it is, even before project start, on high level defined who will define system architecture, and who will perform design, test and integration activities. On the contrary, most project managers are assigned to the projects on the assignment basis, meaning that they get responsibility for execution of particular project or several project phases. Complexity and dedicated scope of today's IS causes development organization to frequently change development area, since it is not likely that any development organization can survive by implementing IS from just one dedicated area [Schwalbe, 2007]. This impacts organization of development project, since project manager must be able to analyze competences of available developers and assign temporary project roles in order to provide foundation for most efficient utilization of available competences. Proper role definition and assignment represent significant challenge on every development project: project manager must found a way to assign project roles according to the personal preferences, but roles must also be assigned in a way to enable efficient cooperation between development team members and proficient software implementation. Typical problems arise when project manager fails to correctly identify available competences and creates inefficient project organization. Amount of change requests that typical project faces during project execution should cause project managers to constantly monitor and adjust project organization [Wysocki, 2006].

Project planning

In stable environment, project planning is relatively simple. Project manager typically assigns analysis team who analyze requirements and customer needs, perform technical investigation and create implementation estimations. Stability of requirements, experience in the development field and used technologies, together with available amount of knowledge about actual customer needs directly determine quality of initial estimations and quality of related project plans [Kerzner, 2006]. Problems arise if the development team doesn’t have sufficient knowledge about development field, used technologies or customer needs and expectations. Project manager should be able to analyze quality of estimations he receives; project planning should in theory be based on accurate estimations created by technically competent developers. However, if the project is required to implement IS with advanced functionality using state of the art technologies, it is certain that most of development team will not be able to create correct estimations since they will have no experience in such environment.

Additional significant impact on project planning has customer's change requests and all changes project faces during execution. Even perfect project plans created on the project start have little value if project faces change requests so significant that the most of IS functionality should be altered [Berkun, 2005]. There are several strategies currently in use which propose IS development projects to be organized in several small development iterations aimed to implement just small parts of IS functionality [Schwalbe,2007], but there is no receipt that can be used to create quality project plans in the environment characterized with high amount of changes. Project organization and planning in such environment directly depends on project manager and development team skills in responding to change requests, adjusting project organization, roles, and development activities towards main goal: delivering changed parts of IS functionality to actual customers as soon as possible.

Quality of project plans is usually measured by the amount of details they contain and the value of estimations they are based on [Kerzner, 2006]. According to our experience, this is not sufficient to judge value of project plans. Often project managers create complicated plans on the project start which contain large number of planned activities which are not even presented to project members or adjusted through the project execution. Proper project plan must clearly define what should be done on the development project, but project plan must also be accepted by all involved persons. This means that project plan must be analyzed and commented by all project members, presented to all involved persons (including IS customers) and constantly monitored and adjusted during the project execution. Project plan must also be stored on common place, so that it is easy accessible for all project members to get information about overall project goals, particular responsibilities and timing for major project phases.

Knowledge utilization and role assignments

Extremely short time frames for today's IS implementation project demand from each project member to efficiently utilize complete set of available competences. Usual analysis of available competences and role assignments so that the roles are assigned to the most appropriate competences is not suitable any more. Short time frames in which IS development project must be prepared for execution (by setting up project organization, gathering initial requirements and analyzing future IS environment), demands that every single preparation activity must be performed as efficiently as possible [Davis, 2005]. Simple role assignments according to the information about available competences usually results with several problems: people are not often motivated to perform efficient activities since they were assigned roles similar to the previous project, their actual capabilities remain unused since they were not considered by role assignments, and most of all, described practice doesn't help to create successful development team since it focuses to personal competences. Knowledge which

project manager receives on the project start about available competences might be completely wrong; in lots of cases and situations development team members do not even receive a chance to perform activities which they are capable of due to the incorrect organization or other issues. If the project manager doesn't consider complete development environment, customer expectations, personal wishes and invest effort to create project organization that will enable each team member to perform activities in which he is interested in, project will face difficulties in execution which will have major impact on project preparation. Environment in which today's IS are implemented usually doesn't leave second chance for project preparation since most of projects simply do not have time and resources to repeat such activities. Major challenge for each project manager today becomes to create efficient project organization. From that perspective, it must be noted that implementation of successful IS under controlled development cost and resources is no longer major goal of the development project; instead, each development project must satisfy two major goals: (1) competence increase of all involved persons and (2) creation and advance of successful teams.

Requirement definition

Fulfillment of project requirements is one of the major responsibilities of each project manager. The role of requirements on development projects is often disregarded: requirements usually do not get properly defined, they are not used properly in development activities, test activities often do not verify which requirements are implemented, and customer changer requests do not update actual requirements used by development team [Davis, 2005]. Requirement discovery and definition so that they represent actual customer expectations is major task of each IS development project. Even properly defined requirements do not guarantee proper IS implementation according to the customer expectations; project must be able to systematically use requirements in each part of development activities to be sure that results of each activity contribute towards satisfying customer expectations. Requirements are often not properly structured: highly important requirements are mixed with not important ones; high level requirements are usually not properly represented with a set of low level requirements which completely clarify that requirement and so on. Projects often fail to properly on-time analyze defined requirements, so that in many cases development team late in execution discovers that the used requirements are incorrect or do not make sense. Proper requirement management practices on development project form a foundation for successful project, but they also require significant resources.

Change management

Changes have become integral part of each IS implementation project. Customers often demand implementation changes at very high frequency, developers

usually discover that the implemented parts of functionality are not properly implemented or that the quality of implemented parts is not sufficient. Typical cause for project failure is inappropriate approach for change management [McCulloch, 2005]. If project manager fails to gather complete control over every single change that is introduced to the project, project will end up in confusion and it will become very difficult to know which functionality is actually implemented. Even worse situation happens when development team members introduce changes on their own hand. To avoid problems related to the change request, and to create foundation for quick change implementation, project manager must establish environment that will enable project members with technical knowledge to accurately analyze each change request and create estimations including all consequences of change implementation. However, decision responsibility about each change request must remain on project manager; only in that way project manager can have control about all implemented changes which will enable IS implementation according to the (changed) customer expectations.

Method for improved project management

Activities required to prepare and execute IS development project in complex environment are presented on the Figure 1. Method is focused to two main parts of the IS implementation: (1) preparation of the development project and (2) execution of actual implementation activities. Method proposes activities that should be performed by project manager through the project. Several important artifacts are presented with respective activities in order to stress valuable outcome of the particular activities.

Project preparation

Proposed method starts with in-depth analysis of the environment in which project is to be executed. Project manager should analyze assignment he receives from its own organization or actual customer. Particular focus should be put into analysis of available competences and skills of project members. Activity aimed to analyze complete project environment is defined to actually analyze all issues that might impact project execution. Goal of that activity is to gather as much as possible information about all project stakeholders, market situation, particular customer, people involved from own organization and any other external influences. Project manager should focus to analysis of the project environment to define set of constraints in which project must be executed. These constraints should be used as a foundation for creating initial project setup and assignment of actual development activities.

In the preparation project phase, project manager should focus to create three efficient project teams: (1) Team responsible for requirement gathering and definition, (2) Team expected to define initial IS system architecture and (3) Team responsible to prepare and perform testing. Roles in the initial project or-

ganization should be assigned according to the competences and interests of project members, but constraints identified in analysis activities should be also considered. Together with initial project setup, project manager has responsibility to define project objectives and present them to all persons involved into the project. The following objectives should be presented to the persons involved in the project: (1) Project goals in terms of main IS functionality and project scope, (2) Project time schedule including exact planned dates for delivery to customer, (3) Estimated project cost, (4) Required competences with the focus to customer requirements and used technologies, (5) Resources required for project execution and (6) High-level competence development plan with presentation of expected competence levels.

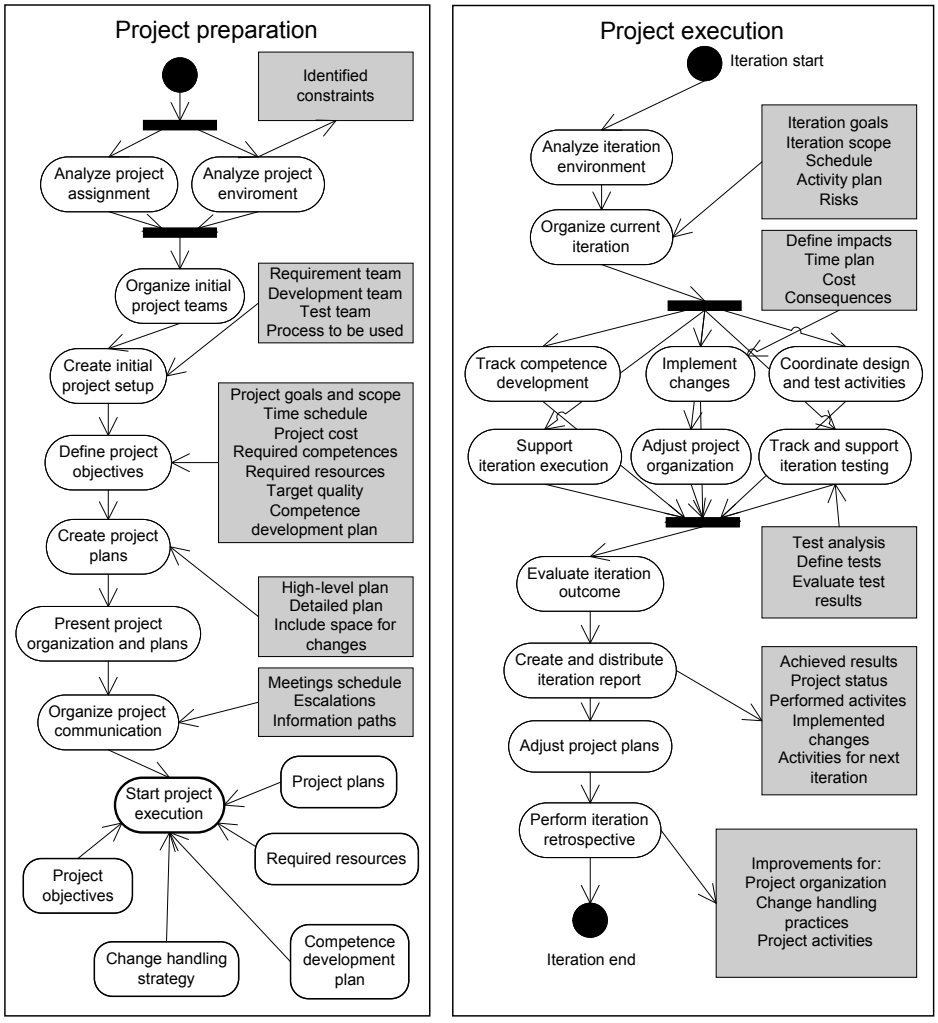


Figure 1: Method for improved project management

When most of project objectives are well-known, project manager and initial project teams can start working on creating project plans. Project plans should be defined on the following two levels: (1) High-level project plan with description of basic project phases, implementation iterations and definition of important project dates and (2) Detailed project plan with definition of particular development activities, their schedule and execution responsibilities. Project manager should assign project teams to create estimations for development activities, but actual project plans should be created only by project manager. It is crucial for complete project team to gain agreement and support for project plan, but creating utilizable project plan is responsibility of the project manager. Both project plans and project organization that will be responsible for execution of project plans should be presented to all involved persons. High-level time plan has the purpose to simply present actual project schedule as viewed from customer perspective to all involved persons, while the value of the detailed project plan is to inform project members about their responsibilities. Both project plans should include about 25% of time and resources for accustomizing changes received from the customer.
Prior to project execution, project manager should establish project communication practices by defining information paths that should be followed in common situation. Project manager should also propose change handling strategy by assigning roles responsible for analyzing change requests, creating and performing impact analysis and actual change implementation. This strategy should be connected with communication plan in order to enable most efficient information exchange in the case of received change requests. Methodology defines that the actual project execution can be started when the following is defined and agreed with complete project team: (1) Project objectives, (2) Project plans, (3) Change handling strategy, (4) Required competences and resources and (5) Competence development plans.

Project execution

Proposed method defines activities that should be performed by project manager in each development iteration. As with project preparation, method starts with environment analysis which is focused to the scope of the current iteration. Environment analysis should be aimed to identify major influences which can impact iteration execution. The following issues should be considered with the analysis: (1) Current situation in the project team, (2) Availability of required competences and resources, (3) Accuracy of project plans, and (4) Change requests which were issued prior to the current iteration. Information from these areas should be used to analyze current project organization and perform organization of current iteration by defining iteration goals, scope, detailed schedule and activity plan. Risk which might impact current iteration should be considered when creating iteration organization in terms that the planned activities are adjusted towards resolving most important risks. When the iteration imple-

mentation starts, the following tasks should be performed by project manager: (1) Track and adjust competence development of involved team members, (2) Organize, track and support implementation of change requests, (3) Coordinate design and test activities to minimize delays in testing by organizing development team to support testing and resolve all problems which occur during the test, (4) Support iteration execution by providing support to all project members and constantly improving project organization and (5) Track and support iteration testing. Method purposely defines separate activity for test support for project manager to stress the importance of iteration testing. Project manager should track test analysis activities, select tests that should be performed in the iteration and evaluate test results.

When the test team confirms that the iteration scope is implemented and when the test results are available, project manager should evaluate results of the performed iteration in terms of satisfied requirements and implemented change requests. It is the responsibility of project manager to confirm that the iteration has succeeded to implement planned scope and to decide if the iteration should be repeated or the project can proceed to further iteration. Project manager should create iteration report to represent current project status with description of all performed activities, achieved results, implemented changes, occurred problems and risks. Report should also describe achieved competence improvements and description of functionality that was added to the IS in the performed iteration, together with description of any functionality problems or limitations. Based on the iteration report and experience from the performed iteration, project manager should adjust project plans to reliably represent further project schedule and actual activities that should be performed. As a last activity in the current iteration, project manager should organize iteration retrospective, a project event where complete project team gathers to discuss achieved results, quality and efficiency of performed activities and all positive and negative experiences from iteration execution. Iteration retrospective should result with proposition for improvements in project organization, development and test activities, and change handling practices.

Results of method application

Proposed method was applied on the development project in the Ericsson Nikola Tesla Company for the development of IS aimed to provide statistics for mobile operator's network. Development project involved 24 developers located on four sites, with preparation phase and three development iterations. Focus on the environment analysis defined in the proposed methodology resulted with in-depth analysis of actual development constraints and several organizational decisions. Project objectives were defined on the project start and explained to all involved project members. By performing preparation activities as defined in the proposed methodology, project manager provided significant amount of information to project members, which were all aware of actual pro-

ject goals, schedule, their role and actual customer expectations at the project start. Each development iteration was organized according to the experiences from the previous one, and project plans were constantly adjusted (several times through iteration). Project faced significant number of change requests, which had major impact on project execution and required additional effort for implementation. By constantly supporting both design and test team, project manager enabled test team to perform all planned tests despite occurred problems in the code design. Iteration retrospectives were performed after every iteration and enabled development team to understand all decisions performed during the iteration. Several valuable improvements in project organization were defined according to the retrospective conclusions.

Conclusion

Current situation in the field of information systems development demands most efficient utilization of available competences and well-organized development activities. Traditional project management practices are no longer suitable, since the amount of change requests faced by the typical project demands from the project manager to actively participate in all project activities and, beside usual project management activities, to coordinate and adjust development and test activities. This work has analyzed areas of project organization to propose method that will enable project managers to focus to the important organizational issues and create efficient project organization that will enable all project members to increase own competences. Quick response to change requests is provided by defined activities which should be performed by project manager and project team members. Results achieved with method application on actual development project clearly show that the project organization can be constantly improved through the project, therefore resulting with better implementation of information systems.

References

Berkun Scott. The Art of Project Management. New York: O'Reilly, 2005

Davis, Alan Mark. Just Enough Requirements Management: Where Software Development Meets Marketing Washington: Dorset House, 2005

McConnel, Steve. Rapid Development, Redmond: Microsoft Press, 1996

McCulloch Erika; Colleen Garton, Fundamentals Of Project Management. Chicago: McPress, 2005

Kerzner Harold, Project Management: A Systems Approach to Planning, Scheduling, and Controlling. New York: Wiley, 2006

Schwalbe Kathy. Information Technology Project Management, Fifth Edition. London: Course Technology, 2007

Wysocki Robert K. Effective Project Management: Traditional, Adaptive, Extreme. New York: Wiley, 2006

Managing Knowledge in the Electric Power Production Sector

Vjekoslav Afrić
Department of Sociology, Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
vafric@ffzg.hr

Alfredo Višković
HEP (Croatian Electro Power Production Company), Zagreb, Croatia
Ulica grada Vukovara 37, Zagreb
alfredo.viskovic@hep.hr

Summary

Knowledge management is a topic of increasing involvements in strategic development of the Electric Power Production Sector in Croatia particularly because of the recent emergence of unification and integration of the European Electric Power market. New ways of thinking about management and organization are a key for Croatian participation in the European Union and in an integrated European Power market. Management of knowledge is the most important point of new sustain development towards the appropriate position of Croatian Electric power production sector in the European integration processes.
The awareness of importance of processing and managing knowledge is of vast importance as a focus on an application capacity of information science. It is a means to enable establishing hard connections between business activities and the development of information sciences.

Key words: Knowledge management, Unification and integration of European Electric power sector, applicative capacity of Information science

Knowledge management is a topic of increasing involvements in strategic development of the Electric Power Production Sector in Croatia particularly because of the recent emergence of unification and integration of the European Electric Power market. New ways of thinking about management and organization are a key for Croatian participation in the European Union and in an integrated European Power market. Management of knowledge is the most important point of new sustain development towards the appropriate position of

Croatian Electric power production sector in the European integration processes.

The awareness of importance of processing and managing knowledge is of vast importance as a focus on an application capacity of information science. It is a means to enable establishing hard connections between business activities and the development of information sciences.

Contemporary world is characterized by sharp development of service sector, market globalization, and technology standards unifications. Knowledge use to have some importance in industrial production from very beginning, but in service sector of economy, knowledge is sine qua non, and that is the main reason because in contemporary market, knowledge and innovation have a most important role ever. Knowledge and innovation today represent increasingly important intangible capital as the most important origin of profit in global market economy.

Intangible or intellectual capital is a subject of new business approach to maintain production and profit by pushing innovative behavior to empower competitiveness, to build consumer trust (brand) and in one word to be successful. Manage intangible capital is managing knowledge, because knowledge is a very core of intangible capital.

Under the new economic circumstances the development of the HEP Group must be well-prepared from the global aspect and the aspect of the local environment in which the corporation operates, as well as from the aspect of the necessary changes in the organization of the business operation to enable the utilization of HEP’s intellectual potentials.

The position of HEP is determined, on the one hand, by the local requirement of further investment in the generation of electricity and of such a price of electricity that will stimulate economic development in Croatia, and in wider region, whereas on the other hand, it is determined by the process of Croatia’s association with the European Union and the directives of the European Union on the common rules for the generation, transmission and distribution of electricity and gas aimed at establishing the European electricity and gas market.

The European Parliament and the European Council have passed Directives 2003/54/EC and 2003/55/EC on the common rules for the single EU electricity and gas market, making them obligatory for all Member States and for candidate countries. The implementation of these two Directives will make the European Union the best integrated energy market world-wide. These two Directives adopted in 2003 are a great step forward towards creating a single electricity and gas market in EU. These documents, which oblige all Member States, set the deadlines for fully opening their individual markets: it was 1 July 2004 for all commercial actors, and 1 July 2007 for all households. All Member States and candidate countries are under obligation to notify the European Commission about the implementation prior to the expiration of the deadlines agreed. The Directives set up common rules for the generation, transmission and distri-

bution of electricity, treated as independent business sectors. Consequently, the Directives provide for strengthening the system of transmission and its independence from other activities (generation and distribution) by legally and functionally separating those activities.

The legal basis of the Directives also entails numerous obligations to be met by national regulations, so national regulators must carefully follow the development of the competition and the levels of investment, and regulate the price level when necessary. This will lead to a much better transparency and provide European companies in the electric energy sector with many more possibilities to forecast, plan and design their own evolution.

There is still much to do for Europe to have an effective and competitive energy market, and it is, therefore, essential that Member States develop, as soon as possible, the necessary legal framework to implement the Directives. Of course, the regulations are only a working framework to enable competitiveness. It is necessary to make a further effort and deal with the nationally dominant position of traditional energy suppliers. The European Commission will continue to monitor the development of the market for some time in the future and to promote new initiatives to make sure the energy market functions as envisaged. National regulations will not be abolished overnight; on the contrary, they will play the central role in effectively establishing the market. It is probable that the European commission will increasingly transfer the responsibility for monitoring and promoting market development to national regulators, eventually.

Electricity and gas are not like other products, they have little, if any, real substitutes. Continuous power supply and its reasonable price are essential to the national economy and to building a modern society. Moreover, electricity and gas markets are most often determined by the framework of national markets and are subject to the domination of one or a small number of companies. New Directives take note of these facts. The introduction of competition does not include a total deregulation of power supply, or a laissez-faire concerning the market fundamentals in terms of prices and services. On the contrary, the Directives require Member States to continuously monitor the market and impose numerous additional obligations to power companies, if need be, concerning the services provided to the public. Such a security is very important during the transition period from fully regulated monopolies towards the situation in which the market will determine the focuses in the generation of electricity such as prices and investment decisions.

New ways of thinking about management and organization are a key for Croatian participation in the European Union and in an integrated European Power market. Management of knowledge is the most important point of new sustain development towards the appropriate position of Croatian Electric power production sector in the European integration processes.

For European companies developing a system of knowledge management is not only a way to their own sustainable development but also a way to the estab-

lishment of the common energy market (Pomeda, J. R., Camacho, 2005). A common electricity and gas market requires the closest technological cooperation and an unimpeded exchange of information. In this sense the development of a system of knowledge management in companies engaging in electricity and gas has a twofold purpose:
- on the micro level, in an individual organization, a correctly organized system of knowledge management is an effective aggregate of innovation, and an appropriate tool to put the innovation emerging from the processes of knowledge sharing and generation to the use of sustainable development of the organization,
-on the macro level, such as the European energy sector, knowledge management is a powerful system for mutual support in the transfer of information's and knowledge with which the single energy market is maintained.
In the energy sector knowledge and experience are cumulated in long years of application practice, which requires constantly keeping the system active. Examples from the practice confirm that it takes about seven years of theoretical and practical education to train an energy specialist to be able to participate in the system, and a life-time education to operate it. This surely shows the importance of knowledge management on the micro level. Close technological cooperation and unimpeded exchange of information are conditions to the existence of the energy market which requires technological compatibility. This is achieved through joint projects of power companies and mutually standardized permanent education and implementation of the knowledge acquired.
The transfer of knowledge and technologies is not a simple process and it will not run spontaneously and randomly. This process requires a developed system of knowledge management that will encourage and manage the processes of production and exchange of corporate knowledge both internally within own organization and externally by shaping the cooperation on the international level, enabling synergy through a collaborative approach to support the development of the single market. Knowledge management is a useful tool in transfer "the best practices" and encouragement of innovation (Pomeda, J. R., Camacho, 2005).
Innovation activities must be incorporated in everyday activities (the phenomenon known as "reutilization") to lower uncertainty in the processes (Jacono, 2000). Particularly in the electricity sector, competitive requirements and the existing pressure of the market underscore the use of systematized approach to innovation processes, trying as much as possible to reduce uncertainties caused by the opening up of the market. In case of technological innovations, there is a practical need to implement the knowledge of the groups of individuals who are solving the practical problems of such technologies. Technology and innovation are becoming key factors of success by re-activating the transformation process and constituting the basic source of the company's competitiveness. However, J. R. Pomeda (Pomeda, J. R., Camacho, 2005) claims that it is not just one's

own innovation that generates the competitive advantage, particularly with regard to the energy market which is based on broad cooperation; what is required is adaptation to the technologies developed by others. The adaptation to new technologies requires a high level of learning including the processes of technological awareness i.e. the processes of adopting and creating knowledge (McElroy, 2002). Knowledge management is a process of generating new knowledge that acts as a catalyst for innovation and creativity, and a process facilitating the collection and influence of knowledge across the industry. "Best practices" cannot possibly be transferred if there are no appropriate processes to make it possible to learn them. It is only through the knowledge management that the efficiency of business approaches incorporating industry knowledge can be successfully transferred (Von Krogh G., Ichijo K., Nonaka I., 2000).

In order to become the leading corporation in this area, HEP must complete the restructuring processes, introducing effective systems for making and implementing decisions, creating merger and takeover department, managing risks, developing and encouraging marketing, defining frameworks for negotiations with partners in the region, in short, HEP must fully use its intellectual potentials to hold its position and to secure its development through building its position on the regional market.

The maximum utilization of the intellectual potentials is only possible by developing a strategy for knowledge management at HEP, and its implementation. This strategy has been adopted in year 2005, by the HEP Group, and its implementation is under way.

HEP possesses the experience and knowledge with which to achieve sustainable development and encourage development in its environment by providing consultancy and other intellectual and technological services, provided it can systematically manage its knowledge.

Creating new knowledge and, perhaps more importantly, utilizing the available knowledge that is there in the organization (Nonaka, I.; Takeuchi H.; 1995), should become the central element of the contemporary business strategy of HEP.

By trying to master the complex processes of knowledge management HEP developed and adopted the strategy of knowledge management. This strategy is the operationalization of HEP's corporate vision by means of which HEP defines its development goals in terms of the growing demand for electricity in Croatia and in terms of the transformation of HEP into a strong regional electric power company optimally harmonized with its environment and strongly partaking in shaping the development of the regional electric power system, capable of utilizing its competitive advantages on the electricity and gas market.

The strategic plan for knowledge management in HEP is a means to secure the strategic approach in decision-making, notably a means to secure human competence in managing business processes and in technically shaping all the undertakings of HEP (Davenport, T. H; Prusak L.; 1998). This plan defines the

functional and technical aspects that can help shape the business space and successful implementation of initiatives in knowledge management in accordance with the Knowledge Management Strategy of HEP. It describes the basic components necessary for implementing a successful KM strategy and it notes the specific joint knowledge management goals of HEP.

The knowledge management strategy is defined as a tool to support the strategic business plan of HEP (Hansen, M. T.; Noria N.; Tierney T.; 1999). The implementation of the strategic business plan of HEP is based on the real power of HEP’s organizational and intellectual resources. These resources depend, to a great extent, on creating and maintaining knowledge bases; on the possibility to attract, train and maintain high capabilities of the staff; and their expertise in utilizing these knowledge bases (Gupta, A.K.; Govindarajan, V.; 2000). The development of shaping the basic business processes and the realization of the strategic business plan rely on the knowledge of people, on the content of such a knowledge, and on the technology by which this knowledge is shared, or in short, on the knowledge management program.

The knowledge management program of HEP includes three dimensions: people, content and technology (Grundling, E.; 2000).

People: those who generate and use knowledge as the basis for making appropriate decisions (Wenger, E.; 1998).

Content: data, information and knowledge relevant to the organization, including processes and procedures (Afrić, V.; Lasić-Lazić J.; Banek Z. M.; 2004), (Storck, J., Hill, P; 2000).

Technology: technical infrastructure and tools making it possible to record, save and deliver the content of knowledge to those who need it when they need it (Borghoff, U. M.; 1999).

The knowledge management program practically includes the entire workforce of HEP. The integral knowledge management program requires the establishment of the business role and responsibility within the human dimension, notably the hierarchy that enables the introduction of management mechanisms in knowledge management processes. Defining new business roles is part of the general restructuring of HEP. Defining roles is not an easy task, it is very important because it enables flexibility at the very heart of the management and practice of the organizational power and politics. Defining business roles also includes defining knowledge communities.

A knowledge community includes those who work on joint groups of tasks within a project or a mission concerning development, operating or implementation plans, application support, maintenance service, project management etc. These common tasks are carried out by teams (knowledge communities) as parts of the organization in which people play individual roles based on their capabilities and skills, rather than on titles and functions (Ruggles, R.; 1997). Teams of practitioners may extend across several different sectors of activity, consisting of individuals from different parts of HEP, and from different areas

of Croatia. It is in this sense that a person may be included in several different knowledge communities as a member of different teams and projects.

Introducing the knowledge management practice and tools in any organization often requires cultural changes, e.g. from the Knowledge-is-Power frame of mind to a culture of knowledge sharing. Without motivation factors it is difficult to have people change the way they do things. There is the fear factor when employees fear their knowledge will be taken from them and they will become irrelevant. A change in behavior may only be brought about when there are rewards for accepting the transition and a shift in focus to a future situation. Pecuniary and non-pecuniary rewards also encourage the effort to accept the desired culture in the organization. This will happen if people are aware that by learning and making effort to contribute with their knowledge to the knowledge of the organization they will be better valued within their company. Creating a culture of mutual trust, teamwork and a strong feeling of common identity, is essential to introducing the knowledge management practice in HEP.

The roles and responsibilities in HEP's knowledge management will vary from informal to highly formal approaches. In most cases the attributed tasks will be treated as related duties. However, responsibilities, managing the content of knowledge, developing tools, maintaining the portal, training for knowledge management etc. will be attributed as formal duties established through the organization as part of the functional business structure. Every member of the knowledge community will have his/her own role in creating the knowledge base. The roles of older and more experienced members of the knowledge community will also be particularly valued in the evaluation and in the generation of new knowledge. In HEP, knowledge management will be institutionalized through virtual organization - the organization whose logic pursues quick growth and enlargement of autonomous networks to improve our business capabilities to know what we know, use and coordinate what we know, and learn something new. As in biological systems, this virtual organization will grow and adapt to its surroundings, rejecting the efforts and capabilities that are not productive, reinforcing and encouraging the initiatives that prove valuable (Stacey, R. D.; 1996). In this virtual way, HEP already used to organize project teams which, as organized groups, only existed at the time of preparing specific projects. This joining together requires a continuous effort, because knowledge management is a journey, rather than a point of destination. Implementing the knowledge management strategy at HEP means care about the human, social and/or cultural capital of HEP.

Human capital is, for the best part, what is meant by intellectual capital, it is hidden knowledge residing in the mind of each employee, same as the future capacity and potential for learning that everyone has. Hidden knowledge includes skills, experience, understanding, intuition and judgment, combined to shape the past and the present. The potential of each employee is an important part of the work resources of HEP.

Social capital is the intellectual capital derived from communications, cooperation and human interrelations (Yli-Renko, H.; Autio E., Sapienza H. J.; 2001). It includes people and virtual networks of social relationships, the relations and interactions throughout such networks and the principles on which such relations are built. The social capital is an expression of the corporate identity and culture. The culture of the organization (cultural capital) is a pattern of beliefs, knowledge, attitudes, and norms of behavior and customs that exist in the organization. A developed culture of the organization defines a developed corporate identity, a strong WE-feeling that motivates people to mutual communication, cooperation and good human interrelations, which results in an increased value of knowledge resources.

The content includes the data, information and knowledge relevant to the organization, as well as processes and procedures. The content includes structured and non-structured information such as databases, magazine subscriptions, websites, news, e-mail, documents, notes, PDF files, summaries, spreadsheets, audio recordings, video recordings, bookmarks, LAN folders, forms, GIS data, transcripts of electronic communication, project sketches, product catalogues, rules, drawings, photographs, graphics etc. Various data, in particular the ones related to energy sector, from the randomly collected to the scientifically systematized or scientific ones, are very important in HEP's business operation. They make up the content of the knowledge that can improve HEP's business operation.

HEP is not interested in entire content of the knowledge created inside and outside HEP, but only in the content that makes the knowledge worth managing. For a content to become the knowledge worth managing, it must belong to HEP's business context. A good management of large quantities of knowledge content significantly reduces information overload, enabling us to have just the information we need, when we need it. The organization, or its knowledge community, cannot manage knowledge unless it is aware of what knowledge it needs and what knowledge it possesses. This is what knowledge maps are for, integrating company specific technologies, required data and priorities into a taxonomy that's serves as a guide to the conceptual and physical organization of information resources, experts, data, information and processes.

Knowledge audits, as content analysis techniques, are a means with which HEP will build knowledge maps for individual life cycles of knowledge (Solomon, P.; 2002) in order to transform non-structured content into data, information or knowledge, and to determine the description of that knowledge in terms of media object attributes, the attribute structure, and the rules for relations between attributes.

Knowledge audit helps the company establish a common vocabulary which is necessary to successfully communicate and cooperate. The common vocabulary of HEP is yet to be created.

Today's Web-centered environment and the geographic dispersion of HEP's workforce leads to the assumption that most HEP's knowledge management practitioners will be using computer-based online environment and render their intellectual contributions through online depositories. The strategy, therefore, discusses, in the most general way, the technologies supporting knowledge management.

The knowledge management development at HEP calls for an organized storage of the data collected and produced by HEP's staff, notably for a further integration of the already stored data. The entire content managed by content management systems can be divided into: 1.) structured data, information, and knowledge, as well as applications generating, managing and maintaining such structured resources, and 2.) non-structured or semi-structured data, information and knowledge (including different types of multimedia objects), as well as applications generating, managing and maintaining them. As opposed to the structured data, the non-structured content has no standardized meta-data structure and there are no standardized procedures for examining, searching or analyzing them. The analysis of the management of structured data/information/knowledge shows a characteristic increase in complexity and integration of this area, as well as the existence of two primary areas of structured content in which the rising trend towards integration and complexity is noticeable (data stored and resources planned). Whereas the data stored are structured by means of a meta-data structure, the resources planned are most often relation databases designed to store particular types of data.

Online portal will be the main tool used in the institutionalization of HEP's knowledge management. In a virtual online space three types of online portals (Firestone, J. M; 2003) will be set up: decision-making portals, collaborative portals and knowledge portals.

Since teams must work together and since persons can be involved more teams simultaneously, knowledge management tools will be based on mutually harmonized standards. HEP will set up and maintain one horizontal knowledge portal and as many vertical knowledge portals as there are project initiatives that may be launched by knowledge management. Standardization (Knowledge Management Handbook, 1999) being essential to the success of the knowledge management program, HEP will adopt a single user interface with intuitive navigation and taxonomy tools that can be readily accepted by all the real and potential users of HEP's portals.

The general structure of HEP's knowledge map will be defined prior to determining the specific interest communities and, additionally, it must define the geographic spread of HEP to every village or town. The general interest communities, like other organization elements, can share approach to the resources through such knowledge maps of the company.

For HEP, the knowledge management strategy means focus on key business needs, further improvement of the project approach, construction of collabora-

tive business culture, stimulating innovation, decentralization of development resources with simultaneous adoption of common standards and purposeful restructuring of the formal organization in accordance with the need to achieve the business vision, the mission and the strategic programs of HEP.

The application of the knowledge management strategy at HEP will enable further restructuring of HEP towards HEP's vision and mission and thereby defined basic values of HEP as strong modern company with market focus, promoting improvement and innovation to achieve the maximum quality under the criteria of profitability. A company that guarantees its staff opportunity to develop on the basis of their competence and professional contribution, i.e. developing collaborative business structure oriented towards its users to whom it guarantees security and reliability of supply and service, providing competitive and quality solutions, caring for the people and healthy natural environment, the development and the prosperity of the entire region in which it operates

The awareness of importance of processing and managing knowledge is of vast importance as a focus on an application capacity of information science. The further development of Information science in Croatia is one of the corner stone of Croatian business development in contemporary circumstances is closely connected with growing of Informational sciences, not only for better processing knowledge, information and data; but for building on knowledge grounded business culture. Competitive potentiality of HEP, and of Croatian power industry, and whole Croatian economy is hard linked with applicability potential of information science in Croatia, and all economy sector must find a way to support that development or to enable establishing hard connections between business activities and the development of information sciences.

References

Afrić, V.; Lasić-Lazić J.; Banek Z. M.: Znanje, učenje i upravljanje znanjem; published in Odabrana poglavlja iz organizacije znanja, Zagreb: Zavod za informacijske studije, 2004

Borghoff, U. M., Pareschi R., Information technology for KM, Berlin: Springer, 1999

Davenport, T. H; Prusak L. Working Knowledge. Harvard: Harvard Business School Press, 1998

Firestone, J. M., Enterprise Information Portals and Knowledge Management, USA: Butterworth-Heinemann, 2003

Grundling, E., The 3M Way to Innovation, in Balancing People and Profit, Tokyo: Kodansha International, 2000

Gupta, A. K.; Govindarajan, V. Knowledge management's Social Dimension, USA: MIT Sloan Management Review, 2000

Hansen, M. T.; Noria N.: Tierney T. What's your Strategy for Managing Knowledge?, Harvard: Harvard Business Journal, 1999

Jacono, G. L'organizzazione basata sulla conoscenza. Verso l'applicazione del KM in azianda. Milano: Franco Angeli, 2000

Knowledge Management Handbook, Edited by Jay Leibowitz, Boca Raton: CRC Press, 1999

McElroy, M. W. The New Knowledge Management, Complexity, Learning and Sustainable Innovation, USA: Butterworth-Heinemann, 2002

Nonaka, I.; Takeuchi H. The Knowledge-Creating Company. Oxford: Oxford University Press, 1995

Pomeda, J. R.,. Camacho, C. Electricity Industry Regulation and Innovation: Benchmarking and Knowledge Management as Appraisal Tools, Madrid: Instituto Universitario ce Administracion de Empresas, Universidad Autonoma de Madrid, Spain, 2005

Ruggles, R., Why Knowledge? Why Now?, 1997, Online Version. Retrieved 08, 09, 2007, from the World Wide Web: http://www.providersedge.com/docs/km_articles/Why_Knowledge_Why_Now.pdf

Solomon, P., Discovering information in Context, Published in: Annual Review of Information Science and Technology, Volume 36, 2002

Stacey, R. D., Complexity and Creativity in Organizations, San Francisco: Berrett-Koehler Publishers, 1996

Storck, J., Hill, P., Knowledge Diffusion Through Strategic Communities. USA: MIT Sloan Management Review, 2000,

Von Krogh G., Ichijo K., Nonaka I. Enabling Knowledge Creation, Oxford: Oxford University press, 2000

Wenger, E. Communities of Practice. Learning, Meaning and Identity. Cambridge: Cambridge University Press 1998

Yli-Renko, H.; Autio E., Sapienza H. J. Social Capital, Knowledge Acquisition and Knowledge Exploitation in Young Technology-Based Firms. USA: Strategic Management Journal, n. 22, 2001

The Issues of Strategic Dissemination of Information in Croatian Tourism

Đani Bunja
Department of Information and Communication Studies
University of Zadar
Dr. Franje Tudjmana 24 I, Zadar, Croatia
gianni@unizd.hr

Marijana Ražnjević
Department of Information and Communication Studies
University of Zadar
Dr. Franje Tudjmana 24 I, Zadar, Croatia
mraznje1@net.hr

Pavica Nekić
Department of Information and Communication Studies
University of Zadar
Dr. Franje Tudjmana 24 I, Zadar, Croatia
pavica0@net.hr

Summary

The paper deals with the problems of strategic dissemination of information in Croatian tourism. The introductory part treats the planning or, precisely, the strategic planning as a generic activity of management. Furthermore, it shows the results of a questionnaire carried out among Croatian tourism management experts about determining the messages disseminated in the tourism market, as much as in-depth interviews about determining the information disseminated in the tourism market. In addition, the paper deals with the issues of formulating, implementing and evaluating the strategies of information dissemination in Croatian tourism. In the concluding part it is asserted that tourist destinations must pay necessary attention to the collecting, processing, storing and disseminating of information, which should result with adequate messages, later conveyed to the tourism market.

Keywords: information, message, dissemination, strategy, Croatian tourism.

Introduction

The future is the source of numerous opportunities and dangers. Therefore, the task for all those who bear responsibility of decision making, i.e. management, is, among else, to anticipate the future and to plan the changes which will be taking effect in the years to come.[1] Planning is one of the basic functions of management. It includes the defining of organization objectives and the ways of reaching them, and it concerns the decision-making in all stages. Planning is a generic activity, whose aim is growth and development. The aim is a desired condition achieved by planning, so its purpose is to define priorities. Planning includes a future period necessary for a series of actions aimed at realizing the strategy comprised in today's decisions.[2] Accordingly, management of tourist destinations aimed at quantity and quality improvement, and growth and development of tourism, seeks to define a strategy, to plan future actions efficiently and to envisage trends, phenomena and events in organizational environment. Strategy can be defined as a means of accomplishing the goal. Strategy can be described as a direction chosen by the organization in order to accomplish the goal. According to Mintzberg[3] there are five ways of using the word strategy: plan – deliberately chosen the course of actions; endeavour – a certain manoeuvre for the purpose of outwitting the rival; pattern – in a series of actions; position – the means of positioning the organization into environment; perspective – an integrative way of perceiving the world. Mintzberg observes the links between the five ways of using the word strategy and concludes that it is necessary to explore various perspectives of a certain organization and its activities described by each of the words. Strategy is based on determining the basic and long-term objectives of a specific project and on the analysis of organizational environment, which should be formalized in the form of a strategic plan. Management should try to avoid crisis situations by pondering upon long-term consequences of each activity. Therefore, it is necessary to ensure guidelines as a basis of not only future activities but also of avoiding unpleasant surprises in case of unforeseen developments. Eventually, it should be pointed out that a quality strategy results in a better acceptance of the organization identity by its actual and potential consumers, and can strengthen the competitive position compared to other organizations.

Research methodology

In its empirical part the research was carried out by means of questionnaires and in-depth interviews. The respondents in the questionnaire were supposed to use

[1] Srića, V., Inventivni menedžer, Croman & MEP Consult, Zagreb, 1994, p.20

[2] Griffin, R.W., Management 3rd ed., Houghton Mifflin Company, Boston, 1990., pp.158-189.

[3] Mintzberg, H., The Rise and Fall of Strategic Planning, Prentice Hall, Englewood Cliffs, New Jersey, 1994.

the following scale: 1- absolutely not, 2 – mostly not, 3 – neither yes nor no, 4 – mostly yes, 5 – definitely yes. In this research there was not a sampling method, but the whole population was interviewed since the respondents were the representatives of tourist boards on a specific level. Although the regulations related to founding, structuring, working methods, objectives, tasks, financing, managing and other issues essential for a proper functioning of the tourist board system, stipulate the obligation for the local government to found tourist boards and to make them operational – this is not the case. According to the data obtained from the Direction of the Development of Tourist Boards and Selective Types of Tourism, controlled by the Ministry of the Sea, Tourism, Transport and Development of the Republic of Croatia, two counties, seventeen cities, and forty-eight municipalities did not establish or activate their tourist boards in the time of the research. The system comprises 328 tourist boards, among which 257 or 78% are operational. Consequently, the non-operational tourist boards from the system do not belong to the population and for that matter they are not relevant for this research. In carrying out the research and interviews, tourist board directors took active part. The questionnaire was filled in by 142 representatives of tourist boards on all levels. This meant 55% of the tourist boards on all levels concerning the whole population.

The in-depth interview was carried out with 7 respondents – the representative of the Ministry of the Sea, Tourism, Transport and Development, the representative of the Croatian Tourist Board and other five representatives of various tourist boards: of the Zadar County and of the cities of Poreč, Opatija, Split, and Dubrovnik. The survey was carried out in cooperation of the Ministry of the Sea, Tourism, Transport and Development and the Croatian Tourist Board's main office. The questionnaires were delivered by the Ministry to the tourist board offices and when filled in later collected by the Croatian Tourist Board's main office and forwarded to the Ministry. The Ministry sent on the filled in questionnaires for the data processing. The research was carried out within the intervals of 45 days, namely from 15 July to 31 August 2004. The data from the questionnaires were brought into a specially constructed electronic matrix and then processed by the computer programme *Statistica 6.0* (Statsoft). The in-depth interview was carried out directly or by phone. The collected data were processed within the intervals of 15 days, namely from 1 to 15 September 2006.

Research results

The questionnaire contained two statements: (1) it is necessary to determine only one message for the specific market and (2) it is necessary to determine more different messages for the specific market. The results of the research for the above mentioned statements (by ordinal numbers) are presented in the following table:

Table 1: The standpoints of Croatian tourist management about determining the messages disseminated on tourism market

statement	number of respondents	medium value	minimum	maximum	standard deviation
1.	142	3.106	1	5	1.230
2.	142	3.472	1	5	1.213

It results from the questionnaire that Croatian tourist management, namely the directors of tourist board offices, in relation to both statements – *it is necessary to determine only one message for the specific market* and *it is necessary to determine more different messages for the specific market* – take ambivalent standpoints. This implies the need for reassessing the efficiency of the present practice of determining key messages disseminated by our tourist management on the market. It also leads to the conclusion that it is possible to improve the present practice of their determining in Croatian tourism. The in-depth interview contained the questions about determining the information disseminated on the tourism market. The results of the in-depth interview show that the determining of information disseminated on the tourism market is carried out on the basis of discussions led within the bodies of tourist boards. In certain cases it is carried out in cooperation with other exponents of tourism policy, experts in the field of communication and specialized public relations firms. Very often, they are determined by public calls.

Formulating the strategy of information dissemination

Efficient communication should be based on the objective evaluation of the position of tourist destination on the market. Regardless the research efficiency, i.e. collecting and processing of information, very important is the attentive observation of changes in the environment. Due to over saturation by information, which affects not only individuals but organizations as well, communication becomes an important factor, on which the efficiency of tourist destination usually depends. The objective insight into the position on the tourism market can help develop strategies of information dissemination, which will make it possible for a specific tourist destination to form its characteristic identity. Before determining the strategy, it is important to know how the tourism market or a group of tourist and potential tourist perceive a tourist destination, i.e. what attributes they use to describe it. The ways of discovering the related information include the carrying out of quality and quantity research of tourists' attitudes, the analysis of former media coverage of a specific destination and its competition, together with internal brainstorming of organization members concerning the tourist destination. After considering the state of affairs regarding internal and external environment, we can continue with setting the goals for the strategy of information dissemination. This is the stage in which we should evaluate the results of the communication process – the reaction of the market on the level of

efficiency of the process of transferring the information to tourists' decisions concerning the choice of tourist destination. Consequently, the goals should be flexible and should be adjustable to changes in internal and external environment. Dissemination of information has certain costs. Too much information about unimportant matters is equally questionable as too little information about important matters. Therefore, the efficiency of communication processes depends on those whose duty is to ensure information, as much as on those who are engaged in delivering messages. A clear picture about targeted groups of recipients is needed in advance. It is easy to make a long list of recipients, but since time and money are limited, it is necessary to rank them in order of importance and influence concerning the goals set in the communication process.[4] Eventually, forming the strategy of information dissemination includes the determining of key messages. By developing the key messages, we should avoid indefinite statements, excessively complex statements and, in general, too much information. It is important to investigate the efficiency of messages on a specific sample of respondents. With respect to the heterogeneous character of the media, it is important to investigate the impact of messages in different forms of transfer. They can be clear and without excessive simplifying.[5] The tourism market is crowded with information and the recipients will simply not be able to memorize too complex messages.

Implementing the strategy of information dissemination

The strategy of information dissemination includes information, i.e. transfer of information, and communication, i.e. presenting the information in the form of key messages adjusted to targeted market segments. The media, and especially TV and radio stations, i.e. electronic media which have become the main means of conveying information, play the most important role in this process. A quick and easy access to the media is indispensable. The members of the team responsible for the implementation of the strategy, or the team PRs, should be available to the media at any time. Websites should become more accessible and more attractive for users and should be updated regularly with new contents. They are more efficient and cheaper than traditional means (catalogues, leaflets, posters, publications, etc). There is also a need to form a team of experts, with necessary professional support, who will be able to respond quickly to problem situations. The services of questions and answers should be dedicated a special attention. The availability of such services should be advertised.[6] The strategy of information dissemination should include the list of all relevant communica-

[4] Communication strategy, Step by Step Guide, Economic & Social Research Council, http://www.esrcsocietytoday.ac.uk, June, 2006.

[5] Ibid.

[6] Adjusted from: Communication Strategy Enlargement, European Commission, 2000.

tion activities in a form of a feasibility plan. Regarding the quality improvement of Croatian tourism, the communication activities can include branding, development of destination identity, media relations, publications, e-communication, events, etc. Generally speaking, by planning communication activities we should focus on the most important and most influential market segments, determine priority activities, and establish the possibilities for the rationalization of the strategy implementation costs. After making the list of priority market segments, it is important to determine the most appropriate channels they can be reached through. In this context, it is equally important to think about various time framework, imperatives, language and objectives of the targeted market segments.[7] If a specific segment is important to Croatian tourism, it does not have to mean that even Croatia as a tourist destination is important to potential tourist of the given market. Many of them feel over saturated with information, and, accordingly, the information should be disseminated by conveying only the relevant ones for them. For success on the offer-saturated tourism market, it is crucial for the content of information to be clear enough. In tourism not even secondary communication channels should be neglected. Maybe not every particular tourist should be reached directly. However, some types of the public, such as the media, are at the same time a channel towards other types of public, i.e. potential tourists, and they are also the public themselves. Therefore, it is wrong to underestimate the strength and importance of the media. Two-way communication is equally important for the development of relations with various segments of the tourism market. Accordingly, channels should be determined, which will be used in order to get feed-back information. The following communication tools can be used in implementing the components of the strategy of information dissemination: brochures, leaflets, posters, fliers, etc; interviews with TV and radio stations, newspapers and the media in general; interactive CDs; articles for selected magazines; interactive websites and information services; letters; personalized and mass electronic distribution of short and simple messages; e-bulletins, short video presentations, newsletters. Having designed the communication plan, we can start with the estimation of the time and budget needed for carrying out the strategy of information dissemination. We must also plan the money for unforeseen developments that can occur during the development of strategy. There are also ways of saving time and money spent on communication. Using e-mail and web communication instead of paper – after initial expenses, can be a cheaper way of communication. It also includes using specialized firms for particular tasks such as organization of conferences, designing and printing of publications and developing of websites. Although this may seem more expensive in short-terms, it can also save precious time of the team members responsible for strategy implementation and

[7] Adjusted from: Communication Strategy Enlargement, European Commission, 2000.

can ensure professional and directed communication by avoiding hidden expenses and inefficiency. Using somebody else's expertise can help with advice and support. Not less important is using only the channels relevant for a specific market segment, focusing on activities of high impact and low costs, and a good estimation of the influence of each activity.[8] A successful implementation of the strategy requires active participation of the public sector, primarily Parliament and the Government of the Republic of Croatia. The Ministry of Tourism should secure a framework for coordinating communication and information activities of all parties concerned who are included in the implementation of the strategy. Within this framework, it is necessary to establish a team for managing communication and information. Communication processes between the members of the team can be programmed or not.[9] Programmed communication is carried out in case of frequent situations. Communication is supported by the programme that defines all the methods and procedures developed to be applied at certain moments. If the members of the team work autonomously on routine tasks, communication should be minimal and automatic. However, in certain cases the members of the team can 'produce' too many reports so that the leader of the team can know what they are doing, and which limits the 'communicational capacity' of the team. Over saturation with information is a common cause why the members of the team avoid being engaged in the communication important for the team. Regarding the technology that makes it possible for us to exchange and to store a large amount of information, we, in fact, do that because we can, without even asking ourselves whether it brings some new value or not. The members of the team must provide each other with sufficient information about the areas where their work is interdependent for the purpose of coordination. The team must agree about methods and procedures for this kind of communication to be maintained and coordinated. Communication is not programmed when there are no adopted procedures which will resolve conflict situations, so communication requires a special treatment. When the members of the team observe some changes, it is of critical importance to inform the whole team about it. Communication should be continuous, comprehensive and in due time. Namely, the changes that occur in one place can be a landmark for something that will affect sooner or later some other parts of the system. The team that do not exchange that kind of information is not efficient. Regardless the fact that the members of a certain team can work together in developing answers to changed conditions, there is always a danger for their communication to become chaotic. Certain skills are needed so that they can communicate with each others in different time and from different places. Especially here, the team

[8] Communication strategy, Step by Step Guide, Economic & Social Research Council, http: www.esrcsocietytoday.ac.uk, June, 2006.

[9] Kimbal, L., Developing the Team's Communication Strategy, Caucus, http://consortium.caucus.com/published.html, June 2006.

should use the main part of communication means that they have at their disposal, i.e. technology, time and energy.

Information dissemination evaluation strategy

The efficiency assessment of information dissemination strategy is inevitable. Implementing the evaluation discipline in the strategy from the very beginnings enables survey and promotion of the strategy as a flexible document. Therefore, the function of evaluation relies on the assessment and feedback, i.e., on the comparison of the goal and the level of accomplishment. It is especially important to point out that the efficiency of evaluation is in a strong correlation with the planning, because only the evaluation affecting and changing people’s behaviour is efficient. In such a way the concentration of tourist management can be directed only towards the issues interesting for Croatian tourism and requiring a certain reaction. With regard to the initial goals and activities, it is important to incorporate some very simple implemental indicators and evaluation methods, such as: research of tourists’ standpoints related to the quality of Croatian offer as a tourist destination ‘before and after’; systematic observation of media coverage (so that the information supply and the transfer of messages can be necessarily adjusted); supervision of website usage. The strategy of information dissemination should include time schedule of the main communication activities. It should include deadlines, sections, and evaluation points. The process of information management requires time, and the best strategy will fail if there is no thinking about the ways of using it. The accurate estimate of time needed for dissemination is important for the feasibility of the strategy. Evaluation is based on information and measurable results of the strategy implementation. Accordingly, so that the strategy can reach the goal, we need the support of information system which will provide Croatian tourist management with necessary information that will help make decisions about improving the strategy of information dissemination, which is the basic element of the information management model functioning in the quality improvement of Croatian tourism.

Conclusion

Tourism (especially its economy effects) is extremely important for Croatia. However, it is doubtful whether it can be recognized as a destination on the tourism market. By considering and determining tourist destination policy, Croatian tourist management must endeavour in trying to act directly on factors that mould the perception of tourist and potential tourists, for the purpose of affecting the results of market competition in the future. Given the fact that tourist are becoming ever more sophisticated, their expectations from tourist destination continue to grow. The processes in tourism are significantly influenced by continuous perfecting information and communication technologies. Tourist are exposed to a large quantity of information, so tourist destinations must seriously

pay attention to collecting, processing, storing and dissemination of information, which should result in adequate messages conveyed to the tourism market.

References

Applegate, L. M., McKenney, J. L. i McFarlan, F. W., Corporate Information Systems Management: Text and Cases, McGraw-Hill Companies, New York, 1999.

Bergeron, P., Information resources management, *Annual Review of Information Science and Technology*, vol. 31, 1996., str. 263-300.

Buhalis, D., Strategic use of information technologies in the tourism industry, *Tourism Management*, vol. 19, n. 5., 1998., str. 409-421.

Communication strategy, Step-by-Step Guide, Economic & Social Research Council, http://www.esrcsocietytoday.ac.uk, lipanj, 2006.

Communications Strategy for Enlargement, European Commission, 2000.

Gill, H.S. i Yatesmercer, P., The dissemination of information by local authorities on the World Wide Web, *Journal of Information Science*, vol. 24, n. 2, 1998., str. 105-112.

Griffin, R.W., Management 3rd ed., Houghton Mifflin Company, Boston, 1990.

Hong-Mei, C. i Sheldon, P.J., Destination information systems: design issues and directions, *Journal of Management Information Systems*, vol. 14, n. 2., 1997., str. 151-177.

Hybels, S. i Weawer, R.L., Communicating Effectively, 6th ed, McGraw-Hill, New York, 2001.

Kimbal, L., Developing the Team's Communications Strategy**,** Caucus, http://consortium.caucus.com/published.html, lipanj, 2006.

Kotler, P., Bowen, J. i Makens, J., Marketing for Hospitality and Tourism, 3rd ed., Prentice Hall, New Jersey, 2003.

Laws, E., Tourism Destination Management, Routledge, London, 1995.

McLeod, R. i Schell, G., Management Information Systems, Prentice Hall, Englewood Cliffs, New Jersey, 2003.

Mintzberg, H., The Rise and Fall of Strategic Planning, Prentice Hall, Englewood Cliffs, New Jersey, 1994.

Page, S., Brunt, P., Busby, G. i Connell, J., Tourism: A Modern Synthesis, Thomson Learning, London, 2001.

Phelps, J.E., Harris, T.E. i Johnson, E., Exploring decision-making approaches and responsibility for developing marketing communication strategy, *Journal of Business Research*, vol. 37, n. 3., str. 217-223.

Sheldon, P.J., Destination information systems, *Annals of Tourism Research*, vol. 20, n. 4., 1993., str. 633-649.

Srića, V., Inventivni menedžer, Croman & MEP Consult, Zagreb, 1994.

E-HERITAGE AND PRESERVATION

The Importance of Usability in Development of Digital Libraries

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

Summary

Digital libraries are among the most interesting electronic information resources of our time. They are the collection of services and the collection of information objects and services that support users in accessing information objects available on the Internet. They are complex information systems consisting of many components which depend on each other in their operation. To achieve the flawless operation of digital libraries their developers should evaluate their achievements during the phase of development as well as during the phase of use. As a result of their efforts, users should get usable and easy to use information systems which are adapted to their needs and preferences. Evaluation and usability are two important concepts which may help librarians and computer specialists to design better information systems and better services for them. Usability testing may include one (usually the user interface) or many components of an information system for evaluation. Evaluation also includes the research of user community. Information acquired in such a way may also significantly help in information system improvement. Since the goal of evaluation is to achieve quality and improvement, the final result should be a superior information system like digital library and satisfied users.

Key words: digital library, usability, evaluation

Introduction

The world of exciting and content rich electronic resources of information available on the Internet is growing constantly. Digital libraries are among these resources. Digital libraries are the collection of services and the collection of information objects and their organization, structure, and presentation that support users in dealing with information objects available directly or indirectly via

electronic/digital means.[1] Still, they are not just collections of information objects as physical libraries are not just places. According to Pomerantz and Marchionini, fundamental characteristics of libraries are systematic access to information resources, the ideas represented by those resources and sets of human stakeholders.[2] In case of digital libraries, information and communication technology provides systematic access to information resources and makes possible their fundamental functionality by giving users an opportunity to access the information resource regardless of their physical location i.e. point of access to electronic information resources. The primary medium for use of the content stored in digital collections is the Internet i.e. World Wide Web. World Wide Web is a window to the world for many Internet users, and due to the number of users accessing digital libraries each day and their differences it is difficult to develop a perfect digital library with a perfect user interface which will satisfy everyone's needs and preferences. At that point, the concept of usability may help librarians and computer specialists to design better digital libraries with more usable user interfaces and better services by collecting data about digital library users. The collected data information can be used as a valuable tool in planning of new services in digital libraries. For the reason, evaluation and usability testing are prerequisites for the development of quality digital libraries.

The role of information technology in digital library

Operation of digital libraries would be impossible without information technology. The role of information technology in digital libraries can be broadly divided into three views:[3]

- Revolutionary view on IT (physical location is irrelevant; more and more activities will be supported by IT; printed publications will vanish; electronic publications and distribution will be become a rule etc.)
- Evolutionary view on IT (network technologies are addition to the previous technologies; digital media are extension of previous media; electronic publishing will be limited only to a certain material with specific purpose; printed and electronic publications will co-exist etc.)
- Co-evolutionary view on IT (in between of the first two views; co-evolution of IT, human behavior and organizations; people use technologies which are helpful for their work; technologies are adapted to user needs etc.).

[1] Koohang, Alex; Ondracek, James. Users' views about the usability of digital libraries. // British Journal of Educational Technology (2005), 3; 408

[2] Pomerantz, Jeffrey; Marchionini, Gary. The digital library as place. // Journal of Documentation (2007), 4; 506

[3] Borgman, Christine. Od Gutenbergova izuma do globalnog informacijskog povezivanja. Lokve;Zadar : Naklada Benja, Gradska knjižnica Zadar, 2002. 1

With the advent of Web 2.0, in theory, co-evolutionary view on IT began to be perhaps the most feasible one since Web 2.0 offers the most similar context for use of IT and is oriented to users and their preferences in organization and use of available electronic information resources on World Wide Web including digital libraries. Such a viewpoint is very valuable for understanding of different aspects of interaction between users and digital libraries. Although it still not clear how significant is the Web 2.0 phenomenon, it will certainly give computer professionals some new ideas for the development of new and useful Web based services.

Usability and digital libraries

Digital libraries are complex information systems consisting of many components depending on each other in their operation in order for the whole system to work flawlessly. To achieve the flawless operation of digital libraries, their developers must adhere to certain standards and plan smartly each phase of digital library development. McCray and Gallagher suggest that every digital library development should follow ten principles:[4]

- Expect change
- Know your content
- Involve the right people
- Design usable systems
- Ensure open access
- Be(a)ware of data rights
- Automate whenever possible
- Adopt and adhere to standards
- Ensure quality
- Be concerned about persistence.

Users expect usable and easy to use information systems adapted to their needs and preferences. That is the reason for considering usability (and evaluation in general) in phases of development and later during the use of digital libraries.

What is usability? For different disciplines, usability has a different meaning. According to Chowdhury, librarians perceive the usability of an information service in terms of efficient and effective access to information.[5] Nielsen defines usability as a quality attribute that assesses how easy user interfaces are to

[4] McCray, Alexa T.; Gallagher, Marie E. Principles For Digital Library Development. // Communications of the ACM (2001), 5; 50

[5] Chowdhury, Sudatta; Landoni, Monica; and Gibb, Forbes. Usability and impact of digital libraries: a review. // Online Information Review (2006), 6; 658

use; the word "usability" also refers to methods for improving ease-of-use during the design process.[6]

In addition, Nielsen points out that usability is defined by five quality components:

- Learnability: How easy is it for users to accomplish basic tasks the first time they encounter the design?
- Efficiency: Once users have learned the design, how quickly can they perform tasks?
- Memorability: When users return to the design after a period of not using it, how easily can they reestablish proficiency?
- Errors: How many errors do users make, how severe are these errors, and how easily can they recover from the errors?
- Satisfaction: How pleasant is it to use the design?

This definition refers to the user interface (the most visible component of a digital library) and design but usability can be tested on some other components of digital library. The user interface is the main access point of every digital library and points of interaction of user and information system and its development ought to be supported by the results of user studies. These results must include user needs and preferences which are often revealed a posteriori, after the development of digital library has been finished. Building and user interface to be attractive rather than to be functional is the result of not taking into account context in which users access the digital library on the Internet (i.e. previous knowledge and experience in use of electronic information resources, choice of a browser, screen resolution, content organization preferences, font type, size and color, accessibility support etc.).

According to McCray and Gallagher, when building a usable digital library system, one must take into consideration difference between (user's) computers and browsers, speed of access, differences among users including Web navigation preferences. The user interface should be as simple as possible with clear navigation mechanisms, and it should provide the user with multiple access points to the content of the visited digital library. Furthermore, accessibility for the users with disabilities should be provided: user access to all content must be given, documents must be clear and simple, user should have control of styles (of content display), information about context and orientation should be also supplied and standard markup should be used.[7]

Van House, Butler, Ogle and Schiff point out that digital libraries can be described and evaluated on three key components: contents, functionality, and in-

[6] Nielsen, Jakob. Usability 101: Introduction to Usability. URL: http://www.useit.com/alertbox/20030825.html (23.7.2007.)

[7] McCray, Alexa T.; Gallagher, Marie E. Principles For Digital Library Development. // Communications of the ACM (2001), 5; 51

terface, and that usability wrongly addresses only or primarily interface design, which is, according to them, too narrow a basis for evaluating something as complex as a digital library.[8]

Additionally, usability is generally characterized as the determining aspect of a system's capability to satisfy the needs and specifications of users. Usability is the degree to which users easily and effectively use a system.[9]

To acquire valuable data from users about the use of an interface of a digital library, a number of methods can be employed: observing users while using different digital libraries, Web sites and collecting data about their activities during these visits, users can verbalize their thoughts while completing necessary tasks on a particular digital library Web site; their activities can also be videotaped and later analyzed. Other methods are: comparative studies of Web sites or their respective components against the list of criteria, use of benchmarks (eg. time necessary for a task completion), asking questions related to the tasks in question to users, user studies (questionnaires, interviews, focus groups) etc.

What importance does the usability have in the world of digital libraries? The results of user studies and usability testing can improve interaction between users and digital libraries on the Internet in most cases. The first digital library projects in the early 1990s were groundbreaking because of the technology they used, electronic material organizational schemes they employed or just because they were the first in their field of knowledge. There wasn't any systematic evaluation since it was more important to get a digital library running rather than to evaluate aspects of its development and use. In contrast, traditional libraries did research their users, collecting valuable data which enabled them to improve the existing services and to develop new ones. However, the important aspect of the traditional libraries – their holdings based on printed material are unchangeable in the physical form. In contrast to physical libraries, digital libraries can be reconfigured comparatively easily after they are built[10] if librarians knew something about their users, which is sometimes very difficult, since the Internet user community represent very heterogeneous population with different motives for visiting digital libraries. To identify and recognize your user is especially important in situations in which users leave digital library Web site if they come across any difficulty during their visits. In such situations usability testing can be helpful to measure how easy is to use a digital library.

[8] Van House, Nancy A.; Butler, Mark H.; Ogle, Virginia; Schiff, Lisa. User-Centered Iterative Design for Digital Libraries: The Cypress Experience. // D-Lib Magazine (1996), 2; http://www.dlib.org/dlib/february96/02vanhouse.html

[9] Koohang, Alex; Ondracek, James. Users' views about the usability of digital libraries. // British Journal of Educational Technology (2005), 3; 408

[10] Pomerantz, Jeffrey; Marchionini, Gary. The digital library as place. // Journal of Documentation (2007), 4; 512

What are actual criteria for usability testing? Saracevic offered a selection of criteria which could be applied to digital library usability testing.[11] He divided criteria into four categories:

- Content: accessibly, availability, clarity (as presented), complexity, (organization, structure), informativeness, transparency, understanding, effort to understand, adequacy, coverage, overlap, quality, accuracy, validity, reliability, authority
- Process (carrying out tasks as search, browse, navigate, find, evaluate or obtain a resource): learnability to carry out, effort/time to carry out, convenience, ease of use, lostness (confusion), support for carrying out, completion (achievement of task), interpretation difficulty, sureness in results, error rate
- Format: attractiveness, sustaining efforts, consistency, representation of labels (how well are concepts represented?), communicativeness of messages
- Overall assessment: satisfaction, success, relevance, usefulness of results, impact, value, quality of experience, barriers, irritability, preferences, learning.

The difficulty with the usability testing in digital libraries is that usually there is no clear decision about what should be evaluated and there is no definite list of criteria for the evaluation, and results from one digital library cannot be directly applied to another digital library, as they sometimes differ significantly. That is the reason why criteria for the evaluation of digital libraries are subject to change and development.

Evaluation

Usability testing is a part of evaluation process in digital libraries. For Marchionini, evaluation is a research process that aims to understand the meaning of some phenomenon situated in a context and the changes that take place as the phenomenon and the context interact. Evaluation specifies what is the research process (metrics and procedures), what is the phenomenon (its mission and salient characteristics), and the context(s) in which the phenomenon occurs.[12]

Evaluation of a digital library is equally important in the initial stages of ots existence i.e. during its development and later during its use. The phase of development is especially important if we want our users to become firmly oriented toward the use of digital libraries in future. To achieve this, digital library

[11] Saracevic, Tefko. Evaluation of Digital Library: An Overview. http://dlib.ionio.gr/wp7/WS2004_Saracevic.pdf (30.08.2007.)

[12] Marchionini, Gary. Evaluating Digital Libraries: A Longitudinal and Multifaceted Vew. // Library Trends (2000), 2; 311

system builders should evaluate often and early and exploit existing good practices within library science.[13]
In his paper on digital library evaluation, Saracevic mentions dilemma of selection of level of objectives of evaluation. He divided objectives of evaluation of digital libraries into seven general classes or levels.[14]

User centered:
Social level. How well does a digital library support the needs and demands, roles, and practices of a society or community?
Institutional. How well does a digital library support the institutional or organizational mission and objectives? How well does it integrate with other institutional resources?
Individual. How well does a digital library (or given services) support information needs, tasks, activities of people as individual users or groups of users with some strong commonalties?
Interface. How well does a given interface provide and support access, searching, navigation, browsing, and interaction with a digital library?

System centered:
Engineering. How well do hardware, networks, and related configurations perform?
Processing. How well do procedures, techniques, algorithms, operations and so on perform?
Content. How well is the collection of information resources selected, represented, organized, structured and managed?
Saracevic points out, that digital libraries are usually evaluated only on one level and that one level can rarely answer question from another. This creates difficulties since digital libraries are complex information systems which require complete understanding of operation of all of its components. In 2003, the same author gave a broadened list of approaches for digital library evaluation:[15]

1. System-centered approach: study of performance assessing effectiveness and/or Efficiency
2. Human-centered approach: study of behavior such as information seeking, browsing, searching or performance in completion of given tasks

[13] Blandford, Ann; Gow, Jeremy. Digital Libraries in the Context of Users' Broader Activities: JCDL 2006 Workshop Report. // D-Lib Magazine (2006), 7/8 http://www.dlib.org/dlib/july06/blandford/07blandford.html (23.7.2007.)

[14] Saracevic, Tefko. Digital Library Evaluation: Toward and Evolution of Concepts. // Library Trends (2000), 49; 363

[15] Saracevic, Tefko. Evaluation of Digital Library: An Overview. http://dlib.ionio.gr/wp7/WS2004_Saracevic.pdf (30.08.2007.)

3. Usability-centered approach: assessment of different features e.g. of portals, by users; a bridge between systems- and human-centered approaches
4. Ethnographic approach: comprehensive observation of life-ways, culture and customs in a digital library environment; impact of a digital library on a given community
5. Anthropological approach: comprehensive observation of different stakeholders or communities and their cultures in relation to a given digital library
6. Sociological approach: assessment of situated action or user communities in social setting of a digital library
7. Economic approach: study of costs, cost benefits, economic values and impacts.

Each of these approaches has its advantages and disadvantages. Since there is no standard model for digital library evaluation, nor is there a comprehensive set of models and toolkits that can be used by digital library evaluators[16] professionals involved in the process of evaluation choose different approaches and methods which seem to be the most appropriate in a particular situation. This makes longitudinal studies in digital libraries difficult and opens a possibility that the results will not be comparable and applicable to future library practice.

Conclusion

Usability and evaluation are two very important concepts. Without the first, information and computer systems would be unusable. Without the second, almost any human activity could not reach the desired level of quality. Digital libraries are results of applied knowledge and experience of professionals from many disciplines. Two among them are very important for the existence of digital libraries: computer science and information science with librarianship. Computer science knows usability well. Librarianship knows evaluation well. By using these two important concepts, digital libraries will most certainly profit by improving quality of their service. Since the users have become the focal point of the recent digital library development efforts, it is expected that their needs and preferences will be integrated in new digital library services. Usability testing will provide digital library developers with information about the level of acceptability of a digital library in a particular user community. Users will define future directions of development of services in digital libraries more than they had chance to do so in the past and that is the reason why usability is an indispensable tool in the development of digital libraries.

[16] Chowdhury, Sudatta; Landoni, Monica; and Gibb, Forbes. Usability and impact of digital libraries: a review. // Online Information Review (2006), 6; 657

References

Blandford, Ann; Gow, Jeremy. Digital Libraries in the Context of Users' Broader Activities : JCDL 2006 Workshop Report. // D-Lib Magazine (2006), 7/8 http://www.dlib.org/dlib/july06/blandford/07blandford.html (23.7.2007.)

Borgman, Christine. Od Gutenbergova izuma do globalnog informacijskog povezivanja. Lokve;Zadar : Naklada Benja, Gradska knjižnica Zadar, 2002.

Chowdhury, Sudatta; Landoni, Monica; and Gibb, Forbes. Usability and impact of digital libraries: a review. // Online Information Review (2006), 6; 656-680

Koohang, Alex; Ondracek, James. Users' views about the usability of digital libraries. // British Journal of Educational Technology (2005), 3; 407–423

Marchionini, Gary. Evaluating Digital Libraries: A Longitudinal and Multifaceted Vew. // Library Trends (2000), 2; 304-333.

McCray, Alexa T.; Gallagher, Marie E. Principles For Digital Library Development. // Communications of the ACM (2001), 5; 49-54.

Nielsen, Jakob. Usability 101: Introduction to Usability. URL: http://www.useit.com/alertbox/20030825.html (23.7.2007.)

O'Reilly, Tim. Web 2.0: Compact Definition? 1.10.2005. http://radar.oreilly.com/archives/2005/10/web_20_compact_definition.html (28.8.2007.)

Pomerantz, Jeffrey; Marchionini, Gary. The digital library as place. // Journal of Documentation (2007), 4; 505-533

Saracevic, Tefko. Evaluation of Digital Library: An Overview. http://dlib.ionio.gr/wp7/WS2004_Saracevic.pdf (30.08.2007.)

Saracevic, Tefko. Digital Library Evaluation: Toward and Evolution of Concepts. // *Library Trends* (2000), 49; 350-369

Tsakonas, Giannis; Papatheodorou, Christos. Critical Constructs of Digital Library Interaction http://www.ionio.gr/~papatheodor/papers/PCI2007_Tsakonas_Published.pdf (28.8.2007.)

Van House, Nancy A.; Butler, Mark H.; Ogle, Virginia; Schiff, Lisa. User-Centered Iterative Design for Digital Libraries: The Cypress Experience. // D-Lib Magazine (1996), 2; http://www.dlib.org/dlib/february96/02vanhouse.html

Xie , Hong. Evaluation of digital libraries: Criteria and problems from users' perspectives. // Library & Information Science Research (2006), 28; 433–452

The Use of Search Engines for Locating Information on Cultural Heritage of Croatia

Danijel Radošević
University of Zagreb, Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
goran.bubas@foi.hr

Goran Bubaš
University of Zagreb, Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
igor.balaban@foi.hr

Igor Balaban
University of Zagreb, Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
miran.zlatovic@foi.hr

Summary

In this paper the authors present an overview of how search engines work and the results of the use of different search engines for locating information on the web related to specific artifacts of the cultural heritage of Croatia. It was found that the Google search engine outperformed the competition in locating web sites with information on specific cultural heritage artifacts. In another analysis, selected web pages with general information on Croatian cultural heritage were examined regarding several elements of search engine optimization. Finally, a group of students of information systems were asked what would be their typical activity when they needed information related to cultural heritage and most of them indicated that they would first search the Internet/web. Recommendation is provided regarding the means for increasing the "visibility" of information related to cultural heritage on the Internet/web and the likelihood of this information being found by the users of this medium.

Key words: search engine optimization, cultural heritage

Introduction

There is considerable interest for developing standards and systems for distributed search and retrieval of cultural heritage information (see, for instance: Moen, 1998; Kando and Adachi, 2004). The worldwide population of Internet

users reached 1.17 billion in June 2007 with Internet penetration of 40% and 69% in the regions of Europe and North America respectively (for the latest statistics see: InternetWorldStats, 2007). The Internet enables cultural heritage information to be shared among a wide range of audiences and this makes immediate access to information, channel capacity, media type, display mode, and adaptability to the type of device of the users some of the important issues that have to be considered in online presentation of cultural heritage (see: Smeulders, 2002). Recently, use of visual navigation with web maps and personalization to user interest (Mac Aoidh *et al.*, 2007), use of the semantic web techniques for searching and annotating cultural heritage collections (Ossenbruggen, 2007) and development of web-based recommender systems (Rutledge *et al.*, 2007) are being developed to facilitate search for cultural heritage information.

The way Internet users locate cultural heritage information on the web is related to their search strategies. Some of the most evident aspects of Internet users web search behaviour can be described as follows (see: White and Iivonen, 1999):

- Users tend to focus on the web sites that are known to them and subsequently proceed in their search by tracing links from those web sites.
- Users browse the web by quickly selecting and following links, perform quick searches and browse through the results when they do not immediately locate what they are searching for.
- Web search often starts with the use of a search engine.
- The simple search terms/statements are preferred in the use of search engines and the modification of those terms/statements is a common reaction to dissatisfying search results.
- The success rate of web searches variable and depends on the search questions and sometimes the users believe that they have located the right information even when that is not true.

More than 60% of adult Internet users in the U.S.A. use a search engine on an average day (Rainie and Shermak, 2005), they trust their search engines (Fallows, 2005) and most of them click on a link within the first page of results, but only 12% of the users continue their search beyond the third page of results (iProspect, 2006). Today, there are a number of web search engines based mostly on keywords that enable full text search. Some of these are world popular, like *Google* and *Yahoo!* etc., but some are local like our Croatian search engines *www.hr* and *Pogodak!*.

This paper will focus on the way search engines collect documents on the web and process them for later retrieval, as well as on web page optimization by use of metadata and other techniques. Also, a brief analysis is performed of what would be typical behavior of an information science student in search of information about popular cultural heritage artifacts of Croatia. Finally, the results of

comparative analysis of several search engine performance in locating selected cultural heritage artifacts is presented.

How search engines work?

Web search engines collect html and other web documents under its domain using so called web crawlers. Web crawlers read web documents, put them into temporary storages (caches) and process them in many ways e.g. by extracting keywords and other features for later retrieval, automatic classification of documents, forming document indexes etc. Therefore search engines try to find needed documents among documents collected by crawlers.

Use of search engines

To facilitate the location of specific documents on the web with a search engine the documents have to be well described. To locate a specific document users try to describe them in a meaningful way and they mostly use queries or topic catalogues for this purpose. Queries consist of keywords – words that describe documents for searching. It is not always easy to find appropriate keywords that describe documents in the best way to perform a web search, but that is a skill which could be improved by practice. On the other hand, topic catalogues are rarely used because the user seldom knows the full topic hierarchy for a specific document, and also there are many ways in which some documents can be classified (e.g. a historical document can be classified using both a geographic criterion and a chronological criterion).

Web crawling

Web crawlers (sometimes called *web spiders*) collect web documents for later searching. Crawling the web is a continuous process because new documents appear on the web daily. Except for that, the documents that have been updated/changed since they were last accessed by the web crawler should also be updated in search engine database.

Crawlers typically use sets of Uniform Resource Locators (URLs; Thompson, 2002.) which represent addresses of web sites and read their contents. Web sites usually have entry documents, like index.html or default.htm, so the crawling of a particular site could start from that document. Except for collecting document texts, it's important to collect all links to other documents. These links will lead the crawling engine to other documents to be crawled. But, the number of links can grow rapidly (e.g. 10 links on the first level could lead to 10 links on each document on the second level, etc.) so the number of links at lower levels could be too big for effective processing. Therefore crawling should be bounded by some parameters like number of levels or by using single domain name (e.g. crawling only links under the same domain as entry URL). Also, a crawler engine should be capable of collecting documents (and information from these documents) in different file formats like html, doc, pdf, ppt etc. Another request

to crawlers is their ability to support different communication protocols like HTTP and FTP.

Search engine indexing

Crawled documents should be indexed for later searching. Off course, the size of indexes has to be significantly smaller than the size of crawled documents, because of the efficiency of searching. There are many different techniques used for indexing, but some of them are most common:

- Ignoring stop words and other extremely common words (Thompson, 2002.). Stop words and extremely common words are useless for searching documents (e.g. “the”, “and”, “it” etc.). Other extremely common words include different formatting words which depend on document type (e.g. html documents use html tags like <html>, <body>, <title>, etc.).
- Using document metatags. Html documents can contain some additional data except for their information content, like keywords, author’s name, time when document was last updated, software used in document production etc.

Html documents often contain keywords which are invisible when viewing through web browser, but visible for web crawlers and for indexing purposes, e.g.

```
<meta name="keywords" content="computing, programming, C++, C#, Pascal, Java">
```

means that this document should be indexed using keywords computing, programming, C++, Pascal, C#, and Java. But, despite metatags make indexing easier, there are some possibilities of their misuse, for instance by using too many keywords, very long keywords etc.

Improving search engine ranking

Users often create web pages using tools which are focused mostly on the appearance of a particular web page. Creating and publishing a web page is often not enough for it to be found by other people. Special techniques are used for web pages to be found by search engines crawlers and favorably ranked in the results of search engine use.

There are several established techniques that can be used to increase web page ranking in the results of a search engine, which makes the web contents of a specific web page more approachable for potential viewers interested in its content (Coopee, 2000):

- submitting a web site to the search engine simply by filling out an online form (even though crawlers collect majority of all web contents, it’s also the way to put information about our site to web search engines and catalogues, especially those which have no crawlers and support manual fill only;

- using metatags (there are more than 50 types of metatags available for use within HTML pages, but most commonly used by search engines for indexing and ranking purposes are the description and keywords tags.

The *description metatag* includes a short description of a site that a search engine would display in a list of search results:

```
<meta name="description" content="Programming and programming languages">
```

The above mentioned tag means that the text *"Programming and programming languages"* will appear next to the URL in search engine results.
When using *keywords tag*, a combination of unique and common keywords should be used (Coopee, 2000). The idea is to anticipate which keywords the potential visitors would use to find the site.
Calculated *linking to the web page* from other web sites (Coopee, 2000) is another technique for influencing page rank. Some search engines (Google) use web page popularity in ranking, which is determined by how often a particular web page is linked to by other pages in the index.
Optimizing by *doorway pages* is also popular by web site promoters. Doorway pages are pages optimized for one search engine and 1-3 keywords (Shapiro and Lehoczky, 2007). Other names for doorway pages are "gateway", "bridge", "entry", "jump" or "supplemental" pages. These pages are separated from the rest of the site (stand on their own), containing only the link to other contents. This method was popular in the past, and often misused for spamming so it should be used carefully to avoid search engines *antispam* protection.
Some frequent mistakes that search engines don't like are *frames*, *dynamic content* and *Flash intros* which make web page indexing difficult for the search engine, while *spam* may cause penalization of the web page by the search engine in relation to rank in search results (Shapiro, 2007). *Spamindexing* is a term that denotes the methods used to increase the web page rank by manipulating the web page ranking algorithms of search engines. The following techniques are usually considered as spam:

- meta refresh tags,
- invisible text and overuse of tiny text,
- irrelevant keywords,
- excessive repetition of keywords,
- overuse of mirror sites,
- submitting too many pages in one day,
- identical or nearly identical pages,
- submitting to an inappropriate category (for directories),
- link farms (a group of web sites that hyperlink to all group members).

An overview of various factors which influence web page relevancy and ranking on different search engines is presented in Table 1.

Table 1. Factors that influence ranking of web pages on Google, MSN and Yahoo! (Shapiro, 2007)

		POTENTIAL POSITIVE INFLUENCE ON RANKING				
SEARCH ENGINE AND WHAT IT LIKES	**Elements that block indexing**	**Content and location of keywords**	**HTML title**	**Meta tags**	**Keyword frequency**	**Link popularity**
Google Link popularity, keywords near each other, keywords in URLs and link text, themes	Not mentioned, spammers	Keywords in text or links; keywords should be close to each other	Not mentioned, but seems to be a factor	No	Not mentioned	Very important, especially from relevant pages
MSN Theme present throughout the site, site popularity	Spammers, frames (<noframes> tag needed)	Not mentioned	Important, should contain keywords	Important, both descrip-tion and keywords	Important, 4-12 times	Important
Yahoo Concise/accurate descriptions and keywords, appropriate categorization	Spammers	Worthy of indexing as determined by editors; in appropriate category	No, but the title filled in plays a role, it should be concise	No, but the descrip-tion and keywords play a role	No	Very important

From the data presented in Table 1 it can be concluded that, after the content and location of keywords, the HTML title of a web page and link popularity are the second most important elements in relation to relevancy ranking by listed search engines. HTML titles can be easily defined by the author of the web page. On the other hand, link popularity is harder to achieve because it means that a specific web site is linked from other web sites.

Link popularity can also be used as a method for manipulating with ranking results. This type of activity is often called a *link bomb* or *Google bomb* when the attempt is to bias the results of the Google search engine by linking to a target web page from words that are not in consistence with the content of this web page (thus a wrong impression of the target can be formed). Simplified, a Google bomb is created if a large number of sites contain an anchor text that leads to a specific target site. Some specific requirements for Google search engine are presented in Table 2.

Table 2. Specific requirements for Google search engine (Shapiro, 2007)

Favored document length	**Keywords**	**Location of keywords**	**HTML title**	**Meta tags**	**What's spam?**	**Other info**
Wide range, from 50-600 words	Weight and proximity matter most	<h> tags, bold text	Keywords here, up to 90 characters	No	Use of link farms, cloaking, excessive repetition	Link popularity is the most important factor

To improve the ranking of a particular web site, the following should be considered (Shapiro, 2007; Heng, 2006, 2007):

- *HTML title* should contain properly placed keywords;
- meta tags should be defined, especially description and keywords tags, despite the fact that some search engines, like Google do not consider them relevant;
- *keyword density* is important because if a particular keyword has a higher density on the web page, then the likelihood of that page obtaining a better search engine ranking increases;
- *keywords in the URL or file name* of the web page;
- *Alt tags* help search engines to recognize what's in the image displayed within the web page;
- *link popularity* is increased if our pages are linked to from other sites; for example, Google uses the *Page Ranking Algorithm* which ranks pages according to the number and quality of links leading to that page;
- *themes* of web pages which are consistent;
- *design* of a web page which makes it easy to navigate;
- no spam or frames since search engines penalize such content;
- dynamically generated pages could be indexed by some search engines, like Google, but it is still recommended to put some links to them from other pages inside the web site.

It must be noted that an experimental analysis proved that using metadata elements increases rank order of web pages (this depends on the quality of metadata description) and that the use of metadata is comparable to using the Doublin core which is much less popular for the authors of web pages (see: Khaled, 2006).

Locating Croatian cultural heritage information with search engines

Different search engines use various mechanisms for web crawlers and different algorithms and principles for indexing and ranking of links to located documents on the web. Therefore, in August 2007 a test was performed of search engines Google, MSN, Yahoo! and Pogodak! to determine which search engine reports more locations on the web in relation to selected names of Croatian cultural heritage artifacts. Also, the first 100 web sites that were linked on the list of search results were examined regarding the usefulness and quality of information on the specific item of Croatian cultural heritage and the more useful links were counted. The results of this test are presented in Table 3.

As can be observed from the data presented in Table 3, the largest number of web sites in relation to most of the items of Croatian cultural heritage that were searched for was reported in the results of Google search engine use. MSN and Yahoo! performed rather poorly in comparison to Google. Finally, on the average, Pogodak! performed slightly better than MSN and Yahoo! search engines.

In addition, for every search term in Table 3 an inspection was performed of the web sites on the list of first 100 links in the search results of all four previously mentioned search engines and, in most cases, a larger number of *quality sites* (with sufficient textual information on a specific cultural heritage artifact) was found in the results of Google search. It can be concluded that Google search engine is probably the best choice for locating information about Croatian cultural heritage artifacts on the web.

Specialized institutional web sites and portals in Croatia provide general information on cultural heritage and links to related web sites. The users of the Internet who are interested in Croatian cultural heritage could locate those portals with a search engine and use of appropriate key words (e.g. “cultural heritage Croatia”).

In the previous analysis (see Table 3) it appeared that Google was the best search engine for locating information on most of the specific cultural heritage artifacts of Croatia. It must be emphasized that Google is the search engine with most search queries (for instance, in August 2007 in the U.S.A. it had about 54% of share of searches, while Yahoo! had 20% and MSN 13% of share of searches; see Nielsen/NetRatings, 2007). Therefore, a Google search was performed for web sites with general information related to Croatian cultural heritage and the results are displayed in Table 4. Also, the web page source was analyzed for the selected web pages and the use of meta tags was assessed together with the count of key words in textual information on the web page.

The data in Table 4 lead to the conclusion that the three web pages with the highest rank had key words “cultural heritage Croatia” either in the title of the web page (marked with “<title>“ in page source) or in the meta tags and document description. However, of those only the *Culturenet Croatia* website, the virtual portal to Croatian culture project funded by the Ministry of Culture of the Republic of Croatia (for more information about this project see: Uzelac, 2005), could be regarded as a web site of a formal cultural or heritage institution, while the other three web sites ranked among the first two pages of Google search results were of the nonprofit organization *Croatian Camping Union (Kamping udruženje Hrvatske),* tourism agency *adriatica.net*, tourism portal *Croatica.net*, and *Croatian National Tourist Board*. Within the first 100 results there were also the web page “Croatia UNESCO Heritage” of the *Find Croatia* tourism portal and several pages with partly related content of the *Ministry of Foreign Affairs and European Integration of Croatia* (not shown in Table 4). The homepage of the more important website of the *Museum Documentation Center* with links to many web sites of Croatian museums was not in the first 100 search results.

Table 3. The number of links found in the results of search engines Google, MSN, Yahoo! and Pogodak! with the use of search terms related to Croatian cultural heritage (the results with and without the use of diacritical symbols in Croatian language are displayed; performed in September 2007)

Searched term*	Google	MSN	Yahoo!	Pogodak!
Bašćanska ploča	**27,000**	**634**	**626**	**4,230**
Bascanska ploca	**31,000**	**430**	**334**	**3,070**
Plominski natpis	**347**	**25**	**157**	**84**
Višeslavova krstionica	**336**	**28**	**209**	**150**
Viseslavova krstionica	**231**	**2**	**14**	**150**
Crkva sv. Križa Nin	**13,000**	**160**	**231**	**504**
Crkva sv. Kriza Nin	**13,000**	**54**	**73**	**505**
Vinodolski zakonik	**46,000**	**576**	**411**	**650**
Povelja kneza Trpimira	**376**	**15**	**64**	**65**
Vučedolska golubica	**798**	**290**	**708**	**386**
Vucedolska golubica	**797**	**14**	**178**	**45**
Knežev dvor	**66,600**	**3,431**	**3,180**	**5,898**
Knezev dvor	**45,000**	**580**	**468**	**5,898**
Katedrala Sv. Jakova	**118,000**	**1,046**	**2,610**	**1,217**
Eufrazijeva bazilika	**29,800**	**1,514**	**18,000**	**1,916**

* Bašćanska ploča *(The Basca Tablet)*, Plominski natpis *(The Plomin Tablet)*, Višeslavova krstionica *(The Baptistry of Višeslav)*, Crkva sv. Križa *(The Church of the Holy Cross)*, Vinodolski zakonik *(The Law Code of Vinodol)*, Povelja kneza Trpimira *(Trpimir's Charter)*, Vučedolska golubica *(Vučedol dove)*, Knežev dvor *(Duke's Palace)*, Katedrala Sv. Jakova *(St Jacobs Cathedral)*, Eufrazijeva bazilika *(Euphrasian Basilica)*.

In search for web sites with general information on the cultural heritage of Croatia it was found that the Ministry of Culture of the Republic of Croatia does not have an English version of its web site (however, it does fund the *Culturenet Croatia* website). Interestingly, the web page "Croatia UNESCO Heritage" of the *Find Croatia* tourism portal had a rather low rank (59) despite the use of keywords in the title of web page and in meta tags. This could have been caused by too many (21) keywords in the meta tag of that web page that may have been perceived as *spamindexing* and penalized by the Google search engine. It must also be mentioned that the occurrence of keywords in the text of the analyzed web pages did not appear to have much impact on page ranking. Finally, one of the most interesting results of the analysis presented in Table 4 was that not more then 0-1 links were found pointing to the listed web pages on Croatian cultural heritage from outside of their web domain. This means that most of the

web sites in Table 4 can be located only by Internet users who already know of them, or by those who perform a search on one of the search engines.

Table 4. Analysis of the web page source regarding use of key words "cultural heritage Croatia" in meta tags and in the text of the web page

Document title for web page description and web address	Rank after Google search [1]	All keywords in title of web page ("<title>")	Document description in meta tags[2]	Number of links to web address[3]	Occurence of text[4] "cultur"/ "heritage"/ "Croatia"
"culturenet.hr – Panorama – Art – Croatian cultural heritage – monuments and sites" http://www.culturenet.hr/v1/english/panorama.asp?id=67	1	Yes	None	0	5/2/3
"Kamping udruženje Hrvatske – Cultural heritage of Croatia" http://www.camping.hr/druga.aspx?stranica=914&pid=69	3	Yes	None	0	4/4/5
"adriatica.net > Tourist guide – Cultural heritage" http://www.adriatica.net/common/destinations/features-culture_en.htm	4	Partly	Very good	0	6/9/12
"Croatia Information – Croatica.net Traveller" http://www.croatica.net/en/portal/hrvatska/	9	No	No	1	9/3/22
"Croatia – Croatian National Tourist Board" http://www.croatia.hr/English/Home/Naslovna.aspx	17	No	No	0	3/3/4
"Croatia UNESCO Heritage" http://www.find-croatia.com/unesco-heritage.html	59	Partly	Good	1	8/5/26
"MDC | Home page" http://www.mdc.hr/index_en.aspx	>100 8*, 55*	No	None	3	3/3/23
"Ministarstvo kulture Republike Hrvatske – KULTURNA BAŠTINA" http://www.min-kulture.hr/default.aspx?id=6	?	No web page in English	No web page in English	1	0/0/0

[1] Keywords used: cultural heritage Croatia; finding of another web page under the same web domain is marked with an asterisk (*)

[2] *None* – no keywords and no description tags are used within page meta tags.

Poor – inadequate number and poor selection of used keywords within page meta tags; no description tags.

Good – adequate number and proper selection of used keywords, but no description tags; or, adequate number but poor selection of used keywords, with description tags.

Very good – proper usage of meta tags within page; adequate number of properly selected keywords and description tags.

[3] A Google search was performed of web address enclosed in quotation marks.

[4] The occurrences were not counted if found in HTML, meta tags, web page title and URLs.

Search for information on cultural heritage by informatics students

It can be assumed that most internet users who needed to find information on cultural heritage would first consult the Internet and then perhaps go to a library or a museum. To verify this assumption a group of 48 students information systems at the Faculty of organization and informatics, University of Zagreb, Croatia, were asked to write in their own words what would be the first, second and third type of activity that they would perform if they needed to find additional information on a Croatian cultural heritage artifact of their choice (they had to select one item from a list of artifacts similar to the one presented in Table 3). The response of 92% of the students was that they would first look on the Internet and 8% responded that they would first go to the library. The list of their second and third choices was rather long and included going to the library, use of Internet/Wikipedia, consulting literature/encyclopedia, asking friends, asking a professor, going to a museum, asking one's parents, going to a tourist agency, etc.

There was much diversity in what students in our brief survey would do as their second and third activity, but the most common sequence of activities reported by 71% of students was to look on the Internet first, and then go to the library. This could reflect the typical behavior on an average Internet user to whom the search of the Internet is an activity that can be performed with much less effort than a visit to the library.

Conclusion

This paper examined various means for influencing the web page rank in search engine results. Since web search often starts with the use of a search engine (White and Iivonen, 1999) and since most Internet users do not proceed in their search for information past the third page of search results (iProspect, 2006) it is important to ensure the highest possible web page rank on the list of results of search engine use for web pages and web sites of culturally valuable information. As can be concluded from the data presented in Table 4, most of the analyzed web pages with content related to general information on cultural heritage of Croatia were not optimized for search engines. Among those sites that were analyzed, the *Culturenet* web site provided a significant amount of information on cultural heritage, but a closer inspection of this web site revealed that it needed improvement regarding usability and visual design to facilitate access to information. The *Museum Documentation Center* web site with its numerous online museums (and texts in Croatian and English language) is the premier resource on the web regarding Croatian cultural heritage and one of the most successful regional projects in its domain. However, this web site was not optimized for search engines at the time of our analysis and difficult to locate with the use of Google search engine.

Since most Internet users would first use this medium to find information about cultural heritage, and because their likely activity would be to use a search en-

gine, it is important not only to present cultural heritage information of the web, but also to optimize the cultural heritage web sites for search engines.

References

Coopee, Todd. *How to climb the search engine rankings*. // InfoWorld Test Center. InfoWorld Media Group, San Francisco, California, U.S.A. 2000. http://www.infoworld.com/articles/mt/xml/00/06/12/000612mtsearch.html (22 August 2007)

Fallows, Deborah. Search Engine Users. Washington, DC, USA: Pew Internet and American Life Project, 2005. http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf (22 August 2007)

Heng, Christopher. *How to Improve Your Search Engine Ranking on Google*. thesitewizard.com. 2007. http://www.thesitewizard.com/archive/google.shtml (22 August 2007)

Heng, Christopher. *Keyword Density in Search Engine Ranking*. thesitewizard.com. 2006. http://www.thesitewizard.com/archive/keyworddensity.shtml", thesitewizard.com. 2007. (22 August 2007)

InternetWorldStats. *Internet usage statistics – The Big Picture: World Internet Users and Population Stats* [Online]. // Internet World Stats, Miniwatts Marketing Group. 2007. http://www.internetworldstats.com/stats.htm (22 August 2007)

iProspect. Search Engine User Behavior Study. Watertown, MA, USA: iProspect.com, inc. 2006. http://www.iprospect.com/premiumPDFs/WhitePaper_2006_SearchEngineUserBehavior.pdf (22 August 2007)

Kando, Noriko; Adachi, Jun. Cultural heritage online: Information access across heterogeneous cultural heritage in Japan. // *International Symposium on Digital Libraries and Knowledge Communities in Networked Information Society, DLKC'04*. Tsukuba, Ibaraki, Japan. 2004. http://www.kc.tsukuba.ac.jp/dlkc/e-proceedings/papers/dlkc04pp136.pdf (22 August 2007)

Khaled, A.F. Mohamed. The impact of metadata in web resources discovering. // *Online Information Review*, 30 (2006), 2, 155-167

Mac Aoidh, Eoin; Bertolotto, Michela; Wilson, C. David. Capturing spatial interactions to personalise cultural heritage access. // *International Workshop on Personalization Enhanced Access to Cultural Heritage*. Corfu, Greece. 2007. http://www.cs.vu.nl/~laroyo/PATCH/Proceedings/7-Aoidh-final.pdf (22 August 2007)

Moen, E. William. Accessing distributed cultural heritage. // *Communications of the ACM*, vol. 41, no. 4, 1998, 45-48

Nielsen/NetRatings. Nielsen//NetRatings Announces August U.S. Search Share Rankings [Online]. Nielsen//NetRatings, The Nielsen Company, Haarlem, The Netherlands. 2007. http://www.netratings.com/pr/pr_070919.pdf (19 September 2007)

Ossenbruggen, Jacco; Amin, Alia; Hardman, Lynda; Hildebrand, Michiel; Assem, Mark; Omelayenko, Borys; Schreiber, Guus; Tordai, Anna; de Boer, Victor; Wielinga, Bob; Wielemaker, Jan; de Niet, Marco; Taekema, Jos.; Orsouw, Marie-France; Teesing, Annemiek. Searching and Annotating Virtual Heritage Collections with Semantic-Web Techniques. // *Proceedings of the Museums and the Web 2007: The International Conference for Culture and Heritage* [Online]. San Francisco, California, U.S.A. 2007. http://www.archimuse.com/mw2007/papers/ossenbruggen/ossenbruggen.html (22 August 2007)

Rainie, Lee; Shermak, Jeremy. Search engine use shoots up in the past year and edges towards email as the primary internet application. [Data Memo].Washington, DC, USA: Pew Internet and American Life Project, 2005. http://www.pewinternet.org/pdfs/PIP_SearchData_1105.pdf (22 August 2007)

Rutledge, Lloyd; Aroyo, Lora; Brussee, Rogier; Stash, Natalia; Wang, Yiwen; Cramer, Henriette; Evers, Vanessa; Ramlal, Satyan. Recommendation with semantics for cultural heritage. // *International Workshop on Personalization Enhanced Access to Cultural Heritage*, Corfu, Greece. 2007. http://www.cs.vu.nl/~laroyo/PATCH/Proceedings/PATCH.pdf (22 August 2007)

Shapiro, Yelena. *How do I Improve Search Engine Rankings?* SearchEngines.com. 2007. http://www.searchengines.com/searchEnginesRankings.html (22 August 2007)

Shapiro, Yelena; Lehoczky, Etelka. Factors that Influence Search Engine Rankings. SearchEngines.com. 2007. http://www.searchengines.com/ranking_factors.html (22 August 2007)

Smeulders, W.M. Arnold; Hardman, Lynda; Schreiber, Guus; Geusebroek, Jan-Mark. An integrated multimedia approach to cultural heritage e-documents. // *4th Intl. Workshop on Multimedia Information Retrieval, in conjunction with ACM Multimedia 2002*. Juan-les-Pins, France. 2002. http://homepages.cwi.nl/~media/publications/Smeulders02a.pdf (22 August 2007)

Thompson, Dave. *The influence of metatags on web-based search retrieval, ranking and relevancy*. Dalhousie University, Halifax, Nova Scotia, Canada. 2002. http://www.ecomm.dal.ca/emec/research/emec-final-april-22.pdf (22 August 2007)

Uzelac, Aleksandra. Cultural Networks and Cultural Portals – is there a difference? // *Colloquium "Interactive Culture: Culture and Online Information"*. Nantes, France. 2005. http://www.culturenet.hr/v1/english/about.asp (22 August 2007)

White, Marilyn Domas; Iivonen, Mirja. Factors Influencing Web Search Strategies. // *ASIS Annual Conference: Proceedings of the 62nd Annual Meeting*. Washington, DC, U.S.A. 1999. http://www.wam.umd.edu/~whitemd/ASIS_pap.pdf (22 August 2007)

The Use of Social Software in the Online Presentation of Cultural Heritage

Goran Bubaš
University of Zagreb, Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
goran.bubas@foi.hr

Igor Balaban
University of Zagreb, Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
igor.balaban@foi.hr

Miran Zlatović
University of Zagreb, Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
miran.zlatovic@foi.hr

Summary

Social software like wikies and weblogs are popular online publication media and a source of information for a substantial percentage of Internet users. An analysis of some components of the cultural heritage of Croatia in the Wikipedia is provided. The links to the related content of the Wikipedia are often among the first ten links offered in the search results when engines like Google are used to locate information on the Croatian cultural heritage. There is growing interest from libraries and educational and cultural institutions to use weblogs and wikies for various purposes. This paper analyzes the potential of those technologies for the online presentation of cultural heritage.

Keywords: social software, Wikipedia, wiki, blog, cultural heritage

Introduction

Management of heritage sites may have an important impact on local communities in developing countries (Hampton, 2005). Heritage sites are of great value for cultural tourism, but they can also profit from tourism because it can generate significant funding for their management. A number of heritage sites in the countries of Central and Eastern Europe could profit from the expansion of cultural tourism, while communication technology (ICT) could be used to attract more visitors by generating interest to view the original sites and objects (Cabrini, 2002). It must be noted that a substantial percentage of travelers who

are Internet users in countries like the USA choose the Internet for travel and destination information (TIA, 2006). Web-based tourism information systems like cultural web portals could provide more information on cultural heritage to tourists, but such information services should be ubiquitous, i.e. enable access to cultural information from any device (PC, PDA, mobile phone, etc.) as was intended in one project for Mediterranean countries (Garzotto *et al.*, 2004)

New media and virtual reality provide considerable potential for the presentation of cultural heritage information, even though the related goals of documentation for preservation purposes, accurate historical interpretation and high quality visualization for the public are rarely fully realized in practice (for the examples of digital technology use, see Addison, 2000). There are various positive impacts from the use of technology in presenting cultural heritage (adapted from Thwaites, 2001):

a) increase of public interest in cultural heritage,
b) greater accessibility to information about remote or closed heritage sites,
c) variety of rich information sources,
d) greater awareness of global humanity,
e) encouragement of virtual tourism,
f) means of recording, preserving, and interpreting heritage, and educating people about it,
g) cross-cultural and inter-cultural communication.

The web is considered as the main communication channel for enabling universal and flexible access to digitalized cultural heritage resources (Kelly *et al.*, 2003). Many heritage sites have web sites that are maintained by the government, a heritage institution, or a local tourism center since the development of such Internet-based information resources can contribute to public awareness and fund-raising efforts (UNESCO, 2007, p. 71). A good example of effective use of the Internet to promote cultural heritage is the "Eternal Egypt" (http://www.eternalegipt.org) web site as one of the most successful cultural heritage projects in Egypt that used virtual reality and other up-to-date web technologies (Saleh and Barakat, 2005). It must also be emphasized that heritage web sites have an important educational potential that can be better realized by proper learning design (Brown, 2006).

Heritage institutions like museums should try to respond to the evolving needs of the public and try to increase interest in their collections by developing the following online services (adapted from Anani, 2005):

- interactive simulations and presentations to enhance pre-visit experiences,
- post-visit online discussion forums,
- collaborative virtual learning environments,
- interactive e-workshops.

The combination of multimedia capabilities with a wireless connection of new mobile devices (handheld computers and 3G mobile phones) creates an addi-

tional challenge for the development of cultural heritage information services (Brelot *et al.*, 2007). Recently, there has been increasing effort to include Web 2.0 technologies in the presentation and interactive/collaborative development and exchange of heritage information (for instance, see Webmoore, 2007). However, to facilitate access to online information on museum collections, it is suggested to optimize the web site for search engines like Google (Chan, 2007). The authors will focus below on the use of the Wikipedia and social software like wikies and weblogs for the presentation of information on cultural heritage.

Wikipedia and Croatian cultural heritage

According to the latest data on global Internet traffic, the Wikipedia is among the 10 most visited global sites (comScore, 2007) and it is also the most visited web site by the populations of Internet users in western countries like the USA, Germany and France. In the USA, 36% of adult Internet users consult the Wikipedia (about 8% of them do this on a daily basis) and for the population of Internet users in the USA the Wikipedia is the most visited web site after they perform a Google search (in fact, about half of all the visits to the Wikipedia result from a Google search; see Rainie and Tancer, 2007).

At the time of the analyses performed for this paper, there was a limited amount of content on the cultural heritage of Croatia (i.e. articles on Croatian history, culture and art) in the English version of the Wikipedia (www.wikipedia.org). For instance, in August 2007, the Wikipedia articles on the major cities of Croatia like Zagreb, Rijeka, Osijek and Split commonly included an overview of their history but in most cases only a brief list of cultural sites. It must be noted that the Wikipedia included separate articles on the *World Heritage Sites* in the towns and cities of Poreč, Šibenik, Trogir, Split, and Dubrovnik. In August 2007, a brief analysis (number of words, images, and links at the end of the article) of the historical and cultural content of the articles in the Wikipedia about the World Heritage Sites in Croatia was performed and gave the following results:

- Euphrasian Basilica, Poreč (240 words, 1 image, 2 listed external links);
- St. Jacob's Cathedral, Šibenik (298 words, 6 images, 3 listed external links);
- Historic town of Trogir (618 words, 2 images, 6 listed external links);
- Diocletian's Palace, Split (939 words, 3 images, no listed external links);
- Old town of Dubrovnik (2280 words, 19 images, 6 listed external links).

For the Euphrasian Basilica in Poreč and St. Jacob's Cathedral in Šibenik, the textual information in the articles of the English version of the Wikipedia was limited to fewer than 300 words. The historic town of Trogir and Diocletian's Palace in Split were presented with only 2 images, and only Dubrovnik was presented with both sufficient textual information and images. These data can easily be verified since each Wikipedia page has a history of its editing which

makes it possible to check its previous versions and analyze how the online content of a Wikipedia page has grown or been modified over time.

The use of search engines is one of the most frequent activities of Internet users. In the USA, on an average day, more than 60% of adult Internet users use *Google*, *Yahoo!*, *MSN Search* or some other search engine (Rainie and Shermak, 2005). Most Internet users trust their search engines and, among other things, they use them to seek information on places, travel, education, humanities, culture, etc. (Fallows, 2005). However, a recent study revealed that 62% of search engine users click on a link within the first page of results, and that 90% of them click on a link within the first three pages of search results, but also that only 12% of users are willing to continue their search beyond the third page of results (iProspect, 2006). The use of a search engine is one of the likely activities of an Internet user who seeks information on a specific object of cultural heritage. Therefore, it would be interesting to check the rank of Wikipedia web pages that contain articles on objects of Croatian cultural heritage after a search is performed on the two most popular search engines (Google and Yahoo!) and also on the search engine *Pogodak!* which specializes in Croatian web space. The results of such searches are presented in Table 1.

The data presented in Table 1 indicate that most of the articles on the web pages of the Croatian version of the Wikipedia (marked with "hr") related to selected objects of Croatian cultural heritage were ranked first or second after a search performed with a Google search engine. For instance, when the key words "Pulska Arena" (without quotation marks) were used for a Google search, the link to the web page of the Croatian Wikipedia on this topic ("Pulska arena – Wikipedija", http://hr.wikipedia.org/wiki/Amfiteatar_u_Puli) was ranked first on the list of search results. Furthermore, if the search was performed on the Pogodak! search engine, which is limited to Croatian web space, the articles on the Croatian Wikipedia were always ranked first. However, the ranks of the Croatian Wikipedia web pages related to cultural heritage in the results of the Yahoo! search were slightly different, but in most cases they had a rank of 1-5 and appeared on the first page of search results.

As can be observed in the data presented in Table 1, among the first 100 results of the Google and Yahoo! search engines, some of the objects of Croatian cultural heritage were also located on the web pages of the English (en), Bosnian (bs) and Serbo-Croatian (sh) versions of Wikipedia. Interestingly, the related web pages of Slovenian, German and Polish Wikipedia articles were also listed in some of the first 100 search results, but these data are not presented in Table 1.

Table 1. Rank of a link to a web article on Croatian cultural heritage objects in several versions of Wikipedia after a search on the Google, Yahoo! and Pogodak! search engines (the searches were performed in August 2007)

	Rank of related article of the Wikipedia (hr, bs, sh, en)†		
Name of heritage object	Rank after a Google search	Rank after a Yahoo! search	Rank after a Pogodak! search
Bašćanska ploča *(The Baska Tablet)*[1]	2 (hr), 35 (bs)	3 (hr), 5*+7* (bs), 10+13* (en)	1 (hr)
Plominski natpis *(The Plomin Tablet)*[2]	2 (hr), 20* (sh)	2 (hr)	1 (hr)
Višeslavova krstionica *(The Baptistery of Višeslav)*[3]	1 (hr), 63* (sh), 68* (sr)	3 (hr), 5* (en)	1 (hr)
Crkva sv. Križa *(The Church of the Holy Cross)*[4]	2 (hr)	5 (hr)	1 (hr)
Vinodolski zakon(ik) *(The Vinodol Law Code)*[5]	4 (hr), 20* (bs), 73 (bs)	3 (hr), 22* (sh)	1 (hr)
Povelja kneza Trpimira *(Trpimir's Charter)*[6]	2 (hr), 40 (sh)	1 (hr)	1 (hr)
Vučedolska golubica *(Vučedol dove)*[7]	1 (hr), 77* (bs)	1 (hr), 6* (sh), 7*+10* (en)	1 (hr)
Pulska Arena *(The Coliseum in Pula)*[8]	1 (hr), 24* (bs)	3 (hr), 15* (bs)	1 (hr)
Katedrala Sv. Jakova *(St Jacob's Cathedral)*[9]	1 (hr)	1 (hr), 23 (en)	1 (hr)
Eufrazijeva bazilika *(Euphrasian Basilica)*[10]	1 (hr)	7 (hr), 17 (en), 42* (sh)	1 (hr)

† Besides the English (en) Wikipedia, there are numerous other versions of Wikipedia including the Croatian (hr), Bosnian (bs) and Serbo-Croatian (sh) version.

* Search results that point to the Wikipedia articles on a different topic, but in which the cultural heritage object is briefly mentioned.

[1] "Bašćanska ploča" is a stone tablet with 13 lines of early text written in Croatian Glagolitic found near Baška on the island of Krk.

[2] "Plominski natpis" is a Croatian Glagolitic text on the wall of St George's Church in Plomin, Istria.

[3] "Višeslavova krstionica" is a Baptistery built by Duke Višeslav at the time of the conversion of Croatians to Christianity.

[4] "Crkva Sv. Križa" is a small church in the coastal town of Nin dating from the 9th century which represents the pre-Romanesque period of Croatian architecture.

[5] "Vinodolski zakonik" is the oldest legal monument written in Croatian Glagolitic in 1288 AD which originates from the coastal town of Vinodol.

[6] "Trpimirova povelja" is the first known charter of a Croatian ruler (Duke Trpimir) written in Latin in 852 AD (preserved only as a transcript dating from the year 1568).

[7] "Vučedolska golubica" is a ceramic pot/vessel in the form of a dove from the Eneolithic period that was found at the archaeological location of Vučedol near the town of Vukovar.

[8] "Pulska arena" is one of the largest and best preserved coliseums in the world dating back to the first century AD.

[9] "Katedrala Sv. Jakova" is a Christian church in the city of Šibenik that was built in the 16th century and has appeared on the UNESCO World Heritage List since 2000.

[10] "Eufrazijeva bazilika" is a Christian basilica built in the 6th century during the period of Bishop Euphrasius and is famous for its mosaics, representations of Byzantine art.

In an analysis of the content of the Croatian version of Wikipedia in relation to the heritage objects listed in Table 1, it was found that they were presented in one primary thematic article and mentioned in 1 to 6 other articles (these other articles were thematically related to the geographical location, such as a city, a town or a village, or to the broader historical or cultural topic: see Table 2). As can be observed from the data presented in Table 2, seven thematic articles on Croatian heritage objects in the Croatian version of Wikipedia consisted of fewer than 300 words, and only the articles on the Vinodol Law Code, St Jacob’s Cathedral, and the Baska Tablet were presented in more detail. Even though the primary thematic articles of these heritage objects in the Croatian version of the Wikipedia were highly ranked on the lists of the search results of the Google, Yahoo! and Pogodak! search engines and were very likely to be selected for viewing by those who performed related searches, most of the articles had only a small amount of textual information about the heritage objects and contained from 0 to 3 photographic images.

Table 2. The number and length (in number of words) of thematic articles on selected Croatian cultural heritage objects in the Croatian version of the Wikipedia and the number of other articles in which the heritage object was mentioned (in August 2007)

Name of heritage object	Number of thematic articles	Number of words in thematic article	Mentioned in other articles
Bašćanska ploča *(The Baska Tablet)*	1	822	6
Plominski natpis *(The Plomin Tablet)*	1	48	0
Višeslavova krstionica *(The Baptistery of Višeslav)*	1	248	1
Crkva sv. Križa *(The Church of the Holy Cross)*	1	169	1
Vinodolski zakon(ik) *(The Vinodol Law Code)*	1	3.621	3
Povelja kneza Trpimira *(Trpimir's Charter)*	1	214	1
Vučedolska golubica *(Vučedol dove)*	1	71	4
Pulska Arena *(The Coliseum in Pula)*	1	229	2
Katedrala Sv. Jakova *(St Jacob’s Cathedral)*	1	1.191	4
Eufrazijeva bazilika *(Euphrasian Basilica)*	1	291	2

Social software tools in the presentation of cultural heritage

The use of the wiki in presenting a museum collection is recommended when there is interest in public participation or in remote collaboration on the development of online content. The posting of general information on the museum on the Wikipedia should also be considered for promotional and other purposes (Bowen *et al.*, 2007).

Social software tools, such as the wiki, have great potential to enable visitors of various cultural heritage institutions to co-create the contents of the exhibitions. In one case of ICT (not a wiki) for similar purposes, visitors' theories and opinions regarding the museum's exhibition objects were recorded and could later be read by other visitors, thus enabling active contribution to the contents of the exhibit (Bannon et al., 2007). Research projects which investigate social interactions related to cultural experience are emerging – one Belgian project called Heritage2.0 (Luyten and Nulens, 2007) aims to link such interaction to a digital environment which is location-sensitive. One of the goals of the Heritage2.0 project is to enrich visitors' experiences before an actual visit, during the visit and after the visit. Wiki technology can also be used in those phases of interaction with visitors of heritage sites, but pre- and post-visit phases seem to be particularly well suited for its use – the wiki makes it easy for visitors to make recommendations, express opinions, conduct discussions, etc.

There are also growing initiatives to use up-to-date ICT to create open interfaces between museum staff and visitors, which support interaction and collaboration in order to connect them in a participatory design process (Hall and Bannon, 2006; Walker, 2007). Wiki tools can easily facilitate such initiatives, especially if a museum or other type of heritage institution already has some kind of traditional web site for presenting its heritage artifacts. Since traditional web sites and interactive wiki sites are based on comparable standards and technologies, their integration is fairly simple, and thus the wiki can easily enhance the existing online presentation of heritage artifacts.

Another potential wiki application is in the collaborative development of online content through interaction between cultural heritage professionals. According to the results of a case study performed by Ontario Ministry of Culture (Bentley, 2007), such interactions can promote preservation of the heritage, enhance the management of artifacts and improve the exchange of knowledge between professionals. Since most of the available wiki tools can be configured to allow editing and reading capabilities only for a closed group of users (such as a group of cultural heritage professionals), wiki technology can be used effectively in such circumstances.

Weblogs, blogs or web-based diaries can also be used to promote cultural heritage. Weblogs are online documents created by social software tools. The content of weblogs is intended for public consumption and their creation does not require knowledge of HTML to enable publication on the web. Because of its simplicity, the technology of weblogs has become widely used, at first as a means of pub-

lishing diaries (personal thoughts, comments, experiences, philosophies, etc.) and more recently for the creation of online content related to almost every field of human interest. A recent study (Sifry, 2006) reported that about 100,000 new blogs were created each day and that *Technorati* (a major blog search engine, www.technorati.com) was tracking more than 57 million blogs.

In the context of cultural heritage, weblogs are mostly used to present historical centers and museums. Visitors can be enabled to write their own impressions and reviews on a specific museum blog, or to comment on the work of other bloggers. In that way, museum staff can gain useful information from visitors by reading their reviews and comments and at the same time visitors can gather information on the personal experience of other visitors about topics of their interest and can interact with them. Furthermore, cultural enthusiasts can be involved in creating and maintaining a museum blog on a specific topic.

An interesting example of a very well organized weblog is the *Museum Blogs* web site (see: http://museumblogs.org/category/cultural-heritage). It is a comprehensive directory and blog covering the latest news from art museums, science centers, and other museum-related bloggers. Its *Cultural Heritage* category contains over 200 searchable museum-related blogs and pieces of news. Visitors can browse blogs by category, search blogs by specific criteria, and subscribe to the most interesting blogs. They can also read news and post comments. *Museum Blogs* is meeting the growing need of cultural heritage organizations all over the world to use social software and web-based technologies. Unfortunately this kind of initiative still does not exist in Croatia.

Conclusion

The results of the analysis presented in this paper indicate that the one of the most effective channels for the presentation of the cultural heritage of Croatia is through the content of the English and Croatian versions of the Wikipedia. As mentioned earlier in this paper, many Internet users consult the Wikipedia which is among the 10 most visited global web sites in western countries. The Wikipedia is also one of the most visited web sites after a Google search has been performed. However, for tourists who plan to visit some of the major towns and cities in Croatia, in most cases there is only limited content on the local cultural heritage in the English Wikipedia articles about those places. Also, for the users of the Croatian version of the Wikipedia, the content of most of the articles on Croatian cultural heritage objects is approximately the size of an abstract of a conference paper (100-300 words). It can be concluded that for the presentation of Croatian cultural heritage, the Wikipedia was ineffectively used at the time the analyses were performed for this paper (August 2007). Bearing in mind that any teacher of history or art at a high school or college institution can give his/her students an assignment to extend these articles, and that the expansion of the articles about the Croatian cultural heritage on the Wikipedia can also be done by professionals, the state of the Wikipedia articles on Croatian

cultural heritage in August 2007 was far from satisfactory. For institutions which have web sites on Croatian cultural heritage, it is opportune to participate in this effort and place links to their web sites in related Wikipedia articles. For museums and institutions which manage heritage sites, it is advised to create Wikipedia articles on their collections and artifacts.

This paper has also provided a brief theoretical analysis of the potential use of social software like wikies and blogs for the presentation of cultural heritage and for interaction with the visitors of museums and other heritage sites. The use of such software is relatively simple and its installation is quite inexpensive for heritage institutions which already manage a web site. Wikies and blogs can enable a more personalized, collaborative and interactive presentation of heritage content. They can attract attention and provide information in a less official way before and after a visit to a museum or heritage site. They also enable visitors to contribute to creating related textual, photographic and video content.

References

Addison, C. Alonzo. Emerging trends in virtual heritage. // *Multimedia*, 7 (2000), 2, 22-25

Anani, Namir. Sustainable engagement in digital heritage – The challenges of learning environments for heritage institutions. // *Museum International*, 57 (2005), 1-2, 142-143

Bannon, Liam; Ciolfi, Luigina; Fernstrom, Mikael. Visitors' contributions as cultural heritage: designing for participation. // *The International Conference for Digital Culture & Heritage ICHIM07 Abstracts* [Online]. Toronto, Ontario, Canada: Archives & Museum Informatics, 2007. http://www.archimuse.com/ichim07/abstracts/prg_335001526.html (22 August 2007)

Bentley, Alan. The impact of advanced information communication technologies on heritage conservation. // *The International Conference for Digital Culture & Heritage ICHIM07 Abstracts* [Online]. Toronto, Ontario, Canada, Archives & Museum Informatics, 2007. http://www.archimuse.com/ichim07/abstracts/prg_335001523.html (22 August 2007)

Bowen, P. Jonathan; Lisney, Eleanor; Filippini-Fantoni, Silvia; Bernal, Isabel. A Museums Wiki. // *Proceedings of the Museums and the Web 2007: The International Conference for Culture and Heritage* [Online]. San Francisco, California, USA: Archives & Museum Informatics, 2007. http://www.archimuse.com/mw2007/papers/bowen/bowen.html (22 August 2007)

Brelot, Marc; Cotamanac'h, Alexandre; Damala, Areti; Kockelkorn, Hub. Nomadic computing in indoor cultural settings: Intelligent connectivity, context awareness and the mobile museum experience. // *Digital Culture & Heritage, ichim05 Proceedings* [Online]. Paris, France. 2007. http://www.ichim.org/ichim05/contenu/PDF/S08-Kockelkorn.pdf (22 August 2007)

Brown, Stephen. Do richer media mean better learning? A framework for evaluating learning experiences in museum web site design. // *International Journal of Heritage Studies*, 12 (2006), 5, 412-426

Cabrini, Luigi. Cultural heritage and tourism development. // *International Conference on "Heritage, New technologies and Local development"* [Online]. Ghent, Belgium. 2002. http://www.unwto.org/regional/europe/PDF/SPEECHES/2002/Ghent%20_Belgium_.pdf (22 August 2007)

Chan, Sebastian. Tagging and Searching – Serendipity and museum collection databases. // *Proceedings of the Museums and the Web 2007: The International Conference for Culture and Heritage* [Online]. San Francisco, California, USA. 2007. http://www.archimuse.com/mw2007/papers/chan/chan.html (22 August 2007)

comScore. comScore Releases Worldwide Rankings of Top Web Properties [Online]. Reston, VA, USA: comScore, Inc., 2007. http://www.comscore.com/press/release.asp?press=1524 (22 August 2007)

Fallows, Deborah. Search Engine Users. Washington, DC, USA: Pew Internet and American Life Project, 2005. http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf (22 August 2007)

Garzotto, Franca; Paolini, Paolo; Speroni, Marco; Proll, Birgit; Retschitzegger, Werner; Schwinger, Wieland. Ubiquitous access to cultural tourism portals. // *Proceedings of the Database and Expert Systems Applications, 15th International Workshop (DEXA'04)*. Washington, DC, USA: IEEE Computer Society, 2004, 67-72

Hall, Tony; Bannon, Liam. Designing Ubiquitous Computing to Enhance Children’s Learning in Museums. // *Journal of Computer Assisted Learning*, 22 (2006), 231-243

Hampton, P. Mark. Heritage, local communities and economic development. // *Annals of Tourism Research*, 32 (2005), 3, 735-759

iProspect. Search Engine User Behavior Study. Watertown, MA, USA: iProspect.com, inc. 2006. www.iprospect.com/premiumPDFs/WhitePaper_2006_SearchEngineUserBehavior.pdf (22 August 2007)

Kelly, Brian; Dunning, Alastair; Guy, Marieke; Phipps, Lawrie. Ideology or pragmatism? Open standards and cultural heritage web sites. // *Cultural Institutions and Digital Technology, ichim03 Proceedings*. Paris, France. 2003. http://www.ichim.org/ichim03/PDF/086C.pdf (22 August 2007)

Luyten, Kris, Nulens, Gert. Social multimedia content distribution in a network of heritage sites. // *The International Conference for Digital Culture & Heritage ICHIM07 Abstracts* [Online]. Toronto, Ontario, Canada: Archives & Museum Informatics, 2007. http://www.archimuse.com/ichim07/abstracts/prg_335001552.html (22 August 2007)

Rainie, Lee; Shermak, Jeremy. Search engine use shoots up in the past year and edges towards email as the primary internet application [Data Memo].Washington, DC, USA: Pew Internet and American Life Project, 2005. http://www.pewinternet.org/pdfs/PIP_SearchData_1105.pdf (22 August 2007)

Rainie, Lee; Tancer, Bill. 36% of Online American Adults Consult Wikipedia [Data Memo]. Washington, DC, USA: Pew Internet and American Life Project, 2005. http://www.pewinternet.org/PPF/r/212/report_display.asp (22 August 2007)

Saleh, Fathi; Barakat, N. Hala. The Global Village of Heritage: The contribution of the Centre for Documentation of Cultural and Natural Heritage (CULTNAT). // *Museum International*, 57 (2005), 1-2, 73-78

Sifry, Dave. State of the Blogosphere, August 2006 [Report]. Sifry’s Alerts, 2006. http://www.sifry.com/alerts/archives/000436.html

Thwaites, Hal. Fact, fiction, fantasy: The information impact of virtual heritage. // *Proceedings of the Seventh International Conference on Virtual Systems and Multimedia VSMM'01*. Washington, DC, USA: IEEE Computer Society, 2001, 263-270

TIA. Leading travel industry consumer survey reports significantly more travelers plan and book trips online [Press release]. Washington, DC, USA: Travel Industry Association, 2005. http://www.tia.org/pressmedia/pressrec.asp?Item=689 (22 August 2007)

UNESCO. World Heritage Series n°20 - Periodic Report and Action Plan, Europe 2005-2006. Paris, France: United Nations Educational, Scientific and Cultural Organization (UNESCO), 2007. http://whc.unesco.org/pg.cfm?cid=162&l=en&&&index=1 (22 August 2007)

Walker, Kevin. Visitor-constructed personalized learning trails. // *Proceedings of the Museums and the Web 2007: The International Conference for Culture and Heritage* [Online] San Francisco, California, USA: Archives & Museum Informatics, 2007. http://www.archimuse.com/mw2007/papers/walker/walker.html (22 August 2007)

Webmoore, Timothy. Taking 'Yahoo!s' seriously: new media and the platform shift in cultural heritage. // *Paper submitted to the Visual Anthropology Review*, (2007). http://anthropology.usf.edu/cma/TimWebmoor-TakingYahoo!sSeriously-AAA2006.pdf (22 August 2007)

Virtual Worlds: Heritage and Educational Challenge

Mihaela Banek Zorica
Department of Information Sciences
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
mbanek@ffzg.hr

Sonja Špiranec
Department of Information Sciences
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
sspiran@ffzg.hr

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

Summary

The authors give a brief overview of the development and current trends in the field of virtual worlds. The emphasis is put on one of the social network virtual worlds - Second life and its usage in educational environment and heritage preservation. Furthermore, the first phases of creating of the extended educational environment of the Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb are shown.

Key words: Virtual worlds, MMORPG, Second life, education, heritage

Introduction

Since its development and inception, ICT has changed every facet of human existence and established itself as a part of the fabric in social and economic enterprises as well as in entertainment, education, culture etc. Heritage communities have recognized these potentials by using networks and multimedia as important tools for the enhancement and upgrade of their services. Nevertheless, those services where purely focused on the presentation of the respective

institutions, their contents and services, in a mainly statical manner and one-directional in nature. This implies that users where supplied with information from the web, making them largely consumers of information resources that are delivered to them, which was largely satisfying, even fascinating for the first generation of web users that where not born to the digital world, but have learned to cope with it - the digital immigrants.

With the generation of digital natives the time was ripe for entering a new, more social and participatory phase of the Web, denoted as Web 2.0. Digital natives have grown up around instantly accessible information and within a networked culture, in crossfire of multimedia stimuli, which resulted in the development of cognitive thinking patterns, expectations and methods of deriving meaning that differ from previous generations. With this new generation, the world of education and the heritage sector have got their new user, which motivated both worlds to transform their Web 1.0 capacities and dive into the Web 2.0 world, initiating the concepts of Library 2.0, Museum 2.0, Education 2.0, Heritage 2.0 etc.

Although the above described concepts have realized the transformation of educational and heritage institutions from places of passive information consumption to dynamic, participative and creative knowledge production spaces, they still just partially offer a sense of ownership, are still passive and flat in nature[1] and therefore do not necessary appeal to users who are more attuned to 3D entertainment software and simulations. This explains the recent emergence of the new, 3.0 hype, focused on translating the 2.0 concepts to tangible, 3-dimensional social networking systems. Those are characterized by rich, cross-institutional, cross-cultural educational opportunities within which the learners themselves play a key role as creators of knowledge artefacts that are shared, and where social networking and social benefit outside the immediate scope of activity play a strong role. The distinction between artefacts, people and process becomes blurred, as do distinctions of space and time.[2]

Virtual worlds

A virtual world is a computer-based simulated environment intended for its users to inhabit and interact via avatars. This habitation usually is represented in the form of two or three-dimensional graphical representations of humanoids (or other graphical or text-based avatars). Some, but not all, virtual worlds allow for multiple users.[3] It is an animated three-dimensional world created with com-

[1] Keats, D. Smith, J. P. The genesis and emergence of Education 3.01 in higher education: the potential for Africa. // First Monday. 12.3.2007. http://www.firstmonday.org/issues/issue12_3/keats/index.html#author

[2] Ibid.

[3] Virtual world http://en.wikipedia.org/wiki/Virtual_world

puter graphics imagining (CGI) and other rendering software. The main feature of these worlds is that they are available to their users 24/7 and run on servers which allow users from all over the world to connect and interact with each other. Although it is computer simulated it functions almost as a real world with all the real world rules like gravity, topography, locomotion, real-time actions, and communication. As the residents of the virtual world are called avatars they represent humans alter-egos' where people create their virtual identities. There are a lot of psychological research on how real life people act in virtual worlds and what kind of virtual identity they give themselves. Furthermore, in most cases they have almost the same needs as in real life i.e. communicate with other, create special interest groups etc.

Kelly describes virtual worlds as separate universe with its own physical and sociological rules. It's a recognizable universe. You can't walk through walls, for example. And you don't fall upward when you slip. But it's also a wholly different universe. If a rock lands on your head, you feel no pain. And if you cast a spell on yourself, you can fly, or breathe underwater, or shoot lightning bolts from your fingertips. They are simultaneously identical to the real world and completely different from it.[4]

If we look at the origin and development of the virtual worlds we can go back to the 1970s and connect them with the IRC (Internet Relay Chat) and chat rooms which have evolved into MUDs (Multi-User Dungeon) and MUSHes (Multi-User Shared Hack) – text based communities and multiplayer games based on role playing adventures. During the middle 1990s 3D virtual worlds started to emerge. The philosophy was almost the same but the environment, due to the ICT advancement, started to change from the pure text based to the visually represented world resembling the computer games.

As virtual world is a fairly vague and inclusive term, currently we are mostly using the term *Massively Multiplayer Online Role-Playing Games* (MMORPG) where the user is either playing a specific character in the game or edits and alters their avatar in order to play a dynamic role. We can divide MMORPGs in two main categories: thematic ones or fantasy-thematic with various subcategories (such as World of Worcraft – WOW, *EverQuest II*, Eve…etc.), and non-thematic or socially oriented (such as Second Life, ActiveWorlds… etc.).

The main goal of the thematic MMORPGs is the development of the avatar or the player's character which is done by performing tasks and actions which give them points and "experience". On the other hand non-thematic ones are more oriented in building and creating worlds and avatars and offer more opportuni-

[4] Kelly, R.V. Massively Multiplayer Online Role-Playing Games: The People, the Addiction and the Playing Experience. McFarland & Company, 2004.

ties to transform real life objects into the virtual ones. The heavily rely on player-created content, including everything from simple animations to complete buildings using player-created textures and architecture. However, these games are very different from the far more popular "standard" MMORPGs revolving around combat and limited character trade skills. Therefore, our main focus will be on these kinds of worlds and the opportunities and challenges they offer to the education and preservation of heritage.

Second life

One of the virtual worlds mentioned earlier is Second Life (SL) developed by the Linden Lab in 2003. Although it is an Internet based virtual world user needs to download a client program (Second Life Viewer) which enables *Residents or Avatars* to interact with each other as well as with the objects created in this world. It is a semi-structured virtual environment where characters undertake activities for the purpose of personal enjoyment. Since opening to the public, it has grown explosively and today is inhabited by millions of Residents from around the globe. Currently there are more than 9 million Residents although not all are the real ones (some have two avatars, some have never logged after creating an account) and there is usually around 1 million users online.

This is a user defined world for general use and it is not a typical game with points, scores, levels etc. Of course one can develops certain skills like building, scripting and create buildings, landscape, vehicles, furniture, and machines to use, trade, or sell. This is a primary source of activity in the economy. Any Resident can also make *gestures* from small animations and sounds from the standard library. Outside *Second Life*, Residents can use various graphics, animation, and sound tools to create more elaborate objects, and upload them into the world.

While the Second Life interface and display are similar to most popular massively multiplayer online role playing games (or MMORPGs), there are two key, unique differences[5]:

1. **Creativity:** Second Life provides near unlimited freedom to its Residents. This world really is whatever they make it, and their experience is what they want out of it.
2. **Ownership:** Instead of paying a monthly subscription fee, Residents can obtain their first Basic account for free or choose a premium account where they pay a monthly fee and get their own land.

Second life and educational institutions

In principle, virtual worlds represent a powerful new media for instruction and education. They can provide new methods for learning evaluation and teacher

[5] Second Life FAQ http://secondlife.com/whatis/faq.php

professional development, including embedded assessment and teacher training linked directly to student performance. A growing number of educational institutions are exploring existing general purpose virtual world platforms as a means to extend and enhance their offerings to students. Typically, educators create an online presence where students can interact, using their avatars to learn about new assignments or create projects that are viewable within the virtual world.

There are now many universities, colleges, schools and other educational institutions researching the use of Second Life as an environment for teaching and learning which offers a community of practice and situated constructivist learning. Among the institutions bringing the use of Second Life into the provision for distance learners is the Open University in the UK which already offers a range of teaching and learning provision on two islands CETLment and SchomeBase. Other Universities that are using these cutting-edge virtual classrooms are Princeton, MIT, University of Derby (UK), Vassar, Harvard, University of North Carolina at Chapel Hill, Ohio University, New York University, Australian Film Television and Radio School, etc. to name just a few.

Under the project *Knowledge organization, management and sharing in electronic learning environment* financed by the Croatian Ministry of Science, Education and Sports we have started testing Second Life platform as an extension to the current Faculty of Humanities and Social Sciences (FHSS), University of Zagreb, electronic educational environment with the goal to test the interoperability between the current VLE OMEGA (based on Moodle) and the SL platform. First phase was based on exploring and experimenting with the platform, as well as investigating and analyzing the best educational practices. The second phase was oriented towards finding a suitable place for our test virtual educational environment and creation of the Faculty's building replica (so that students can recognize it and fell attached).

The problem of finding a place that was not surrounded by the negative aspects of this environment (like gambling, commercial shops etc.) was solved by connecting with the community of educators and librarians which had build user friendly islands for connecting the individuals with the same ideas. Alliance Library System, a consortium of US librarians receiving sponsorship grant to create and maintain two islands Cybarary city I and II, and granted us a place for creating and experimenting with the SL environment (Picture 1).

Picture 1. Building of Faculty of Humanities and Social Sciences, University of Zagreb in Second Life

As we have created a building to perform educational activities we furthered our goal to promotion of the Croatian culture and heritage by creating some interactive objects with the basic information about our land, science and trivia (Picture 2.)

Picture 2. Example of heritage presentation in SL

Furthermore, in order to get a feedback we have gathered the statistical data of residents’ visits and interaction with the materials. Besides, we have created a small survey and asked them to fill it. The results showed that people were very much interested about heritage, and even asked to see more historical and cultural information about Republic of Croatia in this location. This has shown that

although a lot of information can be found on the internet, the new generation of virtual citizens prefer the visual and "tangible" approach to information. Therefore, our next steps will be working with students in creating materials that will promote, transform and adapt the information about the Croatian heritage for the virtual environment, which we hope will be a good combination of education challenges and heritage preservation.

In three months of monitoring the residents interaction with the objects placed in FHSS building we received a feedback from the users in two ways. One was the implemented script which alerted the teacher when the object was touched or used and the other one was the short survey where we wanted to find out more about the users interest. In general, the object that residents were interacting the most were: the replica of *Bašćanska ploča* with the information on the item as well as on the glagolitic alphabet (viewed around 50 times); objects with the information on the Croatian Nobel prize winners and the object with the information on Mohorovičić discontinuity (both viewed around 30 times); and the whiteboard with the several works from the artist Ivan Meštrović and his biography (viewed more than 20 times). Furthermore, the object offering external links to the FHSS services like its library catalogue, VLE, digital library etc. as well as the basic information on the Faculty was also one of the most viewed items.

The survey was filled by the 45 Residents and results of the showed that the majority of the Residents is still from the United States but closely followed by the Europeans who are getting more involved in the SL which has the potential of bringing various cultural influences and information. (Chart 1.)

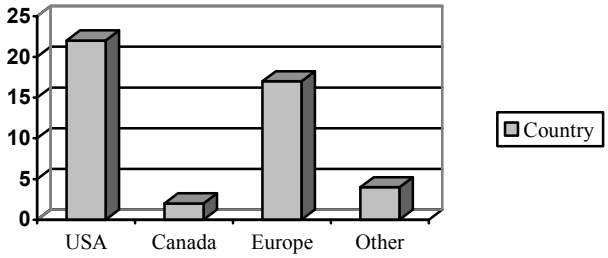


Chart 1. Origin of the users

The main goal of this survey was to find if the *Residents* liked the information found in FHSS building and found it useful, as well as to find out what would be the type of the information that they would prefer to find when and if they come back. (Chart 2.)

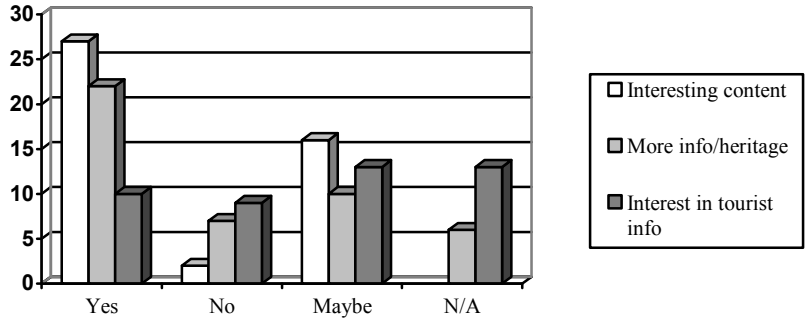


Chart 2. Users interest in materials to be offered in our SL building

The majority of the *Residents* found the content interesting and would like to see even more information on Croatian culture and art, history and language and literature.

Further steps in our project are now exploring the possibilities of creating a connection between SL and Moodle which is popularly called SLOODLE. The main problem that we have faced is that there are no connections between the various versions and upgrades of the Moodle platform but that the connection is possible only with the lower versions.

Educational and preservation of heritage

Preservation of heritage in SL is closely connected with the education and educational institutions. Student attending classes in SL are learning how to build objects, express themselves and even create programming scripts that trigger the objects to do certain actions (for example take you on a tour, teleport you, give you notecards…etc.).

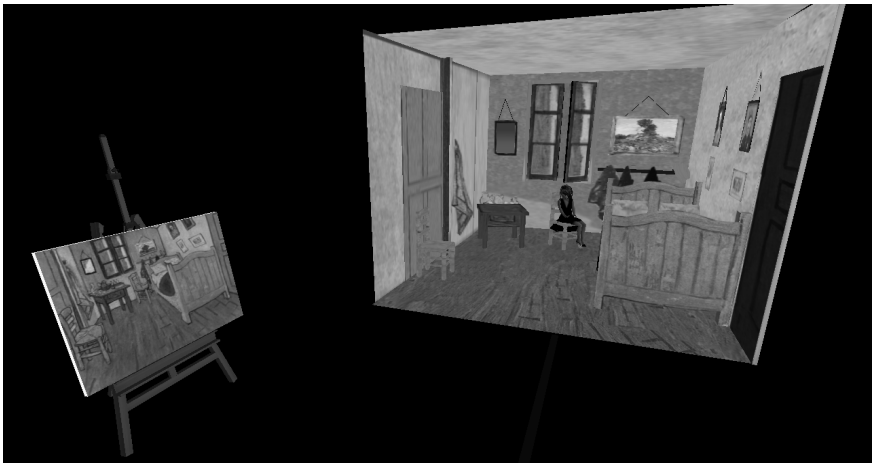
One of the very good examples of combining the educational process and creating a presentation of a local culture and heritage was the Global Outreach Morocco (G.O. Morocco) project which brought students with backgrounds in technology, business, and hospitality together to study economic development in Morocco through the growth of the travel and tourism industry. The G.O. Morocco team created a plan for using Second Life as a promotional tool for the country of Morocco. Over three months, the team created a technical prototype, rebuilding the symbolic Hassan II mosque and the traditional souk on the Campus island in Second Life. The location provides basic information on the country. Further plans are to develop the site in order to recreate the experience of visiting Morocco in Second Life, while promoting cultural education and knowledge.

There are numerous other projects where people have rebuilded either the existing locations (such as Paris and its Eiffel tour, Chichen Itza, London….etc), even in the different periods like the Renaissance island (Picture 3.) for people

to visit in Second life but also fantasy islands based on books like Hogwarts school of witchcraft (based on Harry Potter novels).

Picture 3. Globe theatre at the Renaissance Island

On the other hand artists are exploring the possibilities of presenting their work and creating artistic works in a new environment. Event the museums have started to create their replicas in SL like for example the Second Life Louvre, a virtual representation of the Louvre Museum. Furthermore, different approach to the experience of the artworks of famous artists can also be found on Van Gogh's Starry Night (Picture 4.)



Picture 4. One of the Van Gogh's artwork at the Starry Night Island

It is necessary to emphasize that the main difference is that apart from enjoying the 2D replica of the artists work resident can virtually go into the picture, like in the example shown above (Picture 4.) where we have entered the *ROOM*, explored it and even sat on the chair.

Conclusion

The advancement of ICT combined with the emergence and growing popularity of virtual worlds opened a numerous opportunities for educators and information professionals to utilize their skills and approach the new generation of students and patrons. Currently, Second Life is the most exciting virtual world platform available for extending the traditional real world and Web 1.0 learning and information environment. The possibilities are numerous but one should also be careful of the disadvantages and negative sides of these environments such as the digital divide and technological problems like crashes of the system or so called lags when the usage becomes problematic. We can say that we are still at the very beginnings and the necessary definition of advantages and disadvantages as well as the methodology of its usage in education still has to be done.

References

Freitas,S. Learning in Immersive worlds: A review of game-based learning. JISC, 2006. http://www.jisc.ac.uk/media/documents/programmes/elearning_innovation/gaming%20report_v3.3.pdf (05.09.2007.)

Keats, D. Smith, J. P. The genesis and emergence of Education 3.01 in higher education: the potential for Africa. // First Monday. http://www.firstmonday.org/issues/issue12_3/keats/index.html#author (12.9.2007.)

Kelly, R.V. Massively Multiplayer Online Role-Playing Games: The People, the Addiction and the Playing Experience. McFarland & Company, August 2004.

Proceedings of the Second Life Education Workshop at the Second Life Community Convention // editors Daniel Livingstone and Jeremy Kemp. San Francisco 20.08.2006. http://www.simteach.com/SLCC06/slcc2006-proceedings.pdf (05.09.2007.)

Urban, R. Marty, P. Twidale, M. A Second Life for your Museum: 3D Multi-User Virtual Environments and Museums. http://www.ideals.uiuc.edu/bitstream/2142/1619/2/MiSL_ill.pdf (05.09.2007.)

Virtual world, http://en.wikipedia.org/wiki/Virtual_world (12.9.2007.)

Visual Identity as a Form of Communication in MMORPG

Winton Afrić
Student of Postgraduate Doctoral Study of Information Sciences
Faculty of Humanities and Social Sciences
Križanićeva 5, Zagreb, Croatia
winton.afric@zg.htnet.hr

Summary

The Visual identity of an in game simulated world (virtual reality) is a factor defined and established by the visual identity of all elements of the environment. Each of the (cultural, social, biological) visual elements is part of a system defining all levels of experience within the virtual reality world. Communication forms are crucial in defining the specifics of any particular element witch in turn designs specific experience (for each of the agents immersed in the virtual reality). Specific experience is dependant on visual communication rules emerged through the construction of definite visual identity elements of an in game simulated world, and an understanding of particular game requests that both the visual elements and the visual identity structure as a whole must satisfy. The theme of Visual identity is very important for future development of convincing virtual reality in MMORPG and as a part of communication research for further development of Information science.

Key words: Role playing games, simulation, visual identity, virtual reality, information science

The Virtual Reality Role Playing Game world

RPG systems appeared in the early eighties in America as a new form of socialisation trough play, the reason being people often isolated themselves from social contact in front of computers, regularly missing out on classic forms of socialisation. The concept of RPG is in fact the creation of a virtual reality. The emphasis here lies on the term of virtuality that we are creating. Everything begins with a thorough analysis of societies, social structures, cultures and cultural forms, reality in its whole actually with the purpose of decomposing reality into its key elements and recomposing those elements into a new virtual reality. This virtual reality must with the basic means of recomposition trough which it was born receive a new structure which will determine new cultures, societies and all other aspect that make up a virtual reality. The system of playing RPG is

based on a set of rules by which every player takes on a role of certain individual from this, virtual reality created specifically for the game itself. The character that each player assumes (PC-Player character) is first and foremost determined by all the basic elements of our virtual reality. As we lay out the construction of our virtual reality setting up its basic components trough the use of naturalistic, cultural, and social forms we are in fact creating new cultures, new societies, and new social relations. Of course every RPG character as a part of the system of our virtual reality will be a direct product of some specific part of our virtual cultures and societies. He will be taught to live by the moral understanding and values of his virtual culture and his cultural identity will be clearly defined by his following of their way of life.

The founding elements of virtual reality do not involve only sociocultural aspects of reality but also all the natural elements of environment which define cultures and society. Virtual reality will determine its founding elements in its very structural basis and that basis must constitute from all the elements of reality. So virtual reality needs: a virtual atlas, (climate types, terrain types, meteorological aspects of the climate including catastrophic occurrences, virtual flora and fauna (new forms of life), virtual flow of time (our virtual planet doesn't have to have the same revolution cycle as earth neither it has to keep track of time same as we do), virtual history (determining holidays and heritage), virtual art (can be important to determine social gatherings or cultural elements) and many more elements. Each of those elements will play a key role as a visual identity element although it will certainly derive from a literary basis. Races that inhabit every virtual reality are virtual themselves. Accordingly they don't have to be humans, some of the best known RPG races are Elves and Dwarves. In every virtual reality world as a part of its founding structure, we will surely find, religion, technology, magic (if appropriate), basically cultural forms. Any created virtual culture is bound to worship divinities or forces, bound to have virtual myths and legends, saints, martyrs, heroes and villains, and magic will play as much of a part as it is given to. A simple but well organised form of a Virtual reality creation process is given in the "Dungeon Masters Guide" by the D&D Design team of Wizards of the coast in the Chapter 6: "World building" pg 153 - 164. This example does not give instructions on virtual reality building regarding its visual identity structure; it does however serve as a solid base upon which such a structure can be built.

In any Virtual reality Game world the first goal is to have any agent (player) feel at home. Meaning the level of experience the agent will receive trough their character being involved in the Virtual reality as a whole must be realistic and interesting enough to provide quality uniqueness. In any Game all the communication with the virtual world comes down to the amount of information the player receives trough experiencing the visual elements of that virtual reality. All agent experience is driven by the visual elements and the way they are managed. In a Role playing Game (RPG) the basic goal is to give the agent enough

means to role-play their character as a part of that virtual reality. In a MMORPG (Massive Multiplayer Role Playing Game) all of the agents must be able to do so, not only in communication with the NPC (Non Player Characters) but in communication with each other. The goal of the Virtual reality is not only to give a Visual identity as a feel of a virtual world but as a means of communication that agents can use to help them role-play their characters. Basic levels of communication are agent towards virtual reality and vice versa and agent towards agent. In both cases the Visual elements and the visual identity must be complex enough to enable agents to retain total immersion in the virtual reality with all the means they might find necessary to role-play their characters. So each Race, and culture, each society, creature and object must have their specific visual elements which define certain aspects of the virtual reality and serve not only as a means of communication towards the agent, but also as a tool for the agent to communicate with other agents.

Visual Communication in Games: Meaning and Understanding

Visual communication like any language communication revolves around meaning and understanding. If we are to successfully transmit any kind of information its meaning must be understandable to anyone receiving that information. Only under those conditions we may speak of communication simply because if any of those basic principles fails there will be no communication since the message transmitted wont go trough. Communication in art can be somewhat baffling regarding those two principles. We might stop to think of *L'art pour L'art* for a second. That term implies that the painting, which is in this case a visual communication transmitter between the artist and the public, is not transmitting any message we can read out of what we are simply viewing, which in term requires us to have some kind of background information to be able to fully understand the meaning. Since the meaning is not self implied and under those terms understandable that form of visual transmitter is not acceptable for usage in games. Unlike art visual communication in games must be made to satisfy these basic two principles. It must have a clear message or meaning and it must be understandable to whoever receives it. We cannot however expect to fully circumvent the need for background information but the meaning of almost every thing we can encounter in a virtual reality must have the tendency to be more or less fully self implied.

The background information factor can play a key role in understanding the true means of any visual information transmitter we might encounter. Since we are not talking about isolated elements but elements which comprise a whole of the virtual reality it is logical to assume that the meaning of one visual element can and will serve as background information on another element. The visual elements we are talking of will therefore form visual blocks and the block will ultimately shape the visual identity of our virtual world. Exact examples on the matter can be found in the article "Building a MMO with mass appeal", Sage

publications, by Nicolas Ducheneaut, Nick Yee, Eric Nickel and Robert J. Moore.

Visual Communication Blocks and Block relationships

In a more vivid example a visual communication block might give us information about a certain culture or race in our virtual reality. Let's take a race of Dwarves as an example. First and foremost to consider there is already existing background information on Dwarves and it would be folly to neglect the fact that for instance anyone who read even bits of Tolkiens "The Lord of the Rings" trilogy would have some information on the matter. If we want to stay at least somewhat true to the information we can say that we know that Dwarves in general will be described as stout little strong men with huge beards. They will be exceptional craftsmen and live underground where they are close to ore and gems which they favour in their craftsmanship. On that note some key visual elements must be the usage of metals and gems in designing elements of the dwarven culture. We might deicide at this point that the Dwarves of our virtual world would be separated into Clans, and that each clan would have their colour pallets (which would remain metallic and earthen to stay true to the background information) and their symbols. The symbols would all share similar features, for instance they might all be relief with depicted tools and weapons (hardly ever an animals) and that the dwarven alphabet would be hardly ever round and without fine line fluctuations, rather precisely carved geometrical shapes or lines. With this much background information and, decisions we can easily make restrictions on the character of our visual elements of the dwarven family. Now we have a good idea that we can recognize any dwarven crafted items we may come across in our virtual world and more so draw their lineage back to a certain dwarven clan. This would be a basic example of how a "Visual communication block" is formed. In our case we may call the block the "Dwarven block" and that block in comprised of many smaller blocks. We can divide those blocks in almost any way we choose. On one side we can have certain "Dwarven Clan specific blocks" which will have their own unique visual elements and on the other hand a block might be anything in the line of "Dwarven crafted weaponry" block. These blocks are all in relation because any dwarven marked block is a part of the basic "Dwarf communication block" and having the information from the whole Dwarven block would mean knowing that culture to its fullest in terms of visual experience. Having that information we can safely say that based on our knowledge we may now recognize anything that is not part of the Dwarf block. And thus we may form any kind of a Dwarven to non dwarven visual element relation. These elements can however be a part of the same visual block. Once we determine the basic shape of weaponry which is done trough background information of the type of weaponry used in the type of warfare that we might select for our virtual world we can easily know a weapon from a non weapon element. However since we have our Dwarven culture block

visual knowledge we can now tell apart a Dwarven from a non dwarven weapon.
The whole process of building visual elements and blocks is therefore in line with our basic principles of communication. In other words it is vital to acquire and value all background information, make decisions and set restrictions to further define a visual communication block. Define its meaning accordingly to benefit from the background information and make it understandable and as a part of a whole.

Visual character and characteristics

As we can notice in the process we have just described the decisions we made and our background information helped us to instantly choose defining characteristics of a certain visual block. All those characteristics like stone, metal and gems, master craftsmanship, clan symbols and lineage and so forth instantly provide the information to describe the visual character of that particular visual communication block. There are other aspects to take into consideration when fully defining a visual character of a certain block. In the designing process one often comes faced with requirements like: “That place must look sinister and scary”. In those times we need to step back and consider our background information. Does our information support the requirement or are we building into the information by complying. In any case Visual shape and form, colour and tone, contrast and many other aspects is what sets our emotions in a certain direction. Since we are building aspects of character into an existing block we have the basic Visual identity guidelines from our block. Now is the time to define our sinister block. The best way to approach the situation is to define that block using the same system we used earlier. Get all the background information on “sinister”, make decisions and set restrictions and once we have the guidelines for that particular block we can relate that block to our in example “Dwarven culture block” to make the sinister looking dwarven place.
Since we have just made an exact point in our virtual world using the visual elements and information from separate blocks we can now say that at that place in our virtual reality we have all the means of visual communication in place. As discussed in Bartle R (2004) “Designing virtual worlds”, those elements are with their character and meaning transmitting a certain desired information to the agent and the background information on which they are made, and the fact they belong to a certain family or a larger block makes them understandable as a particular element.

Visual communication as an advertisement

In reality our experience shapes our point of view and all the levels on which we can appraise a certain visual element and therefore conclude what that element means to us. Our experience and background information will play the key role in determining if we find something good or bad for us or any other kind of

posture we might have towards any kind of visual element we come across. In building a virtual reality of a Game one must take a different approach than trying to fully mimic the way things work in reality. It is best to think of an advertisement when trying to find a correct approach. At home, going to work, basically anywhere we are constantly bombarded by all kinds of advertisements trying to set our mind into thinking one way or another about a particular element. We look around and see beautiful people staring at us saying your hair will be like this if you buy a certain product, or your wrinkles will vanish if you try this product or any other kind of thing. Those things may or may not be true but all of those advertisements are giving us information trough communication and all of them are made along the lines of those basic communication rules which are most valuable to us. In a game environment our virtual reality will have severe limitations and it is in fact those very limitations which force us most of all to advertise trough our visual elements and therefore guide the player with the information given trough those advertisements. In example or sinister looking place might be sinister for the reason to make players think twice before going there and with a good reason. The reason might well be that their character will die there if they are not well prepared. In a way the sinister character of the place will serve as an advertisement for players to keep off until they reach a certain level or progress far enough in the game. We can further more use the advertisement method to hint players to look for certain information within our virtual reality. All of this implies that there are rarely coincidental aspects in the visual communication form in a virtual reality of a game. All visual elements must serve as guides and advertise one thing or another to give us information about meanings of elements and teach us world and game mechanics.

Visual advertisement as a plot guide

As we have said earlier the goal of an advertisement is to get us thinking a certain way about certain elements or aspects of elements. In a virtual reality of a game we have to use that method to its fullest to get the most out of getting any information trough to the player. Since all elements are parts of blocks and those blocks form a complex network in which all elements are corresponding we have to consider further the usage of the advertisement method in visual communication in our virtual reality. Often in movies we are led to believe certain things that might prove wrong by the end of the film. Other times we simply don’t have enough information to conclude certain things (like in crime movies – “who the murderer is” situations). All this is due to simply the way the movie is directed. The problem is in the quantity and quality of information we have available to us. If we had all the information we would know who the murderer is from the beginning and the movie wouldn’t be interesting. We can use this shortage of information as a valuable tool in our virtual game world. Therefore we can make misleading elements or simulate coincidences that can

baffle the player into thinking one way or another. It's safe to say that the gaming experience of a virtual reality is directed the same way movies are directed and although some information may exist as a logical part of a whole, it can be misleading or unravelled in a particular element. Those elements can be used as plot guides or mystery elements in our virtual reality but as every mystery it can only exist as a shortage of information on a matter that really is explained and logical. The mystery is enhanced by advertising that element in a certain way directing particular agent experience towards a certain end. Solving such a mystery is self rewarding and gives a great sense of accomplishment resulting in immersing the agent deeper into our virtual reality. The never ending problem is that once the mystery is revealed the agents without the experience on the matter can get all the information from agents that already solved that mystery. That is simply due to the fact that certain advertisement in a certain virtual reality world is always the same. Using these information shortage rules and our visual blocks network we can easily mimic and create mystery wherever and whenever we see fit. In a dynamic MMORPG setting where the virtual world is never finished but constantly built upon the possibilities are endless.

Visual Identity network style

Any visual identity network although built on the same principles can vary much from another visual identity network. The variation is not only due to the limitations a virtual reality as a game might have. The limiting programming and support issues are only a fraction of factors which determine exact Visual Identity style. Another factor which plays a key role in determining the exact character that our Visual network will have is the generic feel of the world we are set upon. In other words we may build a world aimed more at the younger population and make it more cartoony or make a very realistic virtual reality and set it towards another angle. The setting of our world is what makes the greatest impact. The setting is basically the global activity that goes on in our virtual reality and that setting is derived from story elements that comprise any game. A world besieged by war will definitely have a different feel than a peaceful world where all live in harmony. A futuristic setting like the "Bladerunner" will have great visual differences from a fantasy one like J. R. R. Tolkiens "Middle Earth" and so on. So defining the agent type and the setting of our virtual reality is a great place to start when looking for a style of a virtual reality Visual Identity. The style will set restrictions on definite visual elements and will determine the range of variation we are to include in our Visual identity network. All of those elements will then be directed to form agent experience which will ultimately be shaped partially by the Visual Identity style.

Multiple usages of Virtual reality Visual networks

As a step further from a MMORP Game perspective there are many ways we can implement a MMO Virtual Reality Visual Structure into various other

fields. It is important to emphasise that it can be used as a valuable tool in education for instance as a way of simulating various historical events, which can be used in classes. We can therefore "replay" a part of history modelled as accurately as it needs to be: for instance a visit to significant battlegrounds. Here we could use our visual network to emphasise the reasons why all events occurred the way they did but using all visual aspects to better familiarise the agent (student) with any important elements, which could not fit into a lecture because of the time limitations of the class program. We could also simulate future events that would help us determine the odds of any outcome and the scenario that would need to be built to achieve a certain outcome. Virtual meetings in virtual meeting places are another method of using Virtual reality Visual networks. In this situation the reality could be modelled so that the avatars of the agents could be a realistic representation of the agent's real personas. These are only some of the ways we could use a Virtual reality Visual network but all these examples clearly show both the communicational and the educational aspects that are a vital part of this method and we need to be aware of.

Conclusion

In designing a Virtual reality in terms of Visual Identity we must satisfy the basic rules of communication. In order to get the information trough the visual elements must have their meaning which has to be shaped in accordance to other visual elements forming visual blocks and ultimately a network. Those elements must all be understandable individually and as a part of a whole and the information given must be specific and engineered. All visual elements must be advertised in a certain way each fitting a particular element and this will enable us to direct advertising to form specific agent experience. All agent experience is driven by the nature of the visual elements in question and all of those must fit in with the background information, visual Identity style, and the restrictions of a particular virtual reality. Following these guidelines when engineering a virtual reality visual identity will enable us a very unique level of agent experience and an excellent level of directed communication with a focused information flow.

The believability of our Virtual reality is largely dependant upon its components which were made by decomposing reality and recomposing simulated virtual reality. This in itself ensures the reality of simulated worlds. By using an analytical, scientific approach in our Virtual reality building step we are not only supporting logic and meaning in all aspects of a game setting but ensuring the quality of information we are transmitting. Any given structure transmitted to the player in any way will therefore be educational not only in the sense of virtual reality knowledge but also in a general sense because the logic of creating any specific information is based directly as a product of specific real science and information. Building and playing a virtual reality is in its essence the same as testing your knowledge. Thus a Virtual reality of an RPG is impossible with-

out learning. From the agents perspective playing well means having the capability of self-evaluation and evaluation of ones own attributes in regard of the complexity of their surroundings as a whole. The education an agent receives by playing is therefore nonformal but very broad and at the same time very specific. All of those elements are crucial in designing a believable and playable game virtual reality. The theme of Visual identity is very important for future development of convincing virtual reality in MMORPG and as a part of communication research for further development of Information science.

References

Bartle, R. Designing virtual worlds, Indianapolis: New Riders Publishing, 2004

Cook, Monte; Tweet, Johnatan; Williams, Skip. Dungeon Masters Guide. USA: D&D Design Team, **,** October 2000.

Crogan, P. The Question of Computer Games // Games and Culture, Vol. 1; No. 1. January 2006. pp. 72 – 77, SAGE Publications

Ducheneaut, Nicolas; Yee, Nick; Nickel, Eric;.Moore, Robert J. Building a MMO with mass appeal. // *Games and Culture,* Vol. 1 (2006), No. 4; 281 - 317

Koster, R. A theory of fun for game design, Scottsdale: AZ-Paraglyph Press 2005

Moser, M.A.; MacLeod D.(ed). Immersed in Tehnology, Art and Virtual Environment. London: The MIT Press 1995

Turkle, S. Life on the screen: Identity in the age of the Internet, New York: Touchstone Books 1997

Williams, Dimitri. Why Game Studies Now? Gamers Don't Bowl Alone. // Games and Culture, Vol. 1 (2006), No. 1; 13-16

E-INFRASTRUCTURE

Biometrics Identification Process: A Multimodal Approach

Željka Požgaj
Faculty of Economics and Business
Trg. J. F. Kennedy-a 6, Zagreb, Croatia
zpozgaj@efzg.hr

Marko Miloš
Baruna Trenka 16, Zagreb, Croaita
xyzmarmi@yahoo.com

Summary

Rapid progress of biometric technology and its expanded application brings new possibilities in identification process. First biometric identification systems were organized as unimodal systems. It means that process of identification was based on single unique individuals' physical or behavioral traits. Although most biometric systems deployed in real-world applications are still unimodal, multimodal biometric systems represent an emerging trend that attempts to increase the level of security by using more than one biometric trait for identification or authentication.
The paper presents basic characteristics of biometric identification process, discusses various scenarios that are possible in multimodal biometric systems, and analyzes the levels of fusions in biometric systems, modes of operation, security requirements.

Key words: biometrics, identification process, multimodal biometric system

Introduction

The word *biometrics*, an acronym of Greek origin can be literally translated as "the measure of life" (Chirillo et al. 2003). It refers to identification techniques that rely on individuals' measurable physiological (anthropological) or behavioral traits that can be automatically checked. Physiological biometric traits are: fingerprint, hand and finger geometry, vessel pattern, iris and retina pattern, face geometry, facial pattern etc., while behavioral traits reflect individuals' behavior regarding performing of certain actions. They are: gait, keystroke, signature, voice etc.
The implementation of biometric traits in identification process reaches as far as the 2nd century BC The Chinese recognized that every person has their own,

unique bodily traits and started to use fingerprints in signing important documents (Požgaj, 2002.). As a form of identification, biometric traits have been applied up till today. The development of information and Internet technologies in the past two decades has had a great impact on the development and implementation of biometric identification systems. The development of information technologies has influenced the development of specific hardware and software especially adapted to biometric needs, the identification process itself by enabling digitalization of biometric sample, automation of sample comparison and with it the duration of the identification process (by shortening it). The Internet and its technologies are the basis of digital economy. There are new ways of running business (e-business), new ways of making business transactions (e-transactions), new ways of performing everyday activities. New forms of communication are being developed, awareness of value of certain resources like information is rising. Along with the positive changes brought by the new technologies, there are also some negative side effects. The security issue becomes ever more significant in business field as well as in other spheres of everyday life. New forms of protection are being explored that will provide additional security measures in comparison with the existing models. In that context multimodal biometric identification systems are being developed. The first multimodal systems appeared in 1998 (Bearman, 2006), but their full implementation is yet to follow.

From the point of view of security, multimodal systems represent a form of improvement for the identification process in a unimodal system. Unimodal systems are based on a single biometric (physiological or behavioral) trait used in identification and verification process, whereas in multimodal systems more biometric traits participate in an identification process. To be able to speak about specific qualities of multimodal systems it is necessary to become familiar with the qualities of unimodal systems since they represent a genetic pattern of biometric identification systems.

It is an undeniable fact that biometric identification is becoming a generally accepted form of identification. Issues most often arising from implementation of biometric identification systems refer to the choice of biometric system (unimodal or multimodal), choice of biometric trait (or traits) an identification system would be based on, implementation simplicity, implementation costs, system reliability, etc. However, the choice of a particular biometric technology for implementation largely depends on the type of application and the level of security required. This paper concentrates, as the title itself suggests, on multimodal approach to biometric identification process.

Following the introduction, the second chapter discusses basic characteristics of biometric identification/verification process connected to specific qualities of unimodal and multimodal systems. The third chapter deals with specific qualities of multimodal systems. The discussed qualities refer to scenarios of sample creation, operation modes, types of data fusion as well as certain difficulties in

an identification process. Implementation of biometric identification systems is dealt with in the fourth chapter. The fifth chapter is the conclusion.

Basic characteristics of biometric identification/verification process

Basic identification/verification model includes two processes: enrolment and verification (authentication) (Požgaj et al. 2007.).
Enrolment process covers the following steps:

- Taking the initial (identifying) sample
- Transforming the sample to template
- Storing the template (into a database or on a smart card).

In taking the first sample, the person is identified through classical method of identification (identity card) in order to confirm their identity. A biometric trait (initial sample) is taken according to selected methods and characteristics of the identification equipment, transformed to a digital template and stored in a database, on a local reading device or on a smart card.
Verification process can be repeated in every further attempt of identification. It covers:

- Identification process (taking a new biometric sample)
- Verification of the taken and stored template
- Approval or rejection for further actions.

Next time when someone wants to identify themselves they have to pass the identifying process again. The goal of verification is to confirm authentication. Process of verification is successfully completed if the template of the newly scanned part of a person's body corresponds with the stored template. Depending on the result of verification, further activities are either granted or denied.
The described model points at the generic activities typical for the biometric identification system. What separates unimodal from multimodal systems is the way in which the activities are performed. Differences in number of biometric traits from which an initial sample is created are obvious as early as the enrolment process (in unimodal systems one trait is present, in multimodal two or more). The procedure of creating a template itself also differs. In unimodal systems template is created via simple transformation of sample to template, whereas in multimodal systems a template is created following one of the possible scenarios typical of multimodal systems. The matching procedure is identical in both systems. The comparison process is based on sequential searching of the biometric sample database (1: n matching) or on comparison of the initial sample stored on a mobile carrier like smart card and the newly created one (1:1 matching). Verification is a process of confirming the authentication of the person being identified. It is successful if the comparison shows that the newly created sample and the stored sample match (1:1 matching).

Specific qualities of multimodal identification systems

The identification process in multimodal systems depends on multimodality. Since it is a system where identification is based on more than one biometric trait, there are certain scenarios for the final sample creation, fusion, results of matching and reaching a decision.

Possible scenarios for sample creation

As it was said earlier, identifying template in a multimodal system is formaed based on two or more single samples. The way in which a biometric sample is created is determined based on the chosen biometric technology. Regardless of the scenario according to which the data are collected in order to create a sample, the same scenario has to be applied when creating an initial template and the template created whenever identification is attempted. There are four possible scenarios (Ross et al., 2007):

- Single biometric trait, multiple sensors
- Single biometric trait, multiple classifiers
- Single biometric trait, multiple units
- Multiple biometric traits

The Single biometric trait, multiple sensor scenario means that the identification data are collected by scanning a single biometric trait with multiple sensors. A template is created by compressing the collected data. If the chosen biometric trait is hand, it is possible to create a template based on data collected by scanning hand geometry and data collected by scanning blood vessels of the hand (palmprint).

The Single biometric trait, multiple classifiers scenario means that the rough data of a single biometric trait are collected during single scanning. The data are then processed with multiple algorythms. If the biometric trait in question is a fingerprint, the desired template is created by applying minutiae-based and texture-based algorythms.

The Single biometric trait, multiple units scenario means that the biometric sample is created based on more than one biometric unit of the chosen biometric trait. If identification is based on a fingerprint, the final template can comprise fingerprints of more fingers (of all ten of them, of index and ring finger, of five fingers on the left hand, etc.).

The Multiple biometric traits scenario means that the final biometric template comprises of two or more samples of various biometric traits (fingerprint and face traits, fingerprint, face and voice traits, etc.).

Modes of operation

Besides the template creation process itself, in multimodal systems time sequence in which a person is identified is also significant. There are two types of procedures: synchronous and asynchronous. In a synchronous procedure a per-

son is identified simultaneously. Fingerprinting is carried out at the moment of identification by using more than one sensor. It is possible to scan a fingerprint when scanning hand geometry. When an asynchronous procedure is applied, a person is identified sequentially using two or more devices. A person can be identified upon entering the working premises by palmprint and voice trait. It is also possible to combine both synchronous and asynchronous procedures in an identification process.

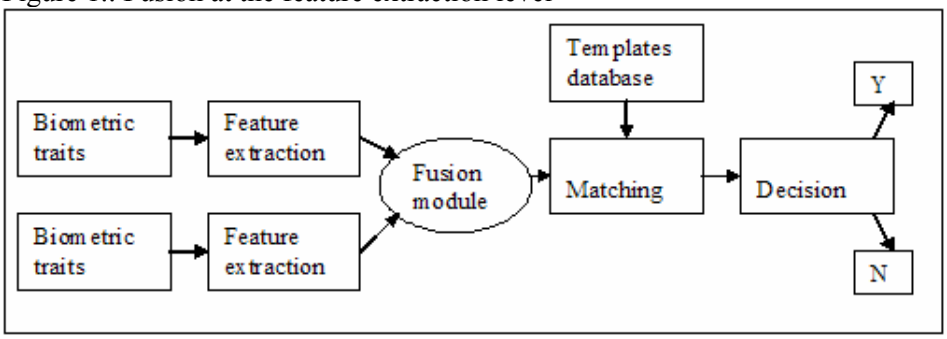
Multimodal system can operate in one of three different modes: serial mode, parallel mode or hierarchical mode (Ross et al., 2006.). In a serial mode the processing of biometric information takes place sequentially. If an identification system involves two biometric traits, a person is first identified based on the first trait (e.g. a fingerprint). In case the identity of a person is authenticated by processing the data collected according to the first trait, further identification is not necessary. If the identity is not authenticated, a person is identified using the following trait (e.g. face geometry). In this way time necessary for identification is reduced, which is extremely important when you have a situation where identification of a great number of users in the shortest time possible is required (for example upon coming to work). If a biometric system functions in a parallel mode, each ID subsystem processes its data independently at the same time and the processed data is combined using an appropriate fusion scheme. Most multimodal systems operate in a parallel mode as their primary goal is to reduce error rates of a biometric system despite reducing the recognition time. If an identification system operates in a hierarchical mode (tree-like architecture) it combines the advantages of both serial and parallel mode. Ross points out the cascade mode. It operates in serial mode but in such a way that the matching process based on the first biometric trait produces the top n matches whereas the second trait determines the identity of the user based on the n retrieved matches. It is possible to choose the order of biometric traits in which identification is carried out.

Data fusion

In multimodal systems data fusion occurs. Fusion can take place prior to or after the matching (Jain et. al., 2007.). Prior to matching fusion can occur either at sensor or feature level.

At sensor level, fusion is carried out at the rough data level. The data refer to the same biometric trait and are collected through one sensor or more compatible sensors. At feature level, fusion is preceded by creation of two (or more) independent feature templates (biometric vectors in Jain et al.). Fusion creates a single new template (vector). It represents a person's identity. Next follows the matching procedure in which a single new template is compared with templates stored in the database. The final decision depends on the results of matching.

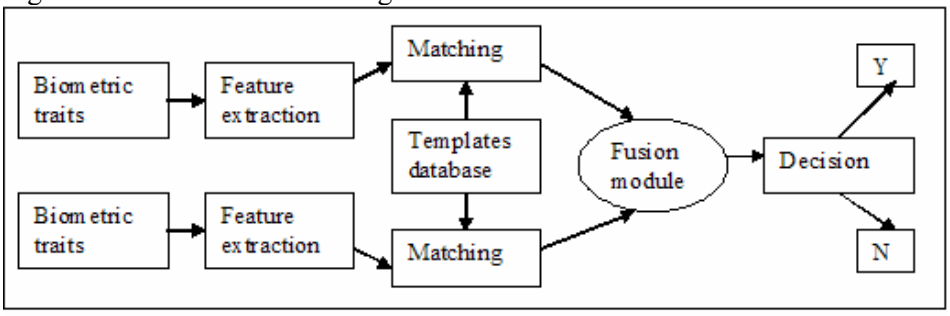
Figure 1.: Fusion at the feature extraction level



Source: Jain K. Anil, Ross Arun, Prabhakar Salil: An Introduction to Biometric Recognition. Modified by authors.

After the matching level, fusion can occur at match score level or decision level. During fusion at match score level (Fig. 2), for each biometric trait features are extracted and an input feature template is created. Matching is carried out in a way that every input feature template is compared with templates from the database. Matching scores are combined using the weighed averaging technique in order to authenticate the veracity of the claimed identity. According to the results a final decision is made.

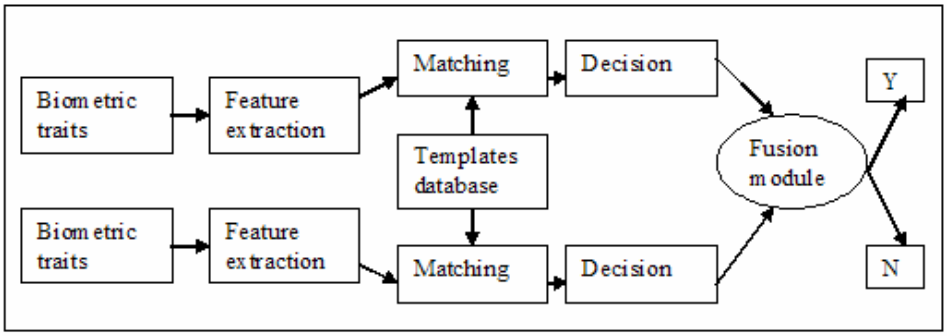
Figure 2.: Fusion at the matching score level



Source: Jain K. Anil, Ross Arun, Prabhakar Salil: An Introduction to Biometric Recognition. Modified by authors.

During fusion at decision level (Fig. 3), each segment of a multimodal system carries out its own identification process during which an input feature template is created and then compared with the templates from the database resulting in a recognition decision. At the end of the procedure single recognition decisions are combined and the identity of the person in question is either authenticated or not.

Figure 3.: Fusion at the decision level



Source: Jain K. Anil, Ross Arun, Prabhakar Salil: An Introduction to Biometric Recognition. Modified by authors.

Difficulties in identification process

Each identification system has certain limitations. In biometric systems these limitations most often refer to (Jain et.al., 2007), (Ross et al., 2006):

- Noise in sensed data
- Intra-class variation
- Inter-class similarities
- Non-universality
- Spoof attack

Noise in sensed data occurs when, for certain reasons, the chosen biometric trait cannot be used in identification. The reasons can be a dirty finger, husky voice, injured hand, but also the dirt on or damage of the sensors.

Intra-class variation and inter-class similarities are possible when the set rules of behavior are not followed in providing an identifying sample (position of finger, hand or face in regard with the sensor i.e. camera). Because of the finger movement or facial mimic it can happen that the created template does not match the real biometric traits of a person that is being identified. This can lead to false results of the verification process and the access can be enabled to an unauthorized person or denied to an authorized person.

Although it is stated that biometric identification methods are universal for everyone, non-universality occurs when certain people due to some physical handicap, illness, etc. cannot be identified through certain biometric traits. According to (Jain et al., 2007.) 4 per cent of the population may have poor quality fingerprint ridges and cannot be identified based on their fingerprints.

Spoof attack is present in all identification systems. In a biometric system the attacker tries to penetrate the system by imitating or forging biometric traits of the person he is posing as.

The described problems (except spoof attack which is possible in all systems) are especially noticeable in unimodal identification systems. The implementation of multimodal systems helps avoid or diminish these problems. So if a per-

son cannot be identified through a certain biometric trait (noise in sensed data or non-universality), other traits in the system can be used for identification (cascade mode identification).The same applies to the case when one of the collected samples cannot be used for comparison (intra-class variation and inter-class similarities). Each spoof attack based on imitation or forging of biometric sample has less chance of success in multimodal systems than in unimodal systems because the identification process is more complex and the system is more reliable.

Reliability is a very important issue in multimodal systems. There are three different metrics to rate accuracy of biometric technology and they are manifested as (Bearman, 2006.), (Požgaj et al., 2003):

- False match (imposter breaks in),
- False non-match (correct user locked out),
- Failure to enrol (user cannot register in system).

False Match Rate (FMR) or rate of false acceptance (FAR) represents the probability that particular user's verification template will be incorrectly judged to be a match for a different user's enrolment template. False Non-Match Rate (FNMR) or false rejection rate (FRR) represents the probability that a user's verification template will be incorrectly judged not to match that same user's enrolment template. Failure to Enrol (FTE) Rate represents the probability that a particular user will be unable to enrol in a biometric system due to insufficiently distinctive biometric sample(s).

These three metrics are strongly related. If FMR decreases or makes system less suitable to importers, legitimate users will be rejected (FNMR) and vice versa. The values of FMR or FAR and FNMR or FRR can form curves whose intersect is called Equal Error Rate (EER) or the crossover accuracy of the system. This is the rate at which the FAR is equal to the FRR. In general, the greater value of the crossover accuracy, the greater inherent accuracy of the biometric. According to the Biometric Consulting Group investigation the greater crossover accuracy is present in retinal scanning (1:10,000,000+), then iris scanning (1:131,000), fingerprints and hand geometry (1:500), signature and voice dynamics (1:50) (Požgaj, 2002.).

The reliability issue in multimodal systems is very complex and although in the end it is expressed with FTE, FAR and FRR, the reliability of the system is influenced by user's behaviour during identification process but also the traits of biometric technology represented through number of biometric traits used in identification, scenario of data collection, fusion and the decision-making process.

Implementation of biometric systems

The advent of biometric systems has meant a great leap forward from the point of view of security issues compared to physical (key, plastic card) and logical

forms of identification (password, PIN). Every day biometric systems are becoming more and more present. On a global level, a clear indicator that biometric technologies industry is becoming a respectable branch of information technologies industry are the predicted annual revenues for 2007 and 2010 stated in Annual Biometric Industry Revenues published by International Biometric Group. The predicted revenue for 2007 comes to around $3010.7 million and in 2010 it could go up to around $5749.2 million (Annual Biometric Industry Revenues, 2006).

There are three areas of implementation for biometric technologies (Jain et al., 2007.):

- Commercial applications
- Government applications
- Forensic applications

Forensic applications deal with a rather specific area of biometric identification and will not be discussed here.

In commercial applications mostly unimodal identification systems are used today. Reasons for that lie in a wide offer of unimodal identification devices based on various biometric traits, acceptable implementation costs, simplicity of use, users' good response to them … The main areas of implementation are:

- Physical access control which could ensure entrance to locations previously intended only for authorized persons (company premises, hotel rooms, schools, kindergartens, prisons …)
- Money transactions where identity control is needed (ATM, credit card payment …)
- Restriction in business activities where identity control is needed (access to particular business resources, access to confidential data …)
- Particular activities that become more secure if biometric identification is implemented (network transactions, logging on to computer, logging on to mobile …).

Comparing unimodal and multimodal systems one can come to conclusion that in multimodal systems the biometric identification technology is more expensive, the procedure of system implementation more complex, implementation forms more demanding, standardization and interoperability problems more clearly noticeable. Because of that, it is necessary to perform a detailed needs analysis when choosing a biometric technology and set certain priorities.

The basic criterion in government applications is increased security. That is why multimodal identification systems are recommended (and worldwide used) when it comes to applications that refer to national ID card, passport control, driver's licence, social security etc.

In Croatia, according to the information of the author, multimodal systems are still not used, but unimodal systems are ever more present. Unimodal systems are mostly used in physical access control with fingerprint as an identification trait.

Conclusion

The aim of this paper was to introduce biometric identification in general and multimodal approach in particular. Since the multimodal approach is in a way an upgrade to the unimodal approach, the paper is mostly based on comparison of the two. Specific qualities of the multimodal approach are represented via ways of collecting biometric samples and forming templates, modes of operation, fusion of relevant data within sensor module, feature module, matching module or decision module. By implementing multimodal systems it is possible to overcome the limitations of unimodal systems that in a way lessen the advantages of biometric identification systems. Multimodal systems are for now mainly used in government applications although it is to be expected that multimodal systems will find its way to commercial applications as well. Reasons for that are many and can be seen primarily in the need of higher security measures when performing business and everyday activities.

References

Bearman Nathan: Multimodal biometrics, http://www.securitysa.com (20.11.2006.)

Annual Biometric Industry Revenues, Biometric statistics in focus, http://www.sciencedirect.com (10.12.2006.)

Chirillo John, Scott Blaul: Implementing Biometric Security, Indianopilis, Willey, 2003.

Jain K. Anil, Ross Arun, Prabhakar Salil: An Introduction to Biometric Recognition, http://www.citer.wvu.edu (20.6.2007.).

Požgaj, Ž.: Biometrics and New Technology, Proceedings of 1st International Conference “An Enterprise Odyssey: Economics and Business in the New Millennium”, Graduate School of Economics & Business, Zagreb, June 27-29, 2002, 975-987

Požgaj Željka, Mateljan Vladimir: Some Sugestion about Biometrics' Technology Selection, Proceedings of the14th International Conference on Information and Intelligent System IIS 2003, Varaždin, September 24-26, 2003. 313-323

Požgaj Željka, Đuretek Ivor: Smart Card in Biometric Authentication, Proceedings of the18th International Conference on Information and Intelligent System IIS 2007, Varaždin, September 12-14, 2007. 319-325

Ross A. Arun, Nandakumar Karthik, Jain K. Anil: Handbook of Multibiometrics, NewYork, Springer, 2006

Ross Arun, Jain K. Anil: Multimodal biometrics: An overview, http://biometrics.cse.msu.edu (20.6.2007)

Značaj skriptnih jezika za administraciju operacijskih sustava

Vladimir Mateljan
Odsjek za informacijske znanosti, Filozofski fakultet
Ivana Lučića 3, Zagreb, Hrvatska
vladimir.mateljan@zg.t-com.hr

Željka Požgaj
Ekonomski fakultet
Trg. J. F. Kennedyja 6, Zagreb, Hrvatska
zpozgaj@efzg.hr

Krunoslav Peter
Filozofski fakultet
Ivana Lučića 3, Zagreb, Hrvatska
kruno_peter@yahoo.com

Sažetak

Skriptni programski jezici, poput jezika `bash` *i* `awk`*, omogućuju rapidni razvoj programskih rješenja koja se koriste za automatizaciju poslova administracije operacijskih sustava, uz minimun uporabljenog programskog koda. Ovaj članak objašnjava osnovni razlog za pragmatičnost skriptnih jezika – njihov dizajn primjeren je za izradu miminalističkih programskih rješenja u formi naloga za naredbeni redak, aliasa i skripti na način da se ostvaruje integracija gotovih komponenti – naredbi i programskih rješenja, iz razvojnog okruženja – ljuske operacijskog sustava. Članak identificira neke dodatne beneficije koje proizlaze iz primjene skriptnih jezika u poslovima administracije operacijskih sustava.*

Ključne riječi: skriptni programski jezik, administracija, operacijski sustav, redak za upis naredbi

1. Uvod

Suvremeni trend u administriranju operacijskih sustava jest da se poslovi administracije operacijskog sustava izvode uz primjenu programskih pomagala iz slikovnog sučelja operacijskog sustava, primjeri kojih su Webmin za Unix i GNU/Linux operacijske sustave ili Microsoft Management Console (MMC) za Microsoft Windows operacijske sustave. Druga mogućnost svodi se na obavljanje poslova administracije u okruženju ljuske operacijskog sustava na način da

se, uz primjenu *skriptnih programskih jezika*, kreiraju i izvršavaju *nalozi* u retku za upis naredbi, *aliasi* i *skripte*. U članku će primjeri takvih programskih rješenja biti prikazani na primjeru skriptnih programskih jezika koji potječu iz okruženja operacijskog sustava Unix:

1. jezik ljuske `bash`.
2. `awk`, programski jezik za pretraživanje i obradu teksta,

Koncepti iz programskog jezika ljuske `bash` implementirani su u ljuskama operacijskih sustava Windows i Mac OS X. `awk` je također je ostvaren za spomenute operacijske sustave.

Drugo poglavlje članka posvećeno je definiranju termina "*skriptni jezik*". Skriptni programski jezici predstavljaju posebnu vrstu programskih jezika; prema svojoj svrsi, osobinama i pragmatici, razlikuju se od konvencionalnih programskih jezika, no s obje vrste jezika, u današnje vrijeme, mogu se ostvariti programska rješenja iste funkcionalnosti. Treće poglavlje donosi sažeti opis dva predstavnika skriptnih jezika, `bash` i `awk`. U opisima spomenutih programskih jezika nastoje se naglasiti karakteristike njihovog dizajna, koje pogoduju izradi programskih rješenja u minimalnom vremenu i uz minimalnu količinu koda.

U promišljanju značaja skriptnih programskih jezika za administraciju operacijskih sustava, polazimo od pregleda proizvoda rudimentarnog programiranja u okruženju ljuske (engl. shell) operacijskog sustava: nalog u naredbenom retku, alias i skripta. Važno obilježje ovih programskih rješenja jest jednostavnost, jer ih u praksi izrađuju administratori i korisnici operacijskih sustava kojima programiranje nije prioritetni posao.

O pragmatici i značaju skriptnih programskih jezika za administraciju operacijskih sustava progovara se u četvrtom i petom poglavlju članka. Primjenom skriptnih jezika za ostvarivanje programskih rješenja možemo automatizirati poslove administracije operacijskih sustava.

Izvorni kod naloga, aliasa i skripti pisan je vrstom slova `Courier`. Nalozi za upis u redak za naredbe pisani su tako da im je početak postavljen znak "**`$`**" koji se ne upisuje, jer predstavlja znak spremnosti (engl. prompt). Neki primjeri u `bash`-u lako se mogu adaptirati za izvršavanje u okruženju ljuske operacijskog sustava Windows (Torres, 2004, 2), npr. naredba `ls` iz Unixa može se zamijeniti naredbom `dir` Windowsu.

2. Skriptni programski jezici

2.1 Svrha i namjena skriptnih programskih jezika

Skriptni programski jezici su interpretirani[1] programski jezici (Ousterhout, 1997), koji su namijenjeni za:

- izradu aplikacija od gotovih komponenti,

[1] Potrebno je napomenuti da se interpretirani programski jezici mogu ostvariti kao prevodioci. Tako je za programski jezik `awk` ostvaren prevodilac s nazivom `tawk`.

- upravljanje aplikacijama koje imaju programabilno sučelje,
- pisanje programa kojih je brzina razvoja važnija od učinkovitosti u izvršavanju (Barron, 2000, 4).

U skriptne programske jezike se ubrajaju jezici ljuski, awk, perl, Python, PHP, Tcl, JavaScript, etc. Većina skriptnih jezika se temelji na paradigmi strukturnog programiranja (bash, awk, etc.); Python je primjer objektno orijentiranog skriptnog programskog jezika.

Skriptni programski jezici obično implementiraju se kao interpretatori (Barron, 2000, 6), izvodeći naredbu iz skripte odmah nakon čitanja, premda se mogu ostvariti kao prevodioci ili prevodioci u trenutku izvršavanja (engl. just in time compiler).

Postoje razne vrste skriptnih jezika. Mogu biti implemetirani samostalno ili u okruženju editora, tabličnih kalkulatora, tekst procesora, etc. U kontekstu ovog članka, pažnja će biti usmjerena na skriptne jezike za kontrolu poslova u operacijskom sustavu (engl. job control languages), u koje se ubraja `bash`, te na skriptne jezike za procesiranje teksta, iz kojih je izdvojen `awk`.

2.2 Bash i awk – predstavnici skriptnih programskih jezika

Cjeloviti opisi jezika `awk` i `bash` mogu biti preobimni čak i za jednu knjigu. U navođenju sintakse naloga oba programska jezika, opcionalni dijelovi naloga pisat će se unutar uglatih zagrada ("`[`" i "`]`"); tri točke ("`...`") označavat će ponavaljanje naloga, koji im prethodi. Naglasak u sažetom opisu `bash`-a bit će na sintaksi naredbi, ulančavanja i preusmjeravanja. U svezi s `awk`-om će se opisati: (1) sintaksa naloga za poziv `awk`-skripte u retku za opis naredbi, (2) sintaksa naloga `awk`-skripte, sastavljena od uzorka i procedure te njegova semantika.

O pragmatici ovih jezika, kao predstavnika skriptnih jezika, raspravlja se jednim dijelom u ovom poglavlju i sljedećim poglavljima članka, uz nastojanje da se naglase svojstva njihovog dizajna koja podržavaju rapidni razvoj programskih rješenja za administraciju sustava uz minimum koda.

Osnovna vrsta podataka nad kojom `bash` i `awk` izvode operacije jest *znakovni niz* (engl. string), odnosno *tekst*. Takav dizajn skriptnih jezika omogućuje implementaciju ulančavanja i preusmjeravanja *tekstovnog tijeka* (engl. text stream) što ćemo promotriti u opisima `bash`-a i `awk`-a.

2.3 Programski jezik ljuske bash

`Bash` je "sustavski interpretator naredbi" (Peek et al, 1998, 82); posjeduje "napredne programske sposobnosti bilo kojeg interpretatora svoje vrste" (Newham, 2005, 81). Naredbe se, u okruženju `bash`-ljuske operacijskog sustava Unix, tipkaju u retku za upis naredbi ili se nalaze u skripti; pojedinačna naredba ima jednostavnu sintaksu (Peek et al, 1998, 9):

```
naredba [prekidač...] [datoteka...]
```

U retku za upis naredbi, naredbe se mogu nizati tako da se prethodna razvoji od sljedeće znakom “**;**” (točka-zarez, engl. semicolon) (Kernighan & Pike, 1984, 71). Na taj način ostvaruje se *nalog*. Sintaksa nizanja naredbi:

```
naredba [; naredba...]
```

Ulančavanje je niz od dvije ili više naredbi povezanih znakom “**|**”:

```
naredba [| naredba...]
```

Semantika sintakse za ulančavanje je jednostavna: podaci sa standardnog izlaza prve naredbe predaju se na standardni ulaz prve naredbe. U okruženju ljuske je uobičajeno da naredbe čitaju podatke sa standardnog ulaza – tipkovnice, a izlazne rezultate prikazuju na standardnom izlazu – zaslonu (Peek et al, 1998, 67); ulančavanjem se izlazni rezultat jedne naredbe može predati drugoj u tekstovnom tijeku.
Sintaksa za ulančavanje naredbi je jednostavna i efikasna – uz minimum koda izražava se poziv druge naredbe i predaju joj se ulazni podaci.
Iz opisa sintakse `bash`-a će biti izdvojena i pojednostavljena sintaksa preusmjeravanja, još jednog koncepta, koji pridonosi rapidnoj izradi programskih rješenja uz minimum koda – izlazni podaci iz naredbe `naredba` pohranjuju se u datoteci `datoteka`:

```
naredba > datoteka
```

Dodavanje sadržaja na kraj postojeće datoteke ostvaruje se primjenom simbola “**>>**”. Ovakva sintaksa, premda jednostavna, ima slojevitu semantiku: upućuje na implemetaciju algoritma za otvaranje datoteke, pohranu znakovnog niza primljenog sa standardnog ulaza i zatvaranje datoteke. Program u konvencionalnom programskom jeziku, kao što npr. C ili Pascal, s ovakvom funkcionalnošću će imati više (desetaka) redaka.
Na primjeru sintakse i semantike ulančavanja i preusmjeravanja dokazuje se tvrdnja da dizajn programskih jezika utječe na njegovu primjenjivost u rapidnom razvoju programskih rješenja i to uz ostvarivanje principa minimuma napisanog koda i maksimalne ekspresivnosti.
U Primjeru 1 prikazana je skripta u kojoj je primjenjeno ulančavanje i preusmjeravanje. Naredba `echo` prikazuje specificirani niz znakova na standardnom izlazu – zaslonu. Znakom “**#**” ispred retka označavaju se napomene u skripti.

Primjer 1. Popis datoteka s nizom znakova "txt" u nazivu

```
#!/bin/bash
# popis datoteka s nizom znakova "txt" u nazivu
echo "TEKSTOVI" > doc.txt
ls | grep txt >> doc.txt
```

Prvi nalog u skripti (`#!/bin/bash`) daje informaciju o interpretatoru skripte, programu koji izvodi skriptu, a to je ljuska `bash`.
Dodatakom naredbi za grananje i petlju te mogućnosti za rad s varijablama sustava (Newham, 2005, 86), programski jezik ljuske `bash` postaje u potpunosti osposobljen izradu programskih rješenja u okviru paradigme strukturnog programiranja. O pragmatici `bash`-a se detaljnije progovara u ostallim poglavljima članka.

2.4 Programski jezik awk

`Awk` je programski jezik za "pretraživanje i obradu teksta" (Aho et al, 1978). Jezik je nastao 1977. godine u Bell Laboratories. Tvorci `awk`-a su Alfred V. Aho, Brian W. Kernighan i Peter J. Weinberger; naziv jezika je kratica za prezimena autora jezika.
Osnovna svrha `awk`-a jest da pretražuje retke ulazne tekst datoteke koji sadrže određene uzorke; nakon što pronađe redak s odgovarajućim uzorkom, `awk` izvodi specificirane radnje (procedure) nad tim retkom (Robbins, 2001, 3). Uzorak može tvoriti proizvoljna Booleova kombinacija regularnih izraza (engl. regular expression) i relacijskih operatora na znakovnim nizovima (engl. strings), brojevima, poljima sloga, varijablama, i elementima asocijativnog jednodimenzionalnog polja (engl. array) (Aho et al, 1978). Procedure mogu također uključivati izraze za prepoznavanje uzoraka, aritmetičke operacije, operacije nad znakovnim nizovima, pohranu i manipulaciju podacima u asocijativnom polju, grananje, `while`- i `for`-petlju te mnostruke izlazne tijekove (Aho et al, 1978).
U Primjeru 2 prikazan je `awk` nalog za upis u naredbenom retku koji prikazuje redni broj ispred naziva datoteke. `Print` je naredba awk-a koja na standardni zaslon ispisuje sadržaj varijable ili niz znakova.

Primjer 2. Awk nalog – prikazuje redni broj ispred naziva datoteke

```
$ ls | awk '{ print NR, $0 }'
```

Skripta se u `awk`-u može, kao što je to vidljivo u Primjeru 2, zadati u retku za upis naredbi.
Sintaksa za poziv `awk`-a sa skriptom zadanom u naredbenom retku je (Robbins, 2000, 24):

```
awk [opcija] 'skripta' ulazna_datoteka
```

Skripta iz datoteke se poziva primjenom prekidača `-f`:

```
awk [opcija] -f skripta ulazna_datoteka
```

Zadavanje skripte `awk`-interpretatoru u retku za upis naredbi predstavlja mogućnost vrijednu naglaska. Na taj način mogu se u vrlo kratkom vremenu, možda čak i u sekundama, izraditi funkcionalna programska rješenja za obradu tekstovnih datoteka (ili tekstovnog tijeka (engl. text stream)). To je razlog da se `awk` ponekad naziva "alatom", premda je on programski jezik.
Sintaksu je `awk` preuzeo od programskog jezika C (Kernighan & Richie, 1988), ali je ona za `awk` u velikoj mjeri pojednostavljena. Nalog se u C-u obavezno treba završiti znakom "`;`" (Kernighan & Richie, 1988), a u `awk`-u je taj znak opcionalan. To pridonosi lakoći i brzini pisanja izvornog koda u `awk`-u.
U odnosu na C ili Pascal, kao predstavnike konvencionalnih programskih jezika, `awk` poznaje samo jednu vrstu podataka – *znakovni niz* (engl. string). Interno se obavlja konverzija stringa u broj, ili konverzija broja u string, kada to kontekst zahtijeva (Aho et al, 1978). U nastavku članka ćemo analizirati beneficije od unifikacije različitih vrsta podataka u znakovni niz.
Osobitost `awk`-a također su ugrađene (engl. built-in) varijable (Aho et al, 1978). Primjer takve varijable je `NR` u Primjeru 2, u koju `awk` automatski pohranjuje broj trenutno učitanog sloga. Ovo je osobina koja pridonosi primjerenosti `awk`-a za rapidni razvoj, uz minimum napisanog koda, jer programera oslobađa potrebe za pisanjem naredbi koda za brojanje učitanih slogova ili drugih pokazatelja.
`Awk` skripta je sastavljena od *naloga* ili *pravila* čija je sintaksa `uzorak { procedura }` (Robbins, 2000, 25). `awk` program je, dakle, niz uzoraka i akcija, koje određuju, što potražiti u ulaznim podacima i što napraviti s pronađenim podacima (Aho et al, 1988). Prema tome, struktura `awk` programa izgleda ovako (Robbins, 2001, 3):

```
uzorak { procedura }
uzorak { procedura }
...
```

Ako je u nalogu izostavljen uzorak, procedura se primjenjuje na sve učitane slogove; ako nedostaje procedura, ispisuje se slog u kojem je pronađen uzorak (Robbins, 2000, 26). Posebni uzorak, `BEGIN`, služi za izvršavanje procedure prije učitavanja prvog zapisa iz ulazne datoteke, koja služi za inicijalizaciju programa (npr. globalnih varijabli). Uzorak `END` izvršava proceduru nakon zadnjeg učitanog sloga zadanih ulaznih datoteka (Robbins, 2000, 27).

Primjer 3. Awk nalog u retku za upis naredbi, koji broji retke u kojima je pronađen uzorak "Linux" ili "linux" i ispisuje broj (frekvenciju) pojavljivanja

```
$ awk '/[Ll]inux/ { i++ } END { print i }' file.txt
```

U Primjeru 3 je prikazana primjena uzorka s regularnim izrazom `[Ll]inux` i uzorka `END` koji se izvršava nakon što se učitaju svi zapisi ulazne datoteke `file.txt`.

U programski jezik `awk`, kao i u druge alate u okruženju ljuske operacijskog sustava Unix, ugrađen je jezik za *regularne izraze*, koji `awk`-u daje izuzetne sposobnosti za pretraživanje teksta i oslobađa programera potrebe da te mogućnosti kodira: znakovni nizovi se mogu pretraživati uz specificiranje uzorka, umjesto prema fiksnim podnizovima znakovnog niza (Robbins, 2000, 2).

Do posebnog izražaja `awk` dolazi u obradi tekst datoteka s podacima, točnije rečeno, sekvencijalnih datoteka koje imaju slogove i polja unutar slogova (Robbins, 2000, 23). `awk` parsira takve ulazne datoteke i automatski dijeli svaki ulazni zapis u polja (Aho et al, 1988). "Budući da su mnoge stvari automatske – ulaz (tj. čitanje ulaznih datoteka), dijeljenje na polja, upravljanje memorijom, inicijalizacija – `awk` programi su obično manji od onih u konvencionalnim jezicima" (Aho et al, 1988). To `awk` čini kvalitetnim skriptnim jezikom i svrstava ga u red jezika, koji su primjereni za rapidni razvoj softvera. On omogućuje sustavskim administratorima da na jednostavan način ostvaruju programska rješenja, i to uz minimun koda (slobodno se može primjetiti da se `awk`-u ne pišu naredbe za otvaranje i zatvaranje ulazne ili izlazne datoteke, niti se kodira algoritam za parsiranje podataka ulazne datoteke).

`Awk` obrađuje ulazne retke, dok ne dosegne kraj ulaznih datoteka (Robbins, 2001, 3) – u programskom jeziku `awk` implementiran je algoritam koji:

- otvara ulaznu datoteku,
- parsira znakove ulazne datoteke (do trenutka nailaska na znak za kraj datoteke),
- zatvara ulaznu datoteku.

Programski jezik `awk` je specijaliziran za pretraživanje i obradu teksta, za razliku od jezika ljuske `bash`, koji je namijenjen za programska rješenja, koja automatiziraju sustavsku administraciju. `awk` se smatra programskim pomagalom za opću namjenu te može nadomjestiti ostale specijalizirane alate i programe (npr. `grep`, `cut`, `tail`, etc.). Često se primjenjuje u je izradi izvještaja (engl. report), transformaciju podataka, ili čak za generiranje izvornog koda (engl. code generator) u drugim programskim jezicima.

3. Nalog, alias, skripta

Nalog uključuje jednu ili više naredbi operacijskog sustava. U Primjeru 4 naveden je nalog koji prekida izvršavanje svakog procesa u čijem je nazivu niz znakova "sleep"; pretpostavka je da naredba `ps` daje podatak o ID-u procesa u prvom, a nazivu procesa u četvrtom stupcu.

Primjer 4. Nalog za prekid izvršavanja procesa koji u nazivu ima niz znakova "sleep"

```
$ ps | awk '$4 ~ /sleep/ { cmd = "kill -9 " $1; print | cmd }'
```

U prikazanom primjeru ostvareno je ulančavanje naredbe `ps` (prikazuje trenutno aktivne procese) i `awk`-skripte koja generira nalog s naredbom `kill` (prekida izvršavanje procesa).
Ukoliko u poslovima administriranja često ponavljamo određeni nalog, možemo ga pohraniti u obliku aliasa.

Primjer 5. Alias za broj datoteka u imeniku

```
$ alias lsx='ls | wc -l'
```

U Primjeru 5 ulančavanjem su povezane naredba `ls` (pokazuje sadržaj imenika) i `wc -l` (broji redove, riječi i znakove).
Niz naredbi možemo pohraniti u u tekst datoteku – *skriptu* (engl. script). Jednostavna skripta je dana u Primjeru 1.
Naziv "skripta" potječe iz ranih 70-ih godina (Barron, 2000, 4); tvorci operacijskog sustava Unix taj su izraz upotrebljavali za "niz naredbi koje se čitaju iz datoteke i redom izvode kao da su zadavane tipkanjem." Korisnici i administratori operacijskog sustava koriste skripte za automatiziranje repetetivnih poslova. Značenje termina "skripta" je kasnije proširено na "niz instrukcija ostalih jezika npr. `awk` skripta, Perl skripta, etc." (Barron, 2000, 4), odnosno na "tekst datoteku, koja je namijenjena izravnom izvršavanju, umjesto prevođenju" (Barron, 2000, 4).

4. Pragmatika skripnih jezika

Naredbe u okruženju ljuske operacijskog sustava Unix čitaju podatke sa standardnog ulaza – tipkovnice, a izlazne rezultate prikazuju na standardnom izlazu – zaslonu (Peek et al, 1998, 67); te komponente poznaju samo jednu vrstu podataka – tekst, odnosno znakovni niz. Tehnikom ulančavanja u ljusci operacijskog sustava Unix – "povezivanje dvije naredbe skupa tako da izlazni podaci iz prve budu ulazni u drugu" (Peek et al, 1998) – moguće je ostvarivati jednostavna funkcionalna programska rješenja – naloge – uz minimum napisanog koda.

Primjer 6 pokazuje ulančavanje naredbi u ljusci operacijskog sustava Unix da bi se ostvario nalog za prikaz broja datoteka koje u nazivu imaju traženi niz znakova:

Primjer 6. Sastavljena naredba koja pokazuje broj datoteka, koje u nazivu imaju niz znakova "txt"

```
$ ls | grep "txt" | wc -l
```

Način povezivanja gotovih komponenti u Primjeru 6 napravljen je primjenom jednostavne sintakse, kojom se na jednostavan način izražava pozivanje komponente i proslijeđivanje podataka u pozvanu komponentu – primjenom znaka "|". Upravo je to jedna od osobina bitnih za skriptne jezike, na kojoj se temelje njihove integracijske sposobnosti – *jednostavnost sintakse* za pozivanje gotovih objekata i proslijeđivanje podataka od jednog prema drugom objektu. Nadalje, da bi se podaci mogli proslijeđivati iz jedne komponente (objekta) u drugu, trebaju imati *standardizirani format podataka* za sve komponente. U okruženju ljuske operacijskog sustava Unix, podaci se proslijeđuju u obliku *znakovnih nizova* – na taj način je ostvarena simplifikacija prijenosa podataka i njihova unifikacija.
Skriptni jezik `awk` također ima mogućnosti izdvojiti dio podataka i predati ih, kao znakovni niz, drugim komponentama operacijskog sustava (Aho et al, 1988, 64) – Primjer 7 (unutar `awk` skripte ulančani su naredba `print` i nareba `sort` iz okruženja ljuske):

Primjer 7. Awk skripta, u retku za upis naredbi, koja izdvaja 1. stupac popisa korisnika sustava i poreda podatke po redoslijedu

```
$ who | awk '{ print $1 | "sort" }' > users.txt
```

Dakle, `awk` nema implementiran algoritam za sortiranje redaka teksta ali njegova sintaksa daje podršku za integraciju s vanjskom komponentom sustava s takvom funkcionalnosti.
Znak "|" semantički ukazuje na proslijeđivanje, a ">" na preusmjeravanje; takva sintaksa je jednostavna za pisanje koda i slikovito ukazuje na značenje.
Na primjerima u skriptnim jezicima `awk` i `bash`, pokazane su integracijske sposobnosti koje posjeduju i ostali skriptni jezici. Mogućnost *integriranja gotovih komponenti* u nekom okruženju presudno je značajna komponenta za rapidni razvoj programskih rješenja.

5. Značaj skripti i skripnih jezika za administraciju operacijskog sustava

Posao administratora operacijskog sustava uključuje ponavljajuće radnje s objektima operacijskog sustava npr. prijepise datoteka iz raznih imenika u jedan određeni imenik, generiranje izvještaja iz logova o događajima u sustavu (engl. event logs) (Stanek, 2003, 3), etc. U svakodnevnom ponavljanju takvih radnji – bilo primjenom miša ili sličnog uređaja u redu s pomagallima u slikovnom sučelju, bilo vlastoručnim upisom i izvršavanjem naredbi, prema određenom redoslijedu – mogu se dogoditi pogreške u izvođenju poslova administracije. Prema Principu automatizacije preporuča se “automatizirati mehaničke i dosadne poslove u kojima se može pogriješiti” (MacLennan, 1999, 10). Rješenje za ovaj problem je izrada i primjena programskih rješenja u skriptnim jezicima; u njih će se pohraniti naredbe i programski konstrukti, koji će izvoditi korake poslova administriranja.

Dakle, aliasi i skripte su, gdjegod su primjenjivi, sredstvo za *automatizaciju* (Stanek, 2003, 3) jednog dijela poslova sustavskog administratora. Automatizacijom repetitivnih poslova, administrator stvara uštedu vremena; oslobođeno vrijeme može posvetiti edukaciji ili kreativnim poslovima, npr. programiranje skripti ili konfiguriranje operacijskog sustava.

Popratna beneficija kreiranja aliasa i skripata je *dokumentiranje* koraka nekog posla u administraciji operacijskog sustava. Nadalje, izvršavanje naloga, aliasa i skripata u retku za upis naredbi ostavlja podatak o izvršenju istih u sustavskim logovima (primjerice *history* u Unix okruženju).

Administratori operacijskog sustava mogu imati znanje o programiranju, ali programiranje nije njihov prvenstveni posao. Njihova prioritetna zaduženja su instalacija i konfiguriranje operacijskih sustava i softvera, otvaranje korisničkih računa, pohrana i povrat podataka, etc. (Collings, 2005, 3). Budući da za te poslove trebaju kontinuirano akumulirati znanja, programski jezici, za primjenu u administraciji sustava, trebali bi imati jednostavnu sintaksu i semantiku, te pragmatički biti primjereni za jednostavnu primjenu i rapidnu izradu programskih rješenja, uz minimum napisanog koda. Skriptni jezici, poput `awk`-a i `bash`-a, zadovoljavaju ovaj zahtjev.

Ne samo da se skriptama može automatizirati administracija sustava (Stanek, 2003, 3), nego se uz pomoć skriptnih jezika mogu generirati nove skripte i na taj način izbjeći mukotrpno tipkanje naredbi u skriptu. Posljednji primjer u članku, Primjer 8, ostvaren u programskom jeziku `awk`, pokazuje jednostavnu skriptu koja će od popisa datoteka napraviti naredbe za njihovo kopiranje u datoteke s nastavkom `.old` u imeniku `/arch`:

Primjer 8. Skripta za generiranje izvornog koda skripte za kopiranje datoteka u zadani imenik

```
#!/bin/bash
ls | awk '{ print "cp " $0 " /arch/" $0 ".old" }' > cp2arch.sh
```

Za svaku naziv datoteke koji prikaže naredba `ls`, u izlaznoj će se datoteci generirati naredba npr. `cp dat1 /arch/dat1.old`.

6. Zaključak

Skriptni programski jezici omogućuju izradu minimalističkih programskih rješenja uz pomoć kojih možemo ostvariti automatizaciju poslova administracije operacijskog sustava. Popratne beneficije u ostvarivanju takvih programskih rješenja su dokumentiranje koraka posla u administraciji te zapisi u sustavskim logovima prilikom njihovog poziva na izvršavanje. Skriptni programski jezici omogućuju rapidni razvoj programskih rješenja za automatizaciju administracije operacijskih sustava, uz minimun uporabljenog programskog koda, jer njihov dizajn podržava razvoj programskih rješenja od gotovih komponenti u razvojnom okruženju, a to su naredbe operacijskog sustava. Dizajn takvih jezika ima simplificiranu sintaksu i semantiku te unificiranu vrstu podataka – znakovni niz. Takve osobine čine ih prihvatljivim za primjenu od strane sustavskih administratora.

Literatura

Aho, Alfred V., Kernighan, Brian W., Weinberger, Peter J. Awk – A Pattern Scanning and Processing Language. Bell Laboratories, 1978.

Aho, Alfred V., Kernighan, Brian W., Weinberger, Peter J. The AWK Programming Language. Addison-Wesley, 1988.

Barron, David. The World of Scripting Languages. Willey, 2000.

Collings, Terry, Wall, Kurt. Red Hat Linux Networking and System Administration. Wiley, 2005.

Kernighan, Brian W., Pike Rob. The UNIX Programming Environment. Prentice Hall, 1984.

Kernighan, Brian W., Ritchie, Dennis M. The C Programming Language, 2nd Edition. Prentice Hall, 1988.

MacLennan, Bruce J. Principles of Prog.ramming Languages: Design, Evaluation, and Implementation. Oxford University Press, 1999.

Newham, Cameron. Learning the bash Shell, 3rd Edition. O'Reilly, 2005.

Ousterhout, John K. Scripting: Higher Level Programming for the 21st Century. Sun Microsystems Laboratories, 1997.

Peek, Jerry, Tondingo, Grace, Strang, John. Learning the UNIX Operating System, 4th Edition. O'Reilly, 1998.

Robbins, Arnold. sed & awk Pocket Reference. O'Reilly, 2000.

Robbins, Arnold. Effective awk Programming, 3rd Edition. O'Reilly, 2001.

Robbins, Arnold, Beebe, Nelson H. F. Classic Shell Scripting. O'Reilly, 2005.

Robbins, Arnold, Lamb, Linda. Learning the vi Editor. O'Reilly, 1998.

Slonneger Kenneth, Kurtz Barry L. Formal Syntax and Semantics of Programming Languages. Addison-Wesley, 1995.

Stanek, William R. Microsoft Windows Server 2003 Administrator’s Pocket Consultant. Microsoft Press, 2003.

Torres, Jesse M. Windows Admin Scripting Little Black Book, 2nd Edition. Paraglyph Press, 2004.

INFORMATION SCIENCE EDUCATION

The Status of Information and Communication Technology in National Curriculum for Compulsory Education

Marina Štibrić
Institute for Social Research in Zagreb
Centre for Educational Research and Development
Amruševa 11, Zagreb, Croatia
marina@idi.hr

Branislava Baranović
Institute for Social Research in Zagreb
Centre for Educational Research and Development
Amruševa 11, Zagreb, Croatia
baranov@idi.hr

Summary

Changes brought by the knowledge society have pointed out the importance of information and communication technology (ICT) and made it one of the key contemporary educational requirements. Introduction of ICT into compulsory education was unavoidable, but at the same time it was one of the most complex curricular innovations. Although there is an agreement about the necessity of including and developing ICT in the curriculum, different countries have found different solutions, depending primarily on guidelines for the development of their educational systems. In order to find out the current status of ICT in compulsory education, a comparative analysis of national curricula in sixteen countries was conducted. While in some national curricula ICT is taught as a separate subject, in other curricula it is part of other subjects or is a cross-curricular area. Although the acquisition of knowledge and skills is fundamental, more and more countries are recognizing the advantage of using ICT in teaching other subjects. There is a notable approach to curriculum based on educational outcomes or competences all students should posses by the end of each program or stage of their schooling. By planning to bring the focus on outcomes and to introduce "Information and communication technology" as a cross-curricular area, Croatian curriculum is coming closer to the world standards in education.

Key words: information and communication technology (ICT), national curriculum, compulsory education, students' competence, comparative analysis

Introduction

For quite some time there is a discussion about changes brought to human society by the innovations in information and communication technology (ICT), networking and rapid expansion of the Internet. These changes are nowadays considered to be the key factors of transformation into knowledge society, enabling participation in global economy. ICT has become extremely important as the most effective modern technology. Besides, today ICT is a ground for effective usage and exchange of information and knowledge. It enables people to have constant and free access to information anywhere in the world. Creating networks enables information exchange, therefore improving their development. Those facts made ICT one of the key contemporary educational requirements. Living in the information society based on knowledge calls for people to possess a new sort of competence in order to function successfully in that society. Generations that are currently receiving their education and those yet to come should get prepared for everyday interaction with ICT. This was recognized by the European Parliament and the Council of the European Union (2006) who included digital competence in key competencies which each person needs to posses so he or she can adapt to the rapidly changing world.

Teaching how to find, choose, assess, store, use, but also how to produce, present and exchange information is considered to be one of the key goals of education. This implies constant use of ICT, therefore throughout education students have to learn not only its possibilities, but also the restrictions and implications included. Furthermore, modern technology is developing rapidly and new generations must learn how to constantly adjust to changes. Therefore, the European Parliament and the Council of the European Union (2006) in their definition of digital competence, along with knowledge and skills, include a critical attitude toward a responsible use of ICT.

The growing usage of ICT in schools is happening in the context of, and following wider social changes. Introduction of ICT into compulsory education was unavoidable, but at the same time it was one of the most complex curricular innovations. The role of ICT in curriculum changed with each new demand. Today ICT is a significant part of transforming the teaching and learning process, and for that to happen it is important how the curriculum is perceived and organized. In order to understand the possibilities and limitations of the changes which education is going through, we must comprehend the concept of curriculum (Yek & Penney, 2006). Although there is an agreement about the necessity of introducing and developing ICT in curriculum, different countries have found different solutions, depending primarily on guidelines for the development of their educational systems. The present study reviews the status of the ICT in different national curricula for compulsory education, including the current situation in Croatia, and gives some recommendations about conceptualizing ICT in national curriculum.

The present research

Research methodology

An analysis of national curricula in sixteen countries aimed to find out what is the current status of ICT in compulsory education. Considering the Croatia's efforts to integrate into the European Union, which also requires synchronization in area of education, most of the countries in the study were old and new members of the EU. An important factor in choosing the countries also was their educational heritage and achievements of their students in international comparative evaluations. Countries included in the sample were: England, Scotland, Ireland, Sweden, Norway, Finland, The Netherlands, Austria, Hungary, Slovenia, Canada (Alberta), United States of America (New York), Australia (New South Wales), New Zealand, Singapore and Croatia. In Canada, USA and Australia we analyzed national curricula of only one of their states or territories, so the descriptions do not apply to the whole country. Ways of conceptualizing ICT in each country of the sample were compared to each other, and with the current situation in Croatia. The status of ICT isolated as a separate subject, ICT taught as a part of a few different subjects and ICT as a cross-curricular area were all taken into consideration.

Research results

While analyzing national curricula, some similarities were spotted in defining the role of ICT between the countries, so they were divided into several groups based on the observed similarities.

Finland, Sweden and Norway

Finland, Sweden and Norway have broad curriculum frameworks, with ICT being taught within mandatory subjects. In Finland it is a part of a subject called "Crafts" (Finnish National Board of Education, 2004), in Norway "Art and Crafts" (Norwegian Directorate for Education and Training, 1997), and in Sweden it is a part of two subjects, "Crafts" and "Technology" (Swedish National Agency for Education, 2000; 2006). Timetable is not defined for any of these subjects. While describing the subject, accent is set on defining skills that students must obtain in using and understanding of ICT. Finnish national curriculum also defines main goals of cross-curricular themes called "Media skills and communication" and "Technology and the individual". It is specific about Norway that there is no subject or cross-curricular area that would deal with ICT.

Scotland, Ireland, England, Canada (Alberta), Australia (New South Wales) and New Zealand

The other group of countries consists of Scotland, Ireland, England, Canada (Alberta), Australia (New South Wales), and New Zealand countries that have both national curriculum frameworks and a subject curriculum. In Scotland (Scottish Executive Education Department, UK, 2000) ICT is primarily being

taught through the subject of the same name, while in Ireland (Department of Education and Science, Government of Ireland, 1999) and England (Department for Children, Schools and Families, England, UK, 2000) it is a part of larger cross-curricular areas. The content of each subject and area is mandatory, but timetable for their teaching is not strictly defined. During first four grades in Scotland "Information and communication technology" is a part of an area called "Environmental studies", for which 15% of the timetable is scheduled, and it becomes a separate subject in the 5th grade. While defining the goals of the subject "Information-communication technology" in Scotland, special accent is put on its role in developing problem solving skills.

Besides the cross-curricular areas "Social, environmental and scientific education" and "Art", ICT in Ireland is being taught as part of the subjects "Technology" and "Technical drawing", in which most of the content is related to the use of graphical communication. Irish curriculum is distinguished from others because special attention is given to ICT as a tool for teaching "Science" and "Arts", with a goal of developing creativity and thinking.

In England, ICT content was arranged in a form of two interdisciplinary areas, "Information and communication technology" and "Design and technology". Knowledge, skills and understanding required from students on all three levels are listed along with links to other subjects and curricular areas.

All provinces in Canada agree that ICT is an integral part of education, and that it also has a key role in preparing students for participating in information society. Alberta has a centralized curriculum, which defines what students must learn, for all grades and all subjects (Alberta Education, Alberta, Canada, n. d.). In Alberta ICT curriculum defines general and specific outcomes, but also illustrative examples and assessment framework for each of the four divisions. The "Information and Communication Technology" curriculum provides students with a broad perspective on the nature of technology, ways to use and apply a variety of technologies, and the impact of ICT on themselves and on society. ICT program of studies is structured as a 'curriculum within a curriculum', and it is infused within core subjects of "English Language", "Arts", "Math", "Science" and "Social Studies".

ICT has an important role in curriculum frameworks of all Australian states and territorial units, most frequently as a cross-curricular area (Australian Education Systems Officials Committee, 2006). K-10 Curriculum Framework of Australian federal state New South Wales contains widely defined learning results which briefly describe knowledge, understanding, skills, values and attitudes all students must acquire throughout and upon completion of their schooling (Board of studies, New South Wales, 2002). Plans and programs clearly describe the standards which show what is expected from students to know and be able to do in every stage from kindergarten to 10th grade. ICT statements listed for each subject are organized by stage, then by outcome, and finally by content or other syllabus requirements. Although ICT is a part of all key learning areas,

the central area in primary school in which ICT knowledge and skills are being taught is "Science and technology". In stages 4 and 5 ICT content is a part of a key learning area called "Technological and applied studies", i.e. mandatory subject "Technology" which has a range of 200 lesson-hours. The area is also connected to elective subjects "Information and software technology", "Design and technology" and "Graphics technology".

The New Zealand Curriculum comprises a set of national curriculum statements which define the learning principles and achievement aims and objectives which all New Zealand schools are required to follow (Ministry of education, New Zealand, 1995). The framework sets out the knowledge, understanding, skills, and attitudes to be developed by all students. In the New Zealand Curriculum ICT is a part of essential learning area called "Technology" which is compulsory only until the end of year 10. Achievement aims are expressed through eight progressive levels of schooling. This area of learning has application to all subjects of the curriculum, but some make a particular contribution to learning in technology, such as "Science", "Mathematics", "Home economics", "Social studies", "Workshop technology", "Music", "Art", "Graphics and design", and "The study of information systems" (Ministry of education, New Zealand, 1995).

Austria, Hungary, Slovenia and Singapore

In a group of countries which consisted of Austria, Hungary, Slovenia and Singapore, ICT is separated into subject curricula. These subjects are mandatory, and their timetable is between one and two hours per week. ICT content is represented in the Austrian curriculum through subjects "Technical education" and "Technical education, textile education" (Federal Ministry for Education, the Arts and Culture, Austria, 2003). "Technical education" is created in a way that it contributes to each of the five educational areas: "Man and society", "Nature and technology", "Language and communication", "Creativity and shaping", and "Health and exercise".

In Hungary subjects "Technology and the way of living" and "Information and communications technology" are being taught, and they have a defined list of activities, content and skills that students have to possess in order to continue to follow that subject in higher grades (Ministry of education, Hungary, 2000).

In Slovenia ICT content is a part of mandatory subjects "Perceiving environment" (in grades 1 to 3), "Science and technology" (4^{th} and 5^{th} grade) and "Technics and technology" in 6^{th}, 7^{th} and 8^{th} grade (Ministrstvo za šolstvo, znanost in šport Republike Slovenije, n. d.). Apart from that, ICT is taught through non-mandatory subjects "Robotics and technics" in the eight grade, and "Electronics with robotics" and "Electro technology" in the ninth grade. There is also a cross-curricular area called "Information literacy" represented in different grades. For all the mandatory and non-mandatory subjects and the cross-

curricular area general and operational aims are defined in order to determine ways of acquiring conceptual and procedural knowledge.
Unlike other countries in this sample, Singaporean education system has only six grades of mandatory education, although almost all children continue their education in secondary schools that have differentiated curriculum according to streams (Ministry of education, Singapore, n. d.). Despite this difference, Singapore stands out with a highly-developed and successful economy, which depends heavily on exports of information technology products. Current education system can also be regarded as highly successful, especially if indicators of participation rate are considered, literacy rate and mean years of schooling, but also with a lot of funds spent on educational upgrade (Yek & Penney, 2006). Besides that, Singaporean students have regularly ranked top when competing in international science and mathematics assessments (Gonzales et al., 2004). In primary education one of the aims of the mandatory subject "Mathematics" is using ICT as a tool for studying and application of mathematics. Based on their results on the national Primary school leaving examination at the end of Primary 6, students are placed in different secondary education tracks or streams: Special, Express, Normal (Academic), or Normal (Technical). In Normal (Technical) students take subjects more technical in nature, such as "Computer Applications", "Technical studies" and "Design and technology". "Design and technology" is a compulsory subject in all streams, but in the Normal (Technical) it includes more ICT content than the curriculum for other secondary education streams. At the end of Special, Express and Normal (Academic) stream students take a Singapore-Cambridge General certificate of education Ordinary level (O-level) exam which can include elective subjects such as "Design and technology" and "Computing" that include ICT content. Based on education systems in other societies, such as those in the USA, Singaporean education has put a bigger emphasis on creative and critical thinking and on learning for lifelong skills.

The Netherlands, and USA
Curriculum approach found in the Netherlands, and USA (New York) is different from all the other countries in the sample because it provides for flexibility, and a high degree of decentralization. An important aspect of education system in these countries is that, although the government sets policies and expectations about outcomes and student achievement, including assessment and reporting on outcomes, schools have autonomy in conceptualizing and enforcing the national curriculum. Each school formulates its own curriculum – within and beyond the requirements of national regulation – at the local level. They plan their teaching and learning programs to support the needs of their particular students. Their curriculum is focused on defining learning standards or student achievement.

ICT related content in the Dutch curriculum is taught through learning areas "Technology" and "Drawing, art and crafts" (Dutch Eurydice Unit, Ministry of education, culture and science, the Netherlands, 2006). There are no specifications neither about school levels on which themes are supposed to be taught, about the amount of time it requires, nor if these programs are mandatory or not. Schools in the Netherlands are responsible for all the curriculum innovations, including using ICT in teaching, and the government just stimulates and encourages these innovations.

The state of New York (USA) has a curriculum based on the national learning standards (New York State Education Department, 1996). Standards have two primary dimensions, content and performance, specified for every subject area and each grade level. Relevant area for teaching ICT is "Mathematics, science and technology" which includes Standard 2 called "Information systems" and Standard 5 called "Technology" with a "Computer technology" as its integral part. Through the "Information systems" standard student will learn how to access, generate, process, and transfer information using appropriate technologies.

Situation in Croatia

Croatia has a tradition of a centralized prescriptive subject based curriculum consisted of teaching program for each subject. In 2005 Croatia began preparing its educational system for education that will be coherent with requirements of the knowledge society, and which will contribute to its development. In this way, one of the priority-areas in "The plan for development of the education system in Croatia 2005-2010" (Plan razvoja sustava odgoja i obrazovanja 2005.-2010.; Ministarstvo znanosti, obrazovanja i športa Republike Hrvatske, 2005) was the application of ICT.

Currently, the compulsory school students in Croatia have a chance to obtain ICT related skills and knowledge only if they choose an elective subject called "Information science" between the 5th and the 8th grade, or if extra-curricular activity is enabled in the area of "Information science" between the 1st and the 4th grade (Ministarstvo znanosti, obrazovanja i športa Republike Hrvatske, 2006). The only ICT content mandatory for all students in compulsory school is being taught as a part of a subject called "Technical culture", which is mandatory between the 5th and the 8th grade. In the curriculum for compulsory school (Nastavni plan i program za osnovnu školu; Ministarstvo znanosti, obrazovanja i športa Republike Hrvatske, 2006) it is recommended that while teaching the subject "Information science", teachers should encourage the use of ICT in other subjects. Apart from that, ICT teaching program is separately outlined in a form of a cross-curricular area, with knowledge and skills that students have to acquire during their schooling.

Conclusion

This comparative analysis has indicated some significant conceptual, structural and content differences between the status of ICT in the Croatian curriculum and in national curricula of analyzed countries. While most of the countries have national curriculum framework which serves as a base for developing detailed school curriculum, Croatia still has fragmented subject based national curriculum. Although there are some efforts, Croatian curriculum still isn't oriented towards learning outcomes as most of the analyzed countries. Curricula of some countries like New Zealand, Australia (New South Wales), Canada (Alberta) and England, define competencies that every student has to acquire to enable him or her for work and life. Using ICT through the curriculum is an example of the way to acquire digital competence. Even though approach to planning and curriculum control differs among states, ICT is recognized as an important segment of education in all of the countries.

While in some national curricula ICT is being taught as a separate subject, in other curricula it is a part of a few different subjects or a cross-curricular area. Although the acquisition of knowledge and skills is fundamental, more and more countries are recognizing the advantage of using ICT in teaching other subjects. By planning to bring the focus on outcomes and to introduce "Information and communication technology" as a cross-curricular area, Croatian curriculum is coming closer to the world standards in education.

A general trend to determine the status of ICT which could be drawn from the national curricula of 15 countries can be a useful starting point for the development of national curriculum for compulsory education in Croatia.

Recommendations

Development of information society based on knowledge, technological innovations and new forms of communication is reflecting onto educational system. One could expect that it will be necessary to have highly developed competence in ICT area in order to successfully function in contemporary society. Therefore it is recommended that ICT is included in national curricula as a compulsory cross-curricular area, and it can also be retained as a separate subject. In the constantly changing environment there is a need to possess knowledge and skills, but also attitudes concerning ICT, so every curriculum should emphasize the digital competence (The European Parliament and the Council of the European Union, 2006).

In order for every student to have a chance in obtaining and developing digital competence, it is necessary to elaborate the status of ICT in curriculum of every system, to ensure adequate infrastructure and Internet access in every school, to provide network and multimedia materials, and to develop teachers' skills in using ICT. Teachers ought to be given support in getting those skills in order to fully exercise potential that ICT has in making learning easier. This is confirmed by "The plan for development of the education system in Croatia 2005-

2010" (Plan razvoja sustava odgoja i obrazovanja 2005.-2010.; Ministarstvo znanosti, obrazovanja i športa Republike Hrvatske, 2005) in which using ICT in teaching is a priority area for initial education and permanent training of teaching staff.

ICT is becoming vital part of educational system as a support for class realization and for teaching and learning process. If adequate infrastructure and content are ensured, ICT can make learning faster and easier (Yek & Penney, 2006). Therefore investing in the ICT is a key stake in preparing students to actively take part in knowledge society.

References

Alberta Education, Alberta, Canada. Alberta's High Quality Curriculum: Development and Implementation. (n. d.). http://www.education.gov.ab.ca/k_12/curriculum/ (13.8.2007.)

Australian Education Systems Officials Committee. The Statements of Learning for Information and communication technologies (ICT). Carlton South: Curriculum corporation, 2006

Board of Studies, New South Wales. K-10 Curriculum Framework. Sydney: Board of studies, New South Wales, 2002

Department for Children, Schools and Families, England, UK. National Curriculum for England. London: Department for Children, Schools and Families, England, UK, 2000

Department of Education and Science, Government of Ireland. Primary School Curriculum. Dublin: Department of Education and Science, Government of Ireland, 1999

Dutch Eurydice Unit, Ministry of education, culture and science, Netherlands. Education system in the Netherlands. The Hague: Dutch Eurydice Unit, Ministry of education, culture and science, Netherlands, 2006

Federal Ministry for Education, the Arts and Culture, Austria. Lehrplan der Hauptschule. Vienna: Federal Ministry for Education, the Arts and Culture, Austria, 2003

Finnish National Board of Education. National Core Curriculum for Basic Education. Helsinki: Finnish National Board of Education, 2004

Gonzales, P.; Guzmán, J. C.; Partelow, L.; Pahlke, E.; Jocelyn, L.; Kastberg, D.; Williams, T. Highlights From the Trends in International Mathematics and Science Study (TIMSS) 2003(NCES 2005–005). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office, 2004

Ministarstvo znanosti, obrazovanja i športa Republike Hrvatske. Nastavni plan i program za osnovnu školu. Zagreb: Ministarstvo znanosti, obrazovanja i športa Republike Hrvatske, 2006

Ministarstvo znanosti, obrazovanja i športa Republike Hrvatske. Plan razvoja sustava odgoja i obrazovanja 2005.-2010. Zagreb: Ministarstvo znanosti, obrazovanja i športa Republike Hrvatske, 2005

Ministrstvo za šolstvo, znanost in šport Republike Slovenije. Osnovnošolsko izobraževanje: Devetletna osnovna šola. (n. d.) http://www.mss.gov.si/si/delovna_podrocja/osnovnosolsko_izobrazevanje/devetletna_os/ (16.8.2007.)

Ministry of education, Hungary. Framework curricula for primary education. Budapest: Ministry of education, Hungary, 2000

Ministry of education, New Zealand. New Zealand Curriculum Framework. Wellington: Ministry of education, New Zealand, 1995

Ministry of education, Singapore. Education System. (n. d.). http://www.moe.gov.sg/corporate/eduoverview/Primary.htm (17.8.2007.)

New York State Education Department. Learning standards for New York state. New York: New York State Education Department, 1996

Norwegian Directorate for Education and Training. Curriculum for the 10-year compulsory school. Oslo: Norwegian Directorate for Education and Training, 1997

Scottish Executive Education Department, Scotland, UK. The 5-14 National Curriculum guidelines for Scotland. Edinburgh: Scottish Executive Education Department, Scotland, UK, 2000

Swedish National Agency for Education. Curriculum for the Compulsory School System, the Pre-School Class and the Leisure-time Centre. Stockholm: Swedish National Agency for Education, 2006

Swedish National Agency for Education. Syllabuses for the compulsory school. Stockholm: Swedish National Agency for Education, 2001

The European Parliament and the Council of the European Union. Recommendation of the European Parliament an the Council of 18 December 2006 on key competencies for lifelong learning (2006/962/EC), Brussels: The European Parliament and the Council of the European Union, 2006

Yek, T. M.; Penney, D. Curriculum as praxis: ensuring quality technical education in Singapore for the 21th century. // *Education Policy Analysis Archives*, 14(2006), 26; 20.10.2006. http://epaa.asu.edu/epaa/v14n26/ (19.8.2007.)

Simulation Models in Education

Hrvoje Stančić
Department of Information Sciences
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
hrvoje.stancic@zg.t-com.hr

Sanja Seljan
Department of Information Sciences
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
sanja.seljan@ffzg.hr

Ana Cetinić
Department of Information Sciences
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
anacetinic@gmail.com

Dijana Sanković
Department of Information Sciences
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
dijana.sankovic@gmail.com

Summary

This paper introduces the use of simulation models in an e-learning environment. Nowadays, simulation models are a part of computer-assisted learning and thus an important guide for lifelong education. E-learning systems, combined with the use of simulation models as tool for interactivity, are the best way to provide some kind of virtual reality in education. Important techniques for building a usable simulation model are also presented. A good model has to be accompanied by texts, demonstration material, worksheets, teachers' guides, student manuals, as well as tools for the teacher in order to be able to make changes in the computer simulation program, like a text-editor, a graphic editor as well as a resource editor. In addition, the paper describes why it is important to follow the building scheme and discusses the problem of credibility. Then, these models are explained as a method of learning dependent on and independent from the use of computers and a view of its valuable aspects is shown. After presenting various examples from the primary and secondary schools to

the university and in lifelong learning, examples of their use in information sciences have been presented, as well as need to include the course of building simulation model in the curriculum.

Key words: simulation model, computer simulation, computer assisted learning, lifelong education, building scheme

Introduction

Simulation is a particular type of modelling. Building a model is a well recognised way of understanding the world; it is a simplification of some structure or a system. On the other hand, it can be a prediction, a substitute for experiential learning, or simply for entertainment. Here, we must mention the major difference between simulation and experimentation – in simulation one is experimenting with a model and not with a phenomenon. In our modern world we sometimes don't have the time to deal with a phenomenon, thus new technologies have brought us models which complete Aristotle's saying – *The things we have to learn before we do them, we learn by doing them*.

The use of simulations represents the natural way of "learn by doing". Just as children do simulation activities by role playing, adults use computer simulations in order to understand complex systems, real situation or dynamic processes. Computer simulations allow also analyzing situations or processes that would be difficult, impossible, dangerous, too long or to expensive to perform in real life.

There are different types of simulation models (ref. Wiki: Simulation) physical simulation (referring to physical objects or real things), interactive simulations (physical simulation with human operators included, e.g. flying or driving simulator), computer simulations simulating an abstract model of a particular system, with or without graphical output, etc. Computer simulations are formalized through mathematical models, using mathematical models, have become very useful in economics, chemistry, physics, psychology, in social sciences, in the process of engineering, technology, traffic, etc.

Computer simulation

According to Fishwick, computer simulation is the discipline of designing a model of an actual theoretical system, executing it on a digital computer and analyzing the execution output. A simulation model is actually a mathematical model calculating the impact of certain inputs and decisions on outcomes. Such a model could be created in a programming language, by statements or using formals in spreadsheets. Regardless its expression, the output in the simulation models depends on an uncertain input, and therefore we are dealing with uncertain variables and uncertain functions (according to "Simulations on Internet"). Emulation, in contrast to simulation, duplicates the system so that the second system behaves like the first one. It is related to imitation of another

program (e.g. printers that emulate HP LaserJet printers because then the printer can work with software for HP printer).

Why is computer simulation worth mentioning? It is interactive and it can be performed in different forms. Today it is very important for it to be used by pupils and students of different background knowledge, and different possibilities. That makes education focused at a student and his needs. We think that e-learning is one of the best models of learning environments today and a computer is the tool that connects pupils and students with lifelong education.

Interactive simulation is today's best form of putting simulation models into education in their fullest form. Computer simulation plays a big role in that implementation; it is a cognitive tool and a learning object. Nowadays, computer literacy is one of the key elements in basic education. Thus, computer, and all its possibilities and options, is a prime-mover for experiential and lifelong learning. Computer simulation has become a useful part of modelling many natural systems in physics, chemistry and biology, and human systems in economics, social and information science. A good example of the usefulness of using computers to simulate can be found in the field of network traffic simulation. In such simulations the model behaviour will change each simulation according to the set of initial parameters assumed for the environment. "Computer simulations are often considered to be *human out of the loop* simulations."[1]

The use of computer in education varies; it can be an instrument, an object, a help tool, a learning environment.

Computer simulation for educational purposes, which uses computer as a learning environment, is called computer assisted learning (CAL). Those so-called "training simulations" typically come in one of three categories, according to Wiki (Simulations):

- "live" simulation – real people use simulated equipment in the real world,
- "virtual" simulation – real people use simulated equipment in a virtual reality, and
- "constructive" simulation – simulated people use simulated equipment in a virtual reality.

It is important to notice that in all three cases people deal with simulated equipment and that shows the difference between simulation and experimentation, as mentioned earlier. Information science includes all of the above mentioned uses of a computer and categories of simulation. That is very important for the development of lifelong learning, with a very significant aspect of interactivity.

[1] Simulation. 28 August 2007. http://en.wikipedia.org/wiki/Simulation (6 August 2007)

Virtual learning environment (VLE)

Taking the aspect of interactivity into consideration, we must mention it can only be realized in a virtual environment. That virtual learning environment is the main pillar of today's information science education. It consists of multimedia, it facilitates e-learning, and it enables everyone to take part in lifelong learning. In such an environment, computer can be used in all of its varieties, as a source of information (as an instrument), in computer science or in electronics (as an object). Further, what is most emphasised is the learning environment, in which it can be used as a tutorial, practice, simulation, modelling, gaming, problem solving or an expert system.

Today, in education and especially information science everyone has to know what a computer is; in its latest form: the World Wide Web. A large amount of didactical and informative content is digital and the business of everyday life demands the possibility of distance learning. No one in the academic world can escape the powerful influence of the World Wide Web. It is increasingly being used for the development, distribution and usage of course materials, texts, books, animations; it is a "multimedia library"[2].

Ordinary model-driven computer simulation can be used as a learning tool and teachers and educational planners just love to use such learning tools in education. It is very important to mention the relevance of feedback in such interaction. Feedback, as maybe one of the most important didactical elements in education, therefore in this form of learning cycle also, can be expressed in many ways. On one side there is learning at distance with a possibility of immediate feedback and that is a really solid ground for strengthening the idea of lifelong learning. On the other side there is an individual approach to every student, what is very important for gaining the right evaluation of one's performance.

World Wide Web is a form of virtual learning environment, and in combination with didactical simulation models, it creates an environment suitable for experience learning. On the Web we can learn languages in so many ways and every tool there is, is a kind of a simulation model. There are dictionaries with suggestions, audio samples, interactive translators and all of these are just one part of a simulation. In an educational environment, the goal is to evolve and thus again feedback comes in the bigger picture. There are open learning environments for insight training or exercising, with little or no instruction. There are also learning environments where everything is stored in large complicated data files. Here intelligence and sometimes visual dynamic feedback play a major part. This complex, dynamic, intelligent feedback is very important to keep learning processes going. Sometimes this feedback is wordless, sometimes graphic, sometimes with all kinds of animations and/or animation-objects.

[2] Min Rik. Simulation and Discovery Learning in an age of zapping and searching. http://users.edte.utwente.nl/min/ (13 August 2007)

We could say that in a lifecycle of a simulation the first factor is imagination and the last one realization. In our opinion, to have a chance to put imagination into action is the best way of learning. In the educational process that is very important for the learning process of an individual, for the learning process within a group, where people can share their ideas and meet different opinions. Of course, that would also make the education of today more developed. Before the learning process is over students should be involved with simulations that in virtual learning environment allow them trying out ideas without risks and a possibility to learn on their own mistakes and have a chance to show (in a new attempt) that they had made progress. That is, once again, the role of feedback.

Simulation and learning

Simulation models could be used as a tool in education system, from primary and secondary school implemented from the users with educational and fun purpose, up to a high school system where use of specific simulation models helps certain research, helping in decision-making, or in the course relation to creation of simulation models.

Simulation model programs in education carry many advantages. A number of these do not specifically relate to simulation only but to its validity and credibility. Simulation model programs give

- the opportunity to experiment with phenomena or events,
- a sense of reality in abstract fields, like information science.

"Simulations are excellent discovery learning techniques that often offer insight or 'gestalts' not gained through more traditional didactic methods"[3]. Stated quote mentions discovery learning, that is the oldest form of learning and it is relevant to notice that modern simulations are encompassing that technique.

Further, there are disadvantages and the biggest problem is with credibility, because if credibility is doubtful, there is a problem with the reality that the simulation model presents.

A computer simulation model developed for the education system at the national University of Tres de Febrero enabled to foresee the impact of education policies in cultural, technological, social, educational and economic context, before their real implementation and to conclude whether planning, teaching, research and knowledge transfer matched the aims of the institution.

Learning on creation of simulation models includes basically three basic sub processes: model design, model execution and model analyses. Simulation models include notions such as modelling of the system, technology, safety, testing, training, acquisition of valid source information, key characteristics of the behaviour, simplifications, fidelity and validity of outcomes and education.

[3] Min Rik. Simulation Technology and Parallelism. http://projects.edte.utwente.nl/pi/Book/Contents.html (13 August 2007)

Simulation models can put students in the positions they will find themselves later in their professional lives. The simulation models e.g. can show them how to put together and optimise the computer, how to organise and fine-tune a network or a web-server depending on the number of users or information requests, how to work with a program, how to work with the fragile archival, library or museum materials, how to preserve them most efficiently (deterioration simulation in different environmental conditions), how to conduct the restoration a work of art, how to organise or digitise a collection, to name just a few possible applications. Even this small list shows various levels of expected knowledge.

On the web site Educational simulations students use simulations to understand different social processes, to simulate life events, to develop critical thinking strategies by identifying, recording facts, ideas and events, analyzing and synthesizing information and learning in social studies, English language, technology and geography. Other application enables them to create personal journal chronicling a simulated life for a specific region in the world. That emphasises the importance of individual approach in learning and shows how the possibility of learning at distance, in combination with the methods that use simulation, brings education to people and makes learning anything but a luxury.

The “Electronic United Nations” is an interactive educational simulation, intended for use in primary and secondary school where users by role-plays create classroom country and complete it with constitution, economy and judicial system. After connecting through the web with another classroom country, they interact on issues of global concern. The effect of this simulation method is, beside factual knowledge, learning about leadership, politics, and importance of law, order and politics.

Dynamic simulations are those that change in response to input signals (e.g. traffic simulation models). Continuous dynamic simulations are often used in simulation games (e.g. flight simulators) and in science (e.g. electrical circuits, chemical modelling). Computer simulations are used in a variety of practical context, such as weather forecasting, analysis of air pollution, noise, logistic system, flight simulators), etc. Computer graphics and animations can be for the graphic display and better visualisation.

Recent advances in computer technology have enabled to create three-dimensional (3D) dynamic models and simulate them within 3D environment. The uses of such models are very useful in teaching of simulation models, during construction of processes, in exploitation of visualising.

Simulation models could be also used in the language processing field, such as an intelligent interface helping modellers document simulation and synthesis models with controlled English (ref. Walling, R. C.; Pradeep, V.). The system is designed to enforce the writing of clear, high quality comments for models or programs. The quality of comments is regulated through the constraints on syntax and vocabulary and if the comments meet standards, they are inserted into the model.

Another example is comparison of user simulation models for dialog strategy learning (ref. Ai, H.) exploring what kind of the user simulation model is suitable for developing a training corpus to automatically learn dialog strategies.
There are simulation models developed to teach on new therapies, treatment and diagnosis, or in the use of placebo drugs in medicine. In finance, computer simulations are often used for risk analysis, discount rates, interests, etc.
Simulation models could be also used in primary and secondary education, e.g. at Chevening School, Sevenoaks children are encouraged to use simulations in order to discover patterns, rules and make predictions. One of online models is a simulation "Plant Force" (http://puzzling.caret.cam.ac.uk/game.php?game=plants) simulation where the user or a student working for a plant development company can find the optimum conditions for its growth and generate a profit. In the model "Food chain simulation" (http://puzzling.caret.cam.ac.uk/game.php?game=foodchain%20) children can learn about food chain components, like consumer, producers, time, production and ecosystem.
In the model "Mophases" (http://www.morphases.com/editor/) it is possible to simulate different characteristics of the face. With "Robot Constructor" (http://www.channel4.com/science/microsites/R/robots/) it is possible to create the own robot using different materials and components. With "Supercity Planner" (http://www.channel4.com/science/microsites/B/buildingthebest/game_city.html) having your own budge you can construct you own Supercity, or with "City Creator" (http://www.citycreator.com/) you can build towns and cities. The model "Dumptown" (http://www.epa.gov/recyclecity/gameintro .htm) is a recycling simulation.

Simulation models and curriculum

Learning tools are almost always intended as part of a series of lessons in a curriculum. These learning tools are suitable to motivate pupils and students to research further about the subject under study. In our modern time schools and institutions have increased possibilities, learning tools are more and more focused on designing complete electronic teaching packages which can be downloaded from the web, thus enabling online learning and, most importantly, lifelong learning. One should, however, be aware of the setbacks of poorly designed, integrated, virtual learning environments. In our opinion, setback is when simulation is used as a tool for a visual method of learning, because without the mix of simulation and experimentation, pupils and student cannot actually get to know the real situation. Thus, a method of practical learning, such as writing or some kind of realization of different ideas, is the best method to make the curriculum richer and more suitable for today's society.
Computer simulation as a learning tool is one of the main subjects in information science education. Computer simulation is an "ordinary" simulation, using a computer – the tool of present everyday life.

Implementation of simulation models as learning tools into the educational process can:

- provide an information or demonstrate/visualise a phenomenon,
- enable practice in learning (e.g. learning languages in many different forms),
- examine and test the knowledge / give individual feedback.

What applies to learning tools also applies to a curriculum – it should be well designed, that is as varied as possible. The basis of every well designed curriculum is the creation of a widely differentiated set of learning tools. In order to create a good curriculum with simulations, one should know the target group or designated community and its exact characteristics. Designers should know which students they are dealing with, what previous knowledge they have and what their individual attitude is. Simulation models can be applied throughout the curriculum, at the various study levels, with or without the requirement of the previously studied knowledge. Then insight should be tested to see if the students can apply (passive) knowledge acquired with the help of the simulation model learning tools.

How to build a simulation model

One form of e-learning that piques learners' as well as developers' interest is educational simulations. Simulations range in complexity from a straightforward reproduction of a software program to more intricate representations of complex environments such as an interactive community of workers. Educational simulations are designed to give learners an opportunity to practice their knowledge and skills in a risk-free environment. However, not all training needs can be met by simulations, and building simulations can be challenging for novice developers.

When implementing simulation model, there are some questions to think about, according Simulations on Internet:

- Was the simulation program realistic?
- Does the simulation represent a real life or fantasy situation?
- How realistic is it?
- Did the same things happen – was there a pattern?
- What are the variables, can you think of any others which might make the simulation more accurate?
- Was there more than one solution to the problem?
- What were food points/bad points?
- What did you find out?

A simulation model of a complex system can only be an approximation to the actual system, no matter how much time and money is spent on model building. There is no such thing as absolute model validity, nor is it even desired. Indeed, a model is supposed to be an abstraction and simplification of reality.

Building of a valid and credible simulation model usually follows these steps (Law, A. M.; McComas, M. G.):

1. Formulation of the problem
2. Collection of the information and construction of the conceptual model
3. Checking the validity of the conceptual model
4. Model programming
5. Checking the validity of the programmed model
6. Design, conduction and analysis of the simulation
7. Documentation and presentation of the simulation

1. Formulation of the problem

The problem of interest is stated by the decision-maker. The scope of the model and the system configurations that are going to be modelled has to be declared.

2. Collection of the information and construction of the conceptual model

It is necessary to collect information on the system layout and operating procedures. After that it is important to check the computer constraints, development timeframe and money constraints. There should not be a one-to-one correspondence between the model and the system.

3. Checking the validity of the conceptual model

One should perform a structured walk-through of the conceptual model before an audience that includes the project manager and analysts. If errors or omissions are discovered in the conceptual model, which is almost always the case, then the conceptual model must be updated before proceeding to programming in Step 4.

4. Model programming

The next step is programming the conceptual model in either a commercial simulation-software product or in a general-purpose programming language (e.g., C or C++) and verification (debugging) of the computer program.

5. Checking the validity of the programmed model

If there is an existing system, then the performance measurement of the model should be compared with the performance measurement collected from the actual system. This is called *results validation*. *Sensitivity analyses* should be performed on the programmed model to detect which model factors have the greatest effect on the performance measures. Those factors will later have the greatest impact on the overall model performance.

6. Design, conduction and analysis of the simulation
For each system configuration of interest, decision on tactical issues such as run length and warm-up period should be made. Analysis of the results and the decision if additional experiments are required should be made.

7. Documentation and presentation of the simulation
The documentation for the model should include the conceptual model, a detailed description of the computer program, and the results of the current study. The final presentation for the simulation study should include animations and a discussion of the model building/validation process in order to uphold model credibility.
There are many examples of simulation models in education which make it easier for students and teachers to explain and understand given lessons. There is a driving test good for coordination (http://www.learn4good.com/games/simulation/driverseducation.htm), simulation of credit card interest (http://vam.anest.ufl.edu/maren/interest.html), and various experiments-simulations for physics (http://www.walter-fendt.de/ph11e/).

The role of instructor in e-learning[4]

Web developers and instructional designers have come to realize that an important part of e-learning is the instructor, i.e. a real person or character that guides, instructs, mentors, and entertains the learner. In a classroom program, the instructor serves that role. He or she ensures that everyone is attentive and answers questions. But how can instructional designers transfer this live person to an e-learning model? Many developers think that the best solution is to use character simulations. Quite simply, character simulation is the practice of creating a relevant, interesting personality who appears and instructs throughout an e-learning program.

Three models for character simulations

The idea of a character simulation is to immerse the learner in a social experience in which a "person" speaks, interacts, and guides learning. This character may have a general persona or be based on a real and familiar personality.
These characters take on a personality through their visual impact, choice of language, attitude, and even voice. He or she appears regularly to help learners remember where they are in the program and to reinforce the learning material at the beginning or at the end of each module.
But before building a character, it's important to think about the role it will play. Typically, there are three roles for characters in online learning.

[4] http://www.learningcircuits.org/2004/jun2004/bersin.htm

Peer
In this approach, the character takes on the role of a peer to instruct or coach. For example, a 19-year old sales representative in a retail store will most likely be more receptive to a program that uses a young representative that wears clothes he/she may wear and discusses topics using slang. The character talks to the learner as a peer, and can suggest, joke, and mimic him.

Expert
In this model, the character takes on the role of an expert to instruct. The character may be a sales manager, flight instructor, or even Albert Einstein. For example, the Air Force program that Bersin & Associates reviewed was led by a flight instructor that wore shiny aviator glasses and talked in a tough voice, while explaining in detail everything the learner must do using his experience as a guide.

Authority Figure
In this model, the character is the boss. In one corporate certification program, for example, the character may be based on the company's actual CEO. She may explain to employees that this particular training program is not only mandatory but will be tracked by the executive management.

Suggestions for the development of the simulated characters

There are several basic suggested techniques that should be kept in mind during the designing process of simulated characters. Some of them are mentioned here.

Create life-like characters
Be deliberate about characteristics that seem minor and non-instructional, such as clothing style, speech and idioms, hairstyle, and voice. These aspects of the personality can make a big difference in the target audience's perception of the character. Before you design the whole program, create one or more characters and test whether the target audience likes and accepts them.

Plan the scenes prior to development
Like other forms of animation, character simulations use built-in scenes. In general, limit scenes to one minute in length, and think about pacing them appropriately for the learners. Storyboarding techniques can be very helpful in the planning process.

Check for understanding
Delivering the information without appropriate user interaction is insufficient for learning purposes. When you use character simulations, make sure that the characters interact with the learner to verify whether objectives are being met.

Have the character simulate common situations and use questions to check the learner's comprehension.

Focus on learning objectives
Get to the point quickly. Character-development products are full-featured and fun, but don't let their features and functions distract the learner. This may seem like common sense to a training professional, but it's easy to get sidetracked.
Use plain text before recording the final script
When you create the audio, create a script first and use plain text on screen until your script is finalised. After the script is ready, record the audio with a real voice. For major, high-stakes programs, hire a professional to record the voice.

Consider a multiskilled team
To do a full blown e-learning program that integrates character simulations into the courses and eventually in the curriculum, you will need writer or developer who can create storyboards and scripts, graphic artist who can develop backgrounds and other artwork and Web developer who can integrate all the files, including Flash, wave, and graphic files.

Run a pilot to gather FAQs
In many cases, character simulations act as a substitute for interactions with real people. It's important that the characters can answer normal questions that come up during the course. Use your pilot program as an opportunity to gather such questions.

Conclusion

In this article we have shown the different categories of computer simulation in connection with the learning process. Such simulation models interactively imitate the reality which would otherwise be very difficult to show to the learners or very dangerous to immerse them in it without the prior knowledge how to react or behave. Simulation models are therefore applicable throughout the learning process – from the school and university levels to the lifelong learning situations. No matter what is the target group level, it is always good to know the designated community of users in order to better customise the learning simulation model to them. Their expected knowledge and experience, or lack of it, is one of the key issues as is their feedback which can later help to fine-tune the model more appropriately.
We think that the issue of simulation models in information science education has at least two main aspects.
First, simulation models can and, in our opinion, should be used as a complement to the process of education. Therefore we can conclude that the simulation models can be used throughout the information science curriculum, from the beginning of the study to the very end of it. Furthermore, the usage of simula-

tion models can be taken as an important factor for differentiating the quality curricula from the curricula that do not incorporate simulation models.
Second, in our opinion information science curriculum should, at the higher levels of study, incorporate course(s) related to development of the simulation models. It could even be suggested to form a curriculum module for the development of simulation modules. With the earlier analysis in the article we aimed at this idea since we think it could be of a great importance for the advancement of the curriculum. In this context it is important to know how to organise a process of building a simulation model, and how to implement the e-educator. Since the application field is very wide we think that not only would the students of the information studies benefit from such course(s) or a module but the field of information sciences could advance also. The future information professionals would know how to build simulation models and thus model the needed environments which could in turn further the research in the field of information sciences more efficiently.

References

Ai, H.; Tetreault, J. R.; Litman, D. J. Comparing User Simulation Models for Dialog Strategy Learning. Proceedings Human Language Technologies, NAACL-HLT, 2007.

Educational simulations http://www.educationalsimulations.com/index.html

Fishwick, Paul A. Computer Simulation: The Art and Science of Digital World Construction, 1995. (http://www.cis.ufl.edu/~fishwick/introsim/paper.html)

Frigg, R.; Hartmann, S. Models in Science. Stanford Encyclopedia of Philosophy. (http://plato.stanford.edu/entries/models-science/)

Law, A. M.; McComas, M. G. How to build valid and credible simulation models. Proceedings of the 2001 Winter Simulation Conference, IEEE Computer Society Press, 2001.

Min Rik. Simulation and Discovery Learning in an age of zapping and searching. (http://users.edte.utwente.nl/min/)

Min Rik. Simulation Technology and Parallelism.
(http://projects.edte.utwente.nl/pi/Book/Contents.html)

Simulations on the Internet
(http://www.kented.org.uk/ngfl/software/simulations/index.htm)
(http://hcs.science.uva.nl/QRM)
(http://wilderdom.com/experential/)
(http://cress.soc.surrey.ac.uk/)

Transfer of Simulation Models for Education Policy in the Province of la Pampa, National University of Tres de Febrero, 2006.

Walling, R. C.; Pradeep, V. An Intelligent Interface for Model Documentation with Controlled English. Proceedings of the International Conference on Modelling and Simulation, 2001.

Wiki: http://en.wikipedia.org/wiki/Simulation

Wiki: http://en.wikipedia.org/wiki/socialsimulation

Načela oblikovanja edukativnog multimedijskog sadržaja u online sinkronom Web okruženju

Vladimir Mateljan
Odsjek za informacijske znanosti, Filozofski fakultet
Ivana Lučića 3, Zagreb, Hrvatska
vmatelja@ffzg.hr

Željko Širanović
Financijska agencija – FINA
Zagreb, Hrvatska
zsirano@ffzg.hr

Željka Širanović
Odsjek za informacijske znanosti, Filozofski fakultet
zsiranov@ffzg.hr

Sažetak

Suvremene informacijske tehnologije sve više prodiru u područje obrazovanja te omogućuju bolje i uspješnije učenje. Multimedijska prezentacija nastavne građe omogućuje da čovjek tijekom učenja bolje razumije i zapamti puno više sadržaja. Suvremena informacijska tehnologija uz distribuciju multimedijskog nastavnog sadržaja preko interneta/intraneta omogućuje sudionicima istovremeno i online komunikaciju u realnom vremenu. Ovakav oblik online edukacije u realnom vremenu poznata je i pod nazivom sinkrona Web edukacija, i ona danas pronalazi svoju praktičnu primjenu kako u školama i fakultetima tako i u suvremenim poduzećima. U radu će se dati opće preporuke koje bi trebali poznavati svi sudionici koji sudjeluju u oblikovanju multimedijskog nastavnog sadržaja, koji se prezentira preko interneta/intraneta pa tako i u tzv. sinkronom Web okruženju. Preporuke za oblikovanje ovakve vrste multimedijske građe temeljit će se na Mayerovim načelima multimedijskog učenja, kao i na osnovu didaktičko-metodičkih aktivnosti i etapa koje se pojavljuju u klasičnoj nastavi.

Ključne riječi: sinkrona, edukacija, multimedija, oblikovanje, web, sadržaj, online, načela

Uvod

Povijesno gledajući, počeci multimedijskog izražavanja obrazovnog sadržaja počinje vjerojatno iz doba pojave prvih tiskanih materijala, u kojima su se uz pisani tekst nalazili i slikovni prikazi. Slikovni prikazi trebali su pružiti bolje razumijevanje i shvaćanje sadržaja, te im dati veću edukacijsku vrijednost. U današnje vrijeme razvitak informacijskih tehnologija, računala i interneta pruža još veće mogućnosti, jer se nastavna građa osim tekstom i slikom može još izraziti zvukom, animacijom i videom. Mogućnost izražavanja građe na ovakav višestruko multimedijski način daje puno veće šanse uspješnijem i boljem učenju, pa možemo govoriti u multimedijalnom učenju. Uz to ako se ovakav multimedijski nastavni sadržaj smjesti na neki poslužitelj, CD-ROM ili DVD-u, tada računalom i internetom može biti prostorno i vremenski lako dostupan, što omogućuje veliku individualizaciju i mogućnost učenja bilo kada i bilo gdje.

Današnja i buduća primjena multimedije u edukaciji oslanja se na nove informatičke tehnologije, što zahtijeva od nastavnika i ovladavanje tim tehnologijama. To su prvenstveno nova znanja u korištenju interneta, poznavanje alata i tehnika oblikovanja multimedije. Danas je u izradu kvalitetnih edukativnih multimedijskih materijala najčešće uključeno više stručnjaka poput dizajnera, programera, nastavnika i sl. Važno je naglasiti samo poznavanje suvremenih informatičkih tehnologija ne rješava samo po sebi problematiku izrade kvalitetne multimedije, tehnologija je samo sredstvo. Sudionici izrade multimedije svakako moraju poznavati tehnologiju i njene mogućnosti, no s druge strana morali bi znati kako didaktičko-metodički oblikovati i prilagoditi građu edukativnu multimedijsku nastavnu građu.

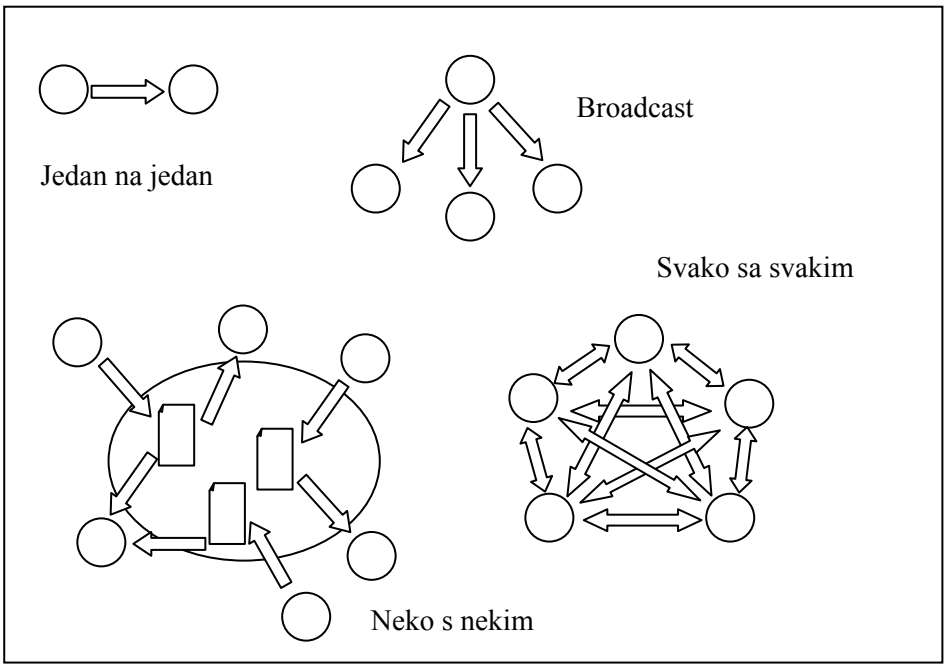
Jedan od ciljeva oblikovanja edukativnog multimedijskog materijala trebao bi biti taj da se građa oblikuje na način da je recipijenti što bolje shvate, razumiju i zapamte. U radu će se navesti neka praktična načela za oblikovanje nastavnog multimedijskog sadržaja temeljena na Mayerovim načelima multimedijskog učenja i nekim praktičnim iskustvima u izvođenju sinkrone web nastave. U nastavku će se kroz osnovne didaktičko-metodičkim etape nastavne jedinice, skladu sa suvremenim teorijama kognitivnog učenja, prikazati osnovna pravila oblikovanja multimedijskog sadržaja.

Sinkrono online okruženje

Učenje preko Web-a u osnovi temelji se na dva modela asinkronom i sinkronom. Kod asinkronog modela za vrijeme učenja između učenika i učitelja nema direktnog kontakta, dok učenik dohvaća multimedijski nastavni sadržaj i uči, nastavnika nema mogućnost nikakvog direktnog utjecaja u realnom vremenu. Kontrole i mehanizmi koji upravljaju učenjem moraju se predvidjeti i ugraditi unaprijed, npr. kod interaktivne animacije neće se dozvoliti nastavak ako učenik ne postupi kako se od njega očekuje. Interakcija između učenika i nastavnika može ići tek naknadno npr. mailom i na neki drugi sličan način.

U sinkronom modelu za vrijeme učenja između nastavnika i učenika postoji on-line interakcija u realnom vremenu (npr. chat, whiteboard, videokonferencija), što daje dodatnu mogućnost direktnog praćenja i usmjeravanja od strane nastavnika (ili drugih sudionika). Sinkronim pristupom moguće je direktno i kvalitativno upravljati procesom učenja u realnom vremenu.

Slika 1. Topologija razmjene multimedijskih poruka – komunikacijska suradnja između sudionika



Sudionici u sinkronom on-line okruženju (slika 1) prema Hortonu mogu međusobno surađivati i razmjenjivati multimedijske poruke na nekoliko načina [12]:

- *Jedan na jedan* – Poruke izmjenjuju dvije osobe (primatelj i pošiljatelj),
- *Broadcast* – Ovakav način omogućuje jednoj osobi najčešće moderatoru slanje multimedijske poruke svakom sudioniku istovremeno (instruktor može poslati prezentaciju, zadatak ili vježbu istovremeno svim sudionicima),
- *Neko s nekim* – U ovakvom načinu rada u nekom trenutku svatko može poslati nekome poruku (instruktor podijeli zadatke, učenici rješavaju zadatak i pojedinci ili grupe međusobno komuniciraju, instruktor nadzire proces),
- *Svako sa svakim* – U ovakvom načinu svi međusobno razmjenjuju poruke (učenici rješavaju zadatak i međusobno komuniciraju, instruktor nadzire proces pomaže i usmjerava ili svi zajedno rješavaju problem).

Ova četiri jednostavna oblika međusobne suradnje sudionika u modelu sinkronom Web edukacije najčešće se međusobno prožimaju te se načini međusobne komunikacije mijenjaju tijekom sinkrone sesije. Na primjer instruktor može u jednom trenutku svim sudionicima poslati test sa zadacima (broadcast), nakon čega sudionici za vrijeme rješavanja zadataka mogu prijeći na druge oblike međusobne suradnje (jedan na jedan ili neko s nekim).

Koji oblik suradnje će sudionici koristiti ovisit će o tehnologiji (npr. brzini i kvaliteti on-line konekcije), njihovoj sklonosti prema pojedinom komunikacijskom alatu (chat, whiteboard, dijeljenje zaslona, audio ili video konferencija) kao i razini usvojenih vještina potrebnih za korištenje pojedinog alata (npr. brzinom tipkanja po tipkovnici) i sl. Horton navodi [12], pozivajući se na istraživanja s Wester Michigan University, da su polaznici najčešće koristili sljedeće alate:

- WEB stranice 100%,
- Online (NetMeeting) 100%,
- Whiteboard 83%,
- Screen sharing 57%,
- Chat 30%.

Bez obzira kako i na koji način recipijenti koriste pojedine alate u sinkronom modelu učenja, važno je da im multimedijska građa bude prezentira na način koji će im olakšati on-line učenje. Oni koji su zaduženi za oblikovanje edukativnog multimedijskog materijala trebali bi znati koje su to aktivnosti odnosno događaji koji dovode do boljeg razumijevanja i zapamćivanja nastavne građe.

Aktivnosti i etape nastave

Dobra prethodna priprema trebala bi garantirati uspješnu Web sinkronu edukaciju, pri čemu se u konačnici očekuje razumijevanje i zapamćivanje određene građe od strane recipijenata, kao i njeno što duže ili trajno zadržavanje. Razumijevanje, zapamćivanje i zadržavanje građe tijekom sinkrone online edukacije zahtijeva aktiviranje vanjskih događaja, Gagne [13][14]:

- *Privlačenje pozornosti*, (izazivanje podražaja i uzbuđenja, često starta problemskim pitanjima ili postavljanjem problema kojeg treba riješiti),
- *Informacija recipijenta o ciljevima učenja,* (istaknuti ciljeve i svrhu učenja te naglasiti praktičnu primjenu i osobnu korist od naučenog),
- *Stimulacija dosjećanja nužnog prethodnog znanja*, (povezati novo sa starim, iz dugoročnog pamćenja „osvježiti" nužna relevantna znanja, pravila, procedure ili vještine).
- *Prezentacija novog gradiva* (izložiti relevantnu građu, tekstove i slike kroz aktiviranje selektivnog pristupa),
- *Usmjeravanje učenja* (organiziranje i semantičko kodiranje informacija),
- *Poticanje na aktivnost* (aktiviranje učenika na reagiranje),
- *Davanje povratnih informacija* (pokrepljivanje),

- *Procjena izvedenog* (aktivacija pretraživanja radi omogućavanja potkrepljenja),
- *Pojačavanje retencije i transfera* (osiguranje znakova i puteva za pretraživanje na slične problemske situacije).

Navedene aktivnosti ne moraju se u online sinkronom okruženju izvoditi navedenim redoslijedom, ali je u fazi pripreme za ovakvu vrstu edukacije potrebno procijeniti koje su aktivnosti nužne, a koje se mogu ispustiti. Da bi se zacrtane aktivnosti mogle ostvariti kroz sinkronu multimedijsku nastavnu jedinicu, slično kao i kod klasične potrebno ju je prethodno dobro pripremiti. Izvođenje jedne takve nastavne jedinice odnosno ostvarenje zacrtanih aktivnosti obavlja se kroz nekoliko didaktičko-metodičkih etapa. Postoje razni modeli izvođenja nastavne jedinice, pojedini autori navode različite didaktičko-metodičke faze ili etape, npr. prema Poljaku [6][11] su to najčešće:

- Pripremanje ili uvođenje učenika u rad,
- Obrada novog sadržaja,
- Vježbanje,
- Ponavljanje,
- Provjera i ocjenjivanje.

Tablica 1. Primjer mogućih etapa i uloge sudionika u sinkronoj multimedijskoj nastavi

Etape	**Uloge sudionika za vrijeme sesije**			**Komunikacija**
	Instruktor	**Polaznici**	**Gost (ekspert)**	
Uvod (5 min) Instruktor uvodi učenike u rad. Najavljuje svrhu i ciljeve.	Prezentira	Slušanje		Broadcast, (NetMeeting)
Prezentacija (20 min) Ekspert izlaže svoj originalni materijal.		Slušanje	Prezentira	Broadcast (NetMeeting)
Pitanja i odgovori (20 min) Instruktor moderira dok recipijenti postavljaju pitanja ekspertu, koji im odgovara.	Moderira	Postavljanje pitanja	Odgovara	Neko s nekim i jedan na jedan (NetMeeting)
Sažetak (5 min) Ekspert rekapitulira izloženu prezentaciju i odgovore na postavljena pitanja.			Sažeta	Brodcast (NetMeeting)
Evaluacija (5 min) Instruktor radi pregled sesije, istakne ostvarene ciljeve i sugerira smjerove za daljnji napredak.	Evaluira			Broadcast (NetMeeting)

Mogući scenarij primjene jedne takve Web sinkrone nastavne sesije u kojoj je osim instruktora i recipijenta sudjeluje i gost-stručnjak (tablica 1) može imati sljedeće etape, Horton [12]:

- Uvod (odgovara Poljakovoj pripremi i uvođenju učenika u rad),
- Prezentacija (obrada sadržaja kojeg će izložiti stručnjak),
- Pitanja i odgovori (učenici će postavljati stručnjaku),
- Sažetak ili zaključak (stručnjak radi rekapitulaciju sadržaja),
- Evaluacija (instruktor rekapitulira cijelu sesiju i usmjerava recipijente na daljnje istraživanje).

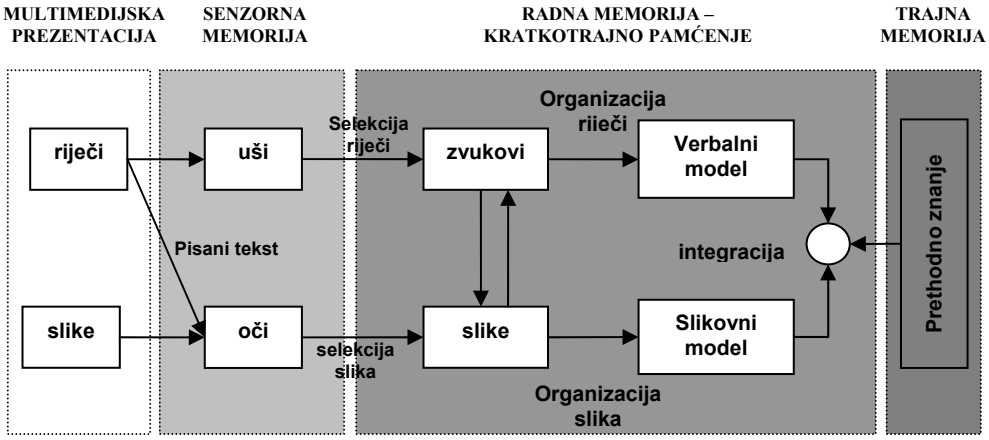
Prezentacija i sva građa koja će biti prezentirana i podržana multimedijom trebala bi biti izrađena prema načelima multimedije, jer će to dovesti do boljeg razumijevanja i zapamćivanja sadržaja. U koliko osobe koje pripremaju prezentacijsku građu ne poznaju tehnike i načela oblikovanja multimedije najčešće neće dobiti maksimalan efekt i jednog ovakvog ne baš jeftinog projekta, a to je bolje razumijevanje i zapamćivanje građe od strane recipijenata. Nije dovoljno da samo poznaju tehnologiji već i ono što će omogućiti da recipijenti bolje shvate i zapamte građu.

Polazište za razumijevanje načela multimedijskog učenja

Za razliku od tehnološkog pristupa oblikovanja multimedijskog sadržaja gdje je u fokusu naglasak na tehnološkoj funkcionalnosti uspješnog prijenosa multimedijske poruke, Mayer [1][2] kao polaznu točku uzima recipijenta. Ovaj pristup ima u fokusu učenika kao polaznu točku kada se promišlja oblikovanje multimedijske poruke. Kod ovog pristupa cilj je pomoći recipijentu da što bolje shvati i zapamti informaciju. Za ovaj pristup Mayer definira načela kako oblikovati multimedijski sadržaj, a onda kako ga prilagoditi i koristiti multimedijsku tehnologiju kako bi se pomoglo ljudskoj spoznaji. Mayerova načela [1][2] temelje se na sljedećim teorijama:

- Teorije dualnog kodiranja koju je postavio Paivio 1986. Prema toj teoriji ljudi posjeduju odvojene nezavisne kanale za procesiranje vizualnih i auditornih informacija,
- Teorije kognitivnog opterećenja koju je postavio Sweller &Chandler 1991., prema kojoj su ljudska bića limitirana u količini informacija koje mogu istovremeno procesirati u svakom pojedinom kanalu,
- Modela radne memorije, postavljenom od strane Baddeleya, 1986.,
- Modela aktivnih procesa (Mayer 1999, Wittrock 1989). Aktivno učenje pri usmjeravanju pažnje na relevantne informacije, organizacijom selektiranih informacija u koherentne mentalne reprezentacije i integracija mentalnih reprezentacija s prethodnim znanjem.

Slika 2. Mayerov model multimedijskog učenja



Ukratko, Mayerov model (slika 2) funkcionira kroz pet glavnih koraka [1,2]:

1. Selektiranje relevantnih riječi (verbalnih poruka) iz okoline i njihova obrada u radnoj memoriji koja obrađuje zvukovne sadržaje,
2. Selektiranje relevantnih slika i njihova obrada u radnoj memoriji zaduženoj za vizualnu obradu,
3. Organizacija selektiranih riječi u verbalni mentalni model,
4. Organizacija selektiranih slika u vizualni mentalni model,
5. Integracija izgrađene verbalne i vizualne reprezentacije s prethodnim znanjem.

Osnovna načela multimedijskog učenja

Prema Mayeru [1][2] recipijent postiže bolje zapamćivanje i razumijevanje građe ako je ona prezentirana tekstom i slikom nego ako je prezentiran samo tekstom. Mayer pod tekstom podrazumijeva [1] govoreni i/ili pisani tekst, dok pod slikama podrazumijeva sve oblike statičkih (fotografije, grafovi, ilustracije, i sl.) ili dinamičkih slika (video i animacije). Ovo načelo je osnovna ideja vodilja oblikovanja multimedijske nastavne jedinice, bez obzira da li se ona izvodi u WEB okruženju sinkrono ili asinkrono. Uz načelo multimedije Mayer [1][2][10] definira sljedeća osnovna načela multimedijskog učenja:

- Prostorna i vremenska povezanost sadržaja,
- Modalitet,
- Zalihost (redundantnost),
- Segmentacija,
- Koherencija,
- Utjecaj individualnih razlika.

Prostorna i vremenska povezanost sadržaja: Razumijevanje i zapamćivanje sadržaja je bolje ako je srodni tekst i slika prostorno što bliže jedan drugomu, te ih treba prikazivati simultano i sinkronizirano.

Prostorna i vremenska povezanost sadržaja su dva srodna čimbenika oblikovanja obrazovnog multimedijskog sadržaja. Prostorna povezanosti naglašava združivanje teksta i odgovarajućih slika prostorno što bliže, dok pravilo vremenske usklađenosti ističe njihovo vremensko zbližavanje. Relevantni tekst koji objašnjava pojedine dijelove ilustracije po mogućnosti treba smjestiti unutar ili tik do statičkih slika, a nikako ih ne treba razdvajati. Ako se kombinira tekst i pokretne slike (animacija), bolje je uz pokretne slike (animaciju, video i sl.) istovremeno koristiti simultanu naraciju nego tekst. Na ovaj način recipijentu je omogućena istovremena čujnost i vidnost informacije u skladu s Mayerovim modelom multimedijskog učenja.

Modalitet: Bolje učenje postiže se iz grafike i naracije (govorenog teksta) nego grafike i pisanog teksta.

Edukacijski sadržaj izražen u vizualnoj formi (naročito pokretne slike) bolje je objašnjavati kroz naraciju nego s puno pisanog teksta na zaslonu, na taj način prema Mayerovu modelu multimedijskog učenja, informacije ulaze u kratkotrajnu memoriju svaka kroz svoj kanal (ne dolazi do „miješanja" slike i zvuka) te se tako primljeni sadržaji bolje obrađuju u radnoj memoriji.

Redundantnost (zalihost): Bolje multimedijsko učenje postiže se ako se iste informacije ne prezentiraju u više formata.

Ovo načelo sugerira da je nepotrebno istu informaciju iskazivati na više načina tako da opterećuju jedan senzorni kanal. Na primjer, ako se multimedijska poruka prikazuje slikom i naracijom, onda nije potrebno tu istu naraciju posebno prikazivati kao pisani tekst, odnosno slika se može objasniti samo tekstom ili samo naracijom. Svejedno je da li će istovrsna informacija biti govorena ili pisana, bitno je izbjeći istovremenu prezentaciju pisanog teksta i njegova naracija, jer prema teoriji dualnog kodiranja govoreni tekst koji ulazi kroz slušni kanal, istovremeno ulazi i kao slika u vizualni kanal i time ga nepotrebno opterećuje.

Segmentacija: Složeniji multimedijski sadržaj bolje se uči ako je razdijeljen na manje dijelove, nego kada je dan kao jedinstvena kontinuirana jedinica.

Ljudi bolje uče ako su složenije multimedijske poruke razdijeljene ravnomjerno u manje jednostavnije dijelove, nego kada je ta poruka dana kao jedinstvena kontinuirana jedinica. Na primjer ako cijela animirana naracija bez prestanka traje duže vrijeme (npr. 120 sekundi), onda bi trebalo logički razdijeliti na segmente (npr. 10 kratkih, svaka u trajanju od 12 sekundi). Pri tome treba omogućiti recipijentu mogućnost samostalnog aktiviranja svakog sljedećeg segmenta (npr. klikom miša).

Koherencija: Bolje učenje se postiže ako su dodatni zanimljivi ali nebitni sadržaji isključeni nego kada su uključeni.

Koherentnost definira odnos između više i manje bitnim sadržajima multimedijskih struktura. Po ovom načelu treba izbjegavati umetanje zanimljivih mul-

timedijskih sadržaja koji su manje bitni ili čak nerelevantni u odnosnu na osnovni sadržaj, jer oni remete procese konstrukcije znanja a pojačavaju emocionalni učinak. Zbog toga je ovakvu vrstu zanimljivih sadržaja bolje stavljati u etapu psihološke pripreme i uvođenja učenika u rad, jer po teoriji emocionalnog interesa podiže početnu razinu motivacije.

Utjecaj individualnih razlika: Dobar multimedijski dizajna ima veći efekt na recipijente s malo prethodnog znanja u odnosu na one s više znanja, te na one koji imaju bolje vizualne sposobnosti u odnosu na one kojima su te sposobnosti slabije.

Prema spoznajnoj teoriji, razumijevanje sadržaja ovisi o sposobnosti povezivanja vizualnih i odgovarajućih verbalnih prikaza koji se istovremeno nalaze u radnoj memoriji. Ako je multimedijska poruka loše oblikovana, recipijent s dobrim predznanjem iskoristit će svoje prethodno znanje da bi nadoknadio nedostatak, a recipijent s lošim predznanjem to neće moći. Ako je multimedijska poruka dobro oblikovana i jedni i drugi će je razumjeti. To znači da će primatelji s lošim predznanjem imati veću koristi od primjene načela oblikovanja multimedijske poruke. To na primjer znači da će za pojašnjenje nekog događaja recipijenti s nižom razinom predznanja trebati više slika (npr. dijagrama) i teksta (npr. naracije), dok će recipijentima s višom razinom predznanja biti dovoljna samo grafički prikaz bez puno teksta. Isto tako recipijent koji ima bolje vizualne sposobnosti bolje će kodirati i izgraditi mentalne slikovne modele i njihove međusobne veze u odnosu na onog koji ima poteškoća s izgradnjom takvih modala. Dakle, kod oblikovanja multimedijske poruke važno je voditi računa o individualnim razlikama između primatelja poruke, posebice o razini njihovog predznanja vezano za konkretni sadržaj.

Zaključak

Multimedijski nastavni sadržaji omogućuje bolje i uspješnije učenje, građa koja se izlaže na ovaj način omogućuje da čovjek tijekom učenja bolje razumije i zapamti puno više sadržaja, što vrijedi i za nastavnu građu koja se izlaže preko Web-a u tzv. sinkronom načinu rada, kada su svi sudionici komuniciraju u realnom vremenu. U radu su dane preporuke za oblikovanje sinkrone multimedijske Web nastave, na Mayerovim načelima multimedijskog učenja, te na osnovu najčešćih didaktičko-metodičkih etapa koje se pojavljuju u klasičnoj nastavi. Najvažnije preporuke koje proizlaze iz Mayerovih načela su:

- Prostorno i vremenski oblikovati sadržaj na način da su tekst i slika prostorno što bliži te vremenski usklađeni,
- Animacije trebaju biti govorene a ne pisana,
- Redundantne sadržaje treba eliminirati,
- Složeniju animaciju treba segmentirati,
- Eliminiranjem zanimljivi ali nebitnih sadržaja postići koherenciju multimedijskog sadržaja,

- U oblikovanju multimedijskog sadržaja uvažavati individualne razlike recipijenata.

Izložena načela u skladu s Mayerovim modelom multimedijskog učenja predstavljaju opće preporuke za oblikovanja multimedijskih sadržaja u Web sinkronom mrežnom okruženju.

Literatura

Clark, R.C., Lyons, C., Graphics for Learning, Pfeifer an imprint of John Wiley&Sons, Inc, 2004.
Desforges Charles: Uspješno učenje i poučavanje psihologijski pristup, Educa, Zagreb, 2001.
Grgin, T., Edukacijska psihologija, Naklada Slap, Jastrebarsko, 1997.
Gugić, I., grupa utora, Priručnik metodike za nastavu računalstva i informatike, Informatička škola Pentium, Vinkovci, 1997.
Horton, W., Designing Web-based training, John & Wiley, New York, 2000.
Mayer, E., Richard, Multimedia Learning, Cambridge University Press, Cambridge, 2001.
Mayer, E., Richard, The Cambridge Handbook of Multimedia Learning, Cambridge University Press, 2005.
Mikelić, N., Modeli i pravila oblikovanja multimedijske poruke i njen utjecaj na zapamćivanje i razumijevanje sadržaja, Magistarski rad, Zagreb, 2003.
Pastuović, N., Edukologija, Znamen, Zagreb, 1999.
Petričević D., skupina autora, Opći priručnik za nastavnike srednjih škola, Hrvatski pedagoško-književni zbor, Zagreb, 1994.
Stevanović M., Interaktivna stvaralačka nastava, Andromeda d.o.o., Rijeka, 2003.
Stevanović M., Modeli kreativne nastave, Andromeda d.o.o., Rijeka, 2003.
Širanović, Ž., A Practical Approach to the Designing of Distributed Learning Solutions,13th International Conference on Information and Intelligent Systems, Varaždin, Croatia, September, 2002.

Internetski izvori

http://edutechwiki.unige.ch/en/Instructional_design
http://edutechwiki.unige.ch/en/Nine_events_of_instruction

Project of Developing the Multimedia Software Supporting Teaching and Learning English Vocabulary

Tomislava Lauc
Department of Information Sciences
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
tlauc@ffzg.hr

Sanja Matić
Department of Information Sciences
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
smatic@ffzg.hr

Nives Mikelić Preradović
Department of Information Sciences
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
nmikelic@ffzg.hr

Summary

Since the Academic Year 2005-2006 we have worked on a project of developing the interactive multimedia software for learning English vocabulary. The software was tested on junior learners in elementary school within the English language course lesson. The results were measured by comparing the test results from the computer-aided lesson with the ones from a traditional one and they showed a great level of disparity both regarding the learners' knowledge and their motivation. Vocabulary of a language does not only consist of separate contextually independent words, but also of contextually bound lexical items, such as collocations. In 2006-2007 we have expanded the software by adding the feature for multimedia collocations. By terms of imitating reality, we have been designing multimedia animated presentations of English collocations comprising picture, text and sound. Our past, present and future phases of the project have purpose of examining the use of animation in educational settings.

Key words: computer assisted language learning (CALL), second language acquisition (SL), multimedia collocation, educational software

Introduction

Entertainment media is more widespread than educational media, especially among younger second language (SL) learners. SL teachers need to arm themselves with the learning content as attractive and motivating as possible, in order to engage their learners in the learning process more effectively. As technology is being more accessible to the masses, more and more teachers approach computer-assisted language learning (CALL) methods as an alternative to traditional time-consuming activities. CALL showed to be useful, not only as a device for drawing the learners' attention, but also as a method for maintaining the learners' interest and long-term memory. Children in Croatia start attending schools with a certain level of English language knowledge. Movies, music videos, video games, etc. play a significant role in SL acquisition. Adjusting the learning content to keep up with the modern media trends proves to invoke higher learners' interest in the learning content and obtains better results.
Most learners perceive learning vocabulary as learning new words, contextually independent and separate. But, faced with various multi-word combinations, the learners often tend to translate literally. This word-for-word translation does not function with lexical items that are language specific, that can be instinctively recognized only by native speakers, and therefore SL teachers have to raise the learners' awareness of the existence of these items.
Collocations are the most typical example of language specific multi-word combinations that have to be learnt by heart by SL learners. Of course, having the maintenance of long-term memory as a primary goal, the approach to such learning process has to be as motivating as possible. Ensuring a natural and spontaneous environment for SL acquisition can indeed compete the strength and influence of mass media.
We have been developing our multimedia educational software for English language vocabulary since the Academic Year 2005-2006, and in 2006-2007, together with our students attending *Multimedia knowledge presentation* course, we have expanded the software by adding features of multimedia collocations.

Current state and research

Until recently, computer-assisted language learning (CALL) was a topic of relevance mostly to those with a special interest in that area. Today, the majority of language instructors must begin to think about the implications of computers for language learning. In the practice courseware, the computer serves as a vehicle for delivering instructional materials to the learner. It is proved that repeated exposure to the same material becomes beneficial or even essential for learning. This makes computer ideal for carrying out repeated exercises, since it does not get bored with presenting the same material and can provide immediate feedback. A computer can also present such material on an individualized basis, enabling learners to proceed at their own pace. CALL is one of few methods proved to make the boring drill process of language acquisition inter-

esting and motivating. Multimedia technology integrated with the computer assisted learning allows the different media (text, graphics, sound, animation, and video) to be accessed on a single machine. This creates a quite authentic learning environment and different skills are easily integrated, since the variety of media makes it natural to combine reading, writing, speaking and listening in a single activity. Also, CALL nowadays provides a lot of opportunities for interaction with different characters and cultural information. Computer technology offers an alternative to time-consuming activities that are part of traditional language learning methods.

The aim of our project is to produce multimedia educational software in English language course that additionally motivates learners to learn faster and get better results. Dealing with learning material in electronic environment accelerates the process of material exchange and feedback on exercises allowing the user to learn in a dynamic and interactive way. Therefore, compared to the classic method of language learning, it conveys better results concerning the learner's acquisition of knowledge [3].

The project idea is motivated by multimedia software Globetrotter [5] that is the first prototype of multimedia software aimed at making the study of the lexicon of a foreign language (English) more efficient and less boring. The tool is destined for the teachers and pupils of the primary and junior high schools and can be exploited by even the least experienced computer technology users. The tool is child-oriented, highly interactive, open and flexible. These features make it particularly interesting from a didactic point of view. It allows the users (teachers and pupils alike) not only to consult but also to enrich the dictionary of the system. The teacher, with the aid of multimedia technology, can prepare efficient and appealing didactic units, while the pupils can explore and increase the material available, by constructing, completing and structuring their lexical knowledge in a foreign language.

Software description

Within our project the software is being designed in the authoring tool Adobe Flash. It comprises of picture, sound, text and animation and is very intuitive, user friendly and children oriented, offering clear and unambiguous navigation. The part developed so far consists of and introductory comic and six activities. The comic introduces the learning material by illustrating a spontaneous everyday-life sequence of events. The protagonists' actions are animated and accompanied with both text and sound. The vocabulary to be learnt is highlighted and is repeatedly presented through the six different activities (fill-in-the-blank, coloring book, spelling exercise, drag-and-drop, make a selection and memory game). It counts 24 words, consisting of contextually linked 8 verbs and 16 nouns designed as a learning content of an English language lesson for SL elementary students aged 9 or 10.

The software's interface is two-dimensional and simply designed in a form of a vector graphic using little memory space. However, having in mind that a purely graphical presentation may fail to aid comprehension, as Sundberg [6] claims, we have accompanied the animated graphics both with text and sound. Our interface is based on a psycholinguistic point of view, where language is perceived as a multi-level, symbolic system connecting audible (phonological) representations with semantic representations.

The two-dimensional presentation is enriched, not only by text and sound, but also by adding another dimension – a dimension of time illustrated by animated frame-by-frame motion, i.e. sequence of events. Text and sound are "played" during the animated presentation, not before, nor after, because, according to Paivio's dual coding theory [5], simultaneous use of both visual and verbal code is preferable. Learning concrete vocabulary items simultaneously with their physical-world counterparts should improve both speed of comprehension and retention in long term memory. With maximum effectiveness being the goal, we introduced the sequencing in presentation of instructional modes – text, sound and animation.

Another important aspect of the software is the interactivity of the exercises with an instant feedback. Learners become active participants of the learning process. Interactivity of the exercises and user-friendly navigation enables the learners to feel in charge, to decide which part should they pay more attention to, to control the time needed to deal with an exercise, to repeat the exercises by trying them several times in order to, either understand them better or to memorize the learning material. Depending on the type of activity, the learners are informed on their success, measured by their results or by an instant feedback on the correctness of their answers. Anitha Devi [1] claims that immediate feedback with a pleasant animation should be a positive reinforcement only for a correct answer and not a wrong answer. Any animation given for a wrong answer as feedback should not be as pleasant as the one given for the correct answer. However, we have deliberately not paid much attention to the difference in feedback mood depending on the correctness of the answer, not only for not discouraging the learners to proceed, but also in order to support the simplicity as much as possible. We believe that any redundancy draws the learners' attention away from the learning content, especially the attention-getting devices which should be used sparingly. Therefore, the cosmetics being of secondary importance to us, we used simple graphics with a limited choice of colors. We set focus on the content presented by imitation of reality, especially within the latest feature of our software – the multimedia collocation. Multimedia collocation feature builds up on our already existent software [3].

Multimedia collocation

As mentioned in the introductory paragraph, language does not only consist of stand-alone vocabulary items and grammar, but it also consists of multi-word combinations, such as collocations. A collocation is an arrangement or juxtaposition of words, especially those that commonly co-occur. Native speakers are spontaneously aware of these language-specific, context-bound and ready-made chunks in their mother tongue, because the sequence of native speakers' language acquisition has a natural flow, in the reality itself. With spontaneity and reality being the formulae for a natural language acquisition, our SL teachers should try to provide our learners with at least illustration of reality. Animation of both lexical segments of a collocation gives an impression of connectivity, i.e. stresses the juxtaposition itself, the motion enriches the picture-only presentation in terms of adding the sense of reality to the presentation, and this aspect enables a learner to perceive a collocation on a non-abstract level. This approach creates an illusion of an incidental, spontaneous and not stressful learning environment, engaging the learners on a latently obligatory level, and we believe this will prove highly efficient in our further research. Studies of language acquisition have increasingly given recognition to the role of memorization and repetition of these units [2] and therefore we added to our software multimedia collocations that develop the learner's ability to retrieve and combine ready-made chunks of language through interactive multimedia usage.

Furthermore, the presentation is also accompanied by text and sound. The interactivity of multimedia collocation exercises is achieved by fill-in-the-blanks activities. According to our previous experience with the fill-in-the-blank method, it proved to be effective in checking the learners' comprehension in a way that it examines the learners' awareness of the contextually linked missing words they have to fill in. This way the learners' attention is concentrated upon these ready-made chunks they have to recognize. As Lewis emphasizes, ready-made chunks are of high importance for language fluency and accuracy. Also, developing the learners' ability to recognize these chunks should be of high importance for SL teachers.

Implementation and research

Once we made the six-part software ready for test usage, we presented the approach for teaching and learning English language vocabulary within the English language course lesson using multimedia educational software. Its usage was tested within a Croatian curriculum with third graders, learners of English as a SL. The level of motivation and achieved results was measured by comparison of the computer-aided lesson with the traditional one. We offered two different presentations of the same content – the traditional lesson and the computer-aided one. All the learners' existent knowledge was tested before the two lessons in order to make sure that the learners' knowledge from both of the groups was in balance. After each of the lessons we gave the learners the final

test to examine their acquisition of the learning material. The learners from the group attending the computer-aided lesson conveyed significantly better results that the learners from the group attending the standard lesson. Not only did our approach make a difference regarding the learners' results, but it also showed a higher level of motivation among the learners dealing with the software.[1]
We had an opportunity to present our software in a workshop on multimedia for elementary, secondary and higher education teachers and the reactions of the attendees were more that welcoming. The multimedia collocations' usage will also be tested with SL learners when the level of difficulty is established. Also, as a further project researching the use of animation in educational settings, we intend to develop multimedia features for learning language tenses and prepositions.

Static versus dynamic

The use of animation in second language learning showed to be instructively efficient, especially with segments of vocabulary involving temporal and spatial aspects, such as prepositions, tenses, multi-word forms, etc. Except for the interactivity of animated exercises engaging the learners' attention and motivation to actively participate in a learning process, illusion of reality is also an important feature of instructional dynamics. By illustrating the natural sequence of events, i.e. by imitating reality, an illusion of spontaneity helps the learners, especially younger ones, to memorize the context-bound lexical items more easily. This way the content becomes non-abstract, and therefore easily understandable. A text-only approach, as in linear textbooks, is limited in engaging the learners' perception. Furthermore, textual content involving pictures is not as efficient in imitating reality as animated content is. Two-dimensional and three-dimensional presentation is highly enriched by including the fourth dimension – time. Images move sequentially within a time frame and this motion follows the natural way of human thinking, and therefore the process of learning becomes less abstract. With the implementation of sound, the list of things and events to be presented is endless. Animated presentation comprises of real objects, real events and real participants of action and therefore fosters contextual comprehension among learners. What we perceive as animation, is a slight change in a sequence of images, presented within a time frame and giving an illusion of connected movement.

[1] For a more detailed description with statistic research data, see: Lauc Tomislava., Matic Sanja., Mikelic Nives. Educational multimedia software for English language vocabulary. // *Current Research in Information Sciences and Technologies- Multidisciplinary approaches to global information systems: VOLUME I* / Guerrero-Bote. Vicente P. (ed.). Merida: Open Institute of Knowledge, 2006, 117-121

Conclusion

The multimedia educational software for English language vocabulary has been developed since the Academic Year 2005-2006.

In Academic Year 2006-2007, we have expanded the software by adding features of multimedia collocations. This new software feature is of high importance, since the animation of both lexical segments of a collocation gives an impression of connectivity. This approach creates an illusion of an incidental, spontaneous and not stressful learning environment, engaging the learners on a latently obligatory level.

The results of our project so far show that the interactive multimedia environment for learning proves to be useful from several aspects; it influences the learners' cognitive abilities by following the natural way of human thinking, invokes the learners' interest in learning the content by drawing and maintaining their attention and provides significantly better results. The higher the motivation for learning the learners have, the better the results they obtain. Furthermore, applicability of computer-assisted language learning in a form of an interactive multimedia interface saves time and energy. Animation can indeed be used in language instruction for more self-reliant learning. Interactivity of multimedia exercises engages the learners to commit themselves attentively to their work and therefore actively participate in the learning process.

Finally, our next step in this project will be the research on use of animation in educational settings, followed by the development of multimedia features for learning language tenses and prepositions.

References

1. Anitha Devi, V. Using animation for teaching phrasal verbs. 8 August 2005. http://www.languageinindia.com/aug2005/animationanitha2.html (25th September, 2007)
2. Janulevieiene Violeta; Kavaliauskiene, Galina. Using the Lexical Approach for the Acquisition of ESP Vocabulary. March 2001. http://iteslj.org/Articles/Kavaliauskiene-LA.html (25th September, 2007)
3. Lauc Tomislava., Matic Sanja., Mikelic Nives. Educational multimedia software for English language vocabulary. // *Current Research in Information Sciences and Technologies- Multidisciplinary approaches to global information systems: VOLUME I* / Guerrero-Bote. Vicente P. (ed.). Merida: Open Institute of Knowledge, 2006, 117-121
4. Lewis, Micheal. The Lexical Approach: The State of ELT and a Way Forward. Hove, UK: Language Teaching Publications, 1993
5. Paivio, Allan. Mental representations: A dual-coding approach. Oxford University Press, USA, September 13, 1990
6. Sundberg, Paul. Animation in CALL: Learning to think in the fourth dimension. // *CALICO'98 Symposium*. San Diego University**,** California, 1998. http://lrs.ed.uiuc.edu/students/psundb/homepage/animCALL.html (25th September, 2007)
7. Turrini, G., Paccosi, A., & Cignoni, L. 2003. GLOBETROTTER: Words with their travelbags. Educational Technology & Society, 6(3), pp.9-23. Available at http://ifets.ieee.org/periodical/6_3/3.html (ISSN 1436-4522)

Service Learning in Information Science: Web for the Blind

Bojan Basrak
Department of Information Sciences
Faculty of Humanities and Social Sciences
Ivana Lucica 3, Zagreb, Croatia
bbasrak@ffzg.hr

Sanja Matić
Department of Information Sciences
Faculty of Humanities and Social Sciences
Ivana Lucica 3, Zagreb, Croatia
smatic@ffzg.hr

Nives Mikelić Preradović
Department of Information Sciences
Faculty of Humanities and Social Sciences
Ivana Lucica 3, Zagreb, Croatia
nmikelic@ffzg.hr

Summary

The biggest problem to blind and visually impaired people poses the integration in various aspects of the social life. Since computer technology with its dominantly visual interface increasingly becomes the prime media for communication, these people are faced with a new obstacle – the ability to work on a computer and use it to communicate with others. Technologies that enable the blind to use the computer do exist, both hardware and software, but they have difficulty learning to use them. The reason for this is lack of standardized educational program so that computer education of blind is usually left to themselves and a few associations that do not have sufficiently qualified staff for such a task. Our service learning project offers the first step to possible solution of the current situation.

Service learning is a methodology where learning occurs when students apply what they learn to community problems. It was introduced into information science curriculum in the academic year 2006/07 for the first time in Croatian University with the goal to transform the old teaching style, give the final year students more place to apply the theory they learned through the study and, most important, to provide an important service to the local community.

"Web for the Blind" is the service learning project that provides a first step towards an optimal solution. The main goal of this student project was to establish an adequate model on which all future projects in this field could be based. At the moment "Web for the Blind" is a stand-alone application that provides the blind and visually impaired with the guides on how to install a web browser, familiarizes them with its interface and gives them some basic information on surfing the web. The application can be used without a mouse and with or without screen-reader software because it contains the tutorials in textual and audio format. The next phase of the project involves intensive testing in the field. It is to be presented to as much users as possible for the evaluation purpose. Finally, the contents of the tutorials need to be gradually updated in both quality and quantity.

Key words: Service learning, Educational change, Public service, Software for blind and visually impaired

Introduction

In this paper we show the results of the implementation of the new educational methodology that improves student learning productivity by linking theory and practice, more precisely, by integrating theory and social needs.

Service learning (SL) is the most common example of experiential education on university [6]. The reason for this is that it is distinguished from other approaches by its intention to equally benefit the provider and the recipient of the service as well as to ensure equal focus on both the service being provided and the learning that is occurring. Internships/practicum, on the other hand, engage students in service activities primarily for the purpose of providing students with hands-on experiences that enhance their learning or understanding of issues related to a particular area of study.

Although practicum is quite common at our university, they are also quite disliked by our students. Most of them do practicum in local government organizations, private enterprise or scientific research centres, but they say they do not benefit from it because their supervisors do not let them to employ their skills effectively, they are only given simple and boring tasks that are beneath their capabilities.

SL was never integrated as a part of the study curriculum in a Croatian university. Since practicum rarely proved to be a powerful experience for our students, in academic year 2006/07 we decided to engage our students in service learning projects that are not only enhancing their learning in particular area of study, but also providing service which addresses a community need.

Integration of service learning in information science curriculum

In the field of Information Science (IS), we integrated service learning with ICT in 12 student projects. In the beginning of the academic year 2006/07 in the De-

partment of Information Sciences, at the Faculty of Humanities, University of Zagreb [5], we included student participation in a non-profit community organization or in public primary school. We had 64 students involved in the SL activity. They started work on their projects in November 2006 and they finished it in mid June 2007.

In the first part of SL activity, our students and community members have worked together to identify community needs and to plan service activities. They also worked together to evaluate the impact of the project at the end of the academic year. There was also a control group of 6 students who performed interviews with the community partners and with each group of students and who controlled the project activities. This provided opportunities for students to voice concerns, share feelings and evaluate the project.

The following is the description of the best service learning project: "*Web for the Blind*".

In the first step of the project students contacted the association for the blind and visually impaired people "*Sismis*" [7] and offered their knowledge and expertise in the field of information science. Alliance of blind and visually impaired students "*Sismis*" is the first university institution that gathers blind and visually impaired students. The goal of the alliance is to provide easy integration for blind and visually impaired students into student and academic community, to provide them with the course material, to enhance the quality of their education and to provide the information technology support.

Talking about the members' needs it turned out that many do not have basic computer skills, and most of those who use computers, use only simple applications. Trying to change that, students decided (in cooperation with the association) to design electronic user manuals that would help the blind and visually impaired to use the Internet and adjust all software applications for the simplest usage as possible. Students realized that the "*Web for the Blind*" user manual would especially motivate those members who reluctantly approach advanced computer usage because the guide would reveal new ways of communication and would offer an approach to entertainment and useful information. The initial plan was to design only a written guide explaining how to install a web browser, how to use and adjust its interface and to give some examples of browsing through web pages.

The blind use computers by the help of a screen reader. The most popular screen reader in Croatia is JAWS, designed by the *Freedom Scientific* company. It is a high-quality and extensive tool transforming all the screen contents into descriptions read by a speech synthesizer incorporated in the JAWS as a plug-in. Speech synthesizer technology has improved on a large scale and the blind are accustomed to listening a digital voice. However, the associates from the "*Sismis*" association emphasized that it would be more acceptable and a lot better for the blind if students could record their guides with a natural voice as

digital audio recordings. Although students did not have a professional narrator, nor professional equipment for studio sound recording, they decided to do the best they could. Furthermore, they abandoned the idea of designing a CD-ROM with textual and audio guides, since it would be difficult for the blind to search and navigate through the structure and numerous files. Instead, they decided to design an application serving as an interface that will allow users to easily approach the desired content through simple navigation.
Since their ideas developed further, their project became larger and larger, and they found necessary to divide it into sections that could be approached separately and solved individually so that the work would be done as fast and efficient as possible. Therefore, students divided the project into four main parts:

1. Definition of goals, communication with users and adjustment to their needs
2. Design of user manuals
3. Sound recording and processing
4. Application design for navigation through content

Software description

Prior to designing the user manuals, it was necessary that students, its designers, become familiar with the way the blind use computers. It turned out to be quite demanding because of the extremely visual orientation of the computer interface and the importance of mouse navigation. The blind use the keyboard (with the shortcuts) and the sound signal from the screen reader as the only means of orientation. Realizing how demanding the first step was, students decided to supplement the user manuals with a list of keyboard shortcuts for easier web browsing. The user manuals are adjusted to the users with poor computer experience in order to encompass the widest spectrum of user needs. In the user manuals, there is a step-by-step description of the screen content, screen options, and guides on how to adjust them. Although students planned to design three separate user manuals, only the first two have been completed: Mozilla Firefox web browser installation user manual and manual for getting to know the browser’s interface and its’ adjustment. The last would include examples and instructions for surfing through web pages, but is not finished yet. Students chose Mozilla Firefox because, according to many studies, it proved to be the most secure one, and because most of the blind using the Internet, use this web browser. Mozilla web browser has screen readers that read aloud all available information in applications and documents or show the information on a Braille display, enabling blind and visually impaired users to use equivalent software functionality as their sighted peers.
Students’ goal was to design the user manuals as detailed as possible, so that the user could have a complete insight into the screen, but at the same time to keep it simple, to reduce possible mistakes. According to the first feedback information, the users are very satisfied with this segment of the project.

The next stage was sound recording of the adjusted user versions of user manuals. They are very detailed in written form, and therefore it was necessary to reshape them in order to make the reading and the listening suggestive and meaningful. Since a professional narrator was not at their disposal, a member from student team did the demanding reading part, while the recording was done at home with minimal equipment (dynamic microphone, preamplifier/mixer and sound card). Having the output signal quite rough, sound editing was necessary. Students used the amateur audio editing tool *Adobe Audition* [1] with several supplements. Of course, they could not match the quality of audio books from the *Croatian library for the blind* [4] to which the users are accustomed to, but students think they did the maximum regarding the possibilities. The users appreciate their work and they are grateful for the achieved sound clearness and tone definition that our students are very proud of.

The last phase of the project was the design of the interface the blind would use to find and open all the existing files.

There were numerous possibilities, e.g. web pages based on HTML or Java or full applications that need to be installed on the computer. Students chose the simplest solution that enables flexibility and further updating – an independent autorun application. This made possible the whole project to be formed as a CD ROM that automatically runs as soon as put in the computer. Furthermore, users do not need to have any contact with slightly confusing file and data structure, but instead they approach all the data directly through the application. The application consists of three manuals. In the first, initial one, there are buttons for approaching Mozilla Firefox installation, audio guides, written guides, list of keyboard shortcuts with some additional information, and finally, the button to exit the application or to return to the main manual. All the buttons are linked to the keyboard keys from one to five, but it is also possible to use the mouse.

In order to ensure the complete application autonomy, each manual automatically runs the recorded speech, which lists the offered options to the user indicating which keys to use to approach them. On a button press, the speech is being stopped, and as an additional option, each manual contains a pause button also activated with the "P" key.

The application is made with the program package *Autoplay media studio* [2] designed by the *Indigo Rose Corporation* company. It is a relatively simple tool for designing similar standalone autorun applications the foundation of which is "Lua" script language. Therefore, the user only inserts the CD ROM containing the application in the computer and the automatically opened manual directs the user by audio instructions. Students think that such a simple approach is very intuitive, user-friendly and motivating for all the blind and visually impaired users.

Implementation and research

The first version is completed and is being tested by the clients from "*Sismis*" association. Students recognized the necessity to evaluate the efficiency of the materials prepared during the academic year 2006/07 and decided to do the software update, when needed. They also intend to further adjust the application to the screen reader so that the users can select the way to use it. Also, in cooperation with other associates, students will compose the guide containing examples of navigation through the web pages that are most necessary for the users (web page of the *Croatian Association of the Blind* [3], web pages of different faculties in *University of Zagreb*, etc). Depending on the interest of the clients and associations dealing with the blind and visually impaired, this project could have a bright future and students hope that it will serve many people.

Conclusion

Considering the fact that we live in a world dominated by values of market competition, economic integration and rapidly developing technology, a world that in the same time has growing requests for promotion of national interests and culture identity, we can see that the new reality imposes an entirely new role on the education sector. Our students definitely need some skills, apart from knowledge.

When students make the connections between their service activities and studies, it deepens their understanding of the curricular material, how it's used, and why it's important. In order to truly understand the impact of the service, students need to observe the impact of the project on different participants. In the process of observing the impact of the project, they recognize the significance of their experience and assess their own learning and the impact of the project on the community being served.

Students should do service learning during their study, since it enables them to move from theory to practice, preparing them for a lifetime of learning, service, and civic engagement. It assists faculty in the transformation of student learning outcomes by strengthening course curriculum and pedagogy through experiential learning. The quality of the service learning activity is that after everything is finished, participants continue to use their new knowledge and skills to make decisions, solve problems, and grow as caring, contributing members of their communities.

Therefore, we believe it can become invaluable academic tool for Croatian students, faculty members and community in general.

References

1. Adobe Audition 2.0 website. http://www.adobe.com/products/audition/(25th September, 2007)
2. AutoPlay Media Studio 6.0 –Visual Multimedia Software Application Development website. http://www.autoplaystudio.com/site/index.php (25th September, 2007)
3. Croatian Association of the Blind website. http://www.savez-slijepih.hr/(25th September, 2007)
4. Croatian Library for the Blind website. http://www.hkzasl.hr/ (25th September, 2007)
5. Department of Information Sciences, Faculty of Humanities and Social Sciences in Zagreb website.
6. http://www.ffzg.hr/infoz/web2/index.php?option=content&task=view&id=363 (25th September, 2007)
7. National Service-Learning and Assessment Study Group. Service-learning and Assessment: A Field Guide for Teachers. Washington, D.C.: Learn and Serve America, Chapter 2, 1999.
8. Sismis Association of the Blind website http://www.udrugasismis.hr/(25th September, 2007)

Electronic Content in Education – "Media Culture" With(out) Media?

Marina Čizmić Horvat
Department of Information Sciences
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
mfruk@ffzg.hr

Predrag Gjuro Kladarić
IT Services
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
gjuro@ffzg.hr

Summary
"Media culture" is a course intended for the students of Information Sciences. It includes two different types of knowledge – knowledge of the subject itself and knowledge of one's personal experience with media. The 21st century is changing the learning process in its essence. The traditional classroom teaching is now replaced by the new way of learning - which means developing competences and collecting information and knowledge concerning current problems. This kind of learning is more likely actualization of knowledge from different data bases, than learning for future, or collecting knowledge for possible cognitive challenges. This way the teachers may become some kind of guides and mentors along the learning path. The paper illustrates our way of combining those two types of knowledge. We present the results of our recent research about the use of electronic content in media learning today.
Key words: media, media culture, electronic content, media learning

1. Introduction

Electronic contents in the study of Information Sciences at the faculty have successfully been used for a long time as an essential part of the teaching process. Students and teachers can use the computers and the internet at the faculty. Evaluation of e-learning systems started in 2002. Moodle was chosen and has been in use since 2005 with great success. Computer logs for "Media Culture" course has been analysed in search of students learning patterns. Every teacher collects his/her own experience during a course which (s)he rarely or never exchanges with other fellow teachers. This conference is a good opportunity to hear the experiences from our colleagues in Croatia and from other countries.

2. Why “Media culture”?

The implementation of the Bologna process started in the academic year 2005/2006, so we started a new elective course named “Media culture” for the first year students of Information Sciences and for the other interested students. Why “Media culture”? The media are in the focus of our life in the new millennium. Our students live in a rapidly evolving media environment. That course should provide students with the skills and understanding necessary for their media future. How to do that? We combine training in critical thinking with instruction in how the media function. The students should find out what happens in the "real media world"! Some graduates will find work in the media and cultural industries, in the world of business, the government, education etc. They will take up jobs that don't even have their names yet.

“Media culture” as the course was organized in total of fifteen meetings, we met once a week, one period of lecture and one of seminar. We should emphasize, that there was no strict boundary between the two. We prefer learning in an active way, because it is more effective than a passive receiving of information. This kind of learning and practicing competence requires a new form of didactic, which is not a scholastic one, meaning acquiring the necessary knowledge and its permanent improving, but a didactic of accessibility, which allows a fast integration into the current working process and the exact achieving of the desired aim. The traditional classroom teaching is now replaced by the new way of learning, which means developing competences and collecting information and knowledge concerning current problems. This kind of learning is more an actualization of knowledge from different data bases, than learning for future, or collecting knowledge for possible cognitive challenges. This kind of teaching demands new specific conditions: limited number of students in a group, large amounts of lessons and the appropriate electronic environment.

3. Electronic learning environment at the faculty

3.1. E-learning on Faculty of Humanities and Social Sciences

History of e-learning on the Faculty of Humanities and Social Sciences of University of Zagreb starts in 2002. with a project of Croatian Ministry of Science, Education and Sport and Department of Information Sciences aimed at investigation and selection of the most appropriate course management system (CMS) among ones available (LearnLoop, ZOPE, WebCT, Blackboard, Moodle). Moodle had been chosen for its set of features and availability and has been installed, customized and translated to Croatian for the use in academic year 2004/2005 under the name Omega (http://omega.ffzg.hr). Acceptance among teachers gradually increased so on August 2007 more than 200 professors from 23 Faculty departments had prepared and used 220+ courses (most numerous among departments being the Department of Information Sciences with 69 courses, followed by Department of English with 22 courses, Department of

History with 20 courses, Department of Sociology with 19 courses and Department of Croatian Language with 15 courses). Total of 3300 students are being enrolled into courses in August 2007.

3.2. "Media Culture" course in numbers

The teacher has chosen to publish a sketch of the media characteristics and let students learn from investigation of suggested sources, preparation of public presentation and discussions with fellow students during the presentations. Students were obliged to upload their presentations to the Omega system, where both teacher and other students can find and use them. The Omega system logs all the student and teacher activities, and some of these are shown here in an attempt to illustrate patterns of use of the system. There were 3,913 logged activities on the course (from enrolling and un enrolling to the course to viewing the course description to enrolled student list to downloading course description and fellow student presentations).

Submission of student presentations took part from December 2006 to February 2007 but students continued to (rarely) visit the course area throughout the academic year. Teacher and student activities form a recognizable pattern regarding an hour of a day: a steep rise in the morning hours with peak just before the lunch time, continuing after the lunch break from 5PM and slightly declining till 1AM (Chart 1).

Of total number of 3,913 activities recorded in log files, 823 took place from computers installed on the Faculty, 1,410 from the CARNet network (with most of them probably from the students dormitories in Zagreb, which are connected directly to the CARNet network) and 1,680 from other places on Internet (Chart 2).

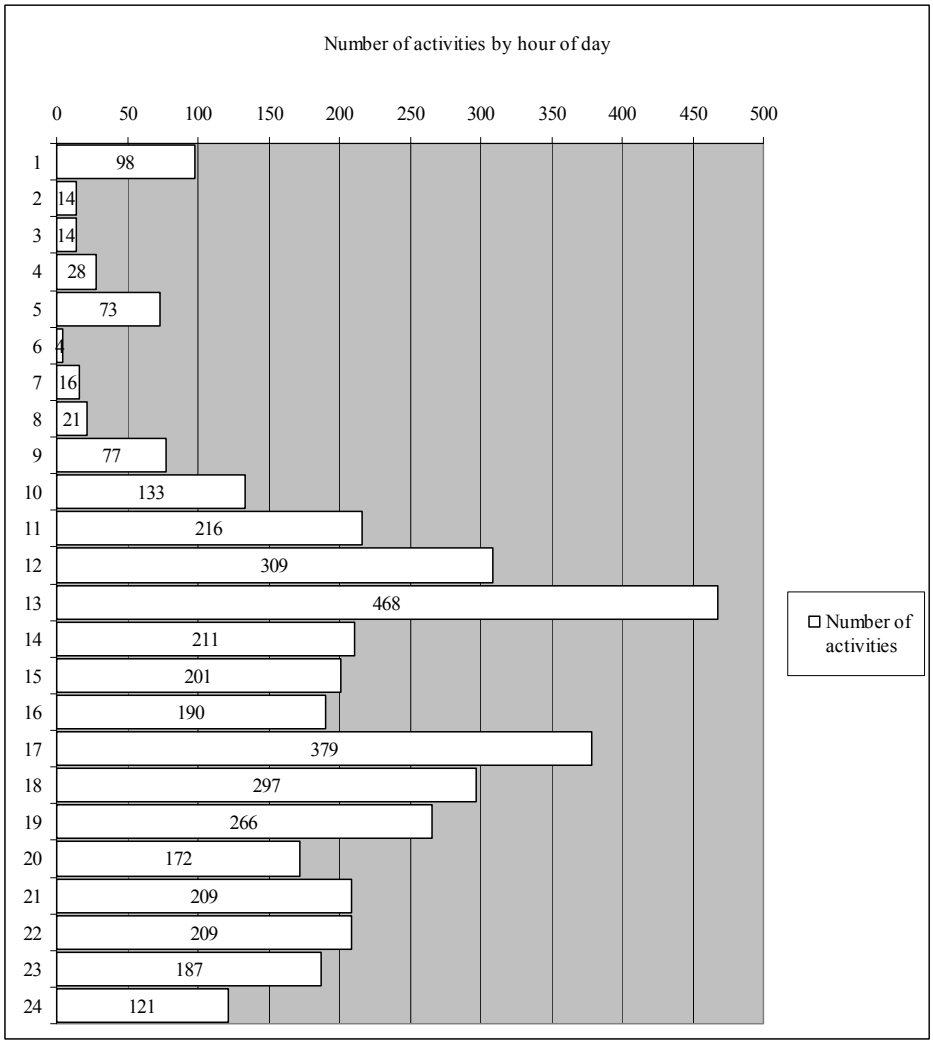
It is interesting to note that attractive theme and lively discussion in the classroom result in an increased viewing rate of some student presentations stored on Omega. The most viewed was Anime, presentation on Japanese animation, followed by Ads, Film and Media role in the consumer sociology and other. The following chart shows number of resource view activities by resource viewed (the first item is the teacher's course description) (Chart 3).

3.3. Future

Faculty of Humanities and Social Sciences does not have enough space to provide students with quiet place to use the Omega system and do the research of course resources. That is expected to change with the new library building where some 100 new publicly accessible computerized workplaces are planned. Future expected enlargement of server capacities and Internet connection speeds will enable both teachers and students to produce more interactive material (which may use much more space on the server). However, there is a need for a support team which would not only support the Omega server but work on im-

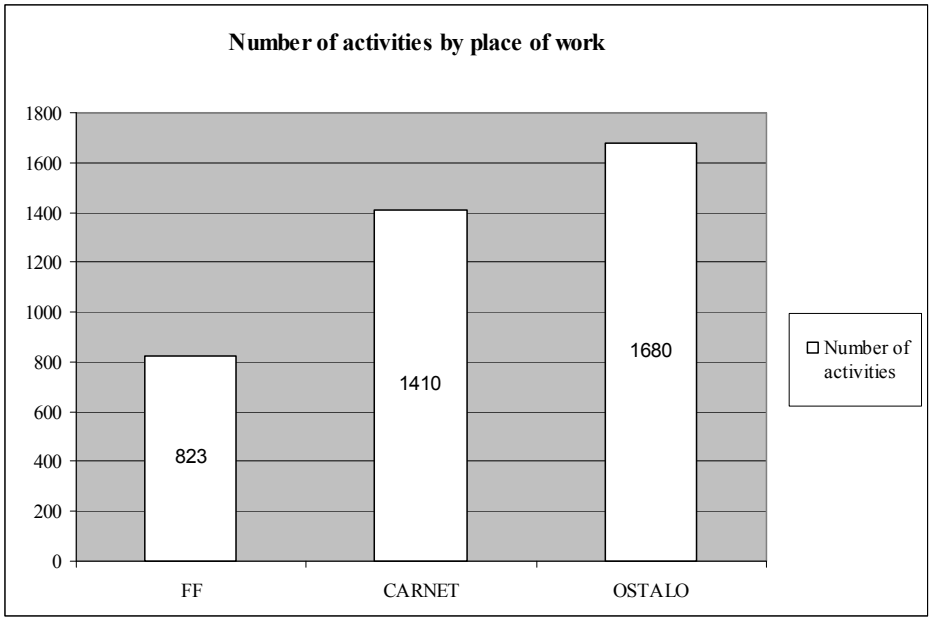
provement of teacher’s use of ICT and assist teachers in production of courses, possibly of more complex and/or multimedia type.

Chart 1. Number of activities by hour of day



Source: computer logs from Omega system for “Media culture” course

Chart 2. Number of activities by place of work



Source: computer logs from Omega system for "Media culture" course

4. Learning in an active mode

We shall present here only one part of the experience that we had doing in the academic year 2006/2007. Our intention is to show, which themes the students chose, which multimedia content they often use and the problems they encountered.

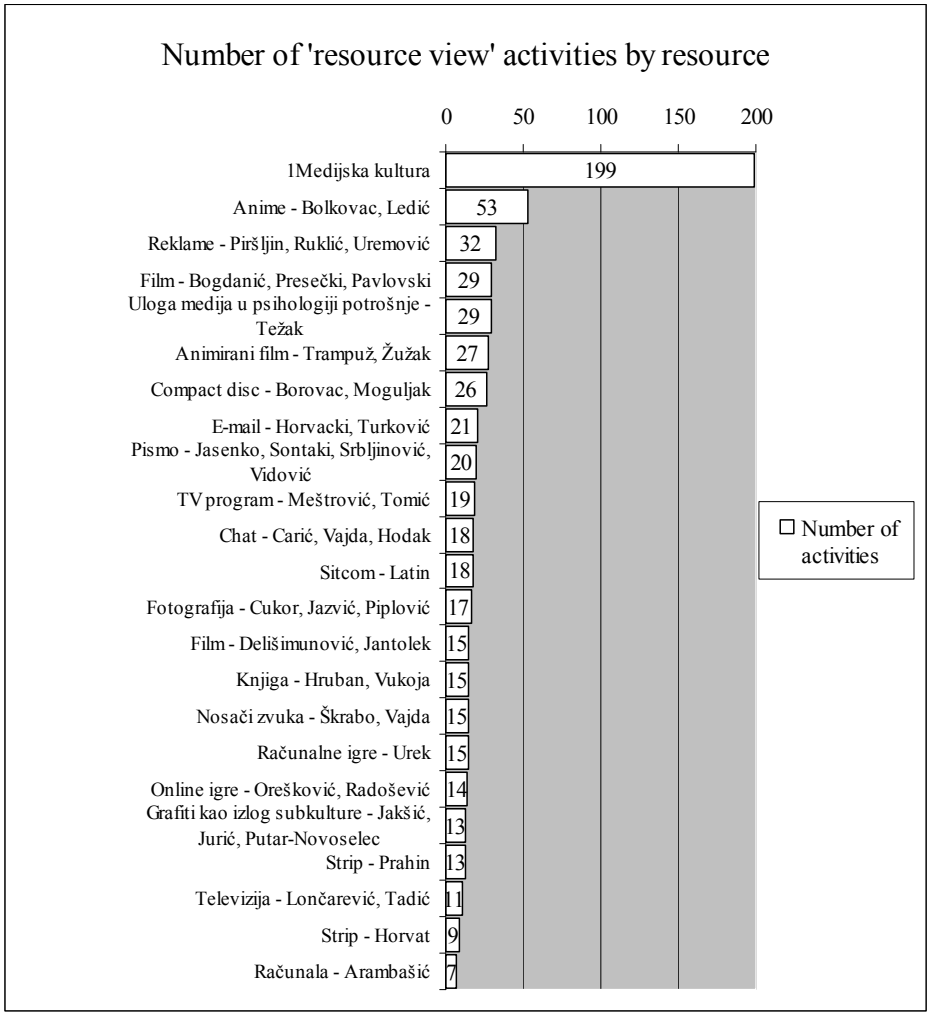
The course "Media culture" includes two different types of knowledge – the knowledge of the subject itself focusing on traditional and digital media and the knowledge of one's personal experience with the media. The teacher will make students more sensitive to the media, they should become aware of the media and develop critical appreciation of the media position in their life today. We prefer an authentic learning method and the learning in an active way. So teachers become guides and mentors along the learning path instead of mere instructors.

4.1. Themes

In the framework of the course students, or a group consisting of three to five students, chose as a theme one kind of media, traditional or digital, of their own choice, and their task was to write a seminar paper or to hold a presentation in front of the class, using some of the popular programmes for creating a multi-

media content - audio, video, text, graphics, animation[1] – for example Microsoft PowerPoint, Macromedia Flash etc. All the students agree they could not imagine their life without the media. The students strongly condemned all kinds of media misuse. They spoke in favour of the need for media education since early childhood.

Chart 3. Number of ‘resource view’ activities by resource



Source: computer logs from Omega system for “Media culture” course

[1] The presentations displayed in Omega don't give a real picture they had to be shortened due to space restriction.

4.2. Electronic content

Most of the students chose the presentation. The presenters and the class show great interest in such presentations and the discussions are sometimes quite heated and full of controversial opinions. The students chose most often one of tertiary or one of digital media[2], which they wished to present. The most popular media-themes are CDs and computers (electronic mail, games). To present these media they used usually audio, text and animation. They presented not only the media evolution and its current usage but also their personal positive and negative experience in using them.

4.3. Problems

The students varied in their competence in using computer technology. They ranged from real beginners to cyber kids, total computer addicts. Because of that the problems which arose while preparing the presentation were very different. Some students complained about the impossibility to do the presentation at the faculty, while others complained about the modest i.e. insufficient presentation possibilities in the classroom. Many were unhappy because they had to give up some great solutions which they had created on their computers at home. Different levels of technological equipment sometimes even caused unpleasant situations when some groups felt being discriminated against. The reasons are: the modest electronic environment at the faculty, too many students in a group and insufficient number of lessons.

5. Conclusion

The online availability of learning resources is much better now with the Omega system than several years ago. Fully functional Omega system allows the encouragement of production of digital learning material and e-learning courses in a standard way which is situated on a standard place and usable in a standard way. This definitely eases the efforts of student learning for both students and teachers.

Our experience has shown that electronic contents in the “Media culture” course are used mostly in two ways. One way is using the Omega system, which is used for communication and the other is the individual work of our students outside the system. Electronic content is not sufficiently present although it plays an important and indispensable part in the process of higher education. We have a dream: smaller groups, better conditions concerning the equipment, reciprocal help, active anticipation in occasional projects at the faculty and “Media culture” with media.

[2] There are numerous media classifications. Here were use Faulstich's classification.

References

Computer Logs for the Media culture course of Omega system. http://omega.ffzg.hr (29.8.2007)

Faulstich, Werner. Medienwissenschaft. Paderborn : Wilhelm Fink Verlag, 2004

Lasic-Lazic, Jadranka; Banek Zorica, Mihaela; Spiranec, Sonja; Klindzic, Jasmin. Using Open source Learning Management System for educating information professionals // *Current Developments in Technology-Assisted Education* / Sevilla, Spain : FORMATEX, 2006, 88-92.

Stangl, Werner. Arbeitsblaetter. http://paedpsych.jk.uni-linz.ac.at/arbeitsblaetter.html (20.8.2007)

Programs of Teaching Information Literacy in Croatian School Libraries

Jasmina Lovrinčević
Arhitectural-geodethic school
Dravska 16a, Osijek, Croatia
jlovrincevic@net.hr

Dinka Kovačević
OŠ "Antun Mihanović" Slavonski Brod
Mihanovićeva 35, Slavonski Brod, Croatia
dinka.kovacevic@sb.htnet.hr

Summary

This paper talks about the beginnings, development and currently used models of teaching information literacy in school libraries.
The foundations of information literacy in Croatia have been set in early eighties of the 20th century by the general transformation of classic school libraries into multimedia centers. The program orientation of activities was focused on, in those days, innovative forms of learning with a clear focus on individual choice of methods and the most appropriate media. Teaching information science has developed with individually developed programs, that came out of the users needs and the models have been executed through group teaching.
In middle 90's of the 20th century, the merge of new technologies and the change of users needs, result with a completely new model of teaching information science, which actually makes the base of the educational program of school libraries and school librarians.

Key words: school library, information literacy, programs, pupils, learning

Introduction

The strategy of development in Croatia called "Croatia in the 21st century" (2000), especially in the first part of this project goal, speaks about education which starting foundation is the concept of lifelong learning (that way it systemizes 3 major forms of education: formal education, informal education of adults and self education – the concept of a "society that learns"). These two complementary concepts are recommend, as the basis for leading a modern educational policy and the reform of education systems, by UNESCO, OECD, ILO and other international organizations that are concerned about education

and are aware that there is an increasing need for individual, independent and technologically supported learning.
Today, in the Croatian educational system, the school library has a law debiting place[1], professionally verified work programs, highly educated school librarian[2] with in many cases the widest and the longest education from all of the employees in their school. Its work is based on the Standard for school libraries functioning (2000), and it is financed by the states Ministry of education and sport. With this clear attributes Croatia has insured a right on information and additional possibilities of learning in school, for all students and all places under equal conditions.
The libraries activities are the backbone of the school's curriculum, and often its creators. Librarians are, for number of years, offered regular professional upgrading on several levels, organized by the Ministry and The Agency for education. At this moment Croatian school libraries are the strongest potential for the change of educational system, although they are still inadequately recognized.

1. The beginnings of information literacy teaching

The school library in its traditional sense (a rich collection and borrowing function) exists in Croatia since 1982. But, the year 1982 was, for Croatian librarianship, a turning point in a sense of a complete transformation of the school library into a center of learning on information sources. The transformation was guided as a extensive republic project which offered, besides the idea, and a clearly set concept that was executed through several fazes. The education of library staff was also carefully organized.
The need for user-student education on the field of working and the usage of the school library came for several reasons. Human knowledge is wider, and the rapid development of the technology for information processing in libraries is demanding an educated user-student. The library holdings have also changed. As information carriers, besides the book fund, the number of periodic publications for students, teachers and parents, has risen. With the entry of film, video, educational programs, CD's, DVD's, computer programs, data bases and the Internet, the perfection of education technology has advanced, but it also imposed the need for education in these new information sources.
Today, we can say that that happened because of the stagnation of tuition and the simultaneous need for changes in the educational process, with which easy going, but sure, changes within the library were caused. The awareness of the

[1] In the law of schooling, it is distinctively stated that "Every school has a library".

[2] School librarians in Croatia are educated on a diploma level on Philosophical colleges, but the majority of have a diploma as a two-subject teacher with a additionally finished diploma study of librarianship.

need for creating a strong backup for tuition and general education took shape in the 80's of the past century, and the school library ‚definitely got a new place and role inside a school. Already then it got an information function[3].The suggested terminology is "Library-information center" or "Multimedia center" which was equally good, because it contained the additional massage that the school library wants to expand on different Medias. The basic idea was learning through collecting different information from various media. That new library could take over the role of cooperators in the education of their users, teachers and students.

The first, basic forms of learning in the modern school library were:

- Individual learning
- Learning in small groups
- Pair learning

Although, in pedagogy already known forms of working, most commonly used in tuition, those forms are now promoted in the school library. It was a time of intensive development of the school library as a library-information center. The tendency was to leave the traditional activities and move forward towards a new goal of active participation in the education process.

In this new library, learning in a different way is in first place.

The starting point was a new environment, for learning that is based on the program of education of students for individual work on information and knowledge sources in the school and other kinds of libraries in their later education. For those reasons it was thought that education in twelve steps should be executed from the first grade of primary school to the fourth grade of high school. Table 1 shows the basic guidelines of this Program.

Learning from sources of information and knowledge was in the first place. Students become active in the educational process at the moment when they use a large number of various materials I order to explore the subject material proscribed in their class curriculum. Librarians have become animators and leaders in the process of learning end they secured the firs class motivation that leads student towards information search and creative problem solving. The final result was a animating of "a learning culture" as an environment of all activities; especially productive learning that is animated in the whole school.

The independence increases when students can recognize what they know about the topic and what else they should know, where to search for the information, how to mark it as a massage, how to evaluate and how to determine their success in the process of information research. The first successfully finished information goal encourages students, animates further learning, information research and develops their self-confidence.

[3] Quote from the book "School library-a step further" (2004) authors D.Kovačević, J. Lasić-Lazić and J. Lovrinčević

Table 1. Basic guidelines

ORGANIZING OF LIBRARY FUNCTIONING (space, library holdings, working hours, working plan-program, information and knowledge sources usage, lending, library holdings protection)	SCHOOL LIBRARY GOALS -educational goal -library-informational and -documentation goal -cultural and public goal
THE PLACEMENT OF LIBRARY HOLDINGS -According to the age of students (from 1st to 8th grade – student holdings) -According to the universal and decimal classification –UDK (teachers library holdings)	INFORMATION AND KNOWLEDGE SOURCES -books -periodicals -AV material -Electronic information sources -catalogs -bibliographies -the referential collection
USER EDUCATION -text papers (written) -notes -term papers -seminar -reading journal -final paper compilation	DATA BASE SEARCH -The renewing of basic information (computer) literacy knowledge (needed in librarianship) -Defining instructions -Skills needed for the search -Skills needed for monitoring the constant changes in the field of research.

2. Information literacy tuition guidelines

The beginnings of learning in the school library are connected to individual learning for classes. The most frequent was making of a seminar, term paper, but also solving a problem produced in a class, in the "classroom". Individually a student studies what he missed, something that he didn't understand, and what he wants to extend or add to his knowledge.

Pair learning (learning in pairs) or a small group is usually in relation to a bigger and more complex seminar. This type of learning has developed usage of various media. That animation came from the school librarian, with a purpose of including teachers also. With a teacher's demand and in cooperation with the librarian, usage of several media could have been inveterate. The next step in promoting the quality of learning was a suggestion and inserting written seminar work through exploring, insisting on a more creative, imaginative work, which had a goal to show the way how did they get to the information.

So, learning from various media (books, pictures, audio, video, class films…) was actually a process of research learning. The result is a written paper, in a similar way prepared presentation[4], also creative, multi-medial.

[4] ibidem

We can see that this form of the school librarian's work with students, (individually, in pairs, in small groups) in base consists of tuition and information literacy skill development. It was very quickly recognized that the librarian's efforts on this forms of learning are easier if the student are initially shown the ways of using the school library resources.
In early 80's of the 20th century an obligation group tutoring of all students was inserted in high school libraries to educate them in various usage possibilities of the school library like:

- Tuition for library and material usage
- Introducing students in seminar making
- Introducing students in final paper making

Through tuition in library usage students were informed in the library materials, resources and the choice of activities. They were taught how knowledge sources were organized from catalogs to shelf-placing, and also in what order knowledge sources are used. The tuition on seminars was used as a general preparation for seminar making, planning work and what phases of work to expect. Tuition on final paper making is a specific tuition for making a distinctive kind or form of final paper, like a graphic paper, film or models. In that phase at the end of education information literate students are taught.
These educational situations are used in further tuition in:

- Individual work
- Small groups work
- Pair work

As the students have basic knowledge, the librarian leads them further through research learning, problem solving, animating creative personal solutions and various media usage. During the making of seminars, the librarian takes over the obligation of tuition and presenting the work (paper).
Learning on assignment in pairs is easily used for "learning of learning" although these skills are actually taught through all forms of work in a library. At the end of the 80′s of the 20th century, working in small groups refers on school project work, which is literally and only affirmed through the work of the school library. It is in order with the ALA's (American Library Association, 1989) presidential committee on information literacy, concise in the form of this manifesto:
"Informational literate people are those who are learning how to learn. They know how to learn because they know haw knowledge is organized, how to find information, and how to use it in a way that others can learn from it. They are prepared for lifelong learning, because they always find the information needed for usage."

Although a new phrase "information literacy"[5] in all its definitions and content, involves a logic cycle of methods which are carried out in the school library, where students are guided through various forms and processes of learning.
"Information literacy" at the same time means:

- Ability to recognize the necessary information for problem solving and idea development
- Crucial question placing
- Various information usage
- Locating of relevant and responsible information
- The ability to use practical and conceptual tools and informational skills
- Research process
- Making and publishing in textual and multimedia formats and the adjustmentto technologies which appear

Ross Todd's definition (1998) marks information technology and describes the process of acquiring skill as:

> "The ability of using information; with purpose and effectively. It is an interactive process of learning that encloses defining, locating, selecting, organizing, presenting and evaluating information from sources, including books and other media, experiences and people. That means to be "up to date" with new knowledge, adding them to earlier knowledge. It also means the application of knowledge, adequate and reliable for solving problems."

In all definitions, in "The big 6" (Eisenberg & Berkowitz, 1990) models or the Kuhlthau (2002) model of "Guided research learning", which are newer, we can recognize the same logic and non avoidable element to which we should teach our students. The shown starting ways of teaching information literacy in Croatian schools and libraries, as we have shown, contained the same, still demanded elements.

3. Development concept of tuition programs

We have developed a well made base in high school libraries spreading content towards new users and the development of new technologies that came to our schools. The school library is again, taking a key place adapting computer and information literacy to each other, so they would insert the usage of new media into old schools.
With the arrival of computers as an individual tool that because of it's possibilities, represses other media, and the process of learning is more and more individual. Even tuition of information literacy is now going towards the individual approach. On the other side, new learning affirmes research learning, creative work, learning through school and other projects, which is connected with al-

[5] Ken Heycock speaks amply on the phrase information literacy and also incorporating the program of acquiring information literacy skills.

ready known ways of working of the school library. With that, more opportunities are offered for systematic tuition and gaining information literacy skills. That is the reason that high school libraries decide on informal and different ways of tuition, in all of their functions, always and in every place and with every encounter with a student. Although the tuition of information literacy is in the "program and plan"(similar to a curriculum) of the school library, in high school libraries firmly set class themes are advocated because of the variety of the curriculum. Different from primary and secondary schools, high schools don't have average but above average students, with defined needs and it is not appropriate to firmly set themes on when and how we should train the needed skills. That is why same goals are achieved in all school libraries, but they are adapted to the student, the school and to the curriculum.

The development of information literacy programs in primary school library are based on the Croatian National Education Standard (CNES/HNOS) that has been implemented in all primary schools in 2006/2007 school year. Goals and tasks of education based on CNES are guided on raising the level of the Croatian primary school system in order to increase the competitiveness of the whole society, based on knowledge and usage of knowledge. The education of motivated young people with developed academic, technical and social abilities is set as the final goal. The motivation and competition influence on the result of education. The result is: the ability of a young person to analyze the situation, planning achieving a goal, valuate its plans, achieve the set goal and evaluate the results.

The task of modern primary school library according to CNES is; to be a support to all educational goals and tasks of the tuition "plan and program" of the school, through communication-informational work. The priority is to ensure access to sources of knowledge which will animate students on various ideas in experience learning and ensure them gaining of creative experience in using and creating information. These are strong enough reasons for starting changes in primary school librarianship that resulted with a unique program called "Librarianship education-Animating reading and information literacy". The program has became a vital part of the new tuition "plan and program" for primary schools.

The education of users has all the elements of information literacy with an accent on library usage and information available in the library. This territory involves understanding and usage of information from modern technologies and modern computer communication sources. The goals and themes of information-librarian knowledge's come from basic goals of education.

General Program goals:

- Students adapt to the library space and environment
- They take a positive stand towards the library and it's holdings

- They develop different skills and abilities (communicational, research, informational) by using knowledge sources and information in the school library. They see problems and learn how to make effective strategies in dealing with them.
- Special attention is given to education for books, reading motivation and esthetic experiencing

Special Program goals:

- Students acquaint library holdings and other sources of information in the school library
- Student are acquainted with bibliographic data on library holdings that they can recognize
- Students get acquainted with primary and secondary information sources and know how to use them for their own information
- Students get acquainted with terms; quote, quotation, reference, note and summary for needs of research project work.

Animation of reading:

From the first grade, skills of reading and writing are animated. Priority for activities in working with students in and with the school library is given to animation of reading as a lingual activity and the base of education. Not so long ago, very little was known on "The psychology pf reading", meaning processes of gaining and developing that skill. A new approach in which learning and reading are understood as information processing (opposite to a computer model) has made new data in this process possible.

The educational environment has a basic role in developing learning, because the need for skills and abilities of orientation and selection, in this fast flow of information and new media growth, are good reasons for animating student, even in first grade, to accept work methods and new forms of learning centered on lifelong learning.

Developing reader literacy animates oral and in written expression of students, gives new knowledge's to students, enriches their vocabulary, and helps them in building a value system and accepting ethic norms that are basic preconditions for successful learning of all classes in all levels of learning. With an approach like that, the library becomes an irreplaceable member of the educational team of the school.

Information literacy:

The school library leads students into the world of information in the fourth grade, teaching them in independent usage of information and knowledge. That doesn't mean that it stops with animating reading. Students are tought to individual usage of various information sources, research skills, evaluating and

processing information for individual research needs. Beside the techniques of primary source research, at this point, they are taught the classic and on-line techniques of secondary source research (book catalogues, registries and bibliographies). The modern information technology, here and in other countries, enables successful library fund research. In order for students to be successful in that, they need knowledge's on data saving and research, and these knowledge's are acquired with tuition through this program. These knowledge's and skills that students acquire from tuition in the school library are basic for successful independent and lifelong learning.

An effective program of tuition in school libraries is based on the best research in practice in a particular area. In that way the program:

- Shows how cooperative tuition, points out information literacy, backs up active and independent learning in schools
- Involves students in learning that will help them in their development of lifelong student that are searching for knowledge outside the formal curriculum
- Shapes a live, cooperative culture of learning and promotes complete action of the whole society

Conclusion

The approach to knowledge and the multicultural wealth of the world, just like lifelong learning and literacy, have become the main goal of our information society and the knowledge society.

According to these standards, the goal of the school library changes:

- To qualify all students for critical thinking
- Enable all students to become active users of information and a part of a wider librarian and information network
- To offer various sources of knowledge and quality services to all users

The orientation of the modern school on research and problem tuition, on team cooperative learning, on larger independence of students in acquiring of knowledge and on the transformation of students from an object to a subject of the educational process, gives the school library a role of a core of information and ability for learning and progress. The school library is available for students in time of their most intensive acquiring of knowledge and learning, developing attitudes and behavior, important for their lives. With its effect, the school library is not only a support for the education; it also opens doors of personal creative development of each student that develops a constant need for lifelong learning.

That is why the basic goal of modern school, together with the school library, is to animate the need for written sources and other forms of library holdings, to develop reading skills and habits and basics of information literacy and informational skills that are the predisposition for effective information research.

With emphasizing of basic school library goals that follows the student from his first steps into information literacy and maturing of students-active participator of the new Croatian school, it's unavoidable role in preparing the student for life and work in the 21st century.

References

Haycock, Ken. Students Information Literacy Needs: Competencies for Teacher- Librarians in the 21st. Century IASL Conference, Proceedings, 1997. http://gataway2.uvic.ca/symposium/laycock.html

Kovačević, D., Lasić-Lazić, J., Lovrinčević, J Školska knjižnica – korak dalje Zagreb: Philosophical College, Institute for information studies, section for information studies: Altagama, 2004, page 174

Kuhlthau, Carol Longitudinal Study Methods in Information Seeking and Retrieval, Libraries in the Digital Age (LIDA) 2002, Dubrovnik, Croatia, May 2002

Lowe, C.A., Eisenberg, M.B. Big6TM for Information Literacy// Theories of Information Behavior/edited by Fisher, K. E., Erdelez, S., McKechnie L. (E. F.). Medford, New Jersey: The American Society for Information Science and Technology : Information Today, Inc., 2005, p. 63-68

Thissen, Frank. Management Games in Intercultural Training. Proceedings of the 7th International Bobcats – Symposium Bratislava, 1999.

Todd, R. (2001). Transitions for preferred futures of school libraries: knowledge space, not information place; connections, not collections; actions, not positions; evidence, not advocacy. Keynote address: International Association of School Libraries (IASL) Conference, Auckland, New Zealand, 2001. Keynote paper, IASL conference 2001 virtual session: paper from Ross Todd, available online at IASL: school libraries online [website]: http://www.iasl-slo.org/virtualpaper2001.html

With knowledge to knowledge: Appendix to the metodics of the school librarian work / Jasmina Lovrinčević, Dinka Kovačević, Jadranka Lasić Lazić, Mihaela Banek Zorica. – Zagreb: Philosophical College, Institute for information studies, Section for information studies, 2005., (Papers of the Institute for information studies; book nr.16) page 314

Digitalni mediji u obrazovanju

Mila Nadrljanski
Faculty of Maritime Studies in Split
Tršćanska35, Split, Croatia
milamika60@yahoo.com

Đorđe Nadrljanski
Faculty of Phylosophy in Split
Tršćanska35, Split, Croatia
djordje44@yahoo.com

Mirko Bilić
Faculty of Maritime Studies in Split
Zrinsko Frankopanska 38, Split, Croatia
mbilic@pfst.hr

Sažetak

Članak se bavi pedagoškim mogućnostima i učincima digitalnih medija u obrazovnom sustavu. Digitalni mediji nalaze široku primjenu u predavanjima, učenju, u programima za vježbu, u bazama podataka, alatima , igrama, u pokusima simulacije za kompleksnu komunikaciju i kooperativno okruženje. Različite aplikacije mogu se primijeniti u programima za učenje i u odgovarajućim aktivnostima.

Kad je riječ o sadržajima za tehnike učenja koje su zasnovane na različitim vrstama kodiranja i odnose se na različite modalitete, postoje i učinci kao što su interaktivni kontakti s objektima koji uče, mogućnosti prilagodbe ovisno o predznanju studenata, povratna veza u različitim aktivnostima, istraživanju i simulaciji, net programiranje u distribucijskim odjelima virtualnog prostora. Postavljaju se pitanja o opsegu uporabe takvih potencijala, o oblicima korištenja te o očekivanjima glede promjene kulture Sva ta pitanja zahtijevaju empirijski odgovor i objasnit ćemo ih u ovome članku.

U svakom slučaju, digitalni bi se mediji trebali primijeniti u obrazovnim sustavima kao što su obrazovanje djece u ranom razdoblju njihova života, rad s mladim ljudima, profesionalno obrazovanje i u poboljšanju sustava za obrazovanje odraslih.. Mogućnosti digitalnih medija ogledaju se u njihovoj kulturnoj biti i u obrazovnoj vrijednosti.

Ključne riječi: digitalni mediji, obrazovanje, studenti, pedagoške mogućnosti

1. Uvod

Ovaj je rad dio opsežne studije koju autori realiziraju u okviru projekta "Novi mediji u obrazovanju." U radu se sintezom izlažu opći stavovi o relevantnim čimbenicima primjene digitalnih medija u obrazovanju. Pritom se u ovom radu ne razmatra kvaliteta sadržaja (obrazovnog softvera) digitalnih medija. "Digitalni mediji imaju potencijal za uspostavu nove mogućnosti za nova iskustva i usluge putem Interneta, u svakom domu i u kompanijama širom svijeta"[1]. Za prijenos digitalnih informacija koriste se elektronički mediji > izvedeni izraz: **digitalni mediji**. Budućnost obrazovnih sustava zasniva se na sposobnosti za samoučenje – to je najvažniji element, a *e-learning* ima u tome značajnu ulogu. U jednom od dokumenata Europskog savjeta kaže se da obrazovanje na daljinu Internetom nije paralelan proces nego je dio djelotvorne integracije informacijsko-komunikacijskih tehnologija u obrazovanje.

Obrazovanje pomoću modernih interaktivnih medija je znatno kvalitetnije u odnosu na klasične metode obrazovanja. Digitalizirane se informacije mogu lakše montirati potpomognute slikom, animacijom i zvukom, istovremeno djeluju na više osjetila dajući potpunu informaciju. Prednost online sustava je u brzini protoka informacija i u mogućnosti njihove aktualizacije. Pomoću raznovrsnih medija omogućen je šarolik transfer znanja. Pojava interaktivnih multimedijalnih sustava je unaprjeđenje u odnosu na dosadašnje obrazovne softvere jer:

- kvalitetno se poboljšava komunikacija između korisnika i računala;
- korištenje sustava kodova i simbola, odnosno istodobno djelovanje na više osjetila pomoću obrazovnih sadržaja omogućava razvitak raznih kognitivnih stilova i učenici su više zainteresirani za učenje u odnosu na druge aktivnosti;
- prikaz sadržaja je zanimljiviji, pragmatičan i doprinosi usvajanju znanja u različitim situacijama;
- interaktivni multimedijalni sustavi omogućavaju transfer znanja istodobnim djelovanjem na više osjetila, a time se ubrzava proces usvajanja gradiva;
- mnogostrukim kodiranjem i modeliranjem moguće je stvaranje realnih i kompleksnih situacija i prikaz nastavnog sadržaja iz drugih perspektiva, u različitim kontekstima i na više razina apstrakcije;
- usvajanje pojmova potiče učenikovu/studentovu aktivnost, a rješavanje problema koje nudi program motivira učenika/studenta na učenje.

Kako bi se moglo govoriti o komunikaciji, potrebna je interakcija, tj. komunikacija postoji samo onda kada komunikacijski partneri stupaju u uzajamni odnos i razmjenjuju mišljenja. Samo tako uspijeva sporazumijevanje. Komunika-

[1] Rekao je Will Poole, potpredsjednik Windows Digital Media odjeljenja pri kompaniji Microsoft. "Windows Media 9 serija je korak naprijed u funkcionalnosti i performansama koje tržište zahtijeva kako bi moglo iskoristiti sve svoje potencijale".

cija je dvostrano zbivanje. Mediji su sastavni dio komunikacije. Oni su važan dio komunikacije u društvu i među ljudima. Mediji se koriste kako bi komunikacija uspjela, i to kako u individualnoj, tako i u javnoj komunikaciji. Mediji su uvijek usko povezani s pitanjem komunikacije. Za svaku je komunikaciju potreban medij, sredstvo pomoću kojega se razmjenjuje poruka između komunikacijskih partnera i koje je temelj za posredovanje između partnera. Isto vrijedi i za izravnu osobnu komunikaciju.

Temeljna važnost medija leži u tome što postižu učinak koji doprinosi uspjehu komunikacije. Taj učinak je u posredovanju. Zato je važno rasvijetliti fenomen komunikacije.

Znanost medije vidi u širem kontekstu, i to kao važne institucije društva:

- Mediji nisu samo kanali komunikacije pogodni za prijenos sustava znakova. Oni su i organizacije, dakle "svrsishodni društveni sustavi", i to kompleksni sustavi. Ovi sustavi imaju veliki utjecaj na sve slojeve društvenog bića. "I to su institucije koje igraju ulogu unutar uređenja društvenog sustava".
- "Da bi uspio proces komunikacije, potreban je medij, tj. prikazivanje jezičnih znakova koji se prenose. Oblik medija zavisi od vrste primijenjenih sustava znakova kao i od vrste društvenog kontakta između partnera u komunikaciji".

Širenjem digitalnih medija u profesionalnom i privatnom životu pojačan je zahtjev za korištenjem tih tehnologija i u obrazovanju. Obrazloženje toga zahtjeva zasniva se na općoj društvenoj važnosti, na budućoj profesionalnoj relevantnosti digitalnih medija, na pedagoškim učincima u poboljšanju učenja i rezultata učenja te na potencijalu promjena koje bi mogle nastupiti u institucijama kao što su škole.

Cilj ovog rada je stjecanje uvida u rad s digitalnim medijima u obrazovanju u našoj zemlji, te na osnovu postojećih iskustava formuliranje polja djelovanja za daljnje aktivnosti na tom području. Baza "podataka" je nereprezentativan izbor empirijskih studija i stručnih razgovora koji su povezani s rezultatima općeg medijsko-pedagoškog, naročito medijsko-didaktičkog istraživanja. Tako nastaje opis situacije rada s digitalnim medijima u obrazovanju koji nije u svakom pogledu reprezentativan i statistički potkrijepljen, ali ipak daje sažetu ocjenu situacije. Ovaj rad opisuje korištenje i važnost korištenja digitalnih medija u obrazovanju, ali i plan za buduću praktičnu primjenu do sada neiskorištenih potencijala. Cilj ovog rada je dati pozitivne impulse za optimizaciju i povezivanje postojećih aktivnosti u području obrazovanja.

2. Važnost i uloga digitalnih medija u obrazovanju

U našem dinamičnom svijetu obrazovanje se kruto drži tradicionalnih shvaćanja, stoga su se u tom području događale minimalne promjene. Sedamdesetih godina prošlog stoljeća, Marshall Mc Luhan u metafori o propasti Gutenbergove galaksije daje kobno proročanstvo o svijetu u kojem apokaliptično ne-

staje svijet pismenih Zapadnjaka.2 Na sreću, Mc Luhanova proročanstva su se samo djelomično obistinila. Ali čak ni on, kao ni ostali teoretičari nisu mogli pretpostaviti u kojoj će mjeri za oblikovanje kulture postati važna uloga Interneta i multimedije i koliko će se uloga elitnih intelektualnih grupa, "vlasnika znanja" marginalizirati u prenošenju, sistematizaciji i skladištenju informacija. Informatička revolucija treba riješiti problem kako bi postala nositelj klasičnih vrijednosti, životnih strategija i normi, a to postavlja i nove zadatke pred suvremeno obrazovanje. Veoma je važna redefinicija ciljeva i zadataka u obrazovnom sustavu.

Ključ je svakog obrazovanja: učenici/studenti trebaju uočiti i razumjeti uzajamne veze između raznih znanstvenih disciplina i moći korištenjem različitih, kompleksnih pristupa rješavati probleme, pronaći sebe kao i način suradnje s drugima. Interdisciplinarnost programa obrazovanja u svojoj biti podrazumijeva kvalitetniji odgoj i obrazovanje i ona mora postati imperativ, jer su i životni problemi obrazovane osobe interdisciplinarni.

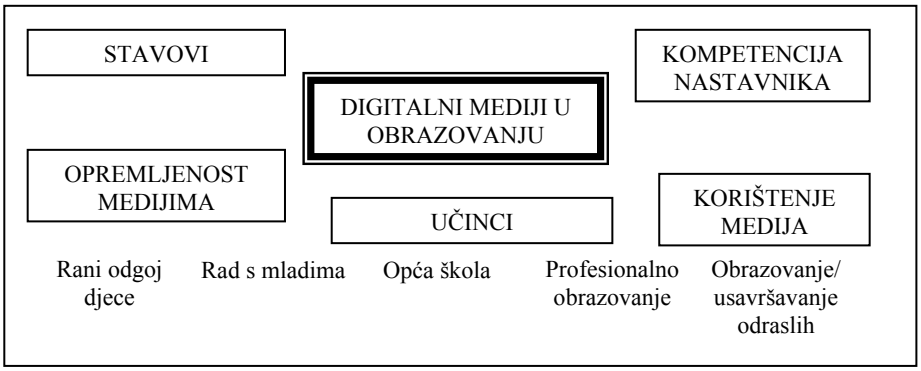
Digitalni mediji u obrazovanju nalaze različite primjene: od nastavnih programa za učenje i vježbe, od baza podataka i alata preko igara za učenje, pokusa i simulacija do kompleksnih komunikacijskih i kooperacijskih okruženja. U skladu s time višestruki su i mogući oblici primjene digitalnih medija u nastavi i s tim u vezi i nastavne aktivnosti. Očekivanja učinaka koji potiču na učenje povezuje se, između ostalog, s određenim funkcionalnostima i osobinama koje su karakteristične za digitalne medije, odnosno njihove posebne sadržaje. Određivanje položaja digitalnih medija zasniva se na evaluaciji empirijskih podataka odabranih studija u području digitalnih medija u školama. Važnost digitalnih medija nije ograničena na školsko podučavanje i učenje, nego igra izuzetnu ulogu i u ranom obrazovanju djece, radu s mladima, profesionalnom obrazovanju i u obrazovanju ili usavršavanju odraslih u kontekstu cjeloživotnog učenja.

Pitanje važnosti digitalnih medija za obrazovanje sigurno zavisi od toga koje potencijale oni zapravo osiguravaju. Ovo je teorijska perspektiva koja se temelji na određenim medijskim i tehničkim osobinama odabranih sredstava koji mogu utjecati na nastavni proces i nastavno- teorijska razmišljanja. Osobni odnos i pristup digitalnim medijima zavisi i od toga kakav stav prema njima imaju učenici/studenti, nastavnici i roditelji. U jednoj studiji o digitalnim medijima u obrazovanju[3], pokazalo se da su važne determinante i oprema i njezino korištenje. Ovo ne vrijedi samo za područje obrazovanja nego i za kućno okruženje, jer se

[2] « ... novi mediji tehnologije kojima se uvećavamo i proširujemo predstavljaju veliki kolektivni kirurški zahvat na društvenom tijelu ... područje reza gubi osjetljivost ... svaki novi zahvat mijenja odnose među osjetilima ... nijedno dosadašnje društvo ovo djelovanje nije osvijestilo u dostatnoj mjeri da bi moglo razviti neki imunitet.» Nadežda Čačinović, Doba slika u teoriji mediologije, Znanost u džepu, Naklada Jesenski i Turk, Zagreb, 2001.

[3] Doktorska disertacija M. Roguljić: Metodološko-metodički modeli obrazovanja na daljinu, Sveučilište u Zadru, 2006.

kreće od toga da ta dva okruženja nisu nezavisna jedno od drugog. Ključno mjesto u kontekstu nastave zauzima i nastavno osoblje čija stručnost, tj. kompetencija u rukovanju digitalnim medijima čini važan preduvjet za uspješnu integraciju digitalnih medija u nastavne procese u obrazovanju. Djelotvornost primjene digitalnih medija čini važno empirijsko pitanje – odnosi se na uspjeh u učenju u stručnim područjima, ali isto tako i na utjecaj na samo obrazovanje.



Slika 1. Model čimbenika u primjeni digitalnih medija u obrazovanju

Širenjem digitalnih medija u profesionalnom i privatnom životu pojačan je zahtjev za korištenjem tih tehnologija i u obrazovanju. Obrazloženje ovoga zahtjeva poziva se na opću društvenu važnost, na buduću profesionalnu relevantnost digitalnih medija, na pedagoške učinke u poboljšanju učenja i rezultata učenja kao i na potencijal promjena koje bi se mogle izazvati u institucijama obrazovanja.
Osim kreiranja sadržaja za učenje koji se zasnivaju na raznim vrstama kodiranja i odnose se na različite osjetilne modalitete, tu su i: interaktivni dodir s objektima učenja, mogućnosti prilagodbe određenim preduvjetima za učenje, feedback uz određene aktivnosti učenja, istraživanje i simulacije ili na mreži zasnovano stvaranje po mjestima distribuiranih sekcija za učenje u virtualnim prostorima.U kojoj mjeri su takvi potencijali iscrpljeni, koji oblici i scenariji korištenja prevladavaju i jesu li ispunjena očekivanja o promjeni kulture i rezultata učenja, empirijsko je pitanje koje treba pojasniti u ovom radu.
Cjelovita ocjena digitalnih medija mora se odnositi i na druge odgojne i obrazovne kontekste kao što je obrazovanje djece ranog uzrasta, rad s mladima, profesionalno obrazovanje i obrazovanje i usavršavanje odraslih. Osim toga, potencijali digitalnih medija ogledaju se u njihovoj kulturnoj važnosti i njihovoj općeobrazovnoj vrijednosti.
Ako se bavljenje digitalnim medijima u obrazovanju ne svede na medijsko-didaktička pitanja – i time u užem smislu na mogućnosti poboljšanja procesa učenja – onda će na vidjelo doći daljnji zadatci, koji se odnose na promjene izražene pojmom "digitalne kulture". Digitalni mediji stvaraju nove uvjete samorazumijevanja i razumijevanja svijeta i djeca i mladi ih koriste naročito u procesu

neformalnog učenja. O takvom obliku stjecanja znanja i mogućnostima povezivanja s institucionaliziranim procesima učenja zna se malo. Bavljenje digitalnim medijima trebalo bi ući u obrazovnu svakodnevicu i za osnovu imati razumijevanje korelacije između kulture i tehnike.

Dosadašnji školski obrazovni procesi karakteristični su po određenoj zatvorenosti. Ta situacija se mijenja kad se nastava otvara uvođenjem digitalnih medija – naročito Interneta. Time se gubi dio spomenute zatvorenosti.

Važno pitanje je i postoje li interesi, motiviranost i spremnost za učenje i spoznaju na takav način, uz posredovanje IKT-a. Ako se pitamo postoje li takvi interesi kod nas, može nam pomoći istraživanje koje je 2003. godine realizirala grupa s Učiteljskog fakulteta u Beogradu. To su još uvijek neobjavljeni rezultati, pa će biti naveden samo ovaj koji je zanimljiv za našu temu – interes za obrazovanje na daljinu. Od 1104 ispitanika – studenata i učitelja Učiteljskog fakulteta, na pitanje: Kada biste imali mogućnost, izabrali biste:

- tradicionalni način učenja
- obrazovanje na daljinu
- kombinaciju tradicionalnog obrazovanja i obrazovanja na daljinu,

849 (76,9%) ispitanika je odgovorilo – kombinaciju tradicionalnog obrazovanja i obrazovanja na daljinu. Ohrabrujuće je to što interes postoji, a naročito to što je on najveći baš za onaj oblik obrazovanja na daljinu koji u svijetu trenutno bilježi najviše uspjeha – za takozvanu distribuiranu nastavu. Ovaj termin odnosi se na kombinaciju tradicionalnog obrazovanja i obrazovanja na daljinu, u kojoj su uključeni najbolji elementi jednog i drugog oblika obrazovanja. Pri interpretaciji ovog rezultata treba uzeti u obzir i to da studenti i učitelji iz Srbije mahom nisu imali iskustva s obrazovanjem na daljinu, ali ipak, pokazuje se, prepoznaju njegove prednosti (ponajprije visok stupanj fleksibilnosti koju ono nudi). Ovo istraživanje je pokazalo i da samo 22% ispitanika potpuno vjeruje u to da postignuća stečena obrazovanjem na daljinu treba vrjednovati ravnopravno s postignućima stečenim tradicionalnim učenjem, dok se dodatnih 40% uglavnom slaže s tom tvrdnjom.

Online informacije koje učenici/studenti koriste nisu redakcijski obrađene i didaktički reducirane, stoga su poseban izazov za nastavnike. Takve informacije moraju se vrjednovati, a zahtijevaju i iznošenje mišljenja. Istodobno mora postojati i cilj – osposobiti učenike za to da o medijskim sadržajima razmišljaju kritički, ocjenjuju ih i po potrebi ih koriste.

Pitanje ovladavanja medijima nije relevantno samo za školski uzrast nego je posebno važno u fazi razvoja u ranom djetinjstvu. Od početnog promatranja medija kao auditivnih i vizualnih izvora izazova preko razvijanja medijskih želja i sklonosti kao i prvih oblika samostalnog ovladavanja medijima do aktivnog rada s medijima, mogu se pratiti različite faze dječjeg korištenja medija. Prvi kontakti se odvijaju u roditeljskom domu i oni svakako zavise od obrazovne razine odgajatelja. To znači da su mala djeca iz neobrazovanih slojeva više izložena opasnosti razvijanja problematičnih medijskih navika i rizičnih sklonosti

prema medijima. Prema tome, dječji vrtići imaju veliku odgovornost da u medijsko-odgojnom smislu sustavno prate djecu pri njihovu ovladavanju medijima. Pritom bi djecu trebalo osposobiti koristiti medije kao mjesta neformalnog učenja, kao sredstva ovladavanja svijetom i kao predmet kritičke ocjene – od čega će imati koristi u organizaciji vlastitog života. To u strukturalnom pogledu zahtijeva unaprjeđenje medijsko-pedagoškog obrazovanja odgajatelja kao i usko povezivanje obiteljskog medijskog odgoja i profesionalnog unaprjeđenja medijske kompetencije u odgajateljskim ustanovama.

Za učenike srednjih škola i studente mediji su mjesta neformalnog učenja izvan institucionaliziranih procesa učenja. Ova populacija koristi medije za orijentaciju u razvoju vlastitih ličnosti, kao izvor informacija i znanja i za razvijanje medijske kompetencije. Nerazmjerne su sposobnosti mladih u korištenju ponude informacija i znanja. To vodi širenju jaza između obrazovanih i manje obrazovanih.

"Digitalna kultura" na poseban način daje pečat i profesionalnoj svakodnevici i stalno je mijenja. Za profesionalno obrazovanje se postavlja poseban zadatak – analizirati i razmišljati o prodoru informacijskih i komunikacijskih tehnologija u sustave rada i o promjeni zadataka koja iz toga proizlazi. Taj prodor utječe sve više na nestajanje granica između rada i učenja. Naglasak nije više na svladavanju nekog sustava rada, nego na razvijanju sposobnosti kojima će se uz pomoć digitalnih tehnologija riješiti određeni problemi u procesu rada. U skladu s tim profesionalno obrazovanje dobiva zadatak podržavati učenje o procesu rada pomoću digitalnih medija prikladnim nastavnim sadržajima, koji uvažavaju kulturu učenja usmjerenu prema posebnim interesima pojedinih zanimanja.

Obrazovanje u smislu stjecanja raznih kvalifikacija ne označava kraj procesa učenja – na to ukazuje dalje usavršavanje i cjeloživotno učenje. Procesi usavršavanja podliježu posebnim okvirnim uvjetima, organizirani su primjerice kao izvanredni studij, kao učenje uz rad ili usavršavanje u slobodno vrijeme. Razumljivo je korištenje mogućnosti digitalnih medija, naročito Interneta, za pomoć u takvim procesima učenja. To iziskuje različite napore na području istraživanja i razvoja, infrastrukture i kvalificiranja. Traže se i odgovori na brojna pitanja, od razvoja posebne didaktike prikladne za poticanje i unaprjeđenje razvijanja kompetencije u procesima učenja zasnovanog na mreži, preko pripreme standardnih tema za usavršavanje do stvaranja obrazovnih profila osoba koje se usavršavaju.

Posebna važnost digitalnih medija za odgojne i obrazovne procese u svim životnim razdobljima može se objasniti posebnom funkcionalnosti medija. U takozvanom Social Software leži veliki potencijal digitalnih medija s obzirom na povezivanje ljudi – kako u školi tako i u profesionalnom obrazovanju, obrazovanju odraslih i daljem usavršavanju ili u radu s mladima. U neformalnim, neinstitucionaliziranim kontekstima osnivaju se samoorganizirajuća društva, u kojima sudionici uz visoku identifikaciju s društvom na osnovu mreže zajedno

konstruiraju znanje i omogućuju dalji transfer. Takvi procesi daju rezultate i u institucionalnim kontekstima te stvaraju poseban obrazovno-politički izazov.

3. Stavovi prema digitalnim medijima

Važan – premda ne i dovoljan preduvjet uspješnog rada s digitalnim medijima u obrazovanju je pozitivan osnovni stav prema novim medijima. Jedno takvo raspoloženje može se shvatiti po pokazateljima subjektivnog mišljenja[4] o važnosti rada na računalu, zainteresiranosti za računalo, društvene važnosti računala za svakodnevni i profesionalni život i mogućih pozitivnih ili negativnih utjecaja.

Naša studija pokazuje da učenici u Hrvatskoj imaju pozitivan osnovni stav prema digitalnim medijima. Nasuprot još nedovoljnom korištenju digitalnih medija u nastavi, djeca i mladi kod kuće često koriste računalo. Godine 2005. je 38% djece starosne dobi između 6 i 13 godina koristilo je računalo u školi redovno najmanje jednom tjedno, a 86% njih redovno je koristilo računalo kod kuće. 17% mladih koristilo je računalo svakodnevno ili više puta tjedno u školi, a 76% njih je to istom učestalošću činilo kod kuće. Kućno korištenje nije ograničeno na izvanškolske interese, nego obuhvaća i aktivnosti učenja za nastavu. Tako je i u međunarodnoj usporedbi[5], pri čemu svakako postoje jasne razlike između dječaka i djevojčica, a u korist dječaka. Uz rast društvene relevantnosti računala i u vezi s tim nužnosti kompetentnog služenja njime, za učenike su nada u poboljšanje nastave i učenja i moguće profesionalne potrebe važni elementi koji utječu na važnost digitalnih medija i na otvoren osnovni stav prema njima. Ukupna pozitivna klima u odnosu na uvođenje digitalnih medija u obrazovanje vlada i kod nastavnika. U brojnim se izvorima u literaturi naglašava društvena važnost i profesionalna relevantnost novih medija za učenike. Digitalni su mediji po riječima mnogih nastavnika doprinijeli poboljšanju motivacije, osposobljavanju za samostalan rad i rad na projektima kao i zanimljivom organiziranju nastave[6]. Pritom nastavnici koji imaju iskustva s medijima u pravilu imaju prema njima izraženiji pozitivan stav u odnosu na one bez takvog iskustva. Nastavnici su nepovjerljiviji prema mogućnostima koje nisu neposredno povezane s medijem, primjerice samostalan rad ili individualna iskustva u učenju. Više se napora usmjerava na poticanje kooperativnog rada s medijima, individualno poboljšanje napretka učenika ili razvoj socijalnih sposobnosti uz pomoć medija.

Ti rezultati ukazuju na nedostatak odgovarajućih iskustava nastavnika i nedostatak predodžbe o potrebama u određenim područjima u kojima su didaktički zahtjevi relativno visoki (npr. u kontaktu s heterogenošću). Na digitalne medije se u principu ne gleda kao na nužan preduvjet dobre nastave, već više kao na

[4] M. Roguljić – Doktorska disertacija

[5] Studija o korištenju digitalnih medija u općeobrazovnim školama u Njemačkoj

[6] Dj. Nadrljanski: Obrazovni softver-hipermedijalni sistemi, Univerzitet u Novom Sadu, 2000.

potencijal koji doprinosi obogaćivanju i poboljšanju nastave i koji treba iskoristiti. O stavu obrazovnih vlasti[7] postoji manje podataka. Stavovi prema primjeni digitalnih medija uglavnom su potpuno pozitivni i otvoreni. Osoblje iz školske uprave često ima manje iskustva u primjeni digitalnih medija u nastavi, a zbog svoje rukovodeće funkcije zavisno od okolnosti ova grupa podliježu izvjesnom očekivanju u smislu socijalne poželjnosti, pa u tom svjetlu treba analizirati i njihove izjave. Ako se tome doda činjenica da i nastavnici potvrđuju otvoren stav školskih uprava, može se uočiti pozitivan trend.

Roditelji u digitalnim medijima vide mogućnosti i korist za didaktičko područje kao podršku procesima učenja, ali su – naročito u svezi s Internetom – u medijsko-odgojnom pogledu kritični i zabrinuti. Ali oni uglavnom – kao što se može pročitati u raznim izvorima – digitalnim medijima pripisuju veliku važnost za svakodnevni život i zanimanje i smatraju da je škola ta koja učenike mora osposobiti za rad na računalu. Posljednjih godina raste i broj roditelja koji su sebi postavili taj zadatak osposobljavanja kao obvezu.

Ako se promatraju studenti[8] i njihovi stavovi prema digitalnim medijima, onda oni posredničku funkciju digitalnih medija ocjenjuju kao apsolutno korisnu – to se naročito odnosi na programe učenja. Sumnjičavi su s obzirom na poboljšanje motivacije i individualizaciju. Stav studenata – kao i kod drugih grupa – zavisi od vlastitog iskustva. Nešto suzdržana ocjena studenata može biti izraz još ne posve realiziranog i prihvaćenog korištenja digitalnih medija u nastavi na fakultetima. Zaključujemo, sudionici u procesu obrazovanja u Hrvatskoj pokazuju pozitivan osnovni stav i osnovno raspoloženje u svezi s društvenom važnosti digitalnih medija i s obzirom na njihovu primjenu u nastavi.

4. Zaključak

Digitalni mediji su dobili veliku važnost u svim područjima života i stalno vode daljnjem razvoju i promjenama u različitim kontekstima. U to se ubrajaju na primjer, postupni razvoj nove kulturne prakse, komunikacija, promjena procesa učenja i poučavanja, nove mogućnosti stjecanja znanja, promjena procesa rada i proizvodnje te nastanak novih tržišta i zanimanja. Pritom je riječ o temeljnoj, radikalnoj i trajnoj promjeni, koja utječe na odnos čovjeka prema njegovu stvarnom i društvenom okruženju kao i prema samom sebi.

Integracija digitalnih medija povezana je osim s nastavom i s utjecajima na razna druga područja obrazovnog sustava, na tehniku, infrastrukturu i osoblje. To znači da se na digitalne medije mora gledati s obzirom na njihovu važnost u procesima razvoja obrazovanja. Ako se u tu svrhu analizira odgovarajuća literatura – u pravilu se radi o studijama pojedinačnog slučaja – onda je jasno zašto u mnogim školama promjene pokreće jedna grupa koju čini nastavnička jezgru.

[7] Anketa s ravnateljima škola u Splitu

[8] Anketa studenata Filozofskog i Pomorskog fakulteta u Splitu

Oni preuzimaju funkciju promotora u tehničkom i pedagoškom pogledu. U promociju se ubrajaju zadaci kao što je motivacija i podrška kolegama, uspostavljanje struktura rada i osiguranje financiranja i protoka informacija u školi i prema vani kao i razvijanje medijskih koncepcija. Jaka podrška školske administracije pokazala se kao važan okvirni uvjet. Podjednako važnim čine se razmjena mišljenja o ciljevima i sadržajima rada s digitalnim medijima i postizanje osnovne suglasnosti. Iako među nastavnicima prevladava uglavnom pozitivno osnovno raspoloženje, ne može se prešutjeti činjenica da osim aktivnih sudionika ima, iako su malobrojni, i onih koji se opiru. Procesi implementacije povezani su s reakcijama koje se ne mogu uvijek objasniti samom stvari, nego se uzrok često nalazi u nepovoljnim okvirnim uvjetima. Kako bi se otklonile teškoće oko prihvaćanja i ovdje je potrebno usuglašavanje oko problema. Aktivno sudjelovanje nastavnika u procesu integracije novih medija u nastavu ne zavisi u velikoj mjeri od postojećih kompetencija nastavnika. U tom kontekstu nastavnici izražavaju potrebu za intenzivnijim usavršavanjem u didaktičkom i tehničkom pogledu i s obzirom na pedagoške sadržaje. Internet je važan most između školskih i kućnih procesa učenja. Vrijeme za razvoj modela koji će omogućiti spajanje toga područja u okruženju integriranog iskustva, učenja i rada. Važan preduvjet za to je djelotvorna mrežna infrastruktura sa širokopojasnim priključcima odnosno pristupima.

Na fakultetima je ponuda medijsko-odgojnih i medijsko-didaktičkih sadržaja u području znanosti o obrazovanju i stručnih didaktika na studiju za obrazovanje nastavnika relativno mala. Tek u malom broju slučajeva postoje takvi sadržaji pomoću kojih se mogu osigurati minimalni standardi[9]. Ako sve uzmemo u obzir, možemo reći da se medijsko-pedagoškim sadržajima u obrazovanju budućih nastavnika ne poklanja odgovarajuća pozornost. Pojedini inozemni fakulteti[10] razvili su posebne fakultetske studije o području medija, tako da studenti ne nalaze samo široku danu ponudu profesionalnog obrazovanja, nego mogu s tim povezati različite oblike dodatne kvalifikacije – od certifikata do dodataka diplomi.

[9] Pedagoški fakultet Svečilišta u Zagrebu i Odsjek za informacijske znanosti Filozofskog fakulteta u Zagrebu nude kolegije i programe za buduće nastavnike u kojima studenti samostalno razvijaju obrazovne softvere i uče teoretska načela oblikovanja multimedijskog materijala u naprednim aplikacijama sa složenim programiranjem i visokokvalitetnim sučeljima.

[10] U Njemačkoj postoje studije profila pedagoga medija; Pedagoški fakultet u Somboru, Republika Srbija ima studije dizajna medija u obrazovanju, autor programa je prof. dr. sc. Djordje Nadrljanski; Postoji inicijativa za medijsko obrazovanje na Sveučilištu u Splitu.

Literatura

Considine, David: Approaches to Media Education. Internet-document. URL: http://www.media-awareness.ca/eng/med/bigpict/medapp.htm#1,1994.

Council of Europe: Report on Media Education. Internet-document: http://stars.coe.int/doc/doc00/edoc8753.htm, 2000.

Čačinović, Nadežda: Doba slika u teoriji mediologije, Znanost u džepu, Naklada Jesenski i Turk, Zagreb, 2001.

European Centre for Media Competence. URL: http://www.ecmc.de/

European Community) Europe Initiative: An Information Society For All. Internet document: http://europa.eu.int/comm/information _society/eeurope/index_en.htm, 2001.

Ferguson, Robert: Global Interculturalism and the Dilemmas of Universalism: Teaching the Media after 2000. A paper to be delivered to the International Congress on Communication and Education, Sao Paulo, Brazil. Internet-document: www.eca.usp.br/nucleos/nce/pdf/140. pdf, 2000.

Hart, Andrew: Paradigms Revisited: Media Education in the Global Village. In: Media Education Centre. Frankfurt 1997.

Instructional System Design http://www.nwlink.com/~donclark/hrd/sat3.html

Joint Hearing between the European Parliament and the Parliamentary Assembly of the Council of Europe: Media Education. Internet-document: http://www.europa.eu.int/geninfo/whatwas new/032000.htm., 2000.

Komenczi Bertalan: Orbis sensualium pictus – Multimédia az oktatásban, Iskolakultúra 1997/ br. 1.Mergel, Brenda, Instructional Design & Learning Theory, http://www.usask.ca/education/coursework/802papers/mergel/brenda.htm

Kumar, Keval. J.: Media Education in India: Strategies, Trends, Visions. Stockholm Media Education Conference. Internet-document. URL: http://home.swipnet.se/~w-52050/konf_keval_kumar.htm, 2000.

Lange, Bernd-Peter: Media Competence in the Age of Global Communication. Internet-document. URL: http://sef-bonn.org/sef/publications/pol-pap/no11/content.html, 1999.

Latzer et. al.: Die Zukunft der Kommunikation. Phänome und Trends in der Informationsgesellschaft. Studienverlag Innsbruck – Wien, 1999.

Latzer, Michael: Mediamatik – Die Konvergenz von Telekommunikation, Computer und Rundfunk. Opladen, 1997.

Lohl, Horst: International Comparisons in Media Education Research: experiences with a small-scale project and perspectives for further research. In: Media Education Centre. Frankfurt Papers, 1997.

Luhmann, Niklas: Die Realität der Massenmedien. Opladen, Westdeutscher Verlag.,1996.

McQuail, Denis: The Future of Communication Theory. In Latzer 1999.

Media Education Centre: The Media Education Symposium at the European Conference on Educational Research. Frankfurt Papers 1997. Internet-document. URL: http://www.soton.ac .uk/~mec/MECWEB/ECERedit.pdf, 1997.

Meier, Werner A.:Was macht die Publizistik- und Kommunikationswissenschaft mit der Konvergenz? In Latzer 1997.

Nadrljanski Roguljić, Mila: Metodološko-metodički modeli obrazovanja na daljinu – doktorska disertacija Sveučilište u Zadru 2006.

Nadrljanski, Đorđe: Obrazovni softver, Univerzitet u Novom Sadu, Tehnički fakultet "Mihajlo Pupin", Zrenjanin, 2000.

Pungente, John: Eight Key Concepts of Media Literacy. Internet-document. URL: http://www.media-awareness.ca/eng/med/bigpict/8keycon.thm, 1989.

Sanders, Franklin: The Global Rise of Ethnic Nationalism. Internet-document. URL : http://www.dixienet.org/spatriot/vol2no2/member4.html, 1998.

Ured visokog predstavnika, Saopštenje za štampu 2. mart 2001. god. Internet dokument: http://www.ohr.int/press/p20010302a.htm

Digitalni mediji u obrazovanju – pregled međunarodnih iskustava

Đorđe Nadrljanski
Faculty of Phylosophy in Split
Tršćanska35, Split, Croatia
djordje44@yahoo.com

Mila Nadrljanski
Faculty of Maritime Studies in Split
Tršćanska35, Split, Croatia
milamika60@yahoo.com

Marko Tomašević
Faculty of Maritime Studies in Split
Zrinsko Frankopanska 38, Split, Croatia
mtomasevic@pfst.hr

Sažetak

Istraživanje i analiza situacije koja se odnosi na digitalne medije u našem obrazovanju bit će uključeni u kontekst međunarodnog razvoja u različitim zemljama. Na temelju analiza međunarodnih studija, metaanaliza, preporuka i Delfi studija nalazimo da je postojeće znanje uz određeni predviđeni razvoj obećavajuće. U usporedbi s međunarodnim iskustvima naši studenti imaju naglašeno pozitivan stav o digitalnim medijima.
Uzimajući u obzir međunarodne usporedbe, primjerice u Njemačkoj, očigledno je da postoje razlike u stavovima dječaka i djevojčica i jako su izražene. Naglašeno pozitivni stavovi roditelja, učitelja i studenata u skandinavskim zemljama mogu se dovesti u vezu s općenito pozitivnim stavovima u društvu u kojemu je integracija digitalnih medija iznimno napredovala.

Ključne riječi: digitalni mediji, obrazovanje, međunarodna iskustva

Uvod

U svijetu se provode mnogobrojne aktivnosti na području obrazovanja, a temelje se na primjeni novih medija. Svake se godine o toj temi objavljuju stotine knjiga. Održavaju se kongresi, konferencije i medijski kampovi. Pokreću se časopisi, fondacije i društva za obrazovanje pomoću digitalnih medija. Traju rasprave o novim temama. Razvijaju se obrazovni alati, video-kasete i nastavna sredstva. Mogu se naći baze podataka, portali, online tečajevi, CD-romovi i

portfoliji. Stimulaciju i motivaciju pružaju modeli novog medijskog obrazovanja koji su se pokazali najboljima u praksi. Navršilo se 48 godina otkako je prvi digitalni medij (računalo) ušao u obrazovni sustav. Time je otpočeo i proces digitalizacije obrazovanja. U jednoj našoj opsežnijoj studiji pod nazivom "Novi mediji u obrazovanju"1, čije sažete izvode izlažemo u ovom radu, sagledali smo neke relevantne čimbenike u vezi s korištenjem digitalnih medija u obrazovanju. Posebna pozornost je usmjerena na predstavljanje različitih stavova i predrasuda s obzirom na učinak digitalnih medija u obrazovanju prema našim i inozemnim iskustvima. Digitalizacija ruši prepreke između tradicionalnih medijskih industrija i briše nekadašnje razlike među njima. Informacijsko – komunikacijsko se područje u posljednjem desetljeću razvija dvostruko brže od prosjeka globalne ekonomije. U okviru toga područja najveći rast ima međunarodno komuniciranje. Danas se koristi univerzalni medij – računalo – opremljen LCD monitorom, zvučnikom, modemom, pisačem, CD/DVD-om. Sinteza tih elemenata dovela je do univerzalnog digitalnog medija koji je istovremeno računalo, mobilni telefon i televizor. Mobilna je telefonija postala polje konvergencije. Konvergencija medijskih tehnologija stvara nesigurnost i na tržištu jer se ruše granice između područja koja su ranije bila relativno stabilni oligopoli. Znalo se tko se bavi televizijom, tko računalima, a tko telefonijom. A danas?

Na svjetskoj se sceni javio novi komunikacijski sustav nazvan medijamatika. On čini glavnu značajku civilizacijskog poretka poznatog kao informacijsko društvo, a rezultat je kombiniranja tehničkih, ekonomskih i političkih trendova digitalizacije, liberalizacije, konvergencije i globalizacije u sektoru komunikacija2. U okviru Europe, na različitim nacionalnim razinama, osmišljene su strategije i mjere politike radi unaprjeđenja informacijskog društva. To se, prije svega, odnosi na liberalizaciju telekomunikacija, uspostavljanje jasnog zakonskog okvira za e-gospodarstvo, kao i za istraživanje i razvoj u ključnim područjima ljudske djelatnosti. Inicijativa za e-Europu "ima za cilj omogućiti online pristup svakom građaninu, svakom domu, školi, tvrtki i upravi", "da time stvori digitalno pismenu i poduzetničku Europu" i "da osigura informacijsko društvo koje obuhvaća sve svoje dijelove"3. Sve je to otvorilo niz pitanja i problema i u sustavu obrazovanja. Kako treba pristupiti medijskom obrazovanju ljudi za informacijsko doba? Jesu li kompjutorska pismenost i pristup tehnologijama prioriteti? Je li važnija svijest o postojanju i ulozi novih medija? Treba li na tim trendovima nastojati u zemljama Jugoistočne Europe samo zato što su takvi

[1] U izlaganju su izostavljeni kvantitaivni podatci s obzirom na ograničenost prostora, a na temelju kvalitativne analize izneseni su stavovi autora.

[2] Latzer, Michael (1997.): Mediamatik – Die Konvergenz von Telekommunikation, Computer und Rundfunk. Opladen.

[3] Council of Europe (2000.) : Report on Media Education. Internet-document: http: //stars.coe.int/doc/doc00/edoc8753.htm

trendovi dio medijske scene razvijenih zemalja ili oni mogu pomoći razvoju društva, demokratizaciji, transparentnosti javnih djelatnosti i službа? Kakva je dinamika razvoja novih medija u obrazovanju u zemljama u tranziciji? Znače li nove tehnologije necenzurirani protok informacija i znanja? Kako zaposlenike uspješno učiti o medijima – da uživaju nove slobode, koriste nove alate u poslu i snose odgovornost za razvoj informacijskog društva u vlastitoj zemlji? Uvažavaju li postojeći redovni i alternativni oblici obrazovanja sadašnje i nagoviještene promjene u području digitalnih medija? Što treba raditi? Čovjek u informacijskom dobu. Mijenja li intenzivna informatizacija navike u svakodnevnom životu? Što se događa s interpersonalnom komunikacijom? Nastojanje na novim izvorima i oblicima informacija – kako naučiti izabrati? Mijenja li se tradicionalni društveni i organizacijski okvir funkcioniranja porodice? Kako će čovjek upravljati vlastitim potrebama, kako će gospodariti radnim i slobodnim vremenom? Sve je to motiviralo autore ovoga rada na istraživanje i prikazivanje nekih relevantnih čimbenika učinkovitosti primjene digitalnih medija u obrazovanju.

Čimbenici primjene digitalnih medija u obrazovanju

U mnogim istraživanjima stavova i iskustava u primjeni digitalnih medija važno je istaći i istraživanja koje započela Hrvatska akademska i istraživačka mreža CARNet4. Ona je još 2000. godine pokrenula program "Obrazovni projekti" čiji je cilj bio ispitati kako se primjena informacijske tehnologije uklapa u obrazovni sustav hrvatskih visokih učilišta. Na osnovu rezultata koji su dobiveni u našem istraživanju5, može se zaključiti da primjena digitalnih medija pospješuje komunikaciju između nastavnika i studenata, u nekim slučajevima zamjenjuje nastavnika, a može motivirati čak i one studente kod kojih su se sve druge metode pokazale manje uspješnima. Veoma je važno to što uporaba računala potiče individualizaciju nastave, čime se povećava kvaliteta učenja. Učenici su obično oduševljeni i prihvaćaju učenje iz elektronskog nastavnog materijala. Svaki učenik obrađuje zadane teme tempom koji mu odgovara, ponavlja i istražuje na svoj vlastiti način. Obrazovanje pomoću digitalnih medija mora se suočiti sa sljedećim:

- brzim rastom medijskih industrija, ubrzanim promjenama uvjeta u kojima mladi ljudi odrastaju i brzim razvojem novih žanrova i formata;
- promijenjenom ulogom škola kao jedinih prijenosnika informacija i izvora u stjecanju znanja, kao i promijenjenom ulogom nastavnika;
- promijenjenim načinom informiranja zbog multimedijskog i hipertekstualnog okruženja u kojemu se informira.

[4] Danijela Horvat: Obrazovni materijali dostupni putem Weba – očekivanja i stavovi studenata CARNet – Časopis Edupoint godište III | broj 16 | Zagreb | 30. 6. 2003. | ISSN 1333-5987

[5] Učinkovitost nastave informatike podržana računalom – sprovedeno u OŠ "Kman – Kocunar" i OŠ "Spinut" u Splitu –Doktorska disertacija M.Roguljić 2006.

- obrazovanjem pomoću digitalnih medija morat će se pronaći načini kako školama prenijeti svijest o važnosti medijske kompetencije u širem smislu.
- promijenjenim i ponekad olakšanim uvjetima za izvođenje obrazovnog procesa zbog multimedijskih aplikacija.

Sljedeći su čimbenici nužni za razvoj obrazovanja pomoću digitalnih medija:

- proširenje područja obrazovanja, posebno zato što neformalno obrazovanje i učenje tijekom cijelog života postaju sve važniji. Cjeloživotno se obrazovanje usmjerava prema odraslim ljudima i budućim odraslim ljudima, a s ciljem da se iz tih aktivnosti ne isključi nijedna društvena grupa;
- suradnja s medijskim profesionalcima i njihovo sudjelovanje u obrazovanju pomoću digitalnih medija, kao i kontakti s medijskim industrijama;
- razvoj novih načina komunikacije i obavljanje obrazovnih zadataka pomoću virtualne pedagogije, obrazovanja na daljinu, otvorenog obrazovanja ili TeleMonitoringa;
- povezivanje na međunarodnoj, europskoj, nacionalnoj, regionalnoj i lokalnoj razini, kao i suradnja s stručnim suradnicima.

Sve to mora omogućiti suvremena metodika nastave i pristup na svim razinama rada te aktivnosti studija e-learninga. Suvremene znanstvene discipline i metode istraživanja su temeljni preduvjeti za stjecanje općih, humanističkih i umjetničkih znanja s jedne, i znanja iz gospodarstva, prava, političkih, menadžerskih i tehničkih znanosti s druge strane. Pojava novih komunikacijskih tehnologija i medija omogućila je promjene u obrazovanju uopće. Postavlja se pitanje odnosa oblika i biti tih promjena. U svijetu posljednjih dvadesetak godina obrazovanje na daljinu postaje vrlo popularno zahvaljujući brzom razvoju informatike i informacijskih tehnologija. Tako je učenje potpomognuto računalima i računalnim mrežama zamijenilo prije poznate metode obrazovanja na daljinu, kao što su dopisne škole i obrazovni programi na radiju i televiziji. Bitnu novost u obrazovanju ipak čini e-learning. Obrazovanje pomoću digitalnih medija može se provoditi na više načina, u većem ili manjem opsegu. Poželjno je široko obrazovanje pomoću digitalnih medija. Čak i manje aktivnosti koje provode pojedini nastavnici, škole i informacijski profesionalci mogu biti učinkovite, što je već dokazano istraživanjem. Svakom su demokratskom društvu potrebni budni, kvalificirani nastavnici i građani koji su informirani i koji iskazuju medijsku kompetenciju. Općenito govoreći, obrazovanje pomoću digitalnih medija u Hrvatskoj se ne treba razlikovati od takvog obrazovanja u drugim zemljama. Već su u uporabi nove tehnologije na europskoj razini. Blackboard, Moodle i slični načini e-učenja i korištenja digitalnih medija se na pojedinim sveučilištima u Hrvatskoj koriste jednako uspješno kao i na poznatim sveučilištima u Velikoj Britaniji, ostatku Europe i u SAD-u, rodnom mjestu e-learninga. U Zagrebu je na Odsjeku za informacijske znanosti FF-a razvijen program Omega zasnovan na Moodleu (omega.ffzg.hr) koji studenti i profesori koriste već nekoliko go-

dina. Pod vodstvom prof. dr. sc. Blaženke Divjak u Varaždinu se odvija projekt e-učenja, koji se nalazi u sklopu međunarodnog FP7 projekta. Koncepti i sadržaji obrazovanja pomoću digitalnih medija zasigurno će biti prilagođeni konkretnoj situaciji i konkretnim potrebama naše zemlje.

Uvjeti za primjenu digitalnih medija u obrazovanju

Preduvjeti za učenje. Kompetencije koje se odnose na medije, kojima učenici/studenti raspolažu na početku obrazovanja, uglavnom su male i odnose se na korištenje standardnih programa, dakle na vještine rukovanja, pri čemu studentice u pravilu imaju manje iskustva od svojih kolega. Stereotipi i predrasude u pogledu spolne podijeljenosti utjecali su na način pristupa digitalnim medijima. Reformom srednjoškolskog obrazovanja nadilaze se te razlike. U odnosu na kompetencije u pogledu stavova prema digitalnim medijima, studenti učiteljskih fakulteta[6] su naročito slabi u odnosu na druge studente. Kako upravo ti studenti u budućnosti trebaju poticati u didaktičkom i odgojnom pogledu i podržavati intenzivno bavljenje djece i mladih digitalnim medijima, važno je da se toj grupi za vrijeme studiranja približe odgovarajući sadržaji. Upravo je taj cilj doveo do izmjena kurikuluma obrazovanja budućih učitelja i do uvođenja novih kolegija iz medijskog obrazovanja.

Medijsko-pedagoška kompetencija. Mada još ne postoje potpuni empirijski podaci[7] o učinkovitosti obrazovanja nastavnika s obzirom na razvoj medijsko-pedagoške kompetencije, ipak prvi rezultati ukazuju na to da se nedostaci koji su prisutni na početku obrazovanja i kompetencije nužne za profesionalnu nastavnu djelatnost ne mogu potpuno nadoknaditi. Čini se da se mnoga nastojanja odnose na razvijanje osnovnih kompetencija tako da se samo pojedinačno mogu promatrati medijsko-didaktička i medijsko-odgojna pitanja.

Potreba usavršavanja. U skladu s tim, kod nastavnika postoji velika potreba za usavršavanjem – ponajprije u tehničkom, zatim u metodičko-didaktičkom području. Kako za one bez iskustva, tako i za nastavnike s iskustvom, u radu s digitalnim medijima važna je podrška u obliku metodičko-didaktičkih poticaja i pomoći. Sa sve širom primjenom medija povećava se potreba za daljom tehničkom softverskom podrškom, što dovodi do spoznaje da će se duljim radom s medijima otvoriti dalji potencijali digitalnih medija i ostvarivati složenije okruženje za koja će biti potrebna dodatna pomoć. Rad s digitalnim medijima za mnoge nastavnike usmjeren je ponajprije na tehnička i softverska pitanja, zatim sve više na didaktičko-metodičke smjernice – kako pomoću digitalnih medija poboljšati nastavu, odnosno kako lakše izlagati stručne sadržaje u nastavi. U

[6] Anketno istraživanje sprovedeno među studentima učiteljskih i tehničkih fakulteta (Novi Sad, Zrenjanin , Beograd i Sombor, R Srbija, vođa istraživanja Đ. Nadrljanski)

[7] U sklopu istraživanja učinaka usavršavanja učitelja za nastavu informatike u osnovnim školama izvedene su navedene konstatacije.

2003. godini polovica nastavnog osoblja[8] često ili povremeno koristi računalo u nastavi, jedna trećina nastavnika u općeobrazovnim školama često ili povremeno koristi Internet. Nastavnici kažu da u budućnosti žele više koristiti medije. Bez obzira na to što 72% njih želi češće koristiti Internet, a 68% češće raditi pomoću računala, još uvijek oko 50% nastavnika nije uopće koristilo digitalne medije u nastavi. Empirijski podaci pokazuju da, ovisno o vrsti škole, jedna grupa od 10 do 30% nastavnika redovito koristi digitalne medije u nastavi. Ove su grupe glavne korisnici online sadržaja za nastavnike. Preko 90% njih koristi računalo za pripremu nastave i za rad poslije nastave svakodnevno ili više puta tjedno, a više od polovice njih koristi računalo u nastavi svakodnevno ili više puta tjedno.

Čini se da se kreće od problema medija, a ne od nastave i aktivnosti učenika. To odgovara i u didaktici često raširenom mišljenju – najprije se moraju postaviti osnove, a onda se može pristupiti rješavanju određenog problema.

Medijsko obrazovanje. Samo neznatnu ulogu ima pitanje kako se može poticati i razvijati svjesna, odgovorna i kritička primjena medija. U skladu s tim je razmjerno mala i potreba za medijsko-odgojnim usavršavanjem.

Vrste usavršavanja. S obzirom na vrstu usavršavanja, nastavnici daju prednost onim oblicima koji se nude u okviru škole, koji su bliski praksi, usmjereni na potrebe i čiji se sadržaji mogu izravno primijeniti u nastavi. Kao i u području obrazovanja nastavnika, tako i u području njihova usavršavanja, do sada nije obavljeno sustavno istraživanje o učincima. Na osnovu evaluacije pojedinih mjera mogu se steći prvi dojmovi. Spomenuta školska interna usavršavanja imaju visoke vrijednosti prihvaćanja, a nastavnici koji sudjeluju u dugotrajnim programima usavršavanja stječu kompetencije i sve veću spremnost da na temelju stečenih kompetencija mijenjaju vlastite didaktičke koncepte.

"Kultura usavršavanja". Rezultati novijih mjera usavršavanja uzimaju u obzir i Internet, a potvrđuju i utjecaj na nastavu u obliku jače motivacije ili veće zainteresiranosti učenika za digitalne medije. Nove su mjere izraz izvjesne promjene u kulturi usavršavanja. Dugo je nastava bila zatvoren sustav, a danas su nastavnici sve spremniji svoju nastavu izlagati diskusijom, sami sebe evaluirati i razmišljati o svojim nastavnim postupcima. Usavršavanja te vrste stavljaju naglasak na komunikaciju i kooperaciju, a nastavnici koji sudjeluju u njima doživljavaju ih kao veoma korisna. Takve mjere počinju konkretnim didaktičkim pitanjima o nastavi i cilj im je poboljšanje kvalitete nastave.

Usavršavanje zasnovano na webu. Veliku važnost za usavršavanje nastavnika dobili su i središnji, nadregionalni portali koji se koriste za pretraživanje nastavnih materijala. Iznimno velik broj nastavnika koji sudjeluju u usavršavanju misli da bi Internet trebali koristiti za suradnju s kolegama ili u Blended-Learning konceptima u okviru usavršavanja. S obzirom na sve, može se zaključiti da

[8] Učinkovitost nastave informatike podržana računalom sprovedeno u OŠ Kman Koncunar i Spinut u Splitu –doktorska disertacija M.Roguljić 2006.

je mali broj nastavnika obuhvaćen usavršavanjem i da je premali broj sadržaja koji se odnose na usavršavanje.

Digitalni mediji u školi u međunarodnoj usporedbi

Školske inovacije u području digitalnih medija ovise, osim o odgovarajućoj opremi, o sposobnosti nastavnika u poticanju i podržavanju nastavnih procesa primjenom digitalnih medija. Vrijeme je kada valja iskoristiti odgovarajuće kompetencije, s jedne strane kao osnovu sveučilišnog obrazovanja, a s druge strane za usavršavanje i daljnji razvoj. Studije o utjecaju digitalnih medija s obzirom na uspjeh u struci, ključne kvalifikacije, razvoj nastave i škole imaju različite istraživačko-metodološke pristup.

Načelno se mogu razlikovati:

- istraživanja općih utjecaja medija, primjerice poput komparativnog istraživanje medijima podržane nastave i tradicionalne nastave;
- istraživanje učinaka posebnih značajki medija, kao vrste kodiranja ili osjetilnih modaliteta;
- studije o evaluaciji.

Zaključak je da pitanje opće prednosti učenja pomoću medija nasuprot učenju bez medija nije od koristi jer je previše raznovrsnih utjecaja koji se ne mogu provjeriti. Eksperimentalne i kvazieksperimentalne studije dale su dragocjene spoznaje o tome kako bi trebalo oblikovati na računalu zasnovane sadržaje – npr. s obzirom na odnos teksta, slike i tona, kako bi se postigli što bolji rezultati u učenju ili u području rješavanja problema. S tim u vezi problematično je to što su eksperimentalne situacije relativno udaljene od školskih nastavnih situacija. Stoga su studije o evaluaciji korisne za istraživanje nastave jer daju odgovor na pitanje vodi li određena mjera pod određenim uvjetima do cilja i koji neželjeni učinci nastupaju u tom slučaju. Takva vrsta istraživanja u pravilu nije reprezentativna, ali ipak može ukazati na uspjeh školskih mjera.

Analiza situacije u vezi s digitalnim medijima u hrvatskom obrazovanju mora se uvrstiti u kontekst međunarodnog razvoja. U toj usporedbi moraju se uzeti u obzir postojeći, dijelom veoma različiti okvirni uvjeti pojedinih zemalja. Međutim, iz analize međunarodnih studija, metaanaliza, preporuka i Delphi-studija mogu se izvesti korisne spoznaje o daljem razvoju situacije u Republici Hrvatskoj.

Stavovi. Kad je riječ o međunarodnim usporedbama, u Njemačkoj su primjerice naročito izražene razlike u stavovima između dječaka i djevojčica. U skandinavskim zemljama pozitivni stavovi roditelja, nastavnika i učenika mogu se usporediti s pozitivnim stavovima prema digitalnim medijima u društvu uopće. Tamo je integracija digitalnih medija znatno napredovala. Roditelji u Danskoj, Švedskoj, Finskoj i Norveškoj imaju izražen pozitivan stav prema digitalnim medijima. Broj nastavnika koji osjećaju pozitivan učinak na uspjeh učenika veći je od broja učenika. Utvrđeno je da nastavnici i učenici u tim zemljama slično procjenjuju učinke.

Opremljenost. Usprkos radu na informacijskom opismenjavanju, opremljenost Hrvatske računalima je u razmjeru 21 učenik:1 računalo. U usporedbi s međunarodnim razmjerima, Hrvatska je među zadnjima. Razmjer 6:1 definira se kao granica kod koje je u međunarodnim regijama s daleko naprednijom IT- integracijom postignut kvalitativan skok. Osim toga, značajni su čimbenici: pristupačnost na svakom mjestu i u svako doba, integriranje računala u učionice, podrška i redovna aktualizacija, odnosno obnavljanje. U odnosu na kućnu opremljenost, Hrvatska se također nalazi ispod OECD – prosjeka, pri čemu mladi iz domaćinstava s nižim socijalno-ekonomskim statusom rijetko kod kuće imaju pristup računalu.

Korištenje. S obzirom redovito korištenje digitalnih medija u školi, 15-godišnjaci u Hrvatskoj nalaze se među zadnjima u OECD-usporedbi. Kućno korištenje računala je očigledno u okviru OECD- prosjeka, što za posljedicu ima razliku između školskog i kućnog korištenja računala petnaestogodišnjaka u Hrvatskoj kad je riječ o usporedbi s međunarodnim podacima.

Public Private Partnership. Čini se da su Public-Private-Partnership programi na školskoj, regionalnoj ili državnoj razini dobra mogućnost poboljšanja infrastrukture u školama. To pokazuju primjeri u SAD-u, Engleskoj i Švicarskoj. Pritom se u aktualnim evaluacijama ističe važnost koordinacije i podrške takvim mjerama na državnoj razini. Značajnu ulogu u okviru ovakvih programa ima ustrajnost kao i podjednako uvažavanje razvoja sadržaja i usavršavanja nastavnika.

Učinci. Osim pokušaja mjerenja učinaka poboljšanja ishoda, neke zemlje danas potenciraju promjene procesa učenja primjenom samoocjenjivanja. I u zemljama u kojima integracija već dugo napreduje, vidljivo je da su i učenici i nastavnici osjećali trajne učinke u odnosu na sudjelovanje, individualizaciju, kreativnost i učinkovitost vremena učenja. Osim toga, ne odnose se utvrđeni učinci samo na tumačenje stručnih sadržaja nego i na promjene kulture nastave, npr. u obliku većeg uvažavanja heterogenih preduvjeta unutrašnjom diferencijacijom. Ove promjene treba popularizirati.

Razvoj škole. S obzirom na utjecaj na razvoj škole, mogu se na međunarodnoj razini identificirati kao čimbenici uspjeha i: djelotvoran i učinkovit, prema potrebama svih sudionika usmjeren menadžment školskih, razvoj osoblja i visok stupanj školske autonomije.

Obrazovanje nastavnika. Ključne strategije u okviru obrazovanja nastavnika su realizacija pojedinih tečajeva i radionica, primjena multimedijalnih elemenata, ostvarivanje partnerstva s školama uz uključivanje mentora, obrada primjera i njihova primjena u praksi, visokoškolska didaktička usavršavanja i poboljšanja infrastrukture kao i pokušaj integriranja u seminare. Pritom se kombinacija što više strategija nameće kao ona koja obećava uspjeh.

Za usavršavanje nastavnika preporučuju se sustavne aktivnosti koje treba započeti na različitim razinama. Ovdje sve veću važnost dobivaju školska interna usavršavanja, koja se orijentiraju na potrebe sudionika i okvirne uvjete na

mjestu događanja, individualna savjetovanja ili Co-Teaching i posebni instrumenti, primjerice digitalni portfolio.

Istraživanje. Na međunarodnoj je razini utvrđeno da postoji daljnja potreba za istraživanjem u svim do sada spomenutim područjima, pri čemu se posebna važnost pripisuje istraživanju procesa učenja, a manje outputima i razvijanju djelotvornih mogućnosti obrazovanja i usavršavanja nastavnika kao i rad na instruktivnim mjerama. Kultura nastave i socijalna pitanja identificiraju se kao važna područja. Pritom se osim kvantitativnih mjerenja naglašava važnost kvalitativnih postupaka kao i kombinacija obaju procesa. Metodički se ističe nužnost provođenja longitudinalne studije u okviru istraživanja nastave i povezivanje raznih disciplina kako bismo mogli što bolje identificirati složene čimbenike utjecaja. Takvu važnost u obrazovnim procesima valja na odgovarajući način uvažavati. Ne samo školsko nego i predškolsko, stručno obrazovanje, daljnje obrazovanje i usavršavanje – ne mogu se više zamisliti bez digitalnih medija kao sredstava i instrumenata poučavanja i učenja, predmeta analize i intenzivnog razmišljanja u didaktičkom i odgojnom smislu. Valja još dodati da su digitalni mediji u izvaninstitucionaliziranom kontekstu postali sustavni dio neformalnih procesa učenja, a time i logičan dio iskustva.

Obrazovno-odgojni ciljevi. U okviru školskih i izvanškolskih obrazovnih nastojanja vrijeme je za osposobljavanje djece i mladih za primjereno, autonomno, kreativno i društveno odgovorno djelovanje. Ta osposobljenost znači da pojedinac sebi pomoću digitalnih medija može približiti svijet i istodobno steći uvid u promjene uvjetovane digitalnim medijima, odnosno tehnološkim razvojem. Taj se zadatak ne postavlja samo za pojedina obrazovna nego i za interdisciplinarna područja.

Tehnološki razvoj. Informacijsko-komunikacijska tehnologija je relativno mlada disciplina koju karakteriziraju izrazito kratki inovacijski ciklusi. U skladu s tim u tom se području ne može govoriti o završenim razinama razvoja, koji se onda primjenjuje u radu u školi. Tu je riječ o stalnom procesu inovacija koji mora biti u povratnoj vezi s obrazovnim procesima. Zemlja s visokim tehnološkim know-how ne može od toga razvoja odvojiti obrazovne procese. U izvanškolskim područjima već postoji uska povezanost. Tako su primjerice u području komunikacije mobilni telefoni za kratko vrijeme postali sustavni dio promijenjenog kulturnog iskustva i koriste se kako bi se zadovoljile potrebe u kontekstu komunikacije i neformalnog učenja.

Cjeloživotno učenje. Ukupni razvoj upućuje na to da određene institucije – naročito škola – više nemaju monopol u procesu učenja, nego se učenje odvija na različitim mjestima i na način koji odgovara individualnim potrebama. Pritom učenje nije više prilagođeno samo određenoj kvalifikaciji već je cjeloživotni proces. U osnovi je takvog shvaćanja predodžba o čovjeku kao aktivnom i samostalnom biću, koje se stalno trudi razotkriti svoju okolinu u procesima percepcije, konstrukcije i komunikacije i sudjelovati u njezinu oblikovanju. Ovaj

proces cjeloživotnog učenja digitalni mediji potiču, a onda ga na prikladan način podržavaju i prate.
Upravo zbog toga što se učenje sve više odvija u takozvanom neformalnom okruženju, važno je pružiti različite mogućnosti onima koji nisu u sprezi s institucionaliziranim kontekstima ili ih ne mogu percipirati zbog drugih ograničenja. Odgovarajući širi utjecaj može se postići sadržajima na Internetu. Internet je istodobno "vidno" mjesto, na kojem je moguće intenzivno se baviti digitalnim medijima u području obrazovanja. To s jedne strane znači stavljanje na raspolaganje sadržaja iz struke nastavnicima i učenicima, a s druge strane korištenje same računalne mreže za prenošenje inovacija.
Obrazovanje na mrežu – obrazovanje preko mreže. Digitalni mediji nisu sami sebi svrha i ne bi privukli takvu pozornost u kontekstu obrazovanja da znatno ne utječu na privatni i profesionalni život. Razumljivo i prihvatljivo je što se procesi implementacije fokusiraju najprije na opremljenost i bavljenje tehničkim okvirima digitalnih medija. To će s obzirom na dalji tehnološki razvoj i nadalje biti potrebno, mada se zbog sve veće primjene i interoperabilnosti kao i nižih cijena novih aparata može očekivati rasterećenje. Zbog svega rečenoga, nužno je obrazovne procese utemeljene na digitalnim medijima staviti u prvi plan. Riječ je o pokretanju i podržavanju obrazovnih procesa preko mreže i s digitalnim medijima i to za razne ciljne grupe. Vrijeme je za daljnji razvoj na širem planu, nužno je dalje razvijati odnosno osigurati infrastrukturu, naročito u obliku djelotvornih širokopojasnih mreža.

Zaključak

Nužan je interes ljudi, a ne samo državnih razina, za osposobljavanje za kompetentno korištenje nove tehnologije – kako u pogledu primjerenog, svrsishodnog korištenja, tako i u pogledu kritičkog odnosa prema tehnici. Od industrije i privrede traži se sudjelovanje u obrazovnim procesima i podržavanje razvoja nove generacije. Ovo nije samo ekonomsko pitanje, nije riječ samo o obrazovanju novih potrošača nego je i pitanje društvene odgovornosti. Trenutna situacija s obzirom na primjenu digitalnih medija u školi nije zadovoljavajuća. O integraciji digitalnih medija u školske nastavne procese ne može se ništa zaključiti jer potencijali digitalnih medija nisu još iscrpljeni. Posljednjih je godina došlo do poboljšanja, ali sadašnje stanje ne zadovoljava ni u pogledu opremljenosti ni u pogledu primjene u obrazovnim procesima. Digitalni se mediji danas u Hrvatskoj koriste uglavnom u kućnom okruženju, a tek onda u školi. Veliki broj djece i mladih kod kuće ima pristup računalu, a to još uvijek izravno ovisi o društveno-ekonomskom statusu. Na međunarodnoj razini, najveća razlika između učestalosti školskog i kućnog korištenja ukazuje na to koliko je nužno spojiti institucionalizirane i neformalne procese korištenja medija i učenje pomoću medija.
Očigledan je pozitivan opći stav u odnosu na digitalne medije. Svi sudionici procesa obrazovanja u školi, učenici, nastavnici i roditelji – pripisuju digitalnim medijima veliku važnost i otvoreni su prema njima. Sada je nastupilo vrijeme za

podizanja osnovnog raspoloženja – i konstruktivno i dosljedno nastavljanje uspjeha zabilježenih u radu s digitalnim medijima. Integracija digitalnih medija u obrazovne procese otežana je zbog nesrazmjera između sporosti procesa inovacija u obrazovnim institucijama i brzine ciklusa tehnoloških inovacija. Ova diskrepancija ima za školu dalekosežne posljedice. Zbog medijskog okruženja u kojemu se učenici kreću postoji sve veća opasnost za njihovo udaljavanje od školskih medijskih okruženja.

U raznim područjima, primjerice u razvoju obrazovnog softvera za učenje ili u obrazovanju za stjecanje zanimanja, postignuti su dobri rezultati. Ipak i sada, kao i u prošlosti, nedostaje široko i stalno utemeljenje. U mnogim studijama pojedinačnih iskustava s obzirom na pozitivne učinka u korištenju digitalnih medija u nastavnim procesima, naši rezultati odgovaraju međunarodnim podacima. Ako Hrvatska želi svoj obrazovni sustav učiniti konkurentnim, i dalje je potreban pojačana aktivnost u tom smislu. Digitalni su mediji integralni dio našeg života i imaju posebnu ulogu u raznim fazama razvoja i obrazovanja. Paralelno s tim razvojem mogu se – s obzirom na rezultate naše studije i ocjene stručnjaka – otkriti različita područja u kojima postoji potreba za istraživanjem i djelovanjem. Kao drugi korak moraju se s obzirom na nadležnosti razvijati mjere koje mogu dovesti do svladavanja opisanih zahtjeva i izazova.

Literatura

Considine, David: Approaches to Media Education. Internet-document. URL: http://www.media-awareness.ca/eng/med/bigpict/medapp.htm#1, 1994.

Council of Europe: Report on Media Education. Internet-document: http://stars.coe.int/doc/doc00/edoc8753.htm, 2000.

European Community eEurope Initiative: An Information Society For All. Internet document: http://europa.eu.int/comm/information _society/eeurope/index_en.htm 2001.

European Centre for Media Competence. URL: http://www.ecmc.de/

Ferguson, Robert: Global Interculturalism and the Dilemmas of Universalism: Teaching the Media after 2000. A paper to be delivered to the International Congress on Communication and Education, Sao Paulo, Brazil. Internet-document: www.eca.usp.br/nucleos/nce/pdf/140. pdf, 2000.

Hart, Andrew: Paradigms Revisited: Media Education in the Global Village. In: Media Education Centre. Frankfurt 1997.

Joint Hearing between the European Parliament and the Parliamentary Assembly of the Council of Europe: Media Education. Internet-document: http://www.europa.eu.int/geninfo/whatwas new/032000.htm., 2000.

Kumar, Keval. J.: Media Education in India: Strategies, Trends, Visions. Stockholm Media Education Conference. Internet-document. URL: http://home.swipnet.se/~w-52050/konf_keval_kumar.htm, 2000.

Lange, Bernd-Peter: Media Competence in the Age of Global Communication. Internet-document. URL: http://sef-bonn.org/sef/publications/pol-pap/no11/content.html, 2000.

Latzer, Michael: Mediamatik – Die Konvergenz von Telekommunikation, Computer und Rundfunk. Opladen, 2000.

Latzer et. al.: Die Zukunft der Kommunikation. Phänome und Trends in der Informationsgesellschaft. Studienverlag Innsbruck – Wien., 1999.

Nadrljanski, Đorđe: Obrazovni softver, Univerzitet u Novom Sadu, Tehnički fakultet "Mihajlo Pupin", Zrenjanin, 2000.

Nadrljanski Roguljić, Mila: Metodološko-metodički modeli obrazovanja na daljinu – doktorska disertacija Sveučilište u Zadru 2006.

Instructional System Design http://www.nwlink.com/~donclark/hrd/sat3.html

Tomašević, Marko; Stanivuk, Tatajana; Ristov, Pančo: Obrazovni softver – kompjutor u matematici kao faktor za preskakanje nedostataka tradicionalne nastave, Konferencija “Društvo i tehnologija” Split 2007.

Progress and Diversity – New Paradigm of Development

Stjepan Mateljan
Department of Information Sciences
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
stjepan.mateljan@gmail.com

Hrvoje Stančić
Department of Information Sciences
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
hrvoje.stancic@zg.t-com.hr

Summary

Cultural diversity is important feature of humankind. Globalisation is latest incarnation of imperialism. Technological progress tends to replicate exact copies, instead to manufacture original functional solutions. In order to keep power, information society standardises world at it's own image. Homogenization of world shifts balance between humankind and nature. Mankind is at unique spot in it's history, not only as a cultural, but also as a biological being. Extinction of wildlife and extinction of languages are two faces of the same process. Technological progress focuses on narrow aspect of the world, while polluting all the others. We can separate technological progress from development. Technological progress is nomadic, and it's property is constant expansion on new 'feeding grounds'. Sustainable development is more evolutive, and focused on expansion in different niches of the same environment.
Technology tends to make world standardized and easy to control, while neglecting it's chaotic behaviour tendencies. From time of industrialization onwards the human impact on the Earth grew, and now is considerable factor of disbalance. Intellect is a limited problem solver. Digital divide is a great problem. While everyone is allegedly invited, to actively participate in the life of information society one must have infrastructure within reach, and some resources to either bid or use while participating. Equal standards in different contexts bring social stratification. Paradoxicaly, the greater the homogenization, the greater the social differences.

Keywords: cultural diversity, digital divide, progress, development, homogenization, meme

Neglecting the need for new paradigm – escapism

The Internet and the World Wide Web have connected human brains in real time. Online communication via computer-changed relationships between people, time and space, and it is not exaggeration to say that at this very moment humankind crosses one of the basic evolution steps. Via the Internet every innovation, every creative thought, every fix of any problem becomes instantly reachable to everyone online. The consequence is progress at the rate never seen before. While knowledge expands at more or less normal rate, expansion of information looks more like an explosion.

The Internet dramatically changes the system of values we are accustomed to. In rich countries, social, political and economical structures are changing from the roots, and that has a lot of impact on the global scale. The need for the resources that forced people to gather around places where they could find them has started to loose meaning. Just as global industry has reoriented during the 1930's from the manufacture to the specialist structure of production, big companies of today are making massive reorganizations, cutting workplaces in production and opening them in fields that are more intellectual. Information societies of today are inducing virtualisation, the process where the value of things is shifted from matter into information. Now, greater worth is given to the context of some good than to that good by itself – somehow similar as it was at the end of the medieval age with replacement of economy based on commodity money with one based on fiat money. This great step changes our relationship towards either society or nature. In the information society, this idea has more market worth than the real thing. From fashion to marketing, sign is everywhere, and people are willing to pay more for it. Today, people do not just buy an item – they buy a sign, an image, the context of an item; they buy information attached to an item. As our prehistoric ancestors had lived within nature, coexisted with it, as we have created civilization as its subset and we are living within, and coexisting with civilization, there is a possibility that our descendants will create a subset of civilization in cyberspace, live within cyberspace and coexist with it. Internet is both the creator and the destroyer of our world. While some people have predicted the golden age of our species in some unknown time before, and have feared that through the history humankind would progressively degenerate, others had an evolutional point of view, where human species would develop in the direction of greater and better self. That is, for now, on horizon of only 17.8% of humankind that have connection to the Internet (or, in Africa, only 3.6%) –presuming that while having a connection, one also have resources to spend or to use them in an adequate or quality way.

The digital divide is the gap between those with regular, effective access to digital and information technology, and those without this access. It encompasses both physical access to the technology and, more broadly, skills and resources, which allow for its use. It widens the gap in economic divisions around the world. Countries with a wide availability of Internet access can advance the

economics of that country on a local and global scale. In today's society, jobs and education are directly related to the Internet. In countries where the Internet and other technologies are not accessible, education is suffering, and uneducated people cannot compete in our global economy. This leads to poor countries suffering greater economic downfall and richer countries advancing their education and economy.

Diversification of uniform world

The main problem lies in hypocrisy of equal rights for all if no regard is taken to the starting conditions. Everything orbits around power – even if civil society do rule over the Internet, what part of it will take the decision making role? If national states do loose control over the media, who could guarantee that the new rulers will not have the same old characteristics and that same old problems will not emerge in different forms. National states will eventually lose control, and international bodies will rule in the name of all humankind – but who could guarantee that old issues will not emerge in new occurrences? We have seen that problem of accumulation of power in hands of individuals equally existed in socialist and capitalist countries, either in context of political, or economical power. Although political power and money are close to each other and often comes in pair, in one system the basis of making that pair was primarily capital, and in the other party membership. When we are talking about arranging a completely new system that will deal with a new unique world, we must think about the idea of progress. Progress has no precise definition: it generally indicates forward moving, without notion of dimension or formation. Something may be good for some, and not for many, or can be good for all in one field, but baneful in many others. When we talk about progress, we most often think about technological and economical progress. Meanwhile we see global ecological crisis grows into ecological catastrophe.

The point is not to stop the development, but to stop the expansion – to stop running always on the new feeding grounds, instead of optimising usage of the old ones. Human species once again must cease to be nomadic, and settle down. Choice of our cancerogenous growth is to settle down or to fade, along with fading the very environment carrying us. We should not think about preserving diversity like a conservative, anti progress movement – closer to the truth, preservation of diversity would be a new face of progress. We cannot preserve traditional cultures from change, no one can – but we must stop their degrading in mere local variants of only one global culture. The concept of cultural diversity, like that of biodiversity envisages the multiplicity of cultures in a systemic perspective where each culture develops and evolves through the contact with other cultures. Diversity is often perceived as disparity, variation, plurality, that is, the opposite of uniformity and homogeneity. This vision has now been superseded, though, as diversity is not defined in opposition only to ho-

mogeneity but in opposition to disparity as well. It is synonymous with dialogue and shared values.

Free trade is the tool of maximising the extraction of resources from the underdeveloped. More than one third of the world trade is the trade among companies, or even within different branches of the same companies, and it often cannot be clearly distinguished from centralised commerce action with goal to improve income[1]. Every culture has dignity and values that should be respected and preserved. Uniformisation of culture might result from the shaping of the information society as a global society, even if it theoretically enables the manifestation of cultural diversity. This is because the new information and communication technologies, far from only being tools, model our ways of thinking and creating. As a result, culture becomes inhabited by technology, in a dialogue with it, sometimes containing it and allowing itself to be elaborated by it. The situation creates a certain inequality and dependency of culture with respect to technology and prevents the manifestation of mentioned diversity that is so vital to the knowledge society. Technology has excluded a whole part of the world population, which continues to live following the principles of nature, does not believe in the state but in the power of ancestors, and does not believe in science but in the traditional knowledge. Cultural diversity considers that there are other ways of thinking, existing and working than the modern anthropocentric, ratio centric way. Although science and technology are easily communicated, are all cultures prepared to accept the mathematical formalism on which technology and its prescribed uses are based?[2] Cultural diversity is the integration, rather than mixing, herding or any other manipulation of cultures, and the information society in which it is expressed a society of shared knowledge. Diverse society cannot be a society made of hierarchy of different cultures, or society where main culture is made of simple coexistence of different cultures – it must be made of different cultures that grow together organically.

Cultural diversity and sustainable development

We cannot aim not to change. It is impossible – we must aim to slow things down because at current pace our actions are widely considered, as progress will change the world faster that the world can adopt. The leading countries need to lead in order to survive in current conditions. In the constantly changing context, it is hard to tell when the spot for u-turn would be – and they refuse even to slow down to be able to perform that turn. It is hard to contain human aggressive nature, and the way of thinking based on power relations and instant gains. Globalisation as it is may be the latest incarnation of imperialism, but it needs to change in order for all humankind to survive global changes. People

[1] Chomsky, p. 229.

[2] Kiyindou.

often forget that all humankind and all its power are only part of world, and that we are far more fragile than the whole Earth. The global ecosystem finds its balance by disposing biomass, counting in percentages and not even considering species. Old system of thinking where the world is vast and resources are infinite is now obsolete. The limits are clear, expansion of human shrank the world, and we are feeling tight – but we have nowhere to go. Even at its best, our progress did not made us greater, only made the world smaller. To grow greater is to grow better, not to replicate big quantities. While we live on Earth as human species, we truly exist as humans in the context of humankind. We must preserve humankind, and balance with the rest of the planet. As we are part of the planet, same doctrines of making balance will be possible to translate to human society. The global information society of tomorrow must be a society of balance. The development needs to be sustainable. UNESCO defines sustainable development as the development that balances the fulfilment of human needs with the protection of the natural environment so that these needs can be met not only in the present, but also in the indefinite future. Its focus is on economic, environmental and social issues. Cultural diversity is closely connected to sustainable development. The Universal Declaration on Cultural Diversity elaborates on the concept of cultural diversity stating that "... cultural diversity is as necessary for humankind as biodiversity is for nature"[3]; it becomes "one of the roots of development understood not simply in terms of economic growth, but also as a means to achieve a more satisfactory intellectual, emotional, moral and spiritual existence". In this vision, cultural diversity is the fourth policy area of the sustainable development, along with economic development, social development, and environmental protection. By analogy with biodiversity, which is thought to be essential to the long-term survival of life on the Earth, it can be argued that cultural diversity may be vital for the long-term survival of humanity, and that the preservation of indigenous cultures may be as important to humankind as the preservation of species and ecosystems is to life in general. In that way of thinking, progress and development cease to be synonyms. In sense of diversity, there are no more or less developed societies, and no referential models of development. For instance, if we consider development to be property of expanded activity in ever changing context of existence while finding the way to preserve balance with active environment while progressing, then technological progress can in most cases be considered exactly the opposite of development. At the same time, the growth of technology can be considered as the mechanism for damage control that is always present in the progress of technology. Technology deals with the deterministic world. Hence, the uniform, the tendency to copy, to unite, and the need for centralised control. Reality comes in variations. Biodiversity exists because if all life would be the same,

[3] UNESCO, 2001.

ideal, and shaped once and not forever changing, only one niche will be populated, of all potential resources only one will be taken, and because of similar metabolism, every disease would be pandemic. Nevertheless, metabolism is intelligent, and ever changing to fit the context, so it is quite impossible for some kind of biodiversity not to develop instantly when we put two identical living materials in the different contexts. Ignoring the environment and consecutively mass-producing exactly the same replicas is a significant human feature – even bacteria within the same colony evolve to fit the context. There are no two same spots. One is always shifted either in space or in time from the other.

Artificially homogenised world

From the start of global expansion of European culture, the world witnessed significant homogenisation. Biodiversity shrank, either by destruction of habitats, or by accidentally or intentionally bringing exotic species to the new habitats. Cultural diversity shrank by establishment, growth and automation of industrial, medical, educational, economic and administration systems. That shrinking was often called technological progress. Some say that the world has lost its variability and became indigenous, and in its indigence became somehow similar to the simple, easy to understand, easy to plan, deterministic world of ideas. But the world is not simple, and it is not easy to understand and to envisage all it’s possibilities – only it’s richness and diversity that make it so resistant to catastrophe created illusion of it as a stabile, easy to manage phenomenon. While the minor changes accumulate, once there are enough of them to break the limit of sustainability the phase shift can be relatively quick. Irish Potato Famine of the 19th century is an example of playing with biodiversity, and two great wars of the 20th century are an example of playing with cultural diversity. Prior to both wars, big populations had only limited ways to act while meeting events very different and complex problems. Diversity is the problem solving issue, and it is a better way to track richness than accumulation. We are primarily biological beings, and only after we are able to meet our biological needs we exist as cultural beings. If we want to preserve our culture as it is, we are doomed because we cannot preserve culture that destroys our biological base. We do not need to regress, but to reorganize. As we know, world hunger is not product of scarce food resources, but of human made divisions. Similar is with overpriced medications, population growth, global warming and extermination of species and languages. While we cannot control if some asteroid collide with the Earth, we are fully responsible if we die out of pollution. Global nuclear war was too powerful, too violent, too concentrated to be ever carried out, and measures to counter it were very high. On the other hand, pollution is sneaky, prolonged in time, often hard to notice until it is too late. Pollution is what monoculture, a culture that focuses only on one aspect of environment, on only few species, only few ways of thinking or on only few languages, produces. We can consider acts of economic propaganda to be aiming at lowering

the diversity of culture, thus trying to homogenise humankind, to produce human monoculture out of all of the humankind, in order to diversify market norms. Economy is chief enemy of cultural diversity, and from economy comes war, poverty, pollution, hunger, overpopulation and the digital divide. In order to preserve the world, we must stop thinking about the profits, personal gain and short term interests.

Big areas of monoculture, either of natural or human kind, are fragile to the influence, hard to keep equilibrium with ever changing context, and keep most of available space empty while overpopulating, and thus damaging, that one little part they inhabit. It is impossible to have a stabile monoculture over big area for a long time. A few centuries is only a blink of time of life on Earth, and while our systems of homogenizing the world are reaching the limit where they will simplify the world beyond the lowest sustainable level of diversity, thus damaging the part over the limit, the world will collapse. While we were simplifying actions based on understanding the world, the complex issues have accumulated. The exponential growth of science discoveries is closely followed by the growth of technological progress – and technological progress is closely followed by its economical impact. Same as the whole Newton's physics is a special case of physics of relativity, Adam Smith's 'invisible hand' can be a special case of new economy theory of sustainability. Key of survival of the humankind is to act towards infinite development and not, as today is the case, towards infinite growth. Growth is simply a matter of numbers, a feature of replication.

While human industrial, medical and school systems caught routine, became huge and have made profits of big scale economy, in natural sciences blossomed linear equities, and in social sciences theories of equilibrium. Although the world is far from being in any sort of simple linear equilibrium, its homogenisation makes its uniform made areas to start to act as linearly equilibristic structures with properties that are possibly to predict and control. Western societies turned objective world (or some of its parts) into type of structure that fits their theories, and so their theories became part of some sort of self-fulfilling prophecy.[4] If we consider the second law of thermodynamics, where entropy tends to grow forever, western type of control over the world is to focus on one aspect to the cost of all the others. Diversity is not entropy – it is the way to live in the real world. Static infinite world is a model, not the real thing. Evolution is a way universe deals with growth of entropy. Maximised entropy always corresponds to the apparent homogeneity in a system. Any random disturbance of a homogeneous system results in no meaningful change, and entropy grows. We can never predict what will happen with 100% accuracy. Our current theories start to implement nonlinear elements, so we are starting to think of heterogeneity as of something worthy, and not as of obstacle of unifi-

[4] De Landa, p. 394.

cation. To old linear concept of causality, we are adding knots of positive and negative back circuits, and by that, we are lowering the level of homogeneity of our theories. After we develop theories that will be able to support change of paradigm, we will have to deal with homogenised gene pools of our domestic plants and animals, and homogenised meme pools of people worldwide.[5] Meme is a unit of information residing in the brain and is the mutating replicator in human cultural evolution. It is a pattern that can influence its surroundings and can propagate. Examples are tunes, catch-phrases, beliefs, clothing fashions, ways of making pots, and the technology of building arches.[6]

Facing the need for the new paradigm – strategy

On one hand, there is a biodiversity that is best presented by variability of species, and species are presented by different contents of gene pool. On the other hand, there is a cultural diversity that is best presented by variability of cultural identities, and cultural identities can be presented by different contents of meme pool. If we want to make real plans for diversifying the world, we must think about the world as a whole. While our world is only virtually relaxed, balanced and open, homogenisation will continue to destroy its richness. As we saw from UNESCO's definition of sustainable development, its goal is to balance the fulfilment of human needs with the protection of the natural environment so that these needs can be met not only in the present, but also in the *indefinite* future. Western world have old tradition of closing itself in a cocoon, and calling inner semi virtual world reality. To limit time or space by straight, understandable line is to treat outer world as the virtual world. Thought, either rational like in science, or irrational like in faith is developed fundamentally as a tool of our survival in the time preceding civilisation. With the advance of civilisation to its technical state we not only survived but started to present danger to other species, while still continuing to progress. The world is random, and the context is always changing, and in the new context, human accumulation of resources and power is danger not only for some species, but also for the whole ecosystem. Often real borders exist only for virtual reasons, and reasons can be virtual for the fact that theories under them are plain wrong or if theories under them are not updated. Intellect is a limited problem solver. The border between human species and ecosystem is a non-existent one, but its existence in human mind ultimately tends to produce the gap between the humans and the nature without concern about the prolonged stability of the world under that paradigm. It is a common misconception between a moment and infinite – while the world is changing constantly (at any moment), and we want to preserve the humankind as long as possible (aiming at infinite), we cannot save it constantly with always

[5] De Landa, p. 395.

[6] Dawkins, p. 245.

the same methods. I.e., if we do not want to evolve in 'outer', real world physically, than we must alter our inner world, the world within us – we most evolve mentally. All our mental activity fitted best in the previous context, and with the context changed the need for a new type of mental activity rose. If we want to survive, we must face the new need or the form of life that fits in the new context better will replace us. We need to break the walls between the humankind in nature, and within the humankind itself. The digital divide can be traced back to the same way of thinking. To focus on just one aspect of the world is, after all, more costly and the world of poor diversity is hard and expensive to sustain – quite opposite to its virtual counterpart. Paradoxically, the higher the homogenisation – the higher the social differences.

Fortunately, we have tool to keep culture diverse while we are slowing down the advance of technological progress and develop in direction of sustainability. The new paradigm should be decentralisation, and the ideal tool of global decentralisation is the Internet. Development of the global network for information exchange can be represented, at its core, as development of a nervous system of living Earth. Internet is mostly self-organised, and as in all self-organized structures, it shows emergent properties. If we look at the Word Wide Web, we can see similar pattern as in the whole Internet: there is no central organisation rationing the number of links, yet the number of links pointing to each page follows the law which says that a few pages are linked to many times and most pages are seldom linked to. A related property of the network of links in the World Wide Web is that almost any pair of pages can be connected to each other through a relatively short chain of links. Although relatively well known now, this property was initially unexpected in an unregulated network. Internet is a neural network, and its size can provide some sort of hive mind intelligent decision-making, and potential development of consciousness, similar as intelligence of every human being is made from the group activity of all of its nervous system. Ant nest does not have a leader and still fulfil all the wishes of ants as its basic components. We must take into account that most of wishes of the western people are produced by technological progress. While education system shapes unbalanced people for support of unbalanced society, unbalanced wishes occur. People of technical civilisation rarely want to make their existence broad, but they make their existence focused. We cannot evolve while remaining the same, but that is what we wish. Maslow is talking about hierarchy of needs, not wishes. The key to override this troubled path is to change the mind of people. We need to break boundaries between people and then the natural homogenisation will come as we will see that we all have the same needs, similar but different in the same way we are not replicas. Self-organisation always lies in line between order and chaos, in equilibrium, i.e. water below the point of freezing will turn to solid state, and above that point will turn to liquid

– pattern of snowflake or frost emerges when the water is in chaotic shift between that two phases.[7]

Capitalism tends to negate democracy, and democracy is what we need. Not to choose main language, or currency, or to create mainstream from any one fixed spot, but to let people organise it by their needs. Equal standards in different contexts bring social stratification. Self-organisation of the Internet is a good example. Some architectural features always emerge. The important feature to keep in mind is the lack of a central spot. It is a hard goal to achieve, because we have evolved as social mammals, and we have instincts of herd and carry our hierarchy all the way from reptiles – but the Internet provides us with the way to develop hive mind. Modern democracy of representatives is more like a herd mind, where representatives present herd leaders. The world is cleaning itself, because if we succeed, the thought will itself become defocused and diverse, and with no need of human brain to carry it. Similar to the Gaia theory, the Internet can be considered as a nervous system of the planet.

Significance for development of the humankind is enormous, because we can clearly see what humans can achieve with the persistent work. It is known that human being is not born, but made. It is possible that we are living in the age of transfer, where the same process is happening on the level of population. Evolutive stabile strategy is not expansion any more, but sustaining. In the context of today, yesterday's wealth of accumulation looks more like cancerous swelling. Wealth of today is in diversity, not because of prestige, but because of sustainability. It is important to differentiate between diversity and accumulation of great range of units (like web sites in web searchers, artefacts in museums, 'small' cultures, 'small' languages, and all types of memes). Harmony of coexistence of units within the system is important, not just plain tolerance between them. In the Internet, old ways of thinking impact users in at least two ways: by long lists of tags on web pages and consequently by long lists of pages recalled with searcher with different design but basically the same information about the wanted term. If we want to overcome our biological base and become vectors of humanity in this world (like we are among our fellow people), we should start first by reconfiguring ourselves. As the usage of written word while searching meme over the Internet provides full service only to people adopted to the written way of expressing, we should start by providing Internet to each and all, and then by changing the doctrine of Internet navigation techniques.

[7] Bentley, p. 124.

References

A Little History of the World Wide Web, <http://www.w3.org/History.html>, 15.8.2007.
Baudrillard, Jean: Inteligencija zla ili pakt lucidnosti. Naklada Ljevak, Zagreb, 2006.
Baudrillard, Jean: Simulacija i zbilja. Naklada Jesenski i Turk, Zagreb, 2001.
Bentley, Peter J.: Digitalna biologija. Izvori, Zagreb, 2004.
Chomsky, Noam: Stari i novi svjetski poretci. Naklada Jesenski i Turk, Zagreb 2004.
Chamacho, Kemly: Digital divide, <http://www.vecam.org/article549.html>, 21.8.2007.
Dawkins, R: Sebični gen. Biblioteka Zodijak, Beograd, 1979.
De Landa, Manuel: Tisuću godina nelinearne povijesti. Naklada Jesenski i Turk, Zagreb, 2002.
Digital Divide – Ushering in the Second Digital Revolution, <http://www.digitaldivide.org/dd/index.html>, 22.8.2007.
Digital native**, <**http://www.digitalnative.org/Main_Page>, 10.8.2007.
Internet World Stats, <http://www.internetworldstats.com/>, 11.8.2007.
Internet Systems Consortium, <http://www.isc.org/index.pl>, 17.8.2007.
Information Society Watch, <http://www.is-watch.net/taxonomy_menu/12/92>, 20.8.2007.
Kiyindou, Alain: Cultural Diversity, <http://www.vecam.org/article597.html>, 21.8.2007.
News on WSIS, <http://cryptome.org/wsis121103.htm>, 20.8.2007.
Nielsen, Jakob: Digital Divide: The Three Stages, <http://www.useit.com/alertbox/digital-divide.html>, 22.8.2007.
Quantcast – Open Internet Ratings Service, <http://www.quantcast.com/>, 17.8.2007.
The Knowledge Management Forum, <http://www.km-forum.org/t000008.htm>, 10.8.2007.
The Sociotechnological Singularity, <http://pespmc1.vub.ac.be/SINGULAR.html>, 16.8.2007.
UNESCO Universal Declaration on Cultural Diversity, <http://www.unesco.ru/files/docs/unesco_universal_declaration_on_cultural_diversity_eng.pdf,> 20.8.2007.
Word Matters, <http://www.vecam.org/article698.html>, 21.8.2007.
World Summit on Information Society, <http://www.itu.int/wsis/index.html>, 22.8.2007.

Sažetak: u teoriji i praksi

Senada Dizdar
Odsjek za komparativnu književnost i bibliotekarstvo
Filozofski fakultet u Sarajevu
Franje Račkog 1, Sarajevo, Bosna i Hercegovina
senadadizdar@gmail.com

Sažetak

Promjene koje su informacijske tehnologije unijele u proizvodnju dokumenata donijele su i probleme u pronalaženju i pristupu digitalnoj građi, zbog čega se ponovo preispituju mogućnosti definiranja elemenata koji se koriste u identifikaciji i opisu dokumenata. Kako je sažetak jedan od bitnih elemenata pogodnih za uključivanje među bibliografske elemente, cilj ovaga rada je da ga sveobuhvatnije predstavi, te da sažetak kandidira da, zahvaljujući njegovoj standardiziranoj formi, on postane nezaobilazan element u identifikaciji i opisu dokumenata. Mada još od druge generacije online kataloga nema tehničkih problema sa uvođenjem sažetka u bibliografske baze podataka (MARC/UNIMARC polje 300), ta praksa ugradnje ovog elementa nije uobičajena. Najbitniji razlozi zbog kojih sažetak, ipak, treba da postane sastavni dio kataloškog zapisa su njegova stroga struktura, prepoznatljive vrste, definiran stil, propisana dužina, odnosno standardizacija koja se primjenjuje u njegovoj izradi. Sve su to osobine zbog kojih se sažetak pokazao pogodnim za uključivanje u standardizirane sisteme svih vrsta. Osim toga, ugradnja sažetka u opis doprinosi poboljšanju analitičke obrade dokumenata i proširenju pristupnih tačaka u pretraživanju, te poboljšanju upravljanja elektronskim dokumentima u mrežnom prostoru.

Ključne riječi: sažetak, standard za izradu sažetaka (ANSI/NISO Z39.14.) proširenje bibliografskog opisa

Promjene koje su informacijske tehnologije unijele u proizvodnju dokumenata donijele su i probleme u pronalaženju digitalne građe i pristupu digitalnoj građi. Stalna pitanja koja su postavljana pri opisu dokumenata: kako osigurati trajne i valjane opise dokumenata, kako opisati veze među dokumentima i kako prilagoditi praksu opisa dokumenata tako da bude efikasna u globalnom kontekstu – postala su aktuelna sada više nego ikada.
Svijet "papirnih kolekcija i papirnih zapisa o njima" nezaustavljivo se mijenja. Digitalne tehnologije promijenile su proizvodnju dokumenata omogućivši potpunu integraciju onoga što se inače smatra različitim medijima. One mogu "be-

šavno integrirati" simboličku notaciju (tekst), grafiku, zabilježiti zvuk i pokretnu sliku. Mogu integrirati tekst i opis teksta u jedan objekt, ali isto tako moguće je imati odvojen digitalni tekst i digitalni zapis. Sve je to potaklo i propitivanje valjanosti tradicionalnih bibliografskih pomagala i pitanja u kolikoj mjeri ta pomagala mogu da odgovore postavljenim zadacima. Da bi se povećala efikasnost u pretraživanju nudila su se razna rješenja. Jedno od rješenja je bilo i obogaćivanje zapisa dodatnim elementima koja se do sada nisu javljala u katalogu, ali su već odavno poznata službama za izradu kazala i sažetaka.

Na tom području biblioteke se susreću i sa službama za izradu sažetaka i kazala. Oduvijek je postojala određena razina međuovisnosti između opisnih zadataka koje su obavljali bibliotečki katalozi i usporednih zadataka koje su obavljala druga opisna pomagala, osobito ona koja se obično nazivaju pomagala za izradu sažetaka i kazala. Stvarnost u kojoj se nalaze biblioteke uvjetuje da katalog *per se* može osigurati samo jedan nivo pristupa bibliotečkom fondu. Većinom se taj pristup usredotočuje na dokument kao takav (to jest dokument kao serijska, omeđena publikacija ili nakladnička cjelina). Tradicionalno smo se oslanjali na proizvođače pomagala za izradu sažetaka i kazala da nam osiguraju analitički pristup sadržaju tih dokumenata. U mrežnom okruženju ta analitička pomagala zauzimaju isti virtualni prostor kao i bibliotečki katalog. Stoga postoji mogućnost da se poboljša djelotvornost i korisnost sučelja između bibliotečkog kataloga i analitičkih pomagala koji osiguravaju bitnu funkcionalnu dopunu kataloga.[1]

Među analitičkim elementima koji se najčešće kandidiraju za proširenje bibliografskog zapisa jeste sažetak, element koji se pokazao kao veoma efiksano sredstvo u pohranjivanju i pretraživanju baza podataka. Sažetak je posebna (tekstualna) forma koja omogućava prikaz dokumenta i pristup informacijama, a nije bio uobičajeno pomagalo u bibliotečkoj praksi, tačnije u obradi, međutim on postaje sve važniji za pretraživanje i zadovoljenje korisničkih zahtijeva. Otuda je i cilj ovoga rada da sveobuhvatnije predstavi sažetak, koji bi, zahvaljujući svojoj standardiziranoj formi, postao nezaobilazan element u identifikaciji i opisu dokumenata.

Sažetak: definicija

Stoljeće i po nakon proizvodnje prvih štampanih knjiga pojavili su se prvi časopisi, a tek u prvoj polovini 19. stoljeća pojavljuju se prvi referalni časopisi koji donose sadržaje članaka iz časopisa. Sažetak, kao obavezan dio uz znanstveni članak javlja se od sredine dvadesetog stoljeća. Prvi Standard za izradu sažetaka publiciran je sedamdesetih godina prošlog stoljeća, a nakon toga pojavljuju još dva, od kojih je posljednji *Smjernice za izradu sažetaka* (Guidelines for Abstrakt ANSI/NISO Z39.14.) publiciran 1997. godine. I mada nosi naziv *Smjer-*

[1] Delsey, Tom. Preispitivanje konvencionalnih paradigmi za opis dokumenta. // Vjesnik bibliotekara Hrvatske. 46, 1-2(2003), str. 42.

nice, po svojoj strukturi i namjeni ima ulogu standarda, pa će u ovome tekstu biti tako i imenovan, dakle kao standard. [2]

Najjednostavnije kazano, sažetak ili abstrakt je kratak sadržaj nekog dokumenta koji specificira njegov karakter, kao i tehnike, pronalaske, te vrijeme i mjesto istraživanja. Šire, sažetak je posebna vrsta vodiča kroz naučnu literaturu i naučne dokumente.

Lancaster definira sažetak kao kratak, ali sažet prikaz sadržaja dokumenta. Prema ISO stanadardu (214:1976) sažetak na skraćen način prikazuje bitan sadržaj dokumenta, bez tumačenja i bez kritike i bez obzira na to ko je autor sažetka. Za J. Rowley, sažetak je sažeto i tačno predstavljanje sadržaja dokumenta koji je stilski sličan originalnom dokumentu.[3]

Osim sažetka postoje i druge forme prikaza dokumenta kao što su anotacija, ekstrakt i rezime.

Ekstrakt (izvod) je skraćena verzija za dokument, sačinjena od rečenica koje su preuzete iz samog dokumenta. Razlika između ekstrakta i sažetka je u tome što ekstrakt direktno citira autora, a sažetak uključuje riječi koje se pojavljuju u dokumentu, ali su to dijelovi teksta koje je sačinio abstraktor.

Anotacija ili bibliografska napomena je kratak komentar ili objašnjenje njegove sadržine ili pak veoma kratak opis dodat kao napomena poslije bibliografske referense dokumenta. Ona ne sadrži podatke koji su u naslovu ili se iz njega mogu zaključiti, ali služi za bolje razumijevanje naslova dokumenta bez obzira na određene potrebe korisnika. Opisna je, ne mora imati potpune rečenice, a može sadržavati samo takve podatke koji se mogu dokučiti direktno iz dokumenta. Uz to, anotacija može sadržavati samo napomenu o obradi i upute za korištenje.

Rezime je kratko preformulirani sadržaj dokumenta u kome su iznijeti najvažniji nalazi i zaključci iz dokumenta. Rezime obično piše autor i nalazi se na samom kraju dokumenta. Budući da se u rezimeu obično ne navode dijelovi iz teksta kao uvod, svrha ili metodologija, onda se ovi surogati dokumenta ne smatraju sinonimima sa sažetkom.

Još jedna vrsta "kratkog sadržaja" dokumenta je *sinopsis.* To je pojam koji se često koristio u starijim časopisima. Sinopsis je sažetak koji je pripremao sam autor dokumenta. Za razliku od sinopsisa izraz sažetak bio je rezerviran za tekst koji je pripremljen od druge osobe koja nije autor teksta. Vremenom se razlika

[2] ISO 214: 1976. Documentation: abstracts for publication and documentation. Geneva : International Standards Organization, 1976.; Guidelines for Apstrakt. (ANSI/NISO Z39.14.). Merylend : NISO, 1996. publiciran 1997. godine.

[3] Lancaster, F. Wilfrid. Indexing and Abstracting in Thery and Practice. 2.ed. Illionois : Graduate School of Library and Information Science, 1998. str. 94.; Rowley, Jennifer; Farrow, John. Oganizing Knowledge : An introduction to managing access to Information. (3. ed.). Aldershot, England: Grower, 2000, str. 67-70; ISO 214: 1976. Documentation: abstracts for publication and documentation. Geneva : International Standards Organization, 1976.

potpuno izgubila, tako da je danas u upotrebi izraz sažetak, umjesto ranijih izraza referat, prikaz i sinopsis.[4]

Sažetak – svrha

Većina autora pravi razliku između dviju osnovnih funkcija sažetka: (1) sažetak kao pomagalo u pretraživanju literature putem kojeg se primarni dokument dovodi u vezu sa upitom korisnika i (2) sažetak kao više ili manje potpuni opis primarnog dokumenta. Prva se funkcija sažetka odnosi na njegovo korištenje u sistemu za pretraživanje i pronalaženje informacija/dokumenata a, zadovoljavajući drugu, sažetak bi trebao pružiti korisniku sve bitne informacije o tome da li je pojedinačan tekst za njega interesantan ili nije. U ekstremnim slučajevima, kako navode Borko i Bernier – autoriteti koji su se bavili kazalama i indeksiranjem, sažetak može biti zamjena za puni tekst, ali to nije poželjna zamjena.[5]

Ove dvije osnovne funkcije sažetka mogu biti proširene i njegovim dodatnim ulogama. Sažetak se pojedinačno može koristiti i u osvjetljavanju sadržaja dokumenata napisanih na nepoznatom jeziku (većina servisa i izdavača prakticira štampanje sažetka na dva jezika), a štampanje i distribucija sažetka zapravo je i efikasno čuvanje svijeta informacija u, i za, određeno polje interesiranja korisnika. Sažetak se može koristiti i kao sredstvo u službama tekućih upozorenja, jer može biti efikasno sredstvo za informiranje o najnovijoj literaturi i dostignućima u određenom području.

Jedna od najznačajnih funkcija sažetka je njegova uloga u kompjuterskim bazama podataka, gdje se koristi za identifikaciju važnih članaka i povećanje pristupnih tačaka u pronalaženju arhiviranih dokumenata (u sistemima u kojima je tekst abstrakta arhiviran u pretraživačkoj formi) i kao podrška u izradi kontrolnih rječnika.

Tekst sažetka može biti dopuna (suplement) indeksnim terminima, te se pretraživanje vrši korištenjem indeksnih termina u kombinaciji sa pojmovima iz teksta sažetka.

Namjena sažetka zavisi od namjene abstraktnog servisa za koji se izrađuje. Analogno tome, u zavisnosti da li se sažetak izrađuje za servise okrenute disciplini (discipline oriented) ili servise okrenute zadatku (mission oriented), zavisit će i izgled sažetka i prosudba o njegovom kvalitetu.

Ipak se može reći da je glavna svrha tradicionalnog sažetka u primarnom i sekundarnom časopisu (odnosno bibliografskoj bazi podataka), da uputi na sadr-

[4] Definicije su urađene koristeći: Chowdhury, G.G. Introduction to modern information retrieval. London : Library Association Publishing, 1999, str. 154**;** Rowley, Jennifer; Farrow, John. Oganizing Knowledge... str. 67-70; Tuđman, Miroslav. Obavijest i znanje. Zavod za informacijske studije, Zagreb, 1990. str. 163.; ISO 214: 1976. Documntation: abstracts for publication and documentation. Geneva : International Standards Organization, 1976.; Guidelines for Apstrakt ANSI/NISO Z39.14.) publikovan 1997. godine.

[5] Lankaster, F. Wilfrid. Indexing und Abstracting... str. 98.

žaj dokumenta, pri čemu se od sažetka zahtijeva tačna, kratka i jasna prezentacija sadržaja dokumenta. Prema istraživanjima koja su sprovedena (Hjørland, B.; Nielsen, L. K.) pokazalo se, što nije neočekivano, da je sažetak efikasnije pomagalo u ocjenjivanju relevantnosti članaka za odabir, nego su to autor ili naslov, te da imaju znatno veći utjecaj na povećanje razine odziva i preciznosti, nego naslovi i ključne riječi. [6]

Razlog za tako visoko ocijenjenu ulogu sažetka treba potražiti u njegovoj posebnoj formi. Na razini makrostrukture to je visoko strukturirana forma, a po stepenu obrade informacija koje su sadržane u sažetku on pripada grupi alata u kojoj je informacija podvrgnuta maksimalnoj obradi. Da bi se razumjelo o kolikoj je obradi informacija riječ, potrebno je sažetak predstaviti kroz njegove vrste, stil, sadržaj elemenata, te upute i propise za njegovu izradu.

Vrste sažetaka i njihova struktura

Sažetak može biti okarakteriziran i opisan na različite načine. Jedan od njih je i podjela sažetka prema vrstama i strukturi karakterističnoj za te vrste. Najpoznatije su dvije vrste sažetaka: informativni i indikativni.

Informativni sažeci generalno se koriste za dokumente koji se odnose na eksperimentalna istraživanja, pretraživanja ili preglede. U tim sažecima navodi se svrha, metodologija, rezultati i zaključci koji su predstavljeni u originalnom dokumentu. Dok većina sažetaka navodi eksperimentalni rad gore navedenim redoslijedom, o najboljem redoslijedu ipak može odlučiti onaj kojem je abstrakt prvenstveno namijenjen. Naprimjer, neki od korisnika na prvo mjesto mogu staviti (najvažnije) rezultate i zaključke, ukoliko su se odlučili za rezultatski orijentiran raspored (redoslijed).

Indikativni sažeci su najupotrebljiviji za manje strukturirane dokumente, kao što su uvodnici, eseji, mišljenja ili opisi; ili za opširne dokumente, kao što su knjige, konferencijski zapisnici, imenici, bibliografije, liste, i godišnjaci. Indikativni sažeci obično se pišu za dokumente koji ne sadržavaju informaciju vezanu za metodologiju ili rezultate. U svakom slučaju sažetak će opisati svrhu ili obim diskusije, ili dati opis dokumenta. Indikativni sažetak također može opisati suštinsko značenje (esencijalni background) materije, korisnički pristup, i/ili argumente predstavljene u tekstu.

U praksi, originalni dokument može sadržavati elemente koji zahtijevaju indikativno-informativni pristup u izradi sažetka.

U Standardu i literaturi među definicijama nalazimo najčešće još dvije vrste sažetaka. To su kritički i namjenski sažetak. Oba tipa sažetka su skupi za proizvo-

[6] J. W. Janes (1991) je istraživao efektivnost apstrakta u odnosu na druge dijelove zapisa: naslov, autora i dr. i dobio očekivane rezultate da je apstrakt više relevantan za pretraživanje, nego drugi dijelovi zapisa. Prema Lancaster, F. Wilfrid. Indexing und Abstracting... str. 98; Do sličnih rezultata došli su i Hjørland, B.; Nielsen, L. K. Subject access points in electronic retrieval // Annual review of information science and technology. 35, 2001, str. 249-298.

dnju, jer se od abstraktora zahtijeva detaljnije znanje predmeta, tako da su oni neobični i vrlo rijetki u publiciranim radovima abstraktnih servisa.

Kritički sažetak je neuobičajena forma sažetka koji sadrži vrijednosne primjedbe na smisao apstrahiranih materijala ili na stil njihove prezentacije. Komentare (primjedbe) pišu abstraktori, koji su obično specijalisti predmetnih oblasti.

Dvije publikacije zahtijevaju kritički sažetak: Mathematical Review i Applied Mechanics Reviews (AMR). Zapis u ovom sažetku je na kraju potpisan i on je kombinacija deskriptivnog sažetka i kritičkih elemenata. Iz iskustva AMR otkriveno je da je vjeran kritički sažetak više izuzetak, nego pravilo.

Namjenski sažetak ima zadatak da predoči dio nečega ili pojedinačno gledišete koje je od interesa za posebnu vrstu korisnika. Namjena sažetka zavisi od namjene abstraktonog servisa za koji se izrađuje. Analogno tome, u zavisnosti od toga da li se sažetak izrađuje za servise okrenute disciplini (discipline oriented) ili servise okrenute zadatku (mission – oriented), zavisit će i izgled sažetka i prosudba o njegovom kvalitetu. Namjenski sažetak se izrađuje u servisima okrenutim zadatku (mission–oriented).[7]

Uz ove vrste sažetaka, koje se najčeće navode u standardima i literaturi, postoje i druge, netipične forme, koje su na prelazu između sažetka i indeksnog niza. To su mini sažetak i telegrafski sažetak.

Mini sažetak je neprecizan termin, smatra Lancaster, i dodaje da bi bilo bolje imenovanje kratki sažetak, kao što je korišteno kod Lunina. Ma kakvo bilo imenovanje, termin referira visoku strukturu abstrakta, doznačenu primarno za kompjutersko pretraživanje. To je, ustvari, vrsta križanja između sažetka i lanca indeksnih termina koje Lunin definira kao "mašinski čitljiv index-abstrakt". Termini napisani u sažecima povučeni su iz kontrolnog vokabulara i stavljeni zajedno u specifične sekvence. Naprimjer, formuliranje "tamo je iznešeno opadanje cinka u ljudskoj krvi izazvano cirozom jetre", u mini sažetku bi bilo predstavljeno:

smanjenje/cink/krv/čovjek/ciroza/jetra[8]

[7] Šesto poglavlje Standarda (ANSI/NISO Z39.14) namijenjeno je vrstama sažetaka i njihovom sadržaju. U ovom Poglavlju su predstavljene dvije osnovne vrste sažetaka (podjela po sadržaju): indikativni i informativni, kao i preporuke kada bi se oni trebali koristiti.

O vrstama abstraktnih i indeksnih servisa vidi kod: Penna, C.V; Foskett, D.J; Sewell P.H. National Library and Ingformation Services : A Handbook for Planners. London...et. cet.: Butterworthes, 1977, str. 192-198.

[8] Primjer je preveden prema : /decr/zinc/blood/humans/cirrhosis/liver Upozorenje! Decr. je skraćeno od decrease što znači smanjenje. Prema Lancaster, F. Wilfrid. Indexing und Abstracting ... str. 105.

Zapis u mini sažetku čine sekvence termina čija je primarna namjena vezana za unapređenje automatskog pretraživanja ili, kako je Lunin smatarao, za "inteligentnog čitača".
Telegrafski sažetak također je neprecizan termin. Ovaj sažetak podrazumijeva vrlo štedljivu prezentaciju sadržaja dokumenta, koja ne mora biti kompletna rečnica (bez sintakse) i liči na telegram dat u obliku lanca termina. Luninov minisažetak po stilu je telegrafski, a koristio se u ranim kompjuterskim pretraživačkim sistemima razvijanim za Western Reserve University[9]

Struktura i sadržaj elemenata sažetka

U tradicionalnom znanstvenom radu postoji globalna distinkcija između dijelova: uvod, metode, rezultati i diskusija, koja je poznata kao akronim UMRD. U uvodu se daju podaci o predmetu, svrsi i opsegu rada, navode se rezultati srodnih radova objavljenih u literaturi, formuliraju se hipoteze. Pod metodom se daje opis tehnike i metodologije rada, a u dijelu – rezultati, donosi se sve što je u radu pronađeno. U diskusiji autor uspoređuje svoje rezultate sa rezultatima drugih, ocjenjuje njihovu relevantnost, izvlači zaključke na osnovu dobivenih rezultata, a ako je u uvodu postavljena hipoteza, ona se u zaključku odbacuje ili prihvata.[10]
Pošto sažetak predstavlja kratko i tačno predstavljen sadržaj dokumenta i njegovog stila, onda on treba biti prikaz i njegove makrostrukture. I zaista, čitav niz članaka i priručnika u kojima se analiziraju terminologija, definicije, povijest, vrste pravila za izradu, svrha i korištenje sažetaka, te raznovrsni standardi – navode kriterije koje sažetak mora udovoljavati u pogledu sadržaja. Prema tim kriterijima sažetak mora sadržavati podatke o svrsi (opis cilja istraživanja ili zašto je članak napisan), metodi (opis eksperimentalne tehnike ili kako je ostvaren cilj istraživanja), rezultatima (opis ishoda rada ili šta je pronađeno), i o zaključcima (interpretacija i ocjena važnosti rezultata).[11]
Makrostrukturiranje sažetka omogućava da se iz dijelova sadržaja teksta prenesu najzačajnije tačke, tako da smisao sažetka odgovara sadržaju teksta. Odluka o tome koje je informacije sažetku još potrebno dodati, ovisi o potrebama koje nameću pojedini servisi za izradu sažetaka. Visoko makrostrukturirani sažeci su najčešće informativni sažeci za koje se smatra da imaju najveću informativnu vrijednost.

[9] Više o tome vidi Lancaster, F. Wilfrid. ... str. 104-107.

[10] Katnić-Bakaršić, Marina. Stilistika. Sarajevo : Ljiljan, 2001, str. 79.

[11] U tekstu Milas-Bracović, Milice, "Zastupljenost ključnih riječi iz naslova i teksta članka u njegovom autorskom sažetku" autorica navodi, da se sadržaj dokumenta ostvaruje na dvije razine: na razini mikrostrukture i na razini makrostrukture. Mikrostruktura se odnosi na proporcije i na njihov linearni slijed, a makrostruktura na strukturu dokumenta kao cjeline ili na strukturu njegovih dijelova kao posebnih cjelina. Vidi Milas-Bracović, Milica. Struktura znanstvenog članka i njegovog autorskog sažetka. // Informatologia Yugoslavica 19, 1-2(1987), str. 51-63.

Stil

Stilske karakteristike sažetaka u uputama za njihovo sastavljanje (npr. Biological Abstracts), kao i u Standardu, definirana su na sličan način. Obično se kaže da sažetak mora biti nekritički informativni pregled koji ne odražava stavove sastavljača i ne donosi sudove o dokumentu. Sažetak treba biti razumljiv bez navođenja referenci o dokumentu. U sažetku treba izbjegavati korištenje fusnota, kao i referensnih listi koje ima tekst originalnog dokumenta, a konciznost i potpunost sadržaja su uslov, nikako tajnost ili opskurnost. Za koherentnost stila, kako je navedeno u Standardu, treba upotrebljavati tradicionalne riječi i fraze.

Dužina

Bitna karakteristika sažetka je dužina. U Standardu za izradu sažetaka (ANSI/NISO Z39.14.-1997) u poglavlju 7. Stil u tačci 7.1 date su upute o dužini sažetka. Navedeno je da dužina sažetka zavisi od vrste dokumenta i veličine, te da je time i uslovljena. Za članke u časopisima predviđa do 250 riječi: za kraće zapise do 100 riječi, pisma urednika i slične vrste teksta do 30 riječi, a za monografije i teze do 300 riječi ili jednu stranu.[12]

I u uputama za izradu sažetaka dati su kriteriji koji se odnose na njegovu dužinu. Tako nam u pomenutoj knjizi Lancaster daje listu od sedam uslova koji mogu da utječu na dužinu abstrakta. Prve četiri tačke odnose se na korelaciju između predmetne materije i dužine sažetka, a to su: kompleksnost predmetne materije, raznovrsnost predmetne materije, važnost predmeta spremljenih u abstraktu, te pristupačnost predmetne materije (zavisi od namjene sažetka). Drugi dio uslova odnosi se na cijenu, te se zalaže za kraći abstrakt i preporučuje da broj ključnih riječi bude između 200 i 500. U tačci sedam Lancaster navodi da je cilj abstrakta da se primarno koristi kao pristup dokumentima za pretraživanje, te da je zbog toga nužno u njemu obezbijediti veći broj pristupnih tačaka, a to svakako utječe na dužinu sažetka.

Većina preporuka koristi broj riječi za mjerenje dužine abtrakta ili znakova, a pojedini autori nam daju preporuke o procentualnoj ovisnosti dužine teksta i dužine sažetka. Tako Borko smatra da je zadovoljavajuća mjera da sažetak zaprema dvanaestinu teksta, a u Uputama za sastavljanje sažetaka "Biological Abstracts" navodi da abstrakt treba da je kratak i da predstavlja 3% od originala.[13]

Strukturiranje (izgradnja) sažetka

U stil sažetka spadaju i generalne upute među kojima su slijedeće preporuke za njegovu izradu. Sažetak treba da bude napisan kao jedinstven paragraf, da ko-

[12] Guidelines for Apstrakt ANSI/NISO Z39.14.-1997…str. 4.

[13] Lankaster F. Wilfrid.: Indexing und Abstracting ... str. 94; Mihajlov, A.I. i Giljarevski R.S. Uvod u informatiku i dokumentaciju. ... str. 51.

risti kompletne rečenice, u prvoj rečenici ne koristi imenovanja, kao naprimjer: ovaj članak opisuje, ova studija predstavlja... [14]
To pokazuje primjer koji slijedi. Početak sažetka: "U radu su analizirani rezultati liječenja 30-ero novorođenčadi radikalno operirane zbog atrezije jednjaka, kao i činioci koji utječu na konačan ishod liječenja." Treba izbjeći: "u radu su", jer bi rečenica bi bila sasvim jasna i sa: "Analizirani su rezultati liječenja 30-ero novorođenčadi radikalno operirane zbog atrezije jednjaka, kao i činioci koji utječu na konačan ishod liječenja".
Zatim je tu i preporuka o korištenju vremena (aktivnog i pasivnog stanja) i na kraju uputstva vezana za terminologiju: izbjegavanje neprihvaćenih termina, akronima, skraćenica i simbola ili prvi put definiranih termina, te citiranje.[15]

Upute za izradu sažetaka

Elementi iz Standarda nalaze se i u upustvima koje daju organizacije i informacioni centri ili abstraktni servisi koji se bave izradom i distribucijom sažetaka. Te preporuke su ujedno i pomoć i podsjetnik abstraktoru. One se mogu svrstati u dvije skupine: (1) preporuke koje su namijenjene širem krugu korisnika koje, naprimjer, daju: Aslib ili Defense Dokumentation Center i (2) preciznija uputstva za izradu sažetaka koje daju abstraktni servisi specijaliziranog predmetnog područja kao naprimjer IEEE (Institute of Electrical and Electronics Engineers) ili uputstva koja distribuiraju informacioni centri kao INSPEC (International Information Services for the Physices and Engineerig Communities). Među ovim uputstvima posebno su zanimljiva uputstva u oblasti medicinskih nauka.
Abstraktni servisi koji pokrivaju oblast medicinskih nauka i njima srodnih oblasti spadaju među najstarije abstraktne servise koji imaju veoma raznovrsna uputstva za izradu sažetaka. Među tim uputstvima nalazimo i uputstvo *Annals of Internal Medicine*. U ovim uputstvima se zahtijeva da sažetak bude "članak sam po sebi", te da se vrste i izrada abstrakta podudaraju sa vrstama članaka.[16]
Suprotna ovim uputstvima za izgled i pisanje sažetka su uputstva koje daje Aslib, a koje ćemo u daljem tekstu navesti u cijelosti. Kvalitet njegovih uputstava je u tome da ona spadaju u vrstu općih uputstava koje je lako prilagoditi pojedinačnim zahtjevima abstraktnih servisa, da daju konkretnija uputstva od većine uputstava za rješenje praktičnih problema prilikom izrade sažetka, da je u njima jasno naznačena veza između indeksnog i abstraktnog procesa, da se dopunjavaju u pogledu dužine i stila sa Standardom za izradu sažetaka (izostavljeni su

[14] U Guidelines for Apstrakt ANSI/NISO Z39.14.-1997, Poglavlja 7.1-7.8 str. 4-5; Audy, Ljiljana... et. al. Činioci uspjeha i neuspjeha u liječenju novorođenčadi s artezijom jednjaka. // Liječnički vijesnik, 100, 4(1978), str. 225.

[15] Guidelines for Apstrakt ANSI/NISO Z39.14.-1997. Poglavlja 7.1-7.8 str. 4-5.

[16] Među najpoznatijim uputstvima su Uputstva International Committee of Medical Journal Editors, Uniform requierments for submitted to biomedical journals i Vankuverska pravila za oblast medicine.

dijelovi koji se mogu naći u Standardu), te da predstavljaju dobru osnovu za analizu sažetka koju ćemo dati u kasnijem dijelu rada.

Sadržaj sažetka

Opći navodi

Upotrebljavati važne pojmove i termine (npr. one koji će povećati povrat slobodnog navoda, one za koje dokument pruža dovoljno informacija, ili ključne riječi).

Indeksni termini

Uskladiti pojmove koji se upotrebljavaju u sažetku sa onim u deskriptorima

1.1 odrediti pojmove u sažetku tako da budu identični sa deskriptorima,

1.2 odrediti pojmove u sažecima tako da dopunjuju deskriptore (npr. relevantne termine izostavljene u indeksiranju deskriptora i u naslovima, termine koji su specifičniji od deskriptora, ili posebne termine koji su važni za područje predmeta kao što su geografska imena),

1.3 odrediti pojmove u sažecima koji dopunjuju i koji su identični sa deskriptorima.

Pojačati indeksiranje neovisno od jezika koji se upotrebljava.

Lista provjere

Slijediti listu relevantnih povratnih elemenata koje treba uključiti u sažetke

Oblici lista provjere

a) kategorije koje bi trebalo uključiti u sažetak (npr. materijali, kategorije, osobine i procesi) i uvjeti pod kojima bi trebali biti uključeni (npr. samo onda ako se o njima opsežno govori, ili bilo kada).

b) Specifične ili posebne odrednice (npr. kad god opisujemo novi proizvod spomenuti mjesto kompanije).

Jezik sažetka

Upotreba autorovog jezika

Upotrebljavati autorov jezik

Ne upotrebljavati autorov jezik

a) Upotrebljavati standardizirane i konkretne termine specifične za predmet.

Upotrebljavati (oba) autorov jezik, a i sinonime.

Odnos sa upotrijebljenim jezikom indeksiranja

Uskladiti (koordinatne) termine u sažecima sa deskriptorima.

Dopuniti deskriptore terminima iz sažetaka (npr. sinonimi ili specifičniji termini).

Upotrijebiti specifične i usvojene termine za određene kategorije (kao što su materijali, procesi, proizvodi).

Izbjegavati
Ne upotrebljavati negativan oblik (npr. umjesto bolestan nije zdrav).
Ne navoditi termine koji u nizu imaju zajedničku posljednju riječ (kao što je niža, srednja i radnička klasa).

Oblici riječi
Slijediti praksu lokalnog jezika (npr. zamjena američkog pravopisa britanskim)
Uvijek točno odrediti termine u određenim kategorijama (kao što su procesi, materijali, proizvodi).
Ako su termin i deskriptor isti, bilježiti termin onako kako je u deskriptoru.
Izražavati oba termina u skraćenom, ali i u punom obliku.[17]

Vrednovanje sažetka

U vrednovanju sažetaka razvijeno je nekolika načina za provjeru njihovog kvaliteta. Pri tome potrebno je naglasiti da nijedan od dva sažetka napisan za dokument neće biti identičan ako su napisani od različitih pojedinaca ili od jednog autora ali u različito vrijeme. Kvalitet sažetka ovisi o tome šta je napisano i ko je to napisao.
Korištene su razne metode u prosuđivanju kvaliteta sažetaka. Među najpoznatije metode spadaju one koje su zasnovane na ispitivanju podudaranja leksike između članka i sažetka (pod leksikom podrazumijevamo skup svih riječi koje se koriste u tekstu članka i njegovom sažetku.) Statističke metode indiciranja sadržaja teksta zasnivaju se na pretpostavci da uočljive formalne karakteristike teksta odražavaju njegov sadržaj, a koriste saznanja do kojih se došlo statističkim proučavanjima jezika (Zipf, Herdan, Guiraud, Mandelbrot itd.) U informacijskim i dokumantacijskim istraživanjima prvi ih je primijenio H. P. Luhn 1957. godine. To je, dakle, metoda koja posredno, preko analize formalnog podudaranja leksike članka i leksike sažatka, omogućava i određivanje njihovog podudaranja po sadržaju.
Nakon Luhna, razni su autori nastavili primjenjivati njegovu metodu uz manje ili veće modifikacije, kako na području kompjutorske izrade sažetka, tako i na području automatskog indeksiranja.[18]

[17] Aslib –Association for Information Management. Tekst Uputstava preuzet od Lankaster, F. Wilfrid. Indexing und Abstracting … str. 132.

[18] P. H. Luhn se dosjetio da upotrijebi podatke o učestalosti pojavljivanja pojedinih riječi u tekstu i da na osnovu njih odabere reprezentativne riječi i rečenice. On je upotrijebio Zipfov zakon kao nultu hipotezu kako bi odredio gornju i donju frekvencijsku granicu iznad i ispod koje se nalaze riječi koje nisu bitne za predstavljanje sadržaja teksta. U gornjem frekvencijskom području nalaze se najčešće korištene riječi koje su prisutne gotovo u svakom tekstu, a u donjem frekvencijskom području riječi rijetko upotrijebljene i samim tim manje bitne za njegov sadržaj. Po Luhnu, riječi koje najbolje predstavljaju sadržaj teksta nalaze se negdje u srednjem frekvencijskom području. Milas-Bracović, Milica; Barany, Istvan, Boras, Damir. Zastupljenost ključnih riječi iz naslova i

Leksička analiza pojmova rađena je i za određivanje dosljednog korištenja pojmova u sažecima. Načelo dosljedne upotrebe termina predstavlja jedan od najvažnijih kriterija za vrednovanje sažetka, a odnosi se na dosljednu upotrebu pojmova u procesu apstrahiranja u odnosu na cjelokupan korpus tekstova jednog predmetnog područja. Dosljedan odabir termina za jedno predmetno područje veoma je važan, jer on obezbjeđuje da ne dođe do "konceptualne rasutosti", koja rezultira gubitkom adekvatnih informacija u, i za određeno polje interesiranja korisnika, i direktno utječe na povećavanje broja relevantnih/nerelevantnih dokumenata.[19] Dosljedno korištenje termina, međutim, može također da smanji opoziv prilikom pretraživanja. Ta paradoksalna situacija proizlazi iz dva osnovna zadatka na koje dobar sažetak treba da odgovori, a to je da se koristi za brzo upoznavanje sa sadržajem dokumenta, i drugo, da se koristi prilikom pretraživanja. Nažalost, ova dva cilja nisu potpuno kompatibilna.

Prvi zadatak sažetka, da pomogne u odabiru koji dokument želimo koristiti a koji ne, zahtijeva da se pojmovi u sažetku koriste dosljedno. Rukovodeći se tim zadatkom dosljednost je podcrtavana kao bitan kriterij u vrednovanju sažetka. Dosljedna upotreba termina je od velike važnosti za čitaoca, jer ne stvara konfuziju u čitanju. Zato je za korisnika poželjna dosljednost, a ne redundantnost, jer dosljednost smanjuje konceptualnu rasutost. Redundantnost pojmova je, međutim, poželjna u pretraživanju, jer povećava vjerojatnošću da će neki dokument biti pronađen. Na primjer, uključivanjem sinonima u trokutasti znak ili znak delte u tekst sažetka, povećava se mogućnost pronalaska dokumenata, ali to povećava i njegovu dužinu.

Pisanje sažetka

Prethodno navedene upute, preporuke i Standardi dati su u svrhu savladavanja dobre prakse pisanja sažetaka. Greške koje se najčešće dešavaju prilikom izrade sažetka odnose se na: greške napravljene prilikom indeksiranja, greške u korištenju jezika sažetka i upotrebe sinonima, i greške koje se odnose na makrostrukturu sažetka.

Uz ove greške, koje zavise od vještine i iskustva, poznavanja propisa i standarda prilikom izrade sažetka, postoji i drugi niz grešaka koje se događaju, a njih uopćeno možemo svrstati u greške tehničke prirode koje se mogu desiti, recimo, prilikom prepisivanja (formula, brojeva itd). I na kraju, treći niz grešaka proizlazi iz nedovoljnog poznavanja predmetne materije koja se obrađuje. Većina autora preporučuje da u eliminiranju načinjenih grešaka sažetak treba biti pregledan od iskusnijeg stručnjaka koji se bavi pripremom abstrakta, pa se prepo-

teksta članka u njegovom autorskom sažetku. // Informatologia Yugoslavica. 17, 3-4(1986), str. 243-265

[19] Milas-Bracović, Milica; Barany, Istvan, Boras, Damir. Zastupljenost ključnih riječi iz naslova i teksta članka u njegovom autorskom sažetku. // Informatologia Yugoslavica. 17, 3-4(1986), str. 243-265.

ručuju liste adekvatnih i konkretnih uputstava (worksheet) kao brza pomoć prilikom pisanja sažetka.[20]

Pisanje sažetka, dakle, zahtijeva: vještinu i iskustvo, poznavanje propisa, standarda i upućenost u predmetnu materiju koja je obrađivana u dokumentu. Jedna od neriješenih nedoumica prilikom pisanja sažetka odnosi se i na pitanje: da li autor, predmetni specijalista ili profesionalni abstraktor, treba pisati sažetak?

Preporuke ko treba pisati sažetak u literaturi su različite. Na strani onih koji smatraju da autor treba biti pisac sažetka je argument da on najbolje poznaje predmetnu materiju koju obrađuje. Oni koji to negiraju (Borko i Bernier) naglašavaju da je to autorima "volonterski rad", te da izradu sažetka treba da rade profesionalni abstraktori, koji imaju objektivniji pristup materiji i iskustvo u izradi sažetka.[21]

I mada dileme oko toga ko je podoban za pisanje sažetka nisu riješene, čini se da sada to i nije presudno pitanje. Ovdje je važnije istaći da se, u većini uputa i smjernica za izradu sažetka, sve više podcrtava uloga sažetka kao sredstva koje se koristi u pretraživanju, te da je to ono što "autorima sažetka" treba biti na umu. Zato se u upute i smjernice za izradu sažetaka sve više unose elementi koji sažetke pripremaju za uključivanje u datobaze, kao što su uspostavljanje semantičkih odnosa među pojmovima u samom tekstu sažetka, kao i strože makrostrukturiranje koje povećava broj pristupnih tačaka za pretraživanje. U tom kontekstu je u Standardu za izradu sažetka (u tačci 4.1 – Svrha) dato i uputstvo da je – prije reprodukcije sažetka za abstraktne servise, tekuće publikacije ili datobaze u kojima će sažetak činiti punu bibliografsku referencu za originalni dokument – potrebno razmotriti njegovu objektivnost. [22]

Tvrdnja da se uspostavljanjem semantičkih odnosa među riječima i strožim makrostrukturiranjem zaista povećava mogućnost boljeg pretraživanja podataka i daje bolji uvid u sadržaj dokumenta, bit će ilustrirana na primjeru sažetka "Hematoencefalna barijera u toku šećerne bolesti".[23]

> Bolesnik sa **šećernom bolesti** je sklon razvoju vaskularnih lezija krvnih žila svih veličina. U toku bolesti se progresivno razvijaju mikroangiopatija i ateroskleroza. Nastavljajući radove Stanleya I, Rapoporta, te drugih istraživača o promjeni permeabiliteta **hematoencefalne barijere** u eksperimentalnim uvjetima, ovim radom se ispitala promjena **barijere** kod pacijenata sa **diabetes mellitusom**. Kod ispitanika sa **diabetes mellitusom**

[20] Lancaster, F. Wilfrid. Indexing und Abstracting... str. 114.

[21] Lankaster, F. Wilfrid. Indexing und Abstracting ...str. 108.

[22] Guidelines for Apstrakt ANSI/NISO Z39.14.-1997, str. 2.

[23] N.N.Hematoencefalna barijera u toku šećerne bolesti. // Medicinski žurnal. 7, 1(2001), str. 36-46.

raste nivo totalnih proteina, glukoze i natrija u likvoru. Kod ispitanika sa **šećernom bolesti** opada nivo hlora i kalija u **likvoru**.

Ključne riječi: Likvor, Totalni proteini, Šećerna bolest, Hematoencefalna barijera

Upoređivanjem citiranog sažetaka sa Aslibovim *Općim uputama za izradu sažetaka*, odmah je uočeno da se u tekstu, sažetku i među ključnim riječima, koriste riječi koje pripadaju latinskim medicinskim nazivima i medicinskim nazivima na bosanskom jeziku, čime je autor sažetka narušio princip dosljedne upotrebe termina, što bi mogla biti greška u ovom sažetku. Uvažavajući preporuke *Uputa* zaključeno je da se greške napravljene prilikom izrade citiranog sažetka odnose na one greške koje su napravljene u procesu indeksiranja, odnosno da nije izvršeno usklađivanje termina iz teksta sa terminima u sažetku i sa ključnim riječima koje su navedene uz tekst.
Međutim, ranije smo naveli da poštivanje principa dosljednosti prati problem smanjenja odziva. Da bi se to prevazišlo koristi se mogućnost upotrebe oba jezika (autorov jezik i sinonimi), pa se problem rješava tako što se pojmovima iz jednog registra pridružuju pojmovi iz drugog (sinonimi), kao što preporučuju Aslibove upute. Ovako napisani tekst sažetka bit će dat u cijelosti nakon izvršenja drugog dijela analize koja se odnosi na vrstu i sadržaj (makrostrukturu) sažetka i njegovog prestukturiranja, uz korištenje preporuka iz Standarda.[24]
Na osnovu ispitivanja strukture navedenog sažetka ustanovljeno je da on pripada indikativnim abstraktima, jer nije u cijelosti strukuriran i posjeduje samo dio elemenata UMRAD strukture: Uvod i Rezultate. S obzirom da prati stručni članak, ocijenjeno je da bi ovaj sažetak bilo korisno podvrgnuti većem strukturiranju i izraditi ga kao informativni sažetak. Izrada takvog sažetka zahtijeva da u njemu budu zastupljeni svi elementi iz UMRD-a što podrazumijeva dodavanje i ostalih elemenata, Metode i Diskusije, čime bi bila zadovoljena visokostrukturirana forma.
Nakon što je analizirani abstrakt podvgnut makrostrukturiranju tekst sažetka bi mogao izgledati ovako:

Uvod: Bolesnik sa *šećernom bolesti* (diabetes mellitusom) je sklon razvoju vaskularnih lezija krvnih žila svih veličina. U toku bolesti se progresivno razvijaju mikroangiopatija i ateroskleroza. **Cilj** je utvrditi da li kod pacijenata sa *diabetes mellitusom* dolazi do promjene permeabilnosti *hematoencefalne barijere (krvno-moždane barijere*). **Metode**: Ispitivanje je vršeno kod sedam pacijenata bez moždanog udara i sa *diabetes mellitusom*, mjerenjem: *totalnih proteina, glukoze, natrija, kalija i hlora* u *likvo-*

[24] Guidelines for Apstrakt ANSI/NISO Z39.14.-1997, Poglavlja 6 i 7.

ru (moždana tekućina). **Rasprava**: Vrijednosti dobijenih rezultata kod ispitanika sa *diabetes mellitusom* su da: raste nivo *totalnih proteina, glukoze i natrija*, a opada nivo ***hlora i kalija*** u **likvoru**, grupirajući se oko donje granice normalnih vrijednosti. **Diskusija/Zaključak:** Na osnovu dobijenih vrijednosti utvrđeno je da dolazi do promjene permeabiliteta *hematoencefalne barijere* kod pacijenata sa *diabetes mellitusom.*

Izrada visokostrukturiranih sažetaka potrebna je jer su istraživanja pokazala da informativni abstrakti imaju veću koncentarciju riječi sa punim značenjem i time povećavaju odziv i preciznost. U prilog njihove izrade je i činjenica da su istraživanja pokazala da pretraživanje po potpunom tekstu često dovodi do velike količine nerelevantnih rezultata, odnosno niske preciznosti, a da su sažeci optimalni za potrebe pretraživanja.[25]

Zaključak

Još od kasnih devedesetih godina prošlog stoljeća preispituju se mogućnosti ponovnog definiranja elemenata koji se koriste u identifikaciji i opisu dokumenata. Kao bitan element pogodan za uključivanje među elemente opisa javlja se i sažetak. Zahvaljujući svojoj strogoj strukturi, prepozanatljivim vrstama, definiranom stilu i dužini, odnosno standardizaciji koja se primjenjuje u njegovoj izradi, sažetak je pogodan za uključivanje u standardizirane sisteme svih vrsta. I mada još od druge generacije online kataloga nema tehničkih problema sa navođenjem sažetka u bibliografske baze podataka (MARC/UNIMARC polje 300), ta praksa nije uobičajena.

Međutim, sve više zahtjeva za analitičkom obradom dokumenata, te potreba za proširenjem pristupnih tačaka u pretraživanju, ponovo aktueliziraju razmišljanje o uvođenju sažetka u opis. Uz to, elektronski dokumenti postaju sve više dio bibliotečkog fonda, a tradicionalni skup elemenata za njihovo arhiviranje i opis postaje nedostatan za upravljanje u mrežnom prostoru. Među "stabilnim" elementima koji se mogu ugraditi u opis elektronskog dokumenta, ali i sačinjavati dio spremišta neke baze podataka ili kataloga koji podržavaju pristup dokumentima, javlja se i sažetak, pa i to doprinosi da se o sažetku razmišlja kao standardnom bibliotečkom elementu.

Zato je za očekivati da će se u paradigmama opisa, kao element, uskoro navoditi i sažetak. Tome u prilog je i mogućnost preuzimanja podataka iz drugih izvora na Mreži, poput onih iz online knjižara, čiji su zapisi obogaćeni i drugim potencijalno korisnim podacima. Kao mogućnost povezivanja novodi se da bi možda bilo dovoljno samo povezati ISBN iz zapisa u bibliotečkom katalogu sa ISBN-om u zapisu u online knjižari i na taj način omogućiti pristup tim podacima, bez potrebe za preuzimanjem čitavog zapisa. U ovome se smjeru na međunarodnoj

[25] Pinto, M.; Lancaster, F. W. Abstracts and abstracting in knowledge discovery. // Library trends. 48, 1(1999), str. 234-248.

razini i kreće suradnja bibliotečke zajednice sa nakladničkim i knjižarskim zajednicama, što su već uradili Britanska biblioteka (British Library) i OCLC, koji su izradili tablice za konverziju nakladničkog i knjižarskog formata ONIX (Online Information Exchange) u knjižnične formate UNIMARC i MARC 21.[26]

Literatura

Audy, Ljiljana... et. al. Činioci uspjeha i neuspjeha u liječenju novorođenčadi s artezijom jednjaka. // Liječnički vijesnik. 100, 4 (1978), str. 225.

Chowdhury, G.G. Introduction to modern information retrieval. London : Library Association Publishing, 1999.

Delsey, Tom. Preispitivanje konvencionalnih paradigmi za opis dokumenta. // Vjesnik bibliotekara Hrvatske. 46, 1-2(2003), str. 32-43.

Hjørland, B.; Nielsen, L. K. Subject access points in electronic retrieval // Annual review of information science and technology. 35, 2001, str. 249-298.

ISO 214: 1976. Documentation: abstracts for publication and documentation. Geneva : International Standards Organization, 1976.

Golub, Koraljka. Predmetno pretraživanje u knjižničnim katalozima s Web-sučeljm (magistarski rad). Zagreb : Filozofski fakultet, 2003.http://eprints.rclis.org/archive/00003379/01/Magisterij-hrv.pdf (17. 10. 2006.).

Grundlagen der praktischen Information und Dokumentation. Band 1. Handbuch zur Einfürung in die Informationswissenschaft und – praxis./ (hrsg.) Rainer Kuhlen, Thomas Seeger und Dietmar Strauch. 5., völlig neu gefasste Ausgabe. München : K. G. Saur, 2004.

Grundlagen der praktischen Information und Dokumentation. Band 2.Glossar. /(hrsg.) Rainer Kuhlen, Thomas Seeger und Dietmar Strauch. 5., völlig neu gefasste Ausgabe. München : K. G. Saur, 2004.

Guidelines for Apstrakt. (ANSI/NISO Z39.14.). Merylend : NISO, 1996. publikovan 1997. godine.

Katnić-Bakaršić, Marina. Stilistika. Sarajevo : Ljiljan, 2001.

Lancaster, F. Wilfrid. Indexing and Abstracting in Thery and Practice. 2.ed. Illionois : Graduate School of Library and Information Science, 1998.

Mihajlov, A.I. i Giljarevski R.S. Uvod u informatiku i dokumentaciju. 2. izd. Zagreb : Referalni centar sveučilišta. 1984.

Milas-Bracović, Milica. Struktura znanstvenog članka i njegovog autorskog sažetka. // Informatologia Yugoslavica. 19, 1-2(1987), str. 51-63.

Milas—Bracović, Milica; Barany, Istvan, Boras, Damir. Zastupljenost ključnih riječi iz naslova i teksta članka u njegovom autorskom sažetku. // Informatologia Yugoslavica. 17, 3-4(1986), str. 243-265.

N.N. Hematoencefalna barijera u toku šećerne bolesti. // Medicinski žurnal. 7, 1(2001), str. 36-46.

Penna, C.V; Foskett, D.J; Sewell P.H. National Library and Ingformation Services : A Handbook for Planners. London...et. cet.: Butterworthes, 1977.

Pinto, M.; Lancaster, F. W. Abstracts and abstracting in knowledge discovery. // Library trends. 48, 1(1999), str. 234-248.

Rowley, Jennifer; Farrow, John. Oganizing Knowledge : An introduction to managing access to Information. (3. ed.) Aldershot, England: Grower, 2000.

Tuđman, Miroslav. Obavijest i znanje. Zagreb : Zavod za informacijske studije, 1990.

[26] Prema Koraljka Golub: Predmetno pretraživanje u knjižničnim katalozima s Web-sučeljm (magistarski rad). Zagreb : Filozofski fakultet, 2003. http://eprints.rclis.org/archive/00003379/01/Magisterij-hrv.pdf (17. 10. 2006.)

Peirceove logičke kategorije kao osnova za meko računarstvo

Alma Granov
Filozofski fakultet
Franje Račkog 1, Sarajevo, BiH
alma.granov@ff.unsa.ba

Sažetak

Meko računarstvo danas predstavlja veliki korak naprijed kada je u pitanju problem koncipiranja vještačke inteligencije i općenito inteligentnih sistema. Kompjuterska semiotika, kao hibridna naučna disciplina kompilirajući postulate semiotike i kompjuterske nauke pokušava modelirati opće koncepte za meko računarstvo. U tom smislu, Peircov koncept logičkih kategorija ispostavlja se kao koncept koji bi mogao polučiti rezultat kada su neki aspekti vještačke inteligencije u pitanju.

Ključne riječi: Peirceove logičke kategorije, meko računarstvo, kompjuterska semiotika

Peirceove logičke kategorije

Teorijski je okvir trima oblicima logičkog zaključivanja u Peircea njegov nauk o trima univerzalnim kategorijama: prvosti, drugosti i trećosti. Peirce u svojoj semiotičkoj teoriji opisuje tri načina zaključivanja: dedukciju, indukciju i abdukciju.

Dedukcija predstavlja način zaključivanja kojim se konstatira da je neko stanje stvari nužno takvo. Dedukcija polazi od nekog općeg pravila i nekog posmatranog pojedinačnog slučaja i dolazi do daljeg znanja o tom pojedinačnom slučaju. No, kako je deduktivni zaključak nužan jer je rezultat zaključivanja već sadržan u samom pravilu, on nikada ne može voditi nekoj novoj spoznaji.

Indukcija je način zaključivanja suprotstavljen dedukciji. U induktivnom postupku zaključivanja od pojedinačnog slučaja i nekog verificiranog ishoda generalizacijom se zaključuje o nekom pravilu. Treba naglasiti da je induktivno stečena spoznaja uvijek samo nekakav vjerojatan, moguć iskaz. Indukcija predstavlja proces potvrđivanja faktičnosti putem sučeljavanja i susretanja s činjenicama.

Abdukcija od ishoda kojem je potrebno objašnjenje zaključuje o dotada nepoznatom. Abdukcija za polazište ima prihvaćanje čiste mogućnosti. Za Peircea je abdukcija ključna u razumijevanju postupka koji se označava kao naučno otkri-

vanje (Hoffmann, 1997:13). Peirce je abdukciju identificirao sa pojmom pukog gonetanja, odnosno identificirao je sa intuitivnom sposobnošću definirajući intuiciju kao "premisu koja sama po sebi nije zaključak" (Peirce, 1868:103). Abduktivni ili hipotetički silogizam se može opisati na sljedeći način:

$$\frac{\text{ako q onda p} \quad \text{q}}{\text{p}} \quad \text{ili} \quad \frac{\text{ako q onda p} \quad \text{ne-p}}{\text{ne-q}}$$

U odnosu na činjenicu da je Peirceova kategorija prvosti definirana kao puka mogućnost, da je drugost definirana kao kategorija golih činjenica koje proizilaze iz nekog odnosa, a da je trećost definirana kao kategorija općih zakonitosti imamo jasnu sliku o abdukciji kao fenomenu prvosti, indukciji kao fenomenu drugosti i dedukciji kao fenomenu trećosti.

Tabela 1: Peirceove logičke kategorije

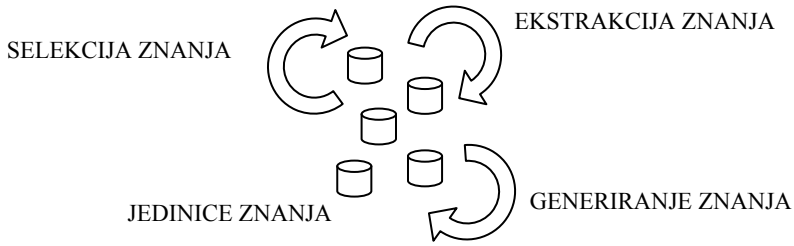
univerzalna kategorija / karakteristike	PRVOST	DRUGOST	TREĆOST
MOGUĆNOST	ABDUKCIJA		
GOLE ČINJENICE		INDUKCIJA	
OPĆE ZAKONITOSTI			DEDUKCIJA

Peirceove logičke kategorije i meko računarstvo

Soft computing (meko računarstvo) predstavlja skup računarskih metodologija oko fuzzy logike (FL), neuro-računarstva (NR), genetičkog računarstva (GR) i probabilističkog računarstva (PR). Unutar mekog računarstva, kako ističe Subašić (Subašić, 1997:189-190) oblasti od interesa fuzzy logike, neuro-računarstva, genetičkog računarstva i probabilističkog računarstva su:

FL – aproksimativno zaključivanje, granulacija informacija, računanje riječima;
NR – učenje, adaptacija, klasifikacija, modeliranje i identifikacija sistema;
GR – sinteza, podešavanje parametara i optimizacija putem sistematskog slučajnog pretraživanja i evolucije;
PR – obrada neizvijesnosti, probabilističke mreže, predviđanje, sistemi zasnovani na kaosu.

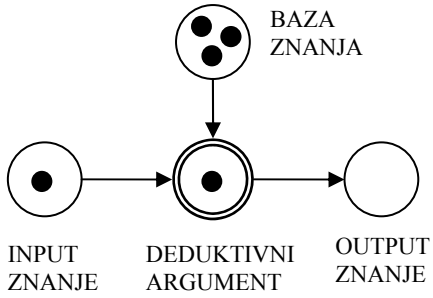
Gudwin i Gomde na tragu Peirceovih logičkih postavki promatraju kompjutersku semiotiku kao pokušaj emulacije kruga semioze unutar digitalnog kompjutera. Godwin smatra da se procesi operiranja znanjem mogu identificirati kao selekcija znanja (abdukcija), ekstrakcija znanja (dedukcija) i generiranje znanja (indukcija) i shematski je to predstavio na ovaj način (Godwin, 1998: 796):



Ilustracija 1: Procesi operiranja znanjem

Naime, Gudwin i Gomide (Gudwin and Gomide, 1997: 3981-3986) su pošli od toga da se transformacije spoznaje (znanja) mogu predstaviti posebnom vrstom znanja koju nazivaju argumentativnim znanjem ili jednostavno "argumenti". Kao tri osnovna argumenta oni prepoznaju, prema Peirceovom modelu, deduktivni, induktivni i abduktivni argument.

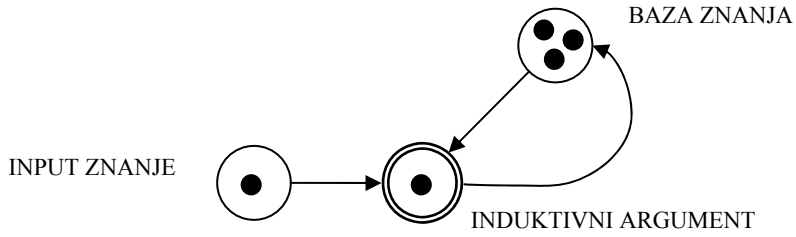
Deduktivni argument, prema njihovom viđenju, koristi transformacionu funkciju koja u osnovi prikuplja znanje iz dva inputa. Prvi konstituira neku vrstu generičkog znanja koja se odnosi na osnove znanja samog sistema. Drugi input je specifično znanje koje će se koristiti kao "ključ" za osnove znanja. Opća shema za deduktivni argument može se prema njihovom mišljenju prikazati ovako:



Ilustracija 2: Deduktivni argument

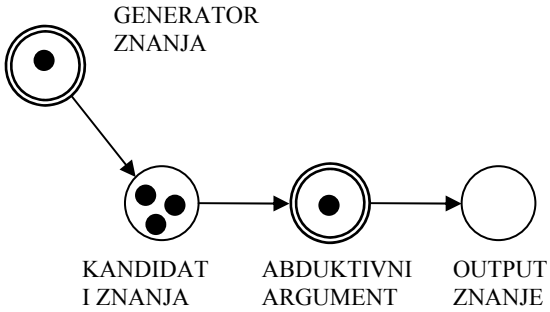
Glavni zadatak deduktivnog argumenta jeste da odabere u bazi znanja ispravan odlomak znanja i da ga koristi sa input znanjem, aktivirajući jednu od svojih unutarnjih funkcija kako bi generirao output znanje.

Induktivni argument se koristi kako bi se generirao novi odlomak znanja koji se može komparirati sa ostatkom znanja koje se već nalazi u sistemu. Osnovnu shemu induktivnog argumenta oni su predstavili ovako:



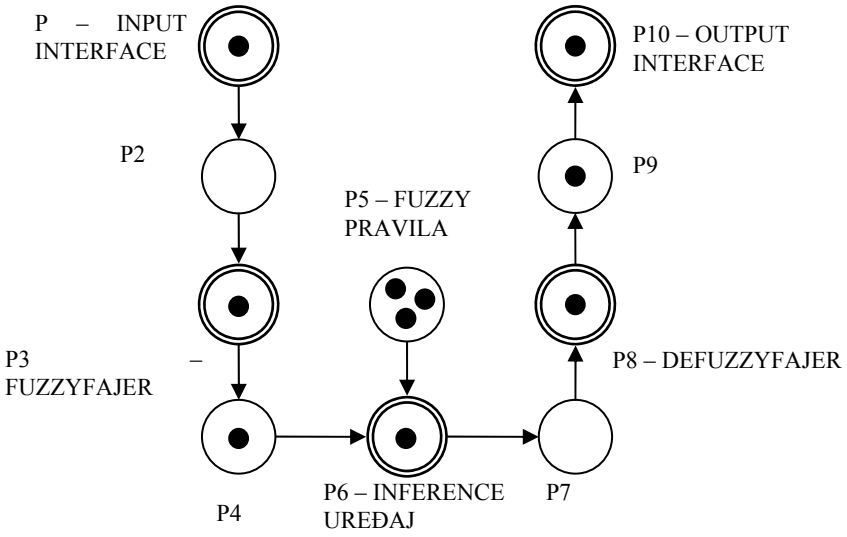
Ilustracija 3: Induktivni argument

Svrha induktivnog argumenta jeste modificiranje dijelova znanja koji su ranije pohranjeni u bazu znanja kao input. U tom smislu, modificirano znanje je u isto vrijeme input i output induktivnog argumenta. Ali to se može prepoznati i dijelom u shemi deduktivnog argument. Razlika među njima se očituje u tome što je znanje generirano deduktivnim argumentom već sadržano u bazi znanja. U tom smislu, ono samo ekstrahuje znanje iz baze znanja. Induktivni argument generira novi odlomak znanja koji nije sadržan u postojećoj bazi znanja.
Abduktivni argument se koristi kako bi izvršio vrednovanje odlomka znanja kojeg je prethodno generirao generator znanja kao što je npr. induktivni argument ili neka vrsta slučajnog procesa. U ovom slučaju ovaj argument ne generira novi odlomak znanja, nego odabire, iz zbira različitih kandidata, odlomak znanja koji je najvjerojatniji. Abduktivni argument oni su shematski prikazali ovako:



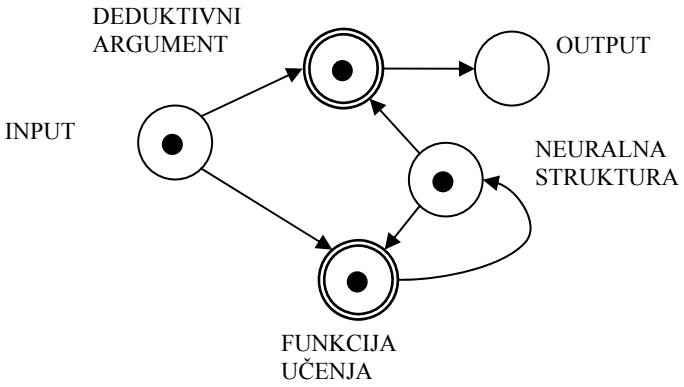
Ilustracija 4: Abduktivni argument

Abduktivni argument koristi svoju selekcionu funkciju kako bi izvršio odabir u inputu i doveo odabrano do outputa
Oni su pokazali kako se ovi semiotički mehanizmi mogu primijeniti u slučaju soft computing tehnika kao što su npr. fuzzy sistemi i neuralne mreže.
Opći fuzzy produkcioni sistem oni su prikazali sljedećom shemom:



Ilustracija 5: Opći fuzzy produkcioni sistem

Glavni argument u fuzzy produkcionom sistemu je deduktivni argument koji je predstavljen kao P6 u datoj shemi, i nosi naziv inference engine (naprava za izvođenje zaključaka). Ostali argumenti u P1, P3, P8 i P10 su također deduktivni elementi, ali njihov jedini zadatak je konvertirati znanje u pogodnu formu.
Shema samoorganizirajuće neuralne mreže u njihovoj interpretaciji semiotičkih argumenata prikazana je na sljedeći način:



Ilustracija 6:Semiotički argument u neuralnoj mreži

Kao što se može vidjeti ova neuralna mreža ima dva glavna argumneta. Deduktivni argument koji se obavlja za vrijeme feedforward faze kada generira output od inputa. Funkcija učenja je induktivni argument koji obavlja samoorganiziranje neuralne mreže. Ona funkcionira promatrajući input i modificirajući neuralnu strukturu obavljajući na taj način njenu fazu učenja.
Možda bi na ovom mjestu bilo interesantno napomenuti da Gudwin inače u svojim radovima (uključujući i doktorsku disertaciju) govori o jedinicama znanja u skaldu sa Peirceovom teorijom. Naime, Gudwin će istaći da ako podaci koje prezentira struktura slijede, odnosno poštuju direktno modeliranje izvanjskog fenomena ta jedinica znanja nosi naziv ikon. Ako oni, pak, pružaju lokalizaciju unutar prostora za prezentaciju neke druge strukture nose naziv indeks. I konačno, ako predstavljaju ključ u konverzionoj tabeli nose naziv simbol.
I što je najvažnije ovo sve se može pretočiti u matematiče opise i time ove postavke učiniti osnovom za kreiranje inteligentnih kompjuterskih sistema. Naime, proces defuzzyfikacije je proces u kojem se fuzzy zaključak pretvara u jedan realan broj. Broj koji se dobije metodom defuzzyfikacije nije proizvoljanon mora da bude realan predstavnik fuzzy zaključka koji smo dobili primjenom fuzzy pravila, te je na taj način prihvatljiv za mogućnost logike sa dvije vrijednosti (Subašić, 1997:97).

Zaključak

Kompjuterska semiotika raspolaže teorijskim i tehnološkim konceptima kojima be se mogli unaprijediti sadašnji modeli mekog računarstva. Periceov semiotički nauk kojega karkterizira pansemiotički pogled na univerzum ispostavlja se dobrom podlogom za pitanja kompjuterske semiotike i to upravo u onom njenom segmentu koji se tiče istraživanja umjetne inteligencije. Peirce je logiku vidio kao disciplinu semiotike i u tom smislu nastojao je definirati semiotički pogled na čovjeka i njegovo spoznavanje čime je stvorio dobru podlogu za one tehnološke koncepte koji nastoje oponašati čovjekov način rezoniranja i njegove sposobnosti obrade semiotičkih elemenata svijeta u kome djeluje. Zato će njegova logika koja počiva na induktivnom, deduktivnom i abduktivnom argumentu moći dobro poslužiti kao iskoristiv koncept u mekom računarstvu.

Literatura

Gudwin, Ricardo; Gomide Fernando. A Computational Semiotics Approach for Soft Computing. // IEEE International Conference on Systems, Man and Cybernetics – SMC'97 – 12-15/October – Orlando,1997, vol. 4, pp. 3981-3986

Gudwin, R. Ricardo. On the generalized deduction, induction and abduction as the elementary Opreators within Computational Semiotics. // Proceedings of ISAS'98/ Intelligent Systems and Semiotics-International Conference, 14-18/September, Gaithersburg, 1998, pp.795-800

Gudwin, R. Ricardo. Contribuições ao Estudo Matemático de Sistemas Inteligentes, Ph.D. Thesis. DCA-FEEC-UNICAMP, 1996

Hoffmann, Michael. Is there a "Logic" of Abduction?. // Forthcoming in the Proceedings of the 6th Congress of the IASS-AIS, International Association for Semiotic Studies in Guadalajara, 1997,pp. 13-18

Nöth, Winfried. Priručnik semiotike, Zagreb : Ceres, 2004
Peirce, Charles Sanders. Questions Concerning Certain Faculties Claimed for Man. // Journal of Speculative Philosophy, 2, 1868; pp.103-114
Peirce, Charles Sanders. Collected Papers, vols. I-VIII (eds. Hartshorne & Weiss; Burks), Cambridge, Mass.: Harvard University Press, 1934-58
Subašić, Pero. Fazi logika i neuronske mreže, Beograd, Tehnička knjiga, 1997
Zadeh, Lotfi. Fuzzy Logic and Approximate Reasoning. // Synthese 30, 1975; 407-428

Konceptualna i primijenjena okosnica studija informacijskih znanosti na Filozofskome fakultetu Sveučilišta u Zagrebu

Jadranka Lasić-Lazić
Odsjek za informacijske znanosti
Filozofski fakultet
Ivana Lučića 3, Zagreb, Croatia
jlazic@ffzg.hr

Marija László
mlaszlo@ffzg.hr

Bulcsú László
blaszlo@ffzg.hr

Summary

The interdisciplinary two-year undergraduate Study of Socio-Humanistic Information Science at the Faculty of Philosophy, University of Zagreb, originated in 1978 in the Department of Linguistics. The corresponding department was established two years later. While the other similar studies worldwide are based on hard sciences, the Zagreb study is connected with social and humanistic sciences in its curriculum. The data from the questionnaire of Zagreb University Committee on Information Science showed that the departments of Faculty of Philosophy offered more subjects in information sciences than any other faculty. Those subjects (just to mention one per department) were: Philosophy with Symbolic Logic, Phonetics with Theory of Information, Comparative Literature with Librarianship, Linguistics with Mathematical Linguistics, Pedagogy with Programmed Learning, History with Archival Science, Art History with Museology, Psychology with Statistics and Sociology with Research Methods. The article deals with the framework of the study and gives the definition of information science.

Key words: information science education, information science concepts, information infrastructure

Uvod

Odsjek za informacijske znanosti ubrzo slavi 30. rođendan. Ustrojen je 1978. pri Odsjeku za lingvistiku i orijentalne studije Filozofskoga fakulteta u Zagrebu

kao Katedra za društveno-humanističku informatiku, a već 1981. postaje temeljem novoosnovanoga Odsjeka za informacijske znanosti. Odsjek se taj razvijao usporedno s razvojem informacijske znanosti i tehnologije u svjetskim razmjerima. Sama je informacijska znanost morala, kao što je i naš Odsjek morao, zbog svoje višestručnosti koju mu je omogućio Fakultet pokazati da se može uhvatiti u koštac i sa znanstvenima i s tehnološkima zahtjevima današnje zbilje.
Svrha je ovoga priloga pokazati posebnost informacijskih znanosti i raspraviti koje će vještine i znanja biti prijeko potrebnima za život, rad i opstanak u informacijskome društvu koje je nastalo u drugoj polovici 20. stoljeća smijenivši pretežno poljodjeljsko s početka stoljeća, potom tvorničko društvo. U godinama se osnutka studija naslućivalo da sazrijeva vrijeme u kojem će se pametni strojevi sve više primjenjivati na svima područjima djelatnosti ljudske. Danas se na pragu 21. stoljeća na njihovoj primjeni izgrađuje društvo znanja kojemu su informacijske su znanosti okosnicom (framework).

Počeci

U tehnološki se razvijenome svijetu već šezdesetih godina 20. stoljeća pojavljuje studij informacijskih znanosti. Primjerice za boravka Lászlóa 1969. Povjerenstvo za informacijske znanosti na Fakultetu za fizikalne znanosti Sveučilišta u Chicagu nudi jednogodišnji studij (uz mogući doktorat ako se ispune opći uvjeti). Povjerenstvo je međustručno pod pročelništvom profesora informacijske znanosti Victora Yngvea, a nastavno mu osoblje tvore tri profesora informacijske znanosti, tri profesora primijenjene matematike i po jedan profesor fiziologije, matematičke biologije i matematičke ekonomije.
Ponuđeni su predmeti studija bili izričito međustručni, i to iz računarskih sustava, numeričkih postupaka, umjetne inteligencije, strojne obrade jezika, operacijske analize, matematičke biologije, numeričke matematike, statističkih metoda, računa vjerojatnosti, knjižničnih sustava i algebarske logike.
I na sveučilišnima ustanovama u Zagrebu pojedini su profesori već u 1970-tima usmjeravali djelatnost u pojedinim predmetima prema sustavnoj obradi koju bi računski stroj mogao prilično olakšati i ubrzati. Bili su to Ekonomski fakultet, Elektrotehnički fakultet, Filozofski fakultet, Fakultet za fizičku kulturu, Prirodoslovno-matematički fakultet, Referalni centar Sveučilišta u Zagrebu i Sveučilišni računski centar. Informatički je odbor pri Sveučilištu proveo anketu da se ustanovi koji su to nastavnici i u kojoj ustanovi rade. Anketa je na opće iznenađenje pokazala da s najvećim brojem predmeta informacijskoga sadržaja raspolaže Filozofski fakultet. Tu su, da spomenemo samo po jedan predmet: filozofija (s formalnom logikom), fonetika (s teorijom informacije), komparativna književnost (s knjižničarstvom), lingvistika (s matematičkom lingvistikom), pedagogija (s programiranim učenjem), povijest (s arhivistikom), povijest umjetnosti (s muzeologijom), psihologija (sa statistikom) te sociologija (s metodama istraživanja).

Tada se je nametnula zamisao da bi se na Filozofskome fakultetu mogao pokrenuti međuodsječni studij društveno-humanističke informatike. Ostali su fakulteti počeli osnivati informatičke studije svojega usmjerenja.Visoka je ekonomska škola iz Varaždina preimenovana u Fakultet organizacije i informatike na prijedlog Radoslava Katičića, člana Savjeta Sveučilišta u Zagrebu.
Dvogodišnji je studij društveno-humanističke informatike i četverogodišnji studij informacijskih znanosti na Filozofskome fakultetu ustrojio i bio voditeljem Katedre i pročelnikom Odsjeka Bulcsú László. Za Katedru je knjižničarstva zaslužan Gajo Peleš, nakon čega je katedru vodila i bila prvi predstojnik katedre dr. sc. Ljerka Markić-Čičuković, za Katedru arhivistike Ivan Kampuš, za Katedru muzeologije Ivo Marojević. Poslije je Miroslava Tuđmana zapalo voditeljstvo Katedre za dokumentalistiku, a Eduarda Hercigonju Katedra za pismene sustave. Napokon je na Odsjeku za opću lingvistiku osnovana Katedra za algebarsku i informatičku lingvistiku pod voditeljstvom Bulcsúa Lászlóa, kojega je potom naslijedio Miro Kačić.
Malo je komu poznato koliko je muke i borbe trebalo da se novouvedeni studij na Filozofskome fakultetu ustali. Za studij su se zalagali i sami slušači omladinci-aktivisti, posebice Zlatko Papeš (sadašnji prof. informatike), ali se je Prosvjetni savjet oglušio na potrebe ondašnjice, dapače, u nedostatku smjernice odozgo, izrijekom zabranio daljnje upisivanje studenata, jer da takav profil stručnjaka našemu društvu nije potreban. Da nije bilo razumijevanja u dekana Milivoja Sironića i prodekana Eduarda Hercigonje studij bi bio obustavljen. Ali obojica njih su se založila za to da Filozofski fakultet vlastitima sredstvima osigura održavanje produžnosti nastave, dok je voditelj studija morao obilaziti poduzeća prikupljajući izjave da su informatički stručnjaci njima, i našemu društvu uopće, prijeko potrebni. Kada je u Fakultetskome vijeću upozorio da se informatika neviđenom brzinom razvija te će se uskoro svi odsjeci i svi nastavnici morati baviti informatikom, članovi su se vijeća u nevjerici smješkali.

Priprema

U Sveučilišnom odboru za informatiku vodila se žestoka rasprava o dodjeli, matičnosti za obećavajuće područje nastave informatike. Riječ je informatika francuska kovanica, tzv. portemanteau riječ, sklopljena od izraza information automatique. Zbog automatike su se za informatiku smatrali matičnima isključivao Elektrotehnički i Prirodoslovni fakultet. Tomu se je uskomu matematičko-tehnološkomu shvaćanju odlučno protustavljao Božo Težak, osnivač i ravnatelj Referalnoga centra Sveučilišta u Zagrebu. Kao prvi i najžešći promicatelj društveno-humanističkoga i poslovnoga shvaćanja informatike pozivao se na široku rusku definiciju informatike, koja pritom obuhvaća i ključnu ulogu matematike i tehnologije.
Kružni se tijek informacija po Težaku može sažeti u prvoslovnoj pokrati ETAkSA. Tu *E* upućuje na emisiju dokumenata (proizvodnju dataka), *T* na transmisiju (prijenos) do odredišta, *Ak* na akumulaciju (prikup) na stanovitome

mjestu, *S* na svrhovitu selekciju (odbir), *A* na apsorpciju (usvajanje) dobivenih informacija za stvaralačku obradu i tvorbu novih znanja, koja se i opet emitiraju.
Službena definicija Američkoga društva za informacijsku znanost i tehnologiju glasi: Informacijska se znanost (vjestoslovlje) bavi proizvodnjom, prikupljanjem, ustrojem, tumačenjem, pohranom, traženjem, razašiljanjem, preoblikom i porabom informacija s posebnim naglaskom na tehnologiju.
Prijepor je oko sadržaja informatike jednim potezom razriješio Stanko Turk s Elektrotehnike dosudivši informatici ulogu nižega reda opsluživanja stroja za rješavanje zadataka, jednom riječju programiranje. Sveučilišne su se ustanove morale dogovoriti za koji će dio široko shvaćenoga informacijskoga područja biti nadležni.
Elektrotehnički se je fakultet usmjerio na računarstvo, informacijsku tehnologiju, na gradnju računskih strojeva, Prirodoslovno-matematički fakultet na primijenjenu matematiku, Fakultet organizacije i informatike na poslovnu informatiku, Sveučilišni referalni centar na obučavanje programiranja i nastave primjenom računskih strojeva, Sveučilišni računski centar na potporu članovima Sveučilišta u primjeni informacijske i komunikacijske tehnologije, a Filozofski fakultet na informacijske znanosti.

Izvedba

Na Odsjeku za informacijske znanosti danas djeluju Katedre za arhivistiku i dokumentalistiku, bibliotekarstvo, društveno-humanističku informatiku, knjigu i nakladništvo, leksikografiju i enciklopediku, muzeologiju i organizaciju znanja. To su naočigled raznolike struke, ali ih sve zajedno veže zajednička potka – stručno znanje kao proizvod informacijske tehnologije:

- **Arhivistika i dokumentalistika** – digitalni arhiv s namjenom čuvanja sve vrste arhivskoga gradiva bez obzira o kojem se nosaču radi i kada je nastao. Gradivo se digitalizira, arhivi povećaju bez prostornoga širenja i čuvaju prošlost daleko pouzdanije i vjernije no što je to ikada bilo moguće.
- **Bibliotekarstvo** – digitalna knjižnica za pristup k proizvedenoj knjižničnoj građi u najširem smislu, izgradnja mrežnih pomagala za opis, pohranu, pristup, pretragu i vrjednovanje građe. Digitalne se knjižnice izdvajaju kao posebna mrežna služba zbog uvriježena obrasca o posebnosti knjižnične građe. Knjiga je povijesno gledajući prvi masovni proizvod informacijske tehnologije.
- **Društveno-humanistička informatika** – izgradnja i oblikovanje obavijesnoga prostora (information space), uspostava mrežnih služaba (network services) za pristup i upravljanje mrežnim sadržajima (network resources, information management).

- **Knjiga i nakladništvo** – znanja i vještine za primjenu informacijske tehnologije u proizvodnji predmeta znanja za digitalne knjižnične, arhivske i muzejske zbirke.
- **Leksikografija i enciklopedistika** – digitalni ustroj jezičnoga blaga i cjelokupnoga činjeničnoga znanja ljudskoga roda. Budući da su nosači znanja oduvijek bile knjige, one su i dalje cijenjeni predmeti znanja zajedno s knjižničarskom znanošću koja njima rukuje.
- **Muzeologija** – digitalni muzej s namjenom predstavljanja i opisa muzejskih predmeta. Svjetske muzejske zbirke u digitalnu obliku čuvaju materijalnu i posredno duhovnu baštinu čovječanstva. U informacijskim se znanostima smatra da svaki zamišljiv predmet može biti nosač obavijesti ili informacije. Muzejski predmeti izravan su dokaz te tvrdnje.
- **Organizacija znanja** – obuhvaća metodologiju navedenih područja i predmete znanja opisuje i prikazuje prateći razvoj područja i tehnologije. Držimo da će upravo organizacija znanja u neku ruku sva ova područja objediniti, tj. izgraditi prihvatljivu zajedničku okosnicu. Proučavanju će se i poučavanju o predmetima znanja moći u budućnosti pristupati na sličan način bez obzira radi li se o arhivskome gradivu, knjižničnoj građi, muzejskim predmetima ili mrežnoj službi.

Studij se informacijskih znanosti na Odsjeku obavlja prema bolonjskim načelima. Nastavno je osoblje uključeno u domaće i međunarodne projekte uz potporu Ministarstva znanosti, tehnologije i športa Republike Hrvatske. Nadalje se njeguje međuodsječna suradnja i čuva zajednička društveno-humanistička i znanstvena zasada Filozofskoga fakulteta na kojoj se temeljilo osnivanje studija. Kroz niž projekata u okviru Odsjeka počeo je i razvoj pomagala za elektronično učenje (e-learning). Odsjek je tako pomogao izgradnju fakultetskoga obavijesnoga prostora s vlastitim predmetima znanja.

Znanja se i vještine potrebne za izvođenje nastave usklađuju s europskim normamama prema zahtjevima novoga načina studiranja. U tome prednjače mlađi naraštaji nastavnoga osoblja koje svoju izobrazbu dopunjavaju na američkim i europskim sveučilištima. Međunarodna suradnja u vidu razmjene studenata i profesora, sudjelovanje na konferencijama i radionicama vezanima za studij, ukazuje da naš Odsjek ravnopravno sudjeluje u razvoju informacijskih znanosti.

Zaključak

Trideset godina nastavničkoga i istraživačkoga rada uzajamno se dopunja i omogućuje daljnje i buduće plodonosno sudjelovanje u razvoju informacijskih znanosti. Cijeli je fakultet umrežen i računarske su učionice mjesta za upoznavanje i primjenu proizvoda informacijske tehnologije. Okosnica se našega pristupa k obrazovanju informacijskoga stručnjaka temelji:

- na vještinama pretraživanja (information retrieval, IR) i procjeni za izbor obavijesti (information extraction, IE),

- na primjeni međunarodnih standarada za označivanje predmeta znanja i njihovo okupljanje u digitalne zbirke (web application formats, web standards, metadata digitalization standards),
- na oblikovanju baza znanja (knowledge database),
- na strojnoj obradi prirodnoga jezika i jezičnima tehnologijama (natural language processing, text segmentation, natural language generation, speech recognition, text-to-speech synthesis),
- na ontologiji ili bazi znanja o informacijskim znanostima koje sada shvaćamo kao razvrstavanje (categorization) informacija u tematske razretke koji su još uvijek posebno imenovani kao arhivistika, bibliotekarstvo, dokumentalistika, muzeologija, leksikografija itd. Svaki se tematski razredak može pretraživati, ako je predmet znanja opisan propisanim i dogovorenim međunarodnim komunikacijsko-tehnološkim standardima i tim načinom bivati dijelom dio informacijskoga prostora (information space).

Upotrebljena građa

Andrijašević, Marin. Povijest Odsjeka. http://www.ffzg.hr/oling/povijest.html (27.10.2007.)

Dragojlović, Pavle. Informatika : uvod u suvremenu obradu informacija s rječnikom stručne terminologije na engleskom i hrvatskom književnom jeziku. Zagreb : Školska knjiga, 1977.

Horvat, Aleksandra. Povijest Odsjeka.
http://www.ffzg.hr/infoz/web2/index.php?option=com_content&task=view&id=3&Itemid=4 (27.10.2007.)

Informacijke znanosti i znanje. Zagreb : Zavod za informacijske studije, 1990. (Radovi Zavoda za informacijske studije ; knj. 1)

Informacijske znanosti u procesu promjena. Zagreb : Zavod za informacijske studije, 2005. (Radovi Zavoda za informacijske studije ; knj. 15).

Modeli znanja i obrada prirodnoga jezika. Zagreb : Zavod za informacijske studije, 2003. (Radovi Zavoda za informacijske studije ; knj. 12).

Obrada jezika i prikaz znanja. (Radovi Zavoda za informacijske studije ; knj. 5). Zagreb : Zavod za informacijske studije, 1993.

Konferenciju su podržali

VJESNIK

Zagreb
CROATIA

Turistička zajednica Grada Zagreba

KRAS

BCC services

Eko-monitor d.o.o.

Zavod za poslovna istraživanja
Institute for Business Intelligence