Vrana, Radovan; Badurina, Boris; Golub, Koraljka

Department of Information sciences, Faculty of Philosophy

I. Lucica 3, 10000 Zagreb, Croatia

Advantages and disadvantages of use of digital collections in the process of education

Introduction

In last few years the Internet has become a true component part of the process of education in many countries in Europe. This global computer network has brought us many types of applications of computer technology suitable for learning. However, there are still some methods and procedures which are part of this global network and are still not sufficiently used. As a result of this opinion, a small group of students of the 4th year of library and information sciences at the Faculty of Philosophy in Zagreb, proposed a project which, they think, would solve the problem of inadequate quantity of printed exam materials at the Faculty necessary for the successful completion of exams by use of a digital collection containing equivalent ready-for-use material in electronic form.

Use of digital collections as a possible solution to the problem of inadequate quantity of printed exam materials

As we are approaching the new millenium, we can look back and say that this last decade has been characterized by the development of the new information sources known as digital libraries, libraries without walls or virtual libraries. The foundation of these new information sources is human knowledge organized according to some principle in form of digital collections. These digital collections can be later used in a local environment for various purposes (mostly for education) and eventually grow in size and combine into larger clusters of knowledge with other local collections thus creating global network of knowledge. Furthermore, recent scientific works agree upon the fact that the future of the knowledge organization will be in form of highly organized structures of knowledge, based on indexed digital collections. On the local level such knowledge structures will share their resources by use of local area networks, while on the global level they will probably use the Internet which has already become a new paradigm for interconnectivity among various information sources in the world today. As it is self-evident, during the last few years, the explosive growth of the the Internet has given immense contribution to the development of digital collections by giving opportunity to authors of such collections to communicate and share their experiences with their colleagues around the world thus creating another type of global network, network of experience and specific problem solving knowledge applicable to the creation of future knowledge structures.

When speaking about the Internet, one can't avoid mentioning World Wide Web as the most popular Internet service today. World Wide Web has come a long way since its early days of the text-only content and is now ready for bigger tasks. The level of maturity reached by Word Wide Web gives us a chance to incorporate this Internet service in

projects which need a medium for accessing organized information resources on remote locations in easy and usually uniform manner. Local information sources can then become more widely accepted and attract more potential users accustomed to the graphical user interfaces of the most popular browsers. This idea has gained its popularity mainly because of the well developed software industry which understood properly potentials of the Internet both in academic and in commercial domain. So called dumb terminals used by skilled professionals are finally things of the past and can no longer threaten any user by their inflexibility. Modern browsers with their functional and ergonomically built graphical user interfaces bridged the gap between the local information sources and their limited acceptance (only) by local users and wider audience, helping people to use large information sources on remote locations in simple and uniform way.

By enumerating these well known facts we have tried to point out some issues and trends which we think are important and are already present in the teaching process at the Faculty of Philosophy in Zagreb and some other institutions in Croatia. Our main idea is to give students at our Faculty at least equal opportunity to access and use the Internet and World Wide Web in their everyday learning routine.

How important this can be showed us the survey conducted at the same Faculty last year. This survey indicated rather small number of students using the Internet as the preferred information source for their term papers or final exam at their respective departments. Their number is slowly increasing at rate at which more and more students have access to the Internet. Due to circumstances, students at the Faculty of Philosophy in Zagreb have limited access to the Internet services. They can access the Internet either in Faculty's computer classroom which is to small to satisfy the needs of all students at the Faculty or from their home PCs. The number of users in both groups is indeed small and many of them still prefer using printed material over the Internet for the preparation of their exams, but they are showing strong inclination to learn more about computers in general and how to access the online content prepared specially for them to help them in their education. This means that there is a large number of future Internet users who explicitly asked for more seminars, exercises and lectures dealing with computer literacy. To round up this thoughts, we can say that these are some of the ideas which contributed to the digital collection project proposal. Another big problem which is not directly related with the computer literacy but is crucial for the success in the exam taking process is the quantity of the printed exam materials. Again, the survey showed us that our public libraries, our national library and some university libraries do not have enough printed material for all students particularly during the period of exams. This major problem is already dealt with but there is still much to be done to meet basic student's needs for exam materials of high quality and appropriate quantity. Possibility of loan or acquisition of articles and books necessary for the successful completion of exams from libraries are rather low due to the inadequate library funding, resulting in restrictions in acquisition of new titles. One possible solution to this problem is of course to raise more money for specific library collections or to propose a project which would bridge the gap between students' demand and libraries' ability to answer that demand. We took the second course with a hope to help the students of librarianship with specially prepared digital collection which would serve as a possible substitute for non existent printed material and would become new source of knowledge.

This assumptions lead us to the main part of this paper about the digital collection project for the students of the library and information sciences. The basic idea for this project came

from a group of students of the 4th year at the Chair of librarianship at the Faculty of Philosophy in Zagreb: Koraljka Golub, Boris Badurina and Mihajlo Zoric who decided to build a digital collection of selected printed and electronic materials necessary for the preparation of exams at the same Faculty. Under the supervision of the assistant professor Tatjana Aparac, Ph. D. and research assistant Radovan Vrana, this project came to light very quickly. The choice of a medium for this project was very obvious, we chose the Internet as an omnipresent environment in which the project would eventually get its full acceptance. We all participated in the last year's survey and we all knew what we should do in order to make things better and easier for the rest of the students population.

Project preparation and phases

Our project has two main phases and each phase has its key points.

Phase one:

- collecting of the relevant information about similar projects in Croatia and in the world
- making of the project plan
- collecting and selection of the printed and electronic material for the collection
- digitization i.e. scanning and OCR
- transformation of the OCRed material into WWW pages
- transfer of the digitized material onto the Faculty's WWW server

Second phase has slightly less key points but they are not less important than those from the first phase:

- solution to the copyright problem
- obtainment of the additional disk space on server for the collection expansion
- creation of the searchable index catalog for the collection

Although one can put some remarks on the choice of the key points in the both phases, there is logical explanation to the each key point of this project. The most obvious deficiency would be the omission of the copyright problem from the first phase. During that first phase which is still in progress, we used printed and electronic materials for which we had clear situation concerning the author's rights. This was possible because we had chosen specific exam materials written by our professors and by people who gave their explicit permission for the use of their material in our digital collection. If it would have been different we wouldn't use their intellectual property to protect their rights from infringement.
When we were setting the main goal of this project, we had in mind solely students of the librarianship at our Faculty and this determined the nature of our collection. All participants were aware of the significant material quantity, so we decided to digitize reading lists for the specific seminars first as our direction mark for the future digitization process. This procedure gave us the exact number of articles waiting to be transformed into the electronic form. Once we finished this phase, we started with the process of digitization of the exam materials. At the very beginning of the project, we knew that the main source for our collection will be articles in various librarianship journals. We also wanted to digitize some

books but then we would have to deal with the copyright problem more extensively than planned, so we decided to postpone it until the second phase of our project. We collected most of the articles quite easily and didn't have any major problem except in situations when all copies of the articles were already loaned.

Digitization

Generally speaking, there are three methods of digitization which we used in our project:

- digitization of the printed materials
- transcribing of the printed or/and written materials
- collecting materials already in the electronic form on disks, CDs etc.

The choice of a method depends directly on the nature of a material which is to be digitized or simply included in the collection if it is already in electronic form. Most of the material included in this project were printed articles from various journals from librarianship and information sciences. The articles contained mostly text or text accompanied by illustrations or photos. When the process of article collection was successfully completed, we selected the method of digitization suitable for each piece of our future digital collection. During the collection creation process we used the first method most frequently and of course we chose the two most popular procedures: scanning and optical character recognition (OCR). The analysis of the usage of these two procedures has lead us to a new problem, the problem of speed. This problem was closely related to the OCR and recognition of special symbols and diacritic signs incorporated in the Croatian written language. Another big obstacle was the slowness of proofreading after the OCR was finished, not to mention the manual correction of errors. We also wanted to have a core literature digitized before the main exam period starts, so the time component was crucial to the success of this project. Nevertheless, we managed not to compromise the quality of the digitized material. As a possible alternative to the OCR we thought of scanning of the printed material and saving it to the Adobe Acrobat .pdf format but we didn't have additional PCs with large disk arrays for temporary storage of large quantity of scanned texts.

The second method - transcribing - was used only with the core material reading lists, where quantity was not an issue. Each reading list contained only articles and books references and this part of the job was quickly and easily finished without major problems.

The third method was the most popular one because we received some materials in electronic form which gave us some additional time to spend on scanning and OCR of the printed materials. We expect that this method will eventually become dominant as soon as we exhaust all sources of printed materials available to us. We also expect authors to give us their work directly on computer disks in order to speed up the process of the collection expansion in the future. This method requires only small amount of time dedicated to the formatting of the text and possible conversion of graphic material into the format appropriate for the World Wide Web.

Output format

One of the most important issues in this project was the question of the output format once the scanning and OCR processes were completed. At the same time, this is one of the biggest problems in the world of digital collections creation because of the interoperability issue. In their development phases digital libraries had served to local (academic) communities and had existed only in local environments, but in time they outgrew that environments and tended to connect with other libraries of similar nature. If their content is saved in different file formats those libraries will have difficult time exchanging their knowledge because they will loose too much time on various types of format conversions. In this project, we concluded that it would be best if we choose the output format right in the beginning of the project thus avoiding the problem of interoperability once other university libraries decide to use knowledge stored in our collection.

The format of our choice is Hypertext Markup Language or HTML. This is very popular kind of language for content structure description which helped us to structure the OCR output and store it as the plain text with HTML tags. A few years earlier our choice would probably have been Standard Generalized Markup Language or SGML, but today, it is HTML and our main reasons for this decision can be described as:

- simplicity of this markup language
- short time necessary for the learning of this language
- large number of available software tools for the creation and editing of documents in HTML
- possibility of expansion of the HTML core element set with new tags in the future
- availability of the most popular HTML browsers such as Netscape Communicator or Internet Explorer.

Each of these points can be percepted as easy choice for someone without experience in the document creation and exchange on the Internet but for others, the choice of the file format is not an easy choice mainly because of the rapid expansion of the Internet and its services. The Internet services are notorious for fast shape changing and one cannot easily determine or decide on the file format in situation in which the chosen format can become obsolete in few months time. HTML is now widely accepted as the markup language and users of digital libraries should benefit from its use.

Simplicity of this markup language is another key argument for its application in our project. There are some deficiencies in HTML but at the same time this markup language offers a lot, especially when the time component is important and this leads us to the new argument, one speaking about ease of learning of the HTML.

HTML can be indeed easily learned because of the numerous tutorials and software applications written specially for this purpose. Students of the librarianship didn't have any formal knowledge of HTML before they decided to propose the project but they learned its basics very fast. In the future, Chair of librarianship plan to offer some workshops in HTML in order to prepare more students for the participation in similar projects or just to help them to find their way on World Wide Web. Of course, there is always the possibility of learning HTML alone but this method has some significant drawbacks although knowledge acquired this way can be later expanded through additional seminars and workshops or even projects like this one.

As it was mentioned before, there is a large number of software applications which can be use in the creation of digital collections. The development of such software is heading towards the more visual tools which do not require any specific or thorough knowledge of HTML elements or tags. This could help us in the future if we decide to ask for more students to participate in project.

Each collection document was coded with the basic HTML element set in order to avoid problems related to the proprietary tags or elements introduced by Netscape or Microsoft. This means that our digital collections can be easily accesses even with the text mode browsers like lynx and of course with more advanced browsers such as Netscape Communicator and Internet Explorer. Another key issue regarding the design of pages was the uniformity or similarity. We wanted our documents in collection to look more or less the same except in cases where it was necessary to make changes which would alter the final design of pages due to some peculiarities like graphics within pages etc.

As this project will move forward, we plan to use more advanced applications of HTML like JavaScript but only when necessary. One such case will be our evaluation survey written in HTML with additional browser windows created in JavaScript in order not to compromise the collection structure and to point out its flexibility for such tasks.

At the end, we plan to use all advanced features of the future HTML standards in situations which will seem appropriate thus expanding our collection in such a way in which our students could benefit best from it.

Copyright

Next important issue is related to copyright. From the very beginning of this project we tried to make copyright central to all our efforts. Solving the problem of another or/and repeated public manifestation of some author's intellectual property seemed to us of utmost importance. We had clear vision what we want and how to do it. An indication of the importance of this issue were situations in which we made somebody familiar with our project, and this person always asked about copyright and how we dealt with it.

Authors usually have fear for their work in electronic environment because electronic works can be easily altered if they are not protected legally and physically from the misuse. This peril isn't fictional at all and it should be treated with great care. This project has strict guidelines concerning the copyright and we will try not to compromise the results of our efforts.

In this special case (of our collection), we contacted all persons whose works were enumerated on the reading lists and asked for their specific permission for the inclusion of their works in our collection. Since this project is still in its early phase, we only had to contact our professors and people from the librarianship community in Croatia. As our project will move forward, we will try to include some works of foreign origin in our digital collection, accompanied by appropriate written permissions for this incorporation if necessary. At the moment of writing of this paper, not a single author whose work is already part of the collection has refused us, on the contrary, they were very eager to participate. We expect this trend to continue.

Access to the collection

As it was stated at the beginning of this paper, access to our digital collection of exam materials is the key issue of the whole project and it reaches into the very essence of the effort. Let us point out once again the fact that the goal of this project is to give the students of librarianship at the Faculty of Philosophy ready for use materials necessary for the successful preparation for their exams. We are hoping that this collection will be also used by other students at Zagreb University, as well as by students out of it, for instance, students of the newly formed Study of librarianship at the Pedagogical Faculty in Osijek and by all other students and scientists whose field of interest will be covered by materials from our digital collection.

Having in mind various different applications of Information Technology (IT) in education and inclination of our University in Zagreb for greater use of computers in the process of education, it was decided that we should make this digital collection public on the Internet, or to be precise, on the World Wide Web. The collection itself is located at the main World Wide Web server at the Faculty of Philosophy in Zagreb which is an Digital Alpha server running Digital UNIX / OSF 1 operating system and Netscape FastTrek WWW server. This server is well equipped with sufficient disk space and with appropriate application software which is of great help in the maintenance of the collection. Faculty of Philosophy is connected by the high speed fiber optic ATM link (155 mbit/sec) to the rest of our Croatian Academic and Research Network (CARNet) i.e. Internet. This kind of support guarantee us future advancement of the project without glitches and problems. There were some suspicions about inadequate amount of the disk space but this was quickly solved and we do not expect such problems in the future.

Although our last year survey showed students' inclination towards the traditional printed sources of knowledge and less towards the use of computers due to objective reasons, we consider this state only temporary and we think that it is absolutely essential to give our students pointers to use computer facilities and our Internet link at the Faculty of Philosophy in greater extent and finally to show them advantages of such electronic knowledge organization structures which will soon become as ordinary for use as our libraries are today. Computer facilities at our Faculty include one major access point and that is our computer classroom whose primary purpose is support for the computer assisted learning (CAL). This computer classroom is very often used by students exclusively for their lectures and seminars which doesn't leave them much free time to explore the Internet on their own or to use computers for the exam preparation and to become more self-reliant when using computers without supervision. This classroom is equipped with 20 new Pentium PCs connected to the local area network and to a Windows NT server equipped with a proxy server connected to the Internet. All PCs have installed necessary Internet software such as Netscape Communicator or Internet Explorer which enables students to use all Internet services according to their needs. Such software configuration eliminates some problems related to the problem of access of the large number of people (students) to the Internet with different hardware and software configurations. This will also help us later when we will evaluate effects and results of our digital collection availability among students' population to eliminate the factor of hardware and software diversity. This access point is almost ideal because it provides equal opportunity for all students to access digital collection through similar browser interface the way it was conceived in our project's plan. Disadvantages of this access point can be summed up in inadequate number of computers in comparison to the total number of students of all four years of the study library and information sciences. In

addition to these problems, we noticed the lack of skill and knowledge in the Internet navigation among some students when accessing electronic information sources.

Another way to access the digital collection is to use a PC and a modem from home. It is very hard to evaluate this type of access to information resources because there is no valid feedback. The user feedback could help us to plan further expansion of the digital collection, improve the access to it and to determine optimum size of material in the collection. Access from home implies different hardware and software configurations which influence the user experience during the access to the collection. If there is too much Java on web pages, or there is a large chunk of text to download, there is always a possibility for user to give up retrieving text from such a collection. In the end, we would like to measure both the students' access from home and at the Faculty so we can get the exact picture of the collection usage.

Project costs

When speaking about digital libraries and the process of digitization, one almost always has in mind project costs. From that point of view, our project is rather unusual because we can't present our expenses in such a way in which digitization of an article page would cost 2 USD or equivalent in some other currency. This project is not financed in any special way either by the Faculty of Philosophy or by the Ministry of Science and Technology. Preparation and selection of material is done at the Faculty but actual digitization is done at students' home due to lack of appropriate equipment at the Faculty. We only use Faculty's World Wide Web server in order to present our work and this is the only piece of equipment financed and maintained by the Faculty.

Experiences and problems

This part of our paper is in direct connection with the its title since experiences and problems we encountered describe advantages and disadvantages of the use of the digital collection in the process of education. Current state of our project singled out some advantages which can be further analyzed during the evaluation process:

- greater accessibility of the exam materials in comparison to the printed materials in our departmental library
- choice of the output of the articles from our collection either in electronic or in printed format
- digital collection offer selected and quality material
- student/user does not necessarily have to be at the Faculty to use the collection, instead, they can use modems and work at home
- there is no access restrictions, everyone (interested) can access the collection
- work with the collection at hand can teach students how to be more independent when using IT

Of course, there are always disadvantages to every project. We expect that more of them will probably show up during the more intensive periods of the collection use, but we will

certainly try to solve all the problems and to improve all segments of the collection. Some of the problems we have encountered are:

- problems with the process of article collection, some journals were missing from the library
- problems with the OCR, some texts were damaged, letters were to small and there was always a problem of manual error correction after the OCR
- lack of the proper digitization equipment, at least one or two quality scanners are required
- small number of access points to the collection
- only three students participating in the project

Evaluation

As this project is still in its development phase, it is not possible for us to make full evaluation of its effects and results. However, if we consider the basic idea behind this project, the one about making the core literature necessary for the preparation of exams available to students on the Internet, we can conclude about some effects of our work. Students at the Chair of librarianship are partially introduced to the collection through contacts with their professors or announcement on the Chair's bulletin board and announcement on the Chair's World Wide Web homepage. The project was initiated before the official exam period in order to prepare core literature for the exams and to receive at least oral feedback from the students.

More thorough evaluation of this project's results will be done by surveying the students' population at the Faculty of Philosophy, especially students at the Department of Information Sciences and Chair of librarianship. This survey will probably be taken during the new academic year when students will be formally introduced to the collection. It is expected that new generations will use the Internet in greater extent and they will be probably more interested in reading of the material in electronic form. Survey seems to be the best method of evaluation because it can be taken during different periods of academic year and the results of these surveys can be later compared to the initial feedback information at the very beginning of the project i.e. now.

Another method of evaluation can be done in the online environment. We think that the first introductory page of our collection will be the right place for a quick online survey. The moment the student accesses the collection, he/she will be presented a survey with several question crucial for the further development of the collection. This type of survey will be taken by use of special browser window written in JavaScript which will open during the access to the collection. It will contain prepared questions and will close as soon as the student answers the questions. This way we won't compromise the structure of the collection and we will successfully finish the evaluation.

The last method of evaluation includes monitoring of the WWW server log files. This is very popular method in the world of information retrieval and can be also used in our project. As soon as we announce to our new generation of students the existence of our digital collection, we will start monitoring log files and analyzing them in order to see which parts of the collection are most interesting to the students.

Conclusion

In this paper we tried to show one possible way of using information technology and the Internet in the process of education at the Faculty of Philosophy in Zagreb. Projects like this one, with students directly participating in them from the beginning, shows us what direction in using information technology in education we should take and what methods we should use. By accepting digital collections as one of their crucial information sources, students accept new technologies introduced to them on everyday basis and also learn how to use the online content for their study. If this project proves to be successful, and we think it will, in the future we will try to attract more students to participate in similar projects, in order to create new digital collections as well as to expand the existing one according to the needs of all students interested in such an information source.

References

1. Cleveland, Gary. Digital Libraries : Definitions, Issues and Challenges. http://www.ifla.org/udt/op (12/6/1998). 8 pp.

2. Faulhaber, Charles. Distance learning and Digital Libaries : Two Sides of a Single Coin. http://www.carl-acrl.org/Archive/Conference95/faulhaber.html. (1/5/1998). 3 pp.

3. Fox, Edward A.; Hix, Deborah; Nowell, Dennis et al. Users, User Interfaces, and Objects : Envision, a Digital Library. Journal of the American Society for Information Science. 8(1993). pp. 480-491

4. Fox, Edward A.; Marchionini, Gary. Toward a Worldwide Digital Library. Communications of the ACM, 4(1998), pp. 29-32

5. Levy, P.; Fowell, S. Networked Learning in LIS Education and Training : the paper presented at the EUCLID-FID/ET Conference, 21-22 November, 1995, Copenhagen

6. Paepcke, Andreas; Chang, Chen-Chuan K; Garcia-Molina, Hector; Winograd. Interoperability for digital libraries worldwide. Communications of the ACM. 4(1998), pp. 33-43

7. Ross, Seamus; Economou, Maria. Information and Communications Technology in the Cultural Sector. D-Lib Magazine. 6(1998). http://www.dlib.org/dlib/june98/06ross.html. (15/6/1998). 9 pp.

Radovan Vrana I was born in 1970. in Zagreb. I finished my primary and secondary education in Zagreb and enrolled the Faculty of Philosophy in Zagreb in 1989. I studied the English language and Information sciences and got my B.A. degree in 1996. I'm now working as a research assistant at the Department of information sciences at the same faculty and preparing my M.Sc. thesis. Fields of interests: library management, building of digital libraries.

Boris Badurina I was born in 1974. in Zagreb. After I finished the High school for electrical engineer I enrolled the Faculty of Philosophy in Zagreb where became a student of sociology. In 1996. I enrolled the Chair of librarianship as a additional study. In addition to being a student, I have a per time job in a library. Fields of interest: surveying the public opinion, digital libraries.

Koraljka Golub I was born in Cakovec in 1975. I finished my primary and secondary school in Cakovec and in 1994 I came to Zagreb to study English language and Information sciences at the Faculty of Philosophy. Now I have finished my lectures and I have a few more exams to take. I am particularly interested in the Internet and translation studies.