Software for digitization of ancient glagolic handwritten manuscripts

Nevena Ackovska, Mile Jovanov,
Darko Aleksovski, Vasil Atanasov, Bojan Endrovski,
Institute of Informatics, Faculty of Sciences, Skopje, Macedonia

Background

Macedonia is the birthplace of the glagolic alphabet. In fact, some of the oldest and most significant texts written in the glagolic alphabet are found in the Republic of Macedonia. Every nation is recognized by its tradition and cultural heritage. Therefore, it is crucial, especially in the age of globalization, for every nation to protect and cherish its national heritage. The most significant historical evidences are written documents.

Problem description

The crucial part of the problem is to transform the written glagolic manuscripts into digital form, so they could be easily processed, manipulated, stored and searched through. For this purpose, a special kind of tools, called Optical Character Recognition (OCR) tools are developed. There are several software tools for optical character recognition, but there is none for the glagolic alphabet. The goal is to develop an OCR glagolic tool for educational purposes.

Solution techniques

There are several approaches to this type of problems. The document should be transformed into digital form, so it could be processed. Digital photography or scan of the document is the most usual technique used. The result that is obtained is a digital image that contains data about the color and intensity of every pixel of the document.

This digital document should be post-processed, to ease the process of the recognition of the glagolic letters. This post-processing consists of:

- binarization – getting the image of grayscale or color into black and white linear picture;
- noise reduction;
- segmentation of lines, words and letters.

The last and most significant phase is the pattern recognition of the letters themselves. Some of the techniques already tested are:

- vectorization, that can be of curve fitting type, poligonization or finding control points (characteristic shapes) and then recognition; or

- classification with classificators (linear classificators or neural nets).

Implementation

As has already been mentioned, the process of text recognition consists of a sequence of several phases. At present, the first phase is not implemented in our application directly; rather, it relies on specialized applications for document scanning. It is recommended that the document be scanned instead of photographed, because, during photography, geometrical distortions often arise, which the application is not able to correct automatically.

The next step is to transform the scanned document into a form from which the characteristic shapes could be extracted more easily. This step commences with the image binarization. Binarization is a process of obtaining an image composed of two colors only – black and white; both grayscale and color images are transformed into strictly black and white images. In the case of an image containing text this means separating the text from the background. The documents written in the glagolic alphabet are often old and paled, which certainly makes this process more difficult. In order to overcome this problem, thresholding is used; the threshold is determined adaptively for the specific document. The image that contains binary information often contains noise as well. This noise can hinder the detection of characteristic shapes in the document, so it is necessary that it be reduced. This is accomplished by using various noise reduction filters.

The next step is the segmentation into rows, words and letters of the document. For successful recognition, the beginning and end of each row containing text must be determined, as well as to segment the words, using blank spaces. If possible, it is advisable to segment the letters from which the word is contained, based on the spaces between them. Unfortunately, in handwritten text that is not simple at all. Namely, it is very difficult, and sometimes impossible, to determine a single bounding box for the rows, words or letters, due to various reasons: very often, during writing, the author binds the letters with one another; very often certain letters from one row intrude into another row of letters; the blank spaces between words are not of equal sizes, etc.

The key phase in the recognition process is the recognition of the segmented text itself. To accomplish this, some of the classification techniques using classificators are used. The classificator receives an input in the form of a 12x12 matrix, representing one letter of the text. This classificator should first decide whether the input is a glagolic letter, and if it is, which letter is it.

The last phase in the entire process is the presentation of the recognized text. The obtained data is stored in a specialized data file.

Word Image Matching in Bulgarian Historical Documents

Andrey Andreev, Nikolay Kirov
IMI, Bulgarian Academy of Sciences, Sofia, Bulgaria,

Keywords: *document text image, bitmap file, word matching, Hausdor_ distance*

1 Introduction

An approach to word image matching based on Hausdorff distance is examined for bad quality typewritten, printed or handwritten Bulgarian documents. A detailed computer experiments were carried out using 49 pages typewritten text, 13 pages printed text and 2 pages of a manuscript. The results of several methods are compared including previously reported methods in the literature. The Hausdor distance used in the paper difers slightly from ones used by other authors and the conclusion from the results is that our method outperforms them despite its simplicity.

Let A and B denote bounded sets on the plane and a and b be points on the plane with coordinates $a = (a_1, a_2)$, $b = (b_1, b_2)$. The Hausdorff distance (HD) between two bounded sets A and B is defned in [1] for the purposes of approximation of discontinues functions as

$$r(A,B) = \max\{h(A,B), h(B,A)\} \tag{1}$$

where

$$h(A,B) = \max_{a\in A} \min_{b\in B} \rho(a,b) \tag{2}$$

$$\rho(a,b) = \max\{|a_1 - b_1|, |a_2 - b_2|\} \tag{3}$$

In 1994 Dubuisson and Jain [2] examined 24 distance measures of Hausdorff type for determination to what extend two point sets on the plane A and B differ. In case when the sets A and B consist of N_A and N_B points along with (3) changed to Euclidean distance they use

$$h(A,B) = \frac{1}{N_A} \sum_{a\in A} \min_{b\in B} \rho(a,b) \tag{4}$$

and claim that among all 24 “distances” examined by them, this “distance” called “Modiffed Hausdorff Distance” (MHD) suites in best way the problem for object matching. Similar approach called “Weighted Hausdorff

Distance" (WHD) is used in [3] for finding word image matching method in English and Chinese document images.

This research has been supported by a Marie Curie Fellowship of the European Community programme "Knowledge Transfer for Digitalization of Cultural and Scientific Heritage in Bulgaria" under contract number MTKD-CT-2004-509754. The work has been done while the authors were at the National Center of Scientific Research "Demokritos", Athens, Greece.

We propose to simplify (4) using

$$h(A,B)=\sum_{a\in A}\min_{b\in B}\rho(a,b) \tag{5}$$

and call the distance (1), (3), (5) Sum (or Simple) Hausdorff Distance (SHD).

2 Distances used in computer experiments

Let us define "word image" - a rectangular image which pixels have values 0 (white) or 1 (black) and "word" - a subset of word image with pixel values 1. The following distances will be tested numerically for estimation of similarity between two words A and B:

1. $L_1(A,B)=\sum_{a\in(A\setminus B)\cup(B\setminus A)}1$;
2. $HD(A,B)=r(A,B)$ where $r(A,B)$ is defined by (1), (2), (3);
3. $HD_1(A,B)=r(A,B)$ where $r(A,B)$ is defined by (1), (5);and
$$\rho(a,b)=\begin{cases}0, & if \quad a=b\\ 1, & otherwise\end{cases}.$$
4. $MHD(A,B)=r(A,B)$ where $r(A,B)$ is defined by (1), (3), (4);
5. $SHD(A,B)=r(A,B)$ where $r(A,B)$ is defined by (1), (3), (5).

Before using a given distance for estimation the difference between two images theymust be adjusted with respect to either their geometric centers or to their mass centers. Forexample if SHD distance is applied combined with geometric center adjustment of images we denote this by SHD^{gc} otherwise we write SHD^{mc}.

Measuring the effectiveness of the distances (or methods connected with them) usually is given by standard estimations *Recall* and *Precision* [4]. Briefly, let us look for a word W in a collection of binary text images in which W occurs N times. Let the method produce a sequence of words

$$\{W_i\}_{i=1,2,\ldots} \tag{6}$$

ordered according to a specific criteria. For a given n (n = 1, 2,…), let $n_1 \leq n$ be the number of words among the first n words of (6) that coincide with W. Note that n_1 is a function of n. Then we define

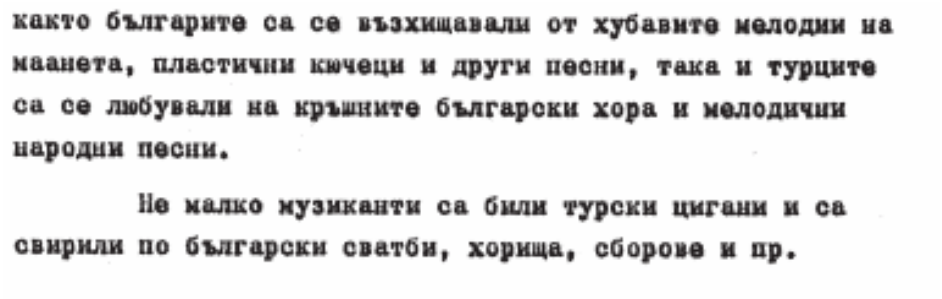
$$Recall(n) = \frac{n_1}{N} \quad \text{and} \quad Precision(n) = \frac{n_1}{n} \tag{7}$$

as functions of n.

3 Experimental results

3.1 Typewritten text

Using the distances defined above we carry out a series of computer word matching experiments. Real Bulgarian documents of typewritten text of 49 pages of bad quality are the material from which a specified word is located and extracted.



Word “песни” (the sixth word from the second row in the text above): occurrence 31 times.

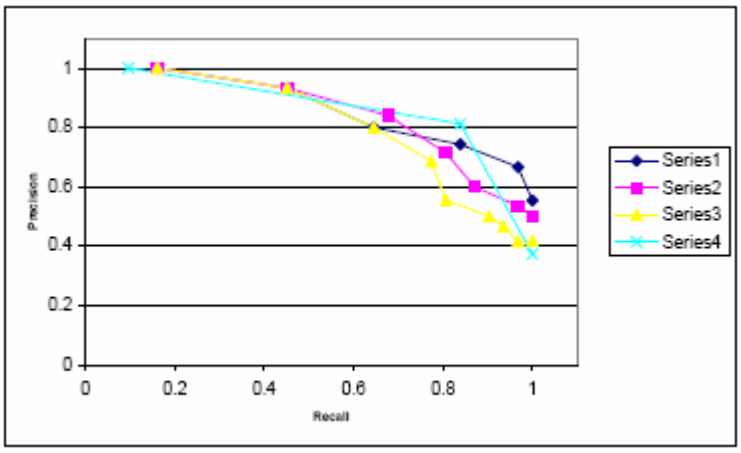
Series2: MHD^{gc}			
n	n_1	*Recall*	*Precision*
5	5	0.16	1.00
15	14	0.45	0.93
25	21	0.68	0.84
35	25	0.81	0.71
45	27	0.87	0.60
56	30	0.97	0.54
62	31	1.00	0.50

Series3: HD$^{gc}_1$			
n	n_1	*Recall*	*Precision*
5	5	0.16	1.00
15	14	0.45	0.93
25	20	0.65	0.80
35	24	0.77	0.69
45	25	0.81	0.56
56	28	0.90	0.59
62	29	0.94	0.47
72	30	0.97	0.42
74	31	1.00	0.42

Series1: SHD^{gc}			
n	n_1	*Recall*	*Precision*
5	5	0.16	1.00
15	14	0.45	0.93
25	20	0.64	0.80
35	26	0.84	0.74
45	30	0.97	0.67
56	31	1.00	0.55

Series4: HD^{gc}			
n	n_1	*Recall*	*Precision*
3	3	0.10	1.00
32	26	0.84	0.81
83	31	1.00	0.37

The results from the last 4 tables above are plotted on the following picture:



3.2 Printed text

We process 13 pages containing printed text written in proportional font.

Бих желал да кажа, че аз като математик дълго време смятах, че тази дуалност има своите ограничения, защото, на пръв поглед поне, би изглеждало, че математиката принадлежи изцяло на първия архетип. Но това е късоглед възглед: погледната отгоре, математиката ясно спада също така и към втория архетип. Може би развитието на тази мисъл да е твърде привлекателно и да заслужава отделен разговор: понастоящем в математиката има

The test word “математика” can be found in the text as a word before the last one. But our searching template “математик” consists only of the first 9

letters form it. Thus we can find all words who have the same 9 letters at the beginning.

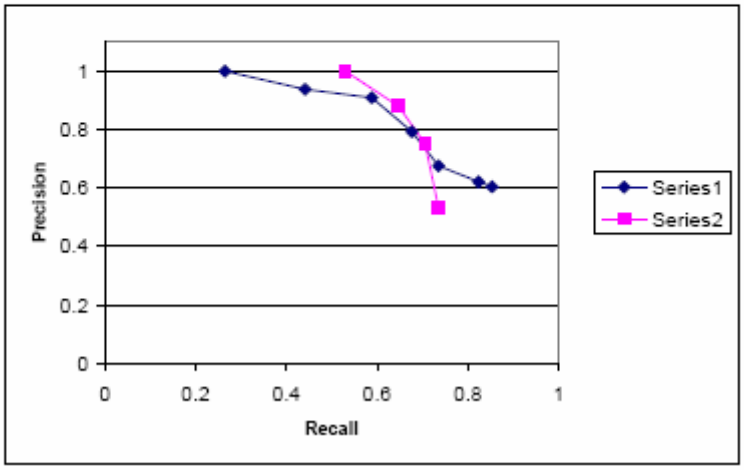
The results of the word recognition process depend on several factors:

- differences in the original document;
- differences in the process of scanning;
- the binarization step (grayscale-b/w conversion and noise removal) can make additional differences;
- the segmentation step may extract same words in a di_erent way.

Our example includes two different binarizations which gives two di_erent recognition results. They are presented in the tables and in the figure as Series1 and Series2.

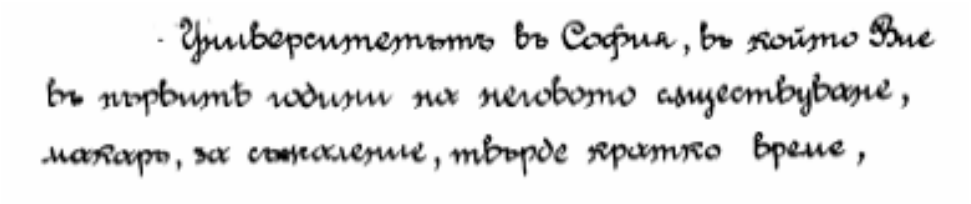
Series1			
n	n_1	Recall	Precision
9	9	0.26	1.00
16	15	0.44	0.94
22	20	0.59	0.91
29	23	0.68	0.79
37	25	0.74	0.68
45	28	0.82	0.62
48	29	0.85	0.60

Series2			
n	n_1	Recall	Precision
18	18	0.53	1.00
25	22	0.65	0.88
32	24	0.71	0.75
47	25	0.74	0.53



3.3 Handwritten text

We process two handwritten pages from a document created in 1929.

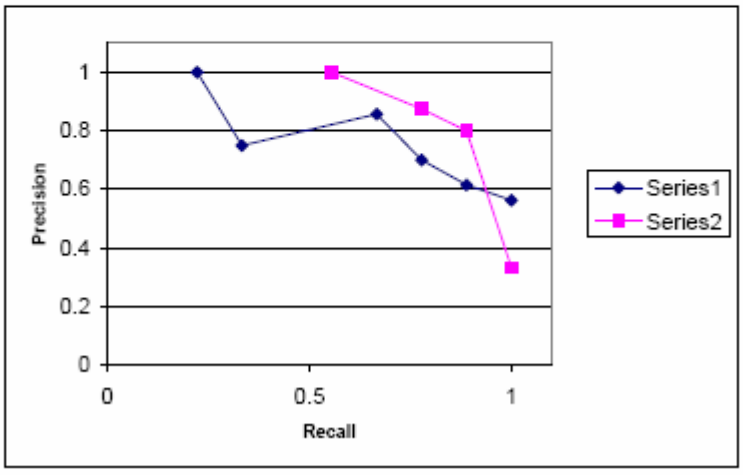


We search for the words, shown below that begin with the pattern “Ваш” (“Ваш” means “Yours”).

The computer results are:

Series1 - SHD^{gc}			
n	n_1	Recall	Precision
2	2	0.22	1.00
4	3	0.33	0.75
7	6	0.66	0.86
10	7	0.78	0.70
13	8	0.89	0.62
16	9	1.00	0.56

Series2 - HD^{gc}			
n	n_1	Recall	Precision
5	5	0.56	1.00
8	7	0.78	0.86
10	8	0.89	0.80
27	9	1.00	0.33



4 Conclusion

We process bad typewritten Bulgarian text, printed text and a manuscript for word matching using various distances. The results show that:

- The distance SHD^{gc} produces better results than other distances and therefore there is no need to complicate the de_nition of SHD (like MHD or WHD).
- Mass centered adjustment *mc* of word images is inappropriate for the purpose of word matching.
- L_1^{gc} distance produces the worst results. HD_1^{gc} method which is a sort of a combination of L_1^{gc} and SHD^{gc} behaves better, but evidently falls back to SHD^{gc}.
- The measurement done by HD^{gc} distance could be considered as a "discontinuity". This explains the deterioration of the results produced by HD^{gc} for values of $Recall(n) \approx 1$. For example, for the short word "песни" with occurrence 31 times HD^{gc} finds:

HD^{gc} distance	No. of words found n	No. of correct words n_1
2	2	2
3	29	23
4	31	5

In this sense the other methods use practically continuous scale for ordering the spotted words.

References

[1] Bl. Sendov, *Hausdorff Approximations*, Kluwer Academic Publishers, 1990.

[2] M.-P. Dubuisson, A. Jain, *A Modified Hausdorff Distance for Object Matching*, In: Proc. 12th Int. Conf. Pattern Recognition, Jerusalem, Israel, 1994, pp. 566-568.

[3] Y. Lue, C. L. Tan, W. Huang, L. Fan, *An Approach to Word Image Matching Based on Weighted Hausdorff Distance*, "6th ICDAR_, 10-13 Sept. 2001, Seattle, USA.

[4] M. Junker, R. Hoch, A. Dengel, *On the Evaluation of Document Analysis Components by Recall, Precision, and Accuracy*, Proceedings ICDAR 99, Fifth Intl. Conference on Document Analysis and Recognition, Bangelore, India, 1999.

Mobile Context in Promotion of Cultural Heritage

Ljupco Antovski, Marjan Gusev, Goce Armenski
Institute of Informatics, Faculty of Sciences,
Ss Cyril and Methodius University, Skopje, Macedonia

Introduction

The impact of culture on the development of new industries is one of the most crucial questions facing the Information Society. The concept of Cultural Heritage is fundamental to the creation and expansion of the global Information Society.

So the information technologies can be fruitfully exploited to permit a wide access to the cultural heritage information. On the one hand, people can perform virtual visits from whatever they are in the world. On the other hand, physical visits can be supported by innovative technologies that provide enhanced services and make the visit more comfortable. If we consider the information about cultural heritage spread over the Internet such as in a grid, it could be very difficult for a user to deal with all the cell of the grid without a loss of time.

Cultural and natural heritage applications have proved to be an attractive vehicle for mobile computing researchers. Several projects have developed data collection tools, museum, site or city visitor guides as a means of demonstrating various concepts including location and context awareness and smart building environments. Together, these represent two ends of the production process of bringing cultural and natural heritage from the research environment to its consumers. The intermediate analytical, archival and curatorial stages are less well represented, but also provide considerable scope for further research.

Mobile Computing and Context

Mobile computing has been an active area of research for the past years. However, research in the field suffers from fundamental methodological weaknesses. The approaches are somewhat empirical, giving a sense that the designers already know what systems to build and what problems to overcome. Real world cultural studies are not considered important and no special effort has been put in delivering solid design methodologies for mobile applications. As a result, various issues remain to be investigated: First of all, effective and efficient positioning and context awareness methods and models. In other words, models to present useful information to

the user with respect to the information communicated to him/her by the environment.

Models to invert the flow of information; concerning context of use in mobile device; i.e. the flow should be inverted compared to a desktop computing environment. Information should be contextualized and personalized according to personal needs, and presented to the user rather than having the user searching endlessly for useful information. This step should be achieved through a deeper understanding of the tasks in certain environments and clarification of cognitive issues related to those tasks. As a result, to achieve interface transparency, the process should be viewed as an effort to seamlessly integrate the computational device to our natural environment.

The mobile devices should not be treated as an autonomous environmental device, but as an artifact seamlessly integrated into the environment. The addressing the question of finding the best ways of making use of the real to support the virtual and visa-versa is of central importance. Therefore, the notion of context is of fundamental importance to anticipate the design challenges in mobile applications.

Existing Mobile Systems

Most of the studied systems use PDAs as a basic mobile interaction device because of the technical characteristics. The PDAs are lightweight, internet-connected, have better presentation capabilities, the user interface is easy managed and offer the possibility of interaction using Bluetooth, IrDA and other emerging short range communication technologies.

The system in the Marble Museum of Carrara was one of the first fully operating systems in a museum. This system stores the information locally in the PDA's memory, uses a map to guide the visitors around the museum and presents content of different abstraction levels (room, section and exhibit).

The "Electronic Guidebook", deployed in the Exploratorium, a science museum in San Francisco. The system tries to involve the visitors to directly manipulate the exhibits and gives instructions and additional science explanations about the natural phenomena people are watching.

The ImogI system is located in the Gallo-Roman Museum of Tongeren and uses Bluetooth to establish communication between the PDAs and the exhibits, and reflects the closest exhibits to the location of the user.

The "Sotto Voce" system is deployed the Georgian Revival House in Woodside, California and gives details about everyday things located in an

old house, by having pictures of the walls on the PDA's screen and asking from the user to select the exhibit he/she is interested in, by pressing it.

The "Points of Departure" system, in San Francisco Museum of Modern Art, gives details, in video and audio form, about the techniques used in an artwork, the message that the artist wants to demonstrate etc., by having 'thumbnails' of several exhibits on the PDA's screen. Also it uses the concept of Smart Tables in order to enrich the interaction.

Technology Approaches

The technology influences the interaction between the user and the environment. The system context focuses in the technological parts that constitute the system, either software or hardware, and in the connections between them and the environment.

In all the examined systems, only one design decision was common for every research group. PDA was the primary device to help the visitor interact with the heritage site exhibits. In order to serve this purpose four different technological approaches, using PDAs, where adopted by the system designers.

In the first approach, the PDA consist the whole system. There are not any other devices or awareness mechanisms involved and the information presented to the user is stored locally in the PDA.

The second approach uses RFID tags to establish communication between the PDAs and the exhibits. It uses cameras to take pictures from the visitors interacting with the exhibits. Through Wi-Fi, the comments that the users wrote in the PDAs and the pictures taken, are send to a web server, stored in the form of web pages, allowing each user to remember her personal visit afterwards.

The next approach uses Bluetooth to establish communication with the exhibits and to deliver the content. The other systems include IrDA technology to estimate the position of the visitor in space. The IrDA tags are placed near every exhibit or in the entrance of each exhibition room. The use of Wi-Fi derives the information to the PDA from a server. Many different additional devices built and integrated in these systems. Some designers placed screens in specific places, either as a standalone device like the 'Smart Table' in 'Points of Departure', or as interacting devices with the PDAs, where the user has the opportunity to transmit sequentially his/her interaction with the system from the PDA to the Screen.

For instance, in the experimental project influenced by the Museum of Fine Arts in Antwerp, the system forces the users to use cameras attached to the PDAs, take an image of an exhibit they wish to have information about

and then see the information in the screen of the PDA. In the Point-it application in Tokyo the visitors use a laser pointer to inform the system about the exhibit they are interesting in, or in the Museum AR a simple look is enough.

When comparing the capabilities of the mobile devices, there are some differences that may result from the decision on selecting between Pocket PC and Palm OS operating systems. Each of these two operating systems has its advantages and disadvantages. Pocket PCs have bigger screens, higher resolution and support the 'landscape mode' of projecting info, if necessary. They can also play MP3s and multimedia in a more effective way compared to the Palm OS. They support four ways of text entry, in contrast to Palm OS, which supports two. In contrary, Palm OS devices are considered easier to use, they have an increased battery life, a lower cost and require fewer computational resources. As a result, depending on application, a careful choice has to be made, in which kind of operating system supports better the needs of a visitor.

Except from hardware, this dimension involves also software components, or more specific, the set of applications that constitute the system. There is an awareness factor that implies that the user must know each time which applications are available, and in which state these are. All the interconnections between applications are part of system context. In such distributed applications, it is important the feedback that the user gets from the system about the existing devices in the system, the existing applications etc.

The metaphor that describes the interaction with the RFID tags is the use of the PDA as a scanner. It can be hard to locate the RFID card, and if the exhibit has a small size, no more than one people can interact with it simultaneously. With the use of IrDA, problems also exist. Use of many IrDA beacons may be prohibitive, since they are more expensive and there can be mixing signals, if they are placed next to each another. They operate in a higher distance than the RFIDs. We must underline that the use of RFID or IrDA, solves the problem of the bond between the physical and the virtual space, but in different ways.

Final Remarks

In this overview we presented how outdoor tracking, mobile computing, 3D visualization, and augmented reality techniques can reconstruct ancient sites, simulate old style of life and revive cultural heritage sites.

The features of the mobile systems used at heritage sites are expected to increase visitors and create additional revenue. At the same time the on-

line tours will help advertise the site and make it better known to the wider public. Finally, researchers will find a useful companion in their research and reconstruction work.

However, some of the systems still being a prototype, it needs further development mainly in 3D tracking and building custom-made mobile devices of compact and lightweight design suitable for carrying especially in a outdoor environments

The Recording and Studying of Serbian Traditional Dances

Vesna Bajić
Senior Student of Ethnomusicology
Faculty of Music, Belgrade

Keywords: *Labanotation, research methods, analysis of the dance kolo in three, digitalization*

The recording of traditional dances in Serbia began in the 1930's with the first thought-out system authored by sisters Danica and Ljubica Janković, pioneers of Serbian ethnochoreology. In eight books entitled *Narodne igre (Folk Dances)*, published in the period from 1934 to 1964, the authors elaborated on and further developed their system of notation, working at the same time on collecting and systematizing traditional dances. Precise, simple, visually clear and easy to use, the system of the Janković sisters met with approval of many domestic and foreign experts. The said system comprises pattern, analysis and melody of the dance, auxiliary signs, diagrams and graphs. However, the system of the Janković sisters proved insufficient for scientific research initiated in Serbia at the beginning of the 1970's. The reason for this was the variability of special signs for recording movements that were recorded by words only, that is, descriptively.

The intensive scientific work of Olivera Vasić, Ph.D., conducted at the end of the 1970's, brought about an increased use in Serbian ethnochoreology of a widely accepted way of recording movements, namely Knust-Laban's notation or *Labanotation* for short. Already in the 1950's experts in the field of dance from the region of the former Yugoslavia became acquainted with the principles of Labanotation and considered the positive and negative aspects of its possible application in the future, albeit without any definite results. It was not until 1990, when the subject Ethnochoreology was introduced at the Faculty of Music as part of the general studies of ethnomusicology, that the students were presented with the opportunity to familiarize themselves, under the guidance of Prof. Vasić, with a new kind of notation, Labanotation, which had by then become standard in Europe. Up until today this system, which has been the basic starting point in ethnochoreological works of different content and scope, has been in use in regular studies.

Labanotation uses a small number of signs that are logically placed in space and can, according to circumstances, be supplemented or modified so as to represent the most diverse movements of man as precisely as possible. The final version of this work will include a presentation of the system of Labanotation adapted to the specificities of Serbian traditional dances.

With the foundation of the *Centre for studying folk dances of Serbia* at the Faculty of Music, Labanotation became available to amateurs, too. As part of a seminar organized by the Centre, courses on Labanotation are held on a regular basis and a booklet is published containing dances of a particular area written in this notation. However, Labanotation is very slowly gaining ground among choreographers, amateurs of traditional dance and amateur researchers. Most of them find it unclear, abstract and unpractical which is why a safer method is still in use, namely description of the dance by words, often accompanied by individual signs. The reason behind this is inadequate training of the individual studying the dance and, on the other hand, the doubt whether Labanotation can actually include every detail of the dance.

The recording of dance and music is art of a particular kind and it requires appropriate education. Unlike notating music, the recording of movements is more complex as dance is a temporal and spatial art. The notation of traditional dance consists of three basic constituents of dance, these being time, space and movement. Such tridimensionality of dance translates into linearity on paper, which is a very complex and time-consuming job. In order to achieve objectivity in the scientific sense and at once avoid subjectivity in description by words, it is necessary to keep developing the sign system. Individual systems of notation, starting from the Janković sisters, must be "translated" into another, standard sign system such as Labanotation in order to avoid heterogeneity of individual notations. Only in this manner is the preservation of cultural heritage possible.

When notating, one should consider every dance from all aspects: examine the title, area of diffusion, historical development, dance event (time and place of the performance, participants, musical accompaniment) and record the dance pattern by the standard system. Great attention should be paid to the recording of steps and movements, their creation, way of transformation, variation, intertwining and dynamic and rhythmic complexities. This is fundamental to analytical work and to studying dance units, micro- and macrostructure (motives, phrases, compositional models).

With the adoption of Labanotation, Serbian ethnochoreology witnessed an increased interest in analytical work. Based on the recorded dances from the times of the Janković sisters and by transposing their

notation into a new visual idiom, as well as based on subsequent research that has continued up to the present, a systematization of the types of dance in Serbian dance heritage took place. Today's most diffused type of Serbian folk dance is *kolo in three*, which is usually shortened to *kolo* among the people. On the example of this dance pattern the final version of the work will show how Labanotation can be used for analytical purposes. Special attention will be called to the development of the dance pattern from dichotomous to trichotomous structure, which is a characteristic of this type of dance. Using a video recording, the author will demonstrate how the dance pattern of *kolo* is recorded and how it has compounded. This process reveals a diversity in embellishment (the enrichment of dance by leg movements), which is peculiar to Serbian dance heritage. Everything that can be detected on the basis of field recordings, even if it includes individual creations of the folk dancer, is studied and recorded. Permeating the collective and individual dance is that which could be termed *style* in ethnochoreology, and it directly points to the specificities of a particular ethnochoreological area.

As the system of Labanotation had been precisely determined, a computer program called *Calaban* was designed in Europe in the 1990's. Unfortunately, the *Centre for studying folk dances of Serbia*, as the only institution in Serbia that archives dance heritage, does not possess this program. Folk dances are still recorded manually, with a pen and a ruler. The gathered field material is published through collections, which certainly is an important albeit not final step in organizing material. Computer processing would facilitate the systematization and use of that material. In addition, the ample video library of the Centre is privately owned by rare enthusiasts, senior associates of the Centre, because an appropriate strategy concerning the archiving of video recordings still does not exist. Forming an archive and its corresponding system is an important step in the development of Serbian ethnochoreology and efforts in that field are yet to be made.

The preservation of cultural heritage is based not only on archiving material, but also on its practical application. Young people have the opportunity to familiarize themselves with the dance heritage of Serbia and other countries by virtue of a large number of cultural clubs, festivals and other events, but an appropriate governmental institution that would educate professional dancers, pedagogues of traditional dance and, above all, ethnochoreologists, still does not exist.

Digitalization of Heritage in Serbia

Ivan Blagojević, Marijana Simu, Ivana Pražić
Beograd, Serbia and Montenegro

Keywords: *Heritage Areas Standardization, Heritage Resources, Systematization, Virtual Research Environment, Network-base, Access, Online Heritage, Virtual Learning Experience*

Background

Present-day situation concerning heritage in Serbia reflects previous fifteen years of general social, cultural, economic and professional deterioration. Dominant approach towards preservation/conservation, presentation and exploitation of cultural and environmental, material and non-material heritage in Serbia still needs to be redefined and remodeled according to contemporary global trends. It is true that certain initiatives by governmental, public, and civil sector have been undertaken in order to revitalize and/or transform the inherited institutional system related to preservation/conservation, and presentation of cultural and environmental heritage. However, lack of systematic approach towards standardization of the fields involved with heritage in Serbia becomes a serious obstacle for its future development and preservation. The most obvious weakness of the mentioned institutional system is reflected in departmentalization of heritage into disassociated fields of environmental and cultural, material and non-material legacy. Just as classification of a cultural/environmental entity as the World Heritage does not rule out its local particularity, integrative treatment of heritage does not preclude its diversification.

Having been recognized as the most systematic means of collecting, preserving and presenting all material and nonmaterial heritage "goods", digitalization is the best solution to the mentioned problems. On the one hand, digitalization represents basic prerequisite in creating an integrative approach towards heritage. On the other, it could also prove to be an essential instrument in achieving the goal of introducing modern and internationally acknowledged standards.

Unfortunately enough, large majority of institutions involved with preservation/conservation, presentation and exploitation of cultural and environmental, material and non-material heritage in Serbia is far from utilizing digitalization. Furthermore, a data-exchange network of the abovementioned institutions has not yet been developed. There is

insufficient digital information about heritage in Serbia, and online catalogues are also not available.

Therefore, the need for representation of all relevant data in a unique interoperable format is quite obvious. Such format
would be based on the internationally acknowledged standards providing it with compatibility and widespread usability.

Short Description of the Project

The project focuses on development of framework for creation of cultural and environmental, material and non-material heritage virtual repository, and software tools for its multi-level efficient usage. Beside this, project implementation includes education of personnel involved in heritage digitalization. In this manner, heritage in Serbia would become more accessible to the experts and the public. At the same time, its integration into broader international community would be made easier. One of the advantages of the project is its low cost software system. Due to this fact, wide usage of this system even by small institutions with limited budgets.

Project Goals

Short-term Goals:

- to build and catalog a virtual repository of heritage resources;
- to provide integrated network-based access to repository;
- to establish electronic metadata link among relevant institutions;

Long-term Goals:

- to enhance public awareness of cultural and environmental, material and non-material heritage;
- to facilitate integration of cultural and environmental, material and non-material heritage in Serbia into broader
- international community;
- to improve performance of relevant heritage institutions in Serbia through introduction of international digitalization
- standards;
- to provide easy-to-use gateway for the researchers and the public;
- to emphasize interactivity in both presenting and understanding heritage;
- to preserve cultural and environmental, material and non-material heritage in Serbia through recording.

Anticipated users

The public—individuals who require detailed information about cultural and environmental, material and non-material heritage in Serbia for various reasons: research, education, information, and commercial use. These are mainly users from different scientific communities, students, journalists, authors, etc. The most important service for this user group will be provision of easy and comfortable access to a large number of items in heritage repository. Access is enabled via web portal with variety of search options and user interfaces supporting interactive learning environment.

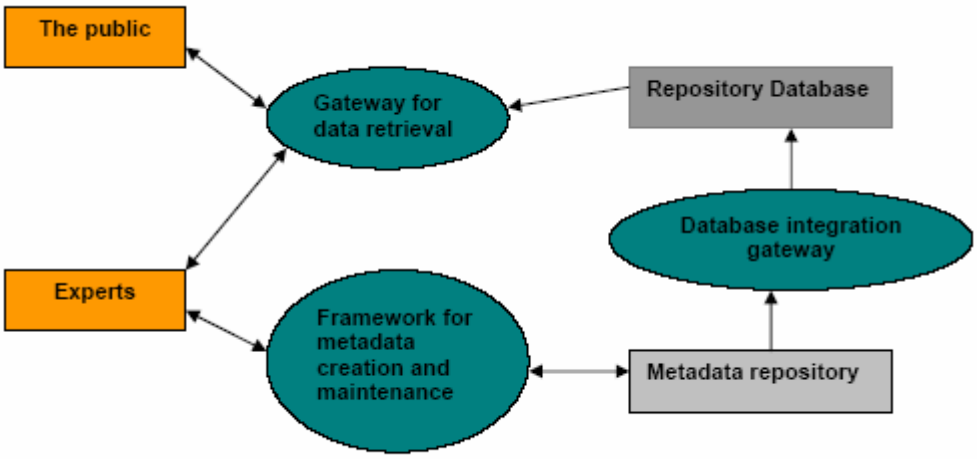
Expert users—museum operators, environmental heritage operators, archive operators, librarians, theoreticians and research staff working in relevant institutions. This group of advanced users will have access to the services for repository data creation and maintenance.

Technical issues

Technical side of the project is rather complex, and this paper will cover only its basic aspects. With strong foundations in XML-based web-applications, adopting the Dublin Core Metadata standards and recommendations, project defines specific metadata format for description of heritage resources. Metadata creation and management is enabled via web-based tools, independent of heterogeneous platforms and accessible to the users via personal computers and web protocols.

Project also aims at developing a WWW gateway, which will provide data retrieval by the public users. The repository content is searchable on many levels that match various data entries of respective items: author, period, title, time and place of its origin/submission, content of the item, etc. Hardware requirements are low both for the data providers (i.e., relevant institutions), and the end users (i.e., the public).

The system architecture design is given below. Both public and expert users use web clients to access application layer. Gateway for data retrieval, framework for metadata creation and maintenance, and database integration gateway are based on Java 2 Enterprise Edition (J2EE) platform and developed using JSP and Java Servlet technology, and XML standards. Metadata repository is set of SGML files (xml, html). Repository database can be any relational database. Project authors suggest free solutions such as MySQL or Javabased hSQL.


The public
Gateway for data retrieval
Repository Database
Database integration gateway
Experts
Framework for metadata creation and maintenance
Metadata repository

The Pythagorean Rule and its Application, According to the Viennese Hellenic Philological Code 65 of the 15th Century (P.11r-126r)

Maria D. Chalkou
Greece

The Viennese Hellenic Philological Code 65 (Codex Vindobonensis phil. gr. 65) is made of paper and dates from the 15th century.

The Code means that the pages (or the volumes) are not joined into a roll, but were simply placed between wooden frames or frames of some other material. It is considered to be a pioneering form of today's book. It appeared in the 2nd century B.C. and due to its ease in use it replaced the cylinder. At about the same time, parchment (usually the hide of a cow of sheep) as writing material, replaced the papyrus. The Code consists of joined volumes and each volume of one (alternating) page number, folded in half.

The author and the origin of Viennese Hellenic Philological Code 65 are unknown. Augerius von Busbeck got the Code when he was Ambassador to Emperor Ferdinand I in the courtyard of Sultan Souleiman II (1555-1562 B.C.) . The pages 126v -140r contain a book of Arithmetic (Arithmetic includes logisticae and geodesia) with solved problems, which was published by H. Hunger and K. Vogel in 1963. The larger part of the Code (p. 11r - 126r) contains an anonymous arithmetic book of which, the preface and the first two chapters were published by J. L. Heiberg in 1899

The eleventh unit of the unpublished section of the Code (chapters 167, 184), of which the first part consists of Geometry, includes problems, which are solved mainly by the use of the Pythagorean rule or "rule of three", as it was called by the author of the Code. The methodologies of the solutions, even if in some cases are probably unknown to the mathematicians of the secondary education, as we ascertain from the examples which follow, adequate common factors with those used today in corresponding problems.

Creating Electronic Critical Editions

R. M. Cleminson
University of Portsmouth / Central European University, England

Keywords: *critical edition, variant, collation, XML, XSLT, TEI*

The paper deals with the creation in electronic form of editions of texts provided with an apparatus indicating variants (popularly known as critical editions, although strictly speaking not all such editions are critical in the proper sense of the word). Existing standards for the encoding of a critical apparatus have been developed with a view to encoding critical editions which have already been created manually, and are therefore very inconvenient for the creation of an apparatus *ab ovo*. In particular, the insertion of new variants is likely to require the manual re-numbering of the entire sequence, just as it would in the revision of a paper edition.

Automatic collation programmes, though they will produce an electronic apparatus, require the encoding of the entire text of each manuscript which it is desired to collate. This may not be desirable or indeed possible, and in any case involves a very considerable amount of work.

Neither approach is therefore ideal for the scholar who wants to edit a text from primary sources with a more or less extensive indication of the variants which they contain.

The proposed solution is to encode the entire text as an XML document (the "base document") with a very simple structure, essentially dividing the text into a linear, non-hierarchical series of segments, each segment consisting of a section of text plus the variants to that section with an indication of the witness(es) in which they are found. The portions of text (either single words or more extensive passages) to which there are variants are then numbered automatically by means of XSLT. Further XSLT scripts are used to mark the passages of text and the corresponding variants on the basis of this numeration, and then to extrapolate the variants from the text into an apparatus. The scripts can be combined in a batch file so that the critical edition can be generated from the base document in a single operation. The result is again an XML document with a relatively simple structure, which can be further transformed into a TEI-conformant document, since its elements correspond to elements within the much more

complex TEI system. Additional mark-up (textual divisions and subdivision, reference systems etc.) may be added at this stage if required.

The advantage of this method is that it allows modifications to be made with the minimum of labour: for example, variants from a new manuscript may be added to the base document, and then a new edition generated with the variants renumbered automatically and no danger of loss of correspondence between text and apparatus. Equally, the scholar in the process of making an edition can use the method to view the output at any stage of its creation.

System Fedora in building digital libraries for works with old books, manuscripts and others cultural collections

Nada Djordjevic
Faculty of Mathematics, Belgrade, Serbia

All cultural institutions today are faced with the challenge of how to promote wider access to, and greater public use of, their collections. The Fedora system (Flexible Extensible Digital Object and Repository Architecture) demonstrates how distributed digital library architecture can be deployed using web-based technologies, including XML and Web services. The new system is designed to be a foundation upon which interoperable web-based digital libraries can be built.

Everything but the smell: toward a more material digital philology

Matthew James Driscoll
Arnamagnæan Institute, University of Copenhagen

Keywords: *digital editions, material philology, textual theory*

Modern textual criticism is defined as 'the science of restoring texts as near as possible to their original state'. It was developed in the 18th and 19th centuries chiefly by classical and biblical scholars, and quickly adopted by scholars in other fields. The method most commonly employed, the enealogical' or 'stemmatic' method, is normally associated with the name of the German philologist Karl Lachmann (1793-1851). Although it has had its critics, most notably the French mediaevalist Joseph Bédier (1864-1938), the method still underlies the bulk of scholarly editorial practice today.

The principal innovation in the area of editorial theory in the last decade or so has been the advent of 'new' or 'material' philology, which is generally recognised as having been launched, in the English-speaking world at least, with the publication in 1990 of a special issue of Speculum: A Journal of Medieval Studies, edited by the romance philologist Stephen Nichols of Johns Hopkins University in the United States. The immediate inspiration for the new philology came in particular from Bernard Cerquiglini's short polemical essay *Eloge de la variante: Histoire critique de la philologie* (Paris, 1989); in a more general way its antecedents include: Paul Zumthor and the concept of 'mouvance', put forward in his *Essai du poetique medievale* (Paris, 1972); the development within Anglo-American bibliography (under the influence of French work in *histoire du livre*) culminating in D. F. McKenzie's Panizzi lectures from 1985, published the following year as *Bibliography and sociology of texts*; work in the German-speaking world from the 80s onward on the history of transmission, *Überlieferungsgeschichte*; and the extensive work on oral literature and questions of orality and literacy.

The key principles of the new/material philological programme may be said to be the following:

1. Literary works do not exist independently of their material embodiments, and the physical form of a text is an integral part of its meaning; one needs therefore to look at 'the whole book', and the relationships between the text and such features as form and layout,

illumination, rubrics and other paratextual features, commentary, marginalia, and, not least, the surrounding texts.

2. These physical objects came into being through a series of processes in which a (potentially large) number of people were involved; and they came into being at particular times, in particular places and for particular purposes (which were socially, commercially, intellectually determined); these factors too influence the form the text takes and are thus also part of its meaning.

3. Instability ('variance', 'mouvance', 'unfixedness') is a fundamental feature of chirographically transmitted literature; it, in fact, what pre-modern textuality is 'about'.

Although the reaction among textual theorists to the new philology has on the whole been favourable, the reaction among those actually involved in scholarly editing has tended to be dismissive, usually on the grounds that 'there is nothing new in the new philology — in fact, we've been doing it all along', or that any kind of philology which does not seek to separate original readings from corrupt ones is no philology at all.

So, is any of this true? Is there nothing new in the new philology? In one sense, no, in that the new philology obviously grew out of the dominant trends in literary critical thought from the 60s onwards, generally subsumed under the heading post-structuralism, which de-emphasised the importance of the author and authorial intention and focused instead on the collaborative nature of literary production, dissemination and reception and the cultural, historical and ideological forces at work in these processes. At the same time, of course, the new philology exhibits features which have always been present (at least potentially) in philological study, but with a very different set of priorities. So, no, there is nothing new in it at all — apart, that is, from its entire orientation.

And have we always been doing it? I think not. Even where we have based our editions on single manuscripts ('best-text' editions), rather than attempting to reconstruct lost archetypes, we have not paid much attention to the physical artefacts themselves or the processes through which they have come into being. Our focus has still been on 'the text' in an abstract sense, and our search essentially still one for origins.

As to whether we ought we to be doing it, I think the answer is definitely yes. And luckily, we now have a means of doing so: electronic editions using XML markup.

Fundamental to the material-philologically-oriented edition would be an awareness of manuscripts as cultural artefacts which — among other things — serve as vehicles for texts. The texts would be presented with as

little interpretation as possible, allowing the reader to appreciate the interplay between form and meaning to a greater extent than has been possible with traditional printed editions. I hasten to add that I am not talking about any 'editorial death-wish', or 'abdication of responsibility'. To these 'level zero' transcriptions levels of interpretation would be added, as indeed they must if an edition is going to be of any use to the user.

Children learning mathematics at school are required to 'show their workings'; they should not, in other words, simply produce a result but also show the process by which this result was arrived at. This seems to me to be something which should also be required of editors. It should be made clear any time there is any form of interpretation. And by interpretation I mean not just corrections or emendations to the text, but also relatively straightforward things such as the expansion of abbreviations, which editors, if they mention it at all, often claim to have done 'as a service to the reader'; as with other services, one should ideally be aware when one is being done one, and also have the option of declining.

The most interesting thing about any manuscript — indeed any text-bearing object — is that it exists. But there are two temporal axes; there is the moment in which a manuscript comes into existence, and then there is the time during which it continues to exists, and so one must ask oneself, what factors led to the production of this manuscript, why does it have the particular form it has, why and how has it come down to us, and what has happened to it along the way?

My paper will look at the ways in which these various aspects of a text's physicality might be presented in the context of an edition.

Notum Sit Omnibus: The new scenario of ICT

Charles J. Farrugia, Angelo Dalli
National Archives of Malta, Santo Spiriuto, Malta

Key words: *Digitization, Linguamine, National Archives of Malta, Intellectual Property Rights, Lund Action Plan, Digital Paper Archive (DPA)*

Malta's cultural and scientific knowledge resources are a precious public asset forming the evolving, collective memory of society. A synergy between archives of historical content and information technologies provides a solid basis for the development of digital content industries in a sustainable knowledge society.

Linguamine performs R&D into next generation search technology that has particular applications to the preservation and access of national heritage. Linguamine has developed a specialised system for the annotation and processing of historical documents, built on top of an existing and proven Digital Paper Archive (DPA) technology platform. This software is currently being tested in a joint project with the National Archives of Malta.

Our project offers a test bed for various innovations in digital image restoration, capture, annotation, enhancement and retrieval. Specialised image processing techniques for historical documents were developed, including techniques that automatically identify inked and pencilled areas across a variety of media (like paper and parchment), while handling unique features such as acidic seep-through. Sentences are automatically identified and segmented from the document image. Once sentence lines have been identified, word boundaries are determined through OCR techniques. Unique problems arise in old handwritten documents since space marks may not always be easy to discern automatically, unlike printed material. Average word length and distribution information for the document language are used to make intelligent guesses in making a final determination of each word's boundaries.

Rule-based systems are finally used to delineate words, sentences and paragraphs, thus attempting to recreate the structure of the original historical document when it was created. The start and end points of different sentences together with paragraphs can be problematic to determine

automatically in old documents, where different writing conventions were often applied (such as having different coloured ink marks or symbols to denote a paragraph.) This important first phase takes place solely from scanned image input, without needing any manual intervention.

Researchers can use a custom-built Historical Document Annotation Interface, that allows researchers to work on a digital replica of the original historical document, and permits semi-automated text transcription to take place directly from the digitised images, cutting down on manual labour and reducing errors while producing a faithful, quality record of the originals.

Various features existing in the Linguamine Digital Paper Archive (DPA) are integrated into the Historical Document Annotation Interface. This allows researchers to utilise the DPA sophisticated search and indexing functions together with web publishing features and the Digital Replica reader. The DPA also allows selected material to be made available automatically on a website without the need for any additional programming or development.

Intelligent character recognition techniques consisting of pattern learning and recognition algorithms can also be applied to extract common keywords, names, dates and abbreviations from the document. Manual training is necessary to train the system to understand the various styles of handwriting and old document printing styles and type faces. Abbreviations and common phrases are handled automatically using Natural Language Processing techniques, and can be expanded to their full form.

The Linguamine DPA offers basic inventory management functionality that has been augmented to cater for the specific needs of archives, namely with support for different international standards (such as RDF, OIA, EAD, CLD and Z39.50) and support for inventory data capture at source using a simple web-based interface.

The software's features and functionality were also assessed in light of the European Union's recommendations given during the Lisbon summit of 2000 and the Lund action plan of 2001 (eEurope, 2001). The Lund action plan makes a series of recommendations for actions that support coordination and add value to digitisation activities in ways that would be sustainable over time. Our joint project proposal fully supports the recommendations in the Lund action plan, positioning itself as an initial pilot project that can be extended, in a slightly modified and enlarged form, as a pan-European project on a similar theme. In particular our project aims to achieve the following eight Lund action plan objectives:

1. Accessible and sustainable heritage. Malta has a wealth of cultural heritage amassed during its long and varied history. Making these resources accessible to the public through the technology developed for this project will be vital for preserving this cultural heritage, while providing significant scientific benefits from the new technologies and R&D products arising from this proposal.
2. Support for cultural diversity, education and content. The Digital Paper Archive, forming the main deliverable of this project, will serve as a key resource for education and researchers.
3. Reducing approach fragmentation, increasing interoperability, decreasing obsolescence. Support for international standards will ensure that the resources produced by this project will be accessible and re-usable by other projects. The use of XML based data storage will also reduce the risk of obsolescence.
4. Lack of simple, common forms of access for the citizen. Many people also find that currently it is very difficult to gain access to most historical documents, and if access is gained, it is quite difficult for a non-expert to be able to adequately read the content of these documents. Both these problems will be addressed and solved by this proposal.
5. Intellectual Property Rights. Linguamine and the National Archives of Malta have a clear agreement on IPR which allows both researchers and the general public alike to enjoy the fruits of our proposal without unnecessary limitations.
6. Lack of synergies between cultural and new technologies programmes. The Lund action plan identifies the need for improved linkages between cultural and new technologies programmes at national and EU level in order to identify priorities and where there is European added value to be gained. This proposal will create scientific and cultural resources at a local level in a collaborative process, while opening an avenue for future pan-European collaboration and re-use of technology by other member states.
7. Take-up of best practices in member states. Feedback from 12 EU countries that maintain competence centres for the digitisation of historical documents – namely Austria, Belgium, Denmark, Estonia, Finland, France, Ireland, Italy, Netherlands, Slovenia, Sweden and the UK – was solicited to ensure that this proposal adheres to best practices while adding new value in the form of new technology and scientific research.

8. Making resources visible and accessible. Inventory management and compliance with standards has been specifically proposed to meet this Lund action plan objective, ensuring unified and affordable access for citizens while providing long-term availability of resources.
9. Advance the dissemination of good practice. The experience gained on this project will be disseminated to other member states through a report that will be distributed by the National Archives of Malta throughout the Minerva and other EU networks, while Linguamine will disseminate information about its R&D through ELSNET and other suitable venues.
10. Improve quality/usability of content, promote unified access for citizens and increase awareness of long-term preservation issues. This proposal addresses all three Lund objectives by:
    - Providing sophisticated search and indexing facilities for historical content that is currently almost unsearchable
    - Providing support for researchers to annotate documents with less effort and fewer errors
    - Making the resources of this proposal easily accessible over the Internet
    - Using digital document replicas to minimise damage to the originals

Game in History, Historical in Game

Aleksandra Fostikov
Institute of History, Belgrade

The text is concerned with relation between game and history, in three various aspects: the idea of game in history from the beginnings till today, the influence of game on human development and objectivity of historical fact in computer games.

Keywords: *Game, history, historical facts, education, computer games.*

This text deals with relation between game and history. It consists of three different parts. First section examines the notion of game in history, the second deals with the importance of game in its contemporary meaning and its influence on the development of one's artistic abilities, knowledge and intellect. The last one deals with historical elements in a game. At the end there are conclusion, abstract and key words.

The game in history

The game, according to some historians, such as *Johann Heuzinga*, represents *the quality of action, different from ordinary life*. Behind this idea hides very wide spectar of human activities. Thus, the Latin word *ludus*, or Old Slavonic word *uɛpý*, have a large number of meanings – from dance to gambling. For example, the Latin verb *ludere* in the first place means *getting out on stage.* Later it received other meanings, like joke and fun. Therefore, noun *ludius* means an actor, or a comedian. So, these words did not necessarily mean only child games, but their usage also implied public character in sense of competitive games. In these games participants showed their skills, necessary for everyday life, like fisical strength, endurance, their effectiveness in handling various weapons (in anthic olimpic games or chivalry duels in feudalism), etc. No matter what kind of game it was, was it only acting or competition, they all included one very important element. It was simulation of reality. Virtual reality.

They also implied free, festivity time for their performance, performing place, playground or theater scene. In this way the game was

closely connected with observing, watching. In the same time there was another latin term for the game – *iocus* (gl. iocari), which means a joke, fun, but also a toy. Therefore, this term was more used for the entertainer in France (jongler), and similar in German (spiel, spielmann). Different developments of connection between terms *ludus* and *iocus*, as well as the rise of new games contibuted to the creation of new categories, such as sport. The noun game has recently earned a new meaning, that of a computer game.

The influence of games on fisical and intellectual development

It is not necessary here to discuss in detail about the significance of games for Homo sapiens as social beings. It is widely known that its influence on humans is very large, especially in attaining knowledge and everyday life experience. This virtual reality could be seen in some child games, while the others are dedicated to the skill development. Thus a girl, imitating her mother, makes cookies for dolls. Double Dutch, no meter how fun and reckless girls look while playing it, actually leads them towards the better coordination. Alexander S. Neill, the author of the book, *Summerhill: A Radical Approach to Child Rearing (*1984*),* dedicated his work to the notion of game as a learning process. In present, a computer game gives us wide opportunities for learning. Computers have become the new interactive toys, while the Internet symbolizes the new playground. For example, *Counter strike* became very big action hit, not only among young ones. It represents the sort of games for which coordination of moves and speeds of reflexes are essential. The older games like *Super Mario* had these characteristics, too. Today, the application of computer games, as exercises for the military purposes, is common practice. Some simplest games, like crosswords, are essentially important, not only for enrichment of native, but also of foreign lexical fond. This concept has already contributed to the formation of numerous similar games for learning the foreign language. Here, we should mention the games which solution also demands the comprehension of combinatory or the mathematic elements. The development in these directions brought to life a special kind of games entitled *Educational games*.

The historical elements and facts in a computer game

During their expansion, computer games were separated by divisions, according to various criteria and depending on players' activities (by Wikipedia, three main groups are major genres, notable genres and superseded genres). We are going to emphases only some notable

subgroups, for analyses of historical elements in them. In relation of game and history, two crucial questions are to be asked. The first one is related to the informative and educational aspects of computer games, and the second one is related to the validity of the historical facts presented in the games.

Certain types of game such as *Simulations* and *Strategies* could be instructive for understanding historical processes. Therefore, while playing the *Tycoon* games (*Railroad*, *Transport*, etc.) we could gain knowledge of transport systems (e.g. railways), which played important role in the history of communications. Famous strategic simulations like *Sid Meier's Civilization* and *Empire Earth* emphasize inevitable correlation between politics, wars, economy, communications, science and progress in history. Negligence of the resource collecting (storage of grain, gold, stones, wood etc.) and their spending exclusively for military purposes brings the empire on the verge of decline and *vice versa*. Some historical events like wars and battles often make basis for shooting games (*First-person, Third-person*), e.g. *Medal of Honor*, *Call of Duty* and others. Historical facts also provide extensive fields of imagination for the creation of *point-and-click* genre of adventures. The solutions for these puzzle-kind of games are closely connected with our historical knowledge. Thus, in an episode of the *Indiana Jones* serial situated in Hitler's time it is necessary to steal *Main Kampf*, while in the game *Loch Ness* it is necessary to open safe-box using the code composed of numbers which represents the year of Scottish-French "Auld Alliance" concluded in the 1295/6 vs. England. The serial *Nancy Drew* is full of historical puzzles. It is designed not only for the young, but also for the adult, depending on game play level. Consequently, in one of these adventures, a player has an opportunity to learn Roman numbers, and in Nancy's notebook, one can find the instruction how to recalculate on the easier level.

The question "Is the historical fact trustworthy or not?" is a special issue. Today there are many critics of computer games. They are accused of being aggressive, bloody, political coloured, and that even a kid by playing a game can learn how to kill a cop. However, they are subjected to historical criticism. If a historian takes a historical fact in its widest sense, whether it is a word, a picture, or something else, he often can't escape the impression that the truth is maybe changed. The example for the apparent mistake can be found in the above mentioned game *Loch Ness*, where on the safe-box stands the inscription *Anno Dominici* instead of *Anno Domini*. The problem in itself is the existence of the game's possibility not only to educate but also to mislead a player. If in the game which take place in antic period of European history a player meets the maize or the usage of paper, especially

younger players might think that these crops or goods then existed. The other problem, which is not of our concern now, is graphical presentation, because the information, given through picture and animation, could easily be misrepresented. At the end, something of the contemporary criticisms is going to be mentioned. Dominating trend of using modern wars as games' topics could easily bring to life very bad influence on the contemporary notion and knowledge of historical events, like any other media.

Conclusion

For everything above mentioned and for objectivity of historical facts it would be good that the games be accompanied with add-on materials like abstracts of the historical period. For example, *Shogun: Total War* introduce us to the history of Japan in XVI and XVII century. It is important because in the games beginning with real historical basis and which incorporate true timeline player can, by his own will and through choices made during gameplay, change its course. In this way, games can have historical educational character.

Preserving and Exploiting the Dodecanese Traditional Songs

Katerina T. Frantzi
University of the Aegean, Rhodes, Greece

Keywords: *Traditional Songs, Corpus Linguistics, Storage and Exploitation*

Culture gives identity and diversity in today's world, preserving and protecting human values. In this work we use technology for the preservation and exploitation of traditional songs. Our application area covers the traditional songs from the islands of Dodecanese, Greece. These songs initially given from grandparents to children and grandchildren orally, later in written form, keep a lot of a society's culture. They reveal peoples' everyday concerns, relations, hopes, beliefs and values.

We create a database and a corpus of traditional songs from Dodecanese that we process using corpus linguistic techniques. Corpus Linguistics is a relatively new area of Linguistics that uses corpora in electronic form, together with techniques (statistical and linguistic), to process language in an accurate and quick way that could not be achieved manually (Biber, et al., 1998; Kennedy, 1998; Sinclair, 1991). Since it is computer-based, it can give quick and accurate results regarding the language characteristics of the corpus. Recently, corpus linguistics finds new areas of application, beyond the borders of pure linguistics. Easy and quick access to information, precise and objective information extraction, are all consequences to the information obtained by corpus linguistic techniques (McEnery & Wilson, 1996; Ooi, 1998). In our application, we reveal both linguistic and cultural characteristics from the traditional songs. Though some of those characteristics could be recognized when simply reading the songs, we use corpus linguistics techniques for their formal and precise exploitation. These techniques can make possible and a lot quicker the accurate and complete study of such work that could not be achieved if done manually. We created our tools (instead of using existing ones), since this way they are a lot more flexible and a lot easier to update (not to mention that most of the ready-to-use programs are not for free, and any kind of alternations would certainly cost extra).

The Greek traditional songs, as an important form of art to reveal cultural elements, have attracted the researchers' interest. A traditional song, combining music and poetry, expresses people's everyday world and feelings (joy, happiness, hope, disappointment, pain, etc.). The difference between "personal" and "traditional" poetry is that the first comes from the poet himself/herself, expressing his/her personal ideas, feelings, life, while the second lies beyond the poet, becoming the expression of those that share the same place and time (Σπυριδάκης, 1975).

The root of "δημοτικά τραγούδια" ("dimotika tragoudia", "traditional songs") comes from the ancient Greek, where "δημοτικά" ("dimotika") derives from the word "δήμο" ("dimo"), meaning "the people of the area", and "τραγούδια" ("tragoudia") from the word "τραγωδία" ("tragodia"), meaning "tragedy". In its common use, "tragedy" covered songs used in celebrations. A smaller sized "tragedy" led to "τραγώδιον" ("tragodion"). It is very hard to find the exact place of the first appearance of a traditional song. Since the songs were taught from generation to generation orally, their original forms were often lost. The reason is that they easily adapt to other areas, getting elements of other cultures, and as such they become part of these cultures. The traditional songs are created by "common people" and women contribute to a great degree to their creation and preservation (Κυριακίδης, 1978). The song used to be more of a "group work", while today their creators are mostly professionals. Strictly speaking, we can say that there is not any more creation of traditional songs (Μπαζιανάς, 1997).

Usually traditional songs talk about nature, family, life and death, love, religion (Παβολίνι, 1981), and they use simple and clear language. Feelings are expressed in the most direct, honest, and full of life way (Διαμαντή, 1981; Σπαταλάς, 1981). The language of the traditional songs shows similarities to that of the Homer Epics preserving a lot of the archaic elements (Μαστροδημήτρης, 2001). There has been a great interest in collecting traditional songs (Πετρόπουλος, 1958; Κουγέας, 1981; Καιροφύλα, 1981). Since it is hard to find the original form of a traditional song and since new elements are added when a song travels to other places and in time, their collection and categorization is not an easy task. Furthermore, there were researchers who, in order to "save" the traditional songs, thought they should "correct" them in various ways. So, some of them changed their language into the formal Greek of the time (Πολίτης, 1925; Πετρόπουλος, 1958; Μελάς, 1981). The traditional songs can be categorized according to their content. The main categories are the following (Κουφός, 1970; Πολίτης, 1978):

Akritic songs: Among the first created. They talk about the life and the fights of Akrites (those protecting the east borders of the Byzantine from the Arabs).

Historical songs: About historical events as well as the life of heroes.

Kleftika: Created when Greece was part of the Ottoman Empire. They expressed the hope of the Greeks for their freedom.

Love songs: They often describe the characteristics of the person the song was created for.

Ksenitias: A very common category. People express their sadness for those who had to leave their place in order to find a job. They are usually words of the mother, the wife and others.

Lament songs: Songs of sorrow for people who died. Many of them created at the time of sorrow.

Wedding songs: Used on weddings and talk about the bride, the groom, their parents and the rest of the relatives. They generally wish the couple to be happy in their new life.

Children songs: Stories to children that usually teach them things. Lullaby songs can be considered of this category.

Celebration/Religion songs: Used in celebrations. Religion songs can be considered of this category.

Paraloges: Generally narrate stories and often contain dramatic elements.

Maxim songs: Express the thoughts of people about life.

Work songs: Sung by workers during their work, making it more relaxing and pleasant.

Satire songs: Specific people as well as specific ideas are satirized in a funny way. They sometimes talk about similarities among the characteristics of the humans and those of animals.

Rimes songs: Common to the islands, and especially Cyprus and Crete. There is usually no music, while emphasis is often given to the personal character of the creator and to the truth of his/her sayings.

Bibliography

Biber D, Conrad S, Reppen R. *Corpus Linguistics - Investigating Language Structure and Use*. Cambridge University Press, Cambridge, U.K, 1998.

Kennedy, G. *An Introduction to Corpus Linguistics*. Addison-Wesley, 1998.

McEnery, T, Wilson, A. *Corpus Linguistics*. Edinburgh University Press, Edinburgh, 1996.

Ooi, V. *Computer Corpus Lexicography*. Edinburgh University Press, Edinburgh, 1998.

Sinclair, J. *Corpus, Concordance, Collocation*, Sinclair, J. and Carter, R. eds.. Oxford University Press, Oxford, England, 1991.

Διαμαντή, Φ. "Η φυσιολατρεία στο δημοτικό μας τραγούδι". Στο *Δημοτικό τραγούδι*, εκδ. ΣΤ΄, εκδ. Δ. Ν. Παπαδήμα. περιοδ. Ελληνική Δημιουργία, Σπύρος Μελάς, τεύχ. 48 τόμ. 5ος, 1η Φεβρουαρίου 1950 Αθήναι (Α΄ αφιέρωμα), Αθήνα, σ. 189-192, 1981.

Καιροφύλα, Κ. "Πώς ο Fauriel μάζεψε τα δημοτικά τραγούδια". Στο *Δημοτικό τραγούδι*, εκδ. ΣΤ΄, εκδ. Δ. Ν. Παπαδήμα. περιοδ. Ελληνική Δημιουργία, Σπύρος Μελάς, τεύχ. 48 τόμ. 5ος, 1η Φεβρουαρίου 1950 Αθήναι (Α΄ αφιέρωμα), Αθήνα, σ. 217-222, 1981.

Κουγέας, Σ. "Η παλαιότατη συλλογή ελληνικών τραγουδιών του Γερμανού βαρώνου Werner von Haxthausen". Στο *Δημοτικό τραγούδι*, εκδ. ΣΤ΄, εκδ. Δ. Ν. Παπαδήμα. περιοδ. Ελληνική Δημιουργία, Σπύρος Μελάς, τεύχ. 50 τόμ. 5ος, 1η Μαρτίου 1950 Αθήναι (Β΄ αφιέρωμα), Αθήνα, σ. 337-346, 1981.

Κουφός, Ι Ν. *Τα Ελληνικά Δημοτικά Τραγούδια*. Επιλογή και Σημειώσεις Ι. Ν. Κουφός, εκδοτικό γραφείον Σείριος, Αθήνα, 1970.

Κυριακίδης, Σ Π. *Το Δημοτικό Τραγούδι - Συναγωγή Μελετών*. Εκδοτική φροντίδα Άλκης Κυριακίδου-Νέστορος, «Ερμής», Αθήνα, 1978.

Μαστροδημήτρης, Π Δ. *Η ποίηση του Νέου Ελληνισμού, Ανθολογία Εισαγωγή - Υπόμνημα - Γλωσσάριο*, εκδ. ΣΤ΄, Αθήνα, 2001.

Μελάς, Σ. "Τρεις Γνώμες για το δημοτικό τραγούδι". Στο *Δημοτικό τραγούδι*, εκδ. ΣΤ΄, εκδ. Δ. Ν. Παπαδήμα. περιοδ. Ελληνική Δημιουργία, Σπύρος Μελάς, τεύχ. 50 τόμ. 5ος, 1η Μαρτίου 1950 Αθήναι (Β΄ αφιέρωμα), Αθήνα, σ. 379-380, 1981.

Μπαζιανάς, Ν. *Για τη Λαϊκή μας Παράδοση Μικρά Μελετήματα*. Εκδόσεις τυπωθήτω, Αθήνα, 1997.

Παβολίνι, Π Α. "Σκέψεις περί της μελέτης της Ελληνικής Δημοτικής Ποιήσεως". Στο *Δημοτικό τραγούδι*, εκδ. ΣΤ΄, εκδ. Δ. Ν. Παπαδήμα. περιοδ. Ελληνική Δημιουργία, Σπύρος Μελάς, τεύχ. 50 τόμ. 5ος, 1η Μαρτίου 1950, Αθήναι (Β΄ αφιέρωμα), Αθήνα, σ. 337-346, 1981.

Πετρόπουλος, Δ. *Ελληνικά Δημοτικά Τραγούδια Εκλογή κειμένων Σχόλια και Εισαγωγή*, τομ. Α΄ - Β΄ Βασική βιβλιοθήκη αριθ. 46 και 47, εκδ. οικ. Ι. Ν. Ζαχαροπούλου, Αθήνα, 1958-1959.

Πολίτης, Ν. Γ. *Εκλογαί από τα Τραγούδια του Ελληνικού Λαού*, 7η έκδοση. Εκδόσεις Ε. Γ. Βογιονάκη, Αθήναι, 1978.

Σπαταλάς, Γ. "Συνθηματικές λέξεις στα δημοτικά μας τραγούδια". Στο *Δημοτικό τραγούδι*, εκδ. ΣΤ΄, εκδ. Δ. Ν. Παπαδήμα. περιοδ. Ελληνική Δημιουργία, Σπύρος Μελάς, τεύχ. 50 τόμ. 5ος, 1η Μαρτίου 1950 Αθήναι (Β΄ αφιέρωμα), Αθήνα, σ. 364-368, 1981.

Σπυριδάκης, Γ Κ. *Ελληνική Λαογραφία (Λαϊκός Πολιτισμός των Νεότερων Ελλήνων),* τεύχ. Δ΄ Δημοτική Ποίησις, εν Αθήναις, 1975.

Digital Archiving of Dancing

Kozaburo Hachimura
Ritsumeikan University, Japan

Recording and preserving a variety of cultural properties with digital technologies, or Digital Archiving, has been attracting considerable attention recently. Planer materials like paintings, photographs, as well as historical documents, are the most fundamental targets of digital archives. Solid materials such as archaeological artifacts, sculptures and other art crafts are also subjected to digitizing with solid shape measuring techniques. Furthermore, intangible cultural properties that include human body motion in dance, entertainment, and hand crafts, have come to be digitized.

At the Art Research Center of Ritsumeikan University, we are undertaking research on archiving materials about dance and performing arts, etc., with the intention of preservation and analysis of these intangible cultural properties. Prerecorded cine and video images about traditional dance are carefully stored and preserved as a film archive in the center. Also, some of the traditional dances have been recorded at the center in a multi-view format by using several video cameras. Furthermore, we have installed an optical-type motion capture system for the purpose of recording precise three-dimensional information about body motion in dance. Some tens of small ball-shaped markers are attached to the appropriate parts of a subject, and are imaged with several high precision video cameras. The exact three-dimensional positions of the markers can be obtained by computer image processing algorithm. We have been investigating the possibility of using this system for digitally archiving of various kinds of dance; Noh plays, traditional Japanese dance, contemporary dance, ballet, and some ethnic dances.

In this talk, topics on digital archiving of dancing as intangible cultural properties by using the motion capturing technique and its application will be presented. Topics include the quantitative analysis of body motion, use of the dance notation called Labanotation, reproduction of dance with CG, as well as the production of multimedia teaching materials of dance. These topics will be briefly described in the following.

The first topic is about our activities on the use of Labanotation methodology for describing motions during dance. Labanotation is a dance score that has been developed and used mainly in the Western dance community, but it is also becoming popular even in Japan. We developed a

graphical editor system called LabanEditor, which can be used for inputting and editing of Labanotation scores interactively. The data for a score can be stored and distributed in the XML format that we designed. Since three dimensional character animation can be generated from the notation, we can verify the movement described by the score easily. Being written in the Java programming language, the system can be run on various PC platforms. Automatic generation of Labanotation from motion capture data is a very challenging and interesting subject. Simple motions can be handled so far, but it proved to be a tough task to apply it to dance motion in general.

When continuing motion capturing of dancing motions, we will face the problem of how to manage vast amounts of motion data. We are able to use metadata-based searching of motion data, but it is not easy to search dance motions that are similar to the motion at hand. For this purpose we are investigating a method of similarity retrieval of motion data. The method employs a DP matching method for judging the similarity between motions that have in general different timing with each other. When retrieving, we can even specify body parts of interest, if we want to retrieve dance motions having similar motions in those parts, e.g. legs.

If a long sequence of dance motion data could be expressed by some of the representative snapshots, these snapshot pictures could be used as an index or an abstract of the motion data. Also, a cut of highlight motion extracted from the original motion will serve as a compact representation, i.e. an abstract, of the original motion. To achieve this end we are using Laban Movement Analysis (LMA) theory for obtaining the quality of motion from motion data, and eventually extracting characteristic poses and highlight motions from motion data.

Quantitative analysis of body motion during dance, for instance, comparing proficiency between an expert and a novice and extracting quantitatively the differences in motions in different categories of dance is a next topic. Currently we are focusing on traditional Japanese dance. Traditional Japanese dance is usually trained by face to face coaching, or by following the master's motion, so it was thought that there was no systematic training methodology. We focused on fundamental motions composing a dance performance, and we measured some of the feature values from the body motion data, and tried to distinguish the difference between types of these motions and to also recognize dancers based on the extracted feature values. The study is still in its infancy, but we believe this approach can be used to quantitatively judge the similarity and/or difference between several kinds of dance motions.

The next topic is the KANSEI analysis of dance movement. KANSEI is a Japanese word expressing the concept related to “sensitivity” or “feeling”. We will obtain some sensitivity information when observing dance motion, for instance happiness, vividness, loneliness etc. We are now trying to extract these types of information from motion-captured body motion. The first step is to investigate the relationship between physical properties of motion data and impressions we get when seeing the motion. Psychological experiments and multivariable analysis have been used for this purpose.

The use of motion-captured data for making CG contents of traditional dance is a kind of by-product of our research, since making CG animation is not a research topic in information science nowadays -- everybody can do this by using current off-the-shelf CG modeling and animation tools. However, we do this to prepare the environment for making the presentation of our research results and also for making some kinds of experiments for using archived motion data for experimental performance. For instance, CG animation generated from motion-captured data can be produced in a virtual reality environment. A dance collaboration among CG characters and real human dancers is a very interesting topic. For this we have to prepare the environment, i.e. a stage or theater, for the presentation. From this view point, the Japan’s oldest historical Noh dance theater was restored by CG. We are thinking of making a dance collaboration on this virtual theater connecting two or three remote sites via a high speed network. We will use an immersive virtual environment installed in our university that is capable of stereoscopic display.

We are also making multimedia contents for educational purposes. Some of the traditional dances in Japan have a long history, and it is not necessarily easy to understand them for today’s people. Especially, the Noh play is very abstract in nature, and since the story of it was taken from classical novels, the narration of the players and songs are very difficult to comprehend. The situation may be similar to classical theater plays like Shakespeare’s in the Western culture. Considering these facts, we are now making multimedia contents of Noh plays by using SMIL language and also an authoring tool for making these kinds of contents. The system is able to handle video, texts and pictures synchronously. We can add some annotations to this information.

We are, thus, using the motion capture system for digital archiving of dancing. However, we must honestly admit that there are some problems. For instance, a dancer is compelled to perform under unusual circumstances, in the sense that the dancer wears special body suits in order to make

measurement of actual body motion. This may cause a negative influence on the performance. Also, although the costume and makeup are important for the record of dance, we cannot record them with current motion capture systems alone. Facial expression and gaze direction are also important but difficult to measure. While motions of hands and fingers can be measured by attaching small markers on the joints of fingers, it is a difficult task to simultaneously measure them together with whole body motion.

Digitally archiving with a motion capture system still has these kinds of problems to be solved. However, by this system we can at least do the quantitative measurement and analysis of body motion of dance. It is also possible to explore the essence of the performing arts by combining quantitative analysis with the qualitative analysis made so far.

Several dance performers of purely traditional Japanese dance, including Noh-play and Kabuki-dance, were asked their impressions about measuring and recoding their own physical body motion by the system, and eventually we received comments that this kind of objective measurement and representation with CG are of value for the analysis of their performance, and that they want to use the system for objectively observing the change of their performance with their aging, for instance. We think that this fact suggests the value of digital archiving of dancing by the motion capture systems.

Digital Archives (e-archives)

Risto Hristov
State Archiv of Macedonia, Skopje, Macedonia

Introduction

The introduction of the global information infrastructure caused globalisation of the information systems. The communication through computers is daily and the amount of information that we have access is bigger. The global development of the interactive contents with application of the new media caused competitive atmosphere where the institutions have to accommodate to the new conditions in order to maintain their reputation. The continuous development of the hardware and software make them implement new conceptual decisions which sometimes are possible without major changes and configuration of the entire information system.
The archives as institutions of social importance are in the focus of changes because they must shape at least a part of their records according to the global interactive communication needs, and at the same time to orientate partly towards the market. They have to offer their contents and services on the Internet most frequently offering only part of the records free, and to gain access to the rest of the records the user has to pay. In this way the archives compensate part of their expenses for digitalization.

The State Archives of the Republic of Macedonia finding itself on the information crossroad can't escape the new trends.

The basic task of The State Archives of the Republic of Macedonia is collecting, storing, protection and preservation of documents not withstanding their form and the medium upon which they have been stored because they are carriers of information concerning the past and the present on the Macedonian ground.

Digitalization of the Archive Records

The most frequently documents are in written form (printed: books, papers, journals; records: manuscripts, reports, minutes, record books, codices; visual records: pictures, photographs, drawings, maps and so on). Some of them have originated a few centuries ago and are of great importance not only for the past of Macedonia, but also for the region. Their damaging or losing is a loss not only for our people but also for the mankind in general which could not be compensated because the documents are the most

frequently in a unique sample. Consequently, their preservation is of primary importance.

Possible kinds of preservation:

- restoration;
- microfilming;
- Photocopying of records in format bigger than A4 is not absolutely allowed;
- digitalization.

E-archives

One of the fundamental tasks of The State Archives of the Republic of Macedonia beside the collection and preservation of archives records is the data transparency in front of the home and world public(better and quicker access to the records).The most efficient way of achieving this function is using the possibilities of the world Internet creating WEB portal called “e-archives”.

A Portal, according to the dictionary of loan words, means "a main doorway or entrance into some construction of imposing dimension like a castle or a palace".

In analogy with the explanation of the term a portal, the term a WEB Portal or an Internet Portal associates with an entry into an Internet and a WEB construction. This construction is a multitude of contents and services. Taking into consideration all above-said, we can establish a definition:

A Web Portal is an independent entry or a part of a Web-site including a combination of information and services. Its purpose is to obtain an individual start as an easy guide into the WEB, offering, at the same time, services of the type of e-mail, forums, browsers, chat rooms, news, purchases etc.

This WEB Portal includes information of the Digital archives of The State Archives of the Republic of Macedonia.

The WEB Portal of the State Archives of the Republic of Macedonia is a dynamic WEB Site because its contents are living material which is being permanently changed and updated. It includes the following modules:

1. Presentation of the State Archives of the Republic of Macedonia

This module offers information about The Archives' history, meaning, function, activity. The contents would include a text, images, video and audio records.

2. Digital archives (e-archives)

Besides the concern of the archive records' preservation, The State Archives of the Republic of Macedonia are obliged to provide maximal transparency of the documents to the users, that is to say to provide quick access, browsing and creating of true copies. It is archived by using so-called electronic or digital archives (e-archives).

The digital archives are a place where the archive records' data are being stored in electronic form. The data on paper documents, microfilms, audio and video records and electronic data are being saved into the Digital archives. The basic data of each document necessary for its recognition and access are being saved into the Digital archives. The Digital archives are an Internet application. The data access and the data browsing is quickly and easily. Using Internet at any time from any place, access to the Digital archives' data could be provided. The data are being protected using security levels. The access to data is being exactly defined for each user. The user gets an allowance to access paying a deposit. Each access to the Digital archives' database and its use lessens the deposit.

Advantages of the Digital archives compared with the ordinary ones:

- It is electronic archive records, and not the ordinary ones that are being researched by the user;
- The access to and the browsing through the electronic archive records is more quickly and simpler;
- The original records are being protected from damage or destruction;
- The research is being performed from any distance;
- The archive records' data are at disposal for all users at any time from any place.

E-archives of the State Archives of the Republic of Macedonia include a base of electronic archive records (electronic documents) and a database concerning the structure (indexing) of these documents (the records include data in accordance with the files' fields *dokumenti.dbf* and *prilozi.dbf*).

E-archives enable:

- Browsing is being performed according to documents, archive collections, fields, periods, authors, headlines and keywords though the archive records into the repositories of The State Archives of the Republic of Macedonia;
- Access and exploring of a certain document on a monitor;
- Downloading of copies of certain documents in printed or electronic form;

- Administering of the base of documents or of their structural database.

3. Services

The users, the services, the using of services in temporal intervals, the deposit for the Portal's use, and the expenses for the use of Portal's services are being notified in the module.

The services of the WEB Portal of the State Archives of the Republic of Macedonia are used until the deposit is being spent.

Conclusion

Archives are preservers of the documents that talk about the past and the present of the ground where they are. They are concerned with the preservation of the documents and their access to all the citizens of the country and the world.

The best way to achieve these two fundamental archive functions is to create e-archive that through digitalisation of the documents preserves them from physical damaging. At the same time using the great possibilities of the Internet it makes the documents accessible to all the users no matter of the place they are and the time they investigate.

The creation of e-archives is an obligation put on to the State Archives of Republic of Macedonia by the world trends in order to outrun the period of transition and to enter in the common European home called the European Union.

System for Storing Data about National Heritage with Advanced Search Techniques

Mile Jovanov, Kindalov Marjan, Mihajlovski Nikolche, Kolevski Goran, Miloshev Pece
Institute of Informatics, Faculty of Sciences,
Skopje, Macedonia

Keywords: *National heritage, digitization, search, database, WEB, server, client*

I. The problem

The culture is a phenomenon that characterizes every person. All of the nations through their existence build a culture they maintain and develop. It makes people feel like a part of a community. But the existence of different cultures divides people into separate groups. To overcome the prejudices, members of one cultural community have to learn characteristics of other cultures. Following the tendency of digital communication as a way of successful information sharing, a need for cultural heritage digitization occurs.

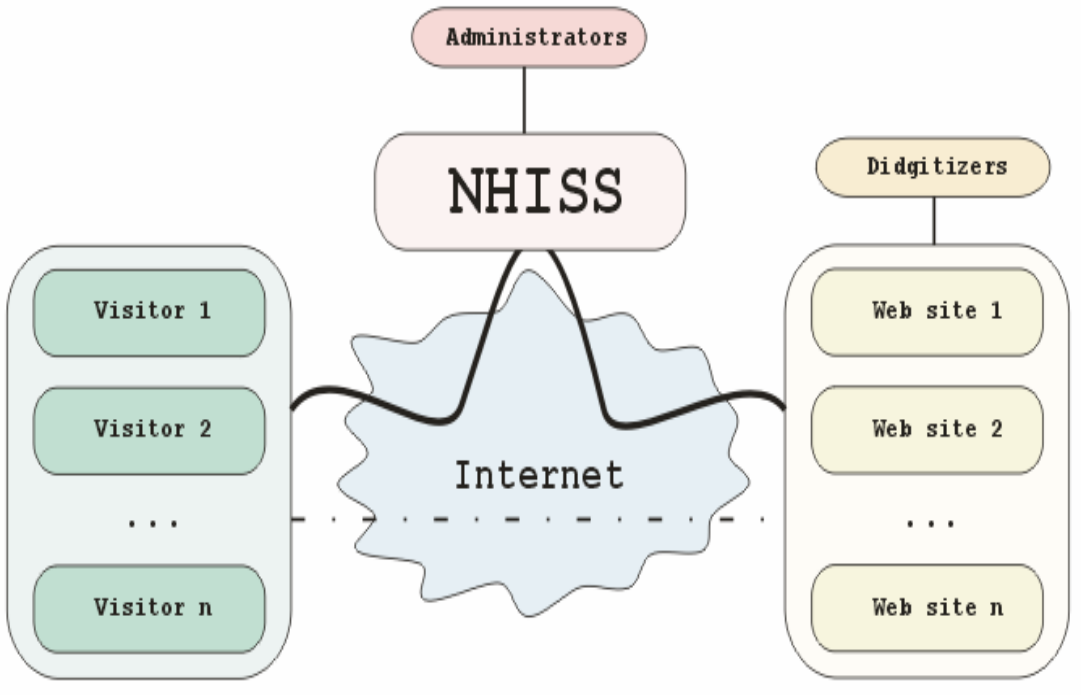


Figure 1. Connection establishmen

But digitization of the cultural heritage solves only a part of the problems. We have to find an effective way of data organization to achieve simple but reliable information access. We need a system with the following characteristics:

- Fast data access,
- Data searching by any of the attributes,
- Simple, but controlled information update,
- Simple information structure upgrade,
- Multi-user data management support.

As we can see on Fig.1 there are different subjects that digitize the cultural heritage. For example, those subjects can be museums, churches, archeologists, collectors, etc. Their exhibits, digitized in some form, are placed on Internet. Global information about the exhibits can be found by any Internet search engine, but some specific information, e.g. churches built in Macedonia, between 1400 and 1540 AD, which occupies area of 200-500 square meters are hard to find.

II. Basic concepts of the solution:

The idea of the NHISS (National Heritage Information Storage System) is to provide the users with answers to the questions of this type. It stores the attributes of the exhibits, so the user can easily search by any of them. Users are categorized into:

- **Digitizers** are the users that import data into NHISS. Information about digitized exhibits they announced on Internet and key attributes are entered into NHISS.
- **Administrators** are responsible for managing NHISS. They add new classes on digitizers' demands and verify the data.
- **Visitors** are the target group of NHISS. They can access it via Internet and search for exhibits. Then NHISS provides them with references to the exhibits they are looking for.

The system consists of four main components:

- **Database**: The main part of the database stores the data entered by digitizers. The second part stores the information needed for database management. The data is structured in classes. Each class has its own attributes which can be inherited by their subclasses. Administrators can add new classes without changing the logic structure of the database.

- **Server application**: This application is interface between the client applications and the database. It receives requests from them and translates those requests into appropriate SQL queries. Then the server translates the SQL answers into messages for the client applications.
- **Client application**: Several instances of this application are connected to the server via Internet or LAN. Administrators can manage the database using the application. By managing we consider:
    - Adding, changing and removing classes of exhibits,
    - Adding, changing and removing administrator accounts[1],
    - Verifying the data entered by digitizers.
- **Web application:** This application is the interface for the visitors and digitizers. Visitors can search for announced exhibits while digitizers can add new information.

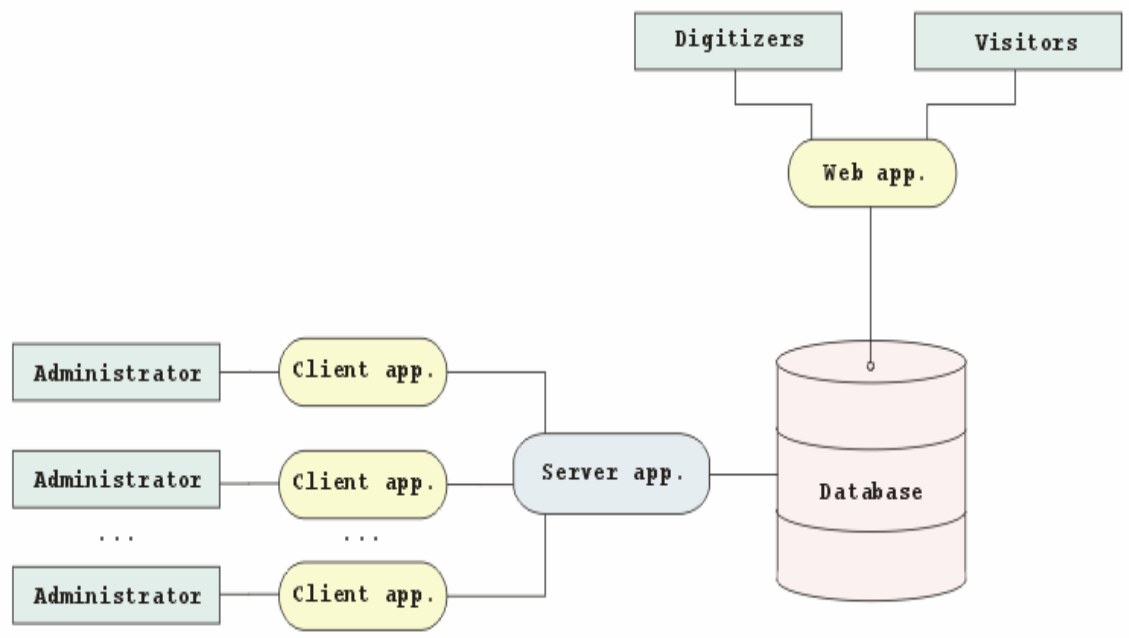


Figure 2. Structure of NHISS

III. Example:

In this section we will give an example of how the whole system works.

Suppose the museum of Bitola wants to announce the information about Heraklea (archeological occurrence in Bitola). They digitize the

1 There are several types of administrators. Some of them can only managed the digitizers data and other can manage system data, e.g. administrator account

information and then they build a website www.heraklea.info.mk. It's up to them to organize their site in the way they think is most appropriate to their research. After building the website they can visit the *web application* of NHISS. With this web application they can enter data into the database, about the artifacts they want to announce. Let consider they want to announce a mask.

Following steps best describes that process:

- **Check if there is a class ARTIFACT and subclass MASK.** The mask is artifact so its class should be a subclass of the ARTIFACT.
- **If ARTIFACT is missing send a request to the administrator for creating this class.**
- **If MASK is missing send a request to the administrator for creating it.** In this request the digitizers should specify the attributes of the class MASK e.g. material, place, shape etc.
- **After all the classes are present the digitizers are filling a form, where they are entering the exact characteristics of the mask.**
- **As final they are entering the link to their website www.heraklea.info.mk**

Next when an administrator will come to work she will logon to the server using the client application. If there are any requests for new classes she can create them or reject them. After the mask is announced she can check if the link is correct and appropriate to the announcement.

Finally some visitor will visit the web application. He can search for artifacts, or specifically masks. The search can be done by any of the attributes of the mask. If he was looking for the mask in Heraklea he will get the link www.heraklea.info.mk where information about the mask can be found.

Wavelets and Word Image Matching[1]

Emil Kelevedjiev, Andrey Andreev
IMI, Bulgarian Academy of Sciences, Sofia, Bulgaria

Keywords: *document text image, bitmap file, word searching, wavelet transform, quantization*

An approach to word image matching (WIM) based on wavelet transform (WT) is examined. A detailed computer experiments was carried out with respect to the type of wavelet filters, types of the wavelet bases used, the best suited frequency band and quantization of wavelet coefficients for WIM. Experimental results for detecting the user specified words in Bulgarian and other typed documents of different quality show good steady results (above 95% right detected words in average).

Introduction

In spite of the fact that newly generated documents are in text format which allows convenient storage and easy remote access to them, the problem for digitizing a huge amount of typewritten (handwritten) scientific, historical and cultural documents of different origin still remains open. The combination of advanced facilities for scanning and conversion such documents to digital images and Optical Character Recognition (OCR) technology which automatically produces from them a computer readable text resolves to a great extent the problem. However it is said in [3]" *It is generally acknowledged that the OCR accuracy requirements for information retrieval are considerably lower than for many document processing applications.*" and motivated by this and some other results in the recent time the goal of this paper is to improve searching and location of a user specified word after OCR or other type of segmentation of the digitized text image. Such problem for word spotting in scanned images is a subject of intensive research both when the text is handwritten or typewritten [1, 2, 3], etc. The wavelet coefficients are used as a basic tool for achieving this goal because taking into account possibilities of:

1. existence of huge amount of different wavelets (filters) that have different support, smoothness, etc.;

[1]This work is prepared under the Marie Curie Action Contract No.: MTKD-CT-2004-509754 – Knowledge Transfer for Digitisation of Cultural and Scientific Heritage in Bulgaria

2. creation of different bases using a fixed filter;

3. quantization of wavelet coefficients in different metrics;

4. using the one of the ”main advantages” of a given wavelet basis to other bases, e.g. the faster diminishing of wavelet coefficients considered by absolute value;

5. lower complexity of wavelet transform with respect to other commonly used transformations, for example Fast Fourier Transform FFT;

6. smoothing properties of wavelets [4] that results in noise removal (noise almost always exists in older documents caused by different sources) using thresholding (quantization) of wavelet coefficients.

All the above listed properties of wavelets are considered more in details and they are tested numerically later in order to find out from practical point of view the best approach for using wavelet coefficients.

Preprocessing

Based on the system and code described in [5] we assume that the binarization and enhancement steps in the sense shown below

source document	$\Longrightarrow$ scanner	digitized image	$\Longrightarrow$ enhancement	improved image
1.books		1.bitmap file	1.Wiener filter	
2.manuscripts		2.text file	2.WT	
3.newspapers		3.word file	3.smoothness	
4.etc.		4.etc.	4.etc.	

are done. Simply, we suppose that the code has done the following:

1. successful word segmentation of the digitized document and a bitmap file of each word;

2. using interactive features of the code the alphabet was built selecting by the help of a scanning technic all small and capital letters in the scanned document. This alphabet will be used later for creation of synthetic image word.

Wavelets and Bitmap image file. Haar, Daubechies and Biorthogonal Wavelets

When representing an image via wavelets their coefficients behave differently with respect to the type of wavelet used and the properties of the image. In order to determine a suitable type of wavelet that reflects the features of the images that are typewritten Cyrillic words obtained by the code given in [5], we carried out series of experiments with 3 different wavelet functions, namely Haar, Daubechies-4 and 10-6 Biorthogonal wavelets. It is important to note that images like that obtained by scanning typewritten texts have a worse smoothness than usual pictures like ”Lenna” used widely by researchers.

If the function *f* is represented by orthogonal wavelet expansion

$$f(x) = \sum_i d_i \cdot \psi_i(x)$$

then the best way to pick *N* coefficients d_{ik} among $\{d_i\}$ making $\| f - f_N^{opt} \|_{L_2}$, where

$$f_N^{opt}(x) = \sum_{i_k=1}^{N} d_{i_k} \cdot \psi_{i_k}(x),$$

as small as possible is simply picking the *N* coefficients with largest absolute value. We determine the smoothness β of the Bitmap image file (as a 2-dimensional function) using the estimation [6], p.111,

$$E_{L_2}^N(f) = \| f - f_N^{opt} \|_{L_2} = \frac{C}{N^\beta} \qquad (1)$$

where the constant *C* in (1) is connected with a Besov norm of a function *f*.

Using Haar wavelets the behavior of a relative error $E_{L_2}^N(f) / \| f \|_{L_2}$ is examined as a function of *N* for two type of images: ”Lenna” type and ”word images”.

The experiments with ”word images” lead to the conclusion that the ”word images” are of lower smoothness and, or what is more important is that such images can be characterized using less wavelet coefficients than must be used for portrait images like ”Lenna”.

Although the ”word images” have lower smoothness, the question which wavelets reflect in ”best way” their essence remains open. As a first

step in this direction, we carry out numerical experiments using three relatively different wavelets - Haar, Daubechies-4 and 10-6 Biorthogonal [7], a computational scheme for Biorthogonal wavelets with symmetric extension, and two types of bases: Regular and Hyperbolic.

Conclusions

1. with fixed number of coefficients the hyperbolic type of wavelets extract more information about the image than regular type of wavelets when applied for word images;
2. Haar hyperbolic wavelet is of the same quality as 6-10 Biorthogonal wavelet if we keep 50-500 wavelet coefficients;
3. taking into account the simplicity and speed of computations using Haar filter it makes sense to use Haar hyperbolic wavelets for word images.

References

[1] Y. Lue, C. L. Tan, W. Huang, L. Fan, ”An Approach to Word Image Matching Based on Weighted Hausdorff Distance”, *”6th ICDAR”* , 10-13 Sept.2001, Seattle, USA.

[2] Y. Lu, C. L. Tan, ”Chinese word searching in imaged documents”, *International Journal of Pattern Recognition and Artificial Intelligence*, Vol.18, no.2, pp.229-246, 2004.

[3] A. Kunde, ”handwritten word recognition using hidden Markov model,” in H. Bunke and P. S. P. Wang (Ed.) *Handbook of character recognition and documents image analysis*, World Scientific, 1977, pp.157-182.

[4] D. L. Donoho, I. M. Johnstone, ”Ideal Denoising in an orthonormal basis chosen from a library of bases”, *C. R. Acad. Sci. Paris*, Ser. *I*, 319(1994), 1317-1322.

[5] T. Konidaris, B. Gatos, K. Nizios, I. Pratikakis, S. Theodoridis, S. J. Perantonis, ”Keyword Guided Word Spotting in Historical Typewritten Documents Using Synthetic Data and User’s feedback”, 2005, (submitted).

[6] SIGGRAPH’95 Course Note, Wavelets and their Applications in Computer Graphics.

[7] A. Cohen, I. Daubechies, J.-C. Feauveau, ”Biorthogonal Bases of Compactly Supported Wavelets”, *Comm. Pure and Applied Math*., 45(1992), 485-560.

The Use of VR panoramas in visualization of cultural heritage on interactive CD ROM "Viminacium lumen meum"

Dr. Miomir Korać[1], Srđan Marković[2], Jelena Obradović[2]
[1]"Arheološki institut SANU", Belgrade,
[2] "Multimediaworx", Belgrade

Keywords: *Virtual Reality, VR panorama, simulation, realism, interactivity, Quick Time*

Virtual simulation of space and objects, as the method of visualization of cultural heritage has substantial significance for accuracy and authenticity of spatial information because it provides "be there" effect – a look at environment or object from every angle.

Important characteristic of photographed VR panoramas lies in their realism and 3D illusion without added 3D information, while modeled and rendered panoramas present the most faithful reconstruction or upgrade of a real ambiance.

Emergence of panoramas, according to Oliver Grau, dates from the antiquity, with frescoes covering the walls of Villa dei Misteri in Pompeii. The pictures, representing figures which participated in Dionysian rituals, offered visitors full 360 degrees vision, thus anticipating dioramas and cycloramas that were to become popular in 19th century. Invention of perspective in renaissance allowed artists to create even more convincing illusion of threedimensional spaces. Baldassare Peruzzi's Salla delle Prospettive in Rome from 16th century transformed the end wall of the room into a colonnaded portico looking over illusionistic perspective of ancient Rome. The idea of painting fully circular canvas in correct perspective was patented by Robert Barker in 1787. and he had it constructed several years later. Grau also notices that appearance of panoramas was characterized by combining media and military history. Early panoramas served in terrain observing and investigating, but their military effectiveness was short lived, and was soon overtaken by their value of propaganda. Panoramic battle scenes, such as the "Battle of Sedan" by Anton von Werner depicting Prussia's victory over the French in 1870, became well-attended spectacles, although produced in an age when manufacturing panoramas was more an industrial than an artistic process.

In the mid 19th century, by joining sets of photographs made from single focal point into cylindrical whole, became dioramas – panoramic rotational images. Industrial modes of creating illusionistic or immersive experiences continued with the invention of film, which became immediate success and huge attraction, and will considerably influence the development of 3D Cinema and Cinerama in the future. Although some of these inventions will be later surpassed, the urge to create fully immersive environments continued with the advent of digital simulation and computer aided interactivity.

Apart from earlier panoramas which were depicting physical world but did not allow any interaction or navigation, computer generated virtual environments permitted this, but were initially limited on what can be represented by 3D computer models. Because of their obvious limitations, such ambiances most usually depicted imaginary or fantasy worlds, and only recently they significantly approached to realism.

By combining artistic and technological methods it is possible to achieve scientifically valid simulation, which would create illusion of realism either by improving environment's characteristics (by postproduction, focusing, lightning correction, detail restoring, retouching…), or by creating precise reconstruction of previous state of the site. Simulation of observing and moving through a real environment, with the addition of zooming onto pictures and objects in maximum quality, together with it's comprehensiveness, becomes priceless documentary material, and in view of it's digital nature, it is possible to be further reused for presentational, educational or scientific purposes.

On the Roman city of Viminacium site, many panoramas of existing situation have been taken, such as: termae, porta praetoria, memoriae. High resolution was required in order to achieve maximum quality and option of full screen viewing. This enabled emphasizing of details, which are additionally enhancing the visual presentation. In case of the panorama of termae, it can be noticed the real nature of photorealistic 360-degree experience, which includes the whole covered site. This enables more obvious look at the real size of the ambiance, composition and orientation of its elements, terrain texture, degree of erosion, details from the excavation site, and even at the exquisite covering construction. By turning in every direction, the illusion of presence is created, and thus is achieved more refined and more extensive approach to the image where its elements are convincingly joined and working together. High resolution allows close

zoom onto specific parts of panorama, which emphasizes significant aspects of location and provides even more information.

In case of non – photorealistic representation of cultural heritage, traditional ways have used mostly illustration and replica sculpting for presentation o f research analysis and hypothesis. Even when computer graphics is available, archaeologists often preferred traditional methods, mostly because of the need for reaching higher exactness or focusing on detail. Sometimes the reason was the intention for obvious confrontation of viewer with the hypothesis, while accurate graphic representation would assume historical truth, which sometimes is not wanted.

However, in the situation when researchers closely collaborate with the artists in the process of visualization, gaining bigger control over data processing and their employ in authentic representation, considerable progress has been made in reconstruction of objects and artifacts, mostly because of careful selection of materials, maps, lightning, complying with plans, sketches, and above all, because of endeavor for creating convincing impression.

Interesting example for this is the virtual reconstruction of the tombs on the site of Viminacium. Frescoes of exceptional archaeological, historical and visual significance had been found in the tombs, and they have been transferred to the Museum in Pozarevac for restoration. Until restoration is finished and frescoes put back on their places, the actual sight of tombs will not be available neither to viewers nor researchers. Therefore, with computer simulation of virtual tombs, there has been achieved their faithful panoramic appearance with modeled structures of the tombs and frescoes mapped onto walls. This approach to representation of an important historic monument is of big importance for documenting actual state of the site, and its contribution lies in making this location accessible and apparent, making very interesting perspective from the centre of the small tomb and from a single nodal point, giving surprisingly broad view with certain sense of bodilessnes.

As research on the Viminacium site continues, it is been planned to develop and upsurge CD presentation as well. Special accent will be put on possibility of making virtual walkthroughs by joining the nodes or points of view, thus gaining another dimension of walking through the site.

On CD ROM “Viminacium lumen meum” we used Quick Time Virtual Reality (QTVR) plug in for viewing virtual panoramas, developed by Apple Computers. It has excellent performance in supporting high resolution panoramas, and Quick Time as “multimedia container” can merge into single package a plenty of diverse material.

There are two classes of QTVR imaging: panoramic VR and object VR. Panoramic image is made when camera is being rotated for precise number of degrees while shooting series of photographs, which will be later joined in complete or incomplete panorama. VR image is being made in similar way, only now the camera is static and object is turning in front of it. Required items for making panoramas are: photo camera (digital or analogue, but analogue photographs should be later accurately scanned) wide angle lens, special rotational head that camera is attached to, and which provides precise turning of camera for chosen number of degrees and under specified angle, and at last, the tripod which can be leveled.

Various software is available nowadays, and most commonly are used: RealViz Stitcher, VR Worx, Helmut Dersch's PanoTools, QuickTime VR Authoring Studio, PhotoVista…

For presenting virtual material, the use of plug in is needed. Depending on desired quality and purpose of panoramas we use: Quick Time VR, Java, VRML, Shockwave, Flash...

The most important thing during the shooting panoramas is finding of nodal point. It is the point where beams, which enter the lens, are crossed. Only by rotating around this point we can achieve wanted accuracy and avoid parallax, which cannot be corrected in postproduction. For professional quality the use of specialized panoramic rotators are essential, because when mounted on tripod they perform exact positioning of camera for rotation by both axes.

While shooting, it is necessary to determine correct number of photographs according to width of the lens, for it is important that they considerably overlap. Recommended overlapping is between 20 and 40 percent. For instance, with the 15mm lens it is enough to take 14 images, whether with 35mm it is necessary to take 50 images. With our 18mm lens we took 24 pictures. Overlapping is important for the program to find enough mutual points on the images, so they can be easily combined.

All pictures taken by wide lens are more or less deformed (so called "Barrel distortion"). For having good quality and seamless merge, this deformation must be corrected. Many programs are used for that purpose (RealViz Sitcher, Helmut Dersch's Pano tools etc).

The use of flash is excluded, because light beams from it, depending on surrounding objects, differently disperse in space, and more or less they are lightening the frame. Such series of images can have big variations in lightning.

There is also a possibility of making hot spots in panoramas. It means that we can create some areas in panoramas which can work as buttons or triggers for some actions, for instance: for loading another panorama (like in virtual walks) or images of magnified object.

With development of new technologies, Virtual Reality in virtual archaeology has obtained a significant place. Visualization of cultural heritage has always made noteworthy difference between research documentation and material suited for presentations or education. With use of photorealistic panoramas this difference has been reduced, hence quality of content in its representation is sacrificed to the minimum. In the "Viminacium lumen meum project" successfully is achieved optimal correlation between artistic interpretation and researcher's practice, keeping the authenticity and relevance of the presented information.

The Role of South East European Prosecutor Advisory Group (SEEPAG) in Harmonizing Cooperation Among SEE And Western Balkan Countries in Combating to Hi – Tech and Organized Crime

Jovan Krstic
Chairman of the SEEPAG, Deputy Prosecutor General of Serbia

Prosecutor representatives of Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Greece, Hungary, FYR of Macedonia, Moldova, Romania, Serbia and Montenegro, Slovenia and Turkey, convened in Beograd, decided at the First Conference on December 11-12, 2003, to constitute the Southeastern European Prosecutor Advisory Group (SEEPAG).The delegates adopted a Declaration of the Prosecutor Representatives of the first meeting of the SEEPAG.

At the last Fourth Conference of the SEEPAG that took place in Belgrade in December 15, 2004, the representatives of the state prosecution offices of the SEE countries put their signature to the main documents of SEEPAG, as stressed by The SEEPAG General Guidelines; the main aim of SEEPAG is to speed-up the judicial cooperation in trans-border crime cases. In this respect SEEPAG Prosecutorial Focal Points were established in each of the above mentioned states. Close cooperation will be ensured with all the relevant regional institution and initiatives, in order to provide necessary assistance to the Task Forces functioning under the SECI Regional Center for Combating Trans-border Crime.

The overall goals for the cooperation of prosecutors with regard to organized crime are to: Enhance the ability of the judiciary to understand the phenomenon of organized crime; enhance the ability of the prosecution to identify, prosecute and judge cases including the effective implementation of national legislation on the basis and in the light of international law and models of best practice; foster the cooperation between the national prosecution services and the police.

Keywords: *Prosecutorial Cooperation, SEEPAG, Organized Crime*

Introduction

The purpose of this article is to elaborate the methodology of extending cooperation across borders of the prosecutors of the SEE region threw the activities of The South East Europe Prosecutor Advisory Group.

It was already recognised that this kind of cooperation is necessary for successful combat trans-border crime during an April 2002 meeting in Ohrid, Macedonia, by the Chief Prosecutors and Deputy Prosecutors from eight of the SECI States and Italy, and high representatives from the Stability Pact, and the Council or Europe and European Commission.

Also, this cooperation is in accordance with the purposes of the SECI Agreement and it is enunciated in the Strategic Plan of the SECI Regional Center for combat to trans/border crime by stating that "SECI *hopes the Agreement lays a foundation for future cooperation among prosecutors in bringing members of criminal organizations to trial."*

SEEPAG is defined as the unique international body with the professional independence and self-managed position, same as it is the position of the prosecutors in the countries of the SEE region in the framework of their national legal systems. We are very well aware of the fact that at the beginning, the success SEEPAG group rests on cooperation with the SECI Center and potentially development of protocols and mechanisms with police and other authorities of the states in the SEE region.

The Declaration

The initiative of Serbia and Montenegro for establishing of this new regional working group of Southeast European countries was accepted and the First Meeting held in Belgrade 11 /12 of December 2003, and the Declaration of the SEEPAG First Meeting. It was a carefully prepared joined project encouraged by all relevant international and regional organizations, and it is also supporting the aim of Serbia and Montenegro to play more important role in combat to organized crime.

As originally anticipated, the SEEPAG seems to be developing along two related lines – as an operational group aiming to improve cooperation in ongoing criminal investigations and prosecutions, and as a legal advisory group seeking to improve and harmonize our criminal codes and procedures by combing our collective knowledge as experienced prosecutors.

The Third meeting was designed to gather the prosecutors from the countries of Southeast Europe nominated to attend and assist in developing the documents for this group. We expect that the members of the SEEPAG are capable to reach the decision on the documents so that they can provide developing and coordinating transnational cooperation and prosecution.

We do remind that at the First Meeting of SEEPAG members said in Declaration that the state prosecutors are legally independent within frameworks of their national judicial systems and with a specific self-managed position in judicial systems of the SEE countries, and that this SEE Prosecutors Working Group must be independent, but we also emphasised that independent does not mean distinct, and SEEPAG also gave the strategic vision shared with relevant regional and international organizations to:

1) implement all conventions, SECI Agreement and other agreements and share information and evidences and make them available to the courts and to obtain the presence of the witnesses threw the related programs of protection in order to achieve more efficient combat to organized crime;
2) ease mutual legal assistance threw: a) SECI Center Focal Points, b) activities of SEEPAG and c) Prosecutorial Focal Points - PFP
3) to propose the protocols for the legal assistance, regional agreement on witness protection and other issues and help the authorities of the countries in the SEE region by releasing legal opinions, advises, recommendation and consulting.
4) cooperate with other international organisations, initiatives and institutions

The Stability Pact as the central to developing the connection to European legal entities and structures, such as EUROJUST and the Council of Europe; and the SECI Center, gave the package of political support to the infrastructure of SEEPAG for operational task forces, and the legal mechanisms under which such cooperation is possible, and offer the framework to prosecutors to cooperate to arrest criminals and gain evidences towards the prosecution of regional organized criminal networks.

General Guidelines and Prosecutorial Focal Points

At the Second Conference in Bucharest, members of the SEEPAG were acting fully as an operational group with concrete cases in which we cooperated to prosecute trans-border criminals. We tried to developed legal and procedural mechanisms for trans-border prosecutor to prosecutor and prosecutor to police cooperation. We were monitoring the special operations against trafficking of human beings and abuse of narcotics, and we are involved in monitoring hi-tech crime and we are facing digital challenges in this field.

For the Fourth Conference of SEEPAG the member states approved the adoption of the basic documents, as they are in accordance to their national laws and admissible in court proceedings, and under the legal authority of the *Agreement to Share Information to Combat Organized Crime (SECI Agreement)* and other relevant international agreements. To do so, at the Third Conference SEEPAG members and experts were discussing the establishment of the Prosecutorial Focal Points in each capital to communicate requests and pass information between prosecutors and with the SECI Center and other organizations. At the Fourth Conference in Belgrade the participants to the meeting had mandate to discuss and adopt the documents of SEEPAG set up within their particular national judicial system.

The Fourth Conference of the SEEPAG that took place in Belgrade in December 15, 2004, and finally the representatives of the state prosecution offices of the SEE countries put their signature to the main documents of SEEPAG, as stressed by The SEEPAG General Guidelines; the main aim of SEEPAG is to speed-up the judicial cooperation in trans-border crime cases. In this respect SEEPAG Prosecutorial Focal Points were established in each of the above mentioned states. Close cooperation will be ensured with all the relevant regional institution and initiatives, in order to provide necessary assistance to the Task Forces functioning under the SECI Regional Center for Combating Trans-border Crime.

General Guidelines defines the main goals and objectives of SEEPAG:

1. to provide operational support and assistance to each other and to relevant law enforcement agencies
2. to provide advice, assistance, and guidance on criminal policy and legal issues related to law enforcement matters in the region if it is so permitted by national legislation and binding international treaties

Prosecutorial Focal Points recommendations define the tasks of PFP:

1. Receiving crime reports and documents
2. Case monitoring
3. Consulting
4. Providing remedies and suggestions
5. Participating in special operations
6. Communicating

At the Round Table of CARDS Regional Programe – CARDS JUDICIARY, in March 30, 2005, State Prosecutors of the six Western Balkan countries signed the Memorandum of Understanding to combat the organized Crime. In the Article 3, the EUROJUST, SEEPAG and other Contact Points were defined as the mechanism of cooperation.

We shared the same vision in the documents of the SEEPAG group and the same ideas about how the SEEPAG network will be organized and what are the next steps for organization of regional Prosecutorial Focal (Contact) Points in each participating state.

The most important legal problems for each of the SEEPAG members in the working groups were also the series of questions within a certain context – the one of the following: Witness Protection Narcotics, Trafficking in Humans, Special Investigative Techniques, and Data Protection. We had no much time for each group to deal with a certain set of issues, but prosecutors mostly were concentrated on witness protection legislation, were we have good national results. Many countries are introducing a new legislation on witness protection in this field of combat to trans-border crime. What is the next goal? It is to develop and propose a regional witness protection program in order to effectively combat organized crime syndicates and to protect witnesses who agree to testify in trials. Same is with the data protection programs and legislation, and other important fields.

To do so SEEPAG have two basic approaches that was identified at the *Brussels Conference on Regional Witness Protection Program*:

1. Establishing a regional program that will allow for the transfer of witnesses among participating states;

2. Setting up a twining program involving country institutions for providing assistance in witness protection programs.

Basic principles for the Regional agreement on Witness Protection should be:

- Witness protection program should be broadly based, and not based on a single type of crime
- The transfer of persons from another country should be taken into consideration and such transfers should be made based upon pre-defined criteria.
- Extremely tight security for the witnesses should be provided.
- The prosecutors in the region could be instrumental in the creation of such a regional witness protection program.

- To the extent possible, witnesses should be placed in their own countries, as witness protection is primarily a domestic problem. The need for regional relocation will apply to few witnesses.
- There are less severe measures that can be taken to protect witnesses, such as temporary relocation, name changes, temporary protection of witnesses in place, and long-term relocation.
- As many witnesses will be criminals, the program should provide for the protection of both incarcerated and free witnesses.
- Liability issues should be considered such as the potential risk created by witnesses to the communities in which they are relocated and the risk assessment programs and procedure similar to the military operations.
- A very rigorous security program is essential to safeguard information (personal data) related to protected witnesses.
- There must be a project director and the central authority that makes judgments on who can be included in the program.

The similar principles should be adapted to the future Regional Memorandum of Understanding on Data Protection and Combat to the Crime of High Technology (Hi-Tech Crime). If we want to save our cultural inheritage, our intellectual property and our digital inheritance, we need to prepare now.

Harmonization of different instruments of prosecutorial cooperation

Enhancing the mutual legal assistance through establishing direct channels of communication between prosecutors in the Western Balkans must be done in harmonized manner. Lack of coordination and compartmentalization of different instruments for fighting organized crime is highlighted by the Council of the EU as one of the key pitfalls in fighting organized crime in the Western Balkans ("Friends of the Presidency Report", Council of the EU, Oct. 2004).

In this respect, the fact that the SEEPAG is an operational mechanism consisted of larger number of the participating countries and has wider scope of activities must be taken into account. Setting up the independent prosecutorial network under the draft Western Balkans MoU, which would be parallel to the SEEPAG, would deepen coordination problem and decrease efficiency of both instruments. In particular, the following issues could arise:

- The ability of the SEEPAG to facilitate the SECI Center related cases would be infringed as far as the operations in the Western Balkans

countries are concerned. Existence of two parallel contact networks would open possibility that their operations are conducted in uncoordinated manner. Considering larger scale of countries participating in the SEEPAG, it is possible that the Western Balkans contact network could facilitate undertaking of acts in the criminal proceedings (e.g. arresting, publishing data etc) that would infringe larger scale operations.

- Experience in the field of judicial cooperation suggests that the Western Balkans contact network would have to develop special institutions (i.e. Secretariat) in order to build its capacity and thus ensure its efficiency. This could be a time consuming and costly process.
- Obtaining and maintaining of statistic data would be done by two networks, what would result by incompatible statistic data and less clear overall picture.

At the CARDS joint meeting of the Regional programs for Western Balkan (CARDS JUDICIARY and CARDS POLICE) – Round Table 2, in Strasbourg May 2005, the recommendations were proposed that the SEEPAG and the network that is to be established under the draft Western Balkans MoU should be harmonized to the greatest extend possible. Such harmonization should reflect the fact that the SEEPAG is consisted of larger number of participating countries, and has wider scope of activities. In particular the recommendations are that two steps should be taken:

- Contact Points that are to be established under the Western Balkans MoU and the SEEPAG Prosecutorial Focal Points should be integrated into the unified prosecutorial units for contact. In ideal case, one person should be contact point for implementation of both instruments.
- The cooperation envisaged under the Western Balkans MoU should use of the services of the SEEPAG Secretary. In particular, the Secretary should facilitate communication, liaison with international counterparts and maintain statistic data.

As each Conference of the SEEPAG takes place, it is the main goal to continually improve prosecutorial effectiveness so that the national prosecution services can show significant progress. This is possible only through the strong and active participation of all of the SEEPAG members. I call upon all of you to help develop and implement a road map for the future of the SEEPAG.

Physicality of Soul
Фи зика на Душата - Fisic an Anama

Mícheál Mac an Airchinnigh
University of Dublin, Trinity College, Dublin, Ireland

Keywords: *blog, digifoto, edutainment, games, google, ontology, Phizika na Dushata, physicality of soul, playing, world-wide web*

The idea of Physicality of Soul was born in South Eastern Europe in 2004. The very strangeness of the phrase, using old words in unusual combination, suggests something very new and of its time. The Bulgarian, *Phizika na Dushata (Fizika na Du{ata)*, tries to capture the sense of this idea from a local cultural perspective. At the other extreme Western Region of Europe, Fisic an Anama, translates the Bulgarian.

Physicality of Soul speaks of the spirit of the Digital re-Discovery of Culture made alive in the playful person, (s)he who is at play in the game of life. The World- Wide Web provides the modern cultural medium by which one begins to know the Other. Such knowledge in becoming is only fully realized in and by the body. This is Physicality of Soul.

Dancing typifies Physicality of Soul. So does play-acting and singing and running. There is play to be found everywhere. How can such playing be conjoined with the digital world? How can the digital world be humanized? This paper explores the possible.

1 Practical Sense

"In . . . the 1970's I organized a cycling tour . . . to Germany, . . . one needed detailed maps . . . a list of Jugendherberge, a telephone, and a practical knowledge of the German language. . . . I remembered stopping one day in a small German town . . . I gave an interview in my best (broken) German to the local newspaper, a copy of which (together with photograph) was sent to me later in Ireland. This typifies for me today, in some way, what it means for physicality of soul. Specifically I mean this : through the tools of map, e-communication (telephone) and language (German) I had set up a "discovery of culture" for myself and my students. But the actual experience of the culture did not take place until I and all the others were physically present in place. Then the soul-anticipation was physically realized." [5,212]

In my mind Physicality of Soul may be considered to be a development, an elaboration, of the idea of the Practical Sense of Philosophizing which

was explored in [6]. To philosophize is to be human. All humans do it. The ways of doing it, of philosophizing, differs from human to human and varies over age (time) and culture (space).The meaning of "practical sense" (an English translation of the French Sens Pratique [1]) was to be conveyed by the way in which every human acquired the Practical Sense of Space (through the developing body from seed to maturity). One might speak therefore of the "human being in becoming" to describe this process of acquiring, among other things, the Practical Sense of Space. Similarly, the developing process of reasoning, of philosophizing, grows with one.

Physicality of Soul is the opposite of "Spirituality of Body." I do not like the sound of the latter.

Physicality of Soul is coined to express the needed bridging between the digital experience and the grounding in physicality, personal reality.

Deliberate separation of the oneness expressed by Physicality of Soul into the Cartesian duality of body and soul, or body and mind, is desirable by militarists who wish the soldiers to kill at a distance without the personal reality of seeing the ones killed. Aerial bombardment is typical. The use of war games trains for such separation.

Let us imagine that you see a photograph of some place on the WWW, such as the Mural

Plate 61 in [11, 32]. See also [8] and [2].

It is an image that brings to mind much. Where is it? Who took the photograph? Why? What is the story it is meant to tell?

Let us imagine that you subsequently go to visit the place. Only then does one have a complete Physicality of Soul experience. Now let us imagine the reverse. This is a photograph of some place to which you have already been.

The photograph on the WWW brings back a memory. The Physicality of Soul experience is in the remembering. This aspect of Physicality of Soul has been captured well by Proust.

But a photograph can be older that one's lifetime. Here is one of the Church in Topcider, Belgrade, Serbia and Montenegro.

Groman's Photo Album 1876–1878

"The church in Topcider, built as a residence, by order of Prince Milos in 1834. On the left side from the church, there is a big building with a second-floor porch. In the foreground, to the right, there are two soldiers and a horseman, dressed in a white fur-bordered tunic; he may be an escort of Princess Natalija." [9].

We know that what is shown is gone. Even the three men are gone. What sort of Physicality of Soul experience might there be in this particular case? I have visited Belgrade in 2004 just as the photographer did more than one hundred years ago. In some way the photograph and the visit bracket the city's history, the history of the people who lived and live there. The experience was mine.

Every photograph is to be read, just as every painting and every tapestry and every book is to be read. In this photograph I am drawn to the figure on the right, the figure in black, smoking a pipe (?) and looking straight into the future though he could not have known it. Sometimes I think he looks just like me.

In the modern Digital re-Discovery of Culture (DrDC) the (digital) photograph plays a central role. Surely it is obvious how manuscripts are digitized? Even Groman's photograph is now digitized. Let us be

deliberately and precisely careless with language and presume that any "modern" photograph is de facto a digital image unless otherwise indicated? No? Then let us be deliberate and precise and use the term digifoto which turns out to be widely used on the WWW.

Anyone can tell their own story. Anyone can tell their own story by writing or singing or dancing or . . . Anyone can tell their own story by constructing (and manipulating) digifotos. We can expect to meet anyone through their stories on the WWW.

Let us consider the problem of encountering the Other through the World-Wide Web.

The Mural above tells many stories. One big story is the clash of neighbouring cultures and dominating cultures. One wonders if the Digital re-Discovery of Culture can bridge cultures in conflict? One wonders if DrDC games can help significantly in this?

The Mural stories reflect an understanding of History in becoming. Each side in a conflict has its own History. It is not surprising therefore that the teaching of History must be the subject of reform where neighbouring cultures are in conflict.

"There was no question that history teaching should be challenging, provoking and questioning, which is better achieved when controversial issues are included in the history curriculum, but, in the divided society of Northern Ireland, the problem of raising controversial issues lay mainly with adults. Indeed, in a divided society, where violence, death and destruction are common, where the wall murals, with their colour, language and slogans, define a territory and tell people who they are, myths become part of the identity of a person. Some studies show that, nearly 20 years after the introduction of new teaching methods, there has been no change in people's minds, mainly because the events have not supported history teaching and violence is in the streets everywhere, and each side sees itself as a victim." Keynote presentation on "Teaching materials: controversial and sensitive issues, multiperspectivity" by Ms Vivien KELLY, Northern Ireland in [15].

2 Digitized Manuscripts

"SEANCHAI: Somewhere on a hill\\ Once called TSAREVETS\\ In a neighbourhood\\ Once known as Veliko Turnovo\\ There are the ruins\\ Of what would appear to be a church\\ Of very strange type.

There is an old story\\ Dating from the AGE of COMPUTERS\\ Long, long ago\\ About the discovery of a great MANUSCRIPT\\ "Seanfhocail Mhíchíl"\\ And its digitization\\ And its distribution\\ To the WHOLE WIDE WORLD.\\ It is believed that the original was buried\\ On the hill.\\ Thanks

to the oral tradition\\ Preserved by humans\\ Some sayings of “Seanfhocail Mhíchíl”\\ Also known as “Old Words of Mícheál”\\ Are in circulation in the vernacular today:

CHORUS: Use your eyes,\\ look and see.\\ Use your heart,\\ feel and love.\\Use your mind,\\ think and reason.\\ Use your soul,\\ be at one. ” [4, 23]

Where is the Physicality of Soul experience for the digitized manuscript?

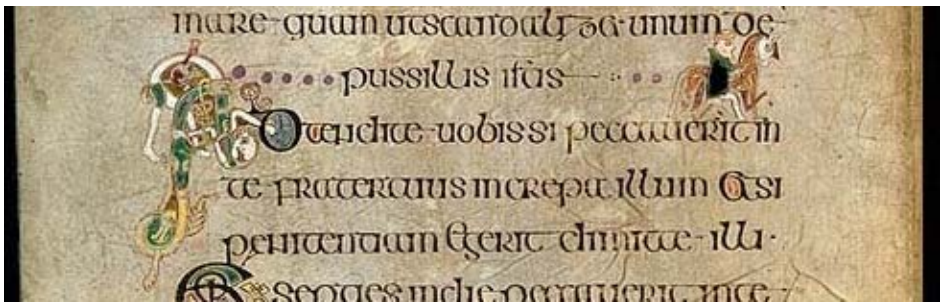
In the opening of his book, The White Castle, Orhan Pamuk [10, 1] arranges his narrator Faruk *Darvinoglu* to give one view of the purpose of manuscript: “I found this manuscript in 1982 . . . I wanted to concentrate on the story for its own sake, rather than on the manuscript’s scientific, cultural, anthropological, or ‘historical’ value.” Does it not seem to you that *Faruk Darvinoglu* does not care very much for “Cultural Heritage” as such? What would his opinion be of the Digital re-Discovery of Culture? Would he not be interested in the digitized thing, whatever it might be, qua thing?

Compare this story of Faruk Darvinoglu with the story of the Book of Kells. Today the Book of Kells is admired for the story of its writers and illuminators, not usually for the story it purports to tell, the “Good News” of the Christian Gospels. (The Mural Image is in a region of the “Island of Ireland” nominally Christian of differing persuasions.)

Consider for example the image of the “green monk”(?) on folio 241 recto which is included in [7].

Is this not a “self-portrait” of the illuminator?

And who is the man (“green monk”?) with the horse on folio 255 verso?



Man and Horse (?); Monk and Horse (?); Why?

Are not these kinds of questions the exact opposite to those of *Faruk Darvinoglu*?

Let us imagine you have seen the digitized copy (CD) of the Book of Kells. Then to see the original Book of Kells and experience something of the usual Physicality of Soul one must travel to Trinity College, Dublin, Ireland, where you will be very welcomed. For some the experience of seeing the original is disappointing, like that of seeing the much acclaimed Mona Lisa in the Louvre.

Alternatively, and although not the same quality of experience, one might be fortunate enough to have access to one of the many facsimile copies of the Book of Kells.

3 DrDC Game on Ohrid

The practical sense of Physicality of Soul is currently being explored in a collaborative research project within the Marie Curie Transfer of Knowledge project KT-DigiCULT-BG which commenced in May 2004. The relevant key component of the project is "Edutainment" and the principal researchers are the author and Ms. Kalina Sotirova, IMI BAS.

The vehicle by which the research is conducted is the Digital re-Discovery of Culture Game. This has been under development of twelve months now and the first results have been published [3].

The game is designed to exhibit the experience of Physicality of Soul.

The game is to be played by anyone intending to visit Ohrid 11th–14th September 2005 for "The First South-Eastern European Digitization Initiative (SEEDI) Conference, DIGITAL (re-)DISCOVERY of CULTURE (PHYSICALITY OF SOUL), — Playing. Digital —". Therefore the target culture of the game is Macedonian, with a specific focus on the location: Ohrid and its surroundings. To experience Physicality of Soul the player needs to be able to get to the objects of the game in the places indicated in the game.

The game ought also be able to surprise those who are Macedonian and even those who come from Ohrid itself.

Here ends the Extended Abstract.

In the first stage of the design it is necessary to collect some important Web pages which deal with Macedonia and Ohrid. Let us begin with the "official" locations:

1. Macedonia [17]
2. Macedonia, Ohrid [14]
3. Macedonia, Ohrid, Cultural and Historical Heritage and Treasures of Art [23]
4. Macedonia, Ohrid, Hotel Inex Gorica [18]

and add in some others:

1. Macedonia, Ohrid, Heritage Trails [21]
2. Lychnidos [19]
3. City of Light [13]
4. St. Clement of Ohrid [22]
5. Macedonian Culture and Information Centre [20]

Irish Cultural Heritage Organization: An Taisce [12]

European Cultural Heritage Organization: ENNHO [16]

Some questions we might ask in designing the game:

Where does the name Ohrid come from? What does it mean? Variants are Ochrida, Okhrida, and Achrida.

Candidate for the Keyword/Keyphrase: **Lychnidos**.

Photo taken near the Digital Hub in Dublin, 2005-06-14.

Table 1: The Ten Names

Raymond McCreesh 21-5-1981 Patsy O'Hara 21-5-1981		Joe McDonnell 8-7-1981 Martin Hurson 13-7-1981
	Bobby Sands 5-5-1981 Francis Hughes 12-5-1981	
Kevin Lynch 1-8-1981 Thomas McElwee 8-8-1981		Kieran Doherty 2-8-1981 Michael Devine 20-8-1981

Corresponding goal: to discover why Ohrid is the City of Light.

4 Acknowledgements
Thérèse Mac an Airchinnigh proof-read the text. Kalina Sotirova advised on the subject matter of Macedonia and Ohrid.

References

[1] Pierre Bourdieu. The Logic of Practice. Stanford University Press, Stanford, California, 1990. Richard Nice (trans.) from the French "Le sens pratique", Les Éditions de Minuit, Paris, 1980. ISBN 0-8047-2011-8.
[2] Kathryn Conrad. Web page. Northern Ireland Photo Gallery: murals, graffiti, marches, and protests. http://people.ku.edu/~kconrad/murals.html, 2005. {20050612}.
[3] Suzanne Gormley. DRDC: A digital discovery/re-discovery of culture and the Semantic Web. Department of Computer Science, University of Dublin, Trinity College, Dublin 2, Ireland. April 2005. A 4th year undergraduate dissertation under the supervision of Dr. Mícheál Mac an Airchinnigh.
[4] Mícheál Mac an Airchinnigh. Web page. The text of a play: Seanfhocail Mhíchíl. https://www.cs.tcd.ie/Micheal.MacanAirchinnigh/Web%20Pages/ME_FEIN/SCRIBHNEOIREACHT/DRAMA/SeanfhocailMhichil1999.pdf, 1999. {20050612}.
[5] Mícheál Mac an Airchinnigh. The Experience at Trinity College Dublin. International Journal Information Theories & Applications, 11(3):211–20, 2004. ISSN 1310-0513.
[6] Mícheál Mac an Airchinnigh. The Practical Sense of Philosophizing:

Why preserve anything at all, even digitally? Review of the National Center for Digitization 4, pp. 111–134, 2004. ISSN 0353-5738. http://www.komunikacija.org.yu/komunikacija/casopisi/ncd/index?stdlang=en {20050406}.
[7] Mícheál Mac an Airchinnigh. The Graven Image — Digitized and Philosophized. Review of the National Center for Digitization 6, pp. 12–44, 2005. ISSN 0353-5738. http://www.komunikacija.org.yu/komunikacija/casopisi/ncd/index?stdlang=en {20050406}.
[8] Jonathan McCormick. Web page. Mural Directory. http://cain.ulst.ac.uk/mccormick/intro.htm, 2005. {20050612}.
[9] Zarko Mijajlović et al. Web page. GROMAN'S PHOTO ALBUM 1876-1878. http://www.ncd.matf.bg.ac.yu/projects/en/groman.html, 2003. Published by The Belgrade City Museum, The Faculty of Mathematics in Belgrade, The Mathematical Institute of SANU. An Electronic Version is available for download in pdf format, {2004-04-14}. The CD is catalogued as ISBN 86-7589-0370-0.
[10] Orhan Pamuk. The White Castle. Carcanet Press Ltd, Manchester, England, 1990. Victoria Holbrook (trans.) from the Turkish Beyaz Kale, Can Yayin Lari Ltd.,1979. The Paperback edition of 1991 is cited. ISBN 0-571-16466-8.
[11] Bill Rolston. Drawing Support 3, Murals and Transition in the North of Ireland. BTP Publications Ltd, Belfast, 2003. ISBN 1-900960-23-0.
[12] Author WWW. Web page. An Taisce. http://www.antaisce.org/about/index.html, 2005. The author is unknown. {20050612}.
[13] Author WWW. Web page. balkananalysis.com — Ohrid Through the Ages. http://www.balkanalysis.com/modules.php?name=News&file=article&sid=376, 2005. The author is unknown. {20050612}.
[14] Author WWW. Web page. City of Ohrid. http://www.ohrid.com.mk/angliska%20verzija/City%20of%20Ohrid.htm, 2005. The author is unknown. {20050612}.
[15] Author WWW. Web page. Council of Europe, Reform of History Teaching. http://www.coe.int/T/E/Cultural_Co–operation/education/History_Teaching/Reform_of_History_Teaching/South_East_Europe/d.Ohrid%20bivar%20black.asp, 2005. Maria Luisa de BIVAR BLACK, Portugal. {20050612}.

[16] Author WWW. Web page. ENNHO: European Network of National

Heritage Organisations. http://www.ennho.org/index.stm, 2005. The author is unknown. {20050612}.
[17] Author WWW. Web page. General facts about Macedonia. http://www.ohrid.com.mk/angliska%20verzija/Macedonia.htm, 2005. The author is unknown.
{20050612}.
[18] Author WWW. Web page. Inex Gorica. http://www.inexgorica.com/index_eng.htm, 2005. The author is unknown. {20050612}.
[19] Author WWW. Web page. Lihnidos. http://www.soros.org.mk/konkurs/076/angver/lihnid.html, 2005. The author is unknown. {20050612}.
[20] Author WWW. Web page. Macedonian Cultural & Information Centre. http://www.macedonia.co.uk/mcic/index.asp, 2005. The author is unknown. {20050612}.
[21] Author WWW. Web page. Ohrid at your fingertips. http://www.culture.org.mk/ePROEKTI.HTM, 2005. The author is "Vertikala – progressive ideas and technologies" -- Skopje. Select "Heritage Trails". {20050612}.
[22] Author WWW. Web page. St. Clement of Ohrid. http://www.mymacedonia.net/language/clement.htm, 2005. The author is unknown. {20050612}.
[23] Author WWW. Web page. treasures of arts. http://www.ohrid.com.mk/angliska%20verzija/treasures%20of%20art.htm#robevi, 2005. The author is unknown. {20050612}.

Usability of ICT-based systems in memory institutions: state-of-the-art and future perspectives

Zinaida Manzhukh
LIS Institute, Faculty of Communication, Vilnius University

Developments in ICT-based systems encouraged new ways of preservation of and providing access to the cultural heritage (CH) in libraries, museums and archives, enabling them to act in co-operation as memory institutions. Digital space offers rich and interactive environment for cultural heritage representation. However, consumption of technology-mediated CH services may become a challenge both for memory institutions and their patrons because of different level of skills, knowledge, ability to use technologies and other important factors that limit initial advantages offered by digital tools. Therefore, usability, which encompasses processes, criteria, methods to support effective, efficient and satisfactory use of ICT-based systems, is of increasing concern to the CH institutions. The aim of this article is to provide an overview of current achievements in usability and delineate research directions for adoption of these principles in memory institutions. For this purpose, the "usability" concept and underlying approaches, interpretations, emerging in the works of leading specialists, corporate guidelines, and ISO standards are discussed. Future perspectives for usability research in memory institutions are highlighted taking into consideration problematic areas, such as issues of shared usability model for memory institutions, current discussions on communication of cultural heritage to the user and the need for multidisciplinary research in designing CH services and their usability. This article is based on the research report *Usability of ICT-based systems: state-of-the-art review,* which is a part of CALIMERA (Cultural Applications: Local Institutions Mediating Electronic Resource Access, http://www.calimera.org) project funded under the European Commission Information Society Technology Programme.

Center for Digitization of National Heritage at Institute of Informatics

Marija Mihova, Vesna Dimitrova
Institute of Informatics, Faculty of Sciences, Skopje, Macedonia

Keywords: *national heritage, standardization, digitization*

Background

The national heritage of Macedonia is big, but a small part of that is in digital form. There is also, not enough information publicly available about the national heritage of Macedonia, especially in digital form. Most of digitized materials are result of the limited projects. This is the reason why on the Internet you can find only partial information about our national heritage and in worst cases not valid information from unreliable sources. The need of centralization, standardization and categorization of all digitization materials is understandable.

Institute of Informatics at Faculty of Science in Skopje, as one of the most important institution in IT education in Macedonia takes some initiative to help in presentation of Macedonian national heritage. For that purpose is established center for digitization of national heritage.

Vision of the CDNH

The vision of the CDNH is to help developing of a digitized presentation of the national heritage of Macedonia via different kind of activities.

The goals of the CDNH

The main goals to be achieved are:

- Building a web portal of the CDNH
- Producing application programs
- Developing and applying suitable programming tools for national heritage
- presentations
- Organizing seminars, workshops and conferences
- Involving new teaching materials in student's education at Institute of Informatics in order the student's conscience of national heritage to be spread out.

The realization of the goals will be obtained in close cooperation with state institutions (Ministries of Culture and Education and Science), central

and local museums, libraries, center for conservations, institutes (Institute of Macedonian Language, Institute of Folklore, Institute of Ethnology,…), faculties (Faculty of Architecture, Faculty of Music Arts, Faculty of Philosophy, Faculty of Philology,…), private companies, individuals and others from Macedonia. The CDNH will also closely cooperate with international community (UNESCO, SEEDI, MINERVA,…)

The program of activities of the CDNH

The program of activities of the CDNH is planned as follows.

- Building a web portal
- Producing application programs
- Mutual discussion with representatives of different institutions and companies
- Organizing student courses
- Organizing seminars, workshops and conferences
- Visiting and active presentations on domestic and international meetings
- Partnerships in domestic and international projects for digitization of national heritage

The preconditions for realization of the program are:

- technical facilities (the Institute of Informatics has some bounded capacity, to be completed from domestic and/or international projects)
- software tools (to be obtained partly from the Institute of Informatics, partly from other domestic and international institutions, private companies, partly to be developed at the CDNH)
- network connections (to be obtained from the Institute of Informatics)
- financial support (to be obtained from Ministries of Culture and Education and Science via domestic projects, from cooperation with private companies, and via partnerships in international projects)
- the willingness of members of Institute of Informatics and members of other institutions to participate at the program
- the student activities at the seminars

Achievements

There are already some activities for realization of existing agenda: the first SEEDI conference is organized by this center, multimedia data base for Web presentation is already in development along with some stand alone and Web applications. There is also collaboration between many other institutions that work with national heritage and companies that work with technical aspect of digitization of national heritage.

From Presentation to Reality, from Preserved to Heard

- Experience in the digitization of analogue sound recordings in the Music Collection of the National and University Library in Ljubljana -

Simona Moličnik Šivic
National and University Library Ljubljana, Slovenija

Keywords: *digitization, metadata, analogue sound recordings, gramophone records, music archives, history of music*

The Music Collection of the National and University Library also stores a set of historical sound recordings representing a rarity in the frames of Slovenia. It consists of approximately 50 old sound recordings, mostly from before the World War II. These recordings are not reproducible any more with classical gramophone equipment, and shellack from which these records are made is extremely fragile. For the purpose of protecting the records themselves and to enable listening to them it was decided to transfer the sound recordings from analogue medium to the digital one.

There are approximately 100 music pieces on these recordings, mostly of solo singers (with piano or small orchestra accompaniment), choruses and smaller vocal groups. Independent instrumental performances are only a few (usually accordion, wind instrument orchestra). The repertoire is limited to adaptations of folk songs and to artistic compositions and partly also to "amusing" scenes with singing. Gramophone records were published by different, mostly foreign publishers, among others Columbia, Elektroton, Edison Bell Radio, His Master's Voice, Odeon, RCA Victor etc. There was no music publisher in Slovenia at that time and Slovenian performers were recorded by foreign studios. Many gramophone records came as a donation from the Slovenian emigration.

Sound recording represent a cross cut of recording in Slovenia and its coverage in a distinct period of Slovenian music history, mainly from the period from before the World War II; as well, they are important because they represent real sound and reveal the state of preserved documents. The dates of the recordings have not been easily determined. There are no data on the materials themselves, nor could they be found in publishers catalogues. Approximate dates were established by the use of different

written sources. The materials from the period of World War II makes us assume that the production was much bigger than known today.

An audio source can represent a starting point for further research of the outline of the significance of this outline of discographic production in Slovenia from the point of view of history, culture and publishing activities. On the level of basic tasks of the national and university library, performed digitization enabled revitalization and access to forgotten and technically inaccessible Slovenian artistic cultural heritage. It reveals interpreting, receptive and stylistic traits of the time which we were so far able to reconstruct only in frames from secondary sources (published reports and critiques, memories, correspondence etc.). For the understanding and study of the history of music this acoustic documentation is of unprecedented value and adds to the central role of the national library, i.e. collecting, preserving and use of important cultural heritage.

To be able to register and preserve our cultural heritage to a great extent, further historical research studies of secondary sources will be needed, as well as additional systematic field research, location of materials with private collectors and use of bibliographies of lesser producers of gramophone records (usually unpublished) or their archives (if accessible). With each new discovery, our digital collection will grow bigger. We are therefore kindly asking all those who know about any existing sound recordings to pass this information on to us.

The present contribution is presenting the concept and methodology of work in detail, as well as the process of the transfer from analogue to digital format, the organization of metadata: as well, it concentrates upon the experience and exposes the results, followed by the description of starting points for further plans for the digitization of music periodicals, printed music, manuscripts and correspondence.

An XML representation of Labanotation LabanXML and its implementation on the notation editor LabanEditor2

Minako NAKAMURA[1], Kozaburo HACHIMURA[2]
[1] Ochanomizu University, [2] Ritsumeikan University, Japan

Keywords: *dance, Labanotation, XML, editor*

Labanotation is widely used as a general body motion notation, because it does not depend on any specific dance. Since Labanotation is based on graphical representations, it is also widely accepted as a readable and understandable notation.

There are number of graphical editors especially designed for Labanotation, but files for those editors are not designed for inputting, searching, and editing, as they are only designed for their internal file representations. Text Representation of Labanotation is required for interchanging Labanotation data via Internet, searching specific motion patterns, analyzing dance movements, and archiving body motion data. There are several types of text representations. Among these, XML eXtensible Markup Language is widely used for the purpose. Especially on the Internet, XML is the choice for most applications. However, XML for Labanotation has not been developed.

In this paper, we discuss an XML representation of Labanotation, called LabanXML.

There is no interchange format for Dance notations. However, as for interchange formats for music notation, there are several formats including SMDL, NIFF, and MusicXML. SMDL, Standard Music Description Language is developed by ISO, International Organization of Standardization. SMDL is a variant of SGML and is universal but too complex. NIFF, Notation Interchange File Format, is designed to exchange only Graphic Notation. Recently, MusicXML is developed by Michael Good based on XML and designed for

Interchanging music data via the Internet and capable of searching.

Dance is usually accompanied by Music. XML representation for Labanotation compatible with MusicXML is required. Thus we designed LabanXML compatible with MusicXML.

Configuration of Labanotation is vertical while Music staff notation is horizontal.

Both proceeds with time line. Thus for the treatment of time, LabanXML follows the same approach taken by MusicXML. In music staff notation, pitch and duration represented by music notes put on a line or between lines. In Labanotation, direction, heights and duration of body movement represented by symbols puts between lines. In Staff notation, lines represent pitch value. In Labanotation, between lines called column represent body parts.

In Labanotation, symbols are drawn on the columns separated by horizontal line. This horizontal line represents measure, that is, time line. Direction of a body movement is represented by a shape of a symbol, level is represented by a texture, duration is represented by its length. To represent a movement of a specific body part is to put this symbol on a column of the staff. More detailed movement of a body part is represented by additional symbols called Body Sign.

For designing LabanXML, we choose that a measure includes columns and is the most basic XML element. A measure ELEMENT has an attribute “num” which represent a number of a measure. A characteristic of LabanXML design is a separation of support column. For representing columns, XML elements, <left>, <support>, and <right> are included by <measure> element.

The root element of LabanXML is the <laban> element. The <laban> element includes the <attribute> and <notation> elements. The <attribute> element includes <time> element. And the <time> element includes <beat> and <beat-type> elements.

Most important part of Labanotation is represented by <notation> element. The <notation> element includes <repeat> element, which describes a repetition, and <measure> element. Columns of Labanotation is classified into <left>, <support> and <right> elements. As explained before, separation of support

column as <support> element and classify other left and right columns into <left> and <right> elements are one of our unique approach to design XML representation for Labanotation. Relationship between body parts are represented by <relationship> element.

XML view of data or information is to recognize data or information as a logical structure of elements. Basic construct of information is recognized as an element in XML. And logical structure is represented by Grouping, Occurrence Indicators, and Connectors. Grouping, Occurrence indicators, and Connectors are SGML terms and SGML, Standard Generalized Markup Language, is the ancestor of XML. XML can be viewed as a simplified SGML.

DTD of the LabanXML is shown on the annex.

Logical tree structure of elements are represented by Grouping. As in ELEMENT <laban>, <laban> includes <attribute> and <notation> elements. The element definition start with less than symbol followed by exclamation mark and capital ELEMENT. Next column is the name of an element. Content of the subject element is the next column, called content model.

Occurrence indicator is represented by asterisk, plus, and no special character after element name. Asterisk means the elements will occur zero or more, plus means one or more, no special character after element name means occurs exactly once in the content model. Connector is represented by Comma or Vertical bar. In this DTD, only Comma is shown. Comma means elements before the comma and after the comma occur in this order. Vertical bar means selection of elements.

XML has very limited abstraction mechanism. Sometimes same contents may appear in several elements. To represent this, ENTITY is the only way to represent this abstraction.

An entity BODYP is defined on the first line and referenced in the left ELEMENT definition. Percent BODYP appeared on the <left> element is replaced by a character string, "hand*, arm*, body*, leg*". #PCDATA means a real value. So, element <level> may has a value of high, middle, or low.

We have developed most of LabanXML for symbols of Labanotation. Analysis of Labanotation and design of LabanXML

has almost been completed. Labanotation score can be generated by using XSLT (eXtensible Style Sheet Language Transformation) and SVG (Scalable Vector Graphic) both developed by W3C.

Preliminary version of LabanXML has been implemented in a system called LabanEditor2. LabanEditor2 is an extension of LabanEditor developed by us. LabanEditor2 is an interactive graphical editor for writing and editing Labanotation scores. By using LabanEditor2, a user can input and edit human body movement of dance and also display animation of a human body model in 3D graphics. LabanEditor2 can now read and write LabanXML data.

Dance analysis and dance archive using LabanXML is the next research challenges.

Annex LabanXML DTD

```
<!ENTITY     % bodyp "hand?, arm?, body?, leg?" >
<!ENTITY     % symbols "bodypart?, direction, level, contraction?,
rpin?,
                hc?, hook?, vl?" >
<!ELEMENT  laban (attribute, notation)>
<!ELEMENT  attribute (beat, beat-type)>
<!ELEMENT  notation (repeat*, measure+)>
<!ELEMENT  repeat (repeatpart)>
<!ELEMENT    measure (relationship*, path*, left?, support?,
right?)>
<!ATTLIST   measure
                num CDATA #IMPLIED>
<!ELEMENT  relationship (others, %bodyp)>
<!ELEMENT  path (direction)>
<!ELEMENT  left (%bodyp;)>
<!ELEMENT  support (%symbols;, turn*)*>
<!ATTLIST   support
                side CDATA #IMPLIED
                ticknum CDATA #IMPLIED
                duration CDATA #IMPLIED>
<!ELEMENT  turn (#PCDATA)>
<!ATTLIST   turn
                type CDATA #IMPLIED>
```

```
<!ELEMENT  right (%bodyp;, head)>
<!ELEMENT  hand (%symbols;)*>
<!ATTLIST    hand
                ticknum CDATA #IMPLIED
                duration CDATA #IMPLIED>
<!ELEMENT  arm (%symbols;)*>
<!ATTLIST    arm
                ticknum CDATA #IMPLIED
                duration CDATA #IMPLIED>
<!ELEMENT  body (%symbols;)*>
<!ATTLIST    body
                ticknum  CDATA#IMPLIED
                duration CDATA #IMPLIED>
<!ELEMENT  leg (%symbols;)*>
<!ATTLIST    leg
                ticknum CDATA #IMPLIED
                duration CDATA #IMPLIED>
<!ELEMENT head (%symbols;)*>
<!ATTLIST   head
               ticknum CDATA #IMPLIED
               duration CDATA #IMPLIED>
<!ELEMENT  repeatpart (#PCDATA)>
<!ELEMENT  beat (#PCDATA)>
<!ELEMENT  beat-type (#PCDATA)>
<!ELEMENT  bodypart (#PCDATA)>
<!ELEMENT  direction (#PCDATA)>
<!ELEMENT  level (#PCDATA)>
<!ELEMENT  contraction (#PCDATA)>
<!ELEMENT  rpin (#PCDATA)>
<!ELEMENT  hc (#PCDATA)>
<!ELEMENT  hook (#PCDATA)>
<!ELEMENT  vl (#PCDATA)>
```

On the Applicability of Protégé/OWL in Building Software Tools for Intelligent Search in Digitized Collections of Manuscripts

Maria M. Nisheva-Pavlova, Pavel I. Pavlov
Faculty of Mathematics and Computer Science,
Sofia University "St. Kliment Ohridski", Sofia, Bulgaria

Keywords: *Manuscript Digitization, Semantic Web, XML, OWL, Protégé*

During the last decades information technologies play a significant role in lots of successful projects directed to digital preservation of cultural and scientific heritage. The growth of the number of digitized collections of manuscripts and printed editions increases the necessity of proper software tools assisting the access to these collections and making the best use of them.

One of the consequences in this direction is the growth in the development of proper search methods and tools. Instead of the facilities supported by the traditional keyword-based search engines, many users prefer to formulate queries in terms of high-level semantic concepts that are more relevant to their professional needs. In these cases the search engine is provided with a phrase which is intended to denote an object about which the user is trying to gather information. The aim is to find a suitable set of documents which together will give him the necessary information.

In [5] we suggest a methodology for development of tools for semantics oriented search in repositories of digitized manuscripts which is designated to assist the search activities in collections that may enlist XML documents which should be:

- catalogue descriptions of manuscripts compatible with the document type definition structure suggested by the project MASTER and adopted by TEI;
- marked-up full texts of manuscripts that may be written in different languages.

It is directed to the development of software environments that will be able to deal with user queries containing words or phrases that are considered as

domain concepts.

The emphasis in our methodology falls on two main types of activities:

- Development of proper ontologies describing the conceptual knowledge relevant to the chosen domain(s). These ontologies define sets of concepts with their basic properties and the relationships (mainly hierarchical in our case) between them. The concepts should be defined in many languages.
- Development of proper intelligent agents for search and processing purposes that are able to retrieve and filter documents by their semantic properties.

The paper presents some considerations related to the implementation of this methodology. The main features of the Web Ontology Language (OWL) and the open ontology development environment Protégé are analyzed from the point of view of their applicability in building software tools for intelligent search. The conclusion is that the OWL Plugin of Protégé is an excellent instrument for the development of tools for semantics oriented search in digitized collections of manuscripts.

As it was discussed above, our methodology is oriented to the construction of proper ontologies relevant to the domain(s) of the user queries and some intelligent agents for search and query processing purposes. An *ontology* is an explicit specification of a conceptualization. Ontologies define domain concepts, their properties and the relationships between them, and thus provide a domain language that is meaningful to both humans and machines. They are formal theories supporting knowledge sharing and reuse which play a significant role in the development of the so-called Semantic Web, "an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation" [1].

Each ontology should adequately represent a specific domain and allow some kind of formal reasoning. Ontologies should be both understandable by humans and processable by software agents. They can be used in particular to annotate Web resources.

There are lots of ontology description languages and ontology development software tools available nowadays.

The Web Ontology Language (OWL) [6] is widely accepted as the standard language for ontology construction and sharing Semantic Web

contents. It is based on a description logic model that makes possible to define and describe concepts and to make reasoning about them.

Protégé [2] is an open ontology development environment with a large community of active users. Protégé's model (the internal representation mechanism for ontologies and knowledge bases) is based on a flexible metamodel, which is comparable to object-oriented and frame-based systems. It basically can represent ontologies consisting of classes, properties (slots), property characteristics (facets and constraints), and instances. Recently Protégé has been extended with support for OWL, and has become one of the leading OWL tools.

Protégé provides functionality for editing classes, properties, and instances. Its user interface consists of several screens, called *tabs*, which display different aspects of the ontology in different views. Each of the tabs can be filled with arbitrary components. Most of the existing tabs provide a tree-browser view of the model, with a tree on the left and details of the selected node on the right hand side. The details of the selected object are typically displayed by means of *forms*. The forms consist of configurable components, called *widgets*. Typically, each widget displays one property of the selected object.

The OWL Plugin [4] is a complex Protégé plugin with support for OWL. It can be used to load and save OWL files in various formats, to edit OWL ontologies and to provide access to reasoning tools based on description logic. The OWL Plugin's user interface provides various default tabs for editing OWL classes, properties, individuals, and ontology metadata. Protégé can save ontology descriptions in various formats (OWL, RDF, CLIPS etc.). As a starting point for the implementation of intelligent search agents we prefer the standard DIG code generated by the OWL Plugin because of its simple and clear structure.

The most important view in the Protégé OWL Plugin is the OWLClasses tab. This tab displays the tree of the ontology's classes on the left, while the selected class is shown in a form in the center. The tree widget of the OWLClasses tab is organized according to the subclass hierarchy. Protégé users can browse, view, and edit the classes from the tree, create new subclasses, and move classes easily with drag-and-drop. The OWL Plugin also allows to navigate and edit ontologies according to other relationships between classes. Thus the user interface of the OWL Plugin of Protégé is very convenient for knowledge engineers and can be accessed without

serious problems by other types of users (philologists, historians, librarians etc.).

The OWL Plugin can interact with any reasoner that supports the standard DIG interface, such as Racer [3]. During ontology design, the most interesting reasoning capability from this type of tools is classification.

Classification is used to infer specialization relationships between classes from their formal definitions. Basically, a classifier takes a class hierarchy including the logical expressions, and then returns a new class hierarchy, which is logically equivalent to the input hierarchy. Protégé can display the classification results graphically. After the user has clicked the classify button, the system displays both the asserted and the inferred hierarchies, and highlights the differences between them. These visualization facilities can utilize the interaction between knowledge engineers and domain specialists and thus should increase the effectiveness of the knowledge acquisition process.

The mentioned reasoning capability associated with description logic is of particular importance because it allows the user to provide intensional definitions for the classes. Using OWL, ontology designers could just add a new concept by describing its logical characteristics, and the classifier would automatically place it in its correct position. Furthermore, it would report the side-effects of adding a new class. In this context Protégé/OWL seems to be a proper tool for building software environments intended to perform some kinds of clustering of manuscript collections in accordance with the features of the authors or scribes of manuscripts.

These considerations can serve as arguments for the conclusion that Protégé and especially its OWL Plugin is a perfect instrument for the development of software tools for intelligent search in digitized collections of manuscripts.

This work has been funded by the EC FP6 Project “Knowledge Transfer for Digitisation of Cultural and Scientific Heritage in Bulgaria” (KT-DigiCULT-BG).

References

1. Berners-Lee, T., J. Hendler, O. Lassila. The Semantic Web. Scientific American, May 2001, pp. 35-43.

2. Gennari, J., M. Musen, R. Fergerson, W. Grosso, M. Crubézy, H. Eriksson, N. Noy, S. Tu. The Evolution of Protégé-2000: An Environment for Knowledge-based Systems Development. International Journal of Human-Computer Studies, Vol. 58(1), pp. 89–123, 2003.
3. Haarslev, V., R. Moeller. RACER user's guider and reference manual. http://www.cs.concordia. ca/˜faculty/haarslev/racer/, 2003.
4. Knublauch, H., R. Fergerson, N. Noy, M. Musen. The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications. Third International Semantic Web Conference, Hiroshima, Japan, 2004.
5. Nisheva-Pavlova, M., P. Pavlov. Tools for Intelligent Search in Collections of Digitized Manuscripts. To appear in: Proceedings of the 9th ICCC Conference ElPub 2005 (Leuven, Belgium, 8-10 June 2005).
6. Smith, M., C. Welty, D. McGuinness. OWL Web Ontology Language Guide (W3C Recommendation). http://www.w3.org/TR/owl-guide/, 2004.

Tools for Search in Electronic Collections of Manuscript Descriptions Realized in the Integrated Environment XEditMan

Pavel I. Pavlov
Faculty of Mathematics and Computer Science,
Sofia University "St. Kliment Ohridski", Sofia, Bulgaria

Keywords: *Manuscript Digitization, XML*

On-line access to information is closely related to the adequate presentation of different cultures in the global information society. On the current European scene, where cultural differences and similarities are playing a key role in the integration of the Old continent, this process is of special importance. While information technologies keep offering cultural institutions a variety of opportunities for presentation and access to resources, this cannot be claimed for cultural heritage collections in Bulgaria. They still cannot be widely accessed in electronic form. One typical example is the mediaeval manuscript heritage.

Bulgarian repositories store about 12,500 manuscripts of Slavonic, Greek, Latin, Ottoman Turkish and other origins. Neither catalogue information on them, nor digital images could be consulted using the Internet. Although work on entering catalogue data is in the focus of interest of various research groups for about 10 years, a consulting materials on local collections is still not possible.

Our analysis shows that in order to offer more powerful tools which would lead to faster preparation and better use of electronic resources on mediaeval manuscripts, we should develop specialised tools for data entry, processing and visualisation. A set of such tools was developed by the author of this paper [1,2]. It helps to produce descriptions in electronic form faster and with better quality, and is of great help for the study of the material by target audiences with various interests. The latter is important in the light of personalisation in the work with Internet.

The paper presents an extension of the specialized editor XEditMan (*XML Editor for Manuscript Data*) which is an XML-oriented tool for editing and browsing catalogue descriptions of mediaeval manuscripts. The original version of XEditMan offers a friendly interface for entering data on mediaeval manuscripts, visualization and execution of queries to the descriptions that are already available. The descriptions are compatible with

the document type definition (DTD) structure suggested by the project MASTER (*Manuscript Access through Standards for Electronic Records*) and adopted by the Text Encoding Initiative. During the data entry the elements which are filled in appear in a sequence which is adopted in the manuscript cataloguing practice. The interface is in Bulgarian and this facilitates preparing electronic descriptions by people who are not acquainted in details with the DTD structure. The tool can be used also for visualization of single descriptions in two modes: complete descriptions or user-selected group of elements. Comparative study of multiple descriptions is achieved through database queries. XEditMan was used in the Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences for the first mass data entry on Old Bulgarian manuscripts. Currently, a collection of 807 descriptions is available.

The visualization of single descriptions is convenient for users that are interested in a particular manuscript. Visualization may be alternated by editing of an existing description and vice versa. The search of a word or phrase in such description is limited by the facilities of the particular browser. In the course of use of XEditMan we realized that it is necessary to have a proper tool enabling the search of a word or phrase in the entire collection of manuscript descriptions. Such global search can give some global information about the contents of the manuscripts and their descriptions. An important question here is how to choose the set of most important elements and attributes in order to restrict the search activities in them due to effectiveness considerations. Interesting phrases as search queries should be the names of historic personalities, authors and scribes of manuscripts (although the names of authors usually are unknown), features of the manuscripts decoration, etc. The additions made by the scribes in the margins of manuscripts can serve as an important source of information. Usually they contain records about natural and social phenomena getting people excited at that time. All these considerations predetermined the further extensions of XEditMan.

Recently XEditMan has been extended by tools for search of key words and phrases in a set of elements and attributes of all XML documents in an existing collection of manuscript descriptions prepared in accordance with the MASTER standard. The chosen set of elements and attributes includes: altName, watermarks, collation, foliation, scribe, script, medium, handDesc, origPlace, origin, acquisition, textLang, decoration, binding, condition, msContents, author, title, incipit, explicit, additions. We think that these elements and attributes are most informative. After the search tool is started, the user is asked to type the key word or phrase representing the

query. Then he/she can choose a specific subset of the elements and attributes listed above. For search purposes we use a built-in function which performs textual comparison. We consider this option as most appropriate because it does not distinguish between capital and small letters and thus allows one to find more appearances of the searched text.

As a result the chosen XML elements and attributes of all documents in the collection containing words or phrases equal to the one given by the user should be properly visualized. More precisely, the displayed search results consist of a sequence of paragraphs, each in its turn containing the name of a XML document found by the search tool, the name of the corresponding manuscript, the name of an element or attribute and the text content of this element or attribute. The discovered words and phrases that match the one given as a user query are highlighted.

Our first experiments have been carried out with the mentioned above collection of 807 catalogue descriptions of mediaeval Bulgarian manuscripts stored in Bulgaria. The search in this collection is completed in approximately 5-6 seconds. We consider the results of these first experiments with the search tool of XEditMan as promising.

This work has been funded by the EC FP6 Project "Knowledge Transfer for Digitisation of Cultural and Scientific Heritage in Bulgaria" (KT-DigiCULT-BG).

References

[1] M. Dobreva, P. Pavlov. How to Enter Data on Mediæval Bulgarian Manuscripts More Easily, or XEditMan: An XML Editor for Manuscript descriptions. Scripta & ***e-***Scripta, Vol. 2 (2004), Sofia, "Boyan Penev" Publishing Center, pp. 79-95.

[2] P. Pavlov, XEDITMAN: A XML Editor for Manuscript Descriptions and its Implementation for Cataloguing of Bulgarian Manuscripts. Review of the National Digitisation Centre of Serbia and Montenegro, Belgrade, Vol. 5 (2004), pp. 49-58.

Digging into the Future of the Past

Zoran Perisic
Mathematical Institute SASA, Belgrade, Serbia and Montenegro

In general, digitalization assumes search and retrieval, in digit form, historic, cultural and scientific heritage and thus provides a huge number of digital data. But what about an enormous quantity of data produced nowadays within complex, long-term research in some natural sciences and mathematics. Beside problems of storage such a number of data, another one is how to transfer them, in a very short time, to the distant labs and computational resources throughout the scientific world. Maybe, the new kind of scientific web, the GRID is a solution. But, management of the Past in these present and future digital times either, and of the Future itself, probably demands a next, new form of computing: quantum computers.

Digital culture practice

(Presentation of 1. CD-ROM Macedonian Antiquities, 2. CD-ROM Ohrid – The Craddle of Beauty, 3. CD-ROM Stoby- 3 dimensional reconstruction)

Irena Petrovska
EIN-SOF, Skopje, Macedonia

Keywords: *digitization of cultural heritage, digital rediscovery of culture, digital culture practice*

Descriptiion of the Major Theme

This paper presents 3 projects in the field of digital culture CD-ROM MACEDONIAN ANTIQUITIES, CD-ROM OHRID – THE CRADDLE OF BEAUTY, CD-ROM 3D STOBY, published or to be published by EIN-SOF and Alef. EIN-SOF is company which has introduced the multimedia in the field of cultural heritage and in general in Macedonia, which has provoked the establishment of the market for cultural multimedia products in Macedonia, which has benchmarked the multimedia production, and which has introduced certified multimedia production by its ISO 9001 certification.

As such, and placed in the scientific context of the Conference, this paper tends to contribute in examining the relationship between the theorethical and practical aspects of the digital re-discovery of the culture.

1. CD-ROM Macedonian Antiquities

... this project has had a number of public promotions and it has caused apparent attention and praise everywhere (Barcelona, Seattle, Washington, Lisbon), so as to win the prestigious THE BEST DIGITAL MEDIA AWARD, in Gifu, Japan, in November 1998, at the 4th International Scientific Conference on Multimedia and virtual reality.

The topic of the project is the cultural history of Macedonia in the period from the Paleolit untill and including, the Middle ages. Project focuses on the main historical events, the change of populations, cultures and civilisations, the flows of the religious and spiritual movements and within it – also on the material culture and arts as their manifestation. The whole content has been structured as a sophistically organized database in different media:

2000 pages Written text

Spoken text

Originally composed Music and sound effects

Drawings (hundreds of maps, chronological tables, grafics, technical drawings…)

1500Photos

10 Short film and video sequences

20 Computer animations

25 Spherical presentations of archeological sites and other cultural monuments

The data, on the one side can be treated as one whole, but on the other side it is the sum of independent elements adapted for a simple and accessible search, which can be organized as different entireties on both general and professional level, in accordance with the users needs.

The general level includes basic information on the cultural history of Macedonia aimed at wide range of users, and the professional level offers in-depth information provided as a result of close cooperation with experts form the University and the compatible public institutions.

This project was pioneer in introducing the 3d computer animation in the rediscovery of the cultural heritage in Macedonia. Among the 3d reconstructions of architect objects included in the project, are the following:

- house from the period of neolit
- several objects from the roman epoch (sqares, public buildings, basilics from stobi, Heraclea and other antient cities)
- middle ages sacral objects (the first church of st. Kliment, authentic profile of Samoil's cathedral on the St. Ahil island, exterior and enterier of St. sofija in Ohrid etc)

Last, but not the least, It is important to underline, that apart from the educational and scientific level of the cd-rom, it provides the users with unique aestehitic experience resulting from the quality of the visual material, the choice and the organization of the presented audio, video, text materials (language, photography, drawings, animations, sequences, texts, music, stile, text composition, pronauncination, fonts).

2. CD ROM "Ohrid – cradle of beauty"

This project has been awarded third prise on the SIPPO best of trado 2005 contest

The CD ROM, " Ohrid – cradle of beauty " is focused on emphasizing the fact that the Lake Ohrid and it's close surroundings is an integral natural, cultural and historical property, one of the very few such regions in the world. Namely, Ohrid region is one of the five site-regions inscribed on the World Heritage List, both according to its' Natural Environment and to its' Cultural

and Historical Aspect at the same time. The CD-ROM affirms two key aspects of its existence: Ohrid as a rare natural phenomena and one of the oldest lakes in the world, on the one side and on the other - Ohrid as an area with cultural living that could be followed since the prehistory, whose quality is certified with the different finds of the material and spiritual culture and with the richness of the cultural monuments of sacral and public character. These two elements equally and simultaneously participate in the profilation of Ohrid's remarkable image, and so they are treated as complement factors in this cd-rom.

Written and oral text, both in macedonian and in english, photos, video scenes, spherical presentations matched with the authentic and original music, all are combined in harmony, to elevate the spectator to a participant in the shown acts.

The Active participation of the experts in the field of cultural and natural heritage in the project realisation provided the Story of Ohrid with the scientific relevancy

The ultimate goal of this project is to encourage the Ohrid and Macedonian tradition and rich spiritual culture to penetrate within the global cultural flows in order to promote and foster the concept of cultural tourism, necessary for the quality promotion of Macedonia in the world.

3. CD-ROM "Stoby- three-dimensional reconstruction"

This project is three-dimensional multimedia revitalization of the urban tradition of the biggest city from the roman times in the north of Macedonnia.

The teams from Alef and eIn-SOF again in collaboration with proffessors from the Faculty of Philosophy – art historins and archeologists, artists and other experts, once again tend to make modern, quality and complex contribution for the affirmation of the Macedonian culture from the ancient times and as well as for the digital preservation of the globally relevant cultural heritage.

In short, among the strengths of this bouth scientifically sustained and visually attractive project we should include

- the experts knowledge on all levels – conceptual, content, technical, aestethical
- virtual walk through Stoby, complex authentic revitalization of the city of stoby, by the implementation of the latest trends in the information technology (integrated media – text, photos, video, audio, spherical presentations, original music, 3d reconstruction of the city)
- bilinguality of the content (Macedonian and English)
- readability of the content for the different target groups

Cultural Heritage Digitization, Archiving, and Dissemination Technologies

Vladimir Radevski[*] and Mathew Palakal[+]
[*] South East European University, Tetovo, Macedonia
[+]Indiana University/Purdue University , USA

Preserving the heritage of ancient civilizations furthers the education and understanding of world cultures that may help to promote world peace. Without digital preservation of important cultural sites, future generations may never know the achievements of past civilizations — our heritage. As pollution, wars, and natural disasters take their toll on these ancient structures, it is increasingly important to accomplish preservation in a timely manner and make the archived information available to those who follow. Current efforts in preservation focus on digitization of documents, artifacts, or buildings and structures located in the midst of modern civilization. These projects overlook the cultural heritage sites that lie hidden beneath the modern world on the brink of extinction. Innovative information technologies need to be developed to systematically collect, archive, and disseminate these heritage treasures for cultural understanding, education, and learning.

1. Digital Documentation Efforts

Several projects currently exist that employ digital documentation of heritage sites. Large to medium efforts include: Mont Saint-Michel [2], the Foundation of the Hellenic World (FHW) [3], and the virtual walkthrough of Fatehpur Sikri [4]. Re-creations of cultural heritage sites are the most common usage of creative digital technology. Examples include the Virtual Tower of London [1]. Projects such as the Digital Michelangelo utilizes 3D laser range finder technology to provide a reliable and accurate digital representation of form and surface texture [5]. The Rome Reborn project developed by UCLA, an interactive recreation of Augustan Rome, allows educators and students to experience a lost environment in all its complexity [6]. The virtual tour of Notre Dame [7] and the Great Buddha Project[8] offers excellent examples of sophisticated virtual tour project.

Online library collections populate web sites worldwide. The Hermitage library collection is an example of a site that involves learning through interactive activities [9]. The site challenges a visitor to find a Gauguin

masterpiece simply by recalling the organization of his subjects or locating a Da Vinci painting by searching for its predominant color vocabulary. IBM's experimental Query By Image Content (QBIC) search technology offers online collections to search for artwork using an artist tool kit [10]. The California Digital Library project provides access to biographies, bibliographies, and critical analyses of authors from any literary discipline [11].

Museums such as the Canadian Museum of Civilization, the Newseum in Arlington Virginia, and Australia's Melbourne Museum have pioneered applications of technology to create a "museum without walls". The Taiwan National Palace Museum housing traditional and historical Chinese arts, Painting, Calligraphy, Ceramics, Jades [12], the Andrew Dickson White Collection of Architectural Photographs presented as digital images and text description [13]; the New York Public Library Digital Collections of Performing Arts in America 1875-1923 [14], the Art Institute of Chicago Collections of Digital images and text description [15], Archnet's Digital Library of Architectural Collections of Images and text [16], and the Asian Historical Architecture of Images [17] are good examples of Museum collaborations in preservation efforts.

2. Data Collection Technologies

Cultural heritage sites normally consist of monument landmarks, gardens and architectures, artifacts, etc. Visual data representation, 2D digital images, video, and 3D models are usually captured for heritage site documentation. Examples of successful representation include *Digital images* [18] for the preservation of ancient fresco in Mogao Cave and restoration of Fresco [19]. Likewise, scanning technology has become prevalent in data representation. The application of *3D Models* [20-26], used lasers to measure 3D shapes of unearthed sculptures in the Museum of Terra-cotta Warriors and Horses in China. The Stanford Digital Michelangelo Project, which scans fine surfaces of the 3D model of Michelangelo's [27], and the measuring of a huge size Buddha using various types of laser [28] are excellent examples of this technology.

Prior efforts related to 3D visual reproduction include graphical modeling, range data scanning, and image-based approach. In graphical modeling, detailed 3D geometric models are manually constructed for individual objects in the scene and represented in a 3D graphic language such as OpenGL and VRML. Range data acquisition using laser or light scanner technology has been widely used to generate high-resolution surface

details of physical objects, such as statues and archaeological artifacts [30]. An image-based approach is used to generate novel views from a set of existing images of a scene [31]. Several variations of image-based techniques exist. The simplest and most commonly used technique is panoramic viewing, such as QuickTime VR and IPIX images [32]. Panoramic viewing presents a field of view from a single viewing position, but is essentially a 2D technique with limited viewing range. The fundamental idea is to use image information to reconstruct a plenoptic function. Since the construction of a complete plenoptic function (5-dimensional) is impractical [33], most methods (*e.g.*light field and lumigraph [34,35]), focus on subsets of the plenoptic function for smaller problem domains, simple scenes and limited view ranges. An extension to lumigraph rendering is also given in a recent work by Buehler *et al* [36]. Another related work is by Aliaga and Carlbom [37] whose method can handle more complex indoor environment than light field and lumigraph methods. However, the requirement for unobstructed space leaves little room for large scale and more complex environments.

3. Application Development

QuickTime VR tours, Streaming Media Technologies, Graphical Recreation of the ruins, Virtual Reality and Immersive Technology, and Augmented Reality of the ruins are just a few examples of applications where one can utilize the preserved data. Some details about these applications are summarized below:

QuickTime VR: This mode of application allows an interactive CD/DVD-based virtual tour. The end-user will have access to the functionality to take a VR enhanced tour through a ruin, examine artifacts in great detail, listen to a narration, see real video, and listen to authentic audio and/or music.

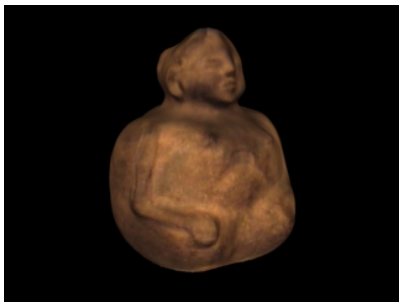
Streaming Media: The applications developed using streaming media will allow virtual tours with multimedia information synchronized with video according to users' needs. This technology can provide better quality video at the expense of higher bandwidth. This method would be an alternate to route panorama, if the user needs only a small segment of video.

Graphical Recreation: At every historical site, natural and manmade calamities have caused damage to structures that are still standing. In this research application, we will graphically recreate the missing parts of the structures and superimpose these phantom treasures on the remaining structures. Once this is done, one will experience a tour of the ruins either as they were created thousands of years ago or as they stand today. Using the

latest laser technology, 3-dimensional video technology can be used for graphical recreation of objects and structures.

Virtual Reality: Applications developed using this technology will allow a user to be immersed in the environment. Users will use special facilities such as CAVE or ImmersaDesk. Technology to support this is currently being refined; however, such a tour, augmented with synchronized music and audio, will provide the user with an unforgettable experience of "being there." The spherical image data will be effectively utilized to create virtual tour of the heritage sites. Fully immersive applications will run on a CAVE or Idesk.

Virtual Gallery of Artifacts: Laser scanners can be used to collect 3D range data of artifacts from the heritage sites. A number of techniques have been developed to reconstruct polygon meshes from dense range data. However, a point-sampled object can produce an extremely large number of polygons, interactivity is difficult to achieve. These methods use adaptive point splatting, as used in the Michelangelo project, or by some combination of textured surface element splatting and Z_buffer techniques. The CLIOH team has proposed to develop a unified solution for the interactive rendering of 3D scenes containing point-sampled objects of different scales as well as 3D synthetic models. This will help to develop and implement haptic feedback technologies to "touch and feel" the 3D artifacts in a virtual museum setting. Users will be able to create virtual galleries using the rendered 3D images of the artifacts. An application tool to create such a virtual gallery will be developed



3D laser scan of an Angel Mound Artifact

4. The CLIOH Project

The purpose of project CLIOH (Cultural Library Indexing Our Heritage) was to provide an online, cultural digital library of indexed and archived digital multimedia, documenting global heritage treasures to users worldwide. Project CLIOH provides a service to a large audience, including, K-12, Higher Ed, Government agencies, Travel industry, Museums and Libraries.

The project assumes through the UNESCO organization's listing of threatened sites that there is a need for information regarding cultural heritage. This research has documented the fact that many of the world's cultural treasures are in danger of being lost or destroyed, causing

irreparable damage to human understanding of manmade treasures and the cultures they represent. Although many ancient sites are photographed by tourists yearly and documented by professionals for the Discovery Channel and National Geographic, yet, little has been done to digitally index and archive it. The goal of CLIOH within the project time-line was to develop technology to digitally archive heritage sites by creating a globally accessible, state-of-the-art digital library and multimedia data bank, using the latest data collecting technology. The data collected are compliant with museum and library approved standards such as the Dublin Core and the Open Archive Initiative.

The CLIOH System: The design of the CLIOH system is now complete and various components of the system were tested. All the Software components used for CLIOH are a combination of open-source and proprietary software tools. Proprietary tools include professional graphics development such as Director from Macromedia. Real Media Server, a proprietary streaming media platform is also used.

Many of CLIOH's applications are open source. These tools, such as Apache server and MySQL database are not only free to use, but also are well documented and supported. One side benefit of open source software is that since it is developed by volunteer effort, the developers often share a mindset that software should not only be freely available but should interact well with other software. Another advantage of open source software is that it is often designed to be customizable by the user; often this process leads to improvements in the software.

The CLIOH Web server: The default web server for CLIOH is the Apache server; it provides the first layer of interaction with the web client by displaying the introductory web pages that precede the inner processes. This page also provide navigation into CLIOH's functions, and will includes a quick search text box to accommodate those who want to begin immediately.

The Search interface: CLIOH's main search interface is built using the advanced graphical capabilities of Macromedia's Shockwave technology. Shockwave was chosen because of its ability to handle still images, audio, video, and virtual reality media such as QuickTime VR and Shockwave3D.

Application Server: In order to implement searches from the web interface, an application server is needed to handle the specialized content, access the database, and return the set of results back to the interface in the proper format. We chose to use Zope as our application server because it combines several useful functions into one software package. On one level,

Zope is a fundamentally new and different environment in which to build and manage websites. On another level it is an application server that simplifies many functions, including interaction with databases.

The CLIOH Database*: Multimedia files, like text documents, need metadata in order for them to be available to users. In the CLIOH system, each file is associated with an application for the Dublin Core metadata element set. This data is then stored inside a database, which can be used to correlate files and provide meaningful search results for the user. CLIOH's database is implemented in MySQL. MySQL is the world's most popular Open Source Database, designed for speed, power and precision in mission critical, heavy load use. MySQL internally uses B-trees structure to index records stored in relational database tables, hence providing an efficient and fast means of retrieving the records from the database.*

The CLIOH Data: At present, CLIOH has several hours of video and hundreds of still images, collected from the ancient Mayan cities of Chichen Itsa and Uxmal, both located in the Yucatan Peninsula in Mexico. Each is being indexed by the Dublin Core metadata descriptors, which are organized into database tables and fields.

In addition to this process, raw files are turned into deliverable products. This step is different for each type of file. Images that were shot on location at the sites need to be transformed into usable file formats. Many of the image files are either stored as photographic slides or recorded on CD using a Kodak proprietary image format; these are converted into the JPEG format for use on the website. In addition, in order to maintain proper archiving practices, a master file is made of each image using the TIFF format. To assist users who may wish to access these master files, we are also making them available in a compressed TIF format, which will speed their download considerably. Each still image used for display must be provided in three sizes: one for the standard display, a larger version which users can choose to view, and a small thumbnail size used as a navigational link to the main image.

***Database design**: The CLIOH team has worked towards building an efficient and elegant database, which is scalable and extensible to future developments or changes. The team has decided to use a relational database that will help in indexing, storing and effectively retrieving multimedia items from the database. The metadata generated to describe each of the items in the digital collection is stored in the databases and lends itself to context – based information retrieval of these items from the database. The CLIOH database indexes these items based on the metadata used to describe them.*

The structure of the database has been designed such that all the records are first classified according to the archeological site they belong to and then secondly, they are again divided according to which topics they represent. These topics are related to each of these archeological sites. The database also uses the Library of Congress vocabulary to aggregate the resources into comprehensive collections.

The database is built on MySQL, which internally indexes all the records in a B-tree format. This aids in the fast retrieval of records when the database is queried. The database can support queries based on keywords, subjects under Library of Congress, archeological sites, Dublin Core elements such as publisher, creator, contributor, date, title, type, format and coverage. The database has been built to be flexible. Hence at any time a table can be added or deleted without having to change the entire schema of the database.

The metadata plan was built around the fifteen Dublin Core metadata elements. The database was in turn designed around the metadata. Applying the Dublin Core standard to multimedia items presents unique challenges that require careful consideration. It is necessary to provide metadata for each individual file and also to maintain metadata that allows different file types, such as a video and its corresponding audio, to be grouped together. DC elements can take on different properties when they are mapped onto different types of data. For example, image files are likely to be many in number, and belong to a logical set. In the case of the photos from Uxmal, the entire group might be considered a single set. Content areas based on the individual architectural work that is their subject further subdivide most of the photos in this set.

5. The Macedonia Digital Museum Pilot study

The Virtual Museums in Cultural Heritage of Macedonia would provide visitors with high quality, vivid, virtual experience that promotes user interest, about the subject of cultural heritage. It will encourage further curiosity and sustain interest that leads to increased attendance at the

Museum and it will bring virtual exhibitions to users who would not be able to visit physical location due to physical limitations, geographical constraints, time or budget. It will expand the museums viewable collections to include augmented information and superimposed media rich content available, anytime.

In addition, the research team will work on determining the unique signature of the artifacts in the collection. While some watermark technique will be implemented in the digital representation of the artifacts, a research will be directed towards assuring the set of elements to determine the artifacts in a unique way - an issue extremely important and actual from the security and protection point of view.

This initiative will provide the Museum of Macedonia with the opportunity to digitize, catalogue and create a 3D representational virtual museum and indexed archive of their collections thus allowing the global users with a highly engaging interactive experience of the museums collection of cultural heritage. Finally, this project will establish an international collaboration among scholars, students, and museum experts to engage and exchange ideas on digital documentation of cultural heritage.

Strategic plan: The development of the virtual museum and the digitization of the Museum Collections involved a multi-phase operation. These phases are divided into: A) Pre-Production/Design; B) Production/Design; and C) Post-Production/Design. In addition to these specific phases, there will be ongoing regular discussions to keep each of these phases on target.

Pre-Production/Design Development in this phase will include discussions by the members of the MK-Team (the Team in Macedonia) and the IUPUI-Team (jointly called the Tech-Team) with expertise in content development database design, 3D scanning and digital photography, interface design, graphic design and 3D animation. This is the initial phase of the project where the Museum's collection is first assigned a keyword identifier (ID) code, then digitally scanned using a Minolta 3D laser scanner and photographed for detail. The Museum's curatorial staff will work with the Tech-Team to ensure that the entire collection is photographed and scanned without damage or consequence. The IUPUI- team has experience with digital documentation of cultural heritage sites and artifacts. This team also has extensive experience in designing, implementing, and bringing to fruition a project of this scope. Some of the advanced techniques previously utilized that have proven to be successful across different applications will be adapted and refined for handling data storage and retrieval in this virtual museum project.

Production/Design This phase involves the continued efforts of the experts, who will be assisted by students from both MK-team and IUPUI-team. This team will build the database and populate it with the ID coded metadata. The IUPUI-team has developed an extensive metadata retrieval system based on the Dublin Core Initiatives and expanded the search features to be more unilateral in compliance with the Dublin Core extracted identifiers such as: year, composition, purpose, and object recognition features (e.g., color, texture, and shape) will be used for identifying and cataloguing. The MK-team will ensure optimal content accuracy and will be responsible for assigning all significant information relevant to each artifact. The MK-team will also provide cross-lingual translation for all the descriptions and metadata. A user will be able to query a searchable database in the digital collection or select the artifact from the virtual museum. The museum's interface design will allow the user to navigate through the exhibition at will and manipulate the artifact in 3D space. As end users traverse and seek information in the virtual environment, different types of data must be quickly retrieved and assembled for presentation. This is a challenging problem because the data is in diverse formats: text, image, 3D, and demonstrative animations. Cutting-edge multimedia data indexing will be used to for efficient storage and retrieval of multi-format data.

Post-Production/Design The final phase of the operation involves the tasks of testing, making adjustments, re-testing and finally implementing and making the website operational. The digital archive created from the Macedonia Museum Collection and the Virtual Museum will be stored in the massive database system that will be physically located at a place that is agreeable to the stakeholders in Macedonia. As a virtual museum experience, it is intended to be multi-tiered, user directed, and information driven. The end user will come away from the website with a complete experience of "cinematic interactivity."

LINA (University of Nantes) - According to its research areas, the LINA (University of Nantes) (through its two members G. RASCHIA and F. TRICHET) is particularly interested to participate to the project "VIRTUAL MUSEUMS OF CULTURAL HERITAGE IN MACEDONIA". More precisely, LINA works on the two following issues.

a) Development of an ontology-based system dedicated to the indexation, the navigation and searching of Macedonian gallery and museum collections.

The main objective of this work is to define an ontology of the art domains concerned with the Cultural Heritage in Macedonia, *i.e.* a semantic layer composed of a semantic network of concepts and their relationships.

Benefits of such an ontology-based approach are: (1) to facilitate browsing, navigating and retrieving multimedia resources by using art concepts linked in a network in addition to the classical searching by metadata, (2) to facilitate enhanced content based searching, (3) to allow other systems to access the Macedonian system using Semantic Web technology (*i.e.* a software agent visiting the system should be able to understand the structure of its knowledge base via the ontology and hence retrieve information automatically and autonomously) or (4) to facilitate the adaptation/personalisation of the navigation in the Virtual Museum.

b) Building short views of Macedonian gallery and museum collections.

The main goal of this topic is to provide user-friendly summaries of the Macedonian gallery and museum collections at different levels of abstraction. Summaries are described thanks to both metadata and content analysis of objects and they represent homogeneous sub-parts of collections. The contribution to the whole project is multiple: i) it provides short descriptions of groups of objects, ii) it supports personalized paths in a virtual museum, with regard to user-friendly criteria (semantic content of objects), iii) it offers a multidimensional index over the data which can be used to originally and efficiently browse and access objects of the all collections and iv) it supports flexible querying due to the fuzzy set-based approach we propose.

6. Acknowledgements

The CLIOH project was supported in part by a grant from the Institute of Museum and Library Services, IMLS Grant # NR-10014. The Macedonia Digital Museum Pilot study was funded in part by a grant from the International Development Fund at Indiana University Purdue University, Indianapolis, Indiana, USA.

References

1. T. O'Neill, Saving Places, *National Geographic*, pp. 58-73, October 2002.
2. J. Bessonnet, F. Daubie, A. Pilpre, T. Ojike. Virtual Mont Saint-Michel. *Proceedings of the VSMM2000 6th International Conference on Virtual Systems and Multimedia ,* Ogaki, Gifu, Japan*, October 3-6, 2000, 156-163*

3. Gaitatzes, D. Christopoulos, A. Voulgari; M Roussou. Hellenic Cultural Heritage through Immersive Virtual Archeology. *Proceedings of the VSMM2000 6th International Conference on Virtual Systems and Multimedia,* Ogaki, Gifu, Japan, *October 3-6, 2000, 57-63.*
4. S. Mudar, V. Punjabi. *Fatehpur Sikri- An Epic in Red Sandstone*. 1998. http://rohini.ncst.ernet.in/fatehpur/
5. http://graphics.stanford.edu/projects/mich/
6. http://www.aud.ucla.edu/~favro/rome-reborn/
7. Virtual Archeology, Great Discoveries brought to life through virtual reality, Colin Renew, Thames and Hudson, 1997 p9
8. http://www.vrndproject.com/
9. VSMM, Ohmsha press,Professor Takeo Oijika Pg 138
10. State Hermitage Museum (Russia)
11. http://www.hermitagemuseum.org/fcgi-bin/db2www/qbicSearch.mac/qbic?selLang=English
12. http://wwwqbic.almaden.ibm.com/http://www.npm.gov.tw/
13. http://rmc.library.cornell.edu/adw/adw.asp
14. http://dlc.nypl.org/lpa/nypl/lpa_home2.html
15. http://www.artic.edu/aic/collections/index.html
16. http://archnet.org/library/
17. http://www.architectureasia.com/CHINA/buildings/oriental.htm
18. M. Barni, F. Bartolini, V. Cappellini, Image processing for virtual restoration of artworks, IEEE Multimedia April-June, 2000, pp. 34-37.
19. J. Y. Zheng, S. Tsuji, Generating Dynamic Projection Images for Scene Representation and Understanding, Computer Vision and Image Understanding, Academic Press, Vol. 72, No. 3, December, pp. 237-256, 1998.
20. J. Y. Zheng et al. Digitizing and virtual recovering excavated relics in a world heritage site, International Conference on Virtual Systems and Multimedia 98, Vol. 2, pp. 654-660, 1998.
21. J. Y. Zheng et. al, Digital Archiving of an Archaeological Excavation Site for Multimedia Display, 14th International Conference on Pattern Recognition, Vol. 2, 1492-1496, 1998.
22. J. Y. Zheng et al, Virtual Recovery of Excavated Archaeological Finds, IEEE Multimedia Systems Conference 98, pp. 348-357,1998.
23. J. Y. Zheng et al, Virtual Recovery of Excavated Relics, IEEE Computer Graphics and Application, May, pp. 6-11, 1999.
24. J. Y. Zheng, Virtual Recovery and Exhibition of Heritage, IEEE Multimedia, April-June, pp. 7-10, 2000.

25. J. Y. Zheng, A flexible laser range sensor based on spatial temporal analysis, 15th International Conference on Pattern Recognition, Vol. 4, pp.749-743, 2000.
26. Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis, Jeremy Ginsberg, Jonathan Shade, and Duane Fulk, The Digital Michelangelo Project: 3D scanning of large statues, SIGGRAPH 2000.
27. Daisuke Miyazaki, Takeshi Ooishi, Taku Nishikawa, Ryusuke Sagawa, Ko Nishino, Takashi Tomomatsu, Yutaka Takase, Katsushi Ikeuchi, "The Great Buddha Project: Modelling Cultural Heritage through Observation," Proc. of the 6th International Conference on Virtual Systems and MultiMedia (VSMM 2000), pp.138-145, 2000.
28. Alonzo C. Addison, Emerging trends in virtual heritage, IEEE Multimedia, April-June 2000, pp. 22-30.
29. M. Levoy, et al, "The Digital Michelangelo Project: 3D Scanning of Large Status," *Proceedings of SIGGRAPH'00*, pp. 131–144, New Orleans, July, 2000.
30. S. Kang, "A survey of image-based rendering techniques," in *VideoMetrics, SPIE Vol 3641,* pp. 2–16, 1999.
31. S. C. Chen, "QuickTime VR – An Image-Based Approach to Virtual Environment Navigation," *Proceedings of SIGGRAPH'95*, pp. 29–38, Los Angeles, CA, August, 1995.
32. L. McMillan and G. Bishop, "Plenoptic Modeling: An Image-Based Rendering System," *Proceedings of SIGGRAPH'95*, pp. 39–46, Los Angeles, CA, August, 1995.
33. M. Levoy and P. Hanrahan, "Light Field Rendering," *Proceedings of SIGGRAPH'96*, pp. 31–42, New Orleans, August, 1996.
34. S. J. Gortler, R. Grzeszczuk, R. Szeliski and M. Cohen, "The Limigraph," *Proceedings of SIGGRAPH'96*, pp. 43–54, New Orleans, August, 1996.
35. C. Buehler, M. Bosse, L. McMillan, S. Gortler and M. Cohen, "Unstructured Lumigraph Rendering," *Proceedings of SIGGRAPH'01*, pp. 425–432, Los Angeles, CA, August, 2001.
36. D. G. Aliaga and I. carlbom, "Plenoptic Stitching: A Scalable Method for Reconstructing 3D Interactive Walkthroughs," *Proceedings of SIGGRAPH'01*, pp. 443–450, Los Angeles, CA, August, 2001.

A New British-Bulgarian Initiative for e-Access and Preservation of Cyrillic Manuscripts

Dilyana Radoslavova
Institute of Literature Bulgarian Academy of Sciences
Sofia, Bulgaria

Keywords: *Cyrillic manuscripts, XML encoding, digitization, searchable catalogues, e-Repertorium Initiative*

Machine-Readable Description and Searchable Catalogues of Cyrillic Manuscripts is a three-year research project (2003-2006) within the frames of the exchange agreement between the British Academy and the Bulgarian Academy of Sciences (BAS). The institutions participating in the collaborative initiative are the British Library (BL), the University of Portsmouth, the Institute of Literature to the BAS (IL), and the Central Library of BAS.[i]

Background

Over the past decade British and Bulgarian specialists have been involved in a fruitful joint work in the field of creating electronic resources for Slavonic written heritage, coordinated by the Commission for Computer Processing of Medieval Slavonic Manuscripts and Early Printed Books to the International Congress of Slavists. Among the initiatives undertaken or accomplished with the participation of specialists from both the countries should be mentioned the project between the IL and the Mediaeval Department of Central European University *Computer-supported Processing of Medieval Manuscripts and Early Printed Books* (1996-1998)[ii] and the project on *Electronic Description of Early printed Cyrillic Books and Rare Materials*[iii]. They were based on implementation of Standard Generalized Mark-up Language (SGML, ISO:8879) and TEI[iv]-conformant DTD's for electronic records of medieval Slavonic material.

Nowadays a worldwide standard for comprehensive electronic descriptions of mediaeval manuscripts, on the whole, is still a task in perspective as the main tendency is to move towards application of Extensible Mark-up Language (XML)[v]. In fact, norms for electronic descriptions of mediaeval manuscripts are emerging from current work within the frames of variety of projects, like MASTER[vi], Digital Scriptorium[vii], etc. Such norms have being applied both at the BL and in the

field of electronic cataloguing in Bulgaria, too. At the BL, *Manuscripts Online Catalogue* (MOLCAT; http://molcat.bl.uk) has brought the information existing in the published manuscript catalogues onto the World Wide Web, while the ongoing *Survey of Illuminated Manuscripts* (http://prodigy/illcat/main.htm) has provided experience of compiling new electronic descriptions.[viii] These data-resources, both Access-based, are to be further enhanced and re-set up according to a standardised access and searching strategy that could ensure their inclusion into the new XML-formatted BL Integrated Library System.

In Bulgaria, since 1995 a group of specialists from various institutions, based at the IL, have been involved in *e-Repertorium* initiative[ix] on creation of a standard for electronic cataloguing and processing of Slavonic manuscripts by the use of mark-up languages. So far over 300 manuscripts have been described within the frames of a few projects. They were created by application of TEI-conformant SGML-DTD, convertible to the format of the recently developed XML Repertorium DTD for electronic description of Mediaeval Slavonic manuscripts.[x]

Aims and Scope of the Project

The new project is aiming on creation of searchable electronic resources based on XML mark-up of a large body of Cyrillic manuscripts in Bulgarian and British repositories. The use of XML in accordance with well-established procedures will ensure the multiple use, portability and long term preservation of the records and the ultimate conformity with the already existing data-bases and the international standards when these are eventually achieved.

In its initial stage the project concentrates on encoding of the BL Slavonic manuscript holdings, which consist of seventy two 14th-19th century parchment and paper books and roles[xi]. The electronic description has being conducted by researchers from the IL, using the XML Repertorium DTD. A work in progress is also the capture of selected digital images of each item of the collection executed at the BL. Simultaneously, microforms of the whole items have been digitised at the Central Library of BAS.

The final goal of the project is a creation of publicly available resources providing highly structured, searchable, and accurate descriptions, supplementing the existing printed catalogues and supplied with at least a digital image, or selection of digital images, and up-to-date bibliography. The digitization of the microphorms aims on both, a long-term preservation

of the manuscript surrogates and on creation of electronic copies of the items providing access to the whole manuscript text.

Thus the aims are consonant with those of the *Survey* and the information stored in the records, which are to be created in the lifetime of the British-Bulgarian project, may be extracted and used for its further extension, as now it contains approximately twenty BL Slavonic manuscripts only. On the other hand, the records are to be used for the further improvement of the BL MOLCAT, which now relays on description files with a minimal structure. It is intended that a new searchable catalogue of the BL Slavonic collection should be maid available on CD-ROM in the Manuscript reading room before the final integration of the *Survey* and the enhanced MOLCAT into the new Library System. Another outcome of the project is the inclusion of the records in the *e-Repertorium* information bank and the refinement of the searching system so that to provide a user interface enabling web-based searching at BAS. It is also foreseen that the electronic descriptions may be converted and added to the catalogue maintained by the MASTER project, or to other electronic catalogues.

The improvement of the record scheme, so that it could enable both a summary and detailed description, is seen as a very important task of the project, together with the development of tools for conversion to the schemes used in other projects. The further work in this perspective will be build upon some research already done both at the BL and in Bulgaria, and will proceed in consultations with groups engaged in similar work elsewhere. Liaison with external projects will be maintained via organizations as the TEI Consortium and the International Committee of Slavists with which members of the new initiative are involved. It is also intended that this aspect of the project will make a contribution to the more distant goal of setting-up standard for electronic description of medieval manuscripts.

---

[1] The project co-ordinators are Ass. Prof. Anisava Miltenova (IL) and Dr Magda Szkuta (BL).

[ii] The project coordinators were Prof. R. Cleminson, by the CEU side, and Ass. Prof. A. Miltenova by the IL side.

[iii] Directed by Prof. Ralph Cleminson, University of Portsmouth; the hard-copy of the electronic records created within the project see in *A Union Catalogue of Cyrillic Books printed before 1701 in British and Irish Collections*. Comp. R. Cleminson, Ch. Thomas, D. Radoslavova, A. Voznesenskij. London 2001.

[iv] TEI:Text Encoding Initiative Consortium; www.tei-c.org.

[v] www.w3.org/TR/REC-xml.

[vi] www.cordis.lu/libraries/en/projects/master.html.

[vii] http:sunsite.berkely.edu/Scriptorium.
[viii]For the history of the project see A.Prescott, M.Brown, R.Masters, "The Survey of Illuminated Manuscripts," *Towards the Digital Library,* London: The British Library, 1998, pp.130-147.
[ix] See http:clover.slavic.pitt.edu/~repertorium.
[x] See in particular *Computer Processing of Medieval Slavic Manuscripts: First International Conference, 24-28 July 1995, Blagoevgrad: Proceedings,* Sofia, 1995; *Medieval Slavic Manuscripts and SGML: problems and perspectives*, ed. by A. Miltenova and D.J. Birnbaum, Sofia, 2000; A. Bojadžiiev, A. Miltenova, D. Radoslavova, "A Unified Model for the Description of Medieval Manuscripts," *Computational Approaches to the Study of Early and Modern Slavic Languages and Texts*. Proceedings of the "Electronic Description and Edition of Slavic Sources" conference, 24-26 September 2002, Pomorie, Bulgaria. Sofia 2003, p. 113–135.
[xi] See the latest printed descriptions of British Library Cyrillic manuscripts in R. Cleminson, *A Union Catalogue of Cyrillic Manuscripts in British and Irish Collections*. London 1988.

Manuscripts, Card Files and All Those Dangerous Things

Kiril Ribarov
Faculty of Mathematics and Physics, Charles University, Prague

I will focus in this talk on problems relative to text processing of old manuscript sources, i.e. manuscripts, printed editions and card files. I will give attention to the content preservation of these sources, which can either be isolated processing of a single sources (manuscripts, excerption cards or similar), or the linguistic, editorial, historic or other type of language connected information structure of a set of (including all) items.

Without any loss of information a manuscript can be digitized by professional scanning. Usually, several scans are obtained including specific ones for the older and seriously damaged manuscripts. The manuscripts are organized in electronic catalogues where all necessary bibliographical information and other text descriptions are added. Libraries, e.g. the Manuscriptorium and Memoria projects of the National Library in Prague, have accustomed procedures and have developed electronic standards for cataloging of the manuscripts and of the old printed editions. The mission and the results are clear: wide accessibility of the electronic catalogues of manuscripts. Although there are still many problems to solve in this area, it may be agreed upon that solid basis are given and it is primarily a matter of financial input of how rich the catalogue items will be as, e.g., whether those items will be accompanied by their scanned images. Typically, the scanned images are indexed on page level and no image content is indexed.

Such a step brings us to a point when one may locate the existence of the manuscript, learn about its characteristics and see the digital image of the manuscript. Undoubtedly this is an inevitable starting point and primer aim of library and archive centers. Nevertheless, the first and foremost interest here is the question related to the next steps of content processing of the catalogued units.

These steps initiate as soon as one tries to type a single word in a text editor or acquire it from OCR software. Unfortunately, time and space do not permit to elaborate inevitable coding problems, which are still the main obstacle for efficient text exchange. I will therefore head towards annotation issues and processing of lexical units.

Processing of manuscript words is a complex task which includes knowledge of the following:

- recognition of words (problems of segmentation),
- rendering of abbreviations and word reconstructions (problems of rendering),
- recording of word variants (problems of variant and editorial recordings).

Frequently, this phase is accompanied by reconstruction of damaged parts which is often connected to series of comparative studies of various cross indexing with similar sources, translation equivalents, the Bible or other types of citation or time parallelisms.

The process of word identification may be approached by two models. The first one has the character as an elementary unit, while the second model has a string of words as elementary unit. Problems of character identification will be discussed and original and rendered word forms will be introduced.

Once a word is captured, I will demonstrate that building of annotated corpora of texts solves many needs of today's studies of the language aspects of the manuscripts. The talk will be accompanied by computer demonstrations of the Annotated Corpora of Text client and demonstration of the unique web portal of annotated Old-Church Slavonic manuscripts, which belong to the Macedonian redaction. One may access these data and tools at http://prometheus.ms.mff.cuni.cz/act.

As multiple variants of word forms need to be distinguished, each of them can be additionally enriched by information relevant to the word itself and pertinent to the context in which the word form is located. The following information is added to each word form (basic linguistic oriented annotation):

- lemma, expanded to keyword mark-up and mark-ups of redaction dependencies,
- morphology (positional and using a general scheme, which can be accustomed to the processing needs),
- multi-word mark-up and its annotation (extension towards syntax and discourse studies; mark-up of phrases, idioms, discourse elements, contexts or other hierarchical and non-hierarchical structures),
- cross-links with other sources (the Bible, other documents),
- translation (translation equivalents).

As annotation is highly time consuming process dependent on professional input, efforts in automation of these processes are applied. The presentation will highlight the application of natural language processing techniques as a solution to automatic and semi-automatic language processing.

Furthermore, one needs to consider the new problems arising from distributed and mass processing of manuscript sources. This opens additional questions relevant to coordination and synchronization of annotation activities and evaluation of the annotation work.

Annotated manuscript texts are excellent objects for profound linguistic studies. Annotation allows that manuscript searches are performed at any step, and outputs, as indexes of various types, are instantaneously generated. Under this framework, card files are solely another view on the manuscript content. It will be demonstrated that direct links between the card files and the manuscripts can be automatically obtained. Such links will significantly support the processing of the card files - a very important problem, i.e. the problem of preservation and accessibility of the content present in the millions of excerption cards.

In addition, this work will underline the necessary technological background and assistance needed for text processing and annotation; current technological obstacles will be as well named and discussed.

Automatic Composition of Ballet Sequences Using a 3D Motion Archive

Asako Soga[1], Bin Umino[2], Jeffrey Scott Longstaff[3]
[1] Ryukoku University, Japan, [2] Toyo University, Japan,
[3] Laban, City University London, U.K.

Keywords: *Dance, Motion Capture, 3D Motion Archive, Automatic Composition, Web*

Introduction

We have developed an automatic composing system for ballet horeographies by using 3DCG animation. Our goal is to develop some useful tools in dance education such as creation-supporting system for ballet teachers and self-study system for students. We discuss our system which is called 'Web3D Dance Composer' (http://www.rikou.ryukoku.ac.jp/~asako/wd/).

Web3D Dance Composer consists of an online archive and user-editable simulation system for ballet steps and step sequences, or '*enchaînement*.' An exhaustive archive of 3D ballet movements can be browsed, previewed, and selected to compose extended sequences. In addition the automatic omposing system allows ballet steps to be combined automatically according to algorithms which describe characteristic features of *enchaînement*. We have developed a system that allows users to select rules for the *enchaînement* algorithm, creating new algorithmic derivates, which can be used to generate new movement sequences for original choreographies for ballet.

Statement of Current System

In collaboration with expert ballet teachers a comprehensive list of 543 ballet steps was identified which comprises a basic vocabulary for women in 'center lesson' for ballet. These were classified into hierarchical categories and assigned with plain text codes. Data for these ballet steps have been motion captured and currently 215 of these have been included in the step archive of Web3D Dance Composer.

We acquired the ballet 3D motion data using motion capture systems. We used both optical and magnetic systems to offer greater possibilities in deriving the data standardization. The data were captured from the performance of professional female ballet dancers in Japan.

The raw data were transformed into manipulated data through several stages. 'Virtual Reality Modeling Language' (VRML) has currently been adopted as the standard data format. We created the original hierarchical skeletal system of a dancer's body which complies basically with 'VRML H-Anim standard' which has three levels of architecture; although we reconstructed it into a single level containing 16 joints and 20 segments. We translated the raw data to fit our skeleton, and then divided the data into three sections for reasons of system efficiency. Rotation data for each segment denotes limb motions, Y-axis data denotes vertical motions of the entire body, and XZ-axis data denotes horizontal motions.

Finally, the sequences of motions were segmented into short pieces so that each segment starts from and ends at one of the 84 basic ballet poses. These motion segments are used as pre-constructed components to create new movement sequences.

The simulation system of Web3D Dance Composer (Fig. 1) was programmed with Java language. Any user through the Internet can access the Java applet. We used 'External Authoring Interface' (EAI) to control the VRML world from the Java applet. The minimum system requirements are only Windows PC (Windows 95/98/Me/NT/2000/XP), Microsoft Internet Explorer, and a free plug-in VRML viewer, such as 'Cortona VRML Client' (http://www.parallelgraphics.com/products/cortona/).

Fig. 1. User interface of 'Web3D Dance Composer Ver. 4.0'

The user interface allows a variety of manipulations. The VRML viewer allows the viewpoint of the performing area to be rotated and panned. The 'Display control panel' allows manipulations of type of dancer, type of stage space, selection of music, number of dancers, and arrangement of dancers. The 'Step catalogue panel' allows selection of positions, steps, and *enchaînement*. The 'Motion control panel' allows selection of duration of steps and speed of performance. This panel also allows steps to be added to the composed sequence, rewound, played, paused, and manually moved forward or backwards with the horizontal scroll bar.

Using the 'Timeline panel', users can select ballet steps one by one, arrange them into a sequence, and combine all steps to make an entire composition. Using the 'Recording panel', users can save their original sequences in plain text code.

The automatic composing system of Web3D Dance Composer displays a selection of logical next steps so that the ending pose of the previous step becomes the beginning pose of the following step. These can be individually selected and added to the composed sequence. In addition, by using the 'Auto' button, the automatic composing system allows the creation of an entire 16 beat *enchaînement* according to an algorithm prescribing its structure.

The algorithm is integrated into the system to create utilitarian choreographies. The algorithm approaches the task of automatic creation of an *enchaînement* by reducing the solution to path selection on a 'directed graph' in which vertices represent the steps and edges represent the connectivity between steps. The system considers categories of steps, physical exertion required, allowable step repetitions, and selects one step from the logical alternative steps at an equal probability. Each subsequent composition by the same algorithm generates a new unique sequence.

We conducted an experiment on our proposed system to evaluate its utility. The results of the experiments were used to assess the efficacy of the *enchaînement* algorithm for ballet education, specifically, whether the algorithm automatically choreographed *enchaînements* appropriate for an elementary level ballet classes. As a result of an evaluation test, we verified that the created choreographies had a possibility to be used in the actual lessons.

Future Plan for Dance Composition

The system will be further developed allowing users to create new modern and unique variations of *enchaînement*. Possible movement sequences will then be automatically constructed from these*enchaînement* within the VRML.

A user interface will be created as part of the Web3D Dance Composer allowing users to select amongst various 'rules' or principles for creation of movement sequences. These rules have been originally derived from analyses of ballet *enchaînement* but will be segmented and generalized into individual factors, each containing a single proposition. New varieties of rules or principles will also be included which had not previously been part of the ballet *enchaînement* algorithm. This list of rules will be displayed in the user interface where many can be selected and combined together into the same algorithm, while others will be mutually exclusive. This will allow the user to create their own personalized, original algorithm governing the creation of new movement sequences.

The user-created algorithms can then be implemented by the automatic composing system of Web3D Dance Composer to automatically generate dance sequences according to the algorithm. Several sequences can be composed, each conforming to the algorithm, automatically generated from different steps selected from the step archive.

This system is valuable for online virtual dance experimentation and exploration by teachers and choreographers involved in creative practices, improvisation, creative movement, or dance composition.

Playing Digital - DrDC game

Kalina Sotirova
University of Dublin, Ireland and
IMI, Bulgarian Academcy of Sciences, Sofia, Bulgaria

Keywords*: edutainment, game, narrative*

Culture as Narrative

Cultural memory of the human society exists and is preserved/presented to the next generation in different languages of one narrative. Mythology, cosmology, art, social/moral systems, and especially games during the ages are in their essence --- storytelling: a story about the world, God, Self and the Other, a story told by stones, paints, toys, and many abstract, language driven, concepts. The Greeks told their story in fables; for the Romans the most powerful, i.e. everyday, barbarian language had to be used (Giambattisto Vico, *The New Science*, Cornell University Press, [1744] 1968). Educational aspect of these narratives necessarily presupposes particular moral system, attached to it.

(Digital) Image as Narrative

In short, in the Digital (re)-Discovery of Culture, the crucial role is given to the story told, to the message of a particular digitized "image". (Graham Clarke, *The Photograph*, Oxford University Press, 1977)

[REF2. We use the word "image" here in the sense of complex presentation of cultural artefact, i.e., digital image, the metadata attached to it, so that it could be searchable in the WWW and especially the Semantic Web, and in some cases text/description of the original object].

Since today everything can be digitized in a standardised way, such an image could be any single element of human cultural / scientific heritage: a mediaeval manuscript, a book, stone monument, art. Digitizing music and dance, and presenting them in WWW, requires more technology tools, in comparison with manuscript, for example.

I will give an example with the *Riverdance* concert I attended on 8th of June, in Dublin [REF3]. If I have to digitize the music, the dance, the lighting, all intercultural links, included in the performance as a whole, and the experience of the audience, filming by digital camera will NOT be enough. This was NOT an Irish performance. The same is valid for "Two

Worlds" music-dance performance of Neshka Robeva (Bulgaria). Both representing mixture nature of post-modern art and culture. They both are deriving their message(s) for the world of today from the roots of yesterday. [REF4] And these roots are cultural, i.e. based on textual/visual memories of the past, starting from ancient times and even before. But in the spreading of our story of the (culture of the) Past, we have to stay open for the Other stories and still to be/discover/live in the present.

Digital Games as Narrative. Why Games?

The digital way in which we present our culture should give enough information to fill all the "slots" in the "syntactic, semantic, and narrative frames" of thinking and perception process. (REF5 For more about frame-theory in human and artificial intelligence see: THINKING, Frame-Theory. Minsky, 343p., 355-377pp.)

Playing as main human activity seems to be good digital way to do this. Edutainment, and especially the DrDC game, can be regarded as an educational "tool" for presenting the culture of the Self and of the Other in serious, but playful way.

Why presenting cultural heritage through Edutainment (entertaining education)?

When I tried to clarify for myself the real goal/meaning of digitizaton of culture (in the frame of KT-DigiCULT-BG project [REF5. www.math.bas.bg/~digicult] and in general), I faced many sad examples of perfectly digitised cultural memory, presented in a manipulative way. This is always the case when "resurrected roots of the past" (see point 2) are used for narrow-focused, chaining purposes. Therefore interdisciplinary, and better, international team is needed to build good, educational example of digitised heritage.

The point I want to make by this is that digitization of culture could not be thought as merely scanning the artifact, attaching some scholarly text to it and putting them on a CD (and/or the WWW). It is intended to have a personal (re-)discovering nature --- personal (re-)Discovery of the Self and of the Other, by digitised textual and visual background of the roots resurrected.

Digitization should not be mummification of cultural artifacts. The concept of digital library/museum is great in principle, but will stay useless, if people are not personally interested in visiting it. The DrDC game emerged as a way to provoke such motivation, to create an interest for personal cultural discovery of the player. After a year's personal experience I can say that it is edutaining game, because its goal is to reveal hidden

meanings, to recall history and to educate on the gaming-board of the WWW.

DrDC Game (to be presented and played in the First SEEDI Conference, September 2005, Ohrid)

Digital Re-discovery of Culture through images, text etc., resulting at its physical experience is at the core of the initial concept of DrDC game. The playing is in fact capturing "images" (text, audio, video material in a size, that can be send by e-mail to the other player) from WWW board and creating a Challenge for the Other. The Other has to discover/answer to the challenge in a creative way. What happens meanwhile is real re-discovery of culture at educational play. This is what DrDC game is all about.
And one always plays as representative of his/her own culture.

APPENDIX A

DrDC game rules [to be further developed]
(edited by Micheal Mac an Airchinnigh and with comments added)

1] * TIME-PERIOD.
A game will last at most 7 days.

2] * WWW TOOL:
GOOGLE search engine

3] * MEDIUM:
The moves will be sent by E-MAIL only.

4] * ORDER OF PLAY.
Players follow one other in giving a key word (or key phrase).
In addition, one url may accompany a key word or phrase.

5] * CONDITION for acceptance/rejection of the key-word given by the partner [a dialogue at the initial level of playing]:
A short introduction (of up to 70 words). Its start and its end to be marked by < and > (e.g. by <quote> and </quote>). It is intended to be an explanation of the physical and psyhological environment for choosing the challenge given. The key words must have their ORIGIN in something interesting/challenging, which is seen/heard during the previous 7 days.

Introduction is NOT a hint, nor interpretation, but facts only, a short STORY.

6] * OPENNESS.
New players can be included at any stage of the process, depending on the GOAL the key-word-giver has in mind. For example, Kalina Sotirova can be included in a game that Michael O'Rahilly starts on Friday.

7] Game Transcript (OWL, Semantic WEB)
8] Formal acknowledgement of receipt of e-mail

APPENDIX B

DrDC Game Example (with notes)

AUTHOR: XYZ

DATE of sending: 2005-05-16.
DATE of receiving: [to be filled in the confirmation receipt by player]
DEADLINE: 2005-05-25

KEYWORD: song
Now the keyword has ontological significance. In this game we are looking for a very specific song as the end-goal.

URL: http://www.users.bigpond.com/garrysmith/location.htm
The web page shows a map which gives the location of a place in Australia called Botany Bay.

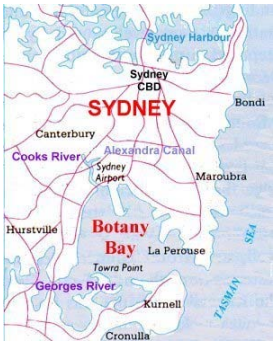


IMAGE:

This image of a Famine memorial is referred to in the text below.

Background Story

<quote>
I sit here all alone, in the Pagoda,
at the centre of the world[1]
like a Buddha[2].

I face in the direction of Sofia[3].
To my right is a poet[4]
and washing women[5].
In front of me is an Irish Volunteer and Dublin[6] Fusilier[7].
Behind me is a poet and playwright[8].
To my left are the blind[9].

Above me there are clouds
of many shades of white, none black.
There are wonderful shades of blue, all forms of black.

---

[1] Le centre du monde est partout.
http://www.leplaisirdapprendre.com/formaprof/stages/ateliers/tv5.php
[2] The Buddha observes.
[3] An exact physical location.
[4] James Clarence Mangan.
[5] The Magdalen Laundry.
[6] An exact physical location.
[7] Kettle.
[8] William Butler Yeats.
[9] A seating area for the Blind.

Beneath me are stones,
reddish brown to match the seat,
set in place by the Unknown Worker.
[Why isn't there a perpetual flame somewhere here
in the centre of the world in memory of him/her?]

In the North Eastern corner of the Green[10] there is a Famine memorial
[see image]
and it brings to mind a very famous song about the Famine.

I was very surprised to learn that this famous song
will be sung[11] in Istanbul on 25th May 2005
but with very different words.
</quote>

Short [up to 70 words] introduction. Its start and its end to be marked by <quote> and </quote>. It is an explanation of the physical and psychological environment for choosing the challenge given. The key words have to have their ORIGIN in something interesting/challenging, which is seen/heard during common work. Introduction is NOT a hint, nor interpretation, but facts only, short STORY.

GOAL: find the name and the music of the song about the Famine

[10] An exact physical location.
[11] By the Liverpool football team in the European Cup.

APOGRAF – Another "semper fidelis" resource to be integrated in the scholarly family of digital scriptoria - A project in the making -

Ştefania-Cecilia Ştefan
Institute for cultural memory, Bucharest, Romania

Keywords: *manuscripts, digital facsimilation, digital library, research*

1. Background

The digitization of old book collections has numerous expressions on the web, and the thouroughest ones also offer precious additional materials to facilitate deeper involvement in scientific study. Depending on the user's profile and interest, these tools are meant not only to provide a "*webccesible*" alternative to often difficult, time-consuming and preconditioned availability of rare imprints or manuscript books, but also to stand for a definitely special, self-sufficient and worthy by itself, work-environment (not to mention its role to preserve and thus prolongue the life of such naturally exposed and frail materials, by saving their identity through a confident, enduring medium in which these values are to be known and perpetuated by the forthcoming generations, also). It would be extremely useful to "typologise" the actual diversity of digital resources in order to turn all the information at hand regarding them, into a sui generis updatable *thesaurus* (which effort would surely mean a lot more than another portal-structured multiple link, however large and systematically envisaged); such a tool would serve to form one of the future bases for interrelating digital libraries and knowledgeable individuals and to promote mutual exchange as information and experience are concerned. The state-of-the art technology has already moved into this far reaching domain (the running MASTER project is a good exemple).

Aiming to join the newest trends in book-digitization and conform to this afore detailed status – that of providing a fundamental instrument of investigation (having a "classic" infrastructure, uniting word and image within a user-friendly document retrieval technique, either for amateur consultation, either for scientific purposes) and that of an expanded and refined tool for worldwide accesibility to one of the famous Romanian collections (which implies certain specifications added to the works in

question, and even original contributions of a refferential kind) – the Institute for Cultural Memory has created **APOGRAF**.

1. a. The Manuscript Collection of the National Batthyaneum Library.

Situated on a hill displaying a wide perspective, close to the North-Western walls of a XVIII-th century fortress erected during the reign of the Austrian emperor Charles the VI-th, The Batthyaneum Library usually opens its massive doors to only a few conoisseurs. In 1792, under the direct supervision of Ignatius Batthyany, a Catholic bishop born in 1741, quite an educated man of his time, theologian, philosopher and historian who studied in Hungarian, Austrian and Italian Universities (Rome) with a keen interest in bibliophilic rarities (he was an unexhausted collectors of old manuscripts and incunabula), the building turns into a scholarly institute. The most precious pieces from the initial collection belonged to the personal library of Bishop Migazzi of Vacz, also cardinal of Viena. The entire collection was bought by Batthyany, Bishop of Alba Iulia (1780-98) and thus the main part of the Batthyaneum Library takes shape, ready for public access; it has been constantly developed over the last centuries, by acquisitions and donations, but the principle treasure consists of the initial corpus. In 1824, the first catalogue numbered 18.201 volumes, approximately. In our days, the library comprises the average quantity of 60.000 volumes, of which – 1230 – are manuscripts, 560 – incunabula, and the rest is formed by a series of documents regarding the history of Transylvania. The manuscripts and the incunabula are well-known to the experts from all around the world.

At Batthyaneum, religious works prevail, but one can find literature, history, geography, natural science, medicine works, as well.

In addition, it should be emphasized that one of the richest collection of Bible specimens of different editions and ages is carefully preserved here, and offered for study.

Among the manuscripts of fine mastership housed at Batthyaneum, several pieces are worth mentioning herewith:

- the oldest and world-famous manuscript of the library, "**Codex Aureus"**, with a particular history of its own, is an outstanding example of manuscriptorial art (its graphic peculiarities are further endowed with wonderful illuminations); only the first part of the Codex is preserved in the Batthyaneum Library, the second half belongs to the Vatican Library;

- Andrea Vendramini's "**Promissio**", a worthy XV-th century (1476) in folio, with somptuous gilded illustrations; the manuscript's covers are the original ones, made of parchment;

- Ptolemeus's "**Cosmographia**" (1482) which contains an atlas with maps coloured by hand; among these maps, one illustrates locations found in present-day Romania.
- a "**Psalterium Davidicum**" , richly illustrated by hand, with gold-decorated miniatures; it is considered, by the bibliographic authorities, as an artistic masterpiece of the XIII-th century European genre;
- a parchment copy after Sallustius's "**De Bello Jugurthino**", dating from the Xth-XIth centuries, special for its uniqueness and antiquity, belonging to the second class of Sallustius manuscripts (known as more accurate and less lacunary) *Etc.*

1.c. Core-project: The Electronic Catalogue of Batthayneum Manuscripts

The electronic catalogue is based upon the conversion of Robert Szentivanyi's work: "*Catalogus concinnus librorum manuscriptorum Bibliothecae Batthyányanae : Editio quarta retractata, adaucta et illuminata*" issued in 1950 in Szeged, Hungary, into a newly elaborated form, first by using ROMARC (Romanian version of UNIMARC with special extensions for older monographies) to transpose the specific data included in each description made by R. Szentivanyi, within an Access database, then allowing the whole work to "rephrase" its contents according the ISBD (A), at the output. Each record in the ROMARC format is divided into two components (according to a pre-established convention): the principal record, namely the *manifestation* (type "0") holding the bibliographic information, and a second one (or more, if there are more than one copies of the item described in the first record) holding the information reffering to certain peculiarities, as : support (parchment or paper or both), binding (its material and artistry), preservation state, exlibris, handwritten notes (marginalia or property statements) etc.. The subsequent record(s) is/are linked to the principal one (which, at its best, displays the full characteristics of the ideal exemplary) . The description of a miscellaneum includes the analytical description of each component work within a different record and the correct conjunction of such record to the principal one, so that the complete picture of the miscellaneic structure is provided as a single layout; this proves specifically useful when each record is accompanied by explanatory notes or augmented refferences, in order to help the identification of the author, period, subject of the work etc.

By a simple click, every individual script of a miscellany *should* be viewed by the user either separately, either as an integral part of the whole corpus, the links between them easing up the search of both variants and/or their specific elements.

This descriptive approach not only enables the cataloger to easily convert, amend and control the material of his/her work, but also to create performant queries and primary hyperlinks, with the view of an online facility where they can be developped and "standardized".

The system this approach is hopefully growing into, is that of an arborescent construction accrued from the idea(l) of a complex and preferably instant communication within the same bibliographic block.

2. The Main Project – APOGRAF

The project we are reffering to in this presentation paper was generated by the **Romanian Ministry of Culture and the Cults**, in association with **WorldBank**. The chief responsible for the technological realisation is **Star Storage**, one of the forefront firms specialized in products and services for document management, archiving and storage. The original manuscripts were manipulated only by the library personnel; the scanner used is a Zeutschel.

At first, a number of 80.000 pages were scanned, and in 2004, another 40.000.

The most valuable items were selected for this job and they already constitute a **Digital Library**. The interface still needs improving, a process being underway.

The library is intended to include another series of 50.000 manuscript pages of non-Batthyaneum provenance. For now, the working version can be seen at **www.apograf.cimec.ro/wx/.**

2.a. Plans for tomorrow

Due to its importance, in the nearest future **APOGRAF** is about to be integrated into a larger project called "**The National Portal of the Libraries**", whose main component is the National Digital Library. Users of this library will have access to: the bank of facsimiles, represented by **APOGRAF**; full text documents; audio or phonographic records or "the speaking books".

A crucial component of the major project is "**The National Partitioned Catalogue**". It will include the metadata of the items stored and described in **The Digital Library**, divided into three cathegories: facsimiles, texts and audible records. In this context, the displayed written page of the document has no less than three different meanings: 1. the digital copy of the original page (see the significance of the term "*apograph*"), 2. the transcription of the original page (a so called "diplomatic edition"), and 3. the reading out loud of the text in question, i.e. the audio document.

Outlier Detection as a Method for Knowledge Extraction from Digital Resources

Eugenia Stoimenova, Plamen Mateev, Milena Dobreva
IMI, Bulgarian Academy of Sciences, Sofia, BG

Keywords: *outliers detection, knowledge extraction, digital resources*

Introduction

One characteristic feature of the digital preservation of and access to cultural heritage is the collection of voluminous data in electronic form which processing is still not a trivial task. These data most often are encoded within the metadata accompanying digital images and full texts. For example, such records are available in the cases of manuscript and archival descriptions, museum collections, etc. Currently, these data are used in visualization of records and in identifying documents, which answer specific criteria. The processing summary information is not very popular, while it could lead to discovering new facts about the cultural and scientific heritage, for example its regional and chronological distribution, items which do not follow the general trends, etc. As an example of innovative approach applied to a collection of manuscript data we could mention the development of proper intelligent agents for search and processing purposes which are able to retrieve and filter data (documents and images) by their semantic properties [4].

In our work we will present the application of outlier detection methods in the studies of data on cultural and scientific heritage resources.

Outliers

In many experimental tasks, in information retrieval, for instance, we have the situation of records, which differ much from the main part of the data. They seem to be surprisingly different, higher or lower, than one would expect from the rest of the records and from the "knowledge" about the underlying process or semantics, which generate the information items. Such records are usually called “outliers” although no formal definition exists.

For instance, in the sequence 0, 1, -1, 3, -2, 103, the last number 103 is an outlier. An outlier may indicate that there was an error in the process, which produced the data, or it may show there is a real abnormality in the system, which we are studying.

An "outlier" is a statistical term. It refers to an observation, record in the database that lies an abnormal distance from other values in a sample from a population. In a sense, this definition leaves it up to the analyst to decide what will be considered abnormal in an assumed context. Before abnormal records can be singled out, it is necessary to characterize normal records.

Of course, it depends on the assumed model if an extreme record is considered as surprisingly different, i.e. if it arises from some other source than the remaining data. The main goal of any statistical analysis is a study of the norm or typical characteristics of a system.

Outliers make any statistical analysis difficult. They might give bias impression about the system and could lead to wrong decision, for instance. When we encounter an outlier, we may be tempted to delete it from the analysis. Before deleting a record, first, we should ask these questions:

- *Was the value entered into the computer correctly?* If there was an error in the data entry, it should be fixed.
- *Were there any problems with collecting data in that record?* For example, if we noted that one document looked false, we have justification to exclude the record resulting from that document without needing to perform any calculations.
- *Is the outlier caused by document diversity?* If each record comes from a different chronology time, the outlier may be a correct value. It is an outlier not because of recording mistake, but rather because that this age documents may be different from the others. *This may be the most exciting finding in our data!*

After answering no to those three questions, we have to decide what to do with theoutlier. There are two possibilities. The first is that the outlier appeared due to chance. In this case, we should keep the value in our analysis. The value came from the same population as the other values, so it should be included. The second possibility is that the outlier appeared due to a mistake – bad recording, forgery, imitation, etc. Since including an erroneous value in the analysis will give invalid results, it should be removed. In other words, the value comes from a different population than the other and is misleading.

The problem, of course, is that we can never be sure which of these possibilities is correct. Clearly, no mathematical calculation will tell us for sure whether the outlier came from the same or different population than the others. However, statistics can answer the questions like these: If the values really were all sampled from a given distribution, what is the chance that we would find one value as far from the others as we observed? If this

probability is small, then we may conclude that the outlier is likely to be an erroneous value, and we have justification to exclude it from our analysis.

Detection Methods

Many methods have been proposed for univariate outlier detection. They are based on (robust) estimation of location and scatter, or on quantiles of the data. A major disadvantage of these methods is that the decision rules are independent from the sample size. Moreover, by definition of most rules (e.g. based on a distance from the mean) outliers are identified even for "clean" data, or at least no distinction is made between outliers and extremes of a distribution.All detection methods first quantify how far the outlier is from the other values. This can be the difference between the outlier and the mean of all points, the difference between the outlier and the mean of the remaining values, or the difference between the outlier and the next closest value. Next, standardize this value by dividing by some measure of scatter, such as the standard deviation (SD) of all values, the SD of the remaining values, or the range of the data. Finally, compute a p-value answering this question: If all the values were really sampled from a Gaussian population, hat is the chance of randomly obtaining an outlier so far from the other values? If the p-value is small, we conclude that the deviation of the outlier from the other values is statistically significant.

The basis for multivariate outlier detection is the Mahalanobis distance. The standard method for multivariate outlier detection is robust estimation of the parameters in the Mahalanobis distance and the comparison with a critical value of the Chi-square distribution [2] (Vandev, 2004). However, also values larger than this critical value are not necessarily outliers, they could still belong to the data distribution.

Experiments

At the present moment (14/06/2005), the work on KT-DigiCult-Bg project foresees collection of metadata on mediaeval manuscripts, archival records (jointly with the General Department of Archives) and mathematical publications of Bulgarian authors.

A collection of 806 XML descriptions of mediaeval manuscripts is already available.

This amount will increase with project evolution and will lead to large data sets of different types.

All these resources contain large data sets of different nature. Some of them contain structured information and metadata where in the future corpora – collections of different size text segments – will be added. Using

more or less intelligent retrieval systems one can extract relatively small subset of items with homogeneous structure. The extracted elements may be presented as vector of features. The text segments usually are presented via indicators of some key words with specified frequencies (of words, collocations, letters, bigramms, trigramms etc.), i.e. as vectors of features too.

Our paper will present an empirical study on a sample of KT-DigiCULT-Bg Resources data sets. Applicability of selected methods for univariate and multivariate outlier detection ([2] and [3]) is tested. Comparative analysis and some unexpected facts will be reported.

Acknowledgement
This research has been supported by a Marie Curie Fellowship of the European Community programme "Knowledge Transfer for Digitalization of Cultural and Scientific Heritage in Bulgaria" under contract number MTKD-CT-2004-509754.

References

[1] KT-DigiCult-Bg project, MTKD 509754.
[2] D. Vandev, *Detecting multidimensional outliers*, Technical report, 2004 (unpublished).
[3] J. D. Jobson, *Applied Multivariate Analysis*, Springer-Verlag, 1991.
[4] M. Nisheva-Pavlova, P. Pavlov, *Tools for Intelligent Search in Collections of Digitized Manuscripts*. In: “From Author to Reader”, 9th ICCC International Conference on Electronic Publishing (eds. M. Dobreva, J. Engelen), Peeters Publishing Leuven, 2005, 145-150.

The Names *Turkey* and *Macedonia* in Old Geographical Maps (1400 – 1700 AD)

Jordan Tabov, Dessislava Dimkova, Emil Kelevedjiev
IMI, Bulgarian Academy of Sciences, Sofia, Bulgaria

A study concerning the presence of names like "Turkey" and "Macedonia" in some old geographical maps (1400 – 1700 AD) is considered. There are examined more than twenty digitized maps of that period, including the maps of Bianco (1436), Ptolemy (1490, 1503, 1511), Mercator (1575), Ortelius (1579) and others. We found some unexplainable lack of names referred to Turkey or the Ottoman Empire and some prevalence of the name "Macedonia" till 17th century on the maps showing the regions of South-Eastern Europe and Asia Minor.

Some aspects of digitizing maps and suitable forms of storing are also discussed. Questions arising about practical use of a resolution level, which is necessary for a proper identification of names on maps, and how to identify them, are presented.

Medieval Manuscript Representation in the Virtual Environment: between Librarianship, Archivistics, Historical Auxiliary Sciences, Philology, and History

Zdeněk Uhlíř
National Library of the Czech Republic, Prague, Czechlands

Keywords: *medieval manuscript, digital copy, full text edition, open shared catalogue, digital library, virtual research environment*

This paper is based on the experience of the National Library of the Czech Republic and on another advanced activities concerning manuscript digitisation as well. Its purpose is considering some perspectives of medieval manuscript digitisation in the context of librarianship and archivistics, historical auxiliary sciences, especially codicology, paleography, respectively diplomatics, as well as philogy and history in a wider sense.

Manuscript digitisation is not only producing images. There is also necessary to provide tools for navigation and orientation withing the growing amount of images. So the images-digital copies of the manuscipt pages must be mutually connected and interrelated as complex digital documents: individual files-digital objects must be gathered onto the virtual book/s. Then the virtual books-complex digital objects must be secondary represented as descriptive catalogue records in order to be possible orientation within the whole resource. And after that the catalogue records must be put into the information system-cataloge in order to be possible navigation through hyperlinks from the record to the individual concrete file-image-page. So the manuscript digitisation means building not only browsable but also fully searchable digital library. Searching within the digital library is basis for the professional and research work.

On the other hand the descriptive catalogue records and images/sets of images is not enough for any advanced and sophisticated work because the most part of the humanities is based on the texts and the work with the texts. Of course, the majority of images are texts in that sense that they are representations of language testimonies. On the ather hand they are not texts in technical sense: they are not machine readable, i.e. they cannot be processed as texts by the using computer. Thus, the practical consequence must be converting these implicit texts to the explicit, i.e. electronic texts. Nowadays it is not possible by the using any of the OCR technologies because there is a huge variety of different types of script as well as

individual writings (so-called scribe hands). So the implicit texts-images must be transcribed manually in order to prepare an available electronic text. Such a machine readable text is then connected both to the descriptive catalogue record and to the respective images-copies of the manuscript pages.

As so as all the data are put into the information system of the digital library they must be standardised; otherwise they would be unprocessable. Standardisation of data first of all means the strict separation between the data on one and the software on the other hand. Then the data can be processed by various software tools and can be put into the various information systems as well. Very useful standard for manuscript describing is MASTER (Manuscript Access through Standard for Electronic Records). Catalogue records can be prepared using very different text processors as well as more complicated systems. Now the MASTER standard is upgraded in order to be compatible with the TEI (Text Encoding Initiative) standard: the manuscript module in the frame of the TEI standard (document type definition and/or relax NG scheme). Consequently the MASTER/TEI manuscript module is very widely used in Europe and America because it enables an easy interoperability between various systems. There is possible on this basis sharing of the data and creating step by step the virtual environment for the manuscript work worldwide.

The same holds good for the fulltext standardisation. Only difficulty is that the TEI standard as the whole is to vast because it is oriented to the text encoding of whatever kind. So there is need to constrict it for the practical usage. The TEI consortium provides a simple tool (Pizza Chef and newly Rome) for generating some closer offers of elements and attributes according to the particular purpose of resource and the individual orientation of researcher. It is very important that all these particular standards are mutually compatible and that they all are compatible also with the full TEI standard. Thus, the interoperability of various systems is ensured. And they are compatible also with the descriptive catalogue records such that the records, the images and the fulltext can be associated within the complex digital document of the higher range. As so as the MASTER as well as the TEI standards are written using the XML markup language (which is a meta-language) it is possible to integrate into the complex digital document also audial documents and so to create the genuine multimedial/multimodal document. It enables a new method of research work which is impossible in the traditional printed environment.

The electronic catalogue goes thgough the digital library to the real research environment such a way. There is a new virtual reality in view. It is

based on the easy integration of other internal and external tools as well. On the other hand the necessary condition is that these internal and/or external tools must be clearly distinguished both each from other and each from the data, i.e. the virtual research environment must be structured in a way of modularity. It is complex but not cumulative: it is possible to do various things within this environment but there should not be any need to do all the things together – the priciple of the virtual environment is selectivity. The internal tool that is a big challenge is a module utilising ontologies. It is the way form the traditional selection language (subject headings etc.) to an enriching of the fulltext searching. The way to the external tools is open, there are no fixed ideas. We can suppose that tools for advanced editing of historical texts should be very useful. On the other hand we can think about another tools of the computational linguistics, e.g. the tools enabling corpora creation, the tools enabling the work with word compexes, the tools enabling citation analysis etc. etc. Now the biggest challenge is seeking more possible ways, attempting new methods, preparing new data.

Of course, the traditional borders between disciplines (i.e. e.g. librarianship, archivistics, historical auxiliary sciences, philology and history) are crossed within the virtual environemnt. The new interdisciplinary and/or transdiciplinary paradigm is coming. It will bring many and many controversies between the "progressivists" and the "traditionalists". The core of these controversies consists in the transition from researching the indivudual phenomenon to researching the collective and/or massive phenomenon. The consequence is that e.g. hstorical auxiliary sciences are challenged to do transition from researching the external features of manuscript material to resaerching the internal features. Or philology is challenged to do transition from preparing the critical editions based on the idea of the fixed text to preparing the pragmatical editions based on the idea of the fluid text. Or history is challenged to transition from seeking *wie es eigentlich gewesen* to seeking the meaning of historical happening. Digitisation of manuscripts in such a wide sense is really a work on presenting the cultural heritage: it gives the further/second/other life to the historical material.

All Platform Digital Collections Presentation: A Simple Solution

Nebojša Vesić, Tamara Butigan-Vučaj,
National Library of Serbia, Belgrade, Serbia and Montenegro

Keywords: *digital document, digital collection, software solution, HTML, DHTML, JavaScript, metadata, Internet presentation*

Introduction

National Library of Serbia started the digitization process in 2002. In these few years the most attention has been paid to the scanning of library materials, not much to the presenting the digital collections. But the bigest challenge in digital libraries building is to make right solution for aceess to the digital contents, for online and offline users. One of the rare advantages of the "late digitization" is a huge experiance of the others, available on the Internet.

There are a lot of commercial software packages that are producing new formats of digital documents, better for viewing on the screen, faster for download, such as DejaVu, MrSid and others. Usually, it is necessary for users to download the plugins for reading these files. The more conventional formats of digital documents (DOC, PDF ...) are often too large to be accessed on Internet.

Open-source systems such as Greenstone, for constructing, presenting and maintaining digital collections present good choice, but often don't allow the full control of digital contents presenting.

In this paper, our solution for online and offline access to the scanned images is presented. It is intended primarily for manuscript collection, but could be also used for periodicals and some special collections.

Requirements

The initial requirements for the solution were:

- Universal accessibility
  The presentation should be visible on all (or almost all) present hardware/software platforms. The special attention should be paid to Windows and Linux operating systems, and the most popular internet browsers – Internet Explorer, Netscape, Opera, Mozilla etc.
- Suitability for Internet presentation

The presentation should be suited for presenting on Internet.

- Suitability for offline presentation
  The presentation should be suitable for presenting on an offline medium (CD, DVD, offline view on local computer or Intranet).
- Suitability for presenting various kinds of digital collection
  It should be possible to present various kinds of digital materials: manuscripts, printed books, photographs, newspapers etc.
- Use of widely accepted format
  "Exotic" presentation format should be avoided as much as possible in order to preserve future readability and usability.
- Possibility of automatic code generation
  It should be possible to generate presentation from internal database containing metadata about digital collection. The process should be automated, in order to avoid manual user intervention as much as possible.
- Use of widely available programming tools
  Programming tools to make presentation with should be chosen from one of widely accepted ones This should minimize probability of changing programming tools should the chosen one became obsolete.

Choosing Solution Technology: Fulfilling All Requirements

Requirements for universal accessibility and *suitability for Internet* have leaded to choice of almost the only one possible format: HTML with client JavaScript. It also fulfills the requirement to **use the widely accepted format.**

Suitability for offline presentation caused any web server technologies (like ASP or PHP) to be avoided. Using server technology is usually connected with some database system on server. Although all of these could be installed locally on (almost) any computer and operating system, it could be inconvenient and too complicated for most end users.

Suitability for presenting various kinds of digital collection could be achieved by careful design of presentation solution no matter what technology is used.

Automatic code generation and use of widely available programming tools have leaded to choice of JavaScript scripting language that is very suitable for dynamic Internet pages.

Solution features

Implemented solution has the following features:

- Navigation through the collection by means of buttons for *first*, *last*, *previous* and *next* page.
- Jump to any page of collection using *goto* button
- Jump to any part of the collection using *contents* button.
- Display metadata about collection as a whole
- Display metadata about every page of collection
- Zoom function
- Choice among different skins enabling different look ‘n feel for various kind
- of digital collection.
- Multilanguage user interface.
- The same code is used for online and offline usage

Some technical details

Presentation solution is written in DHTML and client-side JavaScript. No server technology is used. This enables making offline version of the presentation very easy.

All data about digital materials are contained in multidimensional static arrays in JavaScript code that is downloaded to user computer together with Internet page.

When the user selects a page to view, arrays are searched for the data about the requested page. The page is downloaded from server (or from a local folder in case of an offline presentation) and loaded into viewing frame.

Since no server technology is used, preparing an offline version is very easy. All Internet pages are just copied to an offline medium and delivered to the user.

Conclusion

This simple solution for digital collection presentation is based on basic requirements tending to simplify process of preparing and viewing digital collection. Its main advantages are:

- Platform independency – user is only supposed to have software for viewing standard Internet presentation
- Cheap and fast process of preparing digital collection for presentation.
- Possibility of automatic generation of presentation from various format of metadata
- No additional conversion from online version to offline version are needed

The implemented solution has some disadvantages:

- Used technology has limited means of digital material presentation since it is based on standard HTML
- Not all kinds of digital material are suitable for presenting (i.e. multimedia collection)
- Metadata are embedded into code. Some work (manual or automatic) is necessary to convert metadata to the form required by this solution.

Saying in one sentence, the implemented solution is not ideal, but is optimized in many aspects.