BULGARIAN NATIONAL ETHNOGRAPHIC MUSEUM – MEETING THE CHALLENGES OF DIGITISATION

Detelin Luchev

Abstract: The conservation, spread, comprehension and recreation of traditional culture heritages is one of the main purpose of the National Ethnographic Museum in Bulgaria. As other cultural and scientific heritage institutions, it begins to use new information technologies and strategies for providing access to its cultural heritage treasures. This paper aims to present digital libraries with multimedia content as a modern technological solution for innovative presentation of Bulgarian ethnographical heritage. It includes some basic concepts of digital libraries with multimedia content and a description of three types of architecture. The paper also describes the ideas, conceptual decisions and strategies in the project Experimental Digital Library "Bulgarian Ethnographic Treasury".

Keywords: Digital libraries with multimedia content; Digital library architectures; Cultural heritage; National Ethnographic Museum; Bulgarian ethnographic treasury.

ACM Classification Keywords: H.3.7 Digital Libraries – Collection, Dissemination, System issues.

Introduction

Europe's cultural, historical and scientific knowledge resources are a unique public asset forming the collective and evolving memory of our diverse societies. Resource discovery, accessibility, usability, interoperability, authenticity, quality and trust by all users of the Information Society are essential requirements for the delivery of digital cultural information and services [eEurope and Digitisation].

Digitization has introduced enormous possibilities for preserving, organizing and providing access to cultural heritage objects that were for one or another reason previously inaccessible. Digitalization of cultural and historical artefacts and creation of multimedia information archives; web presentations of valuable artefacts in virtual museums, galleries and digital libraries (DL); 3D virtual realities, which present places of culture and history; digital modelling and simulation, aiding the conservation, restoration, storing and showing artefacts, etc. [Pavlov at al., 2005] are various conceptual and technically feasible solutions for preserving the cultural, historical and scientific heritage of various world nations.

European libraries, archives and museums contain a wealth of information, representing the richness of Europe's history, its cultural diversity and its scientific achievements. By digitising their collections and making them available online, libraries, archives and museums can reach out to the citizens and make it easier for them to access material from the past. The online presence of this material from different cultures and in different languages will make it easier for citizens to appreciate their own culture heritage as well as the heritage of other European countries, and use it for study, work or leisure [Project MINERVAEUROPE]. The "i2010: Digital Libraries" initiative aims at making European information resources easier and more interesting to use in an on-line environment. The Commission adopted on 30/09/2005 the "i2010: Digital Libraries" communication outlining the vision of this initiative and addressing in particular the issues of digitisation, on-line accessibility and digital preservation of our cultural heritage [i2010: Digital Libraries, 2005].

Following this direction this paper aims to present digital libraries with multimedia content as a modern technological solution for innovative presentation of Bulgarian ethnographical heritage and the results of the first stage of the national project Experimental Digital Library "Bulgarian Ethnographic Treasury" (BET)[1]. It presents shortly also the main characteristics of the contemporary digital libraries, three types of DL architecture with different complexity, accentuating on the chosen one for the realization of the project. The last part of this article

[1] The Experimental Digital Library "Bulgarian Ethnographic Treasury" is a project, supported by a grant of the State Agency for Information Technologies and Communications (SAITS), according to contract No 12 / 08.09.2005 between SAITS and Ethnographic Institute with Museum-BAS - <http://ethnography.cc.bas.bg>.

describes the ideas, the conceptual decisions and the first results of the project for creating of BET. The main goals of the project have been the presentation of valuable collections of National Ethnographic Museum in Bulgaria to the wider public, the possibility to use digital library online and the integration of digitized ethnographic heritage into the education processes.

Ethnographical heritage: Traditional knowledge and new information technologies

Europe is well known for its multicultural and multilingual nature. Cultural heritage of different nations and ethnic groups represents an enormous wealth for Europe, but it is often not fully exploited or even neglected, especially in Bulgaria in the last 15 years. T. Blyth argues that “museums have long been hybrids, playing a variety of significant roles as collectors and preservers of material culture, as educators, and as entertainers” [Blyth, 2005]. Museology has itself widened and it is no longer object-concentrated, but phenomenon-centred and the focus has further turned to society. For the collective memory of European cultures, ethnographical collections have a broad social impact, particularly for the Balkan nations and ethnic groups. All of these cultures in unison give us base and background for cultural reconstruction and re-contextualization of our common experiences and memories.

New information technologies and the new museums dissemination strategies are directly related to the multicultural society and the new frameworks of approaching and interpreting diversity. Furthermore, the continuously evolving relationship of ethno-cultural heritage with the shaping of identity remains an important factor for such an approach in Europe and, in particular, on the Balkans. New technologies have the potential to increase worldwide public access to cultural resources and enriching the ways of communication. Combining information technologies and native (national) traditional culture protection can also lead to the appearance of new researches. Digital conservation of culture relics based on digitalization can reduce many of the problems caused by the irresistible disintegration and vanishing of relics.

Museums are still considered as places where “time slept in the corner” not only for many of visitors, but for some of specialist in this area in Bulgaria as well, and this view has to be changed. Globalization of contemporary information environment and international trends of involving of culture in all spheres of public life have impact on ethnographical museums as well. The Web is changing culture and information dissemination. In the globalization process many cultural traditions around the world tend to disappear under the pressure of standardization of practice and content. Cultural diversity seems to recede more and more. In the contemporary society memory institutions experience great changes associated with digitization. It is also a new method for preservation, education and access for many people to their own ethno-cultural (national) past and identity. The conservation, spread, comprehension and recreation of traditional culture heritages is one of the main purpose of the National Ethnographic Museum in Bulgaria.

National Ethnographic Museum is a part of the system of the Ethnographic Institute at Bulgarian Academy of Sciences and today it contains more than 50 000 valued exhibits, which are samples of the Bulgarian traditional folk arts and crafts collected from all territories historically inhabited by Bulgarians over the period mid-17th - mid-20th centuries. The Museum's collections are organized under several items: clothes, goldsmithery, copper objects, agriculture, woodcarving, home furnishing, ceramics, fabrics and embroideries, carpets, ritual objects, foreign art, etc.

National Ethnographic Museum in Bulgaria as other cultural and scientific heritage institutions meet several challenges related to the introduction and using of new information technologies: implementing systems that are prone to handle the increasing volume of heritage content to be digitized and presented in adequate forms; providing access to its cultural heritage treasures; settling a questions related to the archiving and preservation of cultural heritage content; offering personalized, interactive ways to this content; encouraging easy access to its own collections on the international level.

Digital libraries with multimedia content

Some authors argue that there is no need to regurgitate what digital preservation is and why it is important; there are numerous places where digital preservation is defined and discussed, and we presume readers already have a good understanding of the concepts and issues involved [Currall and McKinney, 2006]. In this paper we do not discuss this main digital preservation concepts, we will present contemporary tool for digital presentation of cultural artefacts – digital libraries with multimedia content. An informal definition of a digital library is a managed collection of information, with associated services, where the information is stored in digital formats and

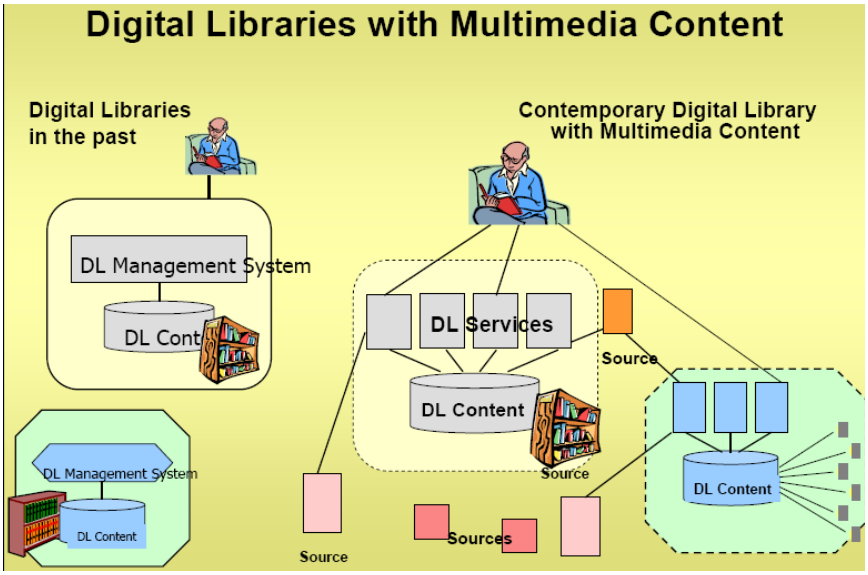
accessible over a network. The fundamental reason for building digital libraries is a belief that they will provide better delivery of information than was possible in the past [Arms, 2003].

Diverse hypertext-organized collections of thematically structured information are stored there for use by many different users and allow the connection between any piece of data and additional data on the same topic. As an addition to the digital objects collection, there are many levels of metadata, indexes, hierarchical links, etc. [Krastev, 2005].

The main characteristics of digital libraries are: the ability to share information; the new forms and formats for information presentation -- semantic annotation of digital resources and collection, for example; the easy information update -- indexing, data and metadata management; the personalization of services for searching, selecting, grouping and presenting digital information multi-layer and personalized search, context-based search, relevance feedback, etc, etc. [Pavlov and Paneva, 2005]. Digital libraries need to distinguish themselves from web search engines in the manner that they add *value* to web resources. This added value consists of establishing context around those resources, enriching them with new information and relationships that express the usage patterns and knowledge of the library community. The digital library then becomes a context for information collaboration and accumulation – much more than just a place to find information and access it [Lagoze at al, 2005].

The new digital libraries will provide and manage complex services, processes and workflows on the basis of existing services. It is expected that these services be heterogeneous, autonomous and distributed. The flexibility, the automatic adaptation, the access anywhere and anytime, the decentralization, the wide variety of digital objects and collections, the information security, etc., will be of the some requirements [IEEE, 2000; Kiernan and Kekhtyar, 2003]. And, as it was explained at the Computers in Libraries conference in Washington, DC, for example, there is a need for libraries to embrace change and innovate in order to meet the need of a new generation of users, who expect that information they seek will be instantly available anyplace, anytime, and preferably via a mobile device [Rainie, 2006].

In 1998, in the International Federation of Library Associations document „Digital Libraries: Definitions, Issues and Challenges" it was simply noted that the technical architecture will be a collection of disparate systems and resources connected through the Internet and integrated within one interface - a Web enabled interface [IFLAI 1998]. The digital library architecture research field is rapidly extending and developing from the time of this definition. At the beginning of our project EDL "Bulgarian Ethnographic Treasury" we observed different kinds of architecture.



Hypermedia digital library can be considered as a database, storing data of different type (text, raster, vector, static and moving (video) images, animation, audio or other media), which is structured in a way to allow easy manipulation and use. Data is stored in the database in the form of objects, usually annotated to facilitate running search queries. To make these procedures automatic, the hypermedia library includes techniques for descriptive presentation of the data semantics as well as services for its management.

Web technologies help organizing hypermedia digital libraries by providing a means to structure and present them in a hypermedia manner. Hypermedia represents hypertext media; therefore it adheres to the hypertext

information organization rules. Users are allowed to quickly move across subject-related topics in a non-linear way. These topics may include sets of objects, such as text, images, audio and other media, which relate to one another via hyperlinks [Paneva at al., 2005].

The Hypermedia digital library is a simplified conceptual solution for presenting complex multimedia content and is found expedient for the realization of first stage of EDL "Bulgarian Ethnographic Treasury" project.

Grid-based infrastructure - The digital libraries are currently undergoing a transition from a statically integrated system to a dynamic federation of services. This transition is inspired by new trends in technology which include developments in technologies like Web services and grid infrastructures as well as by the success of new paradigms like Peer-to-Peer Networking and Service-oriented Architectures. The transition is driven by digital library "market" needs. This includes a requirement for a better and adaptive tailoring of the content and service offer of a digital library to the needs of the relevant community as well as to the current service and content offer, and a more systematic exploitation of existing resources like information collections, metadata collections, services, and computational resources. Such a test-bed digital library infrastructure, for example, has been created for the DILIGENT project (Integrated project funded in part by the European Commission FP6 IST Programme), based on the grid technology [Project DILIGENT].

Hyperdatabase infrastructure - Future digital libraries should enable any citizen to access human knowledge any time and anywhere, in a friendly, multi-modal, efficient, and effective way. A core requirement for such digital libraries is a common infrastructure which is highly scalable, customizable and adaptive. Ideally, the infrastructure combines concepts and techniques from peer-to-peer data management, grid computing middleware, and service-oriented architectures. That infrastructure is offered in the project DELOS "A Network of Excellence on Digital Libraries" funded by the EU's Sixth Framework Programme. Peer-to-peer networks allow for loosely coupled integration of digital library services and the sharing of information such as recommendations and annotations. A service-oriented architecture provides common mechanisms to describe the semantics and usage of digital library services. Furthermore, it supports mechanisms to combine services into workflow processes for sophisticated search and maintenance of dependencies [Project DELOS].

Architecture and main characteristics of the experimental digital library "Bulgarian Ethnographic Treasury"

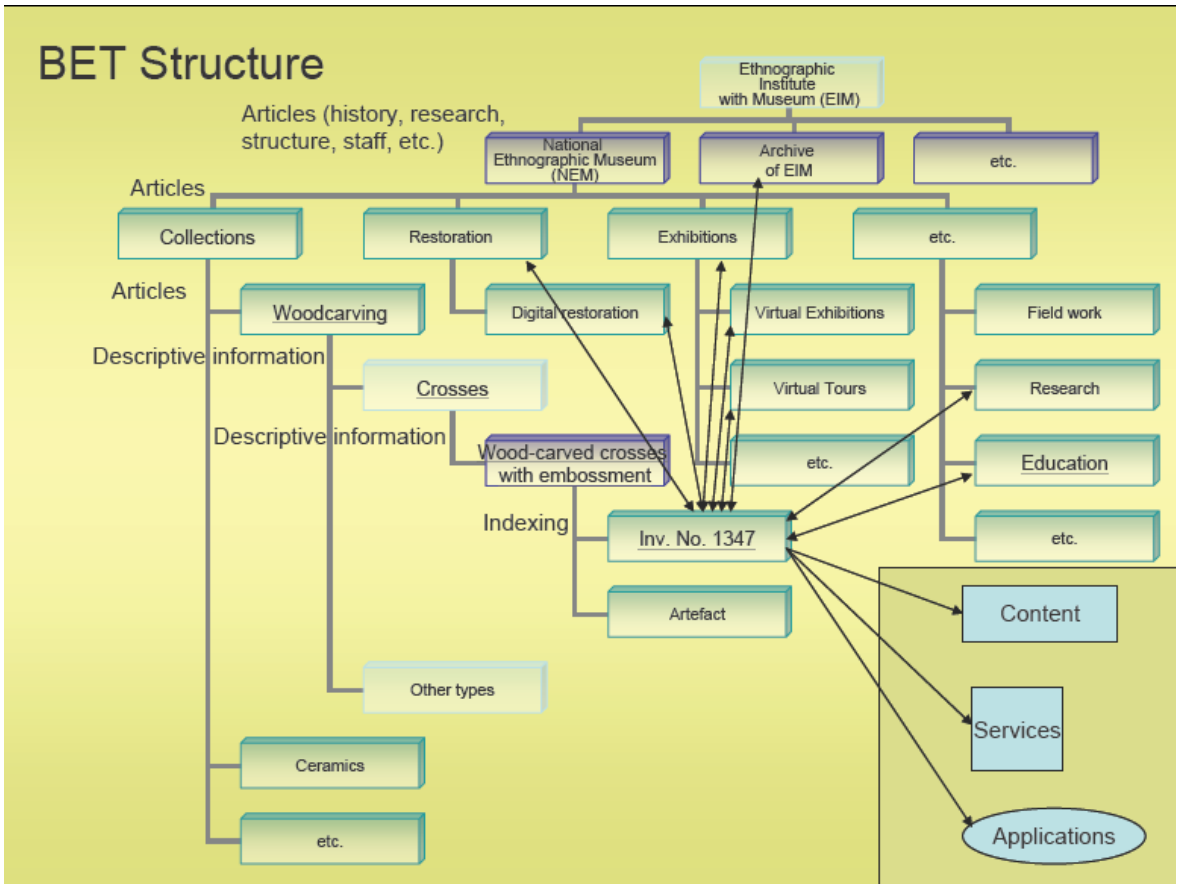
The Experimental Digital Library "Bulgarian ethnographical treasury" sets the beginning of the digitalization of the stock and collection treasure of NEM and is a prerequisite for ttransforming the work of the NEM to a new level consistent with the world trend in this field. The research aims the BET creation and development of the information content and structure compatible with the conditions and the needs of the Ethnographical Institute with Museum at Bulgarian Academy of Sciences. Attempt to digitize and expose in the Internet more of NEM's Bulgarian ethnographic artefacts would increase the accessibility to such a part of world cultural and historical heritage. It is important for presentation of our cultural identity in the time of integration of Bulgaria in EC. That would allow the preservation and even the future digital restoration of a large number of rare and even unique specimens of Bulgarian cultural heritage.

In the beginning of the project, the conditions of the artefacts preservation in the stocks and collections and the reasons of their ageing and ruin have been investigated in details. This research made possible to specify potentially and actually most damaged objects of the stock of the NEM and to focus attention on them by marking down first exhibits for digitalization and inclusion in the BET.

The choice of the most suitable of the existing tools, technologies and methods for making of digital archives was well grounded in order to guarantee quality and actuality of BET for relatively long period of time. The chosen architecture represents a hypermedia digital library. The resources - digital objects of different formats (text, graphics, and other media), are grouped according to their topics into (also real existing) thematic collections (woodcarving, clothes, goldsmithery, copper objects, agriculture, home furnishing, ceramics, fabrics and embroideries, carpets, ritual objects, foreign art, etc.).

The model follows several principles drawn in the process of the research: very wide representativeness of the different types of exhibits, which take place in the stocks; selection of most important for the national heritage types of artefacts; focusing the attention on the most threatened with ageing exhibits and aiming the maximal covering of the different ethnographical regions of the country. Descriptive information includes data about the category, the period, the location, etc. and on higher levels is made in form of short articles containing hyperlinks

to a founds, collections and kind of objects and could be navigated quickly, in a non-linear fashion, within areas of related topics.



Three hundred exhibits are digitalized and included in Experimental Digital Library “Bulgarian ethnographical treasury”. One of the aims of BET was to digitize as many objects, as it is possible, in a relatively short period of time. This attempt contributes for better representativeness and breadth of the particular information.

BET meets users' requirements to provide advanced searching capabilities to them. The NEM requirement is that users will be able to use a variety of searching functionalities so that access to the underlying information will be more effective. As is well-known, information discovery is a complex topic and no one approach satisfies all users or fits all materials. Usually it could be comprehensive search, search of known item, facts, overview and related information, and so on. Because of addressing the BET content to all kind of online users (not only for specialist of the same science field, for example), we decided to make the search service easiest and common in its view and use as more as it is possible. Through the service there could be find any descriptive metadata and data for any object in the digital archives of Ethnographic Institute with Museum (particularly – of NEM), as well as every kind of the institutional information based in BET. All the information offered to the user by the search service is naturally structured following the metadata, data description and presentation of objects, and he/she could start from the any chosen level and go up and down by hyperlinks. A query could be a search term in more than three letters, a full text searching, etc. The service interface is made simple and familiar for the users. Because the information discovery is more than searching, the decision for web-based experimental digital archives BET gives to the users a combination of browsing and systematic searching all the time they explore the information in the library.

At the end, BET has an wider and more complete structure, which represents all the activity of the Ethnographical Institute with Museum at the Bulgarian Academy of Sciences as institution including and centrally presenting National Ethnographic Museum’s collections. This structure outlines also the directions of the work by eventual

continuing of the project, the core of which will be the further building and completing of the digital archives themselves.

Through the project is made possible BET to contribute for the popularization of the Bulgarian culture in Europe and all over the world, to serve as a teaching tool and educational environment for learning and using the characteristics of the traditional Bulgarian culture and to simulate an additional interest in the work in museums. For example, immediately after the ending of the first stage of the project, the results of BET are of use to the training of the students of Ethnology Department of History Faculty at Sofia University "St. Kl. Ohridski".

Conclusion

The completed work is a good base for the continuation of the project aimed at the extension the range of digitalization of the stocks of the National Ethnographical Museum as well as for overall representation of the activity of the institution in the virtual space. The project lays the foundations of the registration, documentation, and the exploration of a practically unlimited number of ethnographic cultural artefacts of NEM stocks.

The realization of the project confirms the opinion that the digitization of collections, ethnological particularly, certainly needs to be based on former experiences in digitization of cultural heritage and on cooperation with experts of various profiles – here it is teamwork of specialists of Ethnographic Institute with Museum and of Institute of Mathematics and Informatics at Bulgarian Academy of Sciences.

The accomplished work outlines also the possibility for the continuation of the project connected with the further building and completing of BET themselves and with the direction to increase the multimedia representation of its components. On the next stages of the work the efforts could be directed to the creation of virtual ethnographical exhibitions as well as virtual tour of the National Ethnographical Museum.

Acknowledgement

The work was made possible with the kind support of Radoslav Pavlov and Dessislava Paneva. I would like to thank them for their encouragement and patience throughout the beginning of my work in this science field and especially for active involvement and the numerous valuable discussions with respect to digital libraries with multimedia content as a technological solution for presentation of Bulgarian ethnographical heritage. I would also like to thank Lilia Pavlova-Draganova and Lubomil Draganov for participation in the project design.

Bibliography

[Arms, 2003] Arms, W. Y. Digital Libraries. Online Edition. 2003 - <http://www.cs.cornell.edu/wya/DigLib/new/Chapter1.html>

[Blyth, 2005] Blyth, T. "Curating for Broadband". In: Museums and the Web 2005. Toronto: Archives & Museum Informatics. <http://www.archimuse.com/mw2005/papers/blyth/blyth.html>.

[Currall and McKinney, 2006] Currall, J. and P. McKinney. Investing in Value: A Perspective on Digital Preservation. In: D-Lib Magazine, 4 (April), 2006. - <doi:10.1045/april2006-mckinney>

[eEurope and Digitisation] eEurope and Digitisation - <http://www.cordis.lu/ist/digicult/eeurope.htm>

[i2010: Digital Libraries, 2005] "i2010: Digital Libraries", COM, 465 final, Brussels, 30/09/2005

[IEEE, 2000] IEEE Multimedia. In: Virtual Heritage, April-June 2000, Vol.7, No.2

[IFLAI 1998] International Federation of Library Associations and Institutions. Digital Libraries: Definitions, Issues and Challenges. The Hague, Netherlands: IFLA, 1998.

[Kiernan and Kekhtyar, 2003] Kiernan, K. and A. Kekhtyar. EPT: Edition Production Technology for Multimedia Contents in Digital Libraries, Presented on Workshop on Multimedia Contents in Digital Libraries, USA, 2003.

[Krastev, 2005] Krastev D.. Central Library of Bulgarian Academy of Sciences – present and future, The libraries of Bulgarian Academy of Sciences, Reference book, Bulgaria, 2005.

[Lagoze at al, 2005] Lagoze, C. D., B. Krafft, S. Payette, S. Jesuroga et al. What is a Digital Library Anymore, Anyway?: Beyond Search and Access in the NSDL. In: D-Lib, November 2005 - <doi:10.1045/november2005-lagoze>

[Paneva at al., 2005] Paneva D., L. Pavlova-Draganova and L. Draganov, Digital Libraries for Presentation and Preservation of East-Christian Heritage, In: Generic Issues of Knowledge Technologies, Proceeding of HUBUSKA Second Open Workshop, Budapest, Hungary, 14 September 2005, 75-83.

[Pavlov and Paneva, 2005] Pavlov R. and D. Paneva. Towards a Creative Exploitation of Digitised Knowledge in eLearning Systems, 2nd CHIRON Workshop, Paris, France, 10-11 October 2005.

[Pavlov at al., 2005] Pavlov R., D. Paneva, L. Pavlova-Draganova and L. Draganov. Digital libraries with multimedia content and applications in Bulgarian cultural heritage (Analytical study), State Agency for Information Technology and Communications, Sofia, Bulgaria, 2005.

[Project DELOS] Project DELOS: A Network of Excellence on Digital Libraries - <http://www.delos.info>

[Project DILIGENT] Project DILIGENT: Digital Library Infrastructure on Grid Enable Technology - <http://www.diligentproject.org>

[Project MINERVAEUROPE] Project MINERVAEUROPE: Ministerial Network for Valorising Activities in digitalisation - <http://www.minervaeurope.org>

[Rainie, 2006] Rainie, L. Life Online: Teens and technology and the world to come (Annual conference of Public Library Association, Boston, March 23, 2006) - <http://www.pewinternet.org/ppt/Teens and technology.pdf>

Authors' Information

Detelin Luchev - Ethnographic Institute with Museum, BAS; Bulgaria, 1000 Sofia, Moskovska str. 6A; e-mail: luchev_detelin@abv.bg

APPLYING GENETIC ALGORITHM IN QUERY EVALUATION PROBLEM

Abdel A. Aly

Abstract: This paper presents an adaptive method using genetic algorithm to modify user's queries, based on relevance judgments. This algorithm was adapted for the three well-known documents collections (CISI, NLP and CACM). The method is shown to be applicable to large text collections, where more relevant documents are presented to users in the genetic modification. The algorithm shows the effects of applying GA to improve the effectiveness of queries in IR systems. Further studies are planned to adjust the system parameters to improve its effectiveness. The goal is to retrieve most relevant documents with less number of non-relevant documents with respect to user's query in information retrieval system using genetic algorithm.

1. Introduction

Several researchers have used the GA in IR and their results seem to indicate that this algorithm could be efficient. In this vein, the main directions concern modifying the document indexing [6] and [27] and the clustering problem [28].

It is not surprising therefore that there have recently appeared many applications of GAs in information retrieval. Most of them use the vector space model, which also seems to be one of the most widely, used models in general [13]. These applications implement learning of the terms and/ or weights of the queries.

Information Retrieval systems are used to retrieve documents that depend on or relevant to the user input query. The growth in the number of documents made it necessary to use the best knowledge or methods in retrieving the most relevant documents to the user query.

Information Retrieval systems deal with data bases which are composed of information items documents that may consist of textual, pictorial or vocal information. Such systems process user queries trying to allow the user to access the relevant information in an appropriate time interval. The art of searching will be in the databases or hypertext networked databases such as internet or intranet for text, sound, images or data, [13]. Thus an information system has its heart a collection of data about reality [29].

Most of the information retrieval systems are based on the Boolean queries where the query terms are joined by the logical operators AND and OR. The similarity between a query and documents is measured by different retrieval strategies that are based on the more frequent terms found in both the document and the query. The more relevant document is deemed to be the query request. The most frequently used measures of retrieval effectiveness are precision, *the percentage of the retrieval documents that are relevant* and recall, *the percentage of the relevant documents that are retrieved*.