I T H E A

International Journal

INFORMATION TECHNOLOGIES & KNOWLEDGE

2007 Volume 1 Number 4

International Journal
INFORMATION TECHNOLOGIES & KNOWLEDGE
Volume 1 / 2007, Number 4

IJ ITA is official publisher of the scientific papers of the members of the Association of Developers and Users of Intellectualized Systems (ADUIS) and the Association for Development of the Information Society (ADIS)

IJ ITK rules for preparing the manuscripts are compulsory.
The **rules for the papers** for IJ ITK as well as the **subscription fees** are given on *www.foibg.com*.
The camera-ready copy of the paper should be received by e-mail: info@foibg.com.
Responsibility for papers published in IJ ITK belongs to authors.

General Sponsor of IJ ITK is the **Consortium FOI Bulgaria** (www.foibg.com).

International Journal "INFORMATION TECHNOLOGIES & KNOWLEDGE" Vol.1, Number 4, 2007
Printed in Bulgaria

Edited by the **Institute of Information Theories and Applications FOI ITHEA**, Bulgaria, in collaboration with the **V.M.Glushkov Institute of Cybernetics of NAS, Ukraine**, and the **Institute of Mathematics and Informatics** and the **Institute of Information Technologies, BAS, Bulgaria**.

Publisher: **Institute of Information Theories and Applications FOI ITHEA**
Sofia, 1000, P.O.B. 775, Bulgaria. www.foibg.com, e-mail: info@foibg.com

ISSN 1313-0455 (printed) **ISSN 1313-048X (online)** **ISSN 1313-0501 (CD/DVD)**

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 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

[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>.

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

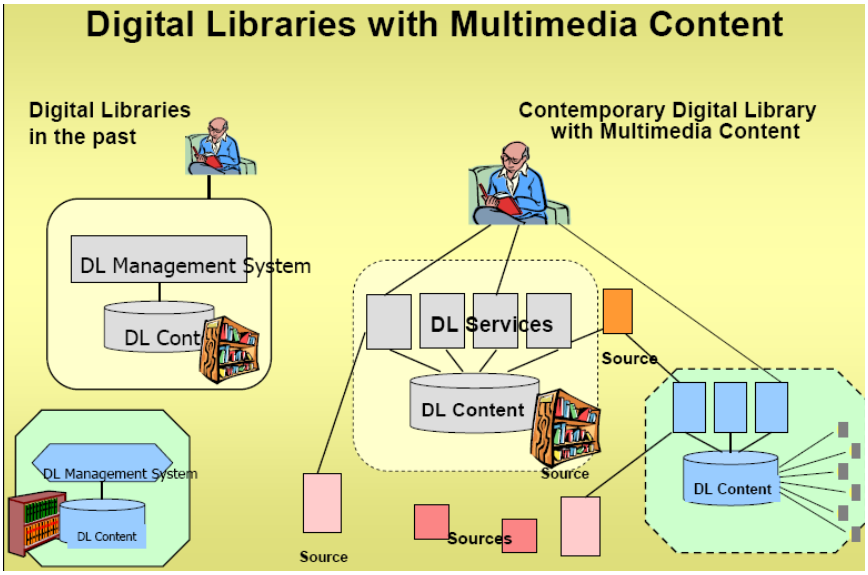
Some authors argue that there is no need to regurgitate what digital preservation is and why it is important; there are numerous places where digital preservation is defined and discussed, and we presume readers already have a good understanding of the concepts and issues involved [Currall and McKinney, 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].

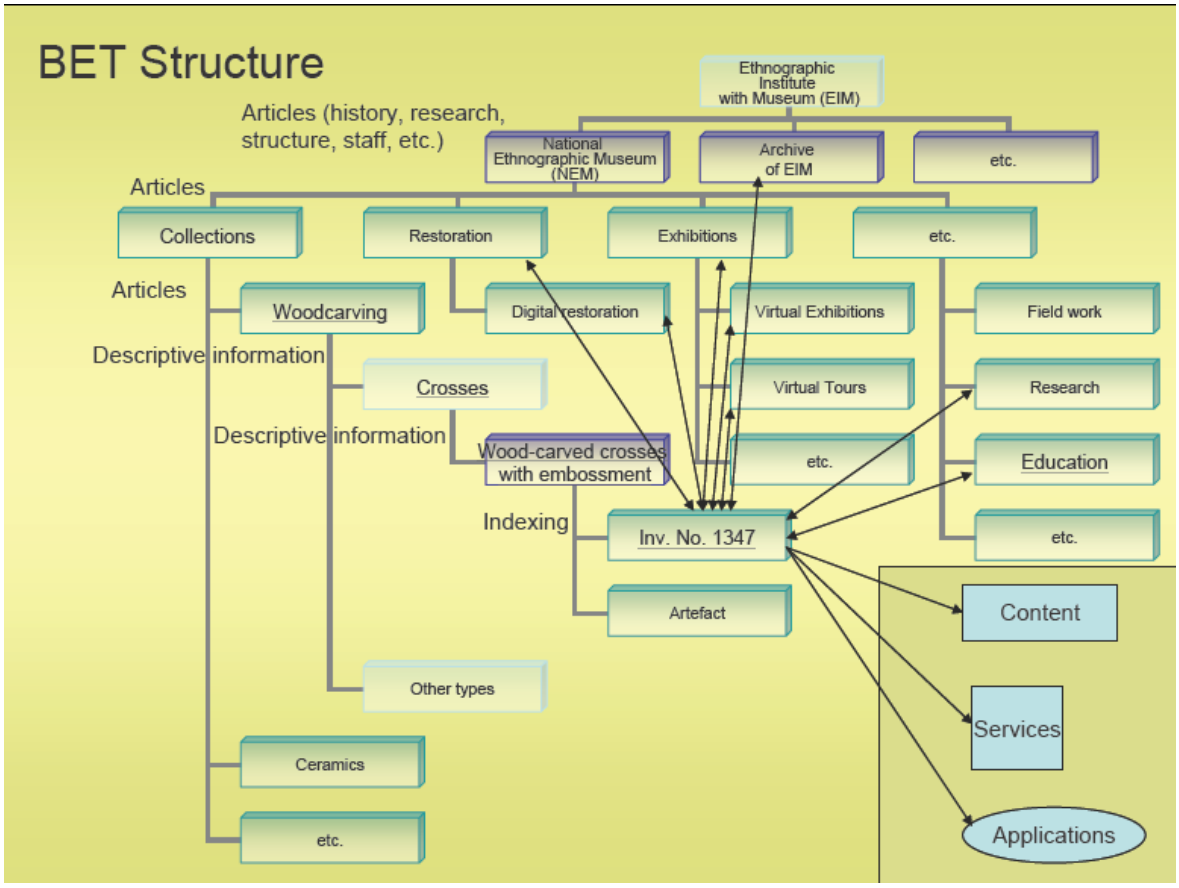
Architecture and main characteristics of the experimental digital library “Bulgarian Ethnographic Treasury”

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 IMPROVEMENT PROBLEM

Abdelmgeid 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 [1] and [2] and the clustering problem [3].

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 [4]. 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, [4]. Thus an information system has its heart a collection of data about reality [5].

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*.

Information retrieval is concerned with collection and organization of texts, responding to the requests of users for the information seeking texts, retrieving the most relevant documents from a collection of documents; and with retrieving some of non-relevant as possible. Information retrieval is involved in:

- Representation,
- Storage,
- Searching,

- Finding documents or texts which are relevant to some requirements for the information desired by a user.

The input of GAs [6] is a population of individuals called chromosomes, which represent the possible solutions to the problem. These are either generated at random or, if one has some knowledge can be used to create part of the initial set of potential solutions [7]. These individuals change (evolve) over successive iterations called generations, via processes of selection, crossover, and mutation. The iterations end when the system no longer improves or when a pre-set maximum number of generations has been reached. The output of the GA will be the best individual of the end population, or a combination of the best chromosomes.

To solve each problem, one has to provide a fitness function *f*. Its choice is crucial for the proper functioning of the algorithm. Given a chromosome, the fitness function must return a numeric value that represents its utility. This score will be used in the process of selection of the parents, so that the fittest individuals will have a greater chance of being selected.

This paper presents a GA based on Cosine fitness function and Classical IR in query optimization problems, and some applications of GA in information retrieval system. This algorithm has been applied on three well-known test collections (CISI, CACM and NPL). The goal is to retrieve most relevant documents with less number of non-relevant documents with respect to user query in information retrieval system using genetic algorithm.

2. Antecedents

2.1. Information retrieval models

Several retrieval models have been studied and developed in the IR area, we analyze some of these models, which are:

Boolean model. In the Boolean retrieval model, the indexer module performs a binary indexing in the sense that a term in a document representation is either significant (appears at least once in it) or not. User queries in this model are expressed using a query language that is based on these terms and allows combinations of simple user requirements with the logical operators AND, OR and NOT. The result obtained from the processing of a query is a set of documents that totally match with it, i.e., only two Possibilities are considered for each document: to be or not to be relevant for the user's needs, represented by the user query [8][9].

Vector space model. In this model, a document is viewed as a vector in an n-dimensional document space (where n is the number of distinguishing terms

used to describe contents of the documents in a collection) and each term represents one dimension in the document space. A query is also treated in the same way and constructed from the terms and weights provided in the user request. Document retrieval is based on the measurement of the similarity between the query and the documents. This means that documents with a higher similarity to the query are judged to be more relevant to it and should be retrieved by the IRS in a higher position in the list of retrieved documents. In This method, the retrieved documents can be orderly presented to the user with respect to their relevance to the query [8].

Probabilistic model. This model tries to use the probability theory to build the search function and its operation mode. The information used to compose the search function is obtained from the distribution of the index terms throughout the collection of documents or a subset of it. This information is used to set the values of some parameters of the search function, which is composed of a set of weights associated to the index terms [10][11].

2.2. Evaluation of information retrieval systems

There are several ways to measure the quality of an IRS, such as the system of efficiency and effectiveness, and several subjective aspects related to the user satisfaction. Traditionally, the retrieval effectiveness (usually based on the document relevance with respect to the user's needs) is the most considered. There are different criteria to measure this aspect, with the precision and the recall being the most used.

Precision (P) is the rate between the relevant documents retrieved by the IRS in response to a query and the total number of documents retrieved, whilst Recall (R) is the rate between the number of relevant documents retrieved and the total number of relevant documents to the query existing in the data base [9]. The mathematical expression of each of them is showed as follows:

$$P = \frac{\sum_d r_d \cdot f_d}{\sum_d f_d} \quad , \quad R = \frac{\sum_d r_d \cdot f_d}{\sum_d r_d} \longrightarrow \qquad (1)$$

with $r_d \in \{0,1\}$ being the relevance of document d for the user and $f_d \in \{0,1\}$ being the retrieval of document d in the processing of the current query. Notice that both measures are defined in [0,1], being the optimal value.

3. Query Definition

This is the most extended group of applications of GAs in IR. Every proposal in this group uses GAs either as a relevance feedback technique or as an Inductive Query By Example (IQBE) algorithm.

The basis of relevance feedback lies in the fact that either users normally formulate queries composed of terms, which do not match the terms (used to index the relevant documents to their needs) or they do not provide the appropriate weights for the query terms. The operation mode involving and modifying the previous query (adding and removing terms or changing the weights of the existing query terms) which taking into account the relevance judgments of the documents retrieved by it, constitutes a good way to solve the latter two problems and to improve the precision, and especially the recall, of the previous query [9].

IQBE was proposed in [12] as ''a process in which searchers provide sample documents (examples) and the algorithms induce (or learn) the key concepts in order to find other relevant documents''. This method is a process for assisting the users in the query formulation process performed by machine learning methods. It works by taking a set of relevant (and optionally, non-relevant documents) provided by a user and applying an off-line learning process to automatically generate a query describing the user's needs.

Smith and Smith [13] propose a GA for learning queries for Boolean IRSs. Although they introduce it as a relevance feedback algorithm, the experimentation is actually closer to the IQBE framework. The algorithm components are described as follows:

- The Boolean queries are encoded in expression trees, whose terminal nodes are query terms and whose inner nodes are the Boolean operators AND, OR and NOT.
- Each generation is based on selecting two parents, with the best fitted having a larger chance to be chosen, and generating two offspring from them. Both offspring are added to the current population which increments its size in this way.
- The usual GA crossover is considered [14]. No mutation operator is applied.
- The initial population is generated by randomly selecting the terms included in the set of relevant documents provided by the user, having those present in more documents a higher probability of being selected.
- The fitness function gives a composite retrieval evaluation encompassing the two main retrieval parameters (precision and recall).

Yang and Korfaghe [15] propose a similar GA to that of Robertson and Willet's [16]. They use a real coding with the two-point crossover and random mutation operators (besides, crossover and mutation probabilities are changed throughout the GA run). The selection is based on a classic generational scheme where the chromosomes with a fitness value below the average of the population are eliminated, and the reproduction is performed by Baker's mechanism.

4. System Framework

4.1. Building an IR System

The proposed system is based on Vector Space Model (VSM) in which both documents and queries are represented as vectors. Firstly, to determine documents terms, the following procedure is used:

- Extraction of all the words from each document.
- Elimination of the stop-words from a stop-word list generated with the frequency dictionary of Kucera and Francis [17].
- Stemming the remaining words using the porter stemmer, which is the most commonly used stemmer in English [4][18].

After using this procedure, the final number of terms was 6385 for the CISI collection, 7126 for CACM and 7772 for NPL. After determining the terms that described all documents of the collection, the weights were assigned by using the formula proposed by Salton and Buckley [19]:

$$a_{ij} = \frac{\left(0.5 + 0.5\frac{tf_{ij}}{\max tf}\right) \times \log\frac{N}{n_i}}{\sqrt{\left(0.5 + 0.5\frac{tf_{ij}}{\max tf}\right)^2 \times \left(\log\frac{N}{n_i}\right)^2}} \longrightarrow \quad (2)$$

Where a_{ij} is the weight assigned to the term t_j in document D_i, tf_{ij} is the number of times that term t_j appears in document D_i, n_j is the number of documents indexed by the term t_j and finally, N is the total number of documents in the database. Finally, the vectors are normalize by dividing them by their Euclidean norm. This is according to the study of Noreault *et al.*[17], of the best similarity measures which make angle comparisons between vectors. A similar procedure is carried out with the collection of queries, thereby obtaining the normalized query vectors. Then, the following steps are applied:

- For each collection, each query is compared with all the documents, using the cosine similarity measure. This yields a list giving the similarities of each query with all documents of the collection.
- This list is ranked in decreasing order of similarity degree.
- Make a training data consists of the top 15 document of the list with a corresponding query.
- Automatically, the keywords (terms) are retrieved from the training data and the terms which are used to form a binary query vector.
- Adapt the query vector using the genetic approach.

4.2 The Genetic Approach

Once significant keywords are extracted from training data (relevant and irrelevant documents) including weights are assigned to the keywords. The binary weights of the keywords are formed as a query vector, and the adapting of the query vector as a chromosome. Then, the GA is applied to get an optimal or near optimal query vector, and the result of GA approach is compared with the classical IRS without using GA. Ga's are characterized by 5 basic components as follows:

- Chromosome representation for the feasible solutions to the optimization problem.
- Initial population of the feasible solutions.
- A fitness function that evaluates each solution.
- Genetic operators that generate a new population from the existing population.
- Control parameters such as population size, probability of genetic operators, number of generation.

4.2.1. Representation of the chromosomes

These chromosomes use a binary representation, and are converted to a real representation by using a random function. We will have the same number of genes (components) as the query and the feedback documents have terms with non-zero weights. The set of terms contained in these documents and the query is calculated. The size of the chromosomes will be equal to the number of terms of that set, we get the query vector as a binary representation and applying the random function to modify the terms weights to real representation. Our GA approach receives an initial population chromosomes corresponding to the top 15 documents retrieved from classical IR with respect to that query.

4.2.2. Fitness Function

Fitness function is a performance measure or reward function, which evaluates how each solution, is good. In our work, we use the cosine similarity as fitness function with equation:

$$\frac{\sum_{i=1}^{t} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{t} x_i^{\,2} \cdot \sum_{i=1}^{t} y_i^{\,2}}} \longrightarrow \qquad (3)$$

4.2.3. Selection

As the selection mechanism, the GA uses ‘‘simple random sampling’’ [20][6]. This consists of constructing a roulette with the same number of slots as there are individuals in the population, and in which the size of each slot is directly related to the individual’s fitness value. Hence, the best chromosomes will on average achieve more copies, and the worst fewer copies. Also, uses the ‘‘elitism’’ strategy [21], as a complement to the selection mechanism. If after generating the new population, the best chromosome of the preceding generation is by chance absent, the worst individual of the new population is withdrawn and replaced by that chromosome.

4.2.4. Operators

In our GA approaches, we use two GA operators to produce offspring chromosomes, which are:

Crossover is the genetic operator that mixes two chromosomes together to form new offspring. Crossover occurs only with crossover probability P_c. Chromosomes are not subjected to crossover remain unmodified. The intuition behind crossover is exploration of a new solutions and exploitation of old solutions. GAs construct a better solution by mixture good characteristic of chromosome together. Higher fitness chromosome has an opportunity to be selected more than lower ones, so good solution always alive to the next generation. We use a single point crossover, exchanges the weights of sub-vector between two chromosomes, which are candidate for this process.

Mutation is the second operator uses in our GA systems. Mutation involves the modification of the gene values of a solution with some probability P_m. In accordance with changing some bit values of chromosomes give the different breeds. Chromosome may be better or poorer than old chromosome. If they are poorer than old chromosome they are eliminated in selection step. The objective of mutation is restoring lost and exploring variety of data.

5. Experiments and Results

To perform the trial, it was first necessary to generate test databases. We created these from three of the best known test collections: the CISI collection (1460 documents on information science), the CACM collection (3204 documents on Communications), and finally the NPL collection (11,429 documents on electronic engineering). One of the principal reasons for choosing more than one test collection is to emphasize and generalize our results in all alternative test documents collections. The experiments are applied on 100 queries chosen according to each query which does not retrieve 15 relevant documents for our IR system. From our experimental observation, the best values for this test documents collections at crossover probability P_c = 0.8 and mutation rate is P_m= 0.7 for GA. The following are the results of applying GA for 100 generations.

5.1. The CISI Documents Collection

The results for the GA are shown in table (1) and figure 1, using non-interpolated average Recall – Precision relationship. From this table and the corresponding figure we notice that GA gives a high improvement than Classical IR system with 11.9%. Also, the average number of terms of query vector before applying GA is 509.61 terms; these terms are reduced after applying GA to 358.84 terms as average number.

Table (1): Shows the experimental results of applying GA on CISI Collection

Average nine-point Recall Precision for 100 query in CISI Collection			
Recall	Precision		GA Improvement %
	Classic IR	GA	
0.1	0.679345	0.877	29.09493703
0.2	0.557805	0.658205	17.99912156
0.3	0.461991	0.584501	26.5178326
0.4	0.400701	0.444153	10.8439959
0.5	0.349373	0.403625	15.52838943
0.6	0.303939	0.310678	2.217221219
0.7	0.25167	0.264587	5.132514801
0.8	0.198868	0.192231	-3.337389625
0.9	0.149076	0.153811	3.176232257
Average	0.37253	0.432088	11.90809502

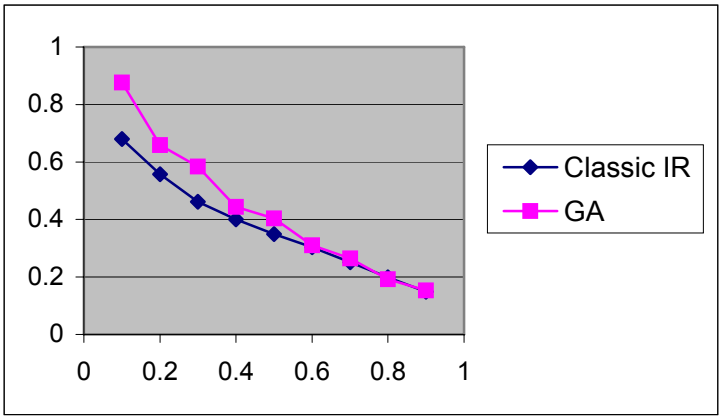


Figure 1. Represents the relationship between average recall-precision for 100 queries on CISI

5.2. The NPL Documents Collection

The results for this experiment are shown in table (2) and figure 2, using non-interpolated average Recall-Precision relationship. From this table and the corresponding figure, we find that the GA gives a higher improvement than classic IR system with 11.5% as average values. Also, the average number of terms of query vector before applying GA is 134.14 terms; these terms are reduced after applying GA to 16.8 terms.

Table (2): Shows the experimental results of applying GA on NPL Collection

Average nine-point Recall Precision for 100 query in NPL Collection			
Recall	Precision		GA Improvement %
	Classic IR	GA	
0.1	0.732924	0.802594	9.505760488
0.2	0.503371	0.550861	9.43439332
0.3	0.435145	0.47663	9.533603741
0.4	0.340466	0.385739	13.29736303
0.5	0.313329	0.363884	16.13479761
0.6	0.239987	0.274637	14.43828207
0.7	0.215388	0.238706	10.82604416
0.8	0.174277	0.193571	11.07088141
0.9	0.145552	0.159113	9.316945147
Average	0.344493	0.382859	11.50645233

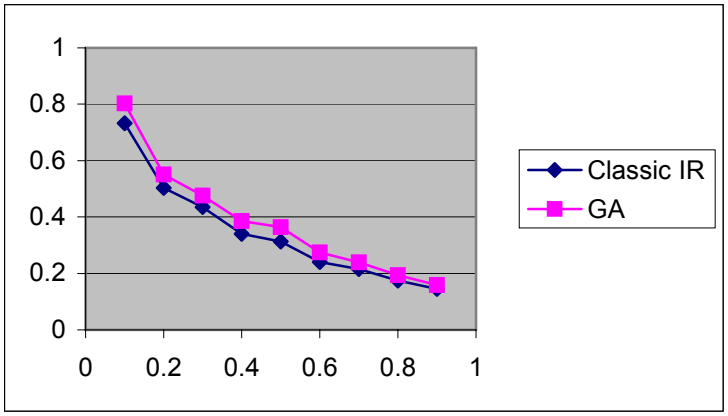


Figure 2. Represents the relationship between average recall-precision for 100 queries on NPL

5.3. The CACM Documents Collection

The results for this experiment are shown in table (3) and figure 3, using non-interpolated average Recall–Precision relationship. From this table and the corresponding figure, we notice that GA gives a higher improvement than that with classic IR system 5.13%, as average values. Also, the average number of terms of query vector before applying GA is 160.7 terms; these terms are reduced after applying GA to 16.83 terms.

Table (3): Shows the experimental results of applying GA on CACM Collection

Average nine-point Recall Precision for 100 query in CACM Collection			
Recall	Precision		GA Improvement %
	Classic IR	GA	
0.1	0.722666	0.788457	9.103929063
0.2	0.416461	0.473779	13.76311347
0.3	0.369358	0.419403	13.54918534
0.4	0.246728	0.273892	11.00969489
0.5	0.212679	0.224543	5.578359876
0.6	0.158008	0.162339	2.741000456
0.7	0.142905	0.139693	-2.247647038
0.8	0.107276	0.104937	-2.180357209
0.9	0.089651	0.085076	-5.103122107
Average	0.27397	0.296902	5.134906305

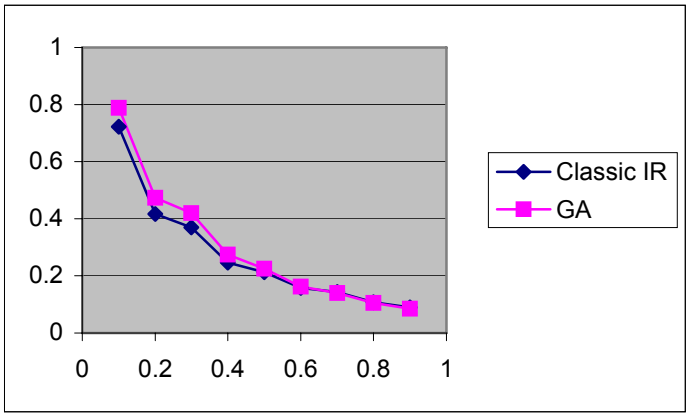


Figure 2. Represents the relationship between average recall-precision for 100 queries on CACM

6. Conclusion

From the previous results, it can be concluded that our GA approach gives more sophisticated results than classical IR system in our test collections. Also, the average number of terms is reduced after applying GA. The experiments developed use three of the relative document collections (CACM, CISI and NPL), and compare the results of two variant systems (Classical IR and GA). The latter algorithm achieves the best performance and it obtains better precision than the first approach. Our GA approach used to adapt the weights of query terms is to get high precision results.

References

[1] Holland, J. H., *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor (1975).

[2] K. A. DeJong, "An Analysis of the Behavior of a Class of Genetic Adaptive Systems", Ph.D. Thesis, University of Michigan (1975).

[3] V. V. Raghavan and B. Agrwal. "Optimal determination of user – oriented clusters: An application for the reproductive plan". In *Proceedings of the second conference on genetic algorithms and their applications*, Hillsdale, NJ (pp. 241-246), 1987.

[4] Baeza-Yates, R. and Ribeiro-Neto, B., *Modern Information Retrieval*, Adisson, 1999.

[5] R. Korfhage Robert. *Information storage and Retrieval*. John Wiley & Sons, Inc. 1997.

[6] Goldberg, D. E., *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, MA. (1989).

[7] Z. Michalewicz. *Genetic algorithms + data structures= evolution programs*. Berlin: Springer- verlag, 1995.

[8] Salton, G. and McGill, M. H., *Introduction to Modern Information Retrieval*, McGraw-Hill, 1983.

[9] Van Rijsbergen, C. J., *Information Retrieval*, second ed., Butterworth, 1979.

[10] A. Bookstein. "Outline of a general probabilistic retrieval model", *Journal of Documentation* 39 (2) 63–72 (1983).

[11] N. Fuhr. "Probabilistic models in information retrieval", *Computer Journal* 35 (3) 243–255 (1992).

[12] H. Chen et al., "A machine learning approach to inductive query by examples: an experiment using relevance feedback, ID3, genetic algorithms, and simulated annealing", *Journal of the American Society for Information Science* 49 (8) 693–705 (1998).

[13] M.P. Smith, M. Smith. "The use of genetic programming to build Boolean queries for text retrieval through relevance feedback", *Journal of Information Science* 23 (6) 423–431 (1997).

[14] J. Koza. "Genetic Programming". *On the Programming of Computers by means of Natural Selection*, The MIT Press (1992).

[15] J. Yang and R. Korfhage. "Query modifications using genetic algorithms in vector space models", *International Journal of Expert Systems* 7 (2) 165–191 (1994).

[16] A.M. Robertson and P. Willet. "Generation of equifrequent groups of words using a genetic algorithm", *Journal of Documentation* 50 (3) 213–232 (1994).

[17] T. Noreault, M. McGill and M. B. Koll. "A performance evaluation of similarity measures, document term weighting schemes and representation in a Boolean environment". Information retrieval research. London: Butterworths (1981).

[18] M. F. Porter. "An algorithm for suffix stripping. Program", 14(3), 130–137 (1980).

[19] G. Salton and C. Buckley. "Improving retrieval performance by relevance feedback". *Journal of the American Society for Information Science*, 41(4), 288–297 (1990).

[20] M. Gordon. "Probabilistic and genetic algorithms in document retrieval". *Communications of the ACM*, 31(10), 1208–1218 (1988).

[21] D. C. Blair. "Language and Representation in information Retrieval". Amsterdam: Elsevier, 1990.

Authors' Information

A. A. Aly – Dept. of Computer Science, Faculty of Science, Minia University, El-Minia, Egypt; Email: abdelmgeid@yahoo.com

SYSTEMOLOGICAL KNOWLEDGE-BASED TECHNOLOGY FOR SOLVING COMPLETE ILL-STRUCTURED PROBLEMS

Dmitriy Elchaninov, Mikhail Bondarenko, Kateryna Solovyova, Sergey Matorin, Vasy Matorin

Abstract: Original method and technology of systemological «Unit-Function-Object» analysis for solving complete ill-structured problems is proposed. The given visual grapho-analytical UFO technology for the fist time combines capabilities and advantages of the system and object approaches and can be used for business reengineering and for information systems design. UFO- technology procedures are formalized by pattern-theory methods and developed by embedding systemological conceptual classification models into the system-object analysis and software tools. Technology is based on natural classification and helps to investigate deep semantic regularities of subject domain and to take proper account of system-classes essential properties the most objectively. Systemological knowledge models are based on method which for the first time synthesizes system and classification analysis. It allows creating CASE-toolkit of a new generation for organizational modelling for companies' sustainable development and competitive advantages providing.

Keywords: systemology, conceptual classification model of knowledge, natural classification.

ACM Classification Keywords: I.2 Artificial Intelligence – I.2.6 Learning: Knowledge Acquisition

Introduction

For optimal decision making, sustainable development of humanity, companies and organizations; for competitive advantage providing new powerful methods and technologies for solving modern complex ill-structured problems in different spheres are necessary. We work with complete ill-structured (symbolic, etc.) qualitative problems in low-formalised problem domains, including systems of the first nature (e.g. social). For solution such problems and analyse such systems we need more powerful system approach and knowledge-based methods and tools. So we use new systemological methodology, which for the first represents features of system methodology of noospheric stage of science development.

This methodology makes it possible to reveal the essential properties of complex objects of an arbitrary nature, and also the reasons for their occurrence, adaptation, and development, and, consequently, to predict and manage any complex system more efficiently. The quality efficiency of this methodology determines the necessary of its use in application fields of an arbitrary nature, including for construction of a picture of the world as a whole [Bondarenko et al, 1996].

Systemology as modern promising system methodology does not regard system as a set but as a functional object which function is assigned by hypersytem. Systemology in particular allows overcoming problems of traditional methods of system analysis at the expense of using conceptual knowledge as well as formalizing procedures of analysis and synthesis of complex systems and creating knowledge-oriented software tools for their simulation [Matorin, 2001a], [Matorin, 2001b], [Matorin, 2001c].

Nowadays there is a variety of information systems and technologies for supporting decision takers in various ways. These technologies and systems include in particular technologies for system analysis and information systems that it automates and represent software CASE-instruments for business systems and business processes simulation. We can give here only few aspects of new technology based on functional systemology.

Let us consider some results of creating original method of systemological «Unit-Function-Object» analysis based on systemology and conceptual models of knowledge [Matorin et al, 2005], [Matorin and Elchaninov, 2002].

Technology of Systemological UFO Analysis

Complexity of business, managerial and production problems is growing constantly. This leads to the fact that at the present stage competitive activity of organizations and making optimal decisions become impossible without designing and reorganizing business processes (business reengineering). To solve the problems of reengineering visual graph-analytical methods of analysis and systems simulation are used. They are the varieties of methods of the system-structural analysis (e.g. SADT, DFD) as well as object-oriented analysis (OOA) using UML. Employment of these methods has been automated with the help of CASE-means (e.g. BPwin, Rational Rose).

At present according to most analytical practitioners and consultants the efficiency of existing traditional CASE-means does not comply with modern requirements, e.g.:

– Using the toolkit for automation of system-structural analysis poorly supports development of object-oriented software applications (most popular and in great demand nowadays), that complicates the choice;
– Using the existing simulation toolkit leads to extra diversity in presenting organizational models;
– Using the toolkit that automates object-oriented analysis methods does not allow adequate representation of business processes in organizations.

CASE-means possibilities are always limited by built-in systems simulation methods. For that matter and for the above mentioned problems activities aimed at creation of theoretical, methodological and instrumental tools for complex systems simulation free from the above drawbacks have been carried out. Investigations in this area have made it possible to develop new ingenious method and algorithm for analysis and simulation of the systems that allow representing the system as a triennium structure "Unit -Function-Object":

– "Unit" is the cross point of input and output connections (flows of any nature) in the structure of simulated system;
– "Function" is the process of conversion of input into output, i.e. the process ensuring balance of "influent" and "effluent" flows of the given node;
– "Object" is the substance, which implements the given function.

The given method (UFO-analysis), which for the fist time combines capabilities and advantages of the system and object approach, is designed for project implementation in business reengineering and may be oriented for simulation and designing information and engineering systems with a help of unique CASE-toolkit ("UFO-toolkit").

Developments in the systemological approach [Bondarenko et al, 1996] to poorly formalized subject areas have led to a new systems theory that overcomes the set-theoretic approach to systems and instead describes specific system properties and relations [Matorin, 2001a]. The concepts from that theory give a formal semantic system with an adaptive alphabet, which serves as the basis of a new systems analysis method (UFO-analysis), which agrees with the requirements and procedures of object-oriented analysis and design (OOAD) in information systems [Matorin, 2001b], [Matorin, 2001c].

This agreement has been obtained by solving the following auxiliary problems:

– setting up a method of identifying a class set as required and suitable for OOAD;
– introducing conceptual classification simulation (CCS) procedures into systems technology.

Both of these amount to setting up a universal conceptual classification model for any subject area. That model constitutes a classification of all the components and properties of a system. The abstract level is considered as a basic one and specifies the category structure, which determines the method of selecting the classes needed for the simulation. The particular (sheet) classes in this classification scheme (and also examples of them) are used to simulate a particular subject area. They also provide the alphabet for the normative system in the new systemological analysis method.

The basic classification is a high-level conceptual classification model (taxonomic and parametric) for forms of system considered as flow objects and also for forms of their functions and structures (fig. 1).

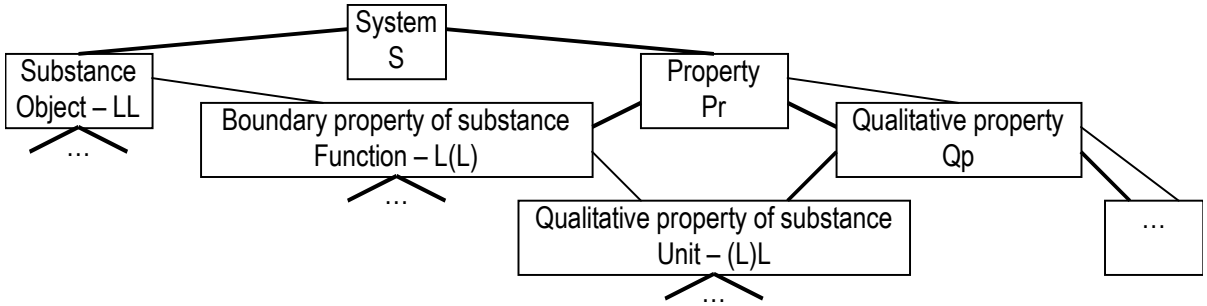


Fig. 1. Basic classification

This classification is a modification of the [Matorin, 2001a] scheme, and the “system” concept is at a limiting level of abstraction, and thus as in [Matorin, 2001a] represents a unique category (root class).

It is best to distinguish primarily objects having boundary properties in the complete system set, of which the most important are the functional properties or “functions”. At the same time, it is best to distinguish properties characterizing qualities (“qualitative properties”).

In systemology, the properties of a system are understood as signs of its activity and are included in these connections and exchange flows with other systems in the supersystem structure. This means that one analyzes system properties by means of functional systemology in terms of an integral object "based primarily on the observation of those fluxes in which it is included as an element of a superobject, i.e., as a flow element in a network of closed exchange flows in the superobject. Naturally, these qualities will be observed simultaneously and a sufficiently complete characteristic of the function of this object is thereby provided, which gives an expression of the integrity, since in qualitative characteristics one cannot get a balance between the incoming and outgoing flows in that case" [Melnikov, 1988].

It is thus best to consider as qualitative properties of a substance or object its capacity to be a unit in a certain structure. The capacity to function at that unit supports the balance between influx and efflux on the incoming and outgoing links, i.e., the capacity to transform the elements of the incoming links into elements of the outgoing ones, which will constitute boundary properties of the object or system (formal ones, in the present case functional ones).

The objects in this hierarchy are thus classified in accordance with the functions they perform. From the formal logic viewpoint and on the basis of the structural properties of that classification, one can write a genus-species definition: “An object is a system characterized by a function”. This completely corresponds with the systemological concept of a system as an object with functions on the basis that a system is always also a flow object. Then also in accordance with the systemological approach, a function of an object is considered as one of its internal determinants, i.e., the cause of certain internal substance characteristics of it.

The functions in this hierarchy are classified in accordance with the flow units that they constitute in the system structure. Consequently, from the viewpoint of formal logic and on the structural properties of that classification, one can write a genus-species definition: “A function is a property or capacity to be a unit in a system structure”. This means that a function is characterized by the elements in the incoming and outgoing links on which flows are exchanged with other objects. The region of definition for that function is the set of elements forming the input flows (links), and also the region of values, namely the set of elements forming the outgoing flows (links).

In accordance with the systemological approach, the flow (link) characteristics of a unit in the supersystem structure are considered as an external determinant for the system, i.e., the cause of a certain internal determinant or current function.

This scheme resembles the [Matorin, 2001a] one in constituting a taxonomic parametric classification, in which the objects are classified in accordance with their properties, forming part of the general class hierarchy. This means that in this hierarchy, as in [Matorin, 2001a], for each object (class or concept) there is not only a generic feature (higher-level class) but also a concept (class) having a species difference in content for the concept, i.e., the property of the object. The dashed lines in fig. 1 show the relationship of a concept with its species difference (class and properties). As this classification is parametric, it enables one to classify system species in accordance with their property species, i.e., to incorporate the natural classification laws [Solovyova, 1999] as metaknowledge. Conceptual classification models of knowledge are based on systemological method (systemological classification analysis) [Solovyova, 1999], which for the first time synthesizes system and classification analysis on the basis of the new criteria of natural classification. They help to investigate deep semantic regularities of subject domain and to take proper account of system-classes essential properties the most objectively. It allows receiving the forecasting parametrical classifications and effective ontologies in different ill-structured problem domains also.

If one considers a system as a flow object whose function is due to a function in a superobject (supersystem), one can represent any particular system S* in terms of a class hierarchy, as in [Matorin, 2001a], with that system being an example of class S, and put formally as the triple S* = <LL, L(L), (L)L>, in which (L)L is a particular unit in the supersystem structure ((L) is an incoming link, L an outgoing link), with L(L) the class of functions that balance the given unit ((L) is the argument and L the function), and LL is the class of objects that realize the given functions (L object input, L object output). We call this triple a UFO-element.

Here L denotes the type of elements with a material or information nature (L = {M, I}, with M a material element and I an information one; M = {V, E}, where V is a substance element and E an energy one; and I = {D, C}, where D is of data nature and C is of control nature), which applies at a certain level in the connected systems, and which these systems use to exchange data through their links, which are considered as flows (fig. 2).

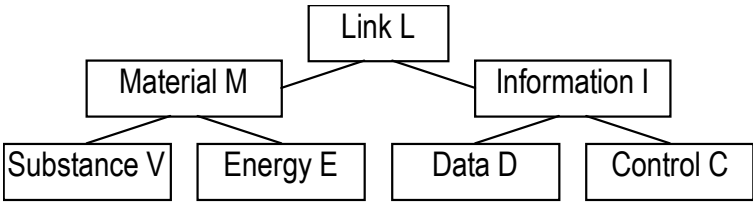


Fig. 2. Initial link classification

This representation of a system as a phenomenon corresponding to an example of a UFO element agrees well with the concept of “generator” (“generatrix”) in pattern theory [Matorin and Elchaninov, 2002]. By “generator” in that theory [Grenander, 1979] one understands an object having certain features α, as well as the incoming and outgoing links (in turn characterized by certain parameters β). In our case (fig. 3), one can consider the examples of classes LL and L(L) as features of the generator, while (L) and L are links whose parameters are of type L (fig. 2). Then the generator g_i as an example of UFO element takes the form $g_i = <L_2^i L_1^i, L_2^i (L_1^i), (L_1^i)L_2^i>$.

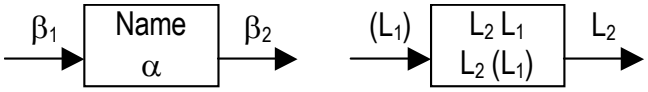


Fig. 3. Graphical formalisms – generator and UFO-element

Pattern theory involves the assumption that is a source that generates the set of generators G = {g_i}. In our case, that source generates UFO elements (generators) and is shown in fig. 1 as part of the basic classification.

The hierarchy there constitutes the basis of a normative system in the new method of systemological analysis. The alphabetic symbols are UFO elements representing particular classes (sheets) in the classification scheme, and also the corresponding examples of those classes. The examples of UFO elements are used here to construct system-object models for systems and also to simulate the functioning of these models.

This alphabet consists of characters for the units, functions, and objects representing a system from the viewpoint of its structural, functional, and substance characteristics respectively. Isomorphism in classifying those three (in accordance with the laws of natural classification) enables one to consider a single combined hierarchy containing UFO elements. A UFO element is a system that corresponds to a certain unit (intersection of links or flows) in the supersystem structure, with a defined function (in general not unique) that balances the flows at that unit and a defined object (in general, not unique for each function) that realizes the given function.

Incorporating the natural classification laws and criteria into the basic classification gives an unambiguous formal-logic definition for each object in terms of the genus and the species difference.

Fundamental constructive criteria and a method of natural classification formulated on the basis of the study and modelling of the regularities of the structure of conceptual system, which reflect the systematic character of reality (a subject field) you can find in details in [Solovyova, 1999]. For example one of them is: ”Essential supporting properties of any system are essential properties of its subsystem and species of its essential functional property. The concepts of essential supporting properties of a system are the concepts of essential functional properties of its subsystems – the species of the concept of the essential functional property of the system”.

On the basis of these criteria, it is possible for any subject field: to construct a classification reflecting essential properties of objects and different aspects of considering them; to estimate any classification in terms of the degree of inclusion of essential properties; to define errors and inaccuracies of any classification and specify their sources and possible remedies; to predict the evolution of any classification; to correct the terminology of subject fields.

A conceptual system is limited to the category and individual concepts. A hierarchy of systems is limited to the supersystem and specific systems. Theses facts should be taken into account in the application of the criteria. A through analysis of a natural classification has established the validity of the criteria for concrete systems as well (individual concepts) if subsystems are defined at functional components of the systems. This fact confirms the important assertion that the classification by “part-whole” principle for concrete systems (functional parts) and the generic-specific classification for external systems (functional genera and species) are essentially analogous and can be viewed in the framework of an integrated natural classification of class systems according to their essential functional properties.

Classification scheme thus acts as an algorithm for the semantics of the normative-system signs, which transforms that system into an algorithmically constructed one. Consequently, this scheme provides an alphabet for that system having not only a completely abstract or strictly mathematical semantics but also an object-oriented form, so that alphabet can be considered as a formal-semantic one, while the normative system is also formally semantic. This implements the current suggestions of informatics experts, who consider that there has

long been a practical need to transfer from formal mathematical analysis of information phenomena to content analysis and the driving force in self-organizing systems of social nature.

Using a classification to generate alphabetic characters that constitute examples of the corresponding classes provides, as in [Matorin, 2001a] a distinctive feature for the alphabet. An ordinary formal system with a finite or infinite formal alphabet uses a finite and restricted number of initial concepts corresponding to the signs in the alphabet. This means that completely different subject areas are simulated by means of the same set of alphabetic characters in the normative system of some traditional system analysis method. On the other hand, specifying an alphabet by means of a classification can change the composition of the initial concepts and the corresponding alphabetic characters, i.e., allows one to adapt the alphabet and the normative system in accordance with the subject area.

This hierarchy (units, functions, and objects) enables one to use a detailed set of simulation facilities, i.e., alphabetic characters, to handle a particular task. For example, to simulate information business (such as organizations in the media), one can distinguish particular classes related to the data links and classes related to substances and energy, as well as ones in the form of abstract classes; to simulate electrical power businesses, it is necessary to specify the classes joined by energy links; and to simulate the transportation companies, one uses classes with substance links; and also to simulate production, one uses classes that simulate obtaining substances of appropriate form. The only thing overlooked here is a principle in accordance with which the properties of objects (functions and units) used to set up an object model are determined by their parametric taxonomic classification, i.e., by the basic class hierarchy incorporating those properties.

Specialization in the UFO element class hierarchy gives models defining the alphabetic elements for systems that incorporate more detailed forms of link on account of the consideration of the forms of material link: substance V and energy E, and also forms of information link: data D and control C. As a result of such specialization at this level, one can obtain for example the following classes of UFO elements (fig. 4), which are defined by the corresponding units and are considered as nonintersecting classes of generators constituting set G.

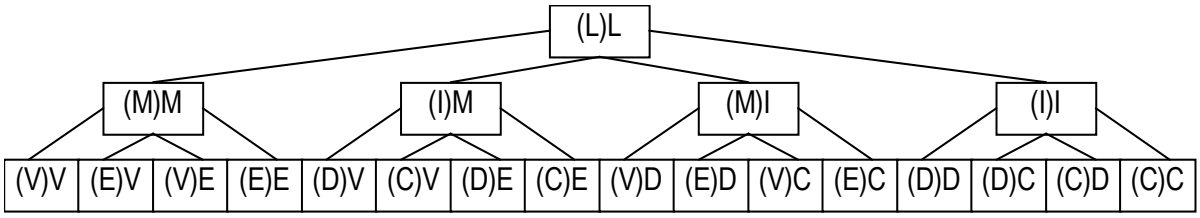


Fig. 4. Classification of alphabetic UFO elements by their units

The UFO analysis algorithm has the following basic steps [Matorin, 2001c]:

– identifying units and links from the functional links of the system as a whole;
– identifying functions performed by the observed units;
– identifying the objects corresponding to this functionality.

The UFO hierarchy adaptation for the initial steps in the algorithm consists in specializing it as regards particular units, functions, and objects in the subject area, i.e., it produces an ontology model. The UFO element specifications in that case will contain the following information:

– For the object (substance parameters): engineering and working characteristics (design, climatic and mechanical working conditions, reliability, necessary and available stocks of energy, materials, and

information, as well as productivity and so on); attributes and quantitative characteristics of input-output ports (including flow capacity) and also the cost and working time. Here it may be useful to rank objects on various features.

– For a function (process parameters): description of the input to output conversion, i.e., the functional protocol; formal description of the functional dependence if it exists and is necessary in the form of a script or macro. In essence, this represents data on the internal determinant of the corresponding system.
– For a unit (structure parameters): type of link. In essence, this represents data about different cases in the use of the object, i.e., on the external determinant of the corresponding system.

If the alphabetic elements are software objects realized in the form of ready-made classes, one can say that UFO analysis is a component of the appropriate technologies under CORBA (Business Object Facility – BOF). In the latter case, the CASE software facility that automates the UFO analysis may function within the framework of the business object component architecture (BOCA) as a framework, which works as a tool for linking business objects into a system and provides a form of convenient working points for carrying out the tasks imposed on it.

To compare the proposed method and developed toolkit against most popular methods and toolkits special comparative simulation has been carried out using SADT/IDEF0 (BPwin), UML (Rational Rose) and UFO-analysis (UFO-toolkit) of the real business process "passing of the contract at the enterprise", as described e.g. at www.devbusiness.ru. Comparative research of methods and toolkits for business-systems simulation have been carried out on known criteria of leading analysts (e.g. G. Kalyanov), and such tool-kit should provide: recording of information on business processes; producing high-level representations of business processes; maintenance of repository; control of business processes description syntax; control of its completeness and model consistency; analysis and verification of processes description and generation of corresponding reports; producing specifications of business processes. Comparison of methods and toolkits for business modelling is presented in the table.

№	Comparison criterion	IDEF0/ BPwin	UML/ Rational Rose	UFO-analysis/ prototype
1	Possibility to orient methodology and toolkit at subject area	-	-	+
2	Possibility to reduce diversity of models representation	-	-	+
3	Possibility of syntactic and semantic control of business system and business process description	+/-	-	+
4	Availability and maintenance of depositary (library)	-	-	+
5	Possibility of component simulation technology support	-	-	+
6	Possibility of software object-oriented design support	-	+	+
7	Number of models types (diagrams) to be designed	3	4-5	1
8	Possibility of models construction automation	-	-	+
9	Simplicity/complexity of using methodology	+	+/-	+/-

If on the other hand technical objects are considered as alphabetic elements, then the UFO analysis will be matched to the CALS technology.

Conclusion

The basic class hierarchy has been used in proposing an alphabet that can be used to construct system-object models for example for organizations. That alphabet together with the system decomposition rules can constitute a formal semantic normative system, whose features are as follows:

– It objectivizes object and class decomposition for the system;
– It provides a range of functional objects (alphabetic symbols) for each form of system in accordance with a unified principle;
– It provides for simulating by computer the properties of the classes and the examples of object models.

That normative system allows one to formulate the UFO analysis algorithm for a system analysis method in which the procedures and results for the first time agree with the OOAD requirements.

Method of systemological classification analysis on the basis of natural classification has advantages in comparison with other method and has been used for conceptual and ontology's modeling and in artificial intelligence systems. Given UFO-method and tool as stated above have a number of advantages in comparison with existing business simulation methods and CASE-means.

Bibliography

[Bondarenko et al, 1996] M.F. Bondarenko, S.I. Matorin, and E.A. Solov'eva. Analysis of systemological tools for conceptual modelling of application fields. Automatic Documentation and Mathematical Linguistics New York: Allerton Press Inc., 1996, vol. 30, No. 2, pp. 33-45.

[Matorin, 2001a] Matorin S.I. An Object-Oriented Approach to Systems Theory: Formalisation and Interfacing Prospects. Automatic Documentation and Mathematical Linguistics. New York: Allerton Press Inc., 2001, Vol. 35, No. 4, pp. 30-37.

[Matorin, 2001b] Matorin S.I. A New Method of Systemological Analysis Co-ordinated With the Object-Oriented Design Procedure. I. Cybernetics and Systems Analysis. Plenum Publishing Corporation, 2001, Vol. 37, No. 4. pp. 562-572.

[Matorin, 2001c] Matorin S.I. A New Method of Systemological Analysis Co-ordinated With the Object-Oriented Design Procedure. II. Cybernetics and Systems Analysis. Plenum Publishing Corporation, 2001, Vol. 38, No. 1. pp. 100-109.

[Matorin et al, 2005] Matorin S.I, Popov S.S. and Matorin V.S. Modelling of the Organizational Systems Using the New Approach "Unit-Function-Object". Automatic Document and Mathematical Linguistics. New York: Allerton Press Inc., 2005, Vol. 39. No. 1, pp. 1-8.

[Matorin and Elchaninov, 2002] Matorin S.I. and Elchaninov D.B. Application of the Pattern Theory for Formalization Systemological UFO-analysis. Automatic Documentation and Mathematical Linguistics. New York: Allerton Press Inc., 2002, Vol. 36, No. 6, pp. 1-12

[Melnikov, 1988] Melnikov G.P. Systemology and Linguistic Aspects of Cybernetics. New York, Paris, Monterey, Tokyo, Melbourne: Gordon and Breach, 1988.

[Solovyova, 1999] Solovyova E.A. Mathematical and Systemological Foundations of Natural Classification. Automatic Documentation and Mathematical Linguistics. New York: Allerton Press Inc., 1999, Vol. 33, No. 4, pp. 42-50.

[Grenander, 1979] Grenander W. Lectures on Pattern Theory. Part 1. Pattern Synthesis [In Russian], Mir, Moscow, 1979.

Authors' Information

Dmitriy Elchaninov - Social Informatics Department, Kharkov National University of Radio Electronics, Lenin Ave., 14, Kharkov, 61166, Ukraine; e-mail: elchaninov@kture.kharkov.ua

Mikhail Bondarenko - Rector, Kharkov National University of Radio Electronics, Lenin Ave., 14, Kharkov, 61166, Ukraine; e-mail: rector@kture.kharkov.ua

Kateryna Solovyova - Social Informatics Department, Kharkov National University of Radio Electronics, Lenin Ave., 14, Kharkov, 61166, Ukraine; e-mail: si@kture.kharkov.ua

Sergey Matorin - Social Informatics Department, Kharkov National University of Radio Electronics, Lenin Ave., 14, Kharkov, 61166, Ukraine; e-mail: si@kture.kharkov.ua

Vasy Matorin - Social Informatics Department, Kharkov National University of Radio Electronics, Lenin Ave., 14, Kharkov, 61166, Ukraine; e-mail: si@kture.kharkov.ua

EVOLUTIONARY TECHNIQUE OF SHORTER ROUTE DETERMINATION OF FIRE BRIGADE FOLLOWING TO FIRE PLACE WITH THE OPTIMIZED SPACE OF SEARCH

Vitaliy Snytyuk, Oleksandr Dghulay

Abstract: In this paper the technique of shorter route determination of fire engine to the fire place on time minimization criterion with the use of evolutionary modeling is offered. The algorithm of its realization on the base of complete and optimized space of search of possible decisions is explored. The aspects of goal function forming and program realization of method having a special purpose are considered. Experimental verification is executed and the results of comparative analysis with the expert conclusions are considered.

Keywords: Fire, Evolutionary techniques.

ACM Classification Keywords: H.4. Information systems applications

Introduction

The search of shorter passage is the discrete optimization problem. Thus determination of optimum route of fire engine following to the fire place has aspects which select him from the general row of such problems. So, practically, it is a problem which solves in critical terms, the still human lives rely on the rightness of its decision. The right chosen route is the necessary condition of prevention of technical and ecological catastrophes. In the conditions of resources deficit of time minimization for fire engine route is the deciding factor of prevention of negative consequences of fire. The far of scientific researches is devoted to solving this task.

The grounds of fire engine departure route are traditionally offered, coming from the criterion of minimization of arrival time of fire brigade to the fire place. In this paper [Pryanichnikov, 1988] the analysis of factors affecting emergency safety of roads is executed: width of carriageway and pavements, quantities of carriagestrips, radius of curvature, visibility, intensity of transport streams. It is offered to determine the coefficient of road terms on a formula:

$$D = \left[\sum_{i=1}^{n} \left(\prod_{j=1}^{m} k_{ij} \right) L_i \right] / L, \qquad (1)$$

where n – quantity of route areas, m – number of factors determining road terms k_{ij} – coefficient of importance of j-th factor of road terms on the i-th route area, L_i – length of i-th area L – general length of fire brigade. Departure of fire computation is assumed on a route having the most value D.

In many scientific publications similar approaches are considered. The necessity of consideration of the fixed, desirable complete set of possible routes is their failing, that practically is difficultly realized. Possibility for values varying of road terms factors importance is not foreseen, that in the conditions of road situation change, repair of road linen, weather terms results in distortion of the route supposed time. Development of adequate model of route time is needed, as dependences on meaningful factors, with possibility of its clarification and adaptation to the changing external terms.

Important to remark that development of time route model is the necessary condition of determination of shorter route for following to the fire place. A method which structurally will allow to define an optimum route is a sufficient condition. As the considered task has a combinatorial nature and, as a result, inevitable there is the problem of calculable complication of algorithm, it is necessary to foresee realization of technology which will allow to shorten the quantity of the analysed routes and optimize the calculations process.

Problem of optimum route determination for fire engine

Without restriction of generality we will assume that the structure of roads is rectangular (fig. 1). We will number every crossing in accordance with a central-radial chart. The location of fire brigade has a zero number, a most number has the farther-most "north-eastern crossing". Number of crossing is N. The matrix of distances between crossing $S=(s_{ij})_{i,j=0}^{N-1}$, where s_{ij} is distance from i-th one to the j-th crossing, corresponds with considered structure of roads. Knowing the middle rate of movement of fire brigade, the distances matrix can put in accordance the matrix of passage time between crossing $T=(t_{ij})_{i,j=0}^{N-1}$.

Factors influencing on a passage time, in a form the presentation of their values possible to divide into three groups: determined, probabilistic-statistical and subjective.

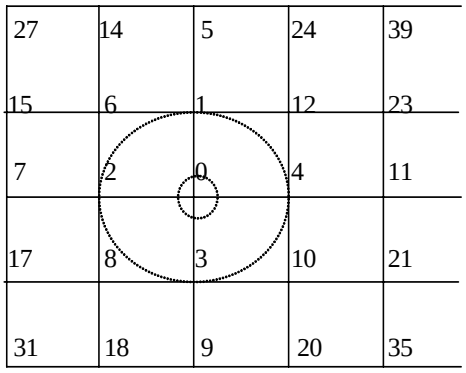


Fig. 1. Central-radial crossing numeration

The least number of crossing K on the passing way is the determined factor, its possible values – natural numbers equal to the number of quasi-concentric circumference (see the fig. 1) and increased fairly removal of crossing from fire station. Work-load of roads U is probabilistic-statistical factor which is characterized to statistical row distributing (tabl. 1), where in overhead part of table time intervals are found, in lower – relative frequencies of cars quantity on the road in these time intervals. Quality of road coverage V is a subjective factor and it is determined by the membership function, which can be both continuous or discrete. Its construction is carried out to one of two methods, the first from which is based on the pair comparisons executed by one expert [Rotshtein, 2002], second – on statistical treatment of experts group opinions [Zadeh, 1965].

Table 1

Statistical row

Intervals	$[t_0,t_1]$	$[t_1,t_2]$	...	$[t_{n-1},t_n]$
Relative frequencies	f_1	f_2	...	f_n

We will assume that the fire place H is found between two crossing n_1 and n_2. Then it is necessary to define an optimum route, that answers the decision of problem [Snytyuk, 2004]:

$$\min_t\{L_{on_1}+L_{n_1H};L_{on_2}+L_{n_2H}\} \qquad (2)$$

where L_{ij} is route from a i-th point to j-th one. The values of matrices elements S; T; $K=(k_{ij})_{i=1,\,j=1}^{N\;\;2}$ are basic data for the decision of problem (2) where k_{i1} – number of final crossing, k_{i2} – the minimal number of crossing, which is needed to pass by passage to k_{i1}; $G=(g_{ij})_{i=1,\,j=1}^{24\;\;2}$, where g_{i1} – number of time interval (days are broken-down on 24 intervals: from 0 hours to the 1-th hour (1), from 1-th to the 2-th hour (2)...), g_{i2} – relative frequencies of cars quantity in a g_{i1}-th time interval, $\sum_{i=1}^{24} g_{i2}=1;\ \sum_{i=1}^{24} g_{i2}=1;,\ \sum_{i=1}^{24} g_{i2}=1;$, where $q_{ij}\in(0,1)$ – coefficients which determine quality of road coverage on an area from the i-th crossing to j-th one. We will remark that a matrix G can have not statistical, but subjective nature. If motion at one and the same time on different areas of road is uneven, the matrix will be three-dimensional, one of measuring of which will correspond to the number of road area a. Depending on the features of concrete city or situation, the quantity of matrices of values factors affecting to the rate of movement of fire brigade can be increase. We will mark that richly in content essence of consideration of other factors will not differ from already considered.

Preconditions for solving the optimum route determination problem by use the evolutionary modeling

The features of socio-economic development of countries are the direct motive for fire number increase and, as a result, death of people and causing of property losses. A personnel and material deficit is, from one side, by the cause of the uneffective extinguishing of fires, and with other – a stimulus to introduction of new information-analytical technologies, that allows to rise efficiency of fire brigade work. **One of** problems requiring application of intellectual models and methods, there is minimization of passage time of fire brigade to the fire place.

We will define initial preconditions its solving. We will remark that such problem has some general aspects with the known problem of traveling salesman. It is known, that the exact method of the given problem solving of any dimension, excepting complete look over of all variants, does not exist. Satisfactory results give the method of branches and scopes [Luger, 2002; Zaychenko, 2000], method of successive analysis of variants [Volkovich, 1993], search of optimum route with the use of the Hopfild's neural network [Wasserman, 1992]. However by application the last method an exact result is got, approximately, in 50% calculations, exactness of the first methods relies on problems dimension, also the hit in local optimums is high-probabilistic.

The feature of optimum route search problem for fire brigade is that the best decision is searched on the criterion of time minimum. Thus it is necessary to take into account the number of crossing on the passage way, work-load of roads (mean number of cars on the road in unit time), their quality. Consideration of other factors also is possible by their special meaningfulness and necessity. We will mark that technology of determination of passage optimum way of fire brigade to the fire place will be realized taking into account subjective and statistical factors. The evolutionary method of shorter route determination is its base element and it consists in the following.

Without limitation of community we will present (2) as a problem of finding

$$\min_{t} L_{0n_1}. \qquad (3)$$

It is obvious, that for the decision of problem (2) it is necessary twice to solve (3) and execute some clarifications of result. We will carry out the search of optimum way by application evolutionary algorithm (EA) of the special kind which allows to find global optimums, in the general case, undifferential functions. We will define its basic principles and base elements.

A general population is the basic concept EA – all set of possible decisions. In our case we will define a general population as set of vectors $X = (x_0, x_1, x_2, ..., x_k, x_n)$, where x_0 – place of fire station, x_n – number of nearest crossing to the fire place. Thus, the sequence of numbers of crossing which it is necessary to pass in order that to arrive in x_n is the values of elements of vector X. We will remark that the number of crossing, in the general case, is variable. The minimum value k is determined by the number of quasi-circumference (see fig. 1), on which lies the crossing x_n, its maximal value can be enough large. All x_i, $i = \overline{0,k}$ are different and neither of them coincides with x_n. On the face of it, optimum there will be those variants at which $x_i < x_j$ for all $i < j$, but implementation of such condition is not obligatory.

Model of goal function

The adequate application EA is related to transformations of number values from two scale of notation in decimal one and vice versa. Thus there is an informative redundancy, as not all two presentations have the analogues in decimal notation. In the general case, it bring to the necessity of enlisting of additional calculable resources and increase of problem solving time [Kislyakov 2000, 2001].

Foregoing facts are indicated on considerable labour-intensiveness and pointlessness of forming of general population. About belonging to it will testify the results of verification. Determination of selective population which must be a characteristic of representation is important procedure [Goldberg, 1989; Werbos, 1974; Isaev, 2000; Jensen, 2001]. The vectors of selective population can have a different number of elements, that it is related to the number of crossing on the passage way. Their generation takes place taking into account contens of matrix S. First and last elements of vectors are identical (crossing, where a fire station and nearest crossing to the fire place are found). Other elements are determined by accidental way, but taking into account implementation of condition, that from the place of fire station it is possible to get on one of the 4th crossing, and from each of them – already on one of the three. We will designate P – number of elements in a selective population.

For forming of goal function (fitness-function) is possible to apply two approaches. In the first case is necessary to have the sufficient set of the statistical data grouped in tab. 2, and to carry out identification of dependence

Table 2

Initial data structure for fitness-function identification

Length of way L	Number of crossing K	№ time interval g	Quality of road coverage q	Time of passage T

$$T = F(L, K, g, q), \tag{4}$$

where T – time of the following of fire brigade to the fire place, K – number of crossing which it passed, g – number of time interval, q – index of road coverage quality, which integrates in itself and weather terms. By correct formalization of problem carrying out identification (4) is simple. It is enough preliminary to execute data normalization and apply a least-squares method for construction of equation of linear regression [Nakonechniy, 1997], the Brandon's method – for nonlinear regression [Chavkin, 2001], methods of models self-organisation – for polinomial dependences (group method of data handling [Ivakhnenko, 1975] or method of successive simplifications [Vasilyev, 2001]).

In second case the goal function forming accomplishing empirically with the use of weighing and correction coefficients. The given matrices T are thus used. Mean time of passage from x_0 in x_n is determined on a formula (on one of routes):

$$T_{cp.} = \sum_{i=0}^{n} \sum_{j \neq i} t_{ij} \cdot \chi(s_{ij} \neq 0), \tag{5}$$

where $\chi(*)$ is function-indicator. As, on the average, time of passing of fire brigade is increased with the increase of number of crossing, we will specify (5):

$$T = w_1 \cdot k_{n2} \cdot T_{cp}, \tag{6}$$

where w_1 – weight coefficient which determines meaningfulness of number crossing parameter. Taking into account quality of road coverage function having a special purpose (5)–(6) is such:

$$T = w_1 \cdot w_2 \cdot k_{n2} \cdot \sum_{i=0}^{n} \sum_{j \neq i} t_{ij} \cdot q_{ij} \cdot \chi(s_{ij} \neq 0), \tag{7}$$

where w_2 – weight coefficient indicative on importance of quality road coverage parameter. As in a different time of days time of fire brigade passing to the fire place will be different, a model (7) must be specified:

$$T_v = \frac{\prod_{i=1}^{3} w_i}{g_{t2}} \cdot k_{n2} \cdot \chi(v = g_{t1}) \cdot \sum_{i=0}^{n} \sum_{j \neq i} t_{ij} \cdot q_{ij} \cdot \chi(s_{ij} \neq 0), \tag{8}$$

where w_3 – weight coefficient of importance of time intervals, v – number of time interval.

We will do the row of remarks. The value of function (8) must be calculated depending on time of fire. Weight coefficients are determined by experts empirically. Thus, the use of offered approach is sujective. Construction of function (4) is carried out analytically and, in most cases, can be in a theory grounded. Dependence (8) is got, coming from empiric deductions, and procedure of its verification is enough protracted. To the receipt of model rationally to take second approach at the small retrospective view of a priori data.

Evolutionary method of optimum route determination for fire brigade following

Taking into account that every top (crossing) is incident only to four other tops, and their common number is enough large (it is used for construction of matrices S and T), to apply traditional binary presentation of vector elements of population (chromosomes) in classic EA is inefficiently. Lets $X_1, X_2, ..., X_p$ – vectors of selective population (contained the great number of routes-crossing), regulated on the elements number, i.e. $|X_i| \leq |X_j|$, $i < j$. For each of them, calculating the value of function (4), we will get $T_1, T_2, ..., T_p$.

Using principle of the successive overcoming of uncertainty, c*rossover* we will conduct on principle of successive selection [Vitkovski, 2003; Alguliev, 2004] in accordance with which large probability of participation in recombination have vectors with the less value of fitness-function. We will assume that it is necessary to define an optimum route to crossing № 39 (see fig. 1). For crossover vectors (0, 1, 5, 24, 12, 23, 39) and (0, 1, 12, 4, 11, 23, 39) are chosen. We determine, whether there are identical elements in these vectors, except for the first two and the last element. Such element is 12, it is the recombination point. Carrying out crossover, we will get two vectors-offsprings: (0, 1, 12, 23, 39) and (0, 1, 5, 24, 12, 4, 11, 23, 39). If identical elements are not present, we abandon one of vectors (with the minimum value of fitness-function) and by accidental appearance (with the use of proportion principle) we choose other vector from a selective population. A zero, one or two vectors, will be the result of crossover. Zero, if $\exists x_i, x_j : x_i = x_j, i \neq j$ in each of vectors; one – if in one; two, if the indicated terms are not executed not for one of vectors-offsprings.

Getting P descendants, among them and among P parents we choose the best vectors. Such selection is named an elite. Except for him, there are other methods of selection: selective, panmiksiya, selection with ousting [Isaev, 2000]. The practical design witnessed advantage of exactly elite selection, as at him optimum vectors-decisions are not lost. From all types of selection only for an elite it is proved [Harti, 1990]in a theory, that the iterative process of search of optimum decision meets.

For prevention of fitness-function hit in a local optimum the mutation procedure is foreseen. It uses with probability 0,01 on such chart. We consider uniformly distribute number on the set $\{1,2,...,P\}$. If $\xi = k$, we carry out the mutation in k-th vector of selective population. If the number of elements in it is d, we choose random number η from set $\{2,3,...,d-1\}$. The mutations are carried out at $\eta = L$ elements, what the random selection from two variants of $(L+1)$-th element is carried out for. Implementation of one the following terms is the criterion of search process ending of optimum decision:

- achievement of necessary value of fitness-function;
- selective population consists of identical elements;
- for any value $\varepsilon > 0 : \left|T_i - T_j\right| < \varepsilon, \forall i, j, i \neq j$.

If are executed the first or third condition, a vector will be the problem decision, the value of fitness-function which is the least.

Such method takes the advantages before classic EA and failing related to the features of problem. Considerable abbreviation of operations number is advantage, that is explained by no the application of transformation procedure of numbers in EA from decimal numeration in the two one and vice versa. Decimal presentation optimizes procedure of crossover due to reduction of forming time of vectors-offsprings. In offered method favour testifies also, that it is not "tied" to the rectangular structure of streets. If on some of them the repair is executed, in matrices S and T it is enough on the proper places to put zeros. To failing we will deliver the problem of selective population forming, that it is related to a different number of elements at vectors-representatives. In addition, procedure of determination of vector's every next element requires the revision of matrix row of distances or time, that at a lot of crossing considerably multiplies time of algorithm work.

The offered technology is oriented to that the application module will work both in active, and in passive modes. In the passive mode for every time interval on the known matrices of crossing number on the fire brigade route and quality of road coverage an optimum route calculates and is written in a data-base. By the fire to fire brigade the order with two variants of routes to the contiguous crossing will be given out. By the change of parameters in one of determining matrices or origin of situation, which a necessity in urgent delivery of information about a route which is not present in a data-base, the system is translated in the active mode of operations and urgently decided a problem.

In the general case, the decision (8) is a local optimum, as a process of its search is determined by the choice of initial point and size of search step. Therefore there is the necessity to use of evolutionary methods which are invariant to such choice.

Optimization technology for space of problem solving search

In the modeling process two problems are exposed. First from them consisted that from every crossing, usually, are present routes to the four other. At the same time, in the matrix of distances, at least, on an order anymore

variants, therefore there is a considerable calculable redundancy. Other problem consists in rational presentation of chromosomes-decisions. In particular, a priori it is impossible to define, what length must be had by a chromosome, the number of elements of which answers the number of crossing which must be passed by fire brigade following to the fire place.

For solving the indicated problems such procedure is offered. In accordance with a fig. 1 and the matrix of distances we build a matrix $N = (n_{ij})_{i,j=1}^{4\ m}$ (table of directions, which crossing contiguous to fixed are shown in) and matrix $L = (l_{ij})_{i,j=1}^{4\ m}$ (table of distances from the fixed crossing to contiguous) (tabl. 3). Obviously, that to the fixed crossing there are a lot of routes, each of which passes through a different number of the intermediate crossing. The least number of such crossing is determined by the number of quasi-concentric circumference passing through the final crossing. The maximal number of crossing is determined by an expert way and, more frequent all, does not exceed the triple number of the minimum crossing in the critical case, and double – in the regular situations.

Table 3

Table of directions

crossing	0	1	2	3	4	5	6	7	8	9	10	11	12
to the left	2	6	7	8	0	14	15	*	17	18	3	4	1
straight	1	5	6	0	12	*	14	15	2	3	4	23	24
to the right	4	12	0	10	11	24	1	6	3	20	21	*	23
backwards	3	0	8	9	10	1	2	17	18	*	20	21	4
Table of distances between crossing													
crossing	0	1	2	3	4	5	6	7	8	9	10	11	12
to the left	1	3	3	5	6	3	2	*	3	2	4	4	2
straight	9	1	1	3	7	*	2	3	1	2	2	3	3
to the right	6	2	1	4	4	2	3	3	5	1	2	*	3
backwards	3	9	1	2	2	1	1	2	1	*	2	3	7

We will define crossing as final № 39 (see fig. 1). He belongs to the fourth circumference, therefore the least length of chromosome is evened four and she will be such:

$x(1)$	$x(2)$	$x(3)$	$x(4)$

There is a starting point (station) $x(1) = 0$ in a chromosome and $x(4) = 39$ – final point. We will set maximal length of chromosome to equal eight. Parallel with implementation of the traditional operations EA, in the offered procedure it is necessary to adhere to such steps. During initialization of selective population it is necessary to provide the even representative of chromosomes of a different length. For this purpose we choose a random evenly distributed number from the set {4, 5, 6, 7, 8} which answers length of chromosome. If this number is 4, first and the last its fragment already known. An auxiliary chromosome consists of four genes. First two genes encode direction of motion from $x(1)$ (accordingly: 00 – to the left, 01 – straight, 10 – to the right, 11 – backwards), other two – from $x(4)$. On admission of decision indicates implementation of limitation which determines that crossing $x(2)$ and $x(3)$ are neighbouring. For longer chromosomes such procedure is recurrent executed.

Analysis of modeling results

Time of experimental modeling without implementation of procedure of search space narrowing on the Pentium 2,0 GHz computer made, on the average, 12-16 minutes. If in the search algorithm auxiliary procedure is executed, time of optimum decision search to 0,8-1,1 minutes diminished due to abbreviation of incorrect steps. If dependence (8) with preliminary set by an expert way weight coefficients is a function having a special purpose, time of passage to the fire place on a route definite by means of modeling, on 7-10% is less, than time, which corresponds with the route offered by experts (by the chiefs of fire brigades) or coincides. Verification of this fact

is achieved by the calculation of goal function on two offered routes by the permanent values of weight coefficients determining the feature of passage.

Inference

The method of determination of shorter route for fire brigade following to the fire place with optimization of search space is technology allowing to evade still human victims and shorten material harm. His effective application supposes the presence of informative base, containing data about the number of crossing, state of roads and road situation, and also its update in the real-time mode. The multiplied number of “corks” on the road underlines actuality of the offered method. The change of information supposes the count of optimum route.

Calculable complication of evolutionary algorithms grounds the necessity of development of the methods directed on the increase of computations speed by unchanging exactness. That is why development of the optimized models of goal functions and procedures of reduction of informative surplus of initial data are perspective. Important to remark that the offered models possess property of openness, i.e. they assume consideration and other meaningful factors, and expediently to divide weight coefficients into local (characterizing areas of roads) and global, being the attributes of road situation on the whole.

Bibliography

[Pryanichnikov, 1988] V.A. Pryanichnikov, V.V. Roenko. Criterion of routes choice of fire brigade following // Organization of works on a prophylaxis and extinguishing of fires: In Proc. VNIIPO. – Moscow: VNIIPO, 1988.– Pp. 89-92.

[Rotshteyn, 2002] A.P. Rotshteyn. Influence of methods of defuzzification on speed of tuning of unclear model // Cybernetics and systems analysis. – 2002. – № 1. – Pp. 34-45.

[Zadeh, 1965] L. Zadeh. Fuzzy sets // Information and control. – 1965. – № 8. – P. 338-353.

[Snityuk, 2004] V.E. Snityuk, A.N. Dghulay. Intellectual technology of route optimization of the fire brigade following to the fire place // ASC and devices of automation. – 2004. – V. 129. – Pp. 41-46.

[Luger, 2002] G.F. Luger. Artificial intelligence. Structures and strategies for complex problem solving. – Addison Wesley: Boston, 2002. – 864 p.

[Zaychenko, 2000] Yu.P. Zaychenko. Operations analysis. – Kiev: Vipol, 2000. – 688 p.

[Volkovich, 1993] V.L. Volkovich, A.F.Voloshin, V.A. Zaslavskiy, I.A. Ushakov. Models and methods of reliability optimization of the difficult systems. – Kiev: Naukova dumka, 1993. – 312 p.

[Wassermann, 1992] F. Wassermann. – Neurocomputing: theory and practice. – Moscow: UNITY, 1992. – 240 p.

[Kislyakov, 2000] A.V. Kislyakov. Genetic algorithms: mathematical analysis of some charts of reproduction // Information technologies. – 2000. – № 12. – Pp. 9-14.

[Kislyakov, 2001] A.V. Kislyakov. Genetic algorithms: operators of crossing and mutation of reproduction // Information technologies. – 2001. – № 1. – Pp. 29-34.

[Goldberg, 1989] D.E. Goldberg. Genetic algorithms in search, optimization and machine learning. – Addison Wesley, 1989.– 196 p.

[Werbos, 1974] P. Werbos. Beyond regression: new tools for prediction and analysis in the behavioral sciences. – PhD thesis: Harvard university, 1974. – 240 p.

[Isaev, 2000] S.A. Isaev. Development and research of genetic algorithms for acceptance of decisions on the basis of multicriterion nonlinear models // PhD thesis −NSU. – Nizhniy Novgorod. – 2000. – 18 p.

[Jensen, 2001] Mikkel. T. Jensen. Robust and flexible scheduling with evolutionary computation // PhD thesis. – University of Aarhus, Denmark. – 2001. – 299 pp.

[Nakonechniy, 1997] S. Nakonechniy, T. Tereshenko, E. Romanyuk. Ekonometriya. – Kiev: KNEU, 1997. – 352 p.

[Chavkin, 2001] A.M. Chavkin. Methods and rational case frames in a market economy. – Moscow: Finances and statistics, 2001. – 320 p.

[Ivahnenko, 1975] A.G. Ivahnenko. Long-term prognostication and management by the difficult systems. – Kiev: Technika, 1975. – 312 p.

[Vasilev, 2001] V.I. Vasilev. Interchangeability of group method of data handling (GMDH) and method of maximum simplifications (MS) // Artificial intelligence. – 2001. – № 1. – Pp. 29–42.

[Vitkovski, 2003] T. Vitkovski, W. Elzvay, A. Antchak. Planning of basic operations of genetic algorithms for planning of production // Problems of management and informatics. – 2003. – № 6. – Pp. 129-138.

[Alguliev, 2004] R.M. Alguliev, R.M. Aliguliev. Genetic approach to the optimum setting of tasks in the distributed system // Artificial intelligence. – 2004. – № 4. – Pp. 79-88.

[Harti, 1990] R.E. Harti. And global convergence proof for class of genetic algorithms. – Wien: Technische Universitaet, 1990. – 136 p.

Author's information

Vitaliy Snytyuk – Taras Shevchenko National University of Kyiv, postdoctorate researcher of cybernetics faculty; Glushkov's av. 2, building 6, Kiev, Ukraine; e-mail: snytyuk@majar.com

Alexander Dghulay – Cherkasy Institute of Fire Safety, senior teacher; Onoprienko Str. 8, Cherkasy, Ukraine; e-mail: djulaj@ukr.net

INFORMATION TECHNOLOGIES OF THE DISTRIBUTED APPLICATIONS DESIGN

Safwan Al Salaimeh

Abstract: The questions of distributed systems development based on Java RMI, EJB and J2EE technologies and tools are rated. Here is brought the comparative analysis, which determines the domain of an expedient demand of the considered information technologies as applied to the concrete distributed applications requirements.

The Distributed Systems' Design Conception

Distributed system, from the information systems point of view, consists of several independent computing devices. The state of distribution concept shares and belongs to data warehouses, hardware and software in equal parts. The application program, whose program code, data warehouses and computing resources are distributed, is called a distributed application. The composite parts of a distributed application interacts between themselves using information telecommunication technologies of seven-level OSI model, where TCP/IP sockets used on lower levels and remote access tools like DCOM, CORBA, RMI on higher levels.

The next problems can appear during the distributed applications design [1-4]: *telecommunications reliability*- communicational tools must provide a reliable interaction between the objects which execute severs and clients functions; *interface dependence*- client and server interfaces have and influence on the queries and answers; *object activation*- during the application execution the necessity of server object activation by client's request can appear; *create and remove*- during the application execution the necessity of object creating and removing can appear; *transaction support*- the control of operations set which executes by the distributed application objects, if at least one operation of the current set cannot be executed then the rest of operations must be also cancelled.

The distributed applications are based on possibility to send objects from one application program to another and to allow an invocation by one application program of object methods, which are located in anther application program. The processing of user's interaction with the database in realized in four levels: web browser, database client and database server. The user's level contains web browser, which displays web pages and collects information for processing. The middle level contains web server and application server, which are the client programs for database. The lowest level contains database servers.

The Technologies of the Distributed Applications Design

The interest to distributed applications is explained by increased requirements to modern program tools. The major of them are:

- *application scalability*- the capability for an effective maintenance of any quantity of clients at the same time;
- *application reliability* to the client application errors and communication failures;

- *transactions reliability*- is the secure system junction in functioning process from one stable and authentic state to another;
- *long-term behavior*- the nonstop run for 24 hours in a week (24x7 run model);
- *high security level* of applications, which guarantees not only the access control to different data, but also guarding in all stages of functioning;
- *high application development speed, simplicity of maintenance* and *possibility to modify* them by the programmers of a medium qualification.

At present time, there are some known technologies for static and dynamic distributed applications realization, which meet the requirements, described above: socket programming, RPC (Remote Procedure Call), DCOM (Microsoft Distributed Component Object Model), CORBA (Common Object Request Broker Architecture) and Java RMI (Java Remote Method Invocation). At the same time the most important of them are the last three-DCOM, CORBA, Java RMI.

The DCOM technology is object-oriented which is supported by the next operation systems: Windows 98, Windows NT, Windows 2000, Sun Solaris, Digital UNIX, IBM MVS, etc. The most important merit of this technology consists in possibility to integrate the applications, realized in different programming systems.

The CORBA technology is the part of OMA (Object Management Architecture), which is developed to standardize the architecture and interaction interfaces of object-oriented applications. The interfaces between the CORBA-objects encode using a special language for interface definitions IDL (Interface Definition Language). Such interfaces can be realized on any language of applied programming and connected to CORBA-applications. In the context of standards it is proposed to connect CORBA-object with DCOM-objects through the special CORBA-DCOM bridges.

The Java RMI applications consist of client and server as usual. Some objects are created on server which can be transmitted through the network or methods which are declared shared to remote application calls. On the client side there are realized applications, which use remote objects. The distinguishing feature of RMI is possibility to transmit through the network 4either methods or objects. This feature provides, finally, mobility (portability) realization.

Today, the Java RMI and CORBA technologies are the most flexible and effective to create distributed applications. These technologies are relatives by their features. The major merit of CORBA is the IDL interface, which unifies communication tools between the applications and the interoperability with other applications. On the other hand, Java RMI is more flexible and powerful tool to distributed applications development using the Java platform, including possibility of mobile applications realization.

The Java RMI technology includes two constituents: Java language instrumental tools and remote method invocation (RMI) to Java-objects. The Java language tools let to create complex distributed network applications, which is possessed of high security level and reliability, realize object-oriented programming, integrated multithreading and platform independence. The RMI technology assigns a set of tools, which let to get access to remote object on server by the special stub-object methods calls.

The Java RMI specification shows:

1) to set a remote interfaces for classes whose methods can be called through the network;
2) to create stub-objects using a special compiler;
3) to get a full copy of remote object, not only the reference to it;
4) to transmit objects in such a way that their behaviors will it not be changed when they are transmitted to another virtual machine;
5) to register a server object in the RMI registry and to support that registry in accessible state with the help of special background process. The clients can access to this registry when they are looking for the needed objects;
6) to produce marshalling and de-marshalling with the help of API serialization which translate object to a byte stream before transmitting and then, after receiving, backwards;
7) to work over the IIOP protocol that gives a possibility to communicate with the CORBA-objects (IIOP protocol can transmit a data of a different types, including structures and unions, even if they contains reverse definitions).

The RMI technology organizes the programmable interface to work with the networks unlike the TCP sockets. Such an interface has higher level, it is based on method invocation and makes an impression allegedly the remote object is being operated locally. RMI is more suitable and more natural than interface based on sockets, however, it requires the Java programs execution on the both side of a connection. The network connection, nevertheless, may use the same TCP/IP protocol.

The Network Information Technologies

There are three main parts in a client-server technologies:

- *user interface* to display an information, realize graphical user interface and form a requests to server;
- *functional logic* to realize required computing and business rules;
- *database* to execute selections, modify data and process them in accordance with received commands.

Depending on ordering method of these components on client and server machines can be 2-tier, 3-tier and n-tier client-server technologies for corporate systems functioning.

The 3-tier technology with distributed services provides independent functional relationships between the user interface, functional logic and database. The user interface and the smaller part of a functional logic are located on client computers. The bigger part of a functional logic is located on application server. The database locates on database server. The applications are independent in the model with distributed services. They interact through the networks with application server.

The object -oriented network technologies of distributed services connect in themselves the models of a distributed databases and services. The software of such kind of systems consists of a set of object units which interact between themselves through the computer network using standard interfaces. This approach lets to use units many times and to spend computer resources more economically. Each object, depending on conditions, can be either client or server in this technology. The object computing architecture, which based on distributed network services, presents a new highly upcoming kind of computer technologies, which are widely uses to distributed corporate systems' design.

The object-oriented approach is perspective to create dynamic web-oriented applications. The problem of "super thin" clients' realization is closely tied with them, when the client-applied program uses the browser environment for execution. The effectiveness of such technology for corporate systems' design is described by the possibility to realize the standard HTML-browser in any operation environment. If we'll take into account that the maintenance cost for one server and the maintenance cost for a thousand of connected to it "thick" clients are not comparable then we can make a conclusion: successfully realized corporate system sharply reduces overhead charges for maintenance.

The Component-oriented Technologies

At present time, for the development of corporate systems with a distributed database branches and remote services (functional logic) the most considerable are integral technologies like an Active X/DCOM, RMI/CORBA, Enterprise Java Beans/CORBA and CORBA/J2EE. The distributed component-oriented Active X/DCOM technology is intended for application server development, registration and management by the distributed program objects. The main its demerits are: in the first place, server computer operates under Windows NT/2000 Server operation system only; and in the second place, the tools of multi-user access for a several servers' operations coordination which is managing by transactions are completely absent.

The remote access technology RMI organizes interconnection with a remote objects, lets to develop a qualitative Internet-applications, possesses by all Java language environment merits, provides an object-oriented programming, guarantees a high security and reliability level, multithreading, multiplatform support, an independence to operation system. The RMI built on the concept of call and its parameters translation to a byte stream for transmits them through the network. The backward operation is carried out on the server side, method invocation and result transmission back to the client.

In the base of the common object request broker architecture lies an idea of uniting multivinder applications in a common tooling environment, than lets to this technology operate in heterogeneous systems, i.e. the network can serve, at the same time, computers of a different types, operating under different operation systems and serving applications written on different programming languages. The CORBA technology provides development

and maintenance of software for distributed corporate systems with a maximal convenience. Its software is called CORBA-application.

Each object requests broker ORB producer, which, practically, is the CORBA, theoretically can propose its own protocol of transport service for data transmission. In this case the expanded name of technology, for example CORBA/IIOP, which reflects the name of the IIOP network protocol. With the aid of IIOP protocol any CORBA/IIOP application can interact with other CORBA/IIOP-applications independent of hardware, software and operation systems producers.

Conclusion

The realized scientific research and practically developed works certificate that in the near future the most perspective are object-oriented Java RMI technologies and component-oriented EJB-and J2EE-technologies for the distributed corporate systems design. The EJB-and J2EE-technologies application is expedient in systems, based in using a powerful DBMS as the systems kernel. For a example, it can be the e-commerce systems, banking systems, business platforms and e-shops.

Bibliography

1- Mike Morgan, Java 2 for Professional Developers, Sams Publishing. 1999, 720 P.;

2- Eckel B, Thinking in Java, Prentice Hall, 2000, 880p.;

3- Perrone P., Chaganti V., Building Java Enterprise Systems with J2EE, Sams Publishing. 2000, 1184P.;

4- Riccardi G., Princeples of Database Systems with Internet and Java Applications, Addison-Wesley Publishing Company, 2001, 480 P.

Author's Information

Dr. Safwan Al Salaimeh – Assistant Professor, Department of Computer Science, Irbid National University, Jordan

USE OF DYNAMIC TECHNOLOGIES FOR WEB-ENABLED DATABASE MANAGEMENT SYSTEMS

Galina Bogdanova, Todor Todorov, Dimitar Blagoev, Mirena Todorova

Abstract: In this paper we consider two computer systems and the dynamic Web technologies they are using. Different contemporary dynamic web technologies are described in details and their advantages and disadvantages have been shown. Specific applications are developed, clinic and studying systems, and their programming models are described. Finally we implement these two applications in the students education process: Online studying has been tested in the Technical University – Varna, Web based clinic system has been used for practical education of the students in the Medical College - Sofia, branch V. Tarnovo

Keywords: Computer systems and Web technologies, Databases, PHP, JSP, education.

ACM Classification Keywords: H.2.4 Systems, H.4 Inforamtion Systems Applications, J.3 Life and Medical Sciences, K.3.1 Computer Uses in Education

Introduction

For some time past World Wide Web is migrating to active websites, which allows to send to the user pages, configured according to his/her own requirements and offer more dynamic experience with browsers. Such sites are created with the help of a combination of programming languages and technologies, which can be used at once, or one by one.

HTML (Hypertext Markup Language) is still the more important technology for visualizing in the Web. HTML is designed in the middle `90s for the creation of text-based documents containing commands for basic formatting (markup) as well as links to other information (hypertext). Although HTML evolved and many improvements have been added, it is in itself still static. The next step is the dynamic Web technologies, which allow the building of active websites [Water, 1998], [Pardi, 1999].

Some dynamic technologies and their application in the creation of Web based systems for database management are considered in this paper. In the second section short overview of the most familiar dynamic Web technologies, is made. In the third section two applications, are described – clinic and studying systems, which use dynamic technologies from the client and from the server side. The project for on-line studying have been implemented in the Technical University – Varna, branch V.Tarnovo. The online clinic system has been used for the practical education of the students in the Medical College - Sofia, branch V. Tarnovo.

Technologies for Building Active Websites

These technologies could be classified as follows:

A Dynamic technologies from the client side
- A1 Java applets
- A2 Active X controls
- A3 DHTML (Dynamic HTML)

B Dynamic technologies from the server side
- B1 Common Gateway Interface (CGI)
- B2 Active Server Pages (ASP)
- B3 Java Servlets and Java Server Pages (JSP)
- B4 PHP
- B5 Patented API's for Web servers (ISAPI and NSAPI)
- B6 Server Side JavaScript (SSJS)

We will overview the most famous dynamic technologies from the client side (A).

The Java applet is a program written in Java, which can be inserted in an HTML page. Because the applet is written in Java it can take all the advantages of the language i.e. it exists separately and is multiplatform.

The ActiveX controls are independent programs, also known as components, which are written in some programming language such as C++, Delphi or Visual Basic. When added to a Webpage they give specific functionality, for example graphs or charts, timers, identity verifiers of clients, or database access. The ActiveX controls are inserted into HTML pages via the <OBJECT> tag, which is now part of the HTML standard. When added to a Web page the ActiveX controls can be executed by the browser or by the server. They are designed by Microsoft and are supported only in ActiveX-enabled browsers. For now they work only in Internet Explorer and therefore cannot be considered as multiplatform.

DHTML allows to use regular HTML, scripts, dynamic object model, absolute positioning, dynamic styles, multimedia filters and many other technologies for dynamic text and graphic manipulation, which HTML shows on the screen.

The scripts are blocks of program source code, which is inserted into the Web pages and is interpreted during execution. This means that a text file with HTML code also contains additional text with commands for the browser.

Internet Explorer supports two languages for writing scripts – Visual Basic Script (VBScript) and JavaScipt. The script languages have some security-related restrictions and are not as fast in the calculations as the compiled programs.

We will now overview the most famous dynamic technologies from the server side.

A CGI program can be written in any programming language. The most popular language for CGI programming is Perl. Web servers that can execute CGI programs act like a gate between the user's request and the data that he/she requests. The main disadvantage of the CGI programs is that scaling is not so well covered. Any time

when the server receives a request a new process is created. This way the server can get very load up and freeze.

ASP combines HTML and writing scripts for the server side in one file, called Active Server Page. When the server receives a request for an ASP he will execute the code that is built up in it and will return an HTML page to the browser. As script languages can be used VBScript, JScript. One big disadvantage of ASP is that they can be used only in web servers running Microsoft operating systems.

JSP are technologies from the server side that use the Java language [Bruce, 1998], [Marty, 2001].

Java Servlet is a program from the server side that handles HTTP requests and returns the result as an HTTP response. Good analogy to a servlet is a non-visual applet that works on the server. The lifetime is comparable to this of the applets and runs inside the Java Virtual Machine (JVM). In contrast to the applets the servlets have no graphical interface. The servlets are a Java technology that corresponds to the CGI programming. One JSP-page contains HTML, Java code and JavaBean components. When the user sends a request for a JSP file the web server will first generate the corresponding servlet, if one does not already exist, execute this servlet and return the received content as a result to the web-browser.

There are two ways to access a JSP-page:

- The client request is sent directly to the JSP. In this case it is assumed that web pages have access to JavaBean components, which perform certain well-defined actions, such as database access.
- The client request passes through a servlet. The servlet generates a dynamic content. In order for the response to be sent the servlet creates a Bean. Then the servlet turns to the JSP, which represents the content along with the one generated from the servlet and stored in the Bean.

JSP and Java Servlets have the same disadvantages that has Java, that is used on the client side: Java is comparatively hard to learn language from the beginner programmers.

PHP works like ASP and JSP: the sections with scripts are framed with the tags <?php..?> and are inside the HTML page. In contrast to ASP, PHP is an independent platform and has different versions for Windows, Unix, Linux and for many Web-servers, including Apache and IIS, as main reason for this is that this product if free and with open source (Scollo, 2001).

There exist many Web servers that can be used with PHP, for example Xitami, but Apache is the only Web server for whom PHP can be compiled as a module.

Some of the advantages of PHP over HTML are:

- The editing of the Web page is made easier by just actualizing the information in the database instead of changing the whole HTML code.
- Creating pages which will be configured in such a way that to show only what the concrete user is interested into.
- Showing and modifying databases contained in the Web page and a possibility for performing manipulations on the data, such as sorting.
- Creating pages, which rotate through series of different graphics.
- Receiving information from a user and returning data to the user according to this information.
- PHP supports application programming interfaces (APIs) for access to different database types, such as Oracle, Sybase, PostgreSQL, MySQL and others.
- Programs written in PHP can use APIs for access to the data in the database when needed. ODBC (Open Database Connectivity) is a standard application programming interface for access to databases that support PHP.

Web Based Database Management Systems that use the Dynamic Technologies

We will present two particular applications: WEB-based clinic system and On-line studying system, during the development of which dynamic technologies from the client and from the server side, are used (A2, A3 и B3, B4).

The data is represented by relational databases – combination of interconnected and stored in one place data with the presence of such minimal excess that allows their use in optimal way for one or more applications [Teorey, 1996]. The relational model of the database is based on two-dimensional tables. The columns are called

fields and the rows - records. For every object or different type information is created a table (relation) in such a way that these requirements are kept: every field in one table is homogeneous, no two identical records exist, the names of the fields are unique etc.

WEB-based system for clinic management "Online clinic"

During the development a user two-tier configuration is used consisting of Application Space and Database. The application space uses a web-browser, which communicates with Apache Tomcat Server through HTTP. The used JSP technology extends the web-server by communicating with an Access database through an ODBC driver (Figure1), (Marty, 2001), (Gruber, 2001).

In the current web-based clinic system a development approach is used – a combination between the two most used techniques - side-centric and servlet-centric. The application is built by a multitude of interconnected JSP pages, which interchange information, but JSP pages without visual elements are used too. They remain hidden to the user and take care of the program logic. The access to a JSP-page is made in the way shown on Figure2.

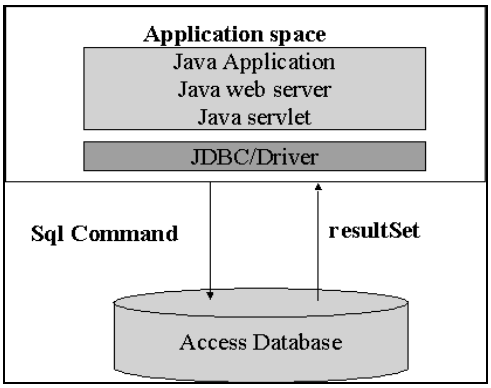


Figure 1

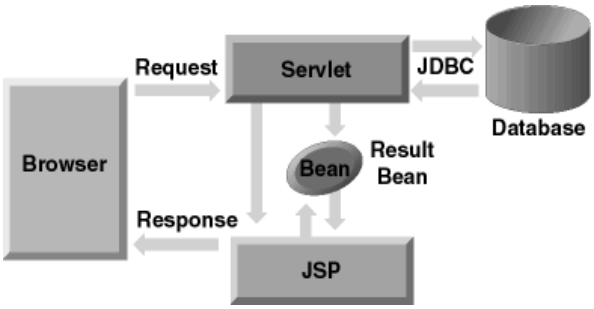


Figure 2

Their action can be summarized in the following steps:

- Reading sent data from the user – most frequently entered in a form field, but can also be sent by a Java applet or an HTTP client.
- Looking for additional information coming from the HTTP request.
- Generating answers – interaction is possible with a database or other applications.
- Formatting the results inside the document.
- Determining the corresponding parameters of the HTTP response – type of the returned document, cache parameters etc.
- Sending the document back to the client.

The developed system represents a client module, which is oriented to the client's ability to communicate with a clinic.

Basic features of the application are:

- Each patient is given a unique username and password with which he/she can use the services of the clinic.
- Each patient can get information for the physicians in the clinic.
- Each registered patient can book at a desired physician in a desired date, or can change or cancel a previously made book.
- Powerful and multi-criteria searching system for information.
- User-friendly interface.
- Fast request processing.

On-line studying system "Mindcheck"

In the development of the on-line studying system "Mindcheck" a two-tier configuration, is used that consists of Application Space and Database, which communicate through PHP code. The application space uses a Web-

browser, which communicates with an Apache server through HTTP. The database is maintained by a MySQL server.

The PHP actions, shown on Figure 3 are:

- Reading the request of the browser;
- Finding the requested page on the server;
- Executing the instructions stored in the PHP code to modify the page;
- Sending back the page to the Internet browser.

Figure 3

The relations in the database are shown on Figure 4.

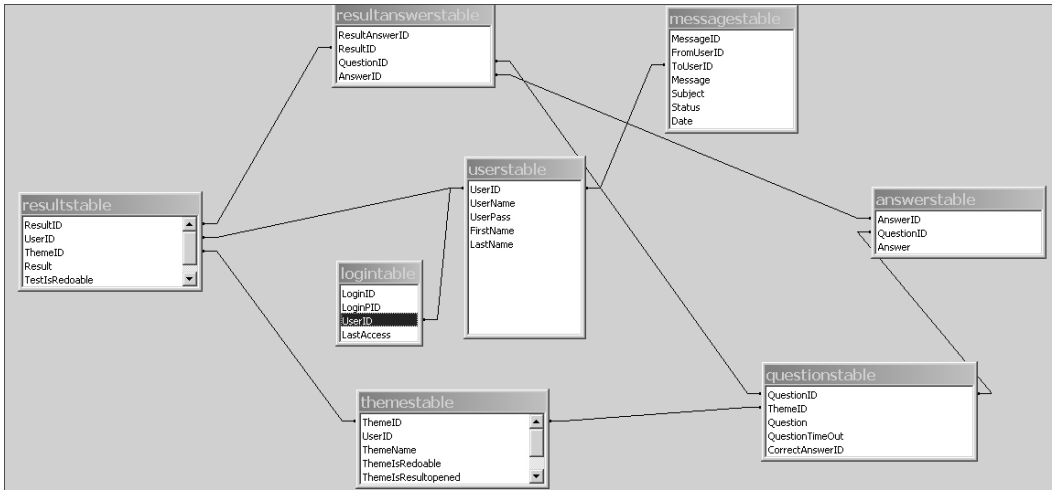


Figure 4

The instruments used in the development of the application are: Microsoft Internet Explorer 7.0, PHP Version 5.2.1 [Scollo, 2001], MySQL 5.0 [Gruber, 2000] and Apache HTTP Server Version 2.2 [Schildt, 2001]. This dynamic technology from the server side is distinguished with its efficiency, flexibility, easy configuration, working with many clients simultaneously. The client is granted security and stability during work, the developers – many tools for error elimination.

An ActiveX technology from the client side is used for studying computer mathematical programs [Bogdanova, 2003].

The Web-based studying system is asynchronous. The study is guided from special instructors or from the creators of the themes themselves [Dragon, 2001]. The users can interact in between through on-line messages or e-mail.

The main features of the application are:

- Each user is granted his own unique username and password.
- Searching for other users registered in the site.
- Easy and convenient user interface.
- Making tests on particular themes.
- Evaluation and comparison of the answers of the tests.
- Users can create their own tests.

- Uploading tests from a text file.
- Creating tests by keyword and level of difficulty.
- Multilanguage support.
- Teachers can supervise the work of the students.
- Test authors can see the results and answers of the users that made their tests.
- Fast request processing.

The on-line testing complements the services in the studying system "WebSchool", included in the project

Implementation of the systems

The project for on-line studying is tested in the Technical University – Varna, branch V.Tarnovo, with three courses for the "Business Information Systems" (Master Degree) – "Business Applications", "Databases", "Macroses in MSOffice". The site is translated into three languages: French, English and Bulgarian. In the "WebSchool" part there are lectures and exercises about "Mathematics", "Informatics", and "Biology".

The project won a special reward at the "Le concours Soft-Qui-Peut" concourse, Poitiers, France 2003.

The first version of web-based clinic system has been developed as a graduated work in the V. Tarnovo University and then has been successfully tested in the practical education of students in the Medical College - Sofia, branch V. Tarnovo

Conclusion

Mindcheck's task is to facilitate not only the students but also the teachers during the processes of learning and testing, while taking into consideration the requirements of the users. In it the users are granted better ways for checking his/her knowledge and the instructor/examiner has more active participation in the studying process. Protecting the studying material with technologies for digital signature is foreseen.

The use of the "Online clinic" will facilitate the physicians when creating their schedule as well as patients when booking and when searching information for the offered services. Priority for the future development on the clinical system is the implementation of the administrative part of the application.

Bibliography

[Bogdanova, 2003] G. Bogdanova, T. Todorov, V. Todorov. Web-based application for coding theory studying. In: The Proceedings of MASSEE International Congress on Mathematics. Borovec, 2003.

[Bruce, 1998] E. Bruce. Thinking in Java. Prentice. Hall PTR, 1998.

[Dragon, 2001] R. Dragon. Electronic education. PC Magazine Bulgaria, Vol. 10, 2001.

[Gruber, 2000] M. Gruber. Mastering SQL. Sybex, 2000.

[Marty, 2001] H. Marty. Servlets and JavaServer Pages. A Sun Microsystems Press, 2001.

[Pardi, 1999] W. Pardi. Dynamic Html in action. SoftPress, 1999.

[Schildt, 2001] H. Schildt. Java 2 A Begginer's Guide. The McGraw-Hill Companies, 2001.

[Scollo, 2001] C. Scollo. Professional PHP Programming. WROX, 2001.

[Teorey, 1996] T. Teorey. Database Modeling and Design: The Fundamental Priciples. Morgan Kaufmann, 1996.

[Water, 1998] A. Water. Programming in WEB. LIO, S, 1998.

Authors' Information

Galina Bogdanova - Institute of Mathematics and Informatics, BAS, Acad.G.Bonthev St., bl.8, Sofia-1113, Bulgaria; e-mail: galina@moi.math.bas.bg

Todor Todorov - Institute of Mathematics and Informatics, BAS, Acad.G.Bonthev St., bl.8, Sofia-1113, Bulgaria; e-mail: todor@moi.math.bas.bg

Dimitar Blagoev - Institute of Mathematics and Informatics, BAS, Acad.G.Bonthev St., bl.8, Sofia-1113, Bulgaria; e-mail: mitaka911@gmx.com

Mirena Todorova - Sofia University St. Kliment Ohridski, Bulgaria; e-mail: mirena_t@yahoo.com

THREAT MODEL FOR USER SECURITY IN E-LEARNING SYSTEMS

Maria Nickolova, Eugene Nickolov

Abstract: This paper suggest a generic threat model for the processes in e-learning systems. After a breaf discussion of the particularities of e-learning systems from the security point of view, the role of threat models in system security is explained. Then a characterization of the attackers is made, together with the assets and entry points identification and threats identification and classification. The most probable attacks for e-learning system are described regardless of the specific implementation, to help the e-learning system developers to eliminate or mitigate these attacks if possible in the design stage not waiting for actual attacks to occur.

Keywords: Security, threat modeling, e-learning.

ACM Classification Keywords: D.4.6 Security and Protection

Introduction

During the past three decades we are witnesses to a radical and transformative technological revolution that has resulted in fundamentally new ways of seeking information, communicating and learning. E-learning has arised as a new way of distributing knowledge without geographic or temporal restrictions and has passed through a several distinct phases – from Computer Based Training through to Learning Management Systems and Courseware Management Systems. There are a number of factors for this evolution, including the following:

- Ongoing development of dedicated e-learning software applications;
- Wide adoption of portals in e-learning;
- Offerings from publishers of value added services to the e-learning market;
- The advent of mobile learning which became a significant area of research and development (e.g., through the European MOBILearn project).

However, this evolution of e-Learning also brings with it a new set of threats for user and application security that impact the design of e-learning content and tools.

Particularities of E-learning Systems from the Security Point of View

E-Learning systems allow multiple users or applications to download, upload and exchange distributed information. Communication issues between end-users' computers and e-learning site (portal) in these systems are very important, as the systems are defined by widely dispersed elements in terms of network topology and physical geography. Additionally, the systems often allow many-to-many communication which provides powerful capabilities and allows many system nodes to have the same communication at any given time.

As noted in [1], a system can be attacked only through its "entry points". Designers of many computing systems can then limit security risks to a large degree by reducing the number of entry points for example by using network communication as little as possible. E-learning system designers obviously could not use this method. Since by definition e-learning communication system is an entity which exists in many physical/logical locations simultaneously, entry points are so prevalent that the system itself can in some ways be defined by them.

Additional unique challenges arise from the dynamic nature of these systems. To be effective in their intended application, it is desirable to allow the most flexibility by enabling dynamic sessions (e.g. on-line tests) in which any process may join or leave the group session at any time. To contrast, in a static session (e.g. read-only processes) all end-users only receive information during the entire session.

Another concern is that we never know in advance which processes will exist on a system at any given time. So not only are there a large volume of potential entry points to a system, but it is impossible to know exactly how

many or even where (in terms of physical/logical location) those entry points will be. Another concern is that a malicious process could always attempt to join a session (by a single user's computer or through the Internet) and conduct attacks from within the system.

Further, the security of the overall system depends not only on the security of the network and sub-networks within the system, but also on the security of each particular member process, the machine(s) on which the process resides, and the communication protocols used between participating processes. Because the identity of an e-learning system user is not necessarily tied to specific machines, it may be appropriate to implement strict authentication criteria and "credentials" verification to better control the intrusion of malicious processes and the possible loss of privacy and confidentiality. However, this may impose some performance decrease and result in a percentage of legitimate processes/users being denied participation, thereby reducing the system effectiveness.

Numerous entry points essentially force e-learning systems to rely heavily on encryption schemes for security, but if not implemented carefully they can lead to additional vulnerabilities. The difficulty in controlling access to entry points, or even identifying entry points, suggests the conclusion that it will be better to make security efforts in restricting an attacker's ability to do anything useful at each entry point. Two principles are used in cryptographic key distribution schemes for confidentiality in e-learning systems, refering to the notions that only current group of learners may possess key information for current traffic - backward secrecy and forward secrecy [2]. Backward secrecy ensures that new learners cannot decipher old traffic. Forward secrecy ensures that ex-learners cannot decipher new traffic. Adherence to these principles essentially requires new keys for the entire group of learners either at frequent intervals, or every time any learner leaves or joins the group.

The Role of the Threat Models in System Security

To design a secure e-learning system it is not sufficient to choose strong authentication and encription system and to implement new security solutions only in response to actual attacks that have recently occurred. The non-systematic nature of this approach could leave unprotected several points of the attack space and allow to break many e-learning systems with little effort. Therefore, a secure e-learning technology must address the issues of security as part of an organized process in the design phase. But a system can't be made "secure" in general, it may be secure only against a specific attack. Therefore, when designing a secure e-learning system, designers must have a clear idea of threats they have to prevent and of technical capabilities the attackers have, i.e. the necessary preliminary step in the secure design of e-learning system is to answer the question "secure against what ?" The answer is the threat model - a set of hypotheses about who (what) and how could attack the system.

Generally, threat models have three main purposes [1]:

- To improve a design's security by anticipating specific attacks and implementing countermeasures in advance.
- To anticipate the varying outcomes of "successful" attacks (for example cracks) and their possible impact.
- To enable the creation of advance response plans to deal with significant attacks as and when they occur.

The intent is that through the threat model as many vulnerabilities as possible could be identified by the developers, rather than to be left for later identification by attackers.

Generally the threat model includes three high-level steps: characterizing the attacker, identifying assets and entry points, and identifying and classifying the threats [1]. Characterizing the attacker consists of identifying goals, motivation, and capabilities. Identification must be made of the system's assets and entry points through which an attacker will seek access to these assets. The threat profile of an e-learning system must describe potential threats arising from the overall threat model and this information must be submitted to risk assessment where a decision-maker will decide to eliminate, mitigate, or accept the risk associated with each threat.

Threat Model for the Processes in e-Learning Systems

Let's now based on the general threat models theory try to create a specific model describing the threats, the motives and the means that represent a risk for the security of e-learning systems. This will be a generic model that will have to be adapted to the real systems.

The Attacker. In reality, it is difficult to answer who the potential attacker to an e-learning system is and what his capabilities may be. Attackers can take many forms. Some of them may perform their actions deliberately. Others may simply be incompetent, legitimate users of the system. Attackers that act deliberately can be divided into many categories: proof-of-concept hackers, crackers, users wanting to get illegal free access to protected resources, disgruntled employees or even bored but technically savvy teenagers. We could try to enumerate common attackers for a specific system but it is impossible to overlook all potential attackers. Their capabilities embrace unbounded time for attack launch (teenagers), extensive knowledge, privileged access (insiders).

The Assets. An asset is any element of an e-learning system which provides critical functionality. Any threat could be defined in part by the asset which the attacker wants to get access to. Of course the goal of any protection solution is not to eliminate the assets but to protect them. However, to protect assets we must first identify them. For any generic e-learning system, the following assets could be targeted by an attacker:

- E-learning content;
- Cryptographic key content;
- User personal data;
- Messages between users;
- Different group membership data;
- Network bandwidth;
- Message integrity;
- Message availability.

Of course, threat sub-models should be created for each of these assets based on the particularities of the specific real system. These sub-models should be designed to answer the following questions:

- What is the damage that can be done to the particular asset?
- What are the vectors of attack that can be used to get access to this asset?
- What conditions would have to be in place for an attack to be successful?

The Entry Points. Entry points to an e-learning system could be defined as points an attacker must use to acquire access to assets. Here is a non-exhaustive list of potential entry points to a generic e-learning system:

- Used network protocols;
- Used communication channels;
- Computers of past, current, or future e-learners;
- Physical network infrastructure;
- Logs gathering data relevant to e-learning sessions.

Now, based on the described above e-learning system's potential attackers, assets, and entry points we could build a threat model of an e-learning system. However, all systems cannot be threat modeled in the same way. Several different approaches exist [3] depending on the particular type of system and the intended purposes of the modeling. A threat model for an e-business system can be based upon the Data Flow Diagram (DFD) of the system. Other systems (e-commerce) may be better modeled using Layered Network Model Approach. We choose the well-known security aspects of Availability, Integrity, Confidentiality and Authentication (AICA) as the basis for our threat modeling of e-learning systems because these characteristics are the most important for the end-user security.

AICA Threat Modeling Approach for E-learning Systems

The four major security aspects in any computing system: Availability, Integrity, Confidentiality and Authentication (AICA) are well known and widely accepted. In this article we provide a description of the applicability of each element of AICA to an e-learning system and classify the different types of attacks in accordance to the related element. Because every attack may affect at least one aspect of the AICA model (and sometimes more), all threats can be covered by this approach (though our list is not exhaustive). Table 1 displays a summary of AICA aspects and attacks independent of the specific e-learning system implementation.

Availability	Integrity	Confidentiality	Authentication
Denial-of-Service	Malicious code attacks	Group session eavesdropping	Brute force attacks
Node attacks	Message injection	Group session traffic analysis	Dictionary attacks
Link attacks	Traffic modification	Group identity disclosure	Login spoofing attacks
Network infrastructure attacks	Traffic deletion		Key management attacks
	Traffic rerouting		Replay attacks
	Traffic misdelivery-rerouting		Man-in-the-middle attacks
	Forgery attacks		Session hijacking attacks
	Stack overflow attacks		Non-repudiation attacks

Let's now start the analysis of the specific attacks in the AICA threat model.

Availability Attacks. Availability attacks attempt to make e-learning services and data (or metadata) unavailable to legitimate users for a period of time. Here, we briefly discuss the two basic types of availability attacks: blocking attacks and flooding attacks. A blocking attack stops authorized users from accessing a resource by physically or electronically destroying the route to the resource. The non-malicious blocking attack (physical denial-of-service) embrace communications lines failure, crash of software/hardware, or inadvertent reconfiguration of systems in a way that prevents access. Malicious blocking attacks may occur when an attacker destroys or delete critical files, invalidates accounts or changes access control protocols.

A flooding attack overloads the e-learning system with a large number of requests to stop authorized users from accessing its resources. A flooding availability attack typically exploits a flaw in the design/implementation of a network protocol, operating system or commonly used application. Common examples of flooding attacks are:

Denial-of-Service. An e-learning system should have reasonable capacity (in terms of bandwidth and connectivity) to meet the peak demands, however, this capacity is finite and can be exhausted. DoS attacks could be very dangerous for e-learning systems because a single message/packet may be replicated to many receivers over many links. These attacks may be malicious, but they can also be caused unintentionally or carelessly. Some are extremely complex, while others rely upon very simple methodologies. These attacks can be conducted in several ways on an e-learning system:

Masquerading Sender DoS Attacks. The attacker may gain access to the e-learning system by joining it through an authentication attack. Once authentificated, a misbehaving source can flood traffic to all other users to disrupt current and future sessions.

Masquerading Receiver DoS Attacks. Once authentificated through an authentication attack, the attacker may join the traffic by creating many end-user processes, thus greatly increasing the overhead of the system but not the traffic into sessions. The size of the session must therefore expand to handle the increased traffic which consumes bandwidth and processing resources.

Insider DoS Attacks. A legitimate end-user becomes a traitor by flooding traffic to all learners or subverting the e-learning system by signaling or creating many receiver processes.

Transit DoS Attacks. Without being authenticated, the attacker may inject unauthorized transit traffic on the same network(s) in order to disrupt communications. The widespread use of UDP as a transport protocol for real-time communications is a vulnerability since currently there is no UDP mechanism to prevent traffic congestion [4].

Availability to Present. A specific learner is denied from participing to a session through a targeted DoS attack [5].

Availability to Receive. A specific learner is denied from receiving information in a session through a targeted DoS attack [5].

SYN flood attacks. They consist in bombarding the e-learning site with SYN packets used to open a connection and could either overload the e-learning system servers or cause them to crash.

Smurf attacks. They use a PING packet sent by an end-user computer with the forged return address of the intended victim. This PING is broadcast to a large number of other users (intermediaries) and if many machines respond by sending PING packets to the intended victim, the result can be severe network congestion or outages, that could potentially make the network unusable. The intermediary can be victimized in his turn.

Distributed denial-of-service attacks multiply the effect of an ordinary denial-of-service attack by launching it simultaneously from multiple addresses. The attacker gains access to a number of other end-user computers in advance and puts code in each to launch the attack at a preset time or on a signal. Using such remote computers makes the attacker invisible, that's why distributed denial-of-service attacks are difficult to trace and prevent.

Node Attacks. By definition, in e-learning communications there are more nodes (senders/receivers) involved in a session communications than point-to-point communications and thus more potential exposure to node-targeted attacks. Each legitimate user node may be attacked through a node-specific vulnerability and be degraded, subverted or otherwise non-functioning member in the e-learning process. This decreases the availability of any information or contribution to joint projects and activities from attacked end-user nodes.

Link Attacks. Communication traffic in an e-learning system uses many more links than point-to-point communications so it is exposed to link attacks [4]. As some legitimate nodes are connected to the network by a single or few links, if these links are successfully degraded by an attack, the affected nodes may no longer be available to the system and the system is no longer available to them. If link attacks are strategically placed, network partitioning and even general system degradation may occur.

Network Infrastructure Attacks. If any part of the e-learning network infrastructure which directly or indirectly supports the communication sessions in progress, is physically/logically attacked or otherwise damaged (power, routers, switches, hubs, servers for DNS, etc.), the system will degrade and may not remain functional.

Integrity Attacks. Integrity attacks attempt to actively modify or destroy information in the e-learning site without proper authorization. Modification may include creating, changing, appending, and deleting both data and metadata. With an integrity attack, authorized users can gain access, but what they find when they get there is not what is supposed to be there. We shall briefly discuss some integrity attacks:

Malicious Code Attacks. Integrity attacks can be non-malicious in origin, as many unintentional causes may corrupt or modify data. However, malicious integrity attacks are becoming both more common and destructive. Malicious code comes in a variety of forms – virus, Trojan horse, worm etc. E-learning system administrators as well as end-users should protect and regularly chech their systems to assure they are malware-free.

Message Injection Attacks. An attacker who joins an e-learning communication, for example via an authentication attack, may then freely inject messages into the system which will be viewed as legitimate system traffic by the other end-users.

Traffic Modification Attacks. An attacker may intercept packet data, rearrange or delete specific bits, and/or forward data as if no changes occurred. This attack does not require knowledge of key data or any particular understanding of the e-learning data itself.

Traffic Deletion Attacks. Similarly to the previous attack an attacker may simply delete data on the communication channels. Again, this requires no particular understanding of the e-learning data itself.

Traffic Mirror-Rerouting Attacks. An attacker may mirror the traffic by rerouting it to unauthorized end-users. This is analogous to inserting a mirror into the stream of observable traffic and re-directing it elsewhere, without affecting the original stream's destination – thus making this attack difficult to detect.

Traffic Misdelivery-Rerouting Attacks. If the attacker reroutes traffic to unauthorized receivers without mirroring, some messages are lost and not received by some or all end-users. This attack is more detectable than traffic mirror-rerouting attacks since end-users and messaging protocols can detect lost messages.

Forgery (counterfeit) Attacks. They make a false representation of data that has come from another address. An attacker can also hijack a session by intercepting a communication session and continuing it in the name of (and with the privilege of) the victimized end-user.

Stack Overflow Attacks. The attacker intentionally supplies a very large amount of input data (for example, 3,000 characters in a limited length field), in order to exceed the space allocated and spill over into adjacent data or code areas, either corrupting other values or inserting new commands to be executed.

Confidentiality Attacks. Confidentiality attacks are passive attacks that expose confidential data to the view of unauthorized readers [4].

Unlike integrity attacks, confidentiality attacks don't alter the e-learning content, but only affect the security level and dissemination of this content, as well as learners' personal data. However, confidentiality attacks may be used as a first step in availability or integrity attacks, as where the attacker obtains confidential passwords by defeating encryption or simply by password guessing. Confidentiality attacks to e-learning systems may take the form of commercial appropriation of confidential information. Most involve intrusions and disclosure of private facts that are not commercially motivated, but are undertaken solely for the amusement of the attacker.

In many e-learning systems, the unauthorised access to confidential data may require the relatively difficult task of gaining access to the system on the administrator level. However, an attacker may also read information without illegitimate privilege escalation (e.g. insider attackers). Storage and timing leaks via covert channels also fall into this class of attack. In an e-learning system exist more eavesdropping opportunities in comparison to most networked systems, because of the multiple channels used through the multicast communication protocols [4]. Further, attackers don't need eavesdrop if they can instead simply become learners – at which point the system views them as legitimate and purposely sends them all relevant traffic. Common protection techniques for this type of attack require extensive use of encryption with cryptographic keys to ensure privacy. The cryptography used may or may not be the same as that used for authentication of the learners. Let's now discuss in short some confidentiality attacks:

Session Eavesdropping Attacks. Traffic between particular learner's processes in an e-learning system is virtually impossible to hide. Because of the dispersal of the processes throughout a typically large network, as well as because the volume of communication channels used, it does not seem feasible to prevent attackers from observing traffic. This is why encryption methods are critical for communications in an e-learning system. The goal is not to prevent enterely the attacker from observing traffic, but rather to ensure that the observable content is unusable or meaningless to him.

Group Session Traffic Analysis. Even if appropriate encryption measures are taken, the attacker may still observe the traffic flow and make corresponding deductions based on when and where messages are sent, message type and volume.

Identity Disclosure Attacks. The identity of e-learners is attacked for unauthorized disclosure [6].

Authentication Attacks. Authentication attacks occur when an attacker masquerades as a legitimate end-user (using a stealed password, key, or credential) or an attack device masquerades as a legitimate device participating in e-learning scheme both usually aiming to get free access to paid e-learning networks and services. But a masquerader can also launch insider attacks to access data/metadata (confidentiality attack), modify data/metadata (integrity attack), and/or deny others data/metadata (availability attack) based on the legitimate user identity authorization capabilities they have taken over. Let's discuss briefly the following attacks:

Brute force attacks. These attacks are unsophisticated and can be effective only if power computer is used. The principle is to attempt every possible combination of characters to satisfy password (key) authentication.

Dictionary attacks. They could be considered as a subset of brute force attacks. Instead of trying all password combinations, a dictionary attack attempts to satisfy a password prompt by trying commonly used passwords from a list or dictionary.

Login spoofing attacks. An attacker can use a fake login program that prompts the legitime user for an ID and password. Instead of logging the user into the requested e-learning system, the bogus program stores or forwards the stolen credentials and returns a notice that the login has failed.

Key Management Attacks. In a system in which the key management scheme does not adequately protect both forward and backward secrecy, an attacker may use information obtained from a former legitime user to gain access to the e-learning system. The attacker may in fact be a former legitime learner aiming to find enough information to break the current encryption scheme and to access the content for free. A basic example may be

procedural mishandling of keys exposing them to disclosure. A centralized server that generates and disseminates keys for legitime users represents a single entry point and an attractive attack target.

Replay Attacks. In certain encryption schemes, it may be possible for an attacker to eavesdrop on a communication channel and to record an encrypted message. This recorded message may later be played back in such a way as to allow the attacker to masquerade as the original sender. The attacker may appear legitimate to the recipient and authorized to access data to which he hasn't a right to.

Man-in-the-Middle Attacks (MITM). In MITM an attacker actively intercepts messages between end-users and servers of the e-learning site and is able to read, modify and retransmit these messages without legitime participators in the communication either knowing that their session has been compromised. The MITM attack may include: eavesdropping (including traffic analysis and known plaintext attacks), chosen ciphertext attacks, substitution attacks, replay attacks, and DoS attacks.

Session Hijacking Attacks. In this kind of attack an attacker intercepts and continues a session started by a legitime user of the e-learning site. In contrast to MITM session hijacking occurs after a session is established and the legitime user's session is terminated.

Non-Repudiation Attacks. Non-repudiation requires the ability to prove to a third party that a particular message is sent or received preventing the source from later denying transmission of the message. As a consequence of existing vulnerabilities in access control, e-learning user may deny (legitimately or illegitimately) having participated in some or all communication sessions and having sent/received specific messages.

Source Non-Repudiation Attacks. An attack targeting an end-user in order to deny sending communications.

Receiver Non-Repudiation Attacks. An attack targeting an end-user in order to deny receiving communications.

Conclusions and Future Work

E-learning systems present a unique challenge to security engineers because of their nature and inherently complex architecture. That's why, it is especially important to use a systematic formal methodology of identifying threats during the development phase. We propose the AICA threat modeling approach for focusing e-learning systems' design efforts on the elimination or mitigation of the threat associated risk, regardless of the specific implementation. The goal of this work is not to provide an exhaustive list of all possible attacks for e-learning systems; it rather tries to provide a conceptual framework by which developers of e-learning systems can decrease the number of overlooked security vulnerabilities at the design stage. The presented AICA threat model needs to be tested in practical e-learning implementations so as to really ascertain its performance.

References

1. Swiderski ,F. and W. Snyder, Threat Modeling (Microsoft Professional). Microsoft Press, 2004.
2. Judge, P. and M. Ammar, "Security Issues and Solutions in Multicast Content Distribution: A Survey," IEEE Network, January/February 2003.
3. Myagmar, S., A. J. Lee, W. Yurcik, "Threat Modeling as a Basis for Security Requirements," In Symposium on Requirements Engineering for Information Security (SREIS), 2005.
4. Ballardie, T. and J. Crowcroft, "Multicast-Specific Security Threats and Counter-Measures," Symposium on Network and Distributed System Security, February 1995.
5. Tannenbaum A., M. van Steen, Distributed Systems: Principles and Paradigms, Prentice-Hall, Inc., 2002.
6. Nickolova, M. and E. Nickolov, User Privacy Problems in DRM protected E-learning Objects – fron Standards to Practice, proceedings or the Second International Workshop on Compiter Science and Education in Computer Science, June, 5-7 2006, Borovetz – Sofia, Bulgaria.

Authors' Information

Maria Nickolova – National Laboratory of Computer Virology; BAS, Acad.G.Bonthev St., bl.8, Sofia 1113, Bulgaria; e-mail: maria@nlcv.bas.bg

Eugene Nickolov – National Laboratory of Computer Virology; BAS, Acad.G.Bonthev St., bl.8, Sofia 1113, Bulgaria; e-mail: eugene@nlcv.bas.bg

DOMAIN MODELING TO SUPPORT ANTI-CYBER CRIME EDUCATION

Yordan Penev, Svetlozar Markov

Abstract: This paper describes an approach to a computer-based learning of educational material. We define a model for the class of subjects of our interest - teaching of investigation and prevention of computer crimes, (those including both theoretical and practical issues). From this model, specific content outlines can be derived as subclasses and then instanced into actual domains. The last step consists in generating interactive documents, which use the instanced domain. Students can explore these documents through a web browser. Thus, an interactive learning scenario is created. This approach allows reusing and adapting the contents to a variety of situations, students and teaching purposes.

Keywords: e-Learning, Cyber Crime.

ACM Classification Keywords: K.3.2 Computer and Information Science Education.

Introduction

Our main purpose is to develop educational activities through Computer based learning. That means creating a series of interactive documents that convey the subject matter in a proper way to our learners. The technology can help to make this work easier in two ways.

On the one hand, it allows a broader, faster and more effective spreading of this material which, in turn, enhances its usefulness for teaching to more students. On the other hand, it facilitates adapting and reusing the resulting material to a variety of situations and for a range of didactic purposes. Therefore, we have set the emphasis in developing a richer and more interactive material rather than copying the usual model based on lecturing. Technology alone can not cope with these goals. It is to be laid upon a solid ground to get the most out of it. Two well-known principles that we have taken into account are the level of abstraction and reusability. They have been widely proposed, for instance, in the work of Merril, Wenger and Mayorga from the Instructional Design perspective, in the works of Maglajlic and Maurer from the educational multimedia viewpoint. Following this line of thought, our idea is to have a domain model separated from the media technologies that will be used to navigate through it.

In fact, we set up taxonomy of domain models from the most generic description to the one that has the actual contents to be learnt. The rationale is to have a description for a class of subjects (a metamodel) which can be refined into a number of subclasses that represent specific domains - domain models, instructional models, cooperative models and structural models. Each subclass can be instanced to hold the actual and usable contents. Structured documents addressing particular educational purposes can be created from these domain instances. These documents include text and a variety of references to objects within the domain model (such as topics, problems, descriptions, hints or common misconceptions among others) which hold the appropriate information. We claim this approach to be both flexible and suited for developing educational activities through computer based scenarios.

The Meta-Model as a Generic Description of Information

A meta-model is a high-level generic description of information that draws a class of subjects. This representation can be refined into any number of domain outlines and is shaped in terms of entities and relationships between them. Thus, it defines a class of models that set up the entities and relationships needed to specify a particular knowledge field. Meta-models contain a number of domains that are also types of models that have their own specific entities and relationships. For each domain, the metamodel has its list of entities and their attributes, which stand for their properties. A domain can incorporate entities that belong to other ones as well as adding some extra attributes to them.

Hence, defining a domain means stating its entities and the relationships that can be held between them. The first ones characterize the objects which are relevant for the subject matter to specify. Any two entities will differ in the

values of their respective attributes. The second ones hold the connections between those objects and can be divided into two classes.

Structural relationships allow setting up taxonomies between entities (such as class-subclass or metonymy). They hold the transitive property, which will allow making inferences through the object taxonomy. As a common rule, there will be, at least, two structural relationships: **type-of** and **part-of**. Domain-specific relationships have a meaning associated only to a particular subject and hence they carry their own semantics. As a metamodel gathers the commonalities of its derived models, it sets the pattern from which all of them inherit their shape. Thus, its usage is, mainly, to be instanced to create a domain model. Nevertheless, it could be useful also for spreading any necessary modifications among its subclasses. The possibility of changing the type of objects and relations at the meta-level makes building up and maintaining the domain easier.

As an example and study case we have created a metamodel for defining subjects in the combating of computer crime field. Those subjects share the feature of having a formal corpus of theoretical topics which should be applied to solve practical problems. We have used three domains to describe that metamodel: *the conceptual domain* holds the contents to be learnt.

The instructional domain includes the objects which will be used to perform the actual teaching and learning processes. The didactic model allows relating the other ones. Its elements are actually references to objects that belong to the other domains to which some new attributes are added. These extra properties permit assessing if the objects suit a particular didactic purpose. In the instructional domain, for example, we can see that a problem can have sub-problems through the structural relationship part-of or that a solution can be linked to a multiple-choice-question or a problem.

Models

A model is an instance of the metamodel and therefore a class derived from it (which is its super-class). It defines a knowledge structure which will be filled with actual data as the model is instanced. It includes a description of a subject matter in terms of elementary units and their relationships. Those constituents have their own meaning and purpose as objects within the context of a specific domain.

The domain model supplies an explicit characterization and a flexible access to the contents which could be retrieved in a variety of ways. They are also defined to be reusable and therefore to ease the creation of new domain models and didactic materials. A library of generic models providing a range of ontology - which could be adapted to a wide variety of subjects-would help authors and is being created.

An Instance of a Model: Network Evidence of Computer Fraud

As an application example of the approach described before we have developed a course on Network evidence of computer fraud. The goal for the students of this subject is to be able to identify locations and collect evidence on network servers where they can see and correctly identify evidence of computer crimes and hostile activity and to prove their correctness. The course includes a theoretical corpus of topics and techniques that are to be applied to investigate those crimes. Hence, theoretical issues are the grounds on which the applied knowledge is based in this practical course. In order to cope with these two perspectives, we have defined two different domain models (conceptual and instructional ones) and a didactic layer which stays upon them, following the meta-models shown in section 2.

The Conceptual Domain Model

This model collects the theoretical corpus stated before. It covers a number of topics dealing with both, declarative and procedural knowledge. There are two kinds of objects belonging to this domain: concepts and activities. The concepts represent the topics stated as declarative knowledge. They describe the basic vocabulary of the field providing accurate definitions for every relevant term. Examples of concepts are specification, precondition, post-condition, predicate or variable.

The activities represent the procedures that are to be learned and then performed by the students. They take a number of concepts and produce a result, which will be another concept. Examples of activities are specify, derive, verify or prove. Examples of how activities and concepts relate to each other could be “To specify means taking a sentence by enunciating the requisites and desired results for a network computer crime investigation and to produce a formal result - collect evidence”. The attributes of both entities are their name (a unique

identifier) and their definition (a rather formal text that describes the topic). The information model for the conceptual domain also includes a variety of relationships.

Structural relationships (part-of, sub-activity) make up taxonomies for the concepts and the activities. There is a range of domain-specific relationships which define the semantics of this subject: Belong-To, Induce, Produce or Apply-To. For instance, a Precondition Belongs-To a Specification, a Predicate Induces a Set-of-States, (To) Specify Produces a Specification (as its result) or (To) Verify Applies-To computer crime.

The Instructional Domain Model

The actual learning of the topics that belong to the conceptual domain model requires a range of objects which are the contents of the instructional domain model. These objects include those ones used for illustrating concepts, practicing procedures or evaluating the student knowledge.

The information model for the instructional domain uses three kinds of objects: explanatory, exploratory and evaluative ones. The explanatory objects contain complementary information to help explaining a given topic. If there was such a type of information available for a topic, one of these objects could be associated to that particular concept or activity. Those explanations can play a number of different roles depending on their instructional purpose. Some explanatory objects can be associated to conceptual domain objects (as in the case of the examples or the common mistakes). Some others relate to other instructional objects (for instance, the hints, which could be associated to problems, or the explanations, which can be associated to any instructional object). Furthermore, there are instructional objects that can accompany any object within the domain (as the descriptions).

The exploratory objects allow the student to navigate through the domain and to practice the procedures that he should learn to apply. For the sake of convenience, we have just one type of such objects: the problem. A problem has as attributes: name, question, which enunciates the exercise to be solved, and location, which refers to a printed collection made available for our students.

Finally, the evaluative objects allow the student to self-assess her proficiency on the domain. A problem can also be an evaluative object; in that case, the student would not be offered the answer in advance. Besides, there are multiple-choice questions which represent small problems with a number of possible answers. The student has to choose just one option as the correct solution to the question. Those objects attributes are: name, question, number of options, list of options and correct answer.

The relationships that connect the instructional entities are Part-Of (which links, for instance, a problem to its sub-problems) and Is-Solution (which relates a problem to one solution). An author can ask the instructional domain about the different attributes of its objects or it is possible to retrieve objects that hold given properties (for instance, “obtain a set of n multiple- choice questions having m options each one”). The relationships allow retrieving the sub-problems or the solutions of a given problem.

The Didactic Domain Model

This model holds the information about the didactic usage and quality of the entities belonging to the other domains. Therefore, it represents a meta-layer over those ones. Its elements are entities which include a reference to conceptual or instructional objects. These units add a number of didactic attributes to those entities. There are two kinds of didactic entities, related to either the conceptual or the instructional entities. The former ones add, as new didactic attributes, the difficulty of the entity and the acquisition level (a measure of the importance of the entity, as a part of the learning process, to the student). The latter ones add the difficulty to the instructional attributes that the entity already had. The didactic relationships can be split into three categories depending on the domains that they connect.

The conceptual entities can be connected by means of the prerequisite relationship which shows conceptual dependencies between its subject and object. If a concept is prerequisite of another one, the former is to be studied before the latter one. The instructional entities can be related by means of two didactic relationships: *Describe* and *Explain*. Describe links a description or example to a problem. Explain connects a hint or an explanation to a problem. This relationship has an attribute called role which shows the intended purpose of the explanatory object. The allowed roles are focus-on, clarify, choose, discard, illustrate and reformulate. Finally, Evaluate connects an evaluative entity to a conceptual one. It allows assessing the knowledge of the student concerning that concept or activity.

The model for the didactic domain allows retrieving data such as “the prerequisites of a concept”, “the available hints to solve a problem”, “is there any available reformulation for a given problem?” or “what is the reason for using that particular solving method”).

We created student profiles by taking into account their performance while they carry out different learning activities. Examples of such documents are self-evaluation tests, programming projects or study guides. The tests help the students to monitor their learning process. The programming projects are structured as questionnaires that the students have to work out and then submit. Study guides propose a tour through the subject matter. They are organized into sections following the topics of the regular course design. They include, for each and every of their sections, a list of the most relevant topics with their descriptions and recommended readings, related exercises, common misconceptions, questions to think about, or some directions to organize their study.

Conclusion

Our environment is a traditional University environment where the high rate of students per teacher prevents a highly individual tuition. Our goal is improving the support to the students. Hence, we bear in mind the introduction of the new information technologies into our set of basic teaching tools. In order to achieve it, we are developing ways of providing interactive learning support for our students. The technology provides the students with the accessibility and the adequate learning material.

Our approach supports the educational authoring process. It provides the author with a wide amount of teaching material and a flexible way to use it. The domain description is based on a conceptualization work that has allowed creating taxonomy of domain models. Modeling means defining the domains in terms of relevant and useful objects that can be retrieved in a number of ways. We first created a metamodel from which domain models can be instanced. Those domain models can be filled with the actual subject matter contents. Document types addressing particular teaching purposes can be created. Those documents define structured templates and learning paths. What we achieved is that now it is possible to create many learning scenarios sharing the same objects but having different teaching purposes and didactic perspectives. Facilitating reusability is a way of reducing the cost of developing applications to teach. It also helps the authors get the most out of the materials they create. These documents have a fixed part of text and a variety of references to objects (topics, problems, descriptions, hints or common misconceptions among others) which hold the appropriate information.

Bibliography

[1] Maglajlic, S., Maurer, H. and Scherbackov, N. (1998), Separating Structure and Content: Authoring Educational Web Applications. In Proceedings of the ED-Media & ED-Telecom 98, pages 880-884.

[2] Mayorga, J.I. and Verdejo, M.F. (1996), Authoring Systems Revisited: The Software Design Cycle Metaphor. In Paiva, A., Brna, P. and Self, J., eds., European Conference on Artificial Intelligence in Education.

[3] Merril, M.D. (1987), The New Component Design Theory: Instructional Design for Courseware Authoring. Instructional Science, 16:19-34.

[4] Schwabe, D. and Rossi, G. (1995), Object-Oriented Hypermedia Design Method. Communications of the ACM, 8 (8):45-46.

[5] Wenger, E. (1987), Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge. Morgan-Kaufman.

Authors' Information

Yordan Penev – Associate Professor, Academy of the Ministry of the Interior, Chief of Department for Combating Crime, Sofia, Mladost-4, Bul. A. Malinov, N 1, ypenev@dir.bg

Svetlozar Markov – Assistant Professor, Academy of the Ministry of the Interior, Sofia, Mladost-4, Bul. A.Malinov N 1, smarkonius@gmail.com

APPLICATION OF METHODS OF MULTI-CRITERIA EVALUATION IN CHOOSING A METHOD OF COMPRESSION AS A MEANS OF HEIGHTENING THE INFORMATION SECURITY OF SYSTEM AND APPLIED OBJECTS

Dimitrina Polimirova, Eugene Nickolov, Cecko Nikolov

Abstract: In this paper is proposed a model for researching the capability to influence, by selected methods' groups of compression, to the co-efficient of information security of selected objects' groups, exposed to selected attacks' groups. With the help of methods for multi-criteria evaluation are chosen the methods' groups with the lowest risk with respect to the information security. Recommendations for future investigations are proposed.

Keywords: Attacks, Methods of Compression, Objects, Information Security, Methods for Multi-criteria Evaluation, Risk Assessment.

ACM Classification Keywords: D.4.6 Security and Protection: information flow controls

The Situation

In the modern society of advanced information technologies a great importance is attributed to different methods of compression of information flows. This is related not only to the growing requirements for reducing their volume, but also to the developing variety techniques for heightening their information security when they are exposed to different attacks.

In recent years, topic of theoretical and experimental researches will be different methods of compression, having for an object to reduce the volume of information flows in the processes of transfer, processing and storage [1].

Another current problem, with respect to information flows' investigations, is heightening their security in relation to different attacks. The successful solving of this problem can be reached by application of different methods of compression to achieve volume reduction of the object and heightening their information security.

The Problem

In [2] a number of experiments had conducted connected with the information security of the object under attack and processed by different methods of compression. Co-efficient of information security is determined and selection of methods with highest co-efficient values of for each object in relation to all attacks is made.

The nest stage is to choose the method with the lowest risk in respect to the co-efficient of information security in relation to the investigated attacks. What has become a priority problem is the set of seriously questions relating to theories, methodologies and practices applied in decision making in relation to implementing a specific safety policy in environment of risk and uncertainly under certain computer, system and network configuration.

The Experiment

The purpose of the current investigation is to propose a model for researching the capability to influence to the co-efficient of information security of objects by methods of compression.

The main purpose of the model is to rate the best alternatives or variants for the decision-maker by calculating multi-criteria problems [3].

The general formulation, the model is based on, is as follows. Let (*M*) is the set of elements (alternatives, real objects), which will be evaluated, compared and among which a selection will be made. All elements of *M* can be described by one set of characteristics (time, size etc.), which is the same for all elements of *M*. A vector of numerical values for characteristics is assigned for each element.

Posed tasks:

1) To design the basic parameters of the model.

In the model are included representatives of the attacks' groups ($a_i \in A_{pot}$), methods' groups ($m_j \in M_{pot}$) and objects' groups ($o_f \in O_{pot}$), which had been determined in [4] by means of matrix transformations, applied on initially build base of relations between maximum attacks' groups (A_{max}), methods' groups (M_{max}) and objects groups (O_{max}). Three representatives for each of the 8th methods' groups (from the set of M_{pot}) from totally 10 methods' groups (from the set of M_{max}) are chosen. Three representatives for each of the 18th attacks' groups (from the set of A_{pot}) from totally 33 attacks' groups (from the set of A_{max}) are included in the model. The methods and the attacks are investigated using three representatives for each of the 7 objects' groups (from the set of O_{pot}) from totally 10 objects' groups (from the set of O_{max}). The attacks' groups, methods' groups and objects' groups are shown in Tables 1a, 1b, 1c.

Table 1a List of attacks' groups

ATTACKS' GROUPS
I. Advertisements
V. Chat
VI. Criminal Investigations
VII. Cracking
VIII. Spying
X. Exploits
XIII. Scanners
XIV. Keyboard Modifiers
XVII. Computer Trojan Horses
XVIII. Computer Backdoors
XIX. Computer Worms
XX. Computer Viruses
XXI. Accessible information
XXIV. Content
XXV. Data Encapsulation
XXVII. Spoofing
XXXI. Social Engineering
XXXIII. Zombie Computers

Table 1b List of methods' groups

METHODS' GROUPS
I. Statistical lossless methods
II. Dictionary lossless methods
III. Image lossless methods
IV. Audio lossless methods
V. Other lossless methods
VI. Dictionary lossy methods
VII. Image lossy methods
X. Audio lossy methods

Table 1c List of objects' groups

OBJECTS' GROUPS
II. Scientific file formats
III. Data file formats
IV. Graphics
V. Sound
VII. Internet related
VIII. Binaries
X. Miscellaneous

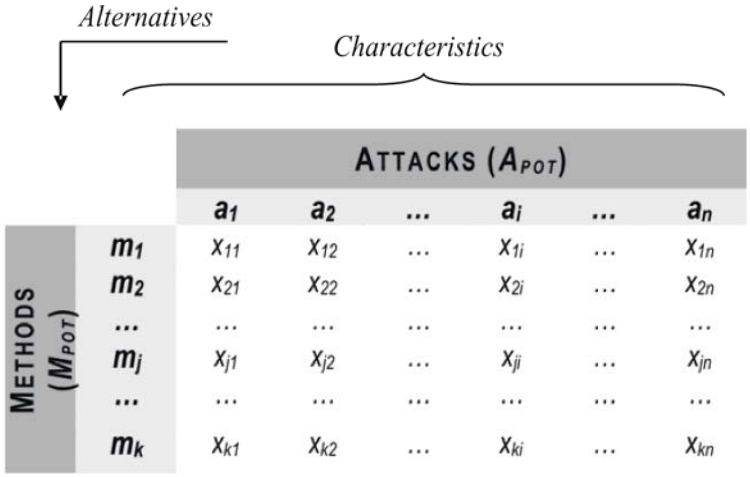
The model makes possible to assess the risk in relation to achieving higher co-efficient of information security when methods of compression are applied to different objects.

2) To build the main database.

It is necessary to systematize the available information to realize this model. For that purpose a matrix $B_{(k,n)}$ is built, which includes the most efficient methods for one objects' group (which are the different alternatives for decision-maker) and the attacks' groups, which can access to these objects' groups processed by these methods' groups (which are the set of characteristics). The vector of numerical values for characteristics, which is assigned for each element, is the co-efficient of information security (K_{INF}). The matrix $B_{(k,n)}$ is built for each of 7th objects' group. On Table 2 is shown a model of this matrix.

In the database of the matrices are included the average values of the obtained coefficients of information security when 24 representatives of the 8th methods' groups are applied to 21 representatives of the 7th objects' groups, exposed to 54 representatives of the 18th attacks' groups.

Table 2 A matrix structure, including co-efficient of information security, obtained when methods' groups are applied to one objects' group, exposed to the corresponding attacks' groups.



		ATTACKS (A_{POT})					
		a_1	a_2	...	a_i	...	a_n
METHODS (M_{POT})	m_1	x_{11}	x_{12}	...	x_{1i}	...	x_{1n}
	m_2	x_{21}	x_{22}	...	x_{2i}	...	x_{2n}
	...	...	...	...	...	...	...
	m_j	x_{j1}	x_{j2}	...	x_{ji}	...	x_{jn}
	...	...	...	...	...	...	...
	m_k	x_{k1}	x_{k2}	...	x_{ki}	...	x_{kn}

To simplify the visualization in Table 3 are shown the average values of the co-efficient of information security, obtained from all 7 matrices.

Table 3 Matrix with average values of obtained coefficient of information security when methods' groups are applied to one objects' group, exposed to the corresponding attacks' groups.

		ATTACKS' GROUPS								
		I	V	VI	VII	VIII	X	XIII	XIV	XVII
		=1=	=2=	=3=	=4=	=5=	=6=	=7=	=8=	=9=
METHODS' GROUPS	I			2,5	3,3	3,62	2,43	3,63		2,76
	II			3,86	4,62	5,26	2,42	5,28		4,13
	III	6,05	6,16	5,36	6,19	6,17		6,22	5,33	5,39
	IV			3,75	4,67	4,68		4,69		4,22
	V			2,76	3,79	3,79	2,06	3,8		3,33
	VI	25,87	25,93	25,17	25,96	25,95		25,97		25,21
	VII	17,53	17,6	16,82	17,62	17,61		17,64		16,87
	X			7,2	8,11	8,12		8,13		7,7
		ATTACKS' GROUPS								
		XVIII	XIX	XX	XXI	XXIV	XXV	XXVII	XXXI	XXXIII
		=10=	=11=	=12=	=13=	=14=	=15=	=16=	=17=	=18=
METHODS' GROUPS	I	2,81	2,72	2,45	2,64	3,32	3,17	4,16	2,54	3,33
	II	4,19	4,09	3,79	4,01	4,64	4,52	6,33	3,9	4,65
	III	5,42	5,36	5,3	5,36		5,86	6,19	5,3	6,25
	IV	4,19	4,19	3,97	4,03		4,51	4,67	3,91	4,69
	V	3,37	3,23	2,92	3,18	3,79	3,7	5,52	3,03	3,8
	VI	25,25	25,17	25,06	25,17		25,73	25,96	25,06	25,99
	VII	16,9	16,82	16,73	16,82		17,38	17,62	16,73	17,66
	X	7,67	7,67	7,45	7,5		7,98	8,11	7,38	8,13

3) To determine methods for evaluation.

The best variant for decision-maker can be determined with the help of the matrix and different methods for multi-criteria evaluation. This variant includes the method of compression with the lowest risk with respect to the co-efficient of information security of the object, which is chosen by the decision-maker in connection to investigating attacks. To find these variants, the following methods for multi-criteria evaluation are used: method of linear

combination of formal criteria and method of maximum guaranteed result. Both methods are based on the same model.

To calculate such a problem according to these methods for multi-criteria evaluation, a transformation, cold normalization (c_{kn}), has to be accomplished for the matrix $B_{(k,n)}$. This means that the maximum value for the column a_i has to be found, where we write 1, and the rest are filled up as the current value of K_{INF} is divided on the maximum value of K_{INF}. The obtained normalized results are filled in matrix $C_{(k,n)}$, which has the same structure as matrix $B_{(k,n)}$ and is part of input data for the model, realizing methods for multi-criteria evaluation and chose of elements from M_{pot} (Table 4).

Table 4 A matrix structure with normalized values

Alternatives ↓ *Characteristics* →

		ATTACKS (A_{POT})					
		a_1	a_2	...	a_i	...	a_n
METHODS (M_{POT})	m_1	c_{11}	c_{12}	...	c_{1i}	...	c_{1n}
	m_2	c_{21}	c_{22}	...	c_{2i}	...	c_{2n}
	...	...	...	...	...	...	...
	m_j	c_{j1}	c_{j2}	...	c_{ji}	...	c_{jn}
	...	...	...	...	...	...	...
	m_k	c_{k1}	c_{k2}	...	c_{ki}	...	c_{kn}
	$\max_n$	max_1	max_2	...	max_i	...	max_n

To simplify the visualization in Table 5 are shown the average values of the normalized coefficients of information security, obtained from all 7 matrices.

Table 5a Matrix with average values of normalized coefficients of information security (left part).

		ATTACKS' GROUPS								
		I	V	VI	VII	VIII	X	XIII	XIV	XVII
		=1=	=2=	=3=	=4=	=5=	=6=	=7=	=8=	=9=
METHODS' GROUPS	I			0,1	0,13	0,14	1	0,14		0,11
	II			0,15	0,18	0,20	0,99	0,2		0,16
	III	0,23	0,24	0,21	0,24	0,24		0,24	1	0,21
	IV			0,15	0,18	0,18		0,18		0,17
	V			0,11	0,15	0,15	0,85	0,15		0,13
	VI	1	1	1	1	1		1		1
	VII	0,68	0,68	0,67	0,68	0,68		0,68		0,67
	X			0,29	0,31	0,31		0,31		0,31
	$\max_n$	**25,87**	**25,93**	**25,17**	**25,96**	**25,95**	**2,43**	**35,97**	**5,33**	**25,21**

Table 5b Matrix with average values of normalized coefficients of information security (right part).

		ATTACKS' GROUPS								
		XVIII	*XIX*	*XX*	*XXI*	*XXIV*	*XXV*	*XXVII*	*XXXI*	*XXXIII*
		=10=	=11=	=12=	=13=	=14=	=15=	=16=	=17=	=18=
METHODS' GROUPS	*I*	0,11	0,11	0,1	0,1	0,72	0,12	0,16	0,1	0,13
	II	0,17	0,16	0,15	0,16	1	0,18	0,24	0,16	0,18
	III	0,21	0,21	0,21	0,23		0,23	0,24	0,21	0,24
	IV	0,17	0,17	0,16	0,16		0,18	0,18	0,16	0,18
	V	0,13	0,13	0,12	0,13	0,82	0,14	0,21	0,12	0,15
	VI	1	1	1	1		1	1	1	1
	VII	0,67	0,67	0,67	0,67		0,68	0,68	0,67	0,68
	X	0,3	0,3	0,3	0,3		0,31	0,31	0,29	0,31
	$\max_n$	**25,25**	**25,17**	**25,06**	**25,17**	**4,64**	**25,73**	**25,96**	**25,06**	**25,99**

To use the multi-criteria evaluation methods a weighted co-efficient (λ_{ji}) for each attack has to be determined. It can be described as the possibility one of pre-chosen attacks' groups to get access to one objects' group from a pre-defined set of objects' groups. The weighted co-efficient of the attacks is shown in Tables 6a, 6b.

Table 6a
The weighted co-efficient of the attacks

ATTACKS' GROUPS	λ
I. Advertisements	0,06
V. Chat	3,2
VI. Criminal Investigations	0,2
VII. Cracking	7,4
VIII. Spying	6,2
X. Exploits	5,2
XIII. Scanners	3,4
XIV. Keyboard Modifiers	7,9
XVII. Computer Trojan Horses	12,15

Table 6b
The weighted co-efficient of the attacks

ATTACKS' GROUPS	λ
XVIII. Computer Backdoors	9,25
XIX. Computer Worms	8,73
XX. Computer Viruses	9,15
XXI. Accessible information	7,2
XXIV. Content	7,7
XXV. Data Encapsulation	0,07
XXVII. Spoofing	5,3
XXXI. Social Engineering	6,8
XXXIII. Zombie Computers	0,09

Method of linear combination of formal criteria

The matrix with normalized values $C_{(k,n)}$ is used. It uses the vector $\vec{\lambda} = \left(\lambda_1, \lambda_2, ..., \lambda_i, ..., \lambda_p\right)$ as an input parameter, whose components are real non-negative numbers and represents weight for decision-making. The following limitation $\sum_i \lambda_i = 1$ has to be observed.

The calculating is:

a) For each alternative (methods' group) is assigned the number S_k, where:

$$S_k = \sum_{i=1}^{p} \lambda_i c_{kn}$$

b) The alternatives are sorted in ascending order by the number S_k, i.e. on the first place is the alternative with the maximum value of S_k; if there are several such an alternatives, their order in the list is arbitrary. Alternatives with lower values of S_k follow, etc.

Method of maximum guaranteed result

The matrix with normalized values $C_{(k,n)}$ is used. It uses the vector $\vec{\lambda} = \left(\lambda_1, \lambda_2, ..., \lambda_i, ..., \lambda_p\right)$ as an input parameter too, for which the following limitation $\sum_i \lambda_i = 1$ has to be observed for.

The calculating is:

a) For each alternative (methods' group) is assigned the number t_k where:

$$t_k = \min_n \left(\lambda_i c_{kn}\right) = \min\left(\lambda_1.c_{j1}, \lambda_2.c_{j2}, ..., \lambda_p.c_{kn}\right)$$

b) The alternatives are sorted in ascending order by the number t_k.

By analogy a variant of these both methods for multi-criteria evaluation can be examined when the values of weighted coefficients are equal. For more detailed analysis can be assumed that the weighted co-efficient λ_i can not always be known. In that case can be made the assumption that all values of λ_i are equal (which is assigned as λ_i^L). Then:

$$S_k^L = \sum \lambda_i^L c_{kn}$$

$$t_k^L = \min_n \left(\lambda_i^L c_{kn}\right)$$

These methods for multi-criteria evaluation are applied for each matrix (i.e. for each objects' group). Thus, for each objects' group we can choose an alternative (methods' group) we give preference to, with respect to the co-efficient of information security to all attacks' groups. The obtained average values are shown in Table 7.

Table 7 Average values when methods for multi-criteria evaluation are applied

S_k	t_k	S_k^L	t_k^L
Dictionary lossy methods	Dictionary lossy methods	Dictionary lossy methods	Dictionary lossy methods
Image lossy methods	Image lossy methods	Image lossy methods	Image lossy methods
Dictionary lossless methods	Audio lossy methods	Dictionary lossless methods	Audio lossy methods
Image lossless methods	Statistical lossless methods	Image lossless methods	Statistical lossless methods
Audio lossy methods	Dictionary lossless methods	Audio lossy methods	Dictionary lossless methods
Other lossless methods	Image lossless methods	Other lossless methods	Image lossless methods
Statistical lossless methods	Audio lossless methods	Statistical lossless methods	Audio lossless methods
Audio lossless methods	Other lossless methods	Audio lossless methods	Other lossless methods

4) To choose an alternative.

On the base of the obtained results in Table 7 a conclusion for the best alternative (methods' group) for decision making with respect to these methods for multi-criteria evaluation can be made. From chosen methods of compression we can make the conclusion, that with respect to the information security attacking objects' groups with different attacks, the lowest risk is when method of compression from the group of Dictionary lossy methods are applied.

5) To determine the methods with lowest risk with respect to the co-efficient of information security.

Matrices and methods for multi-criteria evaluation, which are described higher up, are applied for each objects' group. On the base on the obtained results is made a conclusion, which indicates the methods' groups with

lowest risk with respect to the co-efficient of information security for each objects' group in relation to all attacks' groups. The results are shown in Table 8.

Table 8. Methods' groups with lowest risk with respect to the co-efficient of information security for each group of objects in relation to all attacks' groups

OBJECTS' GROUPS	METHODS' GROUP
II. Scientific	Statistical lossless methods
III. Data	Dictionary lossless methods
IV. Graphics	Dictionary lossy methods, image lossy methods
V. Sound	Audio lossy methods
VII. Internet	Dictionary lossless methods
VIII. Binaries	Statistical lossless methods
X. Miscellaneous	Dictionary lossless methods

Conclusions

The main task of risk management is risk optimization, i.e. to find the moment where the risk and attaining higher level of information security when methods of compression are applied on objects are compensate each other.

Independently from the made expenses, the risk assessment shows that the application of methods of compression to a great extent heightens the information security of objects under attacks.

Recommendations

Investigations have to be made for the influence of the number of methods for multi-criteria evaluation on the order of the alternatives and respectively on the decision for decision-maker.

Future investigations have to be made with respect to the password, which can be applied, and its influence on the co-efficient of information security and the respective decision for the decision-maker.

Bibliography

[1] Milenko Drinić, Darko Kirovski, Hoi Vo, Code Optimization for Code Compression, Proceedings of the International Symposium on Code Generation and Optimization (CGO'03), 2003 IEEE.

[2] Dimitrina Polimirova, Eugene Nickolov, Cecko Nikolov, Research Of The Influence Of Methods Of Compression On The Co-Efficient Of Information Security Of Objects Under Attacks, Proceedings of International Conference Information Theories and Applications: Cyber Security, 28-29 June 2006, Varna, to be published.

[3] Иван Попчев, Боян Метев, Лорета Маркова, Класиране на системни програмни пакети чрез многокритериална оценка, сп. Техническа мисъл, №1р 1987.

[4] Dimitrina Polimirova, Eugene Nickolov, Cecko Nikolov, Investigating The Relations Of Attacks, Methods And Objects In Regard To Information Security In Network TCP/IP Environment, Proceedings of International Conference Information Theories and Applications: Cyber Security, 28-29 June 2006, Varna, to be published.

Authors' Information

Dimitrina Polimirova – PhD Student, Research Associate, National Laboratory of Computer Virology, Bulgarian Academy of Sciences, Phone: +359-2-9733398, e-mail: polimira@nlcv.bas.bg

Eugene Nickolov – Professor, DSc, PhD, Eng, National Laboratory of Computer Virology, Bulgarian Academy of Sciences, Phone: +359-2-9733398, e-mail: eugene@nlcv.bas.bg

Cecko Nikolov – PhD Student, Research Associate, National Laboratory of Computer Virology, Bulgarian Academy of Sciences, Phone: +359-2-9733398, e-mail: nikolov@nlcv.bas.bg

GOOGLE - SECURITY TESTING TOOL

Georgi Staykov

Abstract: Using Google as a security testing tool, basic and advanced search techniques using advanced google search operators. Examples of obtaining control over security cameras, VoIP systems, web servers and collecting valuable information as: Credit card details, cvv codes – only using Google.

Keywords: Google – security testing tool, IJ ITA presentation.

Introduction

The Google search engine found at www.google.com offers many different features including language and document translation, web, image, newsgroups, catalog and news searches and more. These features offer obvious benefits to even the most uninitiated web surfer, but these same features allow for far more nefarious possibilities to the most malicious Internet users including hackers, computer criminals, identity thieves and even terrorists. This paper outlines the more nefarious applications of the Google search engine, techniques that have collectively been termed "Google hacking."

Basic search techniques

The Google search engine is fantastically easy to use. Despite the simplicity, it is very important to have firm grasp of these basic techniques in order to fully comprehend the more advanced uses. The most basic Google search can involve a single word entered into the search page found at www.google.com. Once a user submit a search by clicking the "Submit Search" button or by pressing enter in the search term input box, a result page may be displayed. The search result page allows the user to explore the search results in a various ways.

The top line (found under the alternate search tabs) lists the search query, the number of hits displayed and found, and how long the search took. "Category" link takes you to the Google directory category for the search you entered. The Google directory is a highly organized directory of the web pages that Google monitors.

Main page link takes you directly to the current result web page. Description – the short description of a site. Cached link takes you to Google's copy of this web page. This is very handy if a web page changes or goes down. "Similar pages" link takes you to similar pages on the Google category. "Sponsored links" column lists pay targeted advertising links based on your search query.

Under certain circumstances, a blank error page may be presented instead of the search result page. This page is the catchall error page, which generally means Google encountered a problem with the submitted search term. Many times this means that a search query option was not entered properly. In addition to the "blank" error page, another error may be presented. This page is much more descriptive, informing the user that a search term was missing. This message indicates that the user needs to add to the search query.

Basic Google searches, as I have already presented, consist of one or more words entered without any quotations or the use if special keywords.

Example:

peanut butter

butter peanut

'+' searches

When supplying a list of search terms, Google automatically tries to find every word in the list of terms, making the Boolean operator "and" redundant. Some search engines may use the plus sign as a way of signifying a Boolean "and". Google uses the plus sign in a different fashion. When a Google receives a basic search request that contains a very common word like "the", "how" or "where", the word will often times be removed from the query. In order to force Google to include a common word, precede the search term with a plus (+) sign. Do not use a space between the plus sign and the search term. For example, the following searches produce slightly different results:

Where quick brown fox

+Where quick brown fox

The '+' operator can also be applied to Google advanced operators, discussed in the Advanced Google operators chapter.

'-' searches

Excluding a term from a search query is as simple as placing a minus sign (-) before the term. Do not use a space between the minus sign and the search term. For example, the following searches produce slightly different results:

quick brown fox

quick –brown fox

The '-' operator can also be applied to Google advanced operators, discussed in the Advanced operators chapter.

Phrase searches

In order to search for a phrase, supply the phrase surrounded by double-quotes. Example:

"the quick brown fox"

Mixed searches

Mixed searches can involve both phrase and individual terms. Example:

Macintosh "Microsoft office"

This search will only return results that include the phrase "Microsoft office" and the term Macintosh.

Google Advanced operators

Google allows the use of certain operators to help refine searches. The use of advanced operators is very simple as long as the attention is given to the syntax. The basic format is"

Operator :search_term

Notice that there is no space between the operator, the colon and the search term. If a space is used after a colon, Google will display an error message. If a space is used before the colon, Google will use your intended operator as a search term. Some advanced operators can be submitted to Google as a valid search query. The 'cache:www.google.com' can be submitted to Google as a valid search query. The 'site' operator, by contrast, must be used along with a search term, such as 'site:www.google.com help'.

Advanced Operator Summary

Operator	Description	Additional search argument required?
site:	find search term only on site specified by search_term.	YES
filetype:	search documents of type search_term	YES
link:	find sites containing search_term as a link	NO
cache:	display the cached version of page specified by search_term	NO
intitle:	find sites containing search_term in the title of a page	NO
inurl:	find sites containing search_term in the URL of the page	NO

site: find web pages on a specific web site

The advanced operator instructs Google to restrict a search to a specific web site or domain. When using this operator, an addition search is required. Example:

site:harvard.edu tuition

This query will return results from Harvard.edu that include the term tuition anywhere on the page.

filetype: search only within files of a specific type.

This operator instructs Google to search only within the next of a particular type of file. This operator requires an additional search argument. Example:

filetype:txt endometriosis

This query searches for the word 'endometriosis' within standard text documents. There should be no period (.) before the filetype and no space around the colon following the word "fletype". It is important to note that Google only claims to be able to search within certain types of file. Based on my experience, Google can easily find a word within a file of type ".txt", ".html" or ".php" since the output of these files in a typical web browser window is textual. By contrast, while a WordPerfect document may look like text when opened with the WordPerfect application, that type file is not recognizable to the standard web browser without special plugins and by extension, Google can not interpret the document properly, making a search within that document impossible. Thankfully, Google can search within specific types of special files, making a search like "filetype:doc endometriosis" a valid one. The current list of files that Google can search is listed in the filetype FAQ located at http://www.google.com/help/faq_filetypes.html. As of this writing, Google can search within the following file types:

Adobe Portable Document Format (pdf)

Adobe PostScript (ps)

Lotus 1-2-3 (wk1, wk2, wk3, wk4, wk5, wki, wks, wku)

Lotus WordPro (lwp)

MacWrite (mw)

Microsoft Excel (xls)

Microsoft PowerPoint (ppt)

Microsoft Word (doc)

Microsoft Works (wks, wps, wdb)

Microsoft Write (wri)

Rich Text Format (rtf)

Text (ans, txt)

SQL (sql)

link: search within links

The hyperlink is one of the cornerstones of the Internet. A hyperlink is a selectable connection from one web page to another. Most often, these links appear as underlined text but they can appear as images, video or any other type of multimedia content. This is advanced operator instructs Google to search within hyperlinks for a search term. This operator requires no other search arguments. Example:

link:www.news-panel.com

This query would display web pages that link to news-panel.com's main page. This special operator is somewhat limited in that the link must appear as entered in the search query. The above query would not find pages that link to www.news-panel.com/sitemap for example.

cache: display Google's cached version of a page

This operator displays the version of a web page as it appeared when Google crawled the site. This operator requires no other search arguments. Example:

cache:news-panel.com

cache:http://news-panel.com

These queries would display the cached version of news-panel web page. Note that both of these queries return the same result. I have discovered however, that sometimes queries formed like these may return different results, with one result being the dreaded "cache page not found" error. This operator also accepts whole URL lines as arguments.

Intitle: search within the title of a document

This operator instructs Google to search for a term within a title of a document. Most web browsers display the title of a document on the top title bar of the browser window. This operator requires no other search arguments. Example:

intitle:gandalf

This query would only display pages that contained the word 'gandalf' in the title. A derivative of this operator, 'allintitle' works in a similar fashion. Example:

allintitle:gandalf silmarillion

This query finds both the words 'gandalf' and 'silmarillion' in the title of a page. The 'allintitle' operator instructs Google to find ever subsequent word in the query only in the title of the page, This is equivalent to a string of individual 'intitle' searches.

Inurl: searche within the URl of a page

This operator instructs google to search only within the URL, or the web address of a document. This operator requires no other search arguments. Example:

Inurl:hair

This query would display pages with the word 'hair' inside the web address. One returned result, 'http://www.news-panel.com/en/view.category/category.374/Hair-Loss.html' contains the word 'hair' as the name of a document. The word can appear anywhere within the web address, including the name of the site or the name of a folder. A derivate of this operator, allinurl' works the similar fashion. Example:

Allinurl:hair en

This query finds both the words 'hair' and the 'en' in the URL of a page. The 'allinurl' operator instructs Google to find every subsequent word in the query only in the URL of the page. This is equivalent to a string of individual 'inurl' searches. For a complete list of the advanced operators and their usage see http://www.google.com.help/operators.html.

Example of hacking Google queries

Query	Description
"internal server error" "server at"	*Apache server could reveal admin e-mail address*
intitle: "Execution of this script not permitted"	*Cgiwrap script can reveal lots of information, including e-mail address and phone number*
intitle: index.of dead.letter	*dead.letter Unix file containing the content of unfinished e-mail.*
Filetype: reg reg +intext:"internet account manager"	*Windows registry files can reveal information such as usernames, pop3 passwords, e-mail addresses, and more.*
"Access denied for user" "using password"	*Collecting SQL usernames*
"# Dumping data for table"	*Entire SQL Database dumps (Adding 'username' or 'password' to this query makes things really interesting.)*
"ORA-00933: SQL command not properly ended"	*SQL injection hints*
filetype: inc intext:mysql_connect	*Going after SQL passwords*
filetype: sql "visa | master card"	*SQL data base dump containing credit cards information.*
Filetype: sql "cvv"	*SQL data base dump containing credit cards information.*
Intitle: VNC viewer for java"	*VNC (Virtual Network Computing) allows you to control a workstation remotely.*
Allinurl: index.htm?cus?audio	*One query, many brands of live cams!*

"active webcam page" inurl:8080	*Web cameras.*
intitle: "toshiba network camera -User Login"	*Toshiba Network Cameras*
intitle: "speedstream Router management interface"	*speedstream Router management interface access*
inurl: vswebapp.exe	*Microsoft Virtual Server 2005 access*
inurl: "level/15/exec/-/show'	*Open Cisco Devices*
intitle: "ivista.main.page"	*Security Cameras*
intitle: "everfocus edsr applet"	*My favorite security cameras – most of them using default login details – admin, admin*
filetype: ctt "msn"	*MSN Contact Lists*

Conclusion

The intent of this paper is to educate web administrators and the security community in the hopes of eventually securing this form of information leakage.

Bibliography

www.Google.com
http://johnny.ihackstuff.com, johnny@ihackstuff.com

Author's Information

Georgi Zhechev Staykov - NBU student; Sofia, Bulgaria; e-mail: gstaykoff@gmail.com

USE OF THE MAPLE SYSTEM IN MATH TUITION AT UNIVERSITIES

Tsvetanka Kovacheva

Abstract: The following article explores the application of educational technologies at a University level and their contribution in enhancing the educational effectiveness. It discusses the capabilities of computer algebra systems, such as Maple. It is integrated in the math tuition of the Technical University (TU) in Varna and is used by its students during laboratory exercises.

Key words: education, educational technology, computer algebra systems.

Introduction.

For the purpose of improving the effectiveness of the study process it is necessary to apply a variety of educational technologies (ET). They have two *basic components*:

- *technologies/media used in the study process,* through which different educational resources are introduced such as printed materials, projector equipment, TV & video and all other related resources, computer-based resources (computer algebra systems (CAS), spreadsheets, databases, Power Point, www, email), multimedia, etc.
- *technologies of the study process*, which include planning, organization, carrying out and evaluation of the entire process. They respectively comply with the aims and purposes of education, the design of co-ordinated tuition, integrated curricula, study approaches, principles and methods of evaluation. Appropriate media needs to be selected for the purposes of the study process. This choice is influenced by main factors such as aims and time, media access, human resources, costs. In math education for the engineering speciality computer-based resources play a great role due to the capabilities those resources come with:

- ❖ regarding the requirements towards the educational technologies to offer:
  - enormous storage of information/data,
  - high–speed operation;
  - flexible data handling.
- ❖ regarding the requirements for the educational technologies applied in class to offer:
  - interactive learning;
  - self-study;
  - incorporation of different teaching approaches reflecting the aims of the study process.

Choice of a Math computer algebra system. Main capabilities of the Maple system.

Series of CAS have been released for symbolic (analytic) computations as well as realisation and various numerical methods. Some of the most commonly-used CAS are Mathematica, Maple, Matlab, Mathcad, etc. [1]. The first two are aimed at solving scientific problems while the next two is reserved more for engineering problems.

There is a tendency for integrating all the different systems. For instance, Mathematica and Maple systems both offer good options for visual programming, Matlab also holds an analytical calculations library while Maple and MathCad allow for compatibility with Matlab.

The Maple system was designed by a research group at Waterloo University, Canada [5,13]. It provides for automated computations of any degree of complexity [2,4,7,8,9,10]. More than 3 500 integrated functions are included. The system also allows for Latex documents format conversion. It is available and compatible for all operating systems – Windows, Macintosh, Linux, UNIX. Some of its features that deem it suitable for the study process, of university students in particular, are the following:

- interactive mode of operation;
- handling defined mathematical calculations;
- no need for previous experience with computer equipment;
- an easy to learn and use system;
- uncomplicated and convenient interface;
- user-responsiveness;
- opportunity to cover a great deal of the math tuition for the engineering specialities of the first and second course.

Those are the chiefest considerations in choosing the application of Maple system during laboratory exercises (LEs) with math students at TU. One of the latest versions of the system is used, i.e. Maple 9 [6].

Computations in the system are done in *two ways*: by analytical and numerical methods. In the usage of the first method greater exactness can be achieved. Due to the fact that many problems cannot practically be solved using the analytical method, numerical methods are resorted to, which are built into the system library.

The system of analytical calculations consists of cores, written in C-language procedures, and written in the Maple language library and interface.

The system contains *a manual*, giving information about interface features, symbolic and numerical computations, numerical and symbolic solution of equations, computation of mathematical functions, linear algebra, graphic representation of calculations, programming.

The interface features are as follows (fig.1,fig.2):

- good word processor;
- multi-window tasking;
- representation of incoming and outgoing data in the form of mathematical formulae;
- great graphics capability where graphics can be transferred to and viewed in separate windows or within documents;
- possibility to use hyperlinks and create e-documents;
- easy keyboard management with the help of the main menu and the Tools panel.

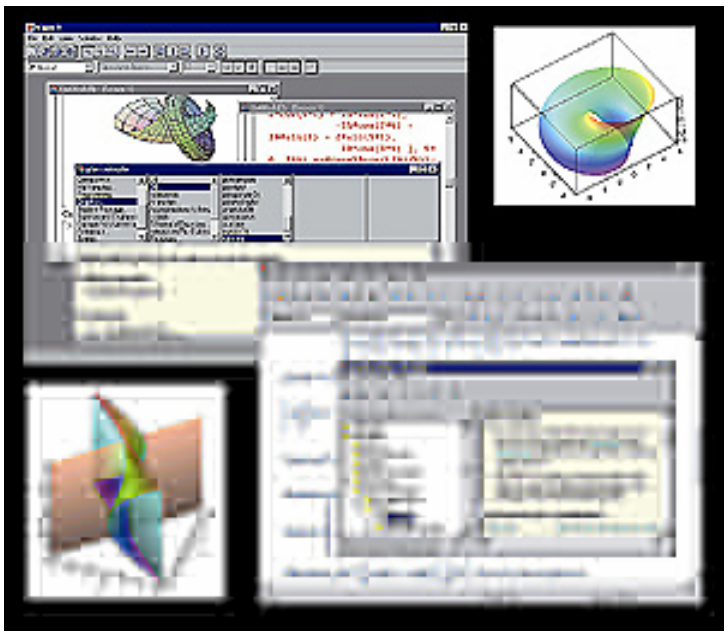


fig.1

The system effectively performs *symbolic and numerical calculations:*

- differentiation of functions;
- numerical and analytical integration;
- calculations of limit functions;
- expansion of functions in the series;
- calculations of sums and products;
- Laplace transforms;
- discrete Z-transforms;
- Fourier transforms;
- work with continuous functions, defined on parts.

The system provides for numerical and symbolic solution of problems. Those include:

- symbolic series calculations;
- solution of linear and non-linear system equations, systems of differential equations, transcendental equations, systems of inequalities.

The system performs *computation of elementary and special mathematical functions*, allowing for computation of all elementary and most special mathematical functions, recalculation of the coordinates of points for a multitude of coordinate systems.

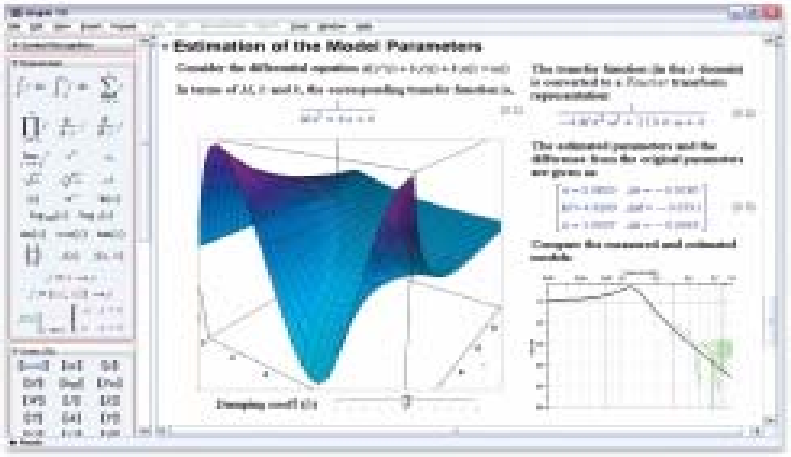


fig.2

In the area of *linear algebra* the system enables for more than 100 operations with vectors and matrices; as well as solution of linear system equations; matrix formation and their transformation; assessment of the eigenvalues and eigenvectors of matrices.

With regards to *graphic representations* it creates:

- multi-function graphic manipulation;
- change of the kind of axis (on a linear or logarithmic scale);
- function graphics in Cartesian and polar coordinate systems;
- special kinds of graphics (points from massive, vector diagrams on a level, etc.);
- graphics allowing the solution of differential equations;
- 3D graphics with functional colouring;
- intercepting objects in space;
- animated graphics, etc.
- Typical *programming features* that accompany the Maple system are:
- powerful inbuilt procedure programming language;
- simple and common syntax of the programming language;
- vast entry of mathematical types of data;
- types of data entered by user;
- comprehensive libraries for expansion of the programming language;
- construction of outer functions and procedures;
- programming interface with programming languages such as C, Fortran and Latex.

Maple 9 Student Edition *System Requirements for* Linux® are [13]:

- Red Hat® Linux 7.1, 7.2, 7.3, 8, 9;
- SuSE® Linux 6.4, 7.0, 7.2, 8.0, 8.1;
- Mandrake Linux 8.1, 8.2, 9.0;
- Intel Pentium III 650 MHz or fully compatible;
- 128 MB of RAM*, 256 MB recommended;
- 250 MB of free disk space;
- CD-ROM drive (for CD installation);
- X11 R5 or R6. Display supporting at least 16-bit color at 800 by 600 resolution;
- Internal TCP/IP connections enabled.

Maple system application in math laboratory exercises.

LEs are an important part in the math tuition for university students [3,12]. They are intended for students in their first and second year in nearly all specialties at TU. Maple system is taught alongside the following subjects:

- Mathematics-1 (M1) – linear algebra (polynomials, matrices, determinants, inverse matrices, systems of linear equations), vectors, eigenvalues and eigenvectors of a matrix, products of vectors (scalar, cross and scalar triple), line equation in plane, conic sections, limits, continuity, derivatives, differentials, analyze of function;
- Mathematics-2 (M2) – undefined and defined integrals, plane and line equations in space, functions of two and higher variables (partial derivatives, differentials, extreme), series, complex analysis (complex numbers, elementary functions of complex variable, analytical functions and Cauchy-Riemann equations);
- Mathematics-3 (M3) – ordinary differential equations, integrals (double, triple, line and surface), complex analysis (integral of function of complex variable, Cauchy's theorem and formula, Residuum's theorem), Fourier and Laplace series and transforms, probability theory and statistics.
- Numerical methods (NM) – systems of linear and non-linear equations and systems, numerical differentiation and integration, approximation of functions, ordinary and partial differential equations.
- Statistics (ST) – combinatorics, probability theory, statistics.

The purpose of the exercises is:

- extending and consolidating of students' knowledge acquired in lectures and seminars;
- introducing students to the Maple system options and getting them used to work with a ready-made system;
- graphic representation, demonstration, modelling and analysing of end results;
- introducing students to methods of numerical solving of problems.

LEs's main task is that they are carried out in accordance with the general study aims of the lectured subject. Appropriate selection of study materials is necessary as well as optimization of the LEs.

For the purpose of raising the effectiveness of the study process there is a preliminary introduction of students to *the topical curriculum* of LEs conforming to the curricula of both lectures and seminar exercises. This ensures suitable choice of methodology and consistency in LEs. Students have the chance to work with the system of LEs and its continuity which enables them to prepare for dealing with all tasks ahead. Planning the term course of LEs allows for establishing the assessment criteria and continuous checks of students' progress.

Each subject holds 7 laboratory exercises. *The structure* of every LE is as follows:

- *theoretical part*, involving introduction to the topic, brief theory and presentation of basic Maple commands;
- *practical exercises,* introducing a step-by-step detailed analysis of a given problem using Maple commands. Students copy and perform the operations under a lecturer's supervision;
- *progress tasks* - problems that require each individual student's work with Maple;
- *progress checks* – questions to consolidate the theoretical material taught.

During LE the lecturer ensures that the study process is carried out in the right manner. Group work method in LEs develops students' skills for team work. Many students' gathering around only one computer is not recommended. Proper assignment of tasks needs to be delegated to avoid idleness in class. Checks and instructions are carried out for the whole group to economise time. Discourse and discussions facilitate the build-up of long-term and in-depth knowledge.

LE is what is known as 'learn by doing it' [11]. Therefore, students should be assisted in taking maximum advantage of this type of study.

At the end of LE each student has an individual check with the lecturer on the validity of their solutions to the given problems, they are then aided in the draw up of conclusions and evaluation of their work. The final assessment of each student is complex and takes into account their progress during LE, marks from the progress checks and/or written assignments.

Students prove to benefit from exercises involving introduction and application of various numerical methods in solving particular problems. They provide a basis for comparison between the analytical and the approximate solutions. The following subjects feature topics such as:

- in M1, solving of system equations (method of simple iteration, the consecutive approximation method, Seidel's method);
- in M2, solving defined integrals numerically (rectangle rule, trapezoidal rule, Simpson's rule);
- in M3, solving of ordinary differential equations (Euler's method, Runge-Kutta method);
- in NM, solving of systems of linear equations (of simple iteration, the consecutive approximation method, Seidel's method), non-linear systems of equations (numerical integration method, Newton's method), approximation of functions (Lagrange's interpolating polynomials and Newton's method), partial differential equations (Libman's method).

During ST LEs there is the option to calculate the statistics characteristics of a given massive of data, to approximate the data through given dependents, to generate random quantities, to numerically estimate statistics functions, to carry out various transformation and to represent them graphically.

Conclusions.

The Maple system holds great potential in assisting the study process of the chapters included in the Math tuition. Every lecturer aims at being most collaborative with their students and in order to achieve efficiency he/she has to creatively make use of the system's capabilities. With the lecturer's consideration, depending on experience, various exercises to suit the purposes of the LEs can be performed. Such could only enhance the in-depth comprehension of the material taught and likewise motivate students' interaction.

The application of Maple in LEs allows for greater amount of solved problems among students, helps them determine certain properties unaided and thence comprehend the content of notions more thoroughly, their properties, theorems and the relations between them.

Students learn to independently solve problems with the help of a CAS and to select methods for their solving. Mastering Maple during the first years at University is beneficial to students not only in solving mathematical problems but also any complex applied technical problems they will encounter.

Bibliography

1. Computer algebra system http://www.answers.com/topic/computer-algebra-system
2. Govoruhin V, Tsibulin V, Computers in Mathematical Researches: Maple, MATLAB, LaTeX. A Course in Teaching, Piter, 2001
3. Educational mathematical site http://www.exponenta.ru/
4. Lessons, Tutorials and Lecture Notes; http://archives.math.utk.edu/tutorials.html
5. "Maple Application Center" http://www.mapleapps.com
6. Maple 9 Introductionary Programing Guide, Maplesoft, 2003; http://www-groups.mcs.st-andrews.ac.uk/~peterc/m9IntroductoryProgrammingGuid.pdf
7. Manson B.M, Maple V Power Edition, Filin, 1998
8. Maple on the Web - Links http://www.df.uba.ar/~jakubi/maple2.html
9. Maple Resources; http://eaton.math.rpi.edu/CourseMaterials/maple.links.html
10. Prohorov G.V., Kolbeev V.V, Zhednov K.I, Lebedev M.A, Maple Mathematical Package, Release 4, Oblizdat, Kaluga, 1998
11. Reis F, Brown S, 500 Tips For the Teacher, Tempus, Sofia, 1995
12. URI Mathematics Online, Teaching Materials ; http://www.math.uri.edu/Center/
13. "Waterloo Maple Software" http://www.maplesoft.com

Author's information

d-r Tsvetanka Kovacheva - Department Mathematics, Technical University – Varna tsetska.kovacheva@tu-varna.acad.bg

ADAPTATION OF A TASK-ORIENTED TRAINING ENVIRONMENT TO ITS USERS

Irina Zheliazkova, Georgi Georgiev, Rumen Kolev

Abstract: The problem of adapting teaching systems to the teacher has not been extensively covered in the specialised literature. The authors present the server-client architecture of a Task-Oriented Environment for Design of Virtual Labs (TOEDVL). The paper focuses on the computational models supporting its base of tasks (BT) and on two groups of behavioural tutor's models for planning training sessions. Detailed examples are presented.

Keywords: task, training, search, sort, adaptation, models

ACM Classification Keywords: J.1 Computer Applications, Administrative Data Processing, Education

1. Introduction

Adaptation of the teaching system to the learner's current level of domain competence, the teaching materials and the context of presenting the information is not a new idea. The learner model ensuring this adaptation was one of the main components of Intelligent Tutoring Systems, developed two decades ago. Different kinds of learner models have been exhaustively considered by *Zheliazkova & Kolev, 2004* in [1]. For the first time adaptation to another user – the teacher was considered by *Peachey & McCalla, 1986.* In [2] they determined the short-term individual learner's plan as a sequence of so called teaching operators and propose to use first order predicates for its execution. The system *EXTERN* (*Paschin and Mitin*, *1985 [3]*) also uses dynamic planning, where the duration of the teaching session and the sequence of teaching blocks are not initially defined. As a criterion for choosing the next block, the different implementations usually use one or more of the following: the volume or relative change of acquired knowledge, the time for learning, the learning speed, etc. Only in recent years the problem of adaptation of the teaching environments to their users is being seriously considered by *Jesshope at al., [4].* The arguments for this are that individual teachers use different strategies for planning the courseware presentation, level of didactic knowledge, types of assessment, scales for evaluation, rules for diagnostic of the learners' knowledge as well as ways for intervention in the learning process.

From the above discussion it follows that the training systems have to be adaptive to the author, tutor, and learner. Although intelligent, some of the well known training systems from the last two decades (*Gonzalez & Ingraham, 1994* [5]; *Chu et al., 1995* [6]; *Vasandani & Govindaraj, 1995* [7]) don't meet this requirement. The algorithm for applying the system RIDES for course development focuses on adapting the training to its author (*Fleming, 1996* [8]).

2. Architecture of a TOEDVL

Figure1 presents the architecture of a TOEDVL. In this figure the following notations are used: double arrows represent data links; single solid lines – control links; dashed lines – HTML documents transported from server to client; dash-dot-dot lines – binary data transferred between client and server. The architecture is domain and task-independent due mainly to the language for knowledge description in the training tasks. It is supposed that the user's registration/access control will be handled by a separate module administrator, also that the learner has already acquired deep structural knowledge about the simulated system by means of a similar environment for design of structural schemes [9]. The teaching material ensuring the feed-back to the learner is prepared by means of standard tools and files e.g. help editor (*.chm*), graphical editor (*.gif*, *.jpg*, *.png*), audio (*.wav*), and video (*.avi*) editors. For presenting tests to the learner, two approaches can be used. Tests can be generated and interpreted by means of specialised tools (the Test editor in Figure 1 and a test interpreter, integrated with the Program interpreter) [10]. Alternatively, they can be in HTML format, prepared by means of HTML content generator, such as *Dreamweaver*. In this case the Test interpreter (not shown in Figure 1) interprets them.

A generated program representing the user's knowledge about a given training task is extracted through visual programming by means of *the program generator.* It stores the syntactically and semantically correct program in a standard text file with an extension *.trn*. *The task manager* implements planning and execution of the sequence of

training tasks in a session (file with an extension *.ses*). This sequence is viewed as a short-term plan for the individual learners. Task execution itself is ensured by means of another tool – *the program interpreter*, which runs the simulation programs and saves the simulation results in a file with the extension *.hst*.

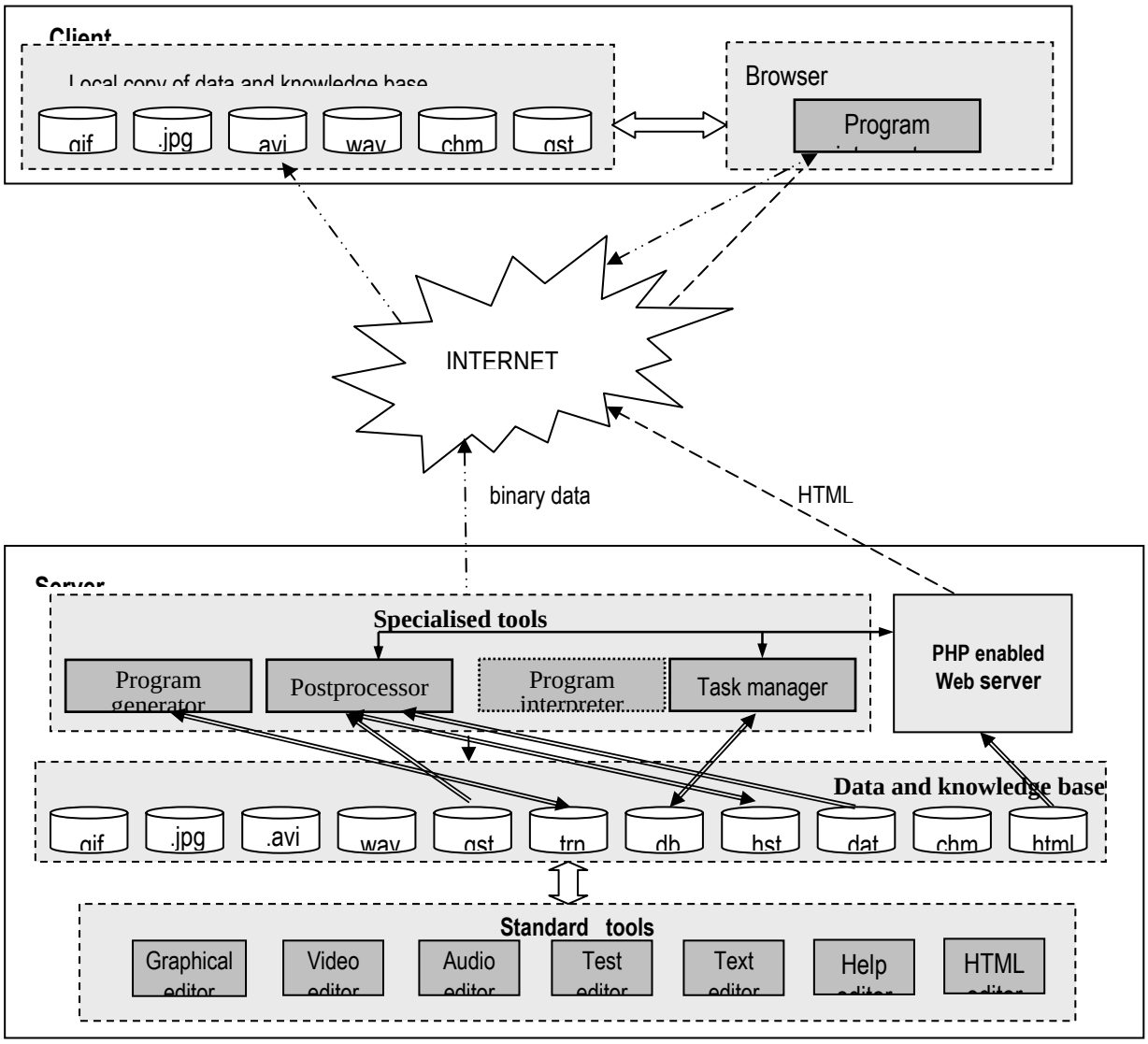


Figure 1. Architecture of TOEDVL

During the execution, the necessary operator's skills in the form of keyboard/mouse activities are added to the declarative and procedural knowledge. The computed or registered task and session parameters are stored in the base of tasks (BT), kept in a database (.db). Each task is run on the same tool once by the tutor and once by each learner. After the author completes a task and his/her *.trn* file is returned to the server, the tool updates the BT. After the learner completes a task and his/her temporal *.trn* file is returned to the server, the tool updates the learner's model, kept in the same database (.db). *The post-processor* is another specialized tool, whose purpose is to implement different standard procedures for representation and processing the simulation results. A sample list of such procedures includes the following: *(1) graphical representation of a functional dependency; (2) tabular representation of such a dependency; (3) graphical representation of a family of dependencies in a common co-ordinate system; (4) evaluation of the type and duration of a transient process; (5) evaluation of the model adequacy; (6) presentation of a test question; (7) comparison of the trainee's actions with those of the* author. Standard text files (*.dat*) are used to store the values of the traced parameters, measured in the real systems. After a given learner's session finishes the learner's session parameters, such as knowledge volume, duration, speed of learning, and so on are accumulated in the database as statistical experimental data, for the needs of an integrated system for individual planned teaching [11].

3. Language for Knowledge Description of Training Tasks

Although the tasks in a real lab vary greatly depending on thfe physical nature of the real objects they can be classified as (1) *calculation tasks* for a given set of output parameters, (2) *monitoring tasks for* dynamic mode of the system, (3) *investigative tasks* for static mode of the system, (4) *drawing tasks* for the transient process in the system, (5) *drawing tasks* for a set of functional dependencies, (6) *optimisation tasks* for choosing the most appropriate system according to some criteria, (7) *control tasks* to cope with abnormal situations, and (8) *diagnostic tasks* to identify failed system components.

Table 1. The common structure of programs

<session_description>::=**SESSION**	<task_description> ::= **SYSTEM** [<string>]
ORGANIZATION <string>	**FILENAME** <string>
DEPARMENT <string>	**DESCRIPTION** <memo>
TEACHER <string>	**DURATION** <integer>
COURSE <string>	**VOLUME** <integer>
TOPIC <string>	**PROMPT** <real>
GOAL <free text>	**DIFICULTY** <real>
DURATION <integer>	{<parameter description>}
VOLUME <integer>	{<dependence description>}
DIFICULTY <real>	[**COLOURS** <integer><integer>]
[<directives_list>]	[**DISCR_STEP** = <real>]
[<criteria_description>]	[{<event description>}]
{<task_list>}	[**TRACE** <list of parameters >**END**]
ENDS	[**SPEED** = <integer>]
<directives_list>::=	[**TIMER** @|<integer><integer><integer>]
ESCAPE | NOESCAPE	[{<procedural operator >}]
PRINT | NOPRINT	[{<operation >}]
SAVE | NOSAVE	**END**
EDIT | NOEDIT	< parameter description > ::= **VAR** <string>
DO | REDO	[**X**= <integer>]
ASSESS | NOASSESS	[**Y**= <integer>]
<criteria_description>::=	[**WIDTH** <integer>]
TYPE SUCCESS|	[**COLOURS_ON**]
FAILURE|	[**INVISIBLE**]
PERCENTAGE|	[**UNITS** <integer>]
SCALE	[**VALUE** = <real>]
CORRECTION <real>	[**STEP** = <real>]
MARK <string>	[**LIMITS** <real> | **@ :** <real> | **@**]
2-FROM: <integer1>**TO**:<integer2>	[**NORMAL** <real> | **@ :** <real> | **@**]
3-FROM: <integer1>**TO**:<integer2>	**END**
4-FROM: <integer1>**TO**:<integer2>	<dependence description> ::= **DEPENDS** <string>=<expression>
5-FROM: <integer1>**TO**:<integer2>	<event description > ::= **IF** < expression > **THEN** {<action1>} **ELSE** {<action2> }
6-FROM: <integer1>**TO**:<integer2>	**END**
END	

We called the special purpose language developed for task knowledge description *SystemScript*. It can be classified as *internal visual very high-level mark-up* language. In Table 1 a common structure of the generated training session's program and task's subprograms is presented using the Backus-Naur notation. Here the keywords of the language are in capital bold letters and the special symbols have the meaning of: **::=** defines a syntactical construction; _ connects the words in a syntactical construction name; **|** divides the alternative constructions; **{ }** enclose a construction, which can be repeated; **[ ]** enclose a construction which is not mandatory; **< >** enclose the name of a syntactical construction, which is not yet defined. In addition to administrative data, parameters of the session and the sequence of the training tasks, the program in *SystemScript* includes six global key directives. They are meant to allow the tutor to intervene during the learner's

performance. Their meanings allow or disallow the learner to: redo the task performance (DO|REDO); give up task performance (ESCAPE| NOESCAPE); printing the *.trn* file (PRINT|NOPRINT); saving this file (SAVE|NOSAVE), editing the file (EDIT|NOEDIT), assessing the learner's performance (ASSESS|NOASSESS). When assessment is allowed, an additional block for criteria description is added to the tutor's program. The tutor can choose between four types of the learner's assessment, e.g. SUCCESS/FAILURE, PERCENTAGE, MARK or PROXIMITY. In the last case the intervals of the traditional mark scale have to be pointed out. In such a way the adaptation of the environment to tutor's preferences are ensured.

In a task's subprogram three kinds of knowledge are embedded, namely: declarative, procedural, and operational. *Declarative knowledge* is represented in the form of parameter, functional, and conditional blocks. *Procedural knowledge* can be seen as a sequence of procedural operators for performing a given task in a standard manner after declarative knowledge comparison. The example list of operators includes: (a) *graphical representation of a functional dependence*, (b) *tabular representation of a functional dependence*, (c) *graphical representation of a set of functional dependencies on a common co-ordinate system*, (d) *evaluation of the character and duration of transient processes*, (e) *evaluation of model validity*. A multiple-choice question could be presented to test whether the learner observing the results of the procedural operator has made the right conclusion. *Operational skills* reflect the way of using declarative and procedural knowledge in order to cope with abnormal situations in pseudo-real time. An author's program for an investigative, control or diagnostic task requires specific learner's actions during its interpretation. Some of the other language constructions serve for flexible control of the simulation process. More detailed information about the language syntax and semantics for task knowledge description can be found in [12].

Table 2. A Sample BT

id	*kw*	*k*	*q*	*p*	*d*	*t*
Task1	simulation, static state, working characteristics, DC motor	2	197	0,36	0,50	9
Task2	examining, transient process, DC motor, DC generator	1	201	0,54	0,50	9
Task3	optimisation, criterion, choose, DC machine, internal parameters, characteristics	5	230	0,47	0,50	9
Task4	exploring, dependencies, effect, external parameters, rotational frequency, relation	3	198	0,36	0,50	9
Task5	monitoring, control, pseudo real time, maintain, rotational frequency	4	236	0,60	0,50	9
	Training session parameters (K, Q, P, D, T):	0.62	1062	0.47	0.50	45

4. Supporting the Base of Tasks

For each task in the training session the *BT* includes the following parameters: identifier (*ID*), set of keywords (*kw*), kind (*k*), knowledge volume (*q*), degree of prompt (*p*), degree of difficulty (*d*) as well as the time expected for its completion (*t*). The analogical parameters of the training session are denoted with corresponding upper letters.

For better illustration of the models proposed in this and next section, Table 2 presents a sample *BT* concerning the well-known system of a DC motor, DC generator and a mechanical connection between them. Some of the parameters have constant values calculated in accordance with the presented formulae, and others have statistical values with initial values shown.

Let Z be the maximal number of the different kinds of tasks, *y(j)* be equal to 1, if the j^{th} kind of task is present in the training session, and to 0, if it is absent. Then the degree of training session variety can be computed as $K = (\sum_{j=1}^{z} y(i)) / Z$. The concept of program tree and the correspondences between its terms and the graph terms was introduced by the authors in [6]. The task performing by the author or learner is viewed as filling in the program tree nodes with keywords, data types, and attributes' values (names, numbers, and text). The numbers of the nodes *V(i)* and links *U(i)* of the tree can serve as a precise and sensitive measure of knowledge volume within the text nodes dimensions, i.e. $q(i)=|V(i)|+|U(i)|\approx 2|V(i)|$.

Let *s(i)* be the number of elements included in the construction **SYSTEM**, *n(l,j)* – the number of elements included in the j^{th} **VAR** construction, o_i – the total number of operators in the j^{th} **IF-THEN-ELSE** construction. If *N(i)* is the number of parameters, *M(i)* – the number of the elements, *k(l,m)* – the number of terminals of element *m*, *Q(i)* – the number of connections, *L(i)* – the number of traced parameters, *P(i)* – the number of functional dependencies, *R(i)* – the number of conditional operators, *I(l,r)* – the number of operators in the **THEN**-part of the r^{th} **IF-THEN-ELSE** operator, *j(l,r)* – the number of operators in the **ELSE**-part of the r^{th} **IF-THEN-ELSE** operator. The nodes to the left of an *S*, *V*, *E, C* or *T* node contain information associated with the subject domain, while the nodes to the right represent the keywords of the attributes associated with the same construction. Then,

$$q(i)=2\left[\begin{array}{l} 2+2s(i)+2N(i)+2\sum_{j=1}^{N(i)}n(i,j)+18M(i)+ \\ 7\sum_{ij-1}^{M(i)}t(i,j)+6Q(i)+L(i)+P(i)+3R(i)+\sum_{i=1}^{R(i)}o(i,j) \end{array}\right]$$

Having *N* tasks planned for the training session, the total volume of acquired knowledge would be $Q=\sum_{i=1}^{N}q(i)$. Let $PT_1(i)$ and $PT_2(i)$ be respectively the author and learner's program tree. Let *a(l,j)* be the number of nodes missing in $PT_1(i)$ but present in $PT_2(i)$, *b(l,j)* – the number of nodes present in $PT_1(i)$ but missing in $PT_2(i)$. The following formula *c(l,j) = (q(i) – a(l,j) – b(l,j)) / q(i)* can serve as a precise and sensitive measure of the degree of proximity between the performance of l^{th} task by the j^{th} learner and that of the author. This parameter also varies between 0 and 1. The formula $c(l,j)^*=c(l,j).(\Delta t_1/\Delta t_2)$ that takes into account the learner's time for task performance Δt_2 relative to the author's Δt_1 presents time correction of *c(l,j)*. So by means of the environment parameter *e*, where $(\Delta t_1/e)\le\Delta t_2\le e\Delta t_1$, the learner who is faster than the author could be encouraged while the slower one could be reprimanded. The calculated parameter and the registered time for performance stored in the learner's *.trn* program present the main part of the third-level learner's model [5].The degree of environment prompt determines what part of the author's knowledge is available to the learner when he/she is performing the task. As it is assumed that the domain names are fixed by the author and the keywords by the environment

$$p(i)=\frac{1+s(i)+N(i)+\sum_{i=1}^{N(i)}n(i,j)+9M(i)+4\sum_{j=1}^{M(i)}t(i,j)+3Q(i)}{q(i)}$$

Now if *M* is the number of trainees who attempted to solve the l^{th} task, the formula for the calculation of the task difficulty becomes precise and simple: $d(i)=\sum_{j=1}^{M}c(i,j)/M$. The degree of difficulty of the session is calculated as the average, i.e. $D=(\sum_{i=1}^{N}d(i))/N$, where *N* is the number of tasks included in the session. Initially *d(i)* = 0.5, for every task, consequently *D*=0.5. In practice, the expected completion time for a task is calculated as the average, i.e. $\bar{t}_i=(\bar{t}_i+t_i)/2$ after each task execution by a learner. Similarly, the average time for completing the session is calculated as $\bar{T}=(\bar{T}+T)/2$ after each session completes. Having an initial session duration set by the tutor as *T*, the initial values of *t(i)* are taken to be the same and equal to *T/N*.

5. Planning of the Training Session

5.1. Search-based Models

Dominant model: This model is used to exclude the least appropriate tasks having extreme values. For example, given the criterion *(q=max) AND (p=max*), task 5 in Table 2 will be excluded.

Restrictive model: Used for reducing the set of tasks to a subset by imposing limits on certain parameters. For example, given the criterion *(p>0,5) AND (q>200*) tasks 2 and 5 will be selected.

Keywords: This model uses the *kw* field as a limiting factor. When more than one keyword is given, a logical *AND* is performed on them. The criterion for example, must be (*kw*="characteristics" AND "DC"), which will yield tasks 1 and 3.

Logical formula: This is a more general model than the preceding one, as it allows a combination of AND, OR, NOT operations as well as brackets. For example, in order to find all tasks, containing the terms "characteristics" but not containing "DC", the criterion will be (*kw*="characteristics" AND NOT "DC"), which will yield task 4.

5.2. Sort-based models

Significance model: The criteria are sorted by significance and the tasks are compared with the most significant one. The task, which outclasses the rest according to the first criterion, is selected. If more than one task has the same value, the next in significance criterion is used and so on. For example, having the following order of criteria: *1) d=max; 2) p=min; 3) q=max*, task 4 in Table 2 will be selected, but only after applying the third criterion.

Weighted model: Depending on the parameters' significance, each is assigned a weight. Then selection is based on the following formula: $A(i) = \sum_{i=1}^{n} w(k).p(i,k)$, where: *A(i)* – rating of the *i*th task; *w(k)* – weight of the *k*th parameter; *p(I,k)* – quantitative evaluation of the *I*th task for the *k*th parameter; *n* – the number of evaluated parameters. Supposing for a certain tutor can assign *t* a weight of 5, the weight of *p* can be 10, while those of *q* 0,01. Then the formula for the weighted sum of the *I*th task becomes: $A_i = 5*t_i + 10*Cp_i + 0{,}01*Q_i$, *(I=1..5).* In accordance with the calculated values of A_i , the tasks in the BT are ordered thus: 1, 4, 2, 3, 5.

Ideal model: Based on the idea that it is not mandatory that the significance of each parameter will grow as its quantitative evaluation. Generally, the selection criterion is: $D(i) = \sum_{i=0}^{n} w(k).|\ p(i,k) - v(i,k)\ |$, where the additional notations used are: D_i – degree of tutor's dissatisfaction with the *I*th task; *v(I,k)*– the ideal value of the *I*th task for the *k*th parameter. The lower the value of D_i, the higher the evaluation of the *I*th task. If it turns out that all parameters match their ideal values, then |*p(I,k)-v(I,k)*|=0, and consequently D_i=0. For example, keeping weights from the previous example and ideal values for a given T of *t*=5 and *p*=0, the tasks will be ordered thus: 3, 2, 5, 4, 1.

6. The Current State of the TOEDVL

Only a part of the above-described ideas have been implemented in the current prototype of the TOEDVL. The program generator is designed to run locally as a standalone WINDOWS application. Due to the need of portability and independency of the quality of the network connection status, the interpreter is implemented by means of ActiveX controls. Currently, we are in the process of implementing the Task Manager that allows easily switching between different sets of keywords, i.e. using different spoken languages.

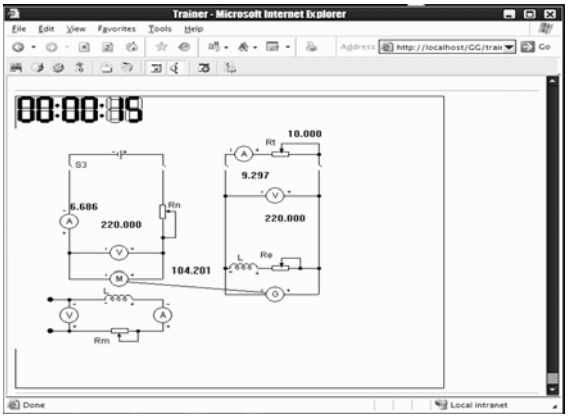


Figure 2. The main window of the program interpreter

A learner's session starts by opening a Web page; in this way the learner contacts the Task manager. After authorization, the learner is presented with the sequence of tasks planned in the *.ses* file. The program interpreter, if not present at the client computer, is downloaded from the server. As it starts interpreting the current *.trn* program, it downloads any additional files, needed for running the current task on the learner's computer. The monitored values of control points appear in the corresponding windows of scheme's bitmap (Figure 2). During execution they are refreshed depending on the model and user interactions. A simulator's clock in the toolbar indicates the elapsed time from the beginning of the simulation. The interpreter-evaluator periodically updates the learner's history file on the server. This file keeps track of all parameter changes, whether due to changes in the status of the modelled lab object, or to trainee's interactions with the model.

The prototype of the TOEDVL has been used in the course of Intelligent Teaching Environments for design of two virtual labs in electrical systems and electronic circuits. The accumulated experience indicates that the proposed environment leads to deep understanding of the teaching material and stimulates learners' motivation and activity in their practical exercises.

7. Conclusion

The problem of adapting a task-oriented training environment to different type of users, (e.g. authors, tutors, and learners) has been discussed. The models for computation of task and session parameters are proposed to support the base of tasks by the authors. Finally, two groups of the tutor's models based respectively on search and sort in the base of tasks are considered. Comparative analysis of the learner and author's program trees produces objective and precise evaluation of the learner's knowledge, taking into account the missing/redundant elements/connections as well as the task completion time.

8. Acknowledgements

The authors wish to thank the reviewer, Prof. Douglas Harms from DePauw University, Indiana, USA for his careful reading of the paper and critical comments and suggestions for its improvement.

Bibliography

[1] I.I. Zheliazkova, R.T. Kolev, Learner Models in the Computer-Based Teaching Systems. A survey, Bulgarian Journal of Information Technologies and Control, 4, 2004, 22-28.

[2] D.R. Peachey, G.I. McCalla, Using planning techniques in intelligent tutoring systems, International Journal of Man-Machine Studies, 24, 1986, 77-98.

[3] E.H. Paschin, A.I. Mitin, Automated Teaching System EKSTERN (MSU, Moscow, 1985, in Russian).

[4] C. Jesshope, E.Heinrich, D-r Kinshuk, Technology integrated learning environments for education at distance, http://www.deanz.org.nz/jeessope.doc

[5] A.J. Gonzalez, L.R. Ingraham, Automated exercise progression in simulation-based training, International Journal System, Man, and Cybernetics,24, 1994, 862-874.

[6] R.W. Chu, G.M. Mitchell, & P.M. Jones, Using the Operator Function Model and OFMspert as the basis of an intelligent tutoring system: towards a tutor/aid paradigm for operators of supervisory control systems, Int.J. System, Man, and Cybernetics, 25, 1995, 1054-1075.

[7] V. Vasandani, T. Govindaraj, Knowledge organisation in intelligent tutoring systems for diagnostic problem solving in complex dynamic domains, Int. J. Systems, Man, and Cybernetics, 25, 1995, 1076-1096.

[8] J. Fleming, The Intelligent Computer-Assisted Training Testbeds (ICATT) Program, 1996, https://www.spider.hpc.navy.mil/html-docs/its-ida/armstrong.htm.

[9] I.I. Zheliazkova, P.L. Valkova, Computer-Aided Teaching and Learning Structural Schemes, Annual Proceedings of the Rousse University, section "Mathematics, Informatics and Physics", Rousse, Bulgaria, 2004 (accepted).

[10] I.I. Zheliazkova, R.T. Kolev, Technology of Using an Intelligent Multimedia Test Generator, Proceedings of the 15th Annual European Conference on Innovation in Education for Electrical and Information Engineering, Sofia, Bulgaria, 2004, 49-58.

[11] I.I. Zheliazkova, R.T. Kolev, A Three-Level Learner's Model for the Needs of an Integrated Environment for Individual Planned Teaching, Proceedings of the International Conference on Computer Systems and Technologies (e-Learning), Rousse, Bulgaria, 2004, IV.13-1- IV.13-6.

[12] I.I. Zheliazkova, G.T. Georgiev, Representation and processing of domain knowledge for simulation-based training systems, International Journal of Intelligent Systems, 2000, 10(3), 255-277.

Author's Information

Irina I. Zheliazkova - Dept. of Computer Systems and Technologies, Rousse University, Studentska Str.8, Rousse, Bulgaria; e-mail: irina@ecs.ru.acad.bg

Georgi T. Georgiev - Dept. of Computer Systems and Technologies, Rousse University, Studentska Str.8, Rousse, Bulgaria; e-mail: gtgeorgiev@ecs.ru.acad.bg

Rumen T. Kolev - Dept. of Computer Systems and Technologies, Rousse University, Studentska Str.8, Rousse, Bulgaria; e-mail: rkolev@ecs.ru.acad.bg

THE AUTOMATION OF ADAPTIVE PROCESSES IN THE SYSTEM OF DISTANCE EDUCATION AND KNOWLEDGE CONTROL

Pavlo Fedoruk

Abstract: The article reveals a new technological approach to the creation of adaptive systems of distance learning and knowledge control. The use of the given technology helps to automate the learning process with the help of adaptive system. Developed with the help of the quantum approach of knowledge setting, a programming module-controller guarantees the support of students' attention and the adaptation of the object language, and this helps to provide the effective interaction between learners and the learning system and to reach good results in the intensification of learning process.

Keywords: distance learning, knowledge control, adaptive systems, knowledge quantum.

ACM Classification Keywords: information systems education, adaptable architectures.

Introduction

Our society undergoes drastic changes connected with the reconsideration of the whole range of scientific, political, and social aspects. They take place in all fields of social life and concern all public institutions including education. As a result of this, the education system also undergoes changes initiated by the system itself as well as under pressure of the changes in other fields. Among the factors which bring about the changes in the field of education, there should be mentioned the process of informatization of the society. This process promoted the appearance and development of different patterns of distance education. Thus, the appearance of distance education is quite a natural stage of the development and adaptation of education to modern conditions. One of the main advantages of the information technologies in the learning process is the possibility of learning individualization. Effective teaching is characterized by the effective use of such pedagogic means of upbringing, studying, and development that are adequate to students' individual peculiarities and help to achieve set educational goals with high effectiveness.

Nowadays such types of education as individual, adaptive, and individualized are distinguished [Mashbits, 1987].

Individual education is education, which is conducted according to the scheme: an instructor (a teacher or an automated educational system) – one student. Opposite to this one is group learning. Computer education can be both individual and group.

Adaptive education is education, which takes into account age as well as individual peculiarities of students. Adaptation may be based on the information collected by the system in the process of studying taking into consideration the learning history of every subject; it can be programmed in advance or it can be the combination of these two approaches.

Individualized education is education, which is based on a student pattern and initiates control actions taking into account this pattern. Educational process reasoning from the goals of individualized education must provide each student with the possibility of an independent choice of means and ways of educational work; methods and strategies of education; content, way, and form of presenting learning material [Iziumova , 1997].

The adaptive system of distance education with the use of information technologies has a number of advantages:

- it helps to reduce the unproductive work consumption of a teacher who, in this case, turns into a technologist of modern learning process in which the key role is assigned not so much to the educational work of a teacher as to teaching students;
- it gives students good possibilities of free choice of their own strategy and tactics of learning;
- it allows both students and a teacher to have effective feedback in the process of learning;
- it increases the efficiency and objectivity of control and the evaluation of the results of learning;
- it guarantees a continuous connection in the relationship "teacher-student";

- it encourages the individualization of educational work (the differentiation of learning speed, the complexity of educational tasks and so on);
- it helps to use a differentiated approach to students, which is based on the fact that different students have their own previous experience and level because every student comes to the process of gaining knowledge with their own store of knowledge which determines the level of their understanding new material and its interpretation. In other words, there is a turn from the acquirement of the same material by all students to the acquirement of "individual" material by different students;
- it increases the motivation of learning;
- it encourages the development of productive, creative functions of thinking in students; the increase of intellectual abilities; the formation of the operating style of thinking;
- it teaches how to work with modern information technologies.

Learning as an active process

In the given system we can present the conducting of education as control over the process of acquiring "skills." The traditional approach, which stated that a simple constant revision of material (with the exception of latent learning when a person acts as information storage) might yield satisfactory results, turned out to be not quite effective. Learning is an active process and gives a result only when there is motivation to learning. And what is more, education foresees certain efforts on behalf of a student. Thus, when the "learning system" interacts with a student, learning process is activated.

In traditional learning systems, a student, as a rule, provides an answer to a posed question or learning problem. To be more precise, they choose one answer out of the presented to them multitude of variants and themselves have to assess the "correctness" of each of the proposed to them variants. And though at present there is a great variety of such learning systems, in the view of cybernetics, all such invariable automated learning systems are "automatic controllers." They suggest the scheme of the feedback of the "known answer" type, which would provide motivation and an unknown (or if learning is based on revision, then – repeatable) set of tasks, which provide studying of certain learning material. Such a method is based on the assumption that there is the best way of learning, and it is embodied in the very system and in the system of finding solutions, which defines the action of the learning system. There are a lot of facts, which are evidence of quite satisfactory results that can guarantee such learning systems. But because of the invariable kernel of such a system, such a method works only for an "average" student, concerning those aspects of the action, which remain invariable even if averaged for a group of individuals.

Thus, the problem of the system adaptation to the needs of a student becomes actual. As the matter of fact, in real life a teacher-instructor, though s/he knows what s/he wants to achieve, has definite cautions concerning the way of achieving the defined aim, and s/he diligently adjusts his/her own methods of teaching to inconstant individual peculiarities of every person. Like a stationary program, s/he also watches students' answers. However, unlike a machine, a person is able to change the way of taking decisions, even the program of the course, and at the same time the interaction has a logical status of conversation which leads to the compromise between the participants at every stage. A personal teacher-instructor is at least an adaptive controller, and there is every reason to believe that a programming module, which corresponds to it, is more effective in practice than a stationary programming module. It is achieved due to introducing different scenarios, which react to a student's individual peculiarities revealed in the process of learning.

An Adaptive Teacher

Let a student have a task to process a knowledge quantum (the least indivisible notional piece of information, e.g. original notion, key word, axiom, definition, etc. [Fedoruk, 2005] [Sirozha, 2002]), which foresees performing some set of operations on the given quantum. A student knows the original state of a quantum (when no operation is performed yet), and the final state of a quantum (when the task is fulfilled) is chosen by him/her out of a multitude of variants-answers and is not known before the fulfillment of the task. Let there be four states of additional information: non-present information, α, β and γ; each of the proposed states characterizes the posed problem on a different side or with a different degree of refinement. As time is of critical value, a student is set an object to solve the problem as quickly as possible, and because of that optional value d is introduced; it

determines the time needed for the fulfillment of the posed task (in fact, it is time rate which is individually determined by a teacher in future). The result of the fulfillment is calculated index of the fulfillment success $\theta(t)$.

It is not very difficult to define restrictions on $\theta(t)$ under obvious conditions: the value must be minimum, if a student simply guesses the order of task fulfillment, and the value must be maximum, if s/he correctly determined the operations to be performed on the quantum and the order of their fulfillment. Let *i* be one of the typical operations the performance of which on the given quantum is possible. Let $\xi_r(t) = 1$ only when operation i must be performed on quantum r, and it was performed by the student, otherwise $\xi_r(t) = 0$. Let $R_i(t)$ be the value inversely proportional to the pending between the stimulus and reaction of the student. Let p_i be the probability of the appearance of necessity for processing quantum *i* in real life, and let $\chi_i(t)$ be the frequency of the appearance of the given quantum at the current stage of learning process. Then at the moment $t = t_0$, we can determine the average of successful task fulfillment considering all the quanta as well as interval $[t_0 - \tau]$ [Gordon Pask, 1961]:

$$\theta^*(t)_0 = \xi_i(t) \cdot R_i(t)[1-(p_i - \chi_i(t))^2]$$

It is certain that the magnitude of $\theta^*(t)$ is only one of numerous possible values; in particular, it gives no information about wrong answers (as at the given point of time we do not know the "importance" of mistakes) and has the value only when the right answer was given during the required period of time. Thus, we must introduce additional information to guarantee the possibility of finding the right solution. But if we do this, we will make the solution of the posed problem easier. The right answer given after receiving the additional information must be valued at lower points than the solution found without any additional information (prompt). So, there is the necessity of determining the "cost" of additional information. Let its value be δ, if such additional information was provided, and 0, if no additional information was given (there is no reason to believe that prompts α or β are more important than γ and vice versa)[5]. And finally, let $\theta(t_0)$ be the average of all magnitudes i and of interval $[t_0 - \tau]$ for the value

$$\theta(t)_r = \xi_i(t) \cdot R_i(t) \cdot \left[1-\left(p_i - \chi_i(t)\right)^2\right] - \delta(t).$$

Learning process in this case looks like the process of a constant search for the right order of performing the operations on the quantum with some volume of additional information, i.e. the sequence of problems. In case of optimally organized learning process, growth level $\theta(t)_r$ will be maximum.

The Adaptive System of Distance Education and Knowledge Control

According to these considerations, let's try and project the work of learning system in such a way that it should "learn" itself to determine the optimal order of learning (processing) separate quanta in order to provide the most effective work. The sequence of performing operations on the quantum determines the successfulness (partial successfulness or unsuccessfulness) of problem solution, and $\theta(t)$ will be used for the positive evaluation of the task fulfillment order. In process of fulfilling the posed problem, a student gets additional information in the form of theoretical material, examples, and previously gained knowledge and acquired skills. Moreover, in case of the correct performance of the $j-$th operation on the quantum, the possibility of operation $j+1$ being also fulfilled correctly increases. Thus, every operation connected with the individual value $\theta_i(t)$ and the system, on the whole, remains stable. There is probability that a student will try to perform operations on the quantum at random; however, the system can report it and set him/her on the right track with the help of additional information. Moreover such a structure of learning creates an additional motivation to study, producing an effect of competitiveness, the award in which is achieving the right answer within the shortest period of time and, consequently, getting higher points for the fulfilled task.

To understand the essence of the process, we should talk about *systems.* The brain has the ability to change in the process of its life, but as any other evolution system, it does not *learn.* A *learning* student is the system, which develops in the brain. When the system is stable, on the whole, its two subsystems (person and machine) are

inseparable and use the possibilities of the machine as own brain to solve the problem. However, it does not mean, yet, that physically they are a single whole. "Communication", which can guarantee such a state of things, leads to two formally separated activities [Gordon Pask, 1961]:

1. A controller must "hold" the attention of a student. The student is a system with a given set of allowable operations, e.g. u, that is, the student *is to* pay attention to certain tasks, and u determines the limits beyond which the data of a certain type must be processed. The solutions of a certain type are acceptable in order that the system could be called a "student." Let us assume that a student is able to focus his/her attention on a posed task. Then the diversity of the problem with respect to a student, to put it differently – its "complexity", is a multitude of operations and solutions, which are to be performed in order to fulfill the posed task (let us assume that the choice of necessary operations is made by truncating deliberately wrong variants until one variant of the answer or necessary action is left). Now in order to hold the attention of a student, the controller is to determine the sequence of the tasks in such a way that the average value of "complexity" equals, at least, u. If this condition is not met, a student will not be able to remain concentrated, and his/her attention will be redirected to another process, different from learning; but even the fulfillment of the posed condition does not guarantee that s/he will remain concentrated. However, if the conditions of point 2 are satisfied, $\theta(t)$ will assess how difficult it is to fulfill the task following the posed conditions.
2. Problems must be adequate to a student. At the lowest (elementary) level it is achieved with the help of the sub-controller of additional information. It gives a great deal of information owing to which a student may start performing the task, and it also determines the order of the entry of such information (either α, at first, or β depending on the needs of a student at a given stage of learning) as $\theta_i(t)$ for the $i-$th quantum increases.

But the problems are not accepted as something separated, their order of delivery is also a part of the "flexible" learning block. In its turn it depends on common rules according to which a controller functions. Any event which may be identified by organism or machine can perform several individual functions concerning probabilities of the behavior of the detector under given circumstances (that works under statistical parameter of the system, which is the detector). This function is its value.

Let's assume that the behavior of a person when s/he makes one decision is determined by matrix of changeable probabilities P_1, when s/he makes another decision – by matrix P_2, and P, the set of all such matrixes, determines the sensible part of his/her behavior. The fact that the stochastically disturbed system, which represents a person, may be described in such a way means that it is divisible. Thus, let us consider sample function $F*$, which can be controlled from the outside by some system with which a person works together, e.g. an instructor. In this case, some of the instructor's messages, which are meant for the student and aim at changing his/her position, incline him/her to a certain choice from set P. By definition, the meaning of such a message is its sample function regarding P. Let us remark that the meaning causes the appearance of the connection between the source of the message and the receiver. Thus, learning problems are used as a message with a certain meaning, as the process of making a decision can change a student's attitude to the problem and its perception, and under the "flexibility" of the lesson we understand the determination of conditions which can essentially influence the meaning of the message or, in other words, the adaptation of the object language of the learning material to the needs of a student.

Condition 1 cannot be fulfilled practically, if only partially condition 2 is not satisfied, i.e. if learning material is not adequate to a student, as $\theta(t)$ is a random value. It is obvious that condition 2 cannot be met without condition 1. So, everything depends upon a student, and no matter what efforts a teacher makes; there is no assurance that a student will go deeply into the subject. But when the "conversation" already started, the stability, which is the result of two-sided adaptation, inheres in it. A student and a computer reach a compromise.

The consequence of such interaction is the growth of value $\theta(t)$, which determines the assessment of its work; however, the result of reaching a certain desired level of interaction with the system is not less important. No doubt it is difficult to demand information potentialities of a human brain from the synthetically created system, even in communication only. The point is that the system develops not like an embryo, but like an autocatalysis. At the first step the presence of the learning machine gives a stimulus to the whole system, the organization which catalyses the appearance of a similar but bigger system. And it generates another one, also catalytic. In

learning systems we demand that the sequence of catalytic systems in its behavior should be directed to the growth of effectiveness, to the development of abilities and skills, in particular.

Conclusion

Using of the proposed technology helps to automate the learning process with the help of the adaptive system. In this system the automated controller should do the following in communication:

1. Hold a student's attention. This action is stimulating, as the widening of the range of problems in t increases the probability that a student will give a wrong answer in t. Besides this induces him/her to studying in order to get higher points at step $t+1$.
2. Adapt the subject language, which for the most part determines the successfulness of interaction.

In practice it is not always easy to separate one from the other; however, it is possible in case of clearly defined step-by-step learning process.

Bibliography

[Mashbits, 1987] Mashbits Ye.I. Psychological grounds of the educational work management. – Kiev: Vyscha shkola, 1987. – 224p.

[Iziumova , 1997] Iziumova S.A. Individual memory and learning process // Transactions of SGU. Series \ "Psychology and Sociology of Education'\. – 1997. – Ed. 4. – P. 10 – 23.

[Fedoruk, 2005] Fedoruk P.I. The technology of creating learning module in the adaptive system of distance education and knowledge control // Mathematical Machines and Systems. – 2005. – Issue 3. – P. 155-165.

[Sirozha, 2002] Sirozha I.B. Quantum modules and methods of knowledge engineering in the problems of artificial intelligence // Artificial Intelligence. Scientific and Technical Journal. – Donetsk, 2002. – Issue 3. – P. 161-171.

[Gordon Pask, 1961] Gordon Pask. An Approach to Cybernetics. London: Hutchinson, 1961. – P. 128.

Authors' Information

Pavlo Fedoruk – Director of The Information Technologies Centre, Precarpathian National Vasyl Stefanyk University, Shevchenka Str. 57, Ivano-Frankivsk 76025, Ukraine; e-mail: pavlo@pu.if.ua

REMOTE DEVELOPMENT OF DISTANCE COURSE USING VIRTUAL LEARNING SPACE 'WEB-CLASS KHPI'

Mykola Savchenko, Ganna Molodykh

Abstract: An approach to a specialized website creation – club of distance courses authors – on the basis of Virtual Learning Space "Web-Class KhPI" is implemented and suggested in the article.

ACM Classification Keywords: K3.1 Computer Uses in Education - Computer-assisted instruction, Distance Learning

Introduction

Nowadays, the majority of distance courses is developed using virtual learning spaces. At the same time the nature of author's work depends mostly on conceptual peculiarities of virtual spaces. If the space operates on basic learning materials as e-book, the work can be done on the author's personal stand-alone computer without any interaction with the learning space. Having some advantages this approach contains a number of considerable limitations which could appear while trying to make changes into the learning materials and the configuration online, to restrict access temporarily to some parts of the course. More problems appear when

organizing the learning process, based on setting rating credits to every learner, regulating access to the materials. These problems can successfully be solved in learning spaces which consider all distance course components as a dynamic structure, allowing the author to make changes into any component online. Such an approach gives a chance to develop the distance course and to control the learning process simultaneously that is a usual thing for teachers in Higher Schools, who have to correct their syllabus constantly, to implement new methodological techniques.

The teachers in our University can get acquainted with distance learning elements either visiting face-to-face lessons or participating in distance courses. Our experience shows that these lessons are not enough for teachers to design their distance courses of high quality. The typical workload for many teachers in Ukraine does not allow creating the course quickly and, consequently, makes impossible finishing this work, because many practical skills are lost as time goes by. The problem of time can hardly

be solved considering the existing workloads, thus, the solution should be found in supporting and developing practical skills in teacher's work with his/her distance course.

The possible solution can be found using Virtual Club Work organization for teachers or des

igners of new distance courses. What is seen here as a 'club' work? The most significant aspects of 'club' work are:

- individual schedule for every teacher;
- constant consultations with leading specialists in distance education;
- possibility to see the other teachers' works;
- objective evaluation of the performed work from the club tutors;
- gradual familiarity with the virtual environment peculiarities, mastering practical skills of work in a distance course as a Tutor;
- restricting access for strangers;
- opportunity for teachers to work from different geographical regions;
- opportunity for teachers to work with multi lingual and cultural peculiarities;
- having opened discussions on topical problems of Higher School.

Structurally the virtual club work can be presented in the form of interacting virtual learning spaces with a definite center - a place of work for all the project participants (See Fig.1). A virtual learning space for new courses is a place for individual work of every participant.

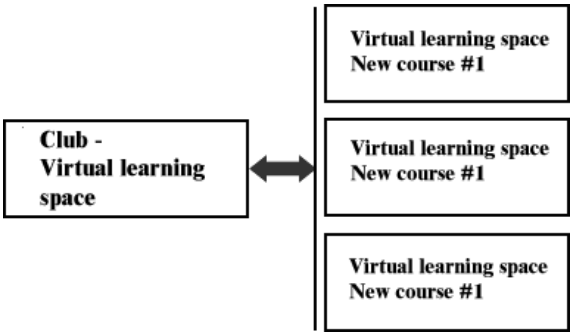


Fig.1. Structure for Virtual Club Work.

It is necessary to note that new courses can be developed by teachers from different educational establishments, who try to implement their conceptual approaches.

Analysis of Latest Research

Websites designers often appeal to a club work which is understood not only as a user registration but also involving him/her into the active development process of the website or its separate components. As an example there is a website for National modern poetry (http://www.stihi.ru/).

It is possible to place poems under any name on this website after the registration. After that poems get to the website rating and trying to reach more popularity and more readerships among the same authors-readers the author is included into an active process of literature communication and reviewing.

Another website (http://www.svistok.ru) is based on so-called large literature environment. This web resource contains many literature communities where authors, critics and readers can share their experience, communicate, present their books, announce their events, and make interesting discoveries in literature.

We think that the time has come to create community for authors of distance courses, directed to unite the distance courses authors and to activate the work in designing high-quality distance courses.

Research Tasks

The main tasks of this research are to create "**The Club for Authors' Distance Courses**" and to study the basis for such website functioning, considering the peculiarities of new distance courses designing. During the development and implementation of the project we studied the questions:

- structure of the basic theoretical material for a website and a club;
- types of practical tasks and conditions of their performance;
- character and forms of tutor-author and author-author interaction;
- forms of community work among participants.

On the whole this research is directed to working out practical recommendations for effective functioning of such websites.

Learning Material Presentation

As a main 'building stone' a Virtual Learning Space 'Web-Class KhPI' has been taken to implement this project.

This Learning Space uses a wide spectrum of learning services, modularity of its logical structure and file implementation. The system can be adapted to implement projects which are close to learning websites by idea content. It can serve as a basis for different interactive websites development. The experience of such work is presented in the distance course 'Creating interactive website on the basis of virtual learning space 'Web-Class KhPI' (http://dl.kpi.kharkov.ua/techn/nvs1/) [Goroshko, E.I., Molodykh, A.S., Savchenko N.V., 2005]. The results of work in this course show the simplicity of learning how to study in 'Web-Class KhPI' and the possibility to transform it to create websites with different informational orientation.

At present time the project is on the implementation stage. The club website can be found on the National Technical University 'KhPI' website (See. Fig.2) http://dl.kpi.kharkov.ua/techn/nvs2/

Fig.2. Starting Page of Distance Course with Club work.

Besides this website there are copies of this learning space to develop individual courses:

http://dl.kpi.kharkov.ua/techn/nvs2_01/ http://dl.kpi.kharkov.ua/techn/nvs2_02/ …

The work under a distance course for every teacher is being built according to individual schedule using consistent performance of these stages:

Stage	Title	Contents
1	Parameters	Parameters setup.
2	Planning	Developing course plan, generating prototypes, connecting resources.
3	Content	Working with learning pages prototypes. Editing separate resources. HTML-code optimization.
4	Testing	Tests development: to test knowledge (X, D-tests), psychological tests (P-test), Net test (N-test). Cards preparation.
5	Questioning	Cards preparation. Developing prototypes. Results analysis. Patterns.
6	Glossary	Glossary resources preparation. Working with Glossary resource database. Key words. Vocabularies profiles. Dictionaries review.
7	Communication	Planning Forums and Chats.
8	Starting Page	Starting Page management.
9	Rating	Development of separate tasks rating.
10	Course management	Tutor's work in the course.
11	Overall results	Summation.

To reach the strategy of consistent moving through the listed stages the teacher is granted a rating which regulates access to theoretical materials. In a case of successful performing a stage Tutor increases the rating. The advantage of the suggested scheme of creating new courses is in the fact that spaces of a Club and a New Course are identical.

Fig.3. Example of the individual Distance Course.

The author does not only create the course but also gets practical skills of working in it. All the listed stages of practical work are performed online. The majority of resources are stored in the course databases. An access to these resources is regulated by the author. It raises the safety of course materials significantly.

Working on creating distance courses, the author arranges the time to perform every suggested stage independently. Considering the fact that the subject of a stage can require significant efforts to create new components for the whole course, it is required to perform only a part of the whole work to increase the rating.

The practice shows that none detailed instruction for the author can replace individual consultations with a specialist. The importance of the consultation is vital in many cases when trying to solve the authors' difficulties.

Working with course authors the dynamic structure of the course staring page is used actively. The replaced news, links to the personal messages optimize the authors' work significantly.

Standard tools are used in the course to organize the communication: Forum, Course Mail, Chat.

Tutor of the Club can view new courses on-the-fly, can make useful corrections and suggest alternative variants to solve arising problems. Now there are attempts to create a collection of terminological vocabularies.

To stimulate Club participants work under their courses we suggested and implemented rating system approach to evaluate the performed work. Rating credits show the author a comparative laboriousness of some task, allow distributing efforts uniformly while performing the current stage.

After the project the author is given a copy of a virtual learning space ''Web-Class KhPI' with all the materials. The authors have all the rights to manage the materials at their discretion.

Conclusions

A website of distance courses authors in a form of a Club is suggested and implemented. The present method allows solving a whole series of contradictions arising when creating new distance courses. The approach allows widening teachers' geography interested in practical aspects of distance learning. The project is on the stage of implementing now, but this information will certainly be interesting for a wide range of teachers in Higher Educational Establishments attempting to inform about themselves in Internet.

Bibliography

[Savchenko, N.V., 2005] Savchenko, N.V. O distantsionnom kurse 'Sozdanie interaktivnogo saita na baze spedy Web-Class KhPI' (About the distance course 'Creating interactive website using virtual space 'Web-Class KhPI'). International Scientific Practical Internet-Conference 'Computer Sciences in Science and Education', Russia, Voronezh, November, 2005 http://www.comp-techn.nm.ru/Section5/doc4.htm

[Goroshko, E.I., Molodykh, A.S., Savchenko N.V., 2005] Goroshko, E.I., Molodykh, A.S., Savchenko N.V. Ispolzovanie virtualnoi uchebnoi sredy 'Web-Class KhPI' v kachestve interaktivnogo saita kafedry Mezhkulturnoi kommunicatsii I inostrannogo jazyka (Using Virtual Learning Space 'Web-Class KhPI' as an interactive website for the Cross-Cultural Communication and Foreign Languages Department). - «Obrazovanie i virtualnost - 2005», Proceedings of 9th International Conference of Ukrainian Association of Distance Education. Edited by V.A. Grebenyuk and V.V. Semenets. – Kharkiv – Yalta: UADO, 2005. – pp.47-50.

[Savchenko, M., Kukharenko, V., Molodykh, A., 2004] M. Savchenko, V. Kukharenko, A. Molodykh. Algorithms of Tutor's Actions While Conducting Distance Courses in Virtual Learning Environment "Web-Class KhPI". Proceedings of the Fourth International Conference Internet-Education-Science IES- 2004, October 5-9 Vinnytsia, vol. 1, pp. 205-208 (UNIVERSUM - Vinnytsya, 2004) http://dl.kpi.kharkov.ua/techn/rle/ (See «Публикации…»).

Authors' Information

Mykola Savchenko – Ph.D. in Physics and Math., Technical Cryophysics Department, National Technical University "KhPI". Address: Frunze st., 21. Kharkiv, Ukraine, 61002. e-mail: nsavchenko@kpi.kharkov.ua

Ganna Molodykh –National Technical University "KhPI"; e-mail: molodykh@kpi.kharkov.ua

E-SUPPORT FOR ENGLISH LANGUAGE COURSE TO ENHANCE UNIVERSITY STUDENTS' LEARNING

Ganna Molodykh

Abstract: The peculiarities of English language teaching for students at higher educational establishment using some elements of distance learning, developed by the author, are described in this article. The results of students' questioning, received at the end of the experimental teaching, are suggested and analyzed. The conclusions are formulated and the further ways of teaching English with e-support are outlined.

Keywords: Distance Learning, methods of teaching foreign languages, course with e-support for teaching, new functions of virtual learning space, questionnaire.

ACM Classification Keywords: K3.1 Computer Uses in Education

Topicality of teaching foreign languages using face-to-face lessons and e-support

Under the conditions of worldwide globalization and a tendency of Ukraine to be integrated into the world community, the knowledge of English at high level becomes one of the compulsory requirements for any specialist in any field of human activities. Teaching English at the Cross-Cultural Communication and Foreign

Languages Department in National Technical University "Kharkiv Polytechnic Institute" (http://users.kpi.kharkov.ua/ccfl/gor) is directed not only to the receiving traditional foreign languages skills – writing, reading, speaking and listening. We also make a stress on using basic skills in e-communication, on the ability to interact with Internet, learning spaces and to work with different communication tools in our teaching, which any young specialist – a graduate of our University – will face with in his/her professional activity.

Nowadays in Ukraine there are some examples of pure distance training of foreign languages specialized courses, such as, for example, in G. Kaluzhna's course [Kaluzhnaya, G., 2006] or B. Shunevich course [Shunevich, B., 2006]. In these courses the stress is made on written English, which could be taught and is easy to be checked at a distance.

The peculiarity of our teaching is the fact that we combine face-to-face and distance learning during the learning process. It is connected with the different level of our students' preparation and, correspondingly, the differences in their needs (speaking, studying grammar, development of listening skills etc.).

To teach beginners a lively talk, intonation, manners of concrete behavior is impossible without regular face-to-face students' contact with each other and the teacher. However, there are no developed methods and instructions for teachers about the combination of face-to-face and distant activities. That's why the article is devoted to solving this question.

Developing methods for foreign languages teaching using e-support

To study a foreign language fully while using traditional Distance Learning (DL), i.e. without face-to-face support, is practically impossible. The presence of live communication with a teacher and other students to receive maximum speaking practice is compulsory.

While implementing the distance course into the foreign languages learning it was noted that learning spaces do not correspond to the specific demands for the learning process being made of foreign language teachers. Specific teaching features of this subject made us to widen the functions of the space used in learning. As a result the virtual learning environment "Web-Class KhPI" [Web-Class KhPI, 2006] has been elaborated to be used in a foreign languages teaching.

The functions being performed now by the learning environment as e-support while teaching foreign languages:

1. Bulletin Board with the information about the homework for the next practical lesson, about the requirements to pass a credit or exam and other announcements;
2. Backed up materials of scanned pages from text-books to be uploaded and printed by students;
3. Providing audio files (*.mp3) with authentic foreign speech to be uploaded and listened. This allows students to perform all the listening tasks independently, at his/her speed, with as many times for listening as needed.
4. Training before e-testing using the developed interactive tasks for students' self-control.
5. E-tests with unlimited or limited times for passing them. Students' interest in being honest is provided by a face-to-face exam or credit at the end of the course (depending on the syllabus).
6. Informing students about the results of assessment according to the rating system;
7. Help while performing home task due to the presence of the built-in e-vocabularies on the web-site and links to the modern e-dictionaries in Internet;
8. Links to the backed up materials for additional home reading indicating the volume of the text in symbols and in megabytes.
9. Possibility for the teacher to replace additional Internet links on the web-site to organize students' project work.

All the mentioned peculiarities above have been developed for the distance course "English for Information Technologies" (see Fig.1).

A question of current importance is: which of these activities could be performed individually by students, which could be performed in a face-to-face mode.

With a face-to-face support all kinds of students' activity should be performed at the beginning of work with a learning space (the first two-three weeks of learning). And later:

- listening (to have discussions according to the materials being heard),
- performing interactive tasks for self-control and training before passing the control testing;
- testing (to provide students' independent work without any help, to receive objective results).

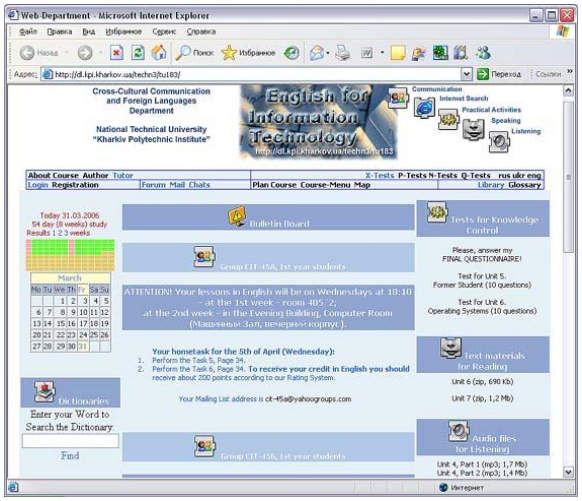


Fig.1. Distance course «English for Information Technologies», http://dl.kpi.kharkov.ua/techn3/tu183

At the same time students' work using computer should necessarily take turns with oral communication to a teacher and other students. Our experience shows that the organizational difficulties appear during a face-to-face lesson if the teacher began it with the work using computers. After that students can hardly switch to face-to-face activities, i.e. working with different speed they cannot finish performing the computer tasks simultaneously. That's why the work using computers should be left till the final stage of a lesson.

For an independent students' work, in fact, any practical activity can be put. However, it is possible only when the teacher is sure that all the students can log into the system and know what and where can be found on the web-site and how to work with it.

Practical Research Results

The Distance Course "English for Information Technologies" is placed in Internet http://dl.kpi.kharkov.ua/techn3/tu183, and also in the local area network of the University, so it can be reached easily and freely by students. 25 students studied in the English course during one term – from September till December, 2005. The compulsory activity in their learning was visiting face-to-face lessons and using some elements of DL (in particular, electronic testing).

At the end of the course the author carried out a questionnaire among students that allowed making some conclusions about the success of using such learning for foreign languages teaching in the Higher School. Let us consider the results of the questionnaire more precisely.

1. Why do you think e-testing was used in our course? Was it helpful for you? Why?

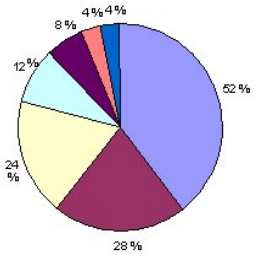


Fig.2. Reasons for using e-testing by students.

52% - it is very convenient, tests could passed any time and in any place;

28% - it helps to improve my knowledge of English;

24% - it helps to put into practice my self-evaluation;

12% - it helps teacher to put into practice the evaluation of students' level;

8% - it saves teacher's time on control implementing;

4% - it helps to get the additional skills of work using computers;

4% - it is hard to say.

2. Did you like studying the English course with distance learning elements? What did you like / did not like and why?

96% of students noted that they liked to study in the experimental course. Among the reasons they mentioned the following:

- there is an opportunity to put questions exactly when they appear, without looking for the teacher at the University or waiting for the next lesson;
- perfect design of the web-site;
- saving time;
- mobility and understanding of the teacher;
- it was possible to check the mistakes if the test were passed unsuccessfully and to correct them;
- it was easy to get access to all the materials and tests sitting at home;
- this term was the most interesting in the whole course of English;
- it was interesting because it was a new form of learning.

Among the things that they did not like, one student mentioned the presence of some misprints in formulations of test questions. It requires further course improving.

3. What was not sufficient in the course?

60% - everything was sufficient;

24% - a number of face-to-face lessons was not sufficient;

8% - a stable and constant access to Internet was not sufficient;

4% - I would like more attention to the English grammar;

4% - I would like more attention to listening.

4. Was the workload in the course sufficient for you during this term?

8% - yes, it was more than sufficient;

60% - yes, it was sufficient;

20% - no, it was not rather sufficient;

8% - no, it was not sufficient.

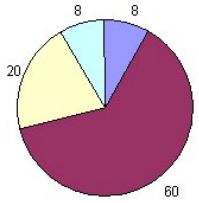


Fig.3. Adequacy of workload in the course (%).

5. What was useful for you in such e-support (distance course) in studying English?

- I became more interested in English and began to study more.
- I received more practice using Internet.
- I could download all the necessary materials easily, without leaving my home.
- I have improved my knowledge of English.
- It was convenient to answer the test tasks.
- I maintained close contact with the teacher.
- I have seen my gaps in my knowledge of English and understood what should be improved.

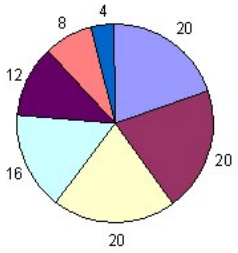


Fig.4. What disturbed you? (%)

6. What disturbed you in studying English?

20% - work, a lack of time;

20% - I felt too lazy and absent-minded;

20% - nothing;

16% - a big workload at the University in other subjects;

12% - poor Internet connection;

8% - uncomfortable schedule;

4% - backward group mates in the level of English.

7. Have your skills of working with computer improved after this course?

40% - no, they have not improved;

36% - yes, they have improved a little;

16% - yes, they have improved.

At the same time among students whose skills, according to their answers, have not improved, there were people who noted that their level of working with computer was very high before the course. That's why we can make a conclusion that using the elements of DL in teaching foreign languages has a positive effect on students' computer literacy improvement.

8. Has your vocabulary in professional terms been improved after this course?

0% - no, it has not been improved;

36% - yes, it has been improved a little;

44% - yes, it has been improved;

20% - yes, it has been improved considerably.

Conclusions

The results of students' answers to the questionnaire allow making a conclusion that it is necessary to use the elements of DL in teaching foreign languages. It is interesting for students, it allows them to control their success in learning better and easier, it improves students' vocabulary in professional terms and improves their computer literacy.

Our experience and questionnaire showed that even in groups with homogeneous participants in the level of English knowledge there are often students whose experience in using English does differ significantly from the experience of other students. That's why the teacher, using e-support in teaching, should allocate the learning materials of different complexity levels, additional materials in grammar, listening, video and audio materials for student's independent study on the web-site. Students' additional activity can be not evaluated in the rating system of the course; however, it allows students who wish to study more to get an access to any necessary materials.

In future the author plans to consider the students' wishes, to elaborate and improve the testing system of the distance course. The necessity of individual approach to any learner requires considering all the wishes that students had, that's why it is planned to add more additional text materials of "Upper-Intermediate" and "Advanced" levels.

Bibliography

[Kaluzhnaya, G., 2006] Kaluzhnaya, G. Distance Course "Fundamentals of Academic Writing: Writing an Essay", http://dl.kpi.kharkov.ua/techn3/gklv

[Shunevich, B., 2006] Shunevich, B. Distance Course "English for Technical Students". http://dl.kpi.kharkov.ua/techn3/tu187

[Web-Class KhPI, 2006] Web-Class KhPI Environment. http://dl.kpi.kharkov.ua

Authors' Information

Ganna Molodykh – Educational Environments Institute, Pedagogical Sciences Academy of Ukraine. Address: Barabashova st., 38-A, room 47, Kharkiv, Ukraine, 61168. e-mail: molodykh@kpi.kharkov.ua

ABOUT MASSACHUSETTS COLLEGES ONLINE JUNE 2006 SHARING BEST PRACTICES CONFERENCE

MICROCOSM OF ISSUES AND DISTANCE LEARNING DEVELOPMENT IN THE UNITED STATES OF AMERICA

Vincent Ialenti

Abstract: This paper summarizes trends and issues in online learning in the United States of America as reflected in the presentations of the Massachusetts Colleges Online "Sharing Best Practices in E-Learning" Conference held on June 13-14, 2006.

The Commonwealth of Massachusetts' nine state-supported four-year colleges and its15 two-year community colleges formed Massachusetts Colleges Online (MCO) in 2001 (MCO, 2006). Among the consortium's core missions were to exchange online courses among institutions and support the growth and development of distance learning in the state-run colleges. A visit to the organization's website shows that the member colleges have developed a pool of close to 1200 courses that are available to students worldwide. In an attempt to foster online learning development among the member institutions, an annual conference was initiated in 2004 with a theme appropriate to its mission: "Sharing Best Practices in E-Learning" (MCO, 2006).

The 48 presentations at the Third Annual MCO Conference, held on June 13-14, 2006 at Middlesex Community College in Lowell, MA, were initiated and offered principally by the faculty and administrators involved in MCO's online learning efforts. The themes and topics of the presentations serve as an excellent indication of the rapid evolution of online learning, and the practices, principles, and issues that currently concern Massachusetts practitioners, and reflect similar interests among other U.S. distance learning educational institutions. The following pages will summarize several of these major issues as indicated by an analysis of the conference's presentations.

There are many signs of online learning's near universal acceptance as a course delivery modality. The 2005 edition of the Sloan Consortium's annual survey "Growing by Degrees: Online Education in the United States" "...concludes that the breadth of online college courses may soon rival traditional face-to-face offerings. Survey results show more than three out of five institutions offering face-to-face undergraduate (63%) or graduate (65%) level courses also offer courses at the same level online. In addition, larger percentages (56%) of chief academic officers agree that online education is critical to their long-term strategy" (Sloan-C, 2005).

The U.S. Department of Education recognized the impact and growth of online learning in its recent report "Assuring Quality in Higher Education: Recommendations for Improving Accreditation" based on a series of discussions with the nation's regional and national college accrediting organizations (USDOE, 2006). A well-attended session at the MCO conference featured a presentation by an executive of the New England Association of Schools and Colleges (NEASC), the organization that accredits MCO colleges. The USDE report was discussed along with NEASC's "Best Practices for Electronically Offered Degree and Certificate Programs" white paper that suggests guidelines and highlights issues that may represent problem areas as college's offering online courses and programs undergo the accrediting renewal process (NEASC, 2006).

NEASC recently issued a "Statement of Commitment by the Regional Accrediting Commissions for the Evaluation of Electronically Offered Degree and Certificate Programs." NEASC and the other accrediting entities have taken a national, as opposed to a regional, approach to developing accrediting standards that will serve as guides, as

colleges develop and assess their online education programs since "… new delivery systems are becoming increasingly important, with institutions developing national and international student populations enjoying only virtual residence, the regional commissions have sought and will continue to seek a significant degree of cross-regional consistency, compatible with their independence and autonomy, in evaluating these activities" (Statement, 2006).

Another presentation reported on the June 2006 National University Telecommunications Network conference devoted to the theme "Managing and Maintaining Quality in Distance Learning" (NUTN, 2006). In addition to the accrediting organizations' efforts, other online quality assurance initiatives have been developed to assist colleges to assess the quality of their distance learning programs such as the State of Maryland's "Quality Matters" Rubric (Quality, 2006),The Interactive Quality Assessment Tool (Iqat, 2006) the Baldridge National Quality Program (Baldridge, 2006), and the Sloan-C Pillars (Sloan-C, 2006).

Technological advancement and its role in online learning is manifest in the fact that seven of the MCO conference presentations dealt with the integration of faculty-produced audio and video sources into online courses. The easiest technology to add to a course is audio. The popularity of portable mp3 audio playback devices, most notably the Apple iPod system (Podcasting, 2006) and the recent development of portable units that deliver video as well as audio playback, is being viewed as a method of combining traditional features of classroom pedagogy into the online courses. Sessions were presented on creating "podcasts" that can be delivered to students via the commonly used course management systems Blackboard or WebCt or through Apples iTunes software.

Faculty incorporation of technology into their online courses is often dependent on the technology's ease of use. Audio "podcasts" have been implemented with minimal technological challenges. Some faculty have recorded portions of their on-campus course lectures and uploaded them into their online courses. Others record lectures, updates, and content explanations specifically for their online courses. iTunes software and RSS subscription feeds can download the content to the students whenever new audio files are posted by the instructor. (Podcasts, 2006)

As video technology becomes simpler for end users, some faculty are self-producing video segments and demonstrations and distributing through their college's streaming video servers using desktop computer video editing and recording software, such as Apple's iMovie (iMovie, 2006). At the opposite end of the video sophistication scale is a movement that is just beginning in the MCO system. The approach is to use high-level video/computer technology to automate the process of recording a faculty member's lectures along with all the visuals used in the class, dividing the content into segments based on the visuals, processing, and uploading the material to a streaming video server with minimal faculty awareness of the technological processes. The current high cost of implementing such automation is preventing its widespread use in state-funded institutions (Anystream, 2006; Accordant, 2006; Telecast, 2006; Viewcast, 2006).

However, other lower cost technologies are beginning to be championed and incorporated into online courses. One presentation demonstrated Microsoft Producer, which is a free download to licensed users of Microsoft Office's PowerPoint 2003 (Microsoft, 2006). The program allows an instructor to easily incorporate audio and video segments along with PowerPoint slides to create rich media presentations that can engage students and provide video online course support. The Microsoft Producer program compresses the final video product so that it can be posted on an ordinary web server.

Another presentation demonstrated Camtasia Studio, which is a low-cost software package that records real, full-motion video of anything on the computer screen with an audio voiceover (Camtasia, 2006). This product is ideal for demonstrating on-screen processes and to creating online tutorials, such as how to search a library database

or use a software program. Instructors have been using the product to reinforce concepts and to recreate for online students experiences that would normally be demonstrated in a face-to-face class. Still another presentation showed how an online instructor uses Macromedia Captivate software to create flash videos of step-by-step accounting and finance problems that are normally done in the classroom (Macromedia, 2006).

One of the principle concerns of NEASC and other accrediting organizations is that an institution offering online courses "…recognizes that appropriate service must be available for students of electronically offered programs, using the working assumption that these students will not be physically present on campus" (Best Practices). Several presentations dealt with one of the most challenging of these concerns: how colleges can provide tutoring services to online students. It appears that many of the MCO institutions try to extend tutoring services by creating electronic components for the activities that are normally delivered on campus, such as library instruction and online writing labs. Pilot projects using commercial firms that provide online tutoring services have been successful for the most part, but the institutions have universally agreed that the cost generally has prevented them from expanding the scope of these services.

One presentation outlined an MCO 2006-07 academic year pilot project in which seven of its members will partner with the Connecticut Distance Learning Consortium's eTutoring program (CTDLC, 2006). This will allow the MCO colleges to join the five-year-old CTDLC program in order to provide MCO students with a full complement of online tutoring services. As participants, the MCO colleges will use a CTDLC designed and developed web-based platform. Professional tutors in a variety of topics will be trained at MCO pilot institutions, and will contribute weekly tutoring hours to the CTDLC pool. The hours are aggregated into a schedule that provides students with online tutoring resources seven days a week from early morning though late evening.

The topic of the MCO Conference keynote presentation is a national and international example of the trend to collaborate and share in the online education world. Executive Director Dr. Gerard L. Hanley of the Multimedia Educational Resource for Learning and Online Teaching (MERLOT) described and demonstrated the project (MERLOT, 2006). MERLOT, designed for higher education faculty and students, is a free, open online service that provides a continually growing collection of online learning materials, assignments, and reviews. According to Hanley "MERLOT helps faculty enhance instruction by providing easy access to online materials connected to strategies for integrating them into teaching and learning" (Hanley, 2006).

In a follow-up workshop, Hanley detailed how publishing in the MERLOT international academic community is becoming part of a faculty member's portfolio for tenure and promotion at many institutions. Among MERLOT's international partners are Canada's Co-operative Learning Object Exchange, Europe's Ariadne Foundation, and Australia's EdNA Online (CLOE, 2006; Ariadne, 2006; EdNA, 2006). Hanley also demonstrated how the MERLOT service allows faculty to browse and search the collection of learning objects, learn from peer reviews of the content, create personal collections of MERLOT resources, and browse and participate in the Virtual Speakers Bureau (MERLOT, 2006).

The above analysis demonstrates that the dynamic nature and rapid growth of the online learning modality is making a strong impact on higher education. Faculty and administrators need to be continually aware of the changes that online learning introduces to the traditional nature of higher education. Conferences such as MCO's "Sharing Best Practices" are invigorating as one becomes excited by the many developments and improvements to course design and delivery. Yet, such summits can also bring one the revelation that it can be a nearly overwhelming challenge to keep up with such change.

Bibliography

1. Accordant Technologies. Retrieved July 3, 2006, from http://www.accordent.com/
2. Anystream. Retrieved July 3, 2006, from http://www.anystream.com/industries-edu.asp
3. Apple iMovie [Computer software]. (2006). Cupertino, CA: Apple Computer.
4. ARIADNE Foundation for the European Knowledge Pool. Retrieved July 3, 2006, from http://www.ariadne-eu.org/
5. Baldridge National Quality Program. Retrieved June 13, 2006, from http://www.quality.nist.gov/
6. Best Practices for Electronically Offered Degree Programs. Retrieved June 25, 2006, from http://www.neasc.org/cihe/best_practices_electronically_offered_degree.htm
7. Camtasia Studio [Computer software]. (2005). Techsmith. Retrieved July 3, 2006, from http://www.techsmith.com/camtasia.asp
8. Camtasia Studio [Computer software]. (2005). Techsmith. Retrieved July 3, 2006, from http://www.techsmith.com/camtasia.asp
9. Co-operative Learning Object Exchange. Retrieved June 21, 2006, from http://cloe.on.ca/
10. Connecticut Distance Learning Consortium eTutoring. Retrieved February 5, 2006, from http://www.etutoring.org/
11. Connecticut Distance Learning Consortium eTutoring. Retrieved February 5, 2006, from http://www.etutoring.org/
12. EdNA Education Network Australia Online. Retrieved July 3, 2006, from http://www.groups.edna.edu.au/
13. Hanley, Gerard. "Taste of MERLOT," keynote address, Massachusetts Colleges Online Conference, Lowell, MA. June 13, 2006.
14. Interactive Quality Assessment Tool. (2006). Retrieved July 3, 2006, from http://www.iqat.org
15. Macromedia Captivate. (2006). Adobe. Retrieved from http://www.adobe.com/products/captivate/
16. Massachusetts Colleges Online. Retrieved July 3, 2006, from http://www.mco.mass.edu
17. MERLOT (Multimedia Educational Resource for Learning and Online Teaching). Retrieved April, 2006, from http://www.merlot.org
18. Microsoft Producer 2003 [Computer software]. (2003). Redmond, Washington: Microsoft. Retrieved May 13, 2006, from http://www.microsoft.com/office/powerpoint/producer/prodinfo/default.mspx
19. Moore, J. C. (2005). Synthesis of Sloan-C Effective Practices, August 2005. Retrieved July 3, 2006, from http://www.sloan-c.org/effective/index.asp
20. Podcasting in Education. Retrieved June 22, 2006, from http://www.apple.com/education/solutions/podcasting/
21. Podcasts. Retrieved July 3, 2006, from http://www.apple.com/itunes/podcasts/
22. Quality Matters: Inter-Institutional Quality Assurance in Online Learning. Quality Matters. Retrieved June 10, 2006, from http://qualitymatters.org/
23. Schray, V. (2006, May 17). U.S. Department of Education. Retrieved July 1, 2006, from http://www.ed.gov/about/bdscomm/list/hiedfuture/reports/schray2.pdf; Assuring Quality in Higher Education: Recommendations for Improving Accreditation
24. Statement of Commitment by the Regional Accrediting Commissions for the Evaluation of Electronically Offered Degree and Certificate Programs. Retrieved July 3, 2006, from http://www.neasc.org/cihe/commitment_electronically_offered_degree.htm; Statement issued jointly by MSA - Middle States Commission on Higher Education; NASC - Northwest Association of Schools and of Colleges and Universities; NCA - Higher Learning Commission of North Central Region; NEASC - New England Association of Schools and Colleges; SACS - Southern Association of Colleges and Schools; WASC - Western Association of Schools and Colleges
25. Telestream. Retrieved from http://www.telestream.net/solutions/markets_edu.htm
26. Viewcast. Retrieved from http://www.viewcast.com/pages.asp?page=niagara-GoStream

Authors' Information

Vincent S. Ialenti – Ed.D., Mount Wachusett Community College; 444 Green Street; Gardner, MA 01440 USA; e-mail: vialenti@mwcc.mass.edu

60th Anniversary of

INSTITUTE OF MATHEMATICS AND INFORMATICS of Bulgarian Academy of Sciences

Acad. G. Bonchev Str., block 8, Sofia 1113, Bulgaria
Tel. (+359-2) 979-3824, Fax (+359-2) 971-3649
http://www.math.bas.bg

The Institute of Mathematics and Informatics (IMI) at BAS was founded in 1947 as Institute of Mathematics. At the beginning about ten research fellows were working at the Institute. In 1961 a computational centre was established as part of the Institute. Later specialist in Mechanics also worked at the Institute, hence and it was named Institute of Mathematics and Mechanics. Its present name dates from 1995

The Institute has considerable achievements in the field of Mathematics that are not discussed here.

The development of the Informatics in Bulgaria started at the Institute. Many researchers have built the career of Informatics specialists.

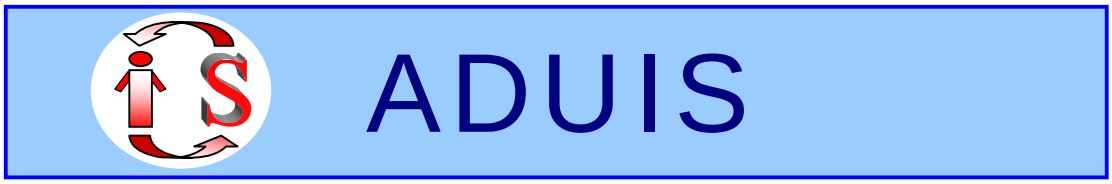
The Institute was the first in Bulgaria to buy an universal analog computing machine MH-7. The first Bulgarian computer was created at the Institute. Soon after that came into exploitation the first imported into Bulgaria computer “MINSK-2”. An original software for this computer – auto code “MIKOD”, operation systems “MID” and “MID-2”, a system for symbol programming “MIKS” and a rich library of programs were created here as well.

The fellows of the Institute also carried out the first Informatics researches in Bulgaria. The Institute has a wide range of activities in Applied Informatics and it continues to produce original software for the solving important problems. Researchers from the Institute organized and taught the first courses in Informatics at the Sofia University “St. Kliment Ohridski” for students in Mathematics. In a short time a major in Informatics was launched with the help of the Institute and later on it became a specialty at the Sofia University. Researchers of the Institute prepared the first syllabus, textbooks, and manuals. The staff of the Institute is also involved in training teachers in Informatics for the secondary school.

In the course of the years the informaticians at focused upon the research activities and many of them are still lecturing Informatics at a number of Bulgarian universities.

Departments of IMI : *Algebra; Artificial Intelligence; Biomathematics; Complex Analysis; Differential Equations; Education in Mathematics and Informatics; Geometry and Topology; Information Research; Laboratory of Mathematical Linguistic; Logic; Mathematical Foundations of Informatics; Mathematical Linguistics; Mathematical Physics; Computational Mathematics; Operation Research; Probability and Statistics; Real and Functional Analysis; Software Engineering; Telecommunications Department.*

15th Anniversary of



ASSOCIATION OF DEVELOPERS AND USERS OF INTELLIGENT SYSTEMS

ADUIS consists of about one hundred members including ten collective members. The Association was founded in Ukraine in 1992. The main aim of *ADUIS* is to contribute to the development and application of the artificial intelligence methods and techniques. The efforts of scientists engaged in *ADUIS* are concentrated on the following problems: expert system design; knowledge engineering; knowledge discovery; planning and decision making systems; cognitive models designing; human-computer interaction; natural language processing; methodological and philosophical foundations of AI.

Association has long-term experience in collaboration with teams, working in different fields of *research and development*. Methods and programs created in Association were used for revealing regularities, which characterize chemical compounds and materials with desired properties. Some thousands of high precise prognoses have been done in collaboration with chemists and material scientists of Russia and USA.

Association can help *businessmen* to find out conditions for successful investment taking into account region or field peculiarities as well as to reveal user's requirements on technical characteristics of products being sold or manufactured.

Physicians can be equipped with systems, which help in diagnosing or choosing treatment methods, in forming multi-parametric models that characterize health state of population in different regions or social groups.

Sociologists, politicians, managers can obtain the Association's help in creating generalized multi-parametric "portraits" of social groups, regions, enterprise groups. Such "portraits" can be used for prognostication of voting results, progress trends, and different consequences of decision making as well.

Association provides a useful guide in technical diagnostics, ecology, geology, and genetics.

ADUIS has at hand a broad range of high-efficiency original methods and program tools for solving analytical problems, such as knowledge discovery, classification, diagnostics, prognostication.

ADUIS unites the creative potential of highly skilled scientists and engineers

Since 1992 *ADUIS* holds regular conferences and workshops with wide participation of specialists in AI and users of intelligent systems. The proceedings of the conferences and workshops are published in scientific journals. *ADUIS* cooperates through its foreign members with organizations that work on AI problems in Russia, Byelarus, Moldova, Georgia, Bulgaria, Czechia, Germany, Great Britain, Hungary, Poland, etc. *ADUIS* is the collective member of the European Coordinating Committee for Artificial Intelligence (ECCAI).

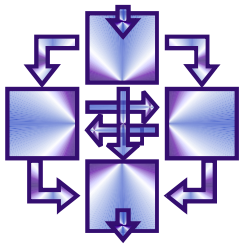
Products developed by ADUIS: **Confor**: Tools for Knowledge Discovery, Classification, Diagnostics and Prediction; **Analogy**: Tools for Solving Problems on the Basis of Analogy; **Manager**: Tools for Decision Support Systems Design; **Discret**: Tool for Discretization of Numerical Data; **Gobsec**: The System for Investment Scheduling.

For contacts: V.M.Glushkov Institute of Cybernetics; National Academy of Science of Ukraine;

Prospect Akademika Glushkova, 40, 03680 GSP Kiev-187, Ukraine;

Phone: (380+44) 5262260; Fax: (380+44) 5263348; E-mail: glad@aduis.kiev.ua

10th Anniversary of



Association for the Development of the Information Society

Acad. G. Bonchev St., block 8, Sofia 1113, Bulgaria
Tel. (+359-2) 979-3813, -3808, Fax (+359-2) 739-808
e-mail: ario@math.bas.bg, adis@einet.bg
http://www.adis.org

The Association for the Development of the Information Society (ADIS) was established in April 1997 and is an independent, non-government, non-profit organization with the non-commercial objective to support the development of the information society in Bulgaria. This objective is extensively defined in the Association's statute and includes:

- Interaction with individuals and organizations working for the development of the information society in Bulgaria and in the world.
- Support of the comprehensive utilization of the capacity of the information infrastructure and information technologies by all layers of society and all ages and professions, as well as by unemployed, ethnic minorities, people with disabilities, etc.
- Development and implementation of national and international projects whose goal is establishing, developing, and governing the information society.
- Participation in the elaboration and implementation of educational, promotional, and demonstration programs dedicated to information society issues.
- Participation in international activities on issues of the development of the information society, and maintenance of ties to and interaction with foreign and international organizations.
- Organization of conferences, forums, workshops dedicated to the information society.
- Publishing of a newsletter distributed among the individual and collective members of the Association.

Besides individual persons, the Association has as collective members from various regions of Bulgaria: Plovdiv University 'Paisii Hilendarski'. Technical University—Gabrovo, the Police Academy, the Institute of Mathematics and Informatics, the Institute of Information Technologies, the Central Laboratory of Computer Security of the Bulgarian Academy of Sciences (Sofia), and other organizations. Societies in the cities of Plovdiv, Shoumen, and Bourgas have been formed as autonomous subsidiaries of the Association. Its membership and associated structures are growing quickly and already include foreign members. The Association has existed since recently but it unites people and organizations with several decades of experience in the field of computer science and information technologies. Since 1999, the Association has organized monthly national seminars in the framework of the Forum Global Information Society. The seminars are devoted to the development of the information society in all fields of the human activities and aspects. Other activities include implementing a project for training disabled (deaf) people to use computers and the Internet, a project for training secondary school teachers in a broad range of computer technologies, participation in the drafting of the Bulgarian national strategy for the Information Society, drafting of models and principals for creating, management and development of public centers for access to Internet, information and communication services and public e-information and e-services for the Bulgarian citizens as well as delivering of talks on Information Society issues at various national and regional events by members of the Association.

The Association gladly welcomes contacts with organizations from abroad whose activities are related to the development of the global information society.

TABLES OF CONTENTS OF IJ ITK – VOLUME 14

Volume 1, Number 1

Volume 1, Number 2

Volume 1, Number 3

Volume 1, Number 4

AUTHORS' INDEX

TABLE OF CONTENTS OF VOLUME 14, NUMBER 4