Editors of Series

Danail Dochev Istvan Simonics
Radoslav Pavlov

Methods and tools for development of semantic enabled systems and services for multimedia content, interoperability and reusability

HUBUSKA Third Open Workshop
Klagenfurt, Austria, 27-28 April 2006

Proceedings

HUBUSKA
Networking Centres of High Quality Research on Knowledge Technologies and Applications

MTA SZTAKI
IMI BAS, IIT BAS
NOVITECH
UNIVERSITY KLAGENFURT

László Böszörményi István Simonics
Editors

Methods and tools for development of semantic-enabled systems and services for multimedia content, interoperability and reusability

ALPEN-ADRIA
UNIVERSITÄT
KLAGENFURT

Open Workshop

Klagenfurt University, Austria
27-28 April, 2006

Proceedings

The workshop presents results of project supported by EU FP6 programme
Specific Support Actions

INCO-CT-2003-003401 HUBUSKA
Networking Centres of High Quality Research on Knowledge Technologies and Applications

Workshop programme and organizing committee

László Böszörményi (*Chairperson and Editor*), Department of Information Technology (ITEC), University Klagenfurt, Austria

Danail Dochev (*Co-Chairperson*), Institute of Information Technologies of the Bulgarian Academy of Sciences – IIT BAS, Bulgaria

István Simonics (*Co-Chairperson and Editor*), MTA SZTAKI eLearning Department, Hungary

Radoslav Pavlov (*Co-Chairperson*), Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences – IMI BAS, Bulgaria

Tatiana Urbanová (*Co-Chairperson*), Novitech Partner Ltd. TeleDom, Slovakia

ISBN 3-9500593-7-7

PREFACE

From April 27-28, 2006 we organized our Third Open Workshop of the HUBUSKA project at the Klagenfurt University, Austria. The workshop was devoted to the subject "Methods and tools for development of semantic-enabled systems and services for multimedia content, interoperability and reusability".

Altogether 13 papers were presented, including a highly interesting invited talk, given by Prof. Johann Günther, and two demonstrations given by members of the Klagenfurt University. The workshop was attended by ca. 20 persons. In accordance with its *open* character, the workshop succeeded to attract several external visitors, in particular from the Technical University of Kosice and from the hosting Klagenfurt University, including excellent talks on running e-learning projects in these institutions.

The Workshop was subdivided into the following four sessions: (1) "Multimedia Content Distribution", (2) "Multimedia Processing and Presentation", (3) "The Development and Implementation of e-Learning/e-Teaching Strategies" and (4) "E-learning Experiences and Security Issues". The demonstrations were part of session 2 (pp. 56-62); session 3 was devoted to the invited talk.

The workshop has been highly useful and efficient. The talks covered a broad topic area, from formal-mathematical optimization issues, over questions of security up to experiences gained by real applications. The workshop offered the participants an excellent opportunity to come into discussion and to exchange ideas for scientific cooperation and for further projects. The organizers thank to all participants for their valuable contributions.

Klagenfurt, May 26, 2006

László Böszörményi and István Simonics
Editors

IN REMEMBRANCE OF DR. OTTÓ HUTTER
1961 – 2006

Ottó Hutter graduated at Budapest Technical University as an Electrical Engineer in 1984. He received Telecommunication Engineering degree in 1985 after postgraduate studies which were available for the best students. He received university doctoral degree in 1989. He has been worked at the Computer and Automation Research Institute of the Hungarian Academy of Sciences – MTA SZTAKI –, continuously since finishing the university. He started his career as a software and hardware developer at the Electronic Department. He became software project leader at the Informatics Department after the reorganisation of Institute in 1990. He was appointed to the Head of Open Learning Center established according his ideas. The eLearning Department was established under Ottó's leadership in 2002 where he made and led relevant technological and methodological developments.

He fulfilled versatile professional and social roles as well beside the activities in the institute. He fostered the domestic dissemination of the Unix culture through the book entitled Unix basics published by OpenInfo which publisher was established by him. He was the founder and the chairman of the national eLearning Committee. He was member of the Digital Curriculum Accreditation Committee of the Ministry of Education. He initiated and co-organised the conference series of eLearning Forums. He was one of the authors and editors of the eLearning Handbook published by the Műszaki Könyvkiadó in 2005. He was author and editor of several national and international publications on the topic of national eLearning standardisation and training material portability. He conducted multiple activities for the dissemination of the national IT culture. Among other activities, he was the editor of CIO column of the Computerworld magazine, the initiator and the chairman of the Programme Committee of the conference series Figyelő-Infopen CIO Forum, and one of the founders and member of the Organisation Committee of the "CIO award of the Year". He was the member of the board of the Hungarian Association of IT Companies from 1996 to 2005.
He received the MTA SZTAKI Institute Award in 1999. He received Dennis Gabor Memorial Award in 2005 for his activities in eLearning at industry level.
He actively participated in several EU projects as a project manager. For example, the aim of the "Adonis" Leonardo project was to develop eLearning training materials and distributed distance learning environment in the field of IT and eCommerce for SMEs. Within the framework of the "KNOSOS" Leonardo project a network of virtual training and resource centres of expertise – New Media Knowledge Village – was created for development of competence in integrated web and digital TV Anytime-Anywhere technologies. The objective of our HUBUSKA FP6 Specific Support Actions project is to support co-operation between research institutes that deal with the development of knowledge technology and eLearning applications, in order to facilitate the distribution of development results and their widespread utilisation.

Ambition for renewal, integration skills, team building and forming the future characterised him both in his professional and social relationships. He built relationship in a very broad extent and his work left enduring traces everywhere. He was continuously following the developments in the electronic education in a wide sense and looking for the areas where substantial contributions were possible. He offered opportunities for young people within the framework of cooperative education.

He was characterised by the eternal optimism, the happiness, and the continuous belief that radiated to his environment as well. This did not change when the illness attacked him one year ago. He gave us hope till the last moment by his behaviour and encouraging words that the illness can not be final and unchangeable. He could tell joke and encourage others even in the hospital bed. It is hard to believe that the sickness was stronger than him. He left us on 21st February 2006. However, he remains an empathic, kind, hopeful smiling friend in our memories as we always have been seeing him.

Table of Contents

Session 1 – Multimedia Content Distribution

AGENT ARCHITECTURE FOR STREAMING MULTIMEDIA CONTENT

Angel Veselinov Velikov
PhD student at IIT-BAS
Angel.Velikov@gmail.com

Abstract

The content providing on the networks has different purposes and uses different tools. With the large development of the networking technologies, now the users have much greater bandwidth at their disposal. Internet services are taking advantage of such great improvements which allow them to provide more information to more customers. The media, such as audio and video, becomes more widely accessible day by day. We are witnessing the first music albums and movies, offered to the public mainly through Internet. In this paper we will propose an architecture, which relies on the agent oriented approach in order to solve the tasks in a world with many content providers and many clients. The architecture is supposed to be on top of the existing solutions, adding new capabilities by monitoring and controlling them. We will describe the different roles which the agents can play and how they fulfil their goals.

1. INTRODUCTION

Networks history

The computer networks origin from the need of the computers to communicate in order to fulfil their tasks. The first experiments with teletype were conducted in 1940. The next milestone in the networks history is the establishing of ARPANet in 1969, which connected several universities in USA. The birth of the current Internet is considered to be in 1983 with the first operational TCP/IP wide area network. The step that moved the network closer to the public was the invention of the HTTP in 1991. Since then the networks and the technologies needed to connect and communicate through and between them continue to drive computer hardware, software and peripherals industries. The expansion is mirrored by growth in the numbers and types of users of networks from researchers and businesses to families and individuals in everyday use. [1] Now more than 1 billion people are using Internet. [2]

Network relationship architectures: client-server, p2p

There are two main approaches in the functional relationship in the networks.
The first approach is that one of the computers is acting passively, as a slave which executes the orders of the master, which is active. Such relationship is called client-server and is widely used for different tasks most of which are some type of content providing. This structure is centralised and the centre is the server computer which can accept connections from many clients on their request. The client-server architecture is useful when we want to share an available processing or storage resources among users which cannot afford to have them individually; when we want to centralise some relationships; when we want to provide some service to the public; and other tasks.
Technically every connection established on the network has initiator which is called client and the counterparty is called server. But when we have computers acting both roles, we are talking about the peer-to-peer or p2p architecture. This architecture is dynamic and the number of the participants in it can change in time. It relies on the computing power, storage and bandwidth of all computers. Nevertheless, most of the solutions use client-server structure for some tasks like searching, and peer-to-peer for the rest. The p2p structure has the advantage that when a new node is joined, it increases not only the demand, but the available resources too. It has distributed nature, which is more robust on failures and usually the data is replicated on many nodes. Another advantage is that the communications are distributed which by the use of different algorithms provides bigger bandwidth for the participants.
The agent architectures are hybrid as structure because they can use both structures. They use centralised client-server structure for searching, but for instance the servers can be connected in p2p manner. The agents can act as clients, searching for a service, as servers providing a service, as mediators relaying a service from

the server to the client, or even dynamically change their behaviour in order to fulfil their goals and optimise their performance.
Their main attributes are the proactiveness, mobility and semantics, which makes them different from the other structures and gives them ability to solve a bigger set of tasks and react more adequately.

Protocols

The layered architecture of the current networks enables them to separate the means from the goals, thus using different physical connections for one task, one physical connection for different purposes, or both. For the different purposes are developed different communication protocols. The protocols are generally sets of standard rules for data representation, signaling, authentication, and error detection required to send information over the network.
One of the most popular uses of the networks nowadays is for providing data from one computer to another. The data can represent different types of content, usually stored as files. For providing and transferring of files we can use the File Transfer Protocol (FTP). Although one of the oldest protocols, it is still widely used.
Another possibility is that the content is generated on user request. Such content is dynamic and usually depends on some user provided parameters. When we browse the world wide web we use the Hyper-Text Transfer Protocol (HTTP), which can provide both static or dynamic (generated) content.
The audio and video content can have special issues, compared to other. When you turn on your radio, you don't hear the programme from the beginning, but from its current point of time. It can play virtually endlessly which means that we cannot estimate its size. It has a special property – bandwidth – which defines the amount of data transferred per second. Usually, due to technical issues, not all of the data is received and the main goal is to keep the bandwidth stable. This data transfer is the multimedia streaming. The Real Time Streaming Protocol (RTSP), Real-time Transport Protocol (RTP) and the Real Time Control Protocol (RTCP) were specifically designed to stream media over networks. Other protocols used for this task are HTTP, Microsoft Media Services (MMS) and Real Data Transport (RDT).

FTP (RFC 959)

- File sharing
- Binary content
- Transfer data reliably and efficiently
- May resume

RTP (RFC 1889)

- Payload-type identification
- Sequence numbering
- Time stamping
- Delivery monitoring

RTSP (RFC 2326)

- Audio/Video
- Play/Pause/Record
- Playback a range
- Multicast

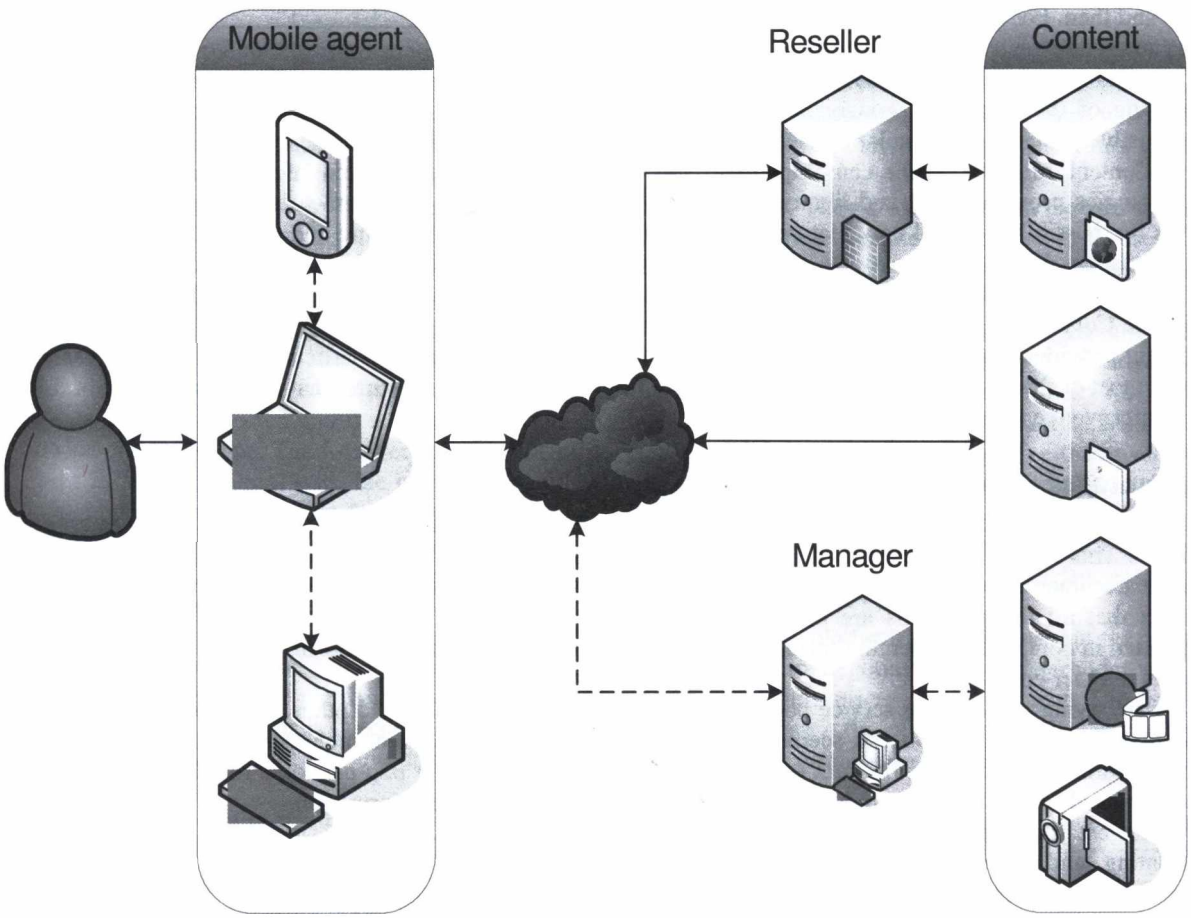
Streaming media - state of the art

There are many existing streaming media solutions. Part of them are proprietary and the rest are open. The most popular are Macromedia Flash, Microsoft Windows Media, Apple Quicktime, Java FORlive, Ogg/Vorbis, Shoutcast, RealNetworks, Icecast and Matroska. They use HTTP and RTSP protocols, as well as other standard and ad hoc protocols for transferring of the data.
The services they provide are various. The basic are the recording, encoding, decoding and converting of multimedia in different formats, which covers the needs of the "video on demand" service. Some are specialised in real time mixing of different content. The webcasting service is used mainly for conference calls and e-Learning. The unicasting is supported by all solutions while multicasting is supported only by some of them. The multicasting is needed for the implementation of a broadcasting service, such as Internet radio and TV. The digital rights management is a feature that is implemented mainly in the commercial software. There are other technical features which are specialised for concrete special purposes. The different user interface is quite similar in all solutions, since all new features are usually copied by the other parties. The p2p streaming solutions usually have the advantage of having searching capabilities. Some of the streaming servers can be set to use several sources, but this is done manually by the administrator – with scripts or through GUI. Other special features some solutions provide are destreaming or dynamic switching between several streams with different time frame to solve the lagging problems.

2. AGENT ARCHITECTURE

Overlook

The proposed architecture is dynamic and in minimum consists of User agent and Content provider agent working in client-server manner. This structure can be expanded in several directions: by adding Reseller agent which can mediate, provide user side fault tolerance, transform content or provide additional value-adding services; or by adding Manager agent and Hosting agent for server side management activities with the goal to improve the quality of service.



Basic use case

The general content providing structure is client-server where we have a user which requests some audio or video and a multimedia source which sends the data. The proposed architecture is not dealing with the process of transferring data but with the finding the requested data, establishing the connection and monitoring it.

The agent oriented approach comes with the solution for searching of particular services. The 'yellow pages' service is basic for agent platforms and allows the agents to advertise their services and search the advertisements.

In this structure our User agent is searching for the service and the Content provider agent is advertising his product. The service enables the agents to define complicated searches including different parameters, metadata and conditions. The agent architecture uses ontologies for the communications. The ontologies contain the semantic information needed by the agents in order understand each other. The searching service is directly affected by the used ontology, as it provides more flexible searching options.

After the search is finished, the User agent can provide the list with the results in a GUI. Here may be applied several interface options for narrowing the search by direct selection from the results, searching again for results similar to the selected or adding conditions. The process is finalised with selecting of particular result.

User agent

- Search for content
- Present content

Content provider agent

- Advertise content
- Provide content
- Predict/plan request and cache (opt.)

The next task in the process is the negotiation of the agents. The agent architecture includes several negotiation protocols which can be used for different cases, and new protocols are being developed every now and then. For the basic case we are describing now we can use the contract net protocol.

When finishing negotiations successfully the agents should exchange all the necessary information for initiating of the contracted connection: streaming protocol, machine addresses, etc. The agents then are responsible to start the particular services which execute the real transfer of data.

The Content provider agent can implement the actual streaming too. Some of the currently popular technologies provide SDKs and libraries thus enabling such solution. A custom solution for the streaming gives the possibility to add new features. Such useful features can be buffering and caching of the streams. This can be done to improve the quality of service, if possible, when the retrieval from the storage experiences sudden slow-downs because of heavy load or bad media. Such service should be loosely coupled with the Content provider agent. The caching feature can be used through a simple "prepare" command. The Content provider issues such order to the service when needs it to try to put particular content in the cache. Since most of the solutions have good user-side caching, this feature is not important and it is not certain if it will improve the results.

Advantages of using agents as mediators

In the basic case, we assume that the User agent and the Content provider agent are using the same searching service. This means that they are on the same multi-agent platform. The platform can be spread on several server machines. It can also be connected to other platforms, sharing services with them. Anyway, we cannot expect that all the platforms are connected with each other. If the existing platforms in the world are separated in components, then the service we advertise in our component of platforms will be unknown to the agents in other components.

If we are trying to make our service popular and widely used the expansion to other platform components is one of the solutions. Although it seems that this problem is due to the agent architecture, actually it is not. The centralised searching is the root of this problem so it is common for p2p networks too. Internet itself is decentralised so practically there is no search engine that includes all the information.

We suggest that the architecture should include a mediating agent. We will call it Reseller agent because of its specific representative tasks. The Reseller agent can be owned by different parties. It can be deployed by the providers to increase their market. It can be used by customer to increase the list of the available offer, thus to increase the possibility to find a better one. The third option is to be specially developed by third party to make profit from the mediation.

Reseller agent (opt.)

- Search for content
- Advertise content
- Relay content (pure salesman)
- Choose the best price provider (consultant) (opt.)
- Provide restricted content (exclusive representative) (opt.)
- Choose between pricing schemes – per item/volume/subscription (opt.)
- De-stream (value-adding reseller) (opt.)

The Reseller agent should have the search related functionalities of both the User agent and Content provider agent. It should be able to search for content and advertise it afterwards. In the simplest case there are no changes in what is found from the search and what is advertised. When mediating between

unconnected platforms, the Reseller agent functionality may be implemented by more than one agent – each situated in different platform and communicating with the others through special channels. We recommend that they use subscription protocol for their initial setup.
When the agent is deployed by the user or the provider, it should have some restrictions in the advertising or searching capabilities in order to offer them only to their owners' agents. This is because we don't want them to serve the concurrence. Their task is just to relay the messages between the user and the provider until the stream is started.
When the agent is deployed by a third party, its task can be quite different. The third party may be the platform owner, trying to make the platform more attractive by offering more services in it. Thus the goal will be to make more contracts as a number. Another goal which we assume will be more widely put is to make bigger profit.
We suggest that the architecture should be concentrated on such economical principles. The free market is a self-adjusting system. The economic relations are subject in the multi-agent researches, as well as the social relations, swarm intelligence, etc. One of the goals for the future experiments on this architecture will be to test whether there are and what are the advantages and disadvantages of having agents with rules which simulate economic behaviour.
We can use several possible approaches to make the reseller profitable. They are copied from the real world. Although not all aspects of the real world situations are planned in this document, our idea is to use this architecture as base for many other future experiments, concentrated on topics like security, trust, negotiating, etc.
One way of making profit is by having a reseller contract. Such contract grants the Reseller agent a reward for every client found for the provider. Even if the Reseller agent and the Content provider agent advertise the same service on the same price, such scheme can be profitable to both, since the Reseller has some chance of being selected and the Content provider increases his chance of getting the contract by listing its service one more time. The Reseller agent can offer prices lower than the original, using part of the reward to cover it, so the contract with the provider should be restrictive to such actions when the Reseller and the Content provider advertise on the same platform.
It is possible that the Reseller agent can have access to other platforms, while the User agent hasn't. In such cases profit can be earned by advertising the best offers from one platform on the other. Thus, by searching the other platforms instead of the users, the agent acts as a consultant.
Also the other platforms may offer content that is not available on this platform. This means that the agent will offer unique product. Another way to get in such position is to negotiate with the provider to be exclusive reseller for the platform. This allows the agent to put premium prices on this content. The prices can be adjusted more easily, because of the simplified economical case – the offering and the demand are both known by the agent and the uniqueness of the content makes it possible to make auctions instead of direct contracting.
A well semantically enabled agent will be able to cope with devising more complicated profiting schemes. For instance, if it uses statistic prediction the agent can take calculated risk and use special pricing contracts with the providers that give better prices but ask for minimum turnover. Planning also helps for better pricing of the clients and increasing of the quality of service.
The Reseller agent can profit by being more involved in the streaming process. It will get more connected to particular technologies and provide additional services to its clients. The additional services should attract the clients. Thus the agent will become a value-adding reseller. The profit of the agent will be based on the better service and its role will shift from the commerce to the servicing. Since we don't want to mix the agent role such situations should be approached with care. We suggest that the service part should be loosely coupled and separate from the agent.

Adding management roles

The server side can benefit from the agent technology too. The minimum solution with server agents that provide the content can be upgraded to more sophisticated structure of agents. The additional agents can help with organizing the mutual work of several servers by gathering information, analyzing it and taking decisions.
The first improvements that come in mind are these realized in the clusters and GRID computing. One agent can be responsible for many hardware systems, centralizing the service and the control. From the view point

of the agent architecture the controlling agent, that we will call Manager agent, is fully trusted by all the controlled agents. They accept and execute all commands that it gives them and provide it with all the information it requests. The Manager agent should have goals that optimize the sum of his assessment and those of his "workers". Again we propose that the assessment function should be the profit of each agent. These conditions implicate that the owner of the Content provider agents and their Manager agent should be one and the same. The owner of several Content provider agents may deploy additional Manager agent to optimize their work. It is also possible several owners to agree to combine their efforts and deploy a mutual Manager agent.

Manager agent (opt.)

- Dispatch requests
- Balance the system
- Keep fault tolerance
- Scale the system (opt.)
- Analyse the market (opt.)
- Adjust economic parameters (price, pricing, packages) (opt.)

Hosting agent (opt.)

- Control MAS platforms

The first task of the Manager agent we think of is to keep the system of servers fault tolerant. This task can be very simple. The agent has to check the availability of the servers and to keep the services (the content) doubled on different machines. There can be many rules applicable for different situations that make the choice of the doubling machine. We want it to be on different electric grid, in different network, in different building, etc. The rules can be applied in particular order until there is only one possible doubling machine available, the next rule returns empty set of machines or there are no more rules. There is another optimization decision to be taken. When the machine which is most appropriate for doubling of the content, but it is already on full capacity, the Manager agent should decide if there is content on this machine, which can be moved to another machine to free some space for the currently doubled content. This decision can be based on the doubling rules only or on additional information for the importance of the content. Also if a new machine is added or one of the machines is removed from the set the Manager agent should check what content needs to be doubled and maybe some content is available on more than two machines and some space can be freed.

The Manager agent should keep index of the content on each machine to fulfill these tasks effectively. The other choice is to query the Content provider agents, which in the first case is done less frequently – only to update the index. If the index solution is chosen, the subscription agent protocol can be used to have even less frequent queries or better to exchange them with information updates from the Content provider agents. We assume that the update increases the message payload insignificantly compared to the 'content update' event and even decreases the total payload of the 'content update' – 'query' – 'update information' message sequence, but is doesn't obsolete the query for active machines.

If the Manager agent can query the Content provider agents and receive information about the services provided – what content was streamed to the clients – this can be the base for the importance measure of the content that we mentioned above. The more requests they have for a particular content, the more important it should be. This information can be used by the Manager agent to spread the important content on more than two machines, thus making it more available. Such function will balance the system according to the customer needs and increase its overall performance.

Another function that is usual for the clusters and GRIDs is to have a dispatcher agent which stands between the clients and the servers much like the Reseller agent. The main differences are that it is owned by the same entity as the servers, the servers trust it, it doesn't make market based but load balancing decisions and can order the servers how to distribute the contents among them.

The more advanced Manager agent should be able to analyze the market. It is important to know what is demanded from the customers, what they search for and how often. It should know the other side too – what is offered by the concurrence, with what quality and prices. Based on such information can be made many different calculations known from the economic theory which will make the Manager agent even more helpful.

The Manager agent can have rules with which to detect when the system needs more machines. Such rules should be based on economical analysis of the demand. The agent can inform a human agent about such events or can be authorized to negotiate for new server machines. Again, the goal of such actions should be increasing the profit, but total for the controlled system.

The new machines can be negotiated automatically only if there is another agent on their side, which controls them. We will it Hosting agent and its main tasks will be to contract with the Manager agents the hiring of its machine and afterwards to allow access and execution only to the authorized agents.
There are other options than the rough scaling of the system. To control the demand and increase the profit, the Manager agent can make changes in the economic parameters. By adjusting the prices it can optimize the load on the system. If the system is working on full capacity, maybe it is time to increase the price. If the market share is small and there is free capacity, lowering prices can help. The prices for the different content can be changed differently based on their performance on the market. Here the Manager agent will also have the task to negotiate with the Reseller agents. It can use different approaches to improve the profit. To the Reseller agents can be offered special pricing schemes – not only per copy, but also per volume. There can be different package offers – for several contents, for subscription or others.

3. CONCLUSION

Most of the currently available solutions are concentrated on the technical implementation of the multimedia streaming task. They provide lots of features suitable for the concrete tasks.
The architecture proposed by us is not intended to implement the features that already exist, nor to make them obsolete by using generally new approach. The main intention of the article is to evolve the current technologies with the goal to make them more suitable for wide commercial use. Now there are offers from different providers for services which provide Internet, television channels and video on demand on one cable. They allow the customer to choose the provider dynamically. Such service introduces new issues for the current technology.
By placing the proposed architecture on top of the existing technologies, we try to satisfy the new demands. The architecture bonds the different technologies. It mediates between the different current standards. It can be used to organize their work, providing scalability, load balancing, fault tolerance or other features that are related to the management and monitoring of the services. One of the special tasks of the architecture is to help for the economical development of the service. It tries to make the service profitable on the provider side and cheaper on the customer side. The main responsibility for this is delegated to the Reseller agent and Manager agent which should give the architecture the advantages of the market economy.
Our future goals are to create a test implementation of the architecture. We intend to use JADE as multi-agent platform, RTSP as main protocol and different multimedia streaming technologies. The experiments we plan to conduct should show the effectiveness of service controlling, mediation, integration, monitoring and improvement of the economic parameters. The agents will use different AI approaches like behaviors, rules or inference.

REFERENCES

[1] http://www.wikipedia.org
[2] http://www.internetworldstats.com

ABSTRACTING AND CHARACTERIZING DISTRIBUTED VOD SERVERS

Péter Kárpáti
Klagenfurt University,
Universitätsstraße 65-67, 9020 Klagenfurt, Austria, E.U.
kpeter@itec.uni-klu.ac.at

Tibor Szkaliczki[1]
Computer and Automation Research Institute of the Hungarian Academy of Sciences
Lágymányosi u. 11. H-1111 Budapest Hungary
sztibor@sztaki.hu

László Böszörményi
Klagenfurt University,
Universitätsstraße 65-67, 9020 Klagenfurt, Austria, E.U.
laszlo@itec.uni-klu.ac.at

Abstract

Multimedia services are becoming widespread over the Internet. Large amount of resources are needed to deliver multimedia data with high quality and it is a quite difficult task to set up and scale video servers to comply with expectations. We developed a formal frame model to support the design of distributed video servers, the evaluation of their behaviour and the control of the server activities. It is a general model that can be customised to different architectures. The model enables giving estimations about the available throughput and defines utilization, balance and economic functions. State transition rules can be defined as well to specify dynamic adaptation steps for the server.

Keywords

modelling, planning, distributed video server, video migration and replication, host recommendation

1. INTRODUCTION

1.1. Video servers and the model

The market for video servers is expanding very dynamically nowadays: ever increasing numbers of video-on-demand service providers are appearing. There are many different video servers in use and new ones are being developed as well. It is not easy to set up and scale such a service according to given expectations and mostly it happens "step-by-step" in that the service infrastructure is modified according to the experiences gained with the system.

Imagine ourselves in the position of system designers who have to install a distributed VoD server . They have at their disposal nodes in different segments inside different LANs. They know the network structures of the LANs and have an idea about the expected clients inside and outside the concerned LANs. They have questions like: how much throughput can the system provide in the best case and in a meaningful worst case? Are there bottlenecks in the system? If so, where are they and how much would it take to get rid of them? How many different videos can be stored in the system? How many requests can be served simultaneously? What is the maximum and minimum number of clients the system could have? Assume that these questions have been answered and that the server is now up and running. The designer most likely wants to know whether the system is working fine, whether it is making a profit and to what extent the clients are satisfied with the service. In addition, she would like to periodically evaluate these characteristics and have a control mechanism that automatically helps to maintain an advantageous state of the system.

[1] Partial support of the Hungarian State Eötvös Scholarship, the Hungarian National Science Fund (Grant No. OTKA 42559 and 42706) and the Mobile Innovation Center, Hungary is gratefully acknowledged.

We have created a mathematic model to act as a tool for planning such a system and evaluating its behaviour. The model tries to be general enough to be applicable to different video server architectures from different manufacturers and for different scenarios. It is a frame model which needs to be completed by a core model and some parameters provided by its user, thus ensuring customizability.
The model consists of two parts: a static and a dynamic part. The static part models the given infrastructure used by the server (i.e. the available computers and local network structure) and analyses the accessible throughput of the system. Usually, the given infrastructure can be found in a LAN; the users can stay inside it or remain outside, separated by the Internet. Since we want to study the manageable service load on the system, we do not consider the location of the external clients and, consequently, the connections (via routers) between them and the subnets. In other words, the static system consists of separate LANs and the limits to throughputs are defined for connections to internal and external clients. We define characteristic functions and some questions (like above) which could be interesting for video-on-demand services. The estimations in the static part of the model help to set up and scale the system.
The dynamic part gives tools to evaluate the working system and to decide whether it operates according to expectations or not. The defined functions can be used in monitoring software which analyses the behaviour of the system and can trigger appropriate reactions to maintain the system in a good, profitable state. In this part, the utilization and load balance of the different system resources are examined and simple models are defined relating to video usage, quality of service, client satisfaction and the cost, income and profit of the system. The model needs knowledge of the expected behaviour of the clients to analyse the system and the user has to define some parameters or relations. To facilitate analysis, the functions of the model were implemented as Matlab functions to try out and compare many different configurations.
The core model defines the state transition rules for the dynamic part of the frame model and thus helps us by modelling a whole system. It is responsible for the host recommendation. Host recommendation defines new configuration of the system when a client request arrives, a new video appears or the system gets into an inappropriate state. The host recommender needs actual information about the terminal capabilities of the clients, the location of the current requests, the required quality of the video, the load of the hosting nodes and the QoS parameters of the network. We present simple and fast algorithms that are especially suitable for large-scale distributed video servers. A more detailed description of the whole model can be found in [1].

1.2. The Adaptive Distributed Multimedia Server (ADMS)

The mathematical model for distributed VoD servers was inspired by the ADMS [2] the modular structure of which makes it possible to automate offensive adaptation strategies [3].

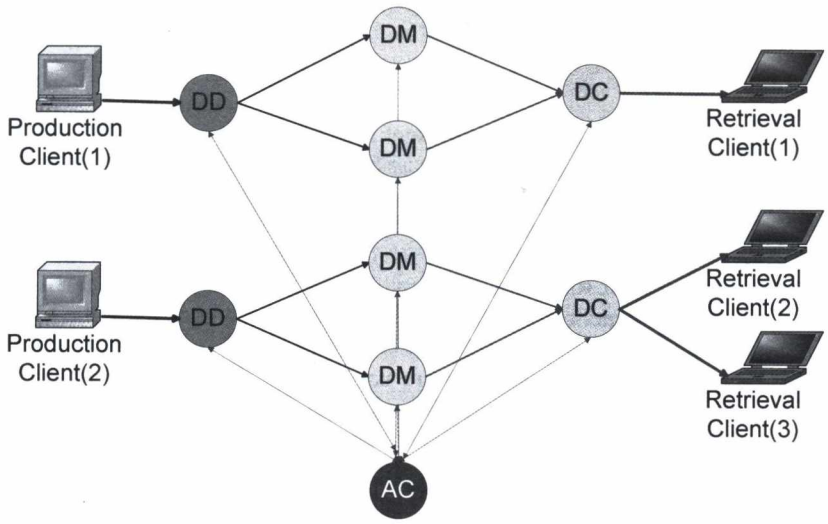


Figure 1 ADMS architecture

The ADMS (Fig. 1) has four components: the Data Distributor (DD) distributes the videos, received from a Production Client, onto Data Managers (DM) which store stripe units of the videos. The Data Collector (DC)

collects these units, assembles them and streams the requested video to a Retrieval Client. Finally the ADMS Controller (AC) organizes how the server works. When a component in the system is overloaded, the server is able to set up new components and also to migrate and replicate videos between DM groups, thus complying with the new requests. To achieve this, the ADMS relies on an appropriate middleware called Vagabond2 [2].
In the mathematical model, a video server consists of storage and streamer components which are supposed to be present in all types of VoD servers. The differences among the many types of video servers (serving stored videos) lies in the way how video storage and streaming are organized, and which kinds of other components are used to optimize the video delivery etc.
Our model can be very easily applied to the ADMS server. The DMs become storage components in the model and the DCs become streaming components. The DDs are not dealt with directly in the model; here the new videos appear in the dynamic system simply through a state transition. The AC is not part of the model; however the proposed control scheme in the dynamic model can be implemented in the AC and some of its aspects can be handled in the core model. (There are more examples in [1] to show how some other video servers fit into the model.)
The paper gives an overview of the work done so far on the related fields of network and resource modelling and host recommendation. The further chapters illustrate the usage of the static, dynamic and core model through an example.

2. RELATED WORKS

2.1. Network models

The internet layer of the networks is usually modelled in connection with the Internet (e.g. in the network emulator EmuNET [4]) because this is the focus of the network applications and this can be monitored using IP datagrams. In our frame model (in the static and the dynamic parts), we cannot really use these models because the Internet is very variable and we do not have any influence on it. A VoD provider can, at most, affect its own computers and the local connections between them and therefore we pay attention to these parts. On the other hand, the models of different LANs are too detailed for our intentions (like the one present in the network simulator NS [5]). We do not need to know which technology is used in the LAN segments or how the different types of segments could be connected, etc. We are only interested in the amount of transferable data which can cross a segment or a connection. In addition, we take advantage of the simpler structure of LANs in contrast with the router infrastructure of the Internet. We let the core model represent the router network.

2.2. Resource modelling

A layered model of a multimedia communication system (MCS) with respect to QoS is presented in [6]. We assume that the task of determining the video streaming abilities of the computers can be done with benchmark tests. One could use, for example, the benchmarks defined and examined in [7]. This paper addresses the problem of mapping the requirements of a known media service workload onto the corresponding system resource requirements and is based on two basic benchmarks introduced in [8]. Each of these benchmarks consists of a set of sub-benchmarks with media content encoded at a different bit rate. The difference between the two basic benchmarks is that in one, the clients access the same file while in the other, different videos are requested. When evaluating the tests, the number of concurrent streams and the delivered bandwidth (both, for the different encoding rates) are used for the fragments in which the stream quality could be maintained. Using the results, a cost function is derived in [8] which reflects the combined resource requirements necessary to support a particular media stream. In addition, two tools are provided: one called *MediaProf* which creates a traffic profile from access logs and the other one *Capacity Planner* which allows a service provider to specify the desirable system performance. To use this approach in our model, we have to have an idea about the characteristic of the expected client load.

2.3. Host recommendation

We introduce a common framework for problems related to host recommendation, such as server selection, host deployment and video replication. These problems are usually examined separately. There has been a significant amount of research recently regarding server selection [9]. Finding the optimal deployment of hosting nodes in a network is a well known problem in the literature. Most of the works, however, deal only with web-proxies [10, 11]. Replica placement is a recent research topic but only few papers consider the cost for delivering objects from the origin to replicas [12].

In earlier papers we dealt with the configuration recommendation algorithms for the offensive adaptation [13, 14]. We assumed that only the streamer nodes (proxies) can be moved because their replication or migration is much less time consuming than that of storage nodes or even video instances. However, replication of videos is also an inevitable task for video servers. For this reason, we extend our model to include these later problems as well. Furthermore the original model is elaborated in more details in order to better reflect to the real world problem.

3. THE STATIC SYSTEM

The static system is the simplest part of our model. It describes only the most relevant parts of a distributed VoD system and enables us to design good initial configurations.

Fig. 2 shows the North-West part of the Hungarian National Academic Backbone Network [15]. We build our examples on this network. It is important to note that although the topology and the maximum bandwidth capacities of the backbone network mirror the real situation but the other parts of the model are hypothetical including available bandwidth, directly connected subnets and client requests.

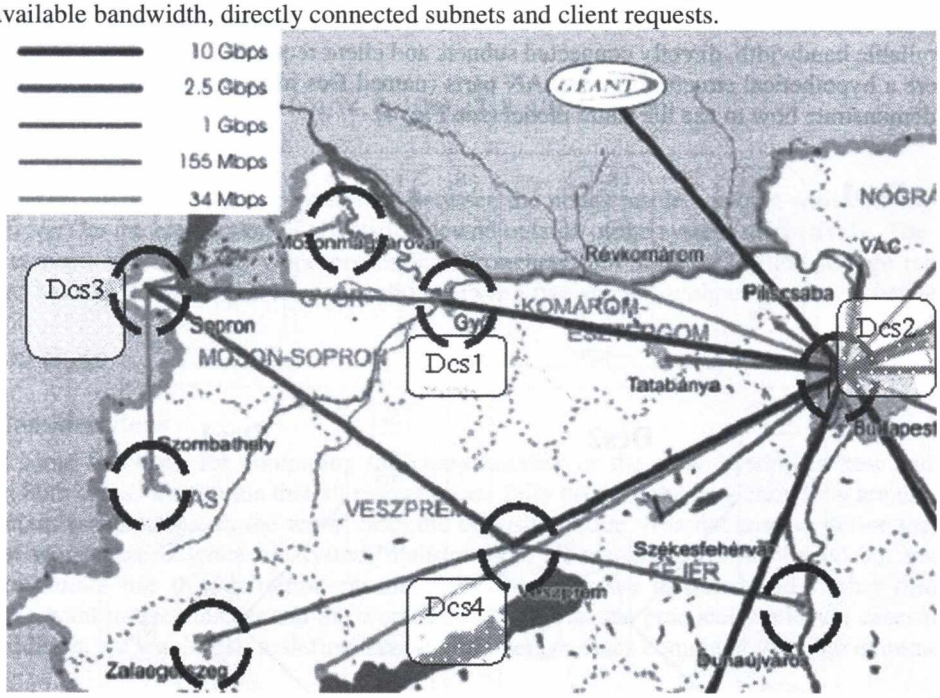


Figure 2 The topology of the sample network. Circles symbolise routers and the rounded rectangles LANs called Dcs in the model.

Fig 3. focuses on the topology of the network depicted in Fig. 2. We show a hypothetical structure of the LAN parts (named Dcs in the model) of our example depicted in Fig. 3 to demonstrate how to use the static model (see Fig. 4). The static system is described in full detail in [1].

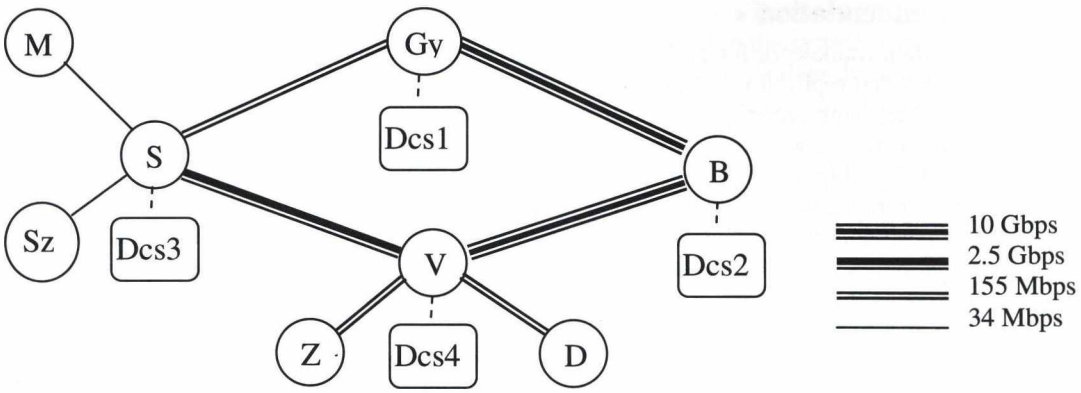


Figure 3 The topology of the sample network. Clients can be connected to the routers everywhere. LANs, available for the video server, are denoted by rounded rectangles which are called Dcs in the model.

3.1. Example (setup)

There is ongoing project to use ADMS on the Hungarian National Academic Backbone Network. Fig. 3 shows the North-West part of the network. The towns are denoted by their initials: B-Budapest, Gy-Győr, S-Sopron, V-Veszprém, M-Mosonmagyaróvár, Sz-Szombathely, Z-Zalaegerszeg, D-Dunaújváros. We build our examples on this network. It is important to note that although the topology and the maximum bandwidth capacities of the backbone network mirror the real situation but the other parts of the model are hypothetical including available bandwidth, directly connected subnets and client requests.

We show here a hypothetical structure of the LAN parts (named Dcs in the model) of our example depicted in Fig. 3 to demonstrate how to use the static model (see Fig. 4).

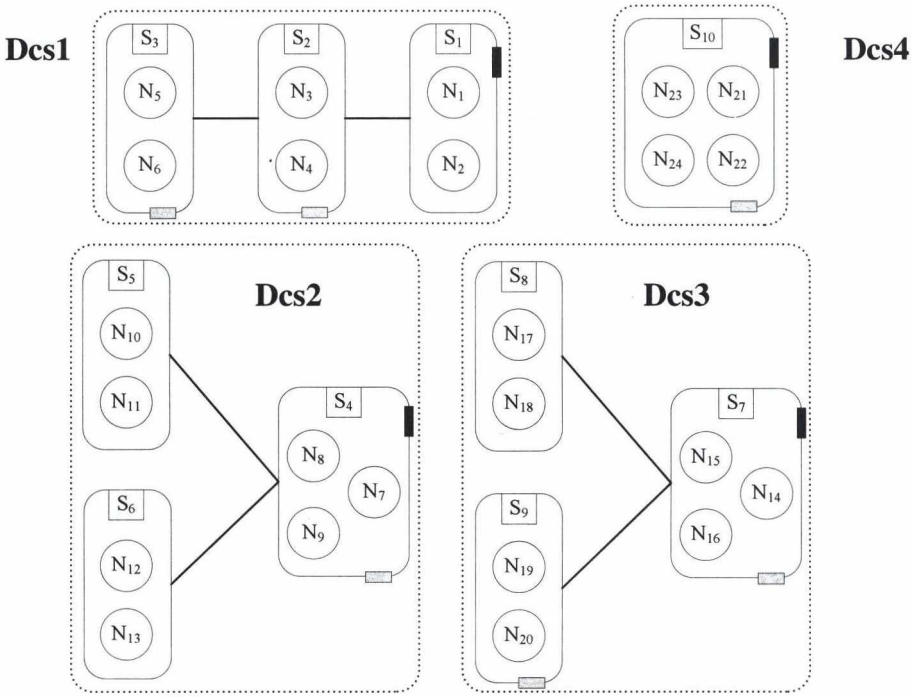


Figure 4 Example LANs (N: node; S: subnet; black rectangle: outgoing connection; grey rectangle: ingoing connection)

Subnet = *SId* × *Nodes* × *InsideNetThp* × *IngoingNetThp* × *OutgoingNetThp*

$Subnet_1$ = (1, {$Node_1$, $Node_2$}, 200, 0, 200)
$Subnet_2$ = (2, {$Node_3$, $Node_4$}, 100, 50, 0)
$Subnet_3$ = (3, {$Node_5$, $Node_6$}, 50, 50, 0)
$Subnet_4$ = (4, {$Node_7$, $Node_8$, $Node_9$}, 200, 200, 200)
$Subnet_5$ = (5, {$Node_{10}$, $Node_{11}$}, 100, 0, 0)
$Subnet_6$ = (6, {$Node_{12}$, $Node_{13}$}, 100, 0, 0)
$Subnet_7$ = (7, {$Node_{14}$, $Node_{15}$, $Node_{16}$}, 300, 150, 150)
$Subnet_8$ = (8, {$Node_{17}$, $Node_{18}$}, 100, 0, 0)
$Subnet_9$ = (9, {$Node_{19}$, $Node_{20}$}, 100, 100, 0)
$Subnet_{10}$ = (10, {$Node_{21}$, $Node_{22}$, $Node_{23}$, $Node_{24}$}, 100, 50, 50)

Node = *NId* × *Sto* × *VideoThp* × *StoRAccThp*

$Node_1$ = (1, 1000, 30, 30)
$Node_2$ = (2, 1000, 80, 80)
$Node_3$ = (3, 1000, 50, 50)
$Node_4$ = (4, 1000, 50, 50)
$Node_5$ = (5, 1000, 80, 80)
$Node_6$ = (6, 1000, 30, 30)
$Node_7$ = (7, 1000, 100, 50)
$Node_8$ = (8, 1000, 100, 50)
$Node_9$ = (9, 1000, 200, 100)
$Node_{10}$ = (10, 1000, 40, 40)
$Node_{11}$ = (11, 1000, 40, 40)
$Node_{12}$ = (12, 1000, 40, 40)
$Node_{13}$ = (13, 1000, 40, 40)
$Node_{14}$ = (14, 1000, 50, 50)
$Node_{15}$ = (15, 1000, 50, 50)
$Node_{16}$ = (16, 1000, 50, 50)
$Node_{17}$ = (17, 1000, 50, 50)
$Node_{18}$ = (18, 1000, 50, 50)
$Node_{19}$ = (19, 200, 30, 10)
$Node_{20}$ = (20, 200, 30, 10)
$Node_{21}$ = (21, 1000, 30, 30)
$Node_{22}$ = (22, 1000, 30, 30)
$Node_{23}$ = (23, 1000, 30, 30)
$Node_{24}$ = (24, 1000, 30, 30)

DirectConn = *SSId* × *ESId* × *NetThp*

(1, 2, 100), (2, 3, 50), (4, 5, 100), (4, 6, 100), (7, 8, 10), (7, 9, 50)

sat = 0.8 (saturation level of network)

The *InsideNetThp* characterizes the connections between the nodes inside a subnet while the *IngoingNetThp* and *OutgoingNetThp* the connection of clients inside and outside of the system respectively. The variable *Sto* defines the storage, *VideoThp* the video processing throughput and *StoRAccThp* the storage reading access throughput of a node. *DirectConn*-s are connection between two subnets without a router in between.

3.2. Functions

3.2.1. Considerations

We defined some functions for computing the characteristics of the static system in best and worst case scenarios. In both cases, we assume that all resources are fully used. In the best case, they are used optimally, based on certain assumptions; in the worst case, the opposite is true. It is not easy to define the worst case. The simplest worst case is when the system malfunctions: all resources are fully used but nothing useful happens. We cannot use this definition meaningfully therefore we distinguished further three scenarios (Table 3.). We want to use the best and the worst cases to restrict the practically relevant cases for using the system. In addition, we would like to define meaningful average cases computed from the extremes.

Type	Description	Model
0.	Resources fully used, nothing meaningful happens	Not in the model
1.	The worst resource affects the whole system.	In the topology independent case
2.	The worst resources affect independent parts of the system.	In the topology independent case

3.	The worst resources affect parts of the system depending on certain assumptions.	In the topology dependent case

Table 1 Worst case types

We defined topology independent (TI) and topology dependent (TD) functions in the model. The TI functions do not consider the (direct) connection structure of the LANs. They provide limit values for the TD functions.
We assume that after reaching a saturation level, a connection becomes useless for streaming because of too big delays or too frequent packet drops during data transport. We introduce ***sat***, the saturation level constant, which defines this level (given in %). Assigning a value to this constant is the task of the system administrator. In the case of the reading access throughput and the video throughput value, we assume that this was determined by a benchmark for the individual computer nodes (as suggested in [7]).

3.2.2. Topology independent case

TI functions are defined for best and for Type-1 and -2 worst cases relating to local resources (storage capacity, storage reading access capacity, video throughput capacity), to network throughput (inside network throughput capacity, direct connections network throughput capacity) and to streaming to client (streaming capacity to internal and to external clients and to both). (See [1] for more details.)
For the Type-2 worst cases the concept of ***directly connected subnets*** is introduced. Let *DCSs* be the set of the directly connected subnets (dcs's) where $dcs \in DCSs$ consists of subnets, connected directly to each other, without crossing a router (there is a connection in *DirectConns* between the two subnets). Practically speaking, a dcs corresponds to a LAN and a *subnet* in a dcs corresponds to a LAN segment.

3.2.3. Topology dependent case

In TD cases, the structure, as defined by the direct connections, is considered as well. Thus, traffic is modelled by the structure of the dcs and by the subnets which it has to cross.
We constructed a recursive function to compute the potential traffic appearing at the connection points to the clients if all nodes generate traffic at their full capacity. The algorithm is based on the choice of the bottleneck resource at the different (recursive) levels in the function. Using this function we can determine the appearing traffic in the best and in Type-3 worst cases for internal and/or external clients. (Additional details are presented in [1].)
Choosing the minimum values in the algorithm picks out bottlenecks. Thus, it is easy to implement a recommendation informing its user about the bottlenecks where the greatest improvement can be achieved. Another possibility is to give a recommendation for the whole dcs part in which all of the bottlenecks are given with suggested improvements. It can be done by repeatedly calling the function mentioned and by summarizing its recommendations.

3.2.4. Example (dissolving bottlenecks)

By applying the algorithm, used to compute the topology dependent cases, the bottlenecks can be detected and the system can be improved by choosing the important ones of them. The improvement recommendations are:

- *Exchange $Node_1$ with $Node_5$*
- *Enhance the InsideNetThp for $Subnet_4$*
- *Change the disks in $Node_{19}$ and in $Node_{20}$ for better StoRAccThp*
- *Change the direct connection between $Subnet_7$ and $Subnet_8$*

	Best case		Worst case	
	TI	TD	TI (Type-2)	TD (Type-3)
Original system	800	648	108	338

Improved system	960	930	130	450

Table 2 The estimated throughput values for internal and external clients

By dissolving some bottlenecks we could reach 43 % improvement for the best topology dependent case and 33 % for the worst, as one can see in Table 4. These are theoretical improvements as far as the experienced improvements are depending on the usage, but it is a quite good hint that these will be effective in the praxis as well.

3.3. Questions to the static model:

Based on the previous functions, we can answer relevant questions related to the planning of a distributed VoD server. We demonstrate the aim and the usage of the model through the example introduced above to give an impression for how it works.

3.3.1. How many different videos can be stored in the system at a given time?

If we assume that the average video size (*avs*) is 50 storage units and the average number of replicas (*copies*) is 2 then the answer is 240 for both the original and the improved systems, because the storage capacities have not been changed by the improvement.

3.3.2. How many requests can be simultaneously served by the system independently from their positions?

The answer for this question can be calculated by similar but more complex formulas as for the previous question (see in [1]). We assume to know the average throughput need of a video request which is 0.5 storage unit per second.

In the general approach, the *average* number of servable internal and external requests is 986 by the original system and 1380 by the improved system (40% more). These values are calculated directly from the results of the recursive function for the topology dependent case. This means that each dcs is used by the proper number of clients according to the computed throughput characteristics.

Let us assume that all dcs's get the same number of requests for videos and this number is the maximum number of requests servable by the weakest dcs. The *minimum* number of servable internal and external requests is 240 and the *maximum* number is 640 for both systems. The reason of the sameness is that we consider the weakest dcs which has the same throughput by both systems.

Let us assume that all dcs's get the same number of requests for videos with average throughput need and this number is the maximum number of requests servable by a hypothetical "average" dcs (computed from the average throughputs of the dcs's). There can be directly connected subnets which cannot serve all the requests and so some of the requests have to be rejected. The *minimum* number of servable internal and external requests is 534 for the original and 735 for the improved system (37 % more). The number of rejected requests is 142 in the first (21% of all requests) and 165 in the second case (18% of all requests). The *maximum* number of servable internal and external requests is 1128 for the original and 1555 for the improved system (37% more again). The number of rejected requests is 168 in the first (8% of all requests) and 308 in the second case (16% of all requests). One could tell that the improved system is able to serve 1145 requests simultaneously on average with less than 20% rejecting ratio by an assumed equal distribution of clients.

3.3.3. How many average clients could the system have?

We assume to know the length of the time period taken by the computations (*period* = 2000), the average throughput load of an average request from an average client (*aload* = 0.5) and the average duration of an average request from an average client (*adur* = 100).

In the worst case, all of clients access the system at the same time; this happens once in a period. Two cases can be distinguished here. In the first, all of the clients attack the same directly connected subnet. In the second case the clients attack each directly connected subnet at the same time. We use the second scenario based on the assumption that all directly connected subnets have their own clients.

In the best case, only as many clients come together as much throughput is available in the system (concerning all of the directly connected subnets) and after finishing, a new group of them appear, once there is place for them in the period.
The *best* average number of the clients is 13608 for the original and 19530 for the improved system (43% more). The *worst* average number of the clients is 7098 for the first and 9450 for the second system (30% more).

4. THE DYNAMIC SYSTEM

The dynamic system extends the static model by state transitions and utilization-, balance- and economic functions. The state space has been extended correspondingly. The dynamic system enables the evaluation how good the system works.

4.1. Functions

4.1.1. State transitions

There are outside and inside events connected to the system which can trigger complex actions within the system accompanied by state transitions. A state transition defines what changes should happen if one of the following inside events happens or one of the elementary actions has to take place.

Streaming begins:
A client requests a new video. The core model finds out which video instances are appropriate for the client. If the request can be fulfilled then streaming will be initiated which consists of two steps: some changes may be needed regarding the roles (a new streaming component setup), and after that the streaming can begin.

Streaming ends:
Streaming to the *i*-th client ends, so the occupied resources become free.

Video deployment begins:
As part of a complex action (initiating video deployment, expanding or reorganizing the system) deployments of video instances are started. This state transition presents the start of one of these video replications. The parameters are the source node, the destination node and the required throughput. (The path inside the dcs's is unambiguous; it does not matter which "gate" will be used to or from the routers; the core model cares about the path between the routers.)

Video deployment ends:
A replication ends; the occupied resources along the route have to be released.

New video appears:
The system administrator uploads a new video onto a set of nodes of the system. The optimal deployment of the video should be arranged by the core model. The execution costs of this action are currently neglected.

Delete a video instance:
A video instance will be deleted as part of a complex action (shrinking or reorganizing). Space at the storage node will be released.

Set up a new node:
A new node will be connected to the server with the given role. It can be the part of complex actions like expanding, reorganizing, video deployments or streaming.

Release a node:
The role of the node will be set to **uu** (during shrinking or reorganizing).

Extend the role of a node:
When a node already has one role (**sto** or **str**), it can be extended to **sas**.

Restrict the role of a node:
When a node has two roles (and is thus an **sas**), it can lose one of its roles.

The state transition function does not deal with the question, how a new state will be determined; this is the task of the core model. When the system is in an inappropriate state, three types of control actions can be taken. Shrinking means releasing the unused parts. Reorganizing means that the videos can be moved and deleted and the roles of the nodes can be changed. By expanding we let new resources join the system which

will take the load off heavily used parts in the future and videos are replicated there. Hints are information from outside which can give the system ideas on where to place videos.

4.1.2. Utilization functions

We defined functions to examine how much are the different resources (storage, storage access, video throughput; inside, ingoing and outgoing network throughput, direct connections) utilized. These functions are based on the functions defined in the static part. In addition, the system throughput utilization and minimum, average and maximum system utilization values are defined as well. (See [1] for details.)

4.1.3. Video utilization functions

Some functions regarding the different aspects of the video utilization are introduced as well. In order to do this, we need to extend the state space with time and data from the past which should be delivered by the core model. We assume that the function of the server is divided into periods. (See [1] for the state space extension.) In the video utilization functions, we examine the usage of the video instances, of the videos, of the video variations and of the streamed throughput. These functions are based on the computation, how many instances of a given variation of a given video could be streamed at most.

4.1.4. Balance functions

We defined balance functions to examine how much the load on the system is balanced using the resource utilization functions. In [1] is shown that the base utilization functions are similarly defined as the expected value in the probability theory and the base balance functions as the variance (both for discrete cases).

4.1.5. Client functions

The client functions are performance metrics connected to the ratio of the served requests and clients' satisfaction. We assume that the clients' satisfaction is proportional to the served quality of the video. Some information are missing to calculate the value of these functions; these should be obtained from the core model or from outside.

4.1.6. Economic functions

Three types of economic functions are defined in the model: cost, income and profit functions, all of them in connection with a given period. We assume that the nodes are rented and costs arise only when they are used (i.e. **sto**, **str** or **sas**). We have income only from the served requests. In this case a mixed model with pay-per-view clients and regular members is presented. The income from a request depends on the following factors: which video, when, (video popularity alters, discount periods at Christmas time, etc.), in which quality it was served and contract type and contract period of the member. The profit is simple defined as the difference between the income and the cost of the system in a given period.

4.2. Evaluating a state

Usually, we would like to answer the general question whether or not the system is in a good state. This is a fairly complex issue. First, the load from the clients, and thus the state of the system, can change rapidly and it can happen that a single state is quite good at one point in time and after a few seconds it is already obsolete (e.g. some streaming will be finished and their videos will not be requested again). In another case a less good state can become desirable when the proactive logic of the system made good predictions and decisions. In summary, the real state evaluation should consider the behaviour of the system over a period of time but as a basis, it needs to evaluate single states as well.

There are *system dependent* and *system independent* aspects of the system state. System independent aspects are utilization and balance because we do not need any additional information about the system, only that which is already in the model. In contrast, system dependent aspects are about income earned and the cost of the system as well as clients' satisfaction with the service; these should be defined by the designer of the system.

The *system works well* or is in a good state if:

1. the resources are exploited to a high degree but their exploitation does not badly affect the quality of the service provided,

2. the distribution of the load in the system is balanced,
3. it provides almost the maximum profit which is possible for it,
4. almost every client with a valid request can be served
5. and the clients' satisfaction level (regarding the served streams) is high.

The first two, system independent, aspects can be evaluated by single system states as well while the other, system dependent, issues are bound to a period of time. More details about the suggested evaluation method can be seen in [1].

5. THE CORE MODEL FOR HOST RECOMMENDATION

The task is to find optimal placements of nodes and video instances. We distinguish two main variations of the host recommendation problem: Light and Heavy Weight Recommendation.

In case of Light Weight Recommendation (LWR), the task is to find optimal places for streamer nodes, collecting stripe units of a video and streaming the collected video to the clients. This kind of recommendation has to be executed immediately after a client request has arrived. The aim of the Heavy Weight Recommendation (HWR) is to place optimally the instances of a video in the system. It is called heavy weight, because large amount of data is transmitted over the network. In this case the running time of the recommendation algorithm is much less critical. We use a common algorithmic framework for LWR and HWR. For the algorithms see [16]. We have selected some algorithms in order to adapt to our problem model and then to implement in the Host Recommendation module of the Adaptive Distributed Multimedia Server.

5.1. The simple algorithm

It assigns each client one after the other to the facility that can serve the best video instance to it and if there are more such facilities it selects streaming route with the best QoS parameters. This method is used for LWR.

Example

The underlying static system is the improved version of our system (Sections 3.1 and 3.2.4, Fig. 3). Let us assume that the network resources are free initially in the dynamic model except Subnets 1, 4 and 7 where the free inside network throughput is reduced to 100. Videos can be found only in one place, namely on the nodes of subnet S4 in Dcs_2. Three video instances are striped on the nodes of this subnet. The available video variations are as follows:

Video Variation ID	Horizontal Resolution	Vertical Resolution	Frame Rate	Color Depth	VSize	VThp
3	640	480	15	16	4	1
2	800	600	25	16	8	2
1	1024	768	30	16	16	4

Table 3 The video variations

Let the client requests come one after the other at Subnet 7 in Dcs_3 and these requests accept each variation of the video.

Let us consider the Simple algorithm. It always selects the streaming route that can serve a single request with the highest capacity. In our case there are four subnets with the same highest capacity in this case the algorithm may select any of them. Let us assume that the first request is served through streaming node N14 in Dcs_3. The free capacity of this node is reduced and the next request is served through streaming node N1 in Dcs_1. The third request is served from town Dcs_2. Each video stream reduces the bandwidth in Subnet 4 in Dcs_2 and Subnet 7 in Dcs_3. For this reason, the algorithm selects them fewer times than the hosts in Dcs_1. The

hosts in Dcs_4 have relatively small local resources but they can be also selected when the free capacities of the selected streaming routes become low.
Let us assume that 12 request arrived. The Simple algorithm would recommend using hosts as streaming nodes in directly connected subnets in the following order: Dcs_3/ Dcs_1/ Dcs_2/ Dcs_1/ Dcs_1/ Dcs_3/ Dcs_1/ Dcs_2/Dcs_1/ Dcs_1/ Dcs_4/ Dcs_1: twice in Dcs_3, twice in Dcs_2, seven times in Dcs_1 and once in Dcs_4.

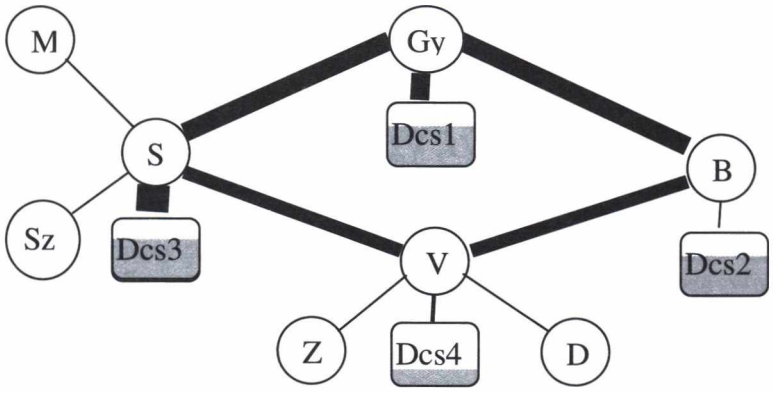


Figure 5 The network serving the clients according to the light weight host recommender. The levels of filling with grey colour represent the utilisations of the inside throughputs of the subnets containing the streaming node in the Dcs. The thickness of the lines represent the needed bandwidth to serve the requests on the backbone links. Routers are denoted by the initials of the towns.

We found video throughput as one of the most relevant attributes for streaming. Only those nodes need to be considered, where a streamer and/or storage node is located. The used video throughput utilization and balance of the above configuration are 0.2666 and 0.92.
Let us consider the case when each streamer node is hosted in Dcs_3. In this case, the used video throughput utilization and balance are 0.384 and 0.976. We can see that the utilisation is now higher because only two DCS's share the same load.

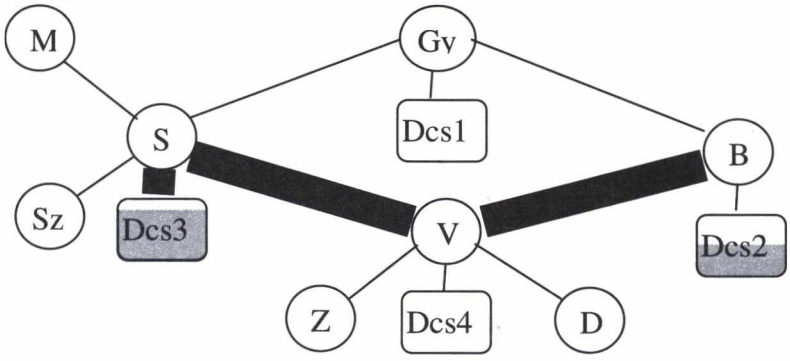


Figure 6 The network serving the clients with one streaming node in Dcs3.

5.2. Best gain per resource usage first

In each step, the algorithm always selects the possible video instance with highest value per unit resource consumption where the value of a video instance is the total value of the client requests servable from it. This algorithm is developed for HWR.

Example

The number of expected clients can be determined by a client behaviour prediction. Let us assume that the client requests in the same town have similar characteristics. Table 11 shows their parameters. The predicted values refer to one week.

Town	Number of expected requests	Appropriate video variation IDs	Value of a request (RValue)	Total value
B	2 000	V1, V2	0.1	200
Gy	100	V1, V2, V3	0.5	50
S	50	V2, V3	0.9	45
M	50	V2, V3	0.9	45
Sz	50	V2, V3	0.9	45
V	50	V1, V2, V3	0.5	25
D	50	V1, V2, V3	0.5	25
Z	50	V1, V2, V3	0.5	25

Table 4 The predicted client requests.

In our example, each client is servable from each hosting node. For this reason, the value of each possible video instance is the same. The algorithm selects hosting nodes in Subnet 1 in Dcs_1 because they had the highest feasibility factor (they have both high video throughput and storage reading access, see Section 3.1). If the value of a placement would incorporate the distance of the clients as well, this would lead to the selection of hosting nodes of Subnet 4 in Dcs_2 as well.

6. SUMMARY

We have presented a mathematical frame model to support design and evaluation of VoD servers. The frame model consists of a static and a dynamic part and needs to be extended with a core model. Automatic control schemes can be built on the top of the frame model to help reconfigure the system in accordance with the client needs and current network and host loads. The unique feature of the core model for host recommendation is that it offers a joint framework for different kind of problems such as server selection, host deployment and video replacement. The usability of the model was illustrated on examples.

REFERENCES

[1] P. Kárpáti, T. Szkaliczki, L. Böszörményi, *Mathematical model for distributed video-on-demand servers*, Technical Reports of the Institute of Information Technology, University Klagenfurt, TR/ITEC/05/2.13

[2] R. Tusch, Design and Implementation of an Adaptive Distributed Multimedia Streaming Server, PhD Thesis, University Klagenfurt, 2004

[3] R. Tusch, L. Böszörményi, B. Goldschmidt, H. Hellwagner, P. Schojer, *Offensive and Defensive Adaptation in Distributed Multimedia Systems*, Computer Science and Information Systems (ComSIS), Vol. 1, No. 1, Feb. 2004, pp. 49-77.

[4] A. Kayssi and A. El-Haj-Mahmoud. *EmuNET: A Real-Time Network Emulator,* In Proceedings of the 19th ACM Symposium on Applied Computing (SAC'04), 2004, pp. 357-362

[5] http://www.isi.edu/nsnam/ns/doc/node144.html

[6] R. Steinmetz: *Multimedia-Technologie,* Springer, 2000, pp. 238-239

[7] L. Cherkasova, W. Tang, S. Singhal, An SLA-Oriented Capacity Planning Tool for Streaming Media Services, Proc. of the International Conference on Dependable Systems and Networks,2004

[8] L. Cherkasova, L. Staley. Measuring the Capacity of a Streaming Media Server in a Utility Data Center Environment, Proceedings of ACM Multimedia, 2002

[9] M. Guo, M. H. Ammar, E. W. Zegura: *Selecting among replicated batching video-on-demand servers*, Network and Operating System Support for Digital Audio and Video, 12th International Workshop, NOSSDAV 2002, Proceedings, pp. 155-163

[10] M. Steen, P. Homburg,, A.S. Tannenbaum: *Globe: A wide-area distributed system.* IEEE Concurrency (1999)

[11] Qiu, L., Padmanabhan, V.N., Voelker, G.M.: On the placement of web server replicas. *INFOCOM.* (2001) 1587-1596

[12] B. Li, M. J. Golin, G. F. Italiano, X. Deng and K. Sohraby *On the optimal placement of web proxies in the internet*, IEEE INFOCOM 1999 - The Conference on Computer Communications, no. 1, March 1999, pp. 1282 – 1290

[13] B. Goldschmidt, T. Szkaliczki and L. Böszörményi: *Placement of Nodes in an Adaptive Distributed Multimedia Server*, In Proceedings of the 10th International Euro-Par Conference, 2004, pp. 776-783

[14] T. Szkaliczki, L. Böszörményi: *Incremental Placement of Nodes in a Large-Scale Adaptive Distributed Multimedia Server*, In Proceedings of International Conference on Distributed and Parallel Systems (DAPSYS 04), 2004

[15] http://www.niif.hu/index.en.php

[16] T. Szkaliczki, L. Böszörményi: Algorithmic Background of the Host Recommendation in the Adaptive Distributed Multimedia Server, In Proceedings of HUBUSKA Third Open Workshop, Klagenfurt, Austria, 2006

ALGORITHMIC BACKGROUND OF THE HOST RECOMMENDATION IN THE ADAPTIVE DISTRIBUTED MULTIMEDIA SERVER

Tibor Szkaliczki[1]
Computer and Automation Research Institute of the Hungarian Academy of Sciences
Lágymányosi u. 11. Budapest H-1111 Hungary
sztibor@sztaki.hu

Balázs Goldschmidt
Budapest University of Technology and Economics,
Magyar tudósok krt. 2, Budapest H-1117 Hungary
balage@inf.bme.hu

László Böszörményi
Klagenfurt University,
Universitätsstraße 65-67, 9020 Klagenfurt, Austria, E.U.
laszlo@itec.uni-klu.ac.at

Abstract

Host recommendation includes problems such as server selection, host deployment and video replication. It is especially relevant for the Adaptive Distributed Multimedia Server (ADMS) which is dynamically able to add and remove its components to different nodes of the network. Host recommendation is a quite complex optimisation problem. The present paper introduces the different variants of host recommendation and gives an overview of its possible mathematical approaches. We put emphasis on the facility location problem and the related approximation algorithms. At last, we present some algorithms selected for implementation.

Keywords

distributed video server, host recommendation, optimisation, facility location problem

1. INTRODUCTION

Host recommendation is a usual task for efficient operation of servers. It answers the question which node to select in a network for hosting a server application or which host to select for a special purpose. Such questions arise in different situations during the operation of a distributed multimedia server: From which server node should a client request be served? Where to put server components in the network? Where to store video instances? It is hard to find the correct answers because of several reasons. First, the operation of the server is influenced by several network and node parameters which are not easy to measure or to calculate accurately. Most of our decisions have effect on the future operation of the server but the real situation is not known exactly in advance. The multimedia server is a quite complex system and the problem model should also be capable to include features such as video replacement or caching. It is unclear in many cases which solution is the best for the host recommendation since the goodness of the system can be evaluated according to several conflicting aspects such as costs, user satisfaction, network load etc. Each video may be delivered with different parameters (resolution, frame rate, audio quality etc.) and it is not enough to select the locations of the hosts, but the video variation also have to be determined. The time-complexity of the algorithm is critical since the distributed video server may extend to large areas through the Internet and the number of the clients may be huge. In many cases, the algorithm should process large problems within strict time constraints.

[1] Partial support of the Hungarian State Eötvös Scholarship, the Hungarian National Science Fund (Grant No. OTKA 42559 and 42706) and the Mobile Innovation Center, Hungary is gratefully acknowledged.

This problem is especially relevant for the Adaptive Distributed Multimedia Server (ADMS) developed by the Klagenfurt University. ADMS is dynamically able to load and remove its components to different nodes of the network and this novel adaptive architecture enables migrating and replicating the server components in the network. Appropriate host recommendation algorithm with short running time is relevant for the distributed server to adapt quickly to the changing network parameters and client demands. We were motivated to deal with the host recommendation in order to improve the performance of the ADMS. Beside the dynamic placement of the server nodes, we studied other problems related to the host recommendation that are relevant for operating any distributed multimedia.
First, we give a brief overview of the ADMS server. Then we introduce the host recommendation problem in our focus and its variants including server selection, host deployment and video replication. Then we continue our paper by presenting the possible mathematical approaches. First of all, we discuss the facility location problem, its different variants and solution methods. Then some further mathematical problems are introduced as well that can be connected with host recommendation such as the network flow and the knapsack problem. At last, we give a brief summary of the host recommendation algorithms that we applied.

2. HOST RECOMMENDATION PROBLEM

2.1. The Adaptive Distributed Multimedia Server (ADMS)

The modular structure of the ADMS [1] makes it possible to automate offensive adaptation strategies [2]. Structure and description of the ADMS can be found in another paper [3] in these proceedings as well but we introduce it shortly here as well. The ADMS has four components: the Data Distributor (DD) distributes the videos, received from a Production Client, onto Data Managers (DM, also called as storage nodes) which store stripe units of the videos. Data Managers storing the stripe units of the same video form a Data Manager group and they are located practically on the same subnet. The Data Collector (DC, also called as proxies or streamer nodes) collects these units, assembles them and streams the requested video to a Retrieval Client. Finally the ADMS Controller (AC) organizes how the server works where the host recommendation runs as well. The nodes that are able to host server components are called as harbours. The server is able to set up new components or remove old ones and also to migrate and replicate videos between DM groups according to the dynamically changing client requests, QoS parameters of the network, and loads of the nodes. To achieve this, the ADMS relies on an appropriate middleware called Vagabond2 [1].

2.2. The optimisation problem

Host recommendation deals with selecting special nodes in the network for a specified purpose. This problem occurs in several different situations during the operation of the server: host recommendation can be used for finding optimal location for the server components in the network or for selecting already occupied host nodes for a given purpose, for example to serve a client request or to store a video.

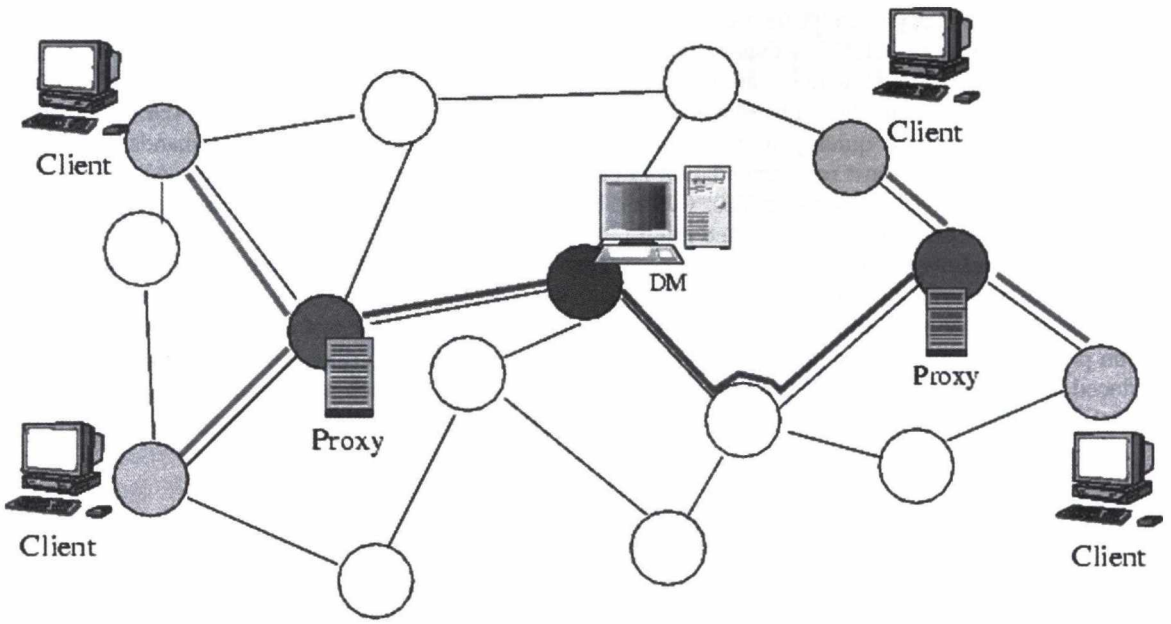


Figure 1 The goal is to find optimal location for multimedia proxies and data managers in a network [4].

Unfortunately, the real problem is extremely complicated. For a detailed list of the input parameters see [5]. The parameters may vary depending on the type of the host recommendation problem (LWR or HWR). We mention some of them in order to demonstrate the complexity of the problem:

- Terminal capability of each client.
- Current or possible locations of the server nodes.
- Locations of the existing or expected possible client requests.
- QoS parameters between the network nodes.
- Possible variations of the videos.
- Resource needs of variations.
- Profits from serving variations.
- Costs of the server nodes.
- Free local resources.
- etc.

The output of the model depends on the type of host recommendation. In case of LWR, the output should tell:

- for each client request, the streamer node serving it.
- for each client request, the storage nodes serving it.
- for each client request, which variation of the video to stream to the client

In case of HWR, it is not enough to determine the storage nodes where to put the video instance. The origin of the video instance is also needed for replicating or migrating the video instance.

The recommendation is done based on several contradictory optimisation criteria whose harmonisation is not a trivial task:

- Maximising the total value of the servable clients.
- Minimising the number of storage nodes.
- Maximising the quality of service between the clients and storage nodes.
- Maximising the quality of delivered video variations.
- Minimising the network load arising from replicating videos.

Multimedia services typically need huge resources. We always have to check whether the available network and the local resources are enough for streaming.

The distributed video server may cover large areas on the Internet and stream data can be delivered to large number of clients. The running time of the host recommendation algorithm becomes crucial for LWR in case

of ADMS since the delivery of the data-streams can start only after the placement of the Data Collector nodes. In this case, the time-consuming algorithms aiming at the solution with exact optimum are not applicable.

2.3. Variants of the problem

The host recommendation is needed in different situations such as server selection, host deployment and video replication.

- Server selection: It assigns server node(s) to the client to serve its request. One client request may be processed at the same time or more. The problem is online in the sense that we do not know the whole series of requests in advance but only the next one.
- Host deployment: It determines nodes in the network where to place individual server components.
- Video replication: It selects storage nodes where the video instances should be placed. The origin of the video instance is also relevant. Usually, it generates huge network load.

We distinguish two main variations of the host recommendation problem: Light and Heavy Weight Recommendation. In case of Light Weight Recommendation (LWR), the task is to find optimal places for data collectors or simply to select one. This kind of recommendation has to be executed immediately after a client request has arrived. The aim of the Heavy Weight Recommendation (HWR) is to place optimally the instances of a video on the DM groups in the system. It is called heavy weight, because large amount of data is transmitted over the network and it is much more time consuming task than replicating or migrating data collectors. In this case the running time of the recommendation algorithm is much less critical. We can distinguish further subcases within these two main categories. Table 1 summarises them. The *static* version of *LWR* is relevant for servers where the location of data collectors is fixed. This is the usual case at multimedia servers. The *dynamic* case refers to the ADMS, where data collectors can be replicated. We speak about *Upload* when a client loads a new video into the system. HWR *Replication* refers to the case when the video is replicated to new storage nodes because the popularity of the video increases.

Type	Client requests	Frequency of calling	Time constraints	Replication	Selection
LWR static	Current	Continuous	Strict	-	DC, DM group, video variation
LWR dynamic	Current	Continuous	Strict	DC	DC, DM group, video variation
HWR Upload	Predicted	Occasional	Moderate	Video (DMs)	DD, DM group, video variation
HWR Replication	Predicted	Periodical	Loose	Video (DMs)	DM group, video variation

Table 1. The main aspects of the host recommendations depending on the time value.

3. RELATED OPTIMISATION PROBLEMS

3.1. Facility Location Problem (FLP)

3.1.1. Overview

Suppose that fire stations are planed to place in a city. The city has a number of sites, which can be used for fire stations. The cost of placing and maintaining them is known at each site. The question is where the fire stations should be located in order to minimise the total establishing and maintaining cost and the average time to get to the potential sites of fire in the city. The facility location problem is looking for answer to such questions. This problem is well known in the operation research for decades. This problem is well known in the operation research for decades. Some early works in this topic include [6-8]. The facility location

problem has many different application areas such as placement of warehouses, telecommunication network design or configuration of distributed servers. To find detailed survey of the problem, see [9-10].
The facility location problem belongs to the optimisation problems. We shall focus on the uncapacitated facility location problem (UFLP), the best-studied variant of the problem. Let us consider its specification.
The input of the problem instance consists of the elements as follows:

- The set of facility candidates (F).
- The set of demand (or client) points (D).
- The fixed cost at each facility candidate i (f_i).
- The transportation cost from the facility candidate i to the demand point j (c_{ij}).

A feasible solution contains a set of facilities selected from the candidates and an assignment of each demand point to one selected facility. The objective of the optimisation is to minimise the total cost that can be calculated as the sum of the fixed costs of the selected facility candidates plus the sum of the transportation costs to each demand point from the facility to which it is assigned.
We remark that the facility location problem has many variants. We give a list presenting some different categories of the problem.

- *Deterministic vs. stochastic*. If some parameters are given by probability distribution, the problem is stochastic, otherwise deterministic.
- *Discrete vs. continuos*, according to the type of the set of the facility candidates and the demand points.
- *Static vs. dynamic*. The problem instance is changing in time in the dynamic case.
- *Capacitated vs. uncapacitated*. In the capacitated case there is an upper bound on the number of demand points that can be assigned to a facility.
- *Multi-capacitated case*. It is an extension of the capacitated case when the different types of the facilities have different capacities.
- *Multilevel vs. one level*. In case of the multilevel facility location problem, the facilities form a hierarchy of k level. In this case the demands are satisfied through the sequence of k different facilities, each of which belongs to different level. This model is useful, for example, when central depots are opened beside warehouses.
- *Metric case*. It is an important special case when the transportation costs are induced by a metric. In this case the costs are nonnegative and satisfy the triangle inequality. Since the demand between two demand points is not defined, the triangle inequality contains the sum of three edges instead of two.
- *Splittable vs. unsplittable demands*. In assigning the demand of customers to facilities, there are two natural variations to consider: the demand of a costumer must be met by a single store (unsplittable demands), and the demand of a customer may be divided across any number of facilities (splittable demands). These variants are distinguished only in the capacitated version, because in the uncapacitated variant each demand is satisfied always by the closest facility.
- *Fault tolerant version* In this case, each demand point has to be connected to a given number of facilities. This variation ensures that if a facility breaks down than its clients still can be served with low cost by the other facilities assigned to the client.
- *Multicriteria FLP* Several objectives may exist for evaluation of different facilities. For a multicriteria optimization problem, usually there are no "optimal" solutions as in the case of single criteria problems, but only preferred solutions are available. The preferred solution must be Pareto optimal (also non-dominated solution, that is, improving any of the objectives is possible only with degrading others).

The present paper focuses on the simplest variation of the problem. However, the above variations may be relevant when we apply the facility location problem to the host recommendation.
We still have to mention here the *k-median problem* as well. This problem differs from the facility location problem that there are no costs for opening facilities, instead a number k is specified, which is an upper bound on the number of facilities that can be opened. Many cases, similar methods can be applied to it as to the FLP.

3.1.2. Complexity results

Unfortunately, this problem is NP-complete, as it is proven by Cornuejols et al. [11]. It means that there is no hope to find an algorithm, which finds the exact optimal solution within acceptable time for large problem instances.

However, some special cases are solvable in polynomial time. Kolen [12] proved that the problem is solvable on trees in polynomial time. The facility location problem is defined on tree if the facility candidates and demand points are located in the nodes of a tree. Nonnegative weights are assigned to the edges of the tree as lengths and the transportation cost between a facility candidate and demand point is the length of the path between the corresponding nodes. Bárány el al. [13] improved the running time of the UFLP algorithm on trees.

3.1.3. Approximation results

Since there is no hope to find the exact solution in polynomial time, the significance of the approximation algorithms has increased. Many constant approximation algorithms have been published for the problem that runs in polynomial time. Unfortunately, there are negative results on the approximation algorithms for the FLP as well. Guha and Khuller [14] proved that it is unlikely to find polynomial time approximation scheme for the FLP, that is, a polynomial time algorithm that finds a solution with cost at most $1+\varepsilon$ times the optimum for any given ε. Sviridenko [15] proved that the guaranteed approximation ratio cannot be less than 1.467 unless P=NP. However, better result can be achieved in special cases. Arora et al. [16] found randomised polynomial approximation scheme if the FLP is defined in the plane. Kolliopoulos and Rao [17] proved the same result to Euclidean spaces of constant dimension.

Hochbaum [18], Lin and Vitter [19] gave algorithms with O(log n) approximation ratio for the uncapacitated facility location problem in general case. They proposed a greedy algorithm and a new technique called the filtering technique, respectively. Many constant factor approximation algorithms were found in the special case when the transportation costs are induced by a metric. The table from Bumb's PhD thesis gives a quick overview of the approximation algorithms on metric UFLP. n shows the number of facility candidates and demand nodes. At the next section, we introduce the techniques in Table 2.

Year	Performance guarantee	Reference	Technique	Running time
1997	3.16	Shmoys, Tardos and Aardal	Filtering+LP rounding	
1999	2.408	Guha and Khuller	LP rounding+greedy augmentation	
1998	1.736	Chudak	LP rounding	
1998	5 + ε	Korupolu et al.	Local search	O (n^6 log (n/ε))
2001	3	Jain and Vazirani	Primal-dual method	O (n^2 log n)
2001	3	Arya et al.	Local search	
2000	3	Mettu and Plaxton	Combinatorial	O (n^2)
1999	1.853	Charikar and Guha	Primal-dual method+greedy augmentation	O (n^3)
1999	1.728	Charikar and Guha	LP rounding + primal-dual method + cost scaling + greedy augmentation	
2001	1.86	Mahdian et al	dual fitting	O (n^2 log n)
2001	1.61	Jain, Mahdian and	dual fitting	O (n^3)

		Saberi		
2002	1.582	Sviridenko	LP rounding	
2002	1.52	Mahdian, Ye and Zhang	dual fitting+greedy augmentation	O (n^3)

Table 2 Approximation results for the uncapacitated facility location problem [10].

3.2. Applied methods

3.2.1. Linear Programming

To derive an integer programming formulation of the facility location problem, two variables have to be introduced as follows:

$X_{i,j}$: (0, 1) variable to indicate whether client i is served by facility j.
Y_j: (0, 1) variable to indicate whether facility j is selected.

The other parameters come from the input of the problem, see the Overview of the FLP.
The Facility Location Problem can be now formulated in the form as follows:

minimize

$$\sum_{i\in D, j\in F} c_{i,j} \cdot X_{i,j} + \sum_{j\in F} f_j \cdot Y_j \qquad (1)$$

subject to

$$\sum_{j\in F} X_{i,j} \geq 1, \quad \forall i \in D \qquad (2)$$

$$X_{i,j} \leq Y_j, \quad \forall i \in D, \forall j \in F \qquad (3)$$

$$X_{i,j} \in \{0,1\}, \quad \forall i \in D, \forall j \in F \qquad (4)$$

$$Y_j \in \{0,1\}, \quad \forall j \in F \qquad (5)$$

Part (1) is the cost function. Constraint (2) ensures that each client is assigned to a facility. Constraint (3) describes that a facility is selected if there is at least one client assigned to it.
However, the integer program is usually hard to solve. For this reason, the LP-relaxation of the problem is solved. In this case, the (0,1) constraints have to omitted:

$$X_{i,j} \geq 0, \quad \forall i \in D, \forall j \in F \qquad (4')$$

$$Y_j \geq 0, \quad \forall j \in F \qquad (5')$$

- **Linear Programming Rounding**

The linear programming relaxation can be solved using efficient algorithms. If each variable is integer then we found the optimum solution for the integer program. However, this is usually not the case. Different rounding techniques can be applied to get integer solution from the fractional one. Shmoys Tardos and Aardal [20] gave the first constant factor approximation algorithm. They achieved their result applying the LP rounding technique and the filtering technique.

- **Filtering technique**

This method consists in constructing a "filtered problem" by setting some variables to zero in the integer programming formulation. Then the feasible integer solution to the filtered program can be found either by LP rounding technique or by using some combinatorial algorithm [19].

- **Primal-dual method, Dual fitting**

Jain and Vazarini [21] applied this method first to the facility location problem. Its running time is quite low $O(m \log m)$, where m is the number of edges. This method is also based on the linear programming. But it

uses the dual pair of the primal linear program as well presented above. The dual pair of the above linear program is as follows:

maximise

$$\sum_{i \in D} v_i \qquad (6)$$

subject to

$$\sum_{i \in D} t_{i,j} \leq f_j, \quad \forall j \in F \qquad (7)$$

$$v_i - t_{i,j} \leq c_{i,j}, \quad \forall i \in D, \forall j \in F \qquad (8)$$

$$t_{i,j} \geq 0, \quad \forall i \in D, \forall j \in F \qquad (9)$$

$$v_i \geq 0, \quad \forall i \in D \qquad (10)$$

In case of dual program, we introduce new variables, $t_{i,j}$ and v_i and we have to maximise the profit function. The special feature of the dual program that for any feasible solution the profit in the dual program is always smaller or equal than the cost in the primal program. Primal-dual methods can achieve good approximation by considering both the primal and the dual problems at the same time. Dual fitting methods also use the primal-dual technique. Mahdian et al. [22] achieved the best approximation ratio (1.52) so far for FLP using dual fitting.

3.2.2. Combinatorial algorithms

- **Local search or greedy**

The local search heuristic for facility location problems is extremely straightforward. The idea is to start with any feasible solution and then to iteratively improve the solution by repeatedly moving to the best "neighbouring" feasible solution, where one solution is a neighbour of another if it can be obtained by either adding a facility, deleting a facility, or changing the location of a facility. This heuristic was proposed by Kuehn and Hamburger [7] and was subsequently shown to exhibit good practical performance in empirical studies (see e.g., [23]).

- **Greedy strategies**

This kind of methods has many variants and it is usually combined with other methods. Mettu and Plaxton [24] developed the first linear time algorithm. The construction applied in the algorithm is similar to the one in the primal-dual method used by Jain and Vazirani [21]. They use a "hierarchically greedy" approach whose basic idea is as follows: Rather than selecting a point based on a single greedy criterion, a region is chosen greedily and then a point is selected recursively within that region. Thus, the choice of point is influenced by a sequence of greedy criteria addressing successively finer levels of granularity.

- **Greedy Augmentation (greedy improvement)**

If either the service cost or the facility cost is very high, greedy local improvement decreases the total cost by balancing the two. It always selects the facility in each step where the ratio of the decrease in the total cost and the facility cost is the highest. The greedy local improvement by itself yields a very good approximation for facility location in $O(n^2)$ time. This method is applied in many algorithms: [22, 25-26].

- **Cost scaling**

Charikar and Guha [26] applied this method combined with greedy local improvement and the primal dual method. The idea is to apply the algorithm to the scaled instance and than scale back to get a solution for the original instance. On several occasions, this technique alone improves the approximation ratios significantly.

- **Simulated annealing**

Simulated annealing is a generic global optimization method. Its name comes from metallurgy. In that case, the hot material is slowly cooled in order to improve a better quality.
In case of global optimisation, the algorithm proceeds in the search space. The subsequent points are not always better in each step: the algorithm may select with nonzero probability even worse solutions. This probability decreases monotonically while the algorithm is running. The initially high probability ensures that high areas are discovered. Later the small probability ensures that the solution remains near to the optimum.

3.3. Network flow

Let us imagine a water plumbing where the individual pipes may have different widths and so different maximum flow per unit time. There is an inlet and an outlet as well and the question is how much is the maximum rate at which water can flow from the inlet to the outlet.
This is a network flow problem which can be specified more formally: Given a graph $G(V,E)$ with nodes V and edges E. $c(e) \geq 0$ denotes the capacity (maximum flow possible) of edge $e \in E$. Source $s \in V$ and sink $t \in V$ denotes special nodes of the graph G. G, s, t and $c()$ specifies a network. Furthermore, there can be specified a flow $f(e)$ on the edges of the network which is at most the capacity of the edge $c(e) \geq f(e)$ for each $e \in E$ and the total the inflow to a node is equal to the total outflow from the node except the two special nodes. The value of the flow is the outflow from the source or the inflow to the sink which is the same.
The network flow problem can be applied even to find the computer networks where the capacities are the available bandwidths in the network, the sink is the client and the source is data supplier node and more than one path may be used to transmit the data to the client at the same time.
This problem can be solved efficiently in polynomial time. It remains polynomial if a cost is assigned to each edge and we are looking for a solution with maximal flow and minimal cost. However, the multi-commodity flow problem is already NP-complete, where multiple source and sink pairs are given, and various "commodities" can flow from a given source to a given sink.

3.4. Knapsack Problem

We found the knapsack problem useful to handle with the local resource capacities. The original knapsack problem is as follows: Given a knapsack with maximum volume and several items with different volumes. The optimisation goal is to put as many items as possible into the knapsack. This problem seems to be simple but it is NP-hard, that is, computationally difficult.
Khan [27] applied its variant, namely the multiple-choice multi-dimension knapsack problem (MMKP) for adaptive multimedia problem. The problem can be described as follows. In this case there are some groups with different numbers of items. Each item has a value and resource need. The resource need can be described as a vector because an item need several resources. Furthermore, we know the amounts of available resources. The multi-dimension knapsack problem is to pick exactly one item from each group while maximising the total value of the selected items, subject to the resource constraints.
This model is useful when the overall quality of the delivered multimedia has to be maximised. In this case, the acceptable variations of the requested video can be viewed as a group of items for each client. Each item can be characterised by the resource needs for delivery to the client and by their qualities.

4. IMPLEMENTATIONS

We mentioned that the host recommendation has different variants and for example. We wanted to avoid duplication of the algorithms. For this reason, the algorithms operate on objects which hide the differences between the LWR and HWR. Due to the underlying common model, the algorithms are able to find solutions for each type of the host recommendation problem but they are not equally appropriate for different cases. For example, facilities are assigned to the client requests to serve the video streams; they are the pairs of streamer and hosting nodes. In case of HWR, the streamer node *StrNId* can be neglected.

4.1. Linear Programming Rounding

We chose an algorithm based on linear programming rounding for the solution of the host recommendation problem. Our linear program is much more complicated than in case of facility location problem. To derive the integer programming formulation of the problem, we had to introduce variables $X_{i,j,k,l}$ to indicate whether the video variation l is served to the client i by DM group k through DC j. Furthermore, S_i indicates whether any request of client i is served or not. Similarly to the facilities, $Y_{j,k}$ indicates whether the connection between DM group k and DC j is used for the service. Weights are assigned to the different optimization criteria. They express their priorities. Each weight is higher than the maximum of the subsequent optimization criteria. We ensure in this way that the subsequent criteria are concerned only if the criteria with higher weights are equal. The exact LP formulation can be found in [28]. We had to add constraints to express the network and node resource constraints.

The possible values of the variables are 0 and 1. Since the time complexity to find the exact solution for an integer linear programming problem is large, we consider the LP-relaxation of the problem, where the possible values of the variables can be any positive real number. To solve the linear program, we applied SOPLEX, an object oriented implementation of the simplex algorithm [29]. This software is integrated into the host recommendation algorithm.

If $X_{i,j,k,l} = 1$ in the solution then let the video variation l of client i be served by server k through proxy j. The possible fractional values of X-type variables do not represent legal solutions. We round them in a greedy manner. We take each X variables with fractional value one after the other and we try to select the current client-DC-DM group route denoted by the variable. We have to check the resource constraints. The route is selected for video delivery if and only if these conditions are fulfilled after the selection of the route.

4.2. Incremental Algorithm

The incremental algorithm takes the facilities one after the other and the facility is selected if it improves the solution. It has been originally developed for LWR. However, some goals for HWR, such as minimising the number of storage nodes can be easily incorporated. The incremental algorithm is a very simple but efficient method. For large problem inputs a time limit can be set to abort running the algorithm when the running time would be too long, and then apply the current recommendation for configuration (this is, why the algorithm is called incremental).

Some optimization criteria such the number of storage nodes can be easily incorporated but in its original form, the incremental algorithm is unsuitable to minimize the network load arising from migration/replication. This is the reason while we introduce its extension, namely Complex Incremental Algorithm, which works on three levels: it tries to add facilities one after the other to the recommendation on the highest level, it examines the possible video instances on the current facility on the next level and at last it tries to assign the client requests to the video instances. For more details see [30].

The scheme of the algorithm:

for each facility **do**
    Decide on selecting the current facility
    if the facility is selected **then**
        for each client connected to the current facility **do**
            Decide on assigning the current client to the facility

4.3. Greedy

In this algorithm we start with an empty set of proxies. In each turn that proxy is added that decreases the cost the most. When calculating the cost, in each turn a new configuration has to be generated based on the previous one and the recently extended proxy set. This configuration setup is made by an inner greedy algorithm that for every client selects the best proxy from the set, and tries to modify the index of the chosen demand by adding or subtracting one. Each proxy contacts the server that gives the least cost. [31]

4.4. Particle Swarm

The particle swarm algorithm is based on the algorithm of Kennedy and Eberhardt [32]. Their original variant uses a set of particles, which particles describe a concrete configuration each. The particles are connected to each other thus forming a given topology.

The particles (i.e. the configurations they describe) are initialized with random values. The particles remember their own best configuration (b_p), and they know which neighbour of theirs (including themselves) gives the least costy configuration (b_n). The problem solving is divided into rounds. In each round the particles calculate their new configuration by combining b_p and b_n, during which they utilize stochastic values also. The algorithm continues until some predefined condition is reached.

The definition of the original combination looks thus:

$$\begin{cases} \vec{v}_i(t) = \vec{v}_i(t-1) + \varphi_1(\vec{p}_i - \vec{x}_i(t-1)) + \varphi_2(\vec{p}_g - \vec{x}_i(t-1)) \\ \vec{x}_i(t) = \vec{x}_i(t-1) + \vec{v}_i(t) \end{cases}$$

where φ_1, φ_2 are random numbers from interval *[0,1)*, $\vec{v}_i(t)$ is the 'velocity' (amount of change) at time *t*, $\vec{x}_i(t)$ the value of the vector at time *t*, $\vec{p}_i$ the least costy vector of particle *i*, $\vec{p}_g$ is the vector of the actually least costy neighbour. When the algorithm reaches the predefined condition, it ends. The result is the vector of the least costy particle.

In our current case (ADMS) the elements of vector $\vec{x}$ are the following:

$$\vec{x} = \begin{bmatrix} \vec{\kappa} \\ \vec{\lambda} \\ \vec{\mu} \end{bmatrix} \quad \begin{array}{ll} \vec{\kappa} \in C, & \text{client } i \text{ connects to DC } \kappa_i \\ \vec{\lambda} \in P, & \text{DC } i. \text{ connects to DM } \lambda_i \\ \vec{\mu} \in Q_i, & \text{client } i \text{ gets demand } \mu_i \end{array}$$

The algorithm, in its original form cannot be applied to problems similar to the one presented in this paper, when the elements of the representation get their values from non-orderable sets, because in these cases the addition in the above formula makes no sense. (E.g. two DCs can not be sensibly added.)

As a solution, we proposed and applied the following vector-modification:

$$\begin{cases} m_l(t) = \dfrac{h(\vec{x}(t-1), \vec{p})}{L} m_g \\ m_a(t) = (1-\alpha) m_a(t-1) + \alpha m_l(t) \\ \vec{x}_i(t) = \begin{cases} \begin{cases} \vec{p}_i & \text{if} \quad \varphi_1 \le m_c \\ \vec{x}_i(t-1) & \text{if} \quad \varphi_1 > m_c \end{cases} & \text{if} \quad \varphi_2 \le m_a(t) \quad \text{(crossover)} \\ e \in S_i & \text{if} \quad \varphi_2 > m_a(t) \quad \text{(mutation)} \end{cases} \end{cases}$$

Where *p* is the vector of the least costy neighbour, $h(\vec{a}, \vec{b})$ is the Hamming distance of the two vectors, *L* is the length of the vectors, *mg* is the global muation rate, *ml* is the local mutation rate, *ma* is the average mutation rate, α is a scaling factor, *mc* is the crossover-rate, *S* is the possible value set of the given vector element.

The meaning of the formula is that the bigger the difference between the vector of a particle and that of its least costy neighbour, the bigger the mutation probability of the vector elements (taking into account the history of the probability). For each element, that did not mutate, a new probability is counted for getting the

respective value from the neighbour. The Hamming distance is used because without that at the end of the calculation a fluctuation can be observed, which prolongs the fulfilment of the stop condition.
The algorithm runs until the cost deviation of the particles is within a given bound.

4.5. The simple algorithm

It assigns each client one after the other to the facility that can serve the best video instance to it and if there are more such facilities it selects streaming route with the best QoS parameters. This method is especially appropriate in LWR for deciding from where to serve a client when the number of clients is small. It is not adequate to handle the facility cost. For an example, see [3] in these proceedings.
The scheme of the algorithm:

for each *client* **do**
    best_facility ← 0
    for each *facility* **do**
        if better *facility* for the client than the *best_facility* **then**
            best_facility ← *facility*
    Assign *best_facility* to the client

4.6. Best gain per resource usage first

It is based on the multiple-choice multi-dimension knapsack problem [27]. In each step, the algorithm always selects the possible video instance with highest value per unit resource consumption where the value of a video instance is the total value of the client requests servable from it. This algorithm is developed for HWR. However, it can be adapted to LWR as well.
The scheme of the algorithm:

best_ gain ← *1*
while *best_gain* > *0* **do**
    for each *v*∈ *possible video instances* **do**

$$ff_v \leftarrow \max_j\left(\frac{needed\_resource_j(v)}{free\_resource_j(v)}\right)$$

        if $ff_v > 1$ **then** $gain_v = \frac{value(v)}{ff_v}$
        else $gain_v \leftarrow 0$
    $best\_gain \leftarrow \max_v(gain_v)$
    if *best_gain* > *0* **then**
        Select *v* for usage as video instance

This algorithm needs some explanation. The video instances can be characterised by the variation of the video and the storage node where it is located. ff_i is called feasibility factor which is at most 1 if there are enough resources. The value of an instance can be for example the expected number of clients that can be served from this instance. The *gain* is the ratio of the value and the feasibility factor. For an example, see [3] in these proceedings.

5. CONCLUSIONS

The area of optimisation algorithms is very rich in different methods and it offers plenty of approaches to solve the host recommendation problem. This problem is much more complex than the well-known optimisation problems. For this reason, the known problems should be extended and different approaches should be combined. In many cases, the running time is more critical than the good approximation and fast

heuristic methods have great significance. In an earlier work [28], we conducted tests to compare four different algorithms (greedy, particle swarm, linear programming rounding and incremental algorithm). The particle swarm algorithm produced the best result while the incremental algorithm found the solution in the shortest time.

REFERENCES

[1] R. Tusch, *Design and Implementation of an Adaptive Distributed Multimedia Streaming Server*, PhD Thesis, University Klagenfurt, 2004

[2] R.Tusch, L. Böszörményi, B. Goldschmidt, H. Hellwagner, P. Schojer, *Offensive and Defensive Adaptation in Distributed Multimedia Systems*, Computer Science and Information Systems (ComSIS), Vol. 1, No. 1, Feb. 2004, pp. 49-77.

[3] P. Kárpáti, T. Szkaliczki and L. *Böszörményi: Abstracting and characterizing distributed VoD servers*, in Proceedings of 3rd Hubuska Open Workshop, Klagenfurt, 27-28 April, 2006.

[4] Ottó Hutter, Tibor Szkaliczki and Balázs Goldschmidt: *Host Recommendation in the Adaptive Distributed Multimedia Server*, ERCIM News, Number 62, July 2005, pp. 34-35

[5] P. Kárpáti, T. Szkaliczki, L. Böszörményi, *Mathematical model for distributed video-on-demand servers*, Technical Reports of the Institute of Information Technology, University Klagenfurt, TR/ITEC/05/2.13

[6] A. S. Manne. Plant location under economies-of-scale-decentralization and computation. Management Sci. 11:213-235, 1964

[7] A. A. Kuehn and M. J. Hamburger. A heuristic program for locating warehouses. Management science, 9, pages 643-666, 1963

[8] J. F. Stollsteimer. The effect of technical change and output expansion on the optimum number, size and location of pear marketing facilities in a California pear producing region. PhD thesis, University of California at Berkeley, California, 1961

[9] P. B. Mirchandani and R. L. Francis, eds. Discrete Location Theory. John Wiley and Sons, Inc., New York, 1990.

[10] A. Bumb, Approximation Algorithms for Facility Location Problems, PhD thesis, Twente University Press, 2002

[11] G. Cornuejols, G. L. Nemhauser and L. A. Wolsey. The uncapacitated facility location problem. In P. Mirchandani and R. Francis, editors, Discrete Location Theory, John Wiley and Sons, New York, pages 119-171, 1990

[12] A. Kolen. Solving covering problems and the uncapacitated plant location problem on trees. European Journal of Operations Research, 12, pages 266-278, 1983

[13] I. Bárány, J. Edmonds and L. A. Wolsey. Packing and covering a tree by subtrees. Combinatorica 6, pages 245-257, 1986

[14] S. Guha and S. Khuller. Greedy strikes back: Improved facility location algorithms. Journal of Algorithms 31 (1), pages 228-248, 1999

[15] M. Sviridenko. Personal communication. Cited in S. Guha, Approximation algorithms for facility location problems, PhD thesis, Stanford, 2000

[16] S. Arora, P. Raghavan and S. Rao. Approximation schemes for Euclidean k-medians and related problems. In Proceedings of the 30th Annual ACM Symposium on Theory of Computing, pages 106-113, 1998

[17] S. Kolliopoulos and S. Rao. A nearly linear-time approximation scheme for the Euclidean k-median problem, In Proceedings of the 7th Annual European Symposium on Algorithms, Springer Verlag LNCS vol. 1643, pages 378-389, 1999

[18] D. Hochbaum. Heuristics for the fixed cost median problem. Mathematical Programming, 22, pages 148-162, 1982

[19] J. H. Lin and J. S. Vitter. ε-approximation with minimum packing constraint violation. In Proceedings of the 24th Annual ACM Symposium on Theory of Computing, pages 771-782, 1992

[20] D. B. Shmoys, E. Tardos, and K. I. Aardal. Approximation algorithms for facility location problems Springer. In Proceedings of the 29th Annual ACM Symposium on Theory of Computing, pages 265-274, 1997.
[21] K. Jain and V. Vazirani. Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and Lagrangian relaxation. Journal of the ACM, 48, pages 274-296, 2001
[22] M. Mahdian, Y. Ye, J. Zhang: Improved approximation algorithms for metric facility location problems, Lecture Notes in Computer Science 2462: 229-242 2002
[23] K. Hidaka, H. Okano: An Approximation Algorithm for a Large-Scale Facility Location Problem, Algorithmica, Vol. 35, pp. 216-224 (2003)
[24] R. R. Mettu and C. G. Plaxton. The online median problem. In Proceedings of the 41st IEEE Symposium on Foundation of Computer Science, pages 339-348, 2000
[25] S. Guha and S. Khuller. Greedy strikes back: Improved facility location algorithms. Journal of Algorithms 31 (1), pages 228-248, 1999
[26] M. Charikar and S. Guha: "Improved Combinatorial Algorithms for the Facility Location and *k*-median Problems". IEEE Symposium on Foundations of Computer Science, pp. 378-388, 1999.
[27] S. Khan, K.F. Li, E.G. Manning, M. Akbar, *Solving the knapsack problem for adaptive multimedia*, Studia Informatica (special issue on Cutting, Packing and Knapsacking Problems) 2 (1), 2002
[28] B. Goldschmidt, T. Szkaliczki and L. Böszörményi: *Placement of Nodes in an Adaptive Distributed Multimedia Server*, In Proceedings of the 10th International Euro-Par Conference, 2004, pp. 776-783
[29] Roland Wunderling: Paralleler und Objekt-orientierter Simplex- Algorithmus, ZIB technical report TR 96-09, Berlin 1996
[30] Tibor Szkaliczki and László Böszörményi: Incremental Placement of Nodes in a Large-Scale Adaptive Distributed Multimedia Server, In Proceedings of International Conference on Distributed and Parallel Systems (DAPSYS 04), September 19th- 22nd, 2004, Budapest, Hungary, Series: The Kluwer International Series in Engineering and Computer Science, Kluwer Academic Publishers, September 2004, Vol. 777
[31] Balázs Goldschmidt and Zoltán László: *A Proxy Placement Algorithm for the Adaptive Multimedia Server*, 9th International Euro-Par Conference, 2003, pp. 1199-1206
[32] James Kennedy and Russel C. Eberhardt. Swarm Intelligence. Morgan Kaufmann, 2001.

Session 2 – Multimedia Processing and Presentation

ON-THE-JOB TRAINING FOR WORKING WITH MULTIMEDIA CONTENT OF THE CULTURAL HERITAGE

Danail Dochev
Institute of Information Technologies – BAS, Bulgaria

Radoslav Pavlov
Institute of Mathematics and Informatics – BAS, Bulgaria

Virgílio Hipólito Correia
Museu Monográfico de Conimbriga, Portugal

Zsolt Markus
Eduweb Multimedia, Hungary

1. INTRODUCTION

This paper presents the main objectives and some results of the LdV project JASON "On the Job e-Training Skills to Deal with Digital Cultural Heritage Content"(2004-2006). The project partners (from Portugal, France, Hungary and Bulgaria) work on JASON learning platform, providing facilities and learning content for OJT e-training aimed at building and enhancing e-competences for specialists in cultural memory institutions to deal with digital objects of the cultural heritage. The project target group covers specific group of knowledge workers - librarians, archivists, museum curators, conservators, historians, archaeologists, art historians, artists, e-publishers, etc. In the last years the professionals in memory institutions work actively to acquire, organise, store, manage and use large amounts of human knowledge in digital form. The modern Information technologies offer them continuously growing set of powerful tools to digitise cultural content and to present it in the global information networks. The fast growth and enrichment of approaches, technological standards, tools and methodological facilities in the multimedia and web technologies require continuous update and acquisition of new knowledge and skills.

On-the-job training (OJT) is a form of vocational e-training - individualized training based on the e-learning paradigm that allows an employee in need of training to receive the necessary knowledge, develop the required skills and improve performance – all while on the job. These characteristics are very important both for the employee and the employer, since they provide the opportunity for quality training while limiting the effect on work plan.

The paper is structured as follows: the first part presents the conclusions of analysing the target groups' educational background, ICT attitudes and learning needs. The second part reviews directions and topics selected by the partners to develop appropriate e-competences for professionals from memory institutions. The third part presents briefly requirements and decisions for implementation of the JASON e-Learning environment.

2. LEARNING NEEDS OF CULTURAL MEMORY INSTITUTIONS PROFESSIONALS

The project partners analysed the project target group's job descriptions and typical educational backgrounds in the partner countries, using especially the experience of the Portuguese Network of Museums. This allowed to outline the following characteristics of the professional groups, directly concerned by the project:

- Curators, archaeologists and other graduate technicians

- They are graduate professionals, commonly also having a MA, Msc. or PhD (more rarely). Their ICT aptitude varies enormously, depending on the individual professional curriculum, from the very familiar with ICT at various levels of use, to the computer inept person. Curators coordinate and perform tasks for researching, inventorying, analysing, exhibiting, disseminating or otherwise organizing cultural heritage and coordinate conservation projects, especially preventive ones. Archaeologists and conservation architects will have a similar job description, but with a stronger emphasis on concrete intervention projects on sites, monuments, collections, etc. Graduate technicians in libraries and archives will have their job description directed more clearly to day-to-day management of collections.

- Technical personnel

It consists of persons with a BSc. in a specific area (Conservation, Photography, Radiology, Archival techniques, Library inventorying, Archaeological excavation). Having more recent and more technical-oriented education ICT aptitude will be more widely spread. Job descriptions tend to be very case specific, directed to the actual performance of a technically defined task (which is normally a part of a larger project designed and coordinated by a professional of a more empowered professional group) and they normally imply the personal responsibility of the technician in the performed job, from the choice of method to the documentation of the task for future reference.

- Professional technicians

They fulfil the various tasks needed to accomplish projects, but lack any actual responsibility for the final quality of the work. They are supposed to work under direct supervision. These professionals normally do not need university education, but generally require professional training (or practice in an institution at a auxiliary level for a number of years, with adequate in-house training, leading to a promotion to this level). ICT aptitude will vary according to the specific field.

The current e-Competences of memory institutions professionals in the partner countries and their picture of the topics they need to learn in the field were additionally investigated by means of common questionnaire containing 9 groups of questions.

3. BASIC JASON E-TRAINING COURSES

The analysis of the target groups job descriptions and available e-competences allows to determine their specific learning needs for ICT knowledge and skills. On the base of this analysis JASON developers determined the scope of necessary e-competences in four sub-areas, considered central for acquiring competences to deal with digital objects of the cultural heritage. As a result a set of topics to be covered by the project eLearning courses was selected.

3.1. Technologies, standards and methods for digitising cultural heritage master pieces.

These courses are intended to introduce the trainees from memory institutions to the modern computer multimedia technologies, available standards and digitising methods. The content of the courses considers the mean information processing background of the learners. The brief structure of the two basic courses from this group is described below.

3.1.1. Introduction to Media Representation

This course consists of 39 HTML pages (plus glossary and bibliography) and 54 pictures/multimedia objects, and has the following structure:

- Digitalization - why use it and digital images
- Introduction of Image File Formats (GIF , JPEG , TIFF, PNG)

- Animation
- Sound (attributes, types, perception of sounds, file formats)
- Video, video streams
- Devices for digitalization (scanner, digital camera; microphone, synthesizer, recorder; video camera, video editing card).
- Types of media manipulation software
- Navigational systems

3.1.2. The Digital Image

This course consists of 99 HTML pages (plus glossary, bibliography, online documents, links) and 242 pictures/multimedia objects, and is structured into the following lessons:

- Introduction to Graphics (proportion and intensity, contrast and dominance, shade and tints, screen resolution etc.)
- Colours (basics, features, combinations, contrast etc.)
- Digital Gallery
- Graphic file formats (raster vs. vector files, GIF, GIF compression, interlaced GIF, JPEG, PNG)
- Graphic File Formats for the Web
- Imaging strategies (interface elements, photographs as GIFs, photographs as JPEGs, diagrams and illustrations as vector graphics)
- Multimedia (introduction, applications, standard formats, web multimedia strategies)
- Digital media processing (audio processing, video processing, delivery, streaming, downloading, drawbacks , design and multimedia)
- Scanning (introduction, scanning for printing or video screens, resolution, screen bit-depth)
- Scanner features (resolution for the video monitor, scan resolution and image size)
- Digital cameras (introduction, vocabulary, optics, viewfinder, storing images).

B/ Processing of digital objects

The new e-competencies to be acquired by the trainees are oriented towards work with multimedia digital libraries, archives and galleries containing diverse multimedia types, with support for 2D and 3D objects. These competencies will permit to the professionals of memory institutions and cultural industry to store and catalogue easier digital objects and to integrate the multimedia content in new complex digital artefacts, thus increasing the knowledge and reducing the costs to produce and publish educational and common-interest materials. The brief structure of the courses, developed up to now is descried below.

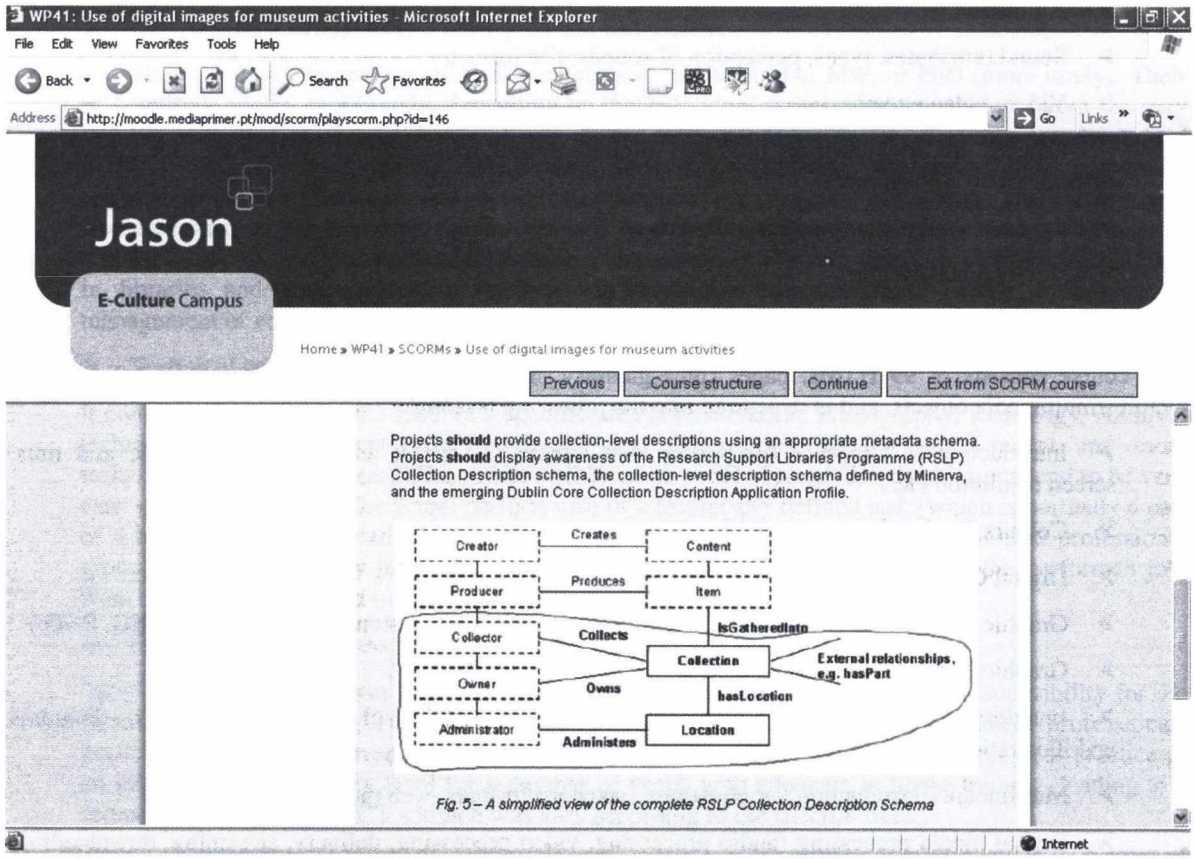


Fig 1. Sample screen of a JASON e-training course /1/

3.1.3. Multimedia Authoring Packages

The course contains 95 HTML pages (plus glossary, bibliography, online documents, links) and 103 pictures/multimedia objects, and consists of the following lessons:

- Authoring tools versus programming tools
- Authoring tools (CD-ROM based, cards or page-based, Icon or object-based, time-based tools. examples – HyperCard, Asymetrix Toolbook, Macromedia Authorware, Macromedia Director)
- Web-based authoring tools (Microsoft FrontPage, Macromedia DreamWeaver, Claris HomePage, Adobe PageMill , Macromedia HomeSite)
- Main opportunities of multimedia authoring tools
- Choosing a right authoring tool (vs. users, projects purpose, multimedia elements, budget etc.)
- Introduction to Microsoft Power Point
- Introduction to Macromedia Director
- Introduction to Macromedia Flash
- Introduction to Adobe Premiere Pro

3.1.4. Use of digital images for museum activities

The course contains 44 HTML pages (plus glossary and bibliography) and 37 pictures/ multimedia objects, and covers the following themes:

- Preparation for digitalization (hardware, software, environment)

- Project planning (management policy for digital assets, defining the audience, evaluating assets)
- Storage and management of the digital collections (capture and storage, preservation strategies)
- Metadata creation/capture (scope, standards, types of metadata, collection-level description)
- Process of publication of digital collections (accessibility, security, authenticity)
- Disclosure of resources (searching and retrieval, browsing, visual and content-based retrieval).
- A sample view of a screen from this course is shown on Fig. 1.

3.2. Using effectively new forms in remote collaborative working and work flow.

This category of courses covers themes such as portal technologies, shared work spaces, collaborative working, approbation strategies and policies for collective publishing etc. The following course is developed up to now.

3.2.1. New forms for remote collaborative working and workflow in memory institutions

The course has the following structure:

- Introduction to the Internet (basics, getting connected, Intranet/ Extranet)
- Using the Internet (browsing, bookmarking, download/upload, cache, data compression)
- Finding information on the Internet (where and how to search, search results)
- Web database (models, implementation, digital archives)
- Information workflow (e-mail client, address book, mailing list, sending e-mails)
- Manage time (meetings schedule)
- Collaborative tools (forums, glossaries, chat, videoconference, virtual networking)

3.3. Effective use of Web digital content (including search by semantic web approach).

One course is developed up to now in this category, considering the overall information processing background of the learners and their. interests. The brief structure of the course is described below.

3.3.1. Semantic Web approach to access information on Internet

The course consists of the following lessons:

- Problems with the current information retrieval in Internet.
- Semantic Web scheme and levels
- Resources and identifiers in the Semantic Web
- Ontologies - concepts, structure, examples
- Ontology description languages
- Problems with actuality, reliability, trust-worthiness of the Internet information
- Information preparation and retrieval in the Semantic Web. Annotations.
- Agent technologies in the Semantic Web
- Perspectives and problems. Integration of information sources
- Example of Semantic Web site for cultural heritage
- A sample view of a screen from this course is shown on Fig. 2.

4. JASON ELEARNING PLATFORM IMPLEMENTATION

According the investigation of JASON e-training forms and learning context, and considering some previous experience of the authors with e-training on the workplace, the following functional requirements for the JASON learning environment were formulated:

- It has to provide sufficient level of interactivity in order to support learning-by-doing operation mode.
- It has to supply facilities for skills acquisition.
- It has to permit fast activation of the learning environment and fast restoring of the working environment.
- It has to allow fast restoring and continuation of the e-training process by storing the current learning status of the trainee on exiting the learning environment.
- It is desirable to simulate operation in apprentice mode, presenting exemplary solutions of practical problems.
- It has to allow easy and natural communication of the trainees with the instructor and with peers.

The JASON environment implementation (Fig. 3) consists of 4 local distant training centres in the partner countries, each integrating two functional software subsystems: virtual training studio for implementation of OJT process, and authors' studio for development of multimedia training materials. The local training centres are developed through modification and customization of an open source Learning Content Management System (Moodle), as it meets the basic JASON requirements. The e-training courses are organized according the SCORM standard for re-usable eLearning materials. There is also a Central repository of courseware materials, keeping the versions of courseware materials in English, using both internal MOODLE format and SCORM format. All local repositories access the central repository.

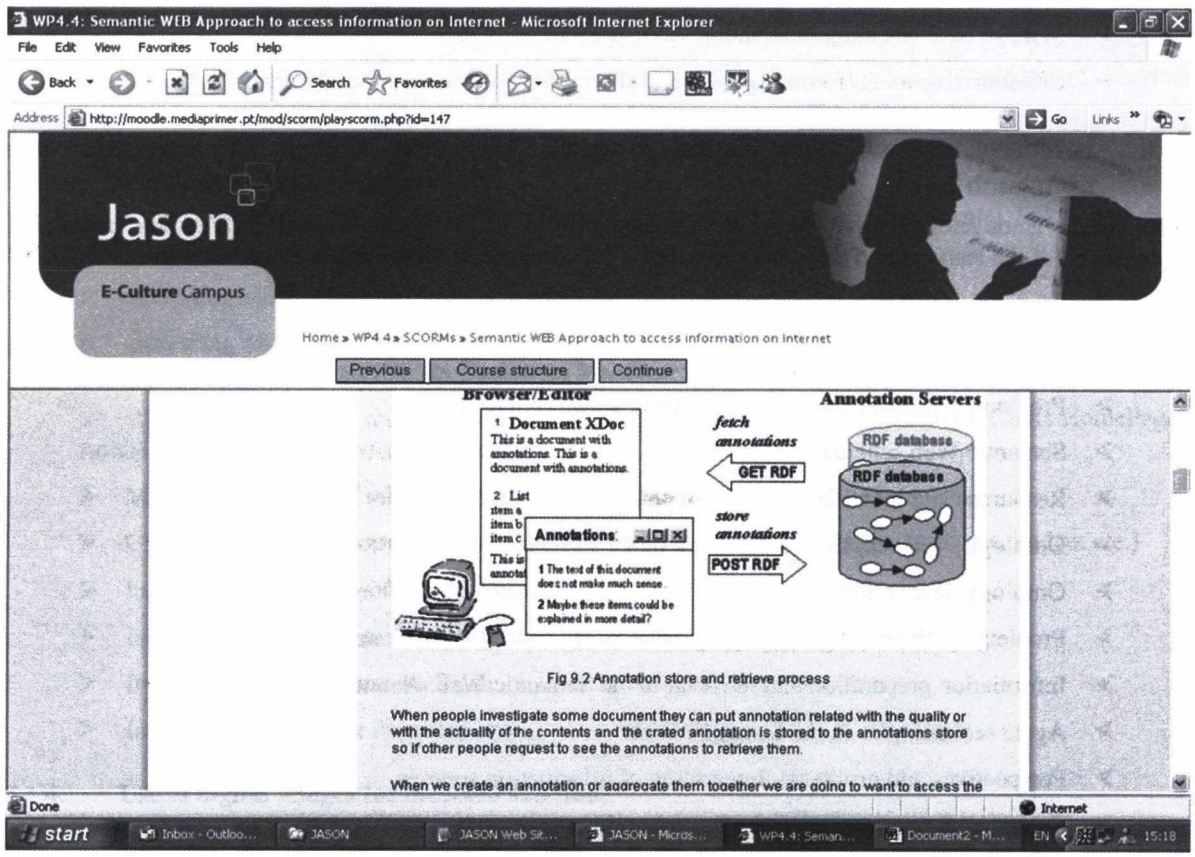


Fig 2. Sample screen of a JASON e-training course /2/

To use JASON environment (http://moodle.mediaprimer.pt/) the end-user needs only to have a multimedia computer connected to Internet. It does not require the installation of any specific software, just a standard software platform: operation system, browser, video, sound reading programs, etc. The JASON environment allows different users to use its resources all the time. The role of the instructor is to assist and guide the learners if they experience difficulties. The environment supports the learning process with synchronous and asynchronous communications and includes context oriented forums, chat rooms, message board. The communications tools are available for interaction both among the learners themselves, and between the learners and the instructors as well.

5. CONCLUSION

As results of their work on the project the JASON developers determined the scope of e-competences, necessary for memory institutions professionals to handle and access digital objects of the cultural heritage. The created e-training courses have just started their experimental use in partners' languages versions. The experimental use has to test the functionality and demonstrate the usability of the JASON environment in order to supply feedback for improvement and enhancement of the training content as well as of the e-training environment implementation.

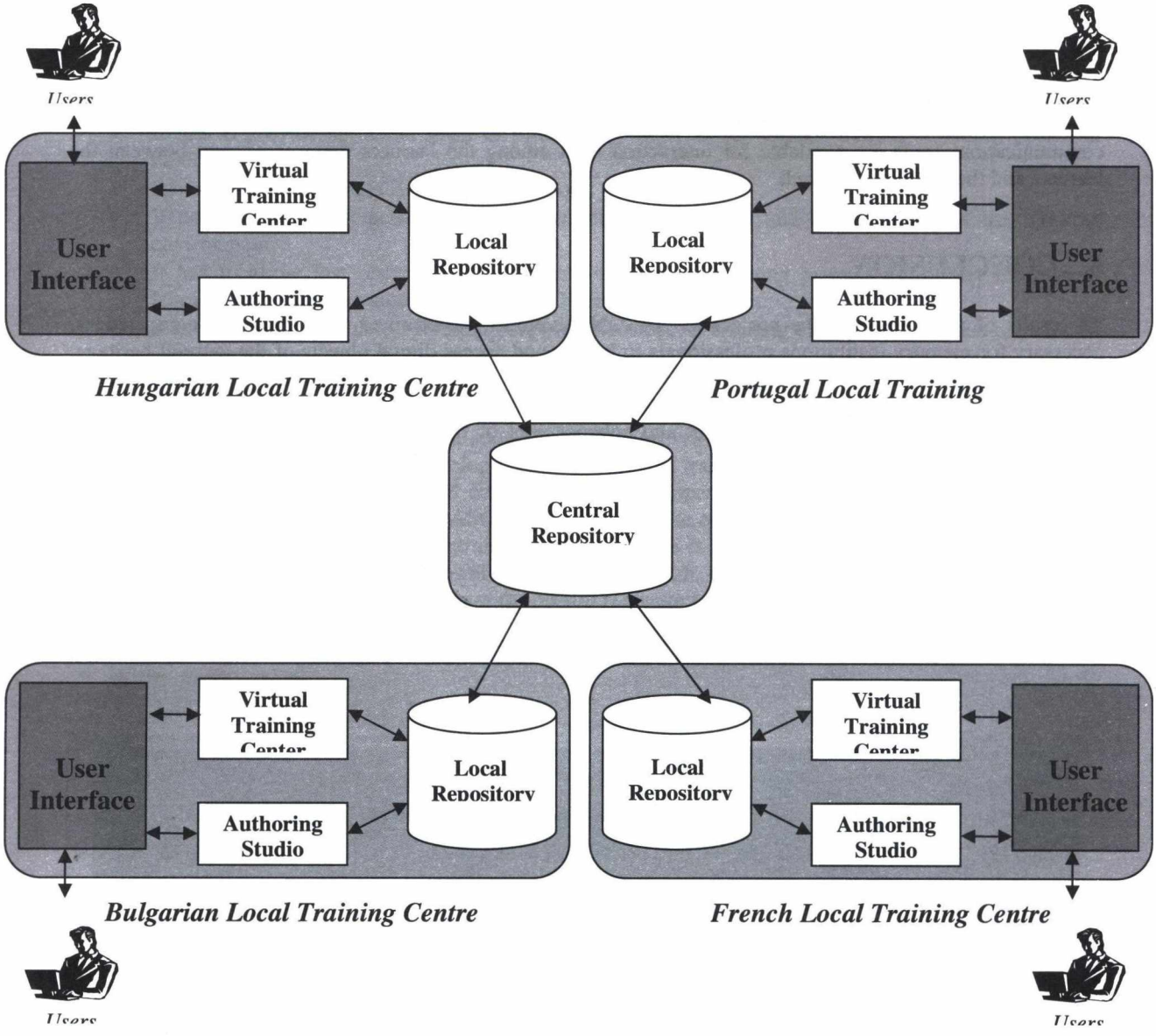


Fig 3. Architectural scheme of the JASON environment

REFERENCES

[1] DOCHEV D., R. PAVLOV, O. HUTTER, I. SIMONICS, P. ARAPI (2004). *Virtual Campuses – from Global Architecture to Local Solutions.* Proceedings of the EDEN Annual Conference 2004 "New Challenges and Partnerships in an Enlarged European Union", Budapest, pp.397-403

[2] DOCHEV D., R. PAVLOV (2005) *E-Learning Solutions for On-The-Job Training.* Proc. of IADIS e-Society 2005 Conference, Malta, 27-30 June 2005, pp. 433-437.

[3] ENCONTRO MUSEUS E EDUCAÇÃO (2002) *Actas*, IPM, Lisboa.

[4] FORUM INTERNACIONAL REDES DE MUSEUS (2002). Actas, IPM, Lisboa.

[5] JACOBS, R.L. AND M.J. JONES (1995). *Structured On-The-Job Training; Unleashing Employee Expertise in the Workplace.* San Francisco: Berrett-Koehler Publishers.

[6] MOODLE (CURRENT VERSION: 1.5.1) (2005), http://moodle.org

VIDEO STREAMING IN A WEB-BASED SEMANTIC SEARCH ENVIRONMENT FOR VIDEO SEGMENTS

HUBUSKA Open Workshop III

Dipl.-Ing. Armin Mueller
ITEC, University of Klagenfurt
armin.mueller@itec.uni-klu.ac.at
April 9, 2006

Abstract

One of the main features of the web-based semantic search environment of the DAHL project is the ability to use adaptive video streaming for the delivery of the video segments to the users. To provide the resulting video segments of a metadata search to users that do not use a video player that supports adaptive video streaming, an additional way of video streaming has been added. The Darwin Streaming Server from Apple[1] is a robust and easy to manage video server that is freely available. This server supports streaming over HTTP, so nearly every video player should support this kind of video streaming. There is also support for the use of SDP files (for segmentation) integrated into the server. In this demo abstract we show the benefits from migrating from a proprietary streaming server solution to the Darwin Streaming Server.

1. OVERVIEW

During the DAHL project (*Delivery with Adaptation and High-Level Video Indexing*) at the ITEC, a web-based semantic video search environment has been developed. This video search [4] is integrated into a special kind of e-learning environment, a historical and technical exhibition about three famous computing scientists (Dahl, Dijkstra and Nygaard), called *People behind Informatics*[2]. It is possible to search for specific terms and semantic data, whereas the results are full videos or video segments that contain the given search criteria. So it is possible to search e.g. for video segments in which Dijkstra talks about the *Shortest Path Algorithm* at a conference at the Newcastle university in 1992. The results of the search are presented with some metadata, like screen capture and semantic information, and a link to a SDP (Session Description Protocol)[1] file that contains all the necessary information for playback (in case of segment the offset to the start of the video). The video streaming server that is used from the ViTooKi project[5] is capable to deliver the video or video segment through adaptive streaming, where the video is adapted dynamically to the constraints of the current bandwidth. This kind of transmission requires a special player that is capable of understanding this technique, so the used player is the MuViPlayer from ViTooKi. The problem of this approach, despite that the MuViPlayer is a proprietary solution, is that the installation of the player requires some technical knowledge from the user (compiling, integration into the web-browser).

2. MIGRATION TO THE DARWIN STREAMING SERVER

The goal of the virtual exhibition *People behind Informatics* is to transfer the knowledge to interested people all over the world. Up to now the retrieval of the video information was bound to the use of the MuViPlayer from ViTooKi. After a short evaluation of other freely available streaming servers, the open-source package Darwin Streaming Server (DSS) was chosen. This media server offers a lot of features for video streaming, but more important is the fact that the widespread QuickTime media player works well together with the advanced features of this server. The DSS is maintained by Apple together with feedback from the open source community, so bugs and errors are relatively fast fixed. For transferring the video to the user, the server (in conjunction with the QuickTime player) uses several protocols, like RTP, HTTP or FTP. The standard way of transferring videos is RTP, where very good performance can be achieved with very good

[1] Apple Darwin Streaming Server, http://developer.apple.com/opensource/server/streaming/index.html

playback behaviour. In case of a firewall being between the user and the DSS, the server falls back to HTTP or FTP streaming (progressive downloading). Tests have shown that even with HTTP streaming the playback starts immediately and playback controls like skipping forwards or backwards in the video work very well. The DSS also supports SDP files, which is important for the use of segmentation features that are part of the semantic video search. The DSS can replace the currently used MuViServer from ViTooKi as the default streaming server. For best delivery circumstances, where enough bandwidth is available, playback with all possible devices that support the properties of the video will work. But in the case where not enough bandwidth is available and jitter or delays can occur, or even if there are client devices with limited playback capabilities, the DSS does it's work as usual, but the user can not view the video without interruptions or sometimes not at all. With the integration of video variations as in proposed MPEG-7 [3], these issues can be solved to some degree. For each result of the semantic video search, there are other links to SDP files available, that each one points to an other version of the same video or video segment. The variations can be bitrate reduced, greyscale encoded, spatial reduced or a combination of them. So the user has to choose the version that fits his needs best, if the standard version does not work very well with his playback environment. The technical steps for the migration to the DSS were, besides the installation of the DSS that was guided through a script, the manual rewriting of the MPEG-7 descriptions, where only the URI had to be changed to the location of the videos on the DSS.

3. CONCLUSION

Through the use of a widespread and open-source video streaming server, we could effectively replace our current video streaming environment. The current environment worked well in our test-scenario, but for technically non-trained end users it was hard if not impossible to get the MuViPlayer working together with the MuViServer from ViTooKi. The DSS accompanies the QuickTime player that is also widespread and easy to use. This player works together with all the features that the DSS provides. The standard way of transferring the video results to the use is RTP video streaming with the capability to use a HTTP or FTP fallback in case of firewall issues. By using MPEG-7 variation descriptions it is also possible to offer the user a kind of adaptation through the providing of pre-encoded versions of the original video. These variations can be used to choose the one for the user that fits its playback capabilities best for retrieval of video information.

REFERENCES

[1] M. Handley and V. Jacobson, *SDP: Session Description Protocol*, RFC 2327, Internet Engineering Task Force (IETF), April 1998, available as http://www.ietf.org/rfc/rfc2327.txt.

[2] ITEC, *People Behind Informatics*, Virtual Computer Science Exhibition web application, August 2003, http://cs-exhibitions.uni-klu.ac.at/.

[3] ISO/IEC JTC 1/SC 29/WG 11 (MPEG), *Multimedia Content Description Interface – Part 5: Multimedia Description Schemes*, October 2001.

[4] A. Müller, *The DAHL Project – Video Search Design*, Deliverable D5-2, ITEC, University of Klagenfurt, 2004.

[5] *ViTooKi – The Video Tool Kit*, open-source project homepage, May 2004, http://sourceforge.net/projects/vitooki/.

THE M3-SYSTEMS MOBILE E-LEARNING GUIDE

Michael Kropfberger
Institute for Information Technology
University of Klagenfurt
Universitätsstraße 65-67, 9020 Klagenfurt, Austria
mike@itec.uni-klu.ac.at

Abstract

Over the last two years, our M3-Systems Research Lab worked on a mobile guide for a local outdoor museum called Minimundus. Scaled models of nearly 140 world-famous tourist attractions like the Eiffel Tower, or the Sidney Opera House, are displayed to over 300,000 people each year.

Until then, there were only booklets or portable audio-guides available, now we offer full information including text, images, audios and videos, for all available objects. All audios and videos are available in four languages and visualized in sign-language, depending on the chosen user profile.

The demonstration of our system setup shows multimedia e-learning content which is available to many different end devices such as handheld Windows PocketPCs, Linux iPAQs, Symbian smartphones and high-resolution tablet PCs. Content delivery is performed via WLAN, where the user's location is tracked via Bluetooth beacons and GPS.
Further, we present an easy-to-use HTML-based CMS (content management system) and a statistics system, which informs the administrators about the user interests and behavior.

1. INTRODUCTION

Minimundus [1], an outdoor miniature model theme park in Carinthia/Klagenfurt has over 140 objects on display. Our M3-Systems Research Lab [2] developed a mobile multimedia guide in cooperation with Minimundus, which offers on-demand, user specific information like video, audio, images and stories for all 140 objects (see Figure 1), in four languages and in Austrian sign language (see Figure 2).

Our special focus lies on the seamless adaptation of the presentation and the video content to different end devices, so we support Linux and Windows Mobile iPAQs, Symbian Smartphones, Nokia 770, Sony Playstation Portable, but also "normal" internet access from any home PC, info terminals, tabletPCs and the easy production of give-away DVDs for off-line consumption.

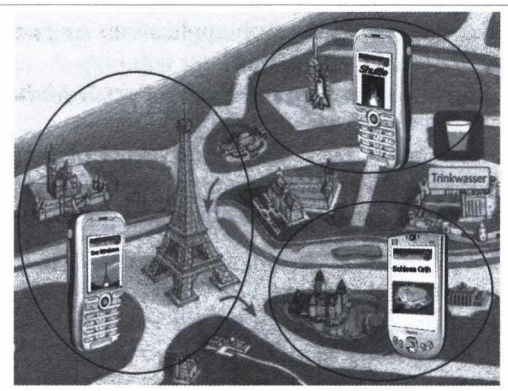


Figure 1: location based multimedia information

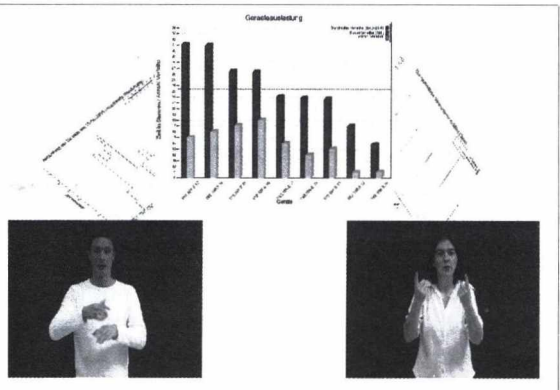


Figure 2: usage statistics and sign language videos

2. TECHNICAL DETAILS

Our system is either connecting via WLAN/UMTS to the different end devices, or the fully adapted content is exported to an offline media, like memory cards or DVDs.

Internally, the content like video, audio, images and text is stored in a file based structure with XML files for detailed metadata information. This is easily managed with our web-based CMS (see Figure 3), and immediately made available to the connected devices (see Figure 4).

To fit the different end devices, the presentation is generated using XSLT, where the platform-independent XML metadata representation is transformed into device-specific HTML. The multimedia content itself is fed to our Transcoding Media Cache (TMC), which adapts eg. incoming video to the exact needs of the connected end device (width, height, frame rate, bandwidth,...).

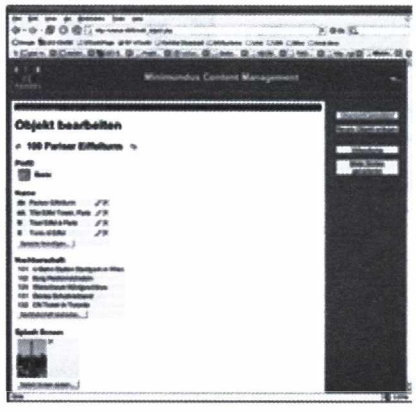


Figure 3: the web-based content management system

Figure 4: some screenshots of different content views

The videos are transcoded, stored in a cache for later re-use, and immediately streamed to the connected end device using HTTP.

So videos can be stored in the CMS in any resolution or framerate in various formats like .mp4, .avi, .flv , .wmv and is transcoded to the exact needs of the device, which may also use another video container like .3gpp or our own streaming-optimized .vit container. The real work is done by various plugins, mainly powered by our own video toolkit ViTooKi [3].

The cache and transcoding status of the TMC is also accessible via a simple web interface as shown in Figure 5.

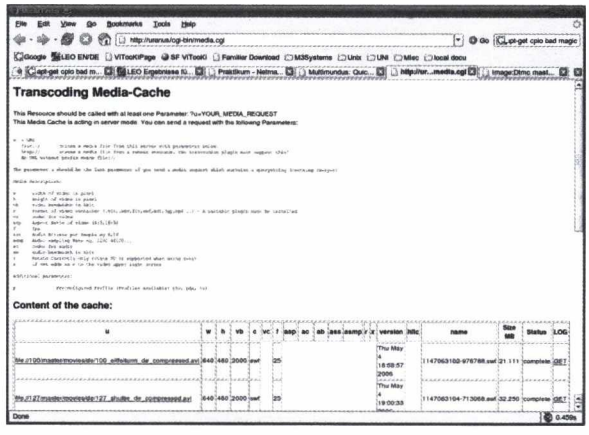


Figure 5: the Transcoding Media Cache (TMC) web frontend

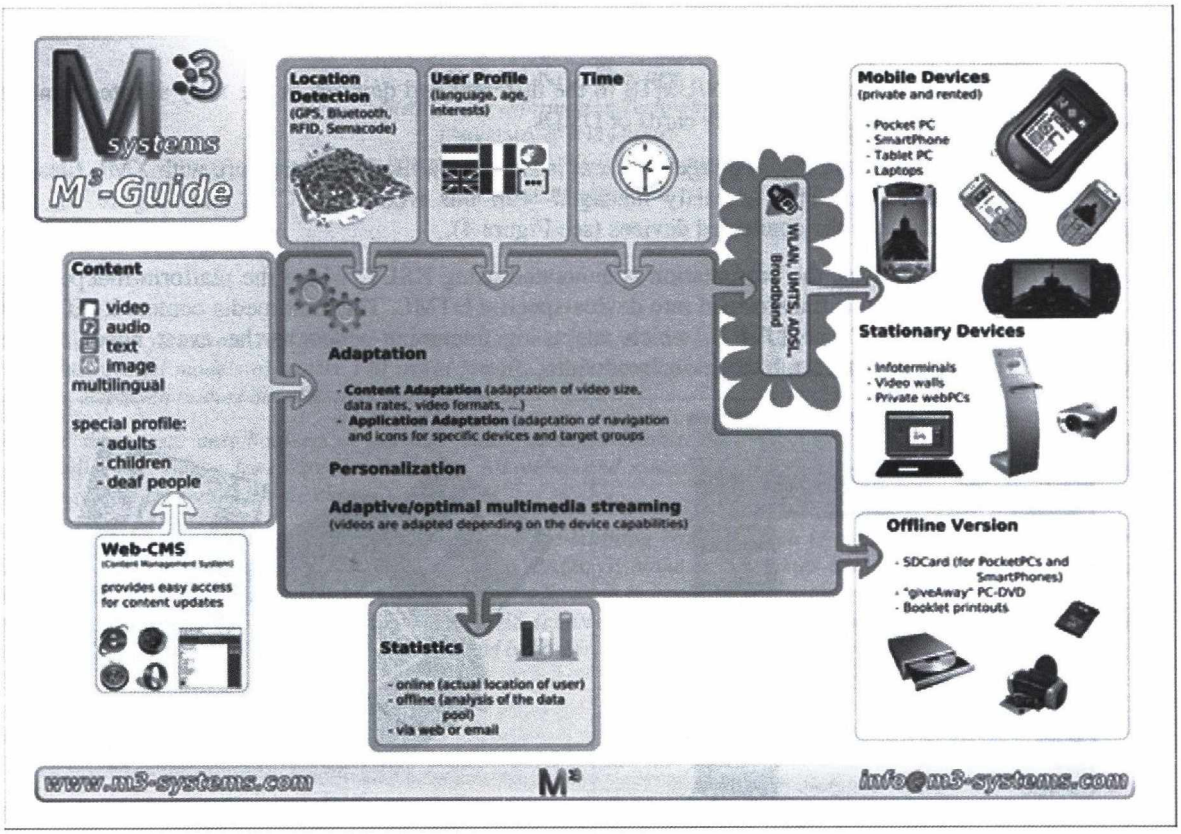


Figure 6: the big picture of the M³-Guide System

Figure 6 gives a complete overview of the M³-Guide System, with all it's various inputs, outputs and computational parts.

3. FUTURE WORK

Right now, we are working on location based technologies like RFID, Bluetooth beacons, GPS navigation, where we detect the user's position and immediately present him/her the closest object of interest. his has to be tested and evaluated on a real user environment this summer.
Further, we have contacted other museums, where we can introduce our system with new features and end devices.

BIBLIOGRAPHY

[1] Minimundus Miniature Themepark, Klagenfurt/Austria, http://www.minimundus.at
[2] M3-Systems Research Lab, http://www.m3-systems.com
[3] ViTooKi, the open-source Video ToolKit, http://vitooki.sourceforge.net

Session 3 – Invited Talk

THE DEVELOPMENT AND IMPLEMENTATION OF E-LEARNING/E-TEACHING STRATEGIES AT UNIVERSITIES AND UNIVERSITIES OF APPLIED SCIENCES

Johann Günther
FH St.Pölten, AT

1. GENERAL TRENDS IN THE EUROPEAN EDUCATION SYSTEM

In recent decades, our world has become global. People's behaviour is becoming similar. The economy sets standards that are the same the world over.
The same trend has now reached the system of education. In this age of liberalism, business and economic thinking rule. Education ministers have coined the term 'employability' and want young people to be educated in such a way that they can be employed in the business world direct and without much additional on-the-job training. Universities[1] have ceased to be the great, free institutes of thought. They have been deprived of their freedom and are now measured on the basis of their output. Their graduates have to be quickly employable in the job market. The Austrian Labour Market Service AMS publishes statistics which specifically show the unemployment rate of graduates of individual higher education institutions. A ranking list of employability has been created. From the point of view of politicians, not the knowledge taught or the final results achieved is crucial but instead the young people's employability.
The increasing internationalisation has led to a unification of education systems. Young people now go to work in foreign countries and on foreign continents. Academic degrees have to be comparable so that Human Resource managers in America know what to make of a European or an Australian university diploma.

Accordingly, efforts are being made to achieve a globally uniform system of tertiary education. The European Union is switching to the Anglo-American system with its 3 stages:

- Baccalaureate
- Master
- PhD.

Study periods, too, are being equalised to match the formula of '3 – 2 – 3':
3 years' Baccalaureate
2 years' Master
3 years' Doctorate (PhD).

The switch to the 3-tier system produces a first layer which has not so far existed in Europe. Education institutes which have not hitherto been called academic now become 'undergraduate'. Thus military academies, pedagogical academies, health and social academies become higher education institutions and offer programmes at Baccalaureate level and over. This leads to an additional number of graduates with a degree from a higher education institution which, however, is not yet fully recognised as 'academic' by the Austrian government.

[1] The term 'university' is meant to denote the traditional university. 'University of applied science' is the term used for *Fachhochschule*, a more recent type of higher-education institution which is characterised by vocationally orientated training.

2. AUSTRIA'S TERTIARY SYSTEM OF EDUCATION

Austria has approximately 8 million inhabitants. 4 million Austrians are working. Unless adjustments are made, the Ageing Society will lead to a diminution in the number of people working:

	2000	2050	
Inhabitants (sum)	8.1	8.2	
3rd Generation (aged 60 and over)	1.7	2.9	
2nd Generation (60 or younger)	5.0	4.2	
1st Generation (15 or younger)	1.4	1.1	
Working people	3.7	3.1	
Gap			1.4

Source: Statistik Österreich
Numbers refer to millions.

Independent of this development, there are now 300,000 people in Austria with an academic degree. This number will rise with the introduction of the Baccalaureate.
300,000 students are shared between the higher education institutions as follows:

Institution	**Students**
Universities	185,438
Universities of Art	7,954
Universities of Applied Sciences	20,591
Pedagogical Academies	10,236
Vocational Teacher-Training Colleges	1,524
Theological Academies	1,111
Agricultural Academies	113
Social Academies	844
Colleges	5,222
Health Academies	2,982
Total	**236,025**

2003/2004
Source: Statistisches Taschenbuch 2004, BMBWK Austrian Ministry of Education, Science and Culture

The universities of applied sciences are growing by 270 newly granted study places every year. But this is only a theoretical figure. The study places approved by the ministry are not actually fully made use of. There is not enough demand for them, particularly in the technological field; this is a trend that can be seen all over western Europe.
In the academic year 2005/06, there were 25,700 students – 2,300 more than the year before. There were 8,200 beginners and 4,200 students had graduated in 2004/05. This meant an increase of 700 beginners and 2,300 graduates compared with the previous year.

The number of students at universities stagnates:

	Winter Term 2000	Winter Term 2003
Austrian	193,649	155,861
Foreign	27,856	29,577

Source: Statistisches Taschenbuch 2004, BMBWK Austrian Ministry of Education, Science and Culture

The percentage of foreign students at Austrian universities of applied sciences is lower:

Austrian	19,721
Foreign	870

Source: Statistisches Taschenbuch 2004, BMBWK Austrian Ministry of Education, Science and Culture

3. RESEARCH

Research is an important prerequisite for teaching. Universities possess a long tradition of research, whereas universities of applied sciences have not yet fully established their research acitivites in the brief period of their existence. Programmes like FH Plus are meant to help them do so. In 2004 Austria spent 2.22% of its GNP on research, which corresponds to a rising trend in this sector:

1998: 1.18
2000: 1.95
2002: 2.19
2004: 2.22 (= €5,273,800)

By international standards, Austria's expenditure on research is above the European and EU average but below that of the OECD countries.
Because of a lack of tax incentives, the proportion of private and public funding in the private business sector is not yet optimal.

	% GNP	Public	Private
Austria	2.07	38.2	41.8
OECD countries	2.29	29.1	63.2
EU	1.93	34.4	55.9

Source: Statistisches Taschenbuch 2004, BMBWK Austrian Ministry of Education, Science and Culture

4. PRIVATISATION AND LIBERALISATION

Due to changing conditions public subsidies are being redefined. The new situation has been brought about by

- the change in the world economy
- globalisation
- Austria's membership in the EU
- increased affluence
- demographic changes, and
- budget cuts by the national government.

At the political level, the aspiration is for Austria to develop into an efficient and service-orientated state that nevertheless practises social inclusion.

Because globalisation is progressing fast, an individual nation state can no longer intervene and regulate by itself. It is inevitable for the countries of the European Union to act together. Political-economic competency has been transferred from the state to the EU.

In the academic year 2005/06, it became apparent that open borders can lead to problems when a huge number of German citizens started studying medicine in Austria. Similarly, Belgian universities count more French than Belgian students in some disciplines. The state of Belgium is paying for the French students' education. The Austrian taxpayers have to bear the costs of the medical training of many Germans. This is where international regulation is required.
The situation becomes even more problematic in the area of distance learning, which has grown with the rise of telecommunications and the Internet.

Politicians regard the education system increasingly from an economic angle. In the tertiary sector, universities of applied sciences were from the start conceived as private enterprises. They are higher education institutions operated as private limited companies – companies dealing not in goods but in education programmes.
This sector was liberalised, too. Anybody could apply for a 'licence'. A supervisory committee watches over the standards laid down by the state.
In the business world, liberalisation has had a long tradition. Already Empress Maria Theresa and her successor Joseph II created an economically favourable situation for the Danube monarchy by liberalising many areas. The subsequent counter-reformations did not diminish this advantage.

The process of liberalisation continued with the licensing of private universities and reached its climax with the University Statute of the year 2000 and the so-called 'autonomy' of the universities.

Thus Austria now has private, public and state universities side by side with each other. However, the system of accreditation and the continual evaluation examines different criteria.

- Universities of applied sciences: individual study programmes are monitored continually; the institution as a whole, once accredited, is no longer evaluated.
- Private universities are accredited as institutions. Study programmes are not accredited or examined individually. The institution is evaluated continuously.
- State universities are not evaluated or accredited.

While a free market economy has been created by law in the tertiary sector of education, in reality the individual areas are judged by different standards and an influence is exerted on the free rules of the market.
State universities are 'autonomous' on the one hand, but on the other hand target agreements between Ministry and university managers determine the academic fields and their scope.
Universities of applied sciences are run as private business enterprises and at the same time subsidised to the tune of a pre-defined number of students per programme. The institution receives a specific sum per student.

5. INTERNATIONALISATION

As our economy is becoming internationalised, so are university studies. National and international organisations promote the exchange of students and lecturers. The introduction of the European Credit Transfer System has led to a rise in mobility and students are now given the chance to do parts of their studies in various countries and at various universities.
Alongside, the exchange of teaching staff is increasing, too.
'Incoming students' add to the internationalisation at home and 'outgoing students' gain international experience abroad.

6. ORIENTATION BY OUTPUT

The economy is increasingly switching over from an input-orientation to an output-orientation. In the days of input-orientation, work was measured on the basis of time. Employees were paid on the basis of their time at the workplace, not for their productivity.
Output-oriented pay means that only actual performance is remunerated. Tasks are laid down in a target agreement and it is no longer relevant how much time it takes to perform them. Fast workers are therefore paid better than slow ones. In intellectual jobs, smart and clever employees are paid better. They need to spend less time on performing a specific task and reaching a specific goal.

A similar change has taken place in the education sector. Teaching has been and still is measured in hours per week ('teaching units'). What counts is the hours a lecturer teaches, 'reads' to the students. The measuring unit is now switching to ECTS (European Credit Transfer System) points. Through this system academic work is recognised across borders and students can get credit for work done in another country or at another university. In addition, the system entails that lectures are no longer defined by teaching time but by the work involved for the student.
An ECTS point expresses the amount of time an average student needs to spend on acquiring a specific teaching unit. This includes the contact time with the lecturer and the time the students needs for self-study and revising.
Four ECTS points can consist of a one-hour lecture for which the student has to do an additional 3 hours of self-study. The four points, however, may equally consist of a four-hour lecture which does not require any additional work by the student.
It is only through ECTS points that e-Learning and distance learning have become measurable. In distance learning ECTS points may not require any contact hours at all.

This output-orientated measurement of teaching brings with it several advantages:

- Internationalisation is made easier; teaching units become transportable and are recognised by all institutions of tertiary education in Europe.
- Students who are fast learners receive more ECTS points for less time expended; poor students must put in more time.
- Distance learning becomes measurable and is fully compatible with 'live' teaching.

7. STRATEGIC PROCESS

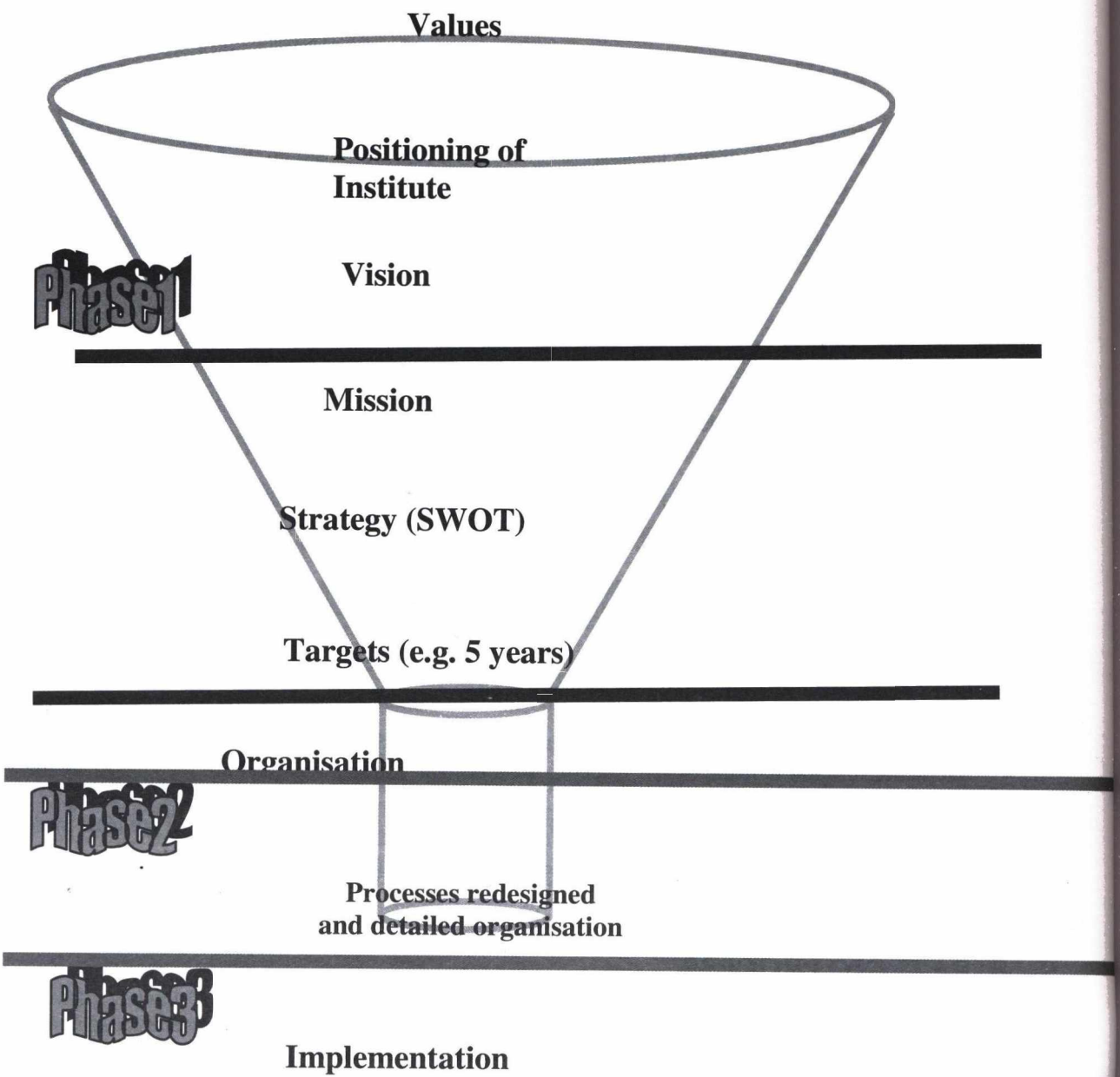
All these changes in the European and Austrian education systems lead to an enhanced economic influence on universities and make them business-orientated. Because higher education institutions are run on economic principles, it becomes necessary for them to develop strategies like enterprises do.

8. PHASES OF CHANGE

The process of developing one's own strategy happens in several phases.

- **Values**
  In the first phase a value system for the enterprise has to be installed – values that govern communication with the students and with other members of staff. Value systems form the basis of co-operation in the educational institution.

- **Positioning the Institute**
  On the basis of its values, the educational institution can position itself on the market and become distinctive.



- **Vision**
  On the basis of the values defined, a vision can be worked out, i.e. an idea or a goal for what one wants to achieve within 5 or 10 years. The vision should be achievable and yet ambitious. The goal should be achievable with substantial effort.
- **Mission**
  The mission expresses how the enterprise sees itself, its *raison d'être* and its goals.
- **Strategy**
  Only after the values have been laid down can a strategic positioning take place. What does one want to achieve? What can be achieved? How can it be achieved? Where do the strengths, the weaknesses lie?

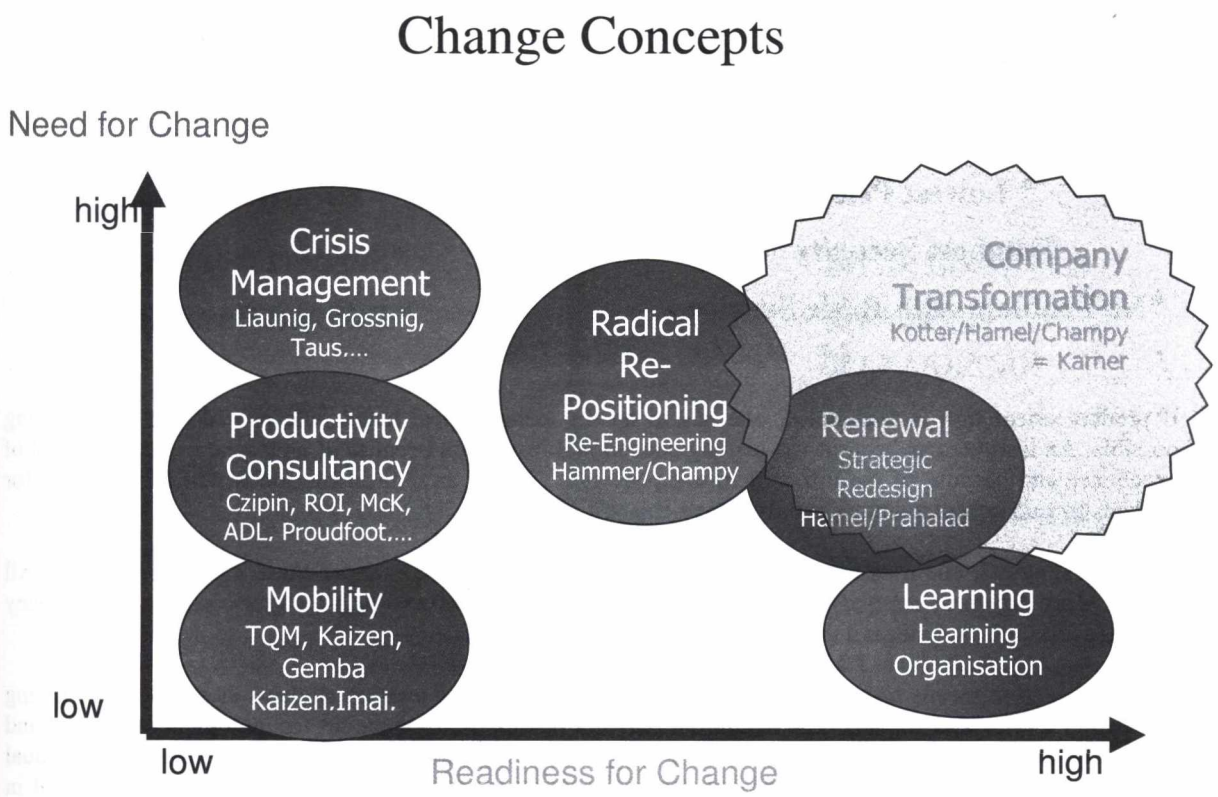
- **Targets**
  Targets are part of the strategy. The vision is expressed in quantitative terms, e.g. the number of students the institution wants to have in five years' time.
- **Organisation**
  In order to reach the targets, an organisation must be defined which fits these requirements.
- **Implementation**
  In the last Phase – Phase 3 – the implementation takes place.

9. WILLINGNESS TO EMBRACE CHANGE

Every strategic process is based on a process of change. A precondition for the implementation is that the organisation is willing to embrace change.
This willingness can be sparked by

- a crisis,
- a necessary consolidation of productivity, or
- mobility arising from the introduction of a quality management system.

The willingness to change depends on the initial situation which leads to its necessity. In a crisis a radical repositioning is required. If only a rise in productivity is sought for, it is sufficient to reform the given situation. Through improvements in the area of quality management and such like, an organisation can change itself.

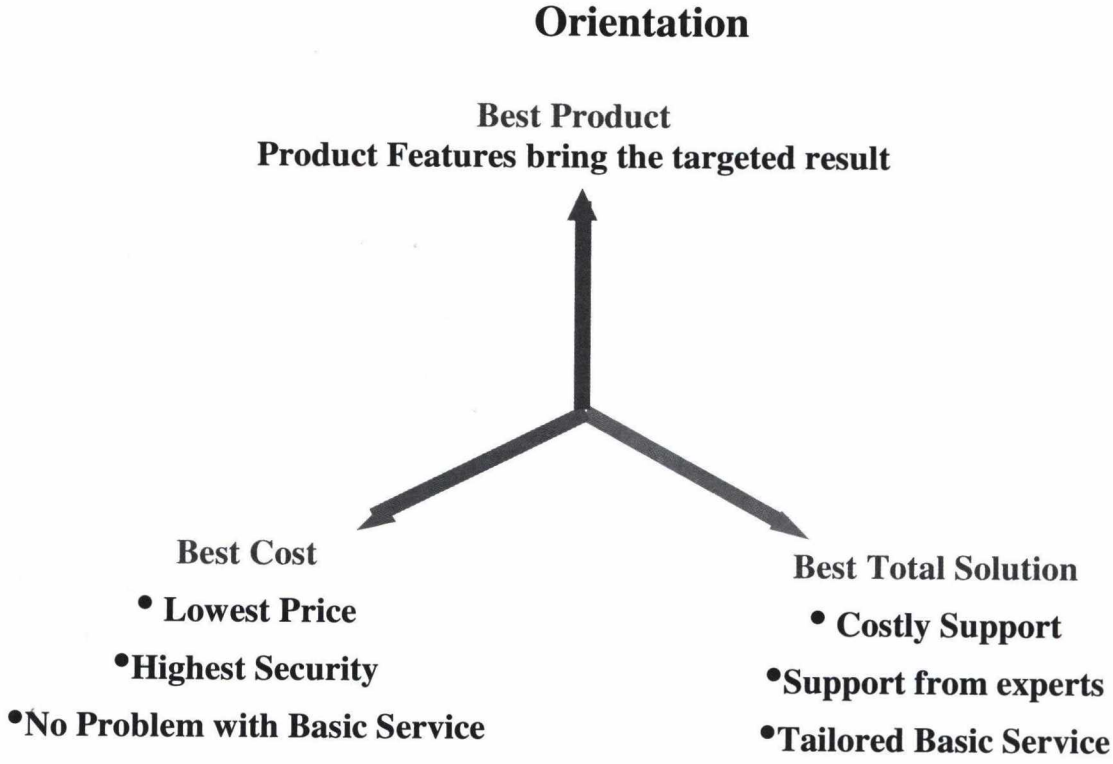


10. POSITIONING

There are three ways of positioning onself:

- through product orientation,
- cost orientation, and
- client orientation.

The educational institution must reach a minimum value in each of these three areas. It should at any rate reach the market average. One area, however, must be foregrounded.



If product orientation is the choice, one has to be the technological leader and provide the best training possible. An institute can become the leading educational institution because of its teachers and their level of excellence and because of the infrastructure it possesses. This is an elite institution which it is difficult for students to gain access to and who only selects the best.

In cost orientation, one tries to produce a maximum of output with low expenditure and few resources. All measures are driven by the question of cost. The result is a mass university. A lecturer costs the same money for 100 as for 500 students and the cost factor is reduced to a fifth.

Client orientation means trying to achieve a maximum of output by treating students as partners and giving them good service and support. The institution may not possess state-of-the-art technological equipment and it may not count the most eminent academics among its teaching staff. But it tries to arrive at high individual results by a good deal of commitment and personal contact to each student. Every student is supported in reaching his or her aim.

The first strategic decision to be taken is about positioning oneself. Among scientific personnel it is common to aspire to product leadership above all things. Everyone wants to be the best in the sector and wants to be

among the leaders. One soon realises, however, that this also requires financial and other resources which are often lacking.

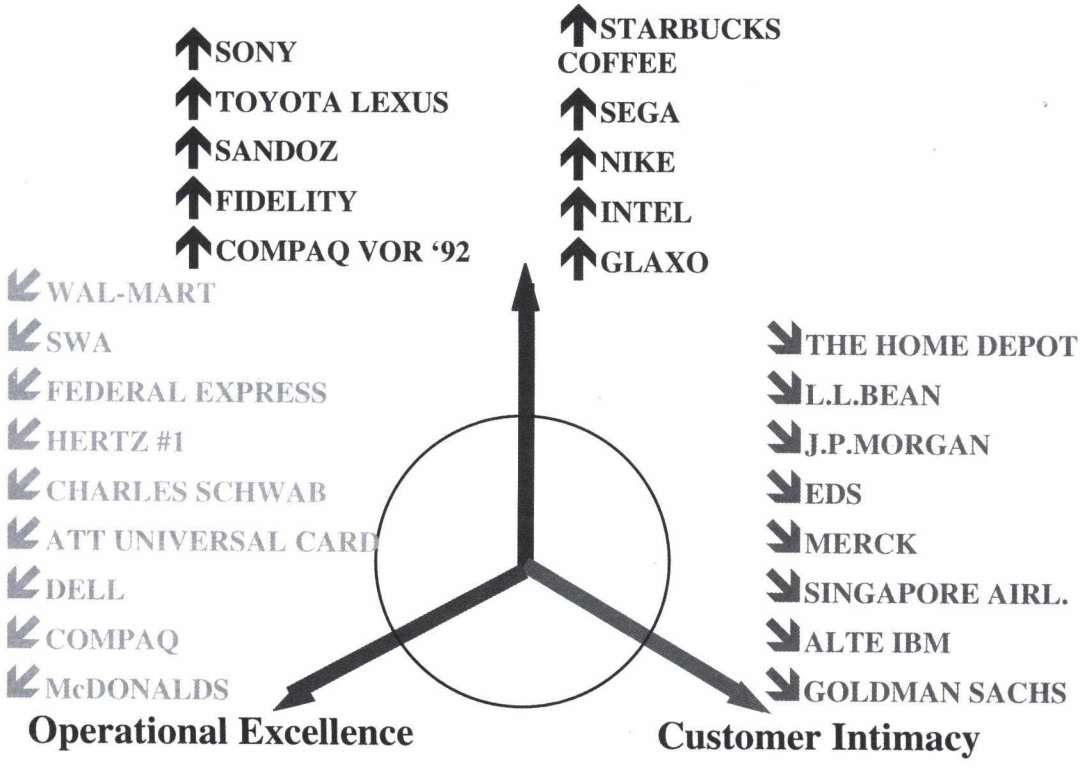
Cost orientation seems the least desirable at first sight as it carries overtones of education 'on the cheap'. Nobody wants to work at a 'cheap' university that is solely or primarily geared towards the cost factor. Such a university would decide even the introduction of academic programmes on the basis of the costs involved in setting them up. Expensive technological programmes would be out of the question because of the costly laboratories required. There are such higher education institutions – even in Austria – but one does not recognise them at first sight. Only after taking a closer look can one find out what they offer and in what manner.

Client orientation requires more personnel. There have to be enough professors and assistants to fulfil the high standards of personal tuition and support. In a client-orientated university, the professors have no office hours and are always approachable for the students.

The following examples from business illustrate the idea of positioning and the resulting message to the outside world:

Examples:

Product Leadership



11. VISION

The vision represents an image of the future in 5 to 10 years' time which depicts in what way the organisation will be different:

- different from today
- different from the others today
- different from the others in the future
- different due to the breaking of rules
- different through re-inventing the sector or at least regenerating the strategy.

Abstract thinking is often missing. Renewal cannot be derived from the past. Exploratory marketing is necessary. Instead of perpetuating the past, a new approach must be found that breaks with existing rules. At the same time one must not forget that other education institutions and other organisations are going through a similar process. It is not enough only to overtake the competitors at the level they are currently at.

12. MEASURES OF IMPLEMENTATION

In the following, an eight-step model of implementation will be presented. This is an example of how vision and strategy can be put into action in an organisation.
It can be assumed that among the staff,

- ⅓ are pioneers who quickly accept innovations and join in the process of implementing them,
- ⅓ are followers, i.e. employees who follow the opinion leaders, and
- only appr. ⅓ are resistant to innovation and cannot accept it. For them one must find ways of integrating them into the new process while they retain their old views.

1st step
Creating Awareness for the Implementation of Change

Staff must be made aware that change is vital. The degree of urgency can be ascertained by analysing the conditions of the market and the competition. Real and potential crises as well as opportunities must be identified so that they can be discussed with the decision makers in the organisation.

2nd step
Forming a Group of Opinion Leaders

Alongside the institution's leadership – rector or executive director – a group of staff members who are willing to embrace change must be formed whose task is to accompany the process and serve as catalysts for the rest of the organisation. Coalitions must be formed so that all the various interest groups in the organisation can align themselves towards the common goal.
The catalyst group must be strong enough to be able to lead the process of change. Its members must be recruited among the pioneers. They accelerate the process of implementation. A staff member unwilling to embrace change would slow down the introductory phase. The members of the leadership group must act as a team and not as solitary agents working alongside each other.

3rd step
Defining the Vision

In order for the change to become goal-orientated, a vision must be created. The vision must be the common goal for all the staff members of the organisation.
As the next step, the vision is carried out with the help of strategies.
The vision is the objective that one wants to reach. Objective-orientated organisations with staff who have visions are more successful than those which lack orientation.

4th step
Publishing the Vision

The vision must be clear and easy to understand. It must be understood by all employees. It must be made known to everybody without exception – equally to external partners. To ensure this, all available means of communication should be used. Since vision and strategy must be understood by everyone, the process of communication is crucial.
The group of opinion leaders are a convenient testing ground for the visions and strategies developed. Their conduct should be observed closely, and the feedback gained might be beneficial for the whole process. Finally the jointly defined vision can be communicated with conviction by all the opinion leaders.

5th step
Living the Vision

The vision is a theoretical construct. It must be put into action in everyday life. Staff must be directed – and empowered – to live the vision.
Once an organisation has opted for an egalitarian, client-orientated mode of working together, authoritarian decisions should be avoided in favour of a democratic decision-making process.
Staff members who do not act according to the joint visions and values have to be called to account just like those whose performance is deficient.
Often a number of obstacles must be overcome before the goal is reached. Not all staff members are prepared to make additional efforts. If somebody undermines the objective he or she needs to be disciplined so that the process of change can proceed unhindered.
The pioneers must be particularly encouraged to introduce unconventional ideas, carry out innovations and take risks.
Of nine things carried out in everyday life six turn out to be wrong with hindsight. Therefore one avoids mistakes by doing less. However, the percentage of wrong decisions remains the same. Processes of change will also naturally involve some wrong decisions and swift corrections.

6th step
Short-term Successes

Short-term successes increase the motivation of the workforce. If successes are in a too distant future, people lose the will to overcome obstacles. Short-term successes are therefore necessary.
Short-term successes must be planned and created just like long-term ones. Staff members who have attained a short-term success should be rewarded publicly. This is a signal to the whole staff that achievement is recognised and encourages them to follow suit.

7th step
Consolidation

The consolidation of the process of change always forms a new basis for further changes. The process of change usually leads to considerable nervousness among the workforce. One can rebuild secure foundations by consolidating at specific intervals. Thanks to their feeling of security staff will regain confidence in the leaders who set the targets. Each consolidation must be followed directly by a new process of change with new interim goals, new projects, new topics and a new agenda of change.

8th step
Institutionalisation

Vision and strategy are implemented fully when all staff members are clearly aware that there is a connection between the success of the enterprise on the one hand and, on the other hand, the action determined by the values, goals set by the vision, and quantitative targets set by strategy.
The opinion leaders' function, however, must go even further. Team members may change, but the vision requires continual redefinition. Objectives must be located in the future and must not slide back into the past. In this sense the process of vision and strategy never ends.

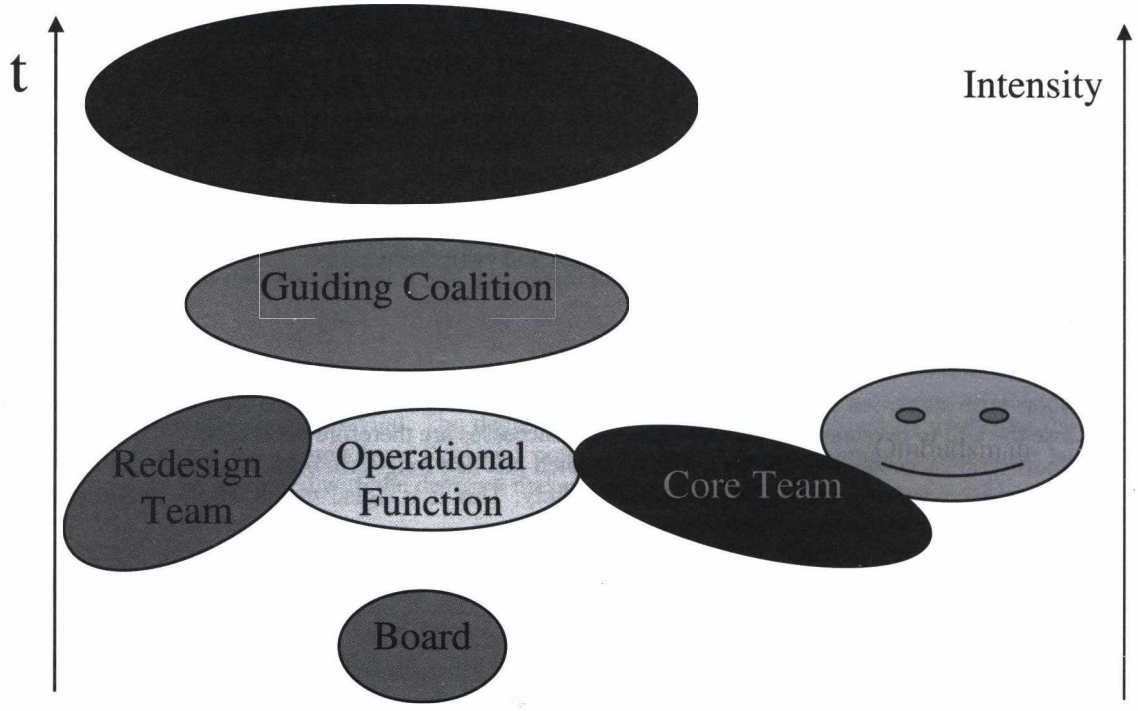
13. PENETRATION

Once the vision and strategy have been institutionalised, it is important to reach a high degree of penetration. Innovations must be shared by many.
The process of penetration proceeds from

- the senior managers, rector or executive board via
- the chosen core team made up of opinion leaders and further down to the
- multipliers, i.e. pioneers who are among the upper third of staff members, until finally it reaches
- all employees.

If the process has been implemented fully, a team of 'redesigners' takes over the task of the core team. All ideas are continually queried and re-examined. The actual work done towards achieving the aims is measured and evaluated and, if necessary, adapted and improved.

The Function of Core Team & Guiding Coalition



14. E-LEARNING STRATEGY

Once an organisation's strategy and vision have been defined they must be adapted by the individual areas. Partial visions and partial strategies develop which, however, are subordinate to the overall direction the organisation wants to take.
This is also necessary in the area of e-Learning. Especially for a new and fast changing area like e-Learning, a modern strategic orientation is required. E-Learning exerts more influence on the overall orientation of an educational institution today than do most other areas. An excellent modern system of accounting is surely important in the running of an enterprise, but modern teaching methods are of the utmost importance in an enterprise whose business is education.

BIBLIOGRAPHY

[1] Anon.: Specific educational measures to promote all forms of giftedness in Europe, Working Document drafted for the Informal Meeting of Education Ministers, 16-17 March 2006 Vienna, 2006

[2] Anon.: Auf dem Weg zu einem europäischen Qualifikationsrahmen für lebenslanges Lernen, Arbeitsunterlage der Kommissionsdienststellen, Brussels 2005

[3] Anon.: ECTS User's Guide European Credit Transfer and Accumulation System and the Diploma Supplement, Brussels 2005

[4] Anon.: Die österreichischen Universitäten im europäischen Hochschulraum, Salzburger Nachtstudio, 1.2.2006, ORF, OE1

[5] BARTELS, Andrew: Trends 2006: Six Economic Shifts Affecting IT, FORRESTERS, 2006

[6] FLUCH, Helga, et al.: Analphabetismus, St. Pölten 2006-04-02

[7] GÜNTHER, Johann: "Der vernetzte Egoist. Telekommunikation und Computer verändern den Menschen", Studienverlag Innsbruck, Vienna, Munich, Bolzano 2004

[8] GÜNTHER, Johann: "Die neue Mobilität der Gesellschaft", Studienverlag Innsbruck, Vienna, Munich, Bolzano 2002

[9] GÜNTHER, Johann: "Videokonferenz in der Lehre", Braumüller Universitätsverlag, Vienna 2001

[10] GÜNTHER, Johann: "VEN Virtual Education Network", Krems 2000

Session 4 – E-learning Experiences and Security Issues

AMEISE
AN INTERACTIVE ENVIRONMENT TO ACQUIRE PROJECT-MANAGEMENT EXPERIENCE

Roland Mittermeir[1)], Andreas Bollin[1)], Elke Hochmüller[2)], Susanne Jäger[1)], Daniel Wakounig[1)]

[1)] *Institut für Informatik-Systeme*
Alpen-Adria Universität Klagenfurt
Klagenfurt, Österreich
{roland,andi, susi, daniel}
@isys.uni-klu.ac.at

[2)] *Studiengang Telematik/Netzwerktechnik*
Fachhochschule Technikum Kärnten
Klagenfurt, Österreich
E.Hochmueller@cti.ac.at

Abstract

This paper reports on experience with AMEISE, a simulation environment for the management of software engineering projects. AMEISE allows trainees to act as project managers in a virtual development situation.
The system keeps track of every decision and of all relevant simulation states. This allows for automatic as well as discursive assessment of the students performance as project managers.
The paper presents the motivation and didactical considerations of developing and using such a system as well as the experiences gained by using it in various courses during the last years.

1. MOTIVATION

Learning how to manage a Software Engineering Project cannot be achieved by only reading books or listening to lectures. It requires a combination of theoretical knowledge and practical skills. While the former can be acquired in practical lectures, practical skills are not to be imparted that easily. Notably, students have not the chance to revoke decisions that are considered problematic at a later time of their project and failures usually involve high social cost. Though there are well known best-practices (e.g. [Hump 98]), content presented in lectures quite often suffers from credibility problems, even when enriched with empirical data. This deficiency has many roots, but mainly it is due to the fact, that such lectures are based on data from other people's projects – projects that are considered to be different anyway.

Gaining practical experience is, for sure, inevitable. But experience gained in lab situations is usually characterized by a lot of guidance. To be realistic, students should be entrusted far more responsibility. Consequently, they should be allowed to make a complete mess of the project, to try to save it from decline, and to learn from past mistakes. In general it means to allow for a trial and error approach, which might lead to very time-consuming courses [Hoch 02]; and, who dares to conduct real-world projects with students and to let them fail? One way-out is simulation.

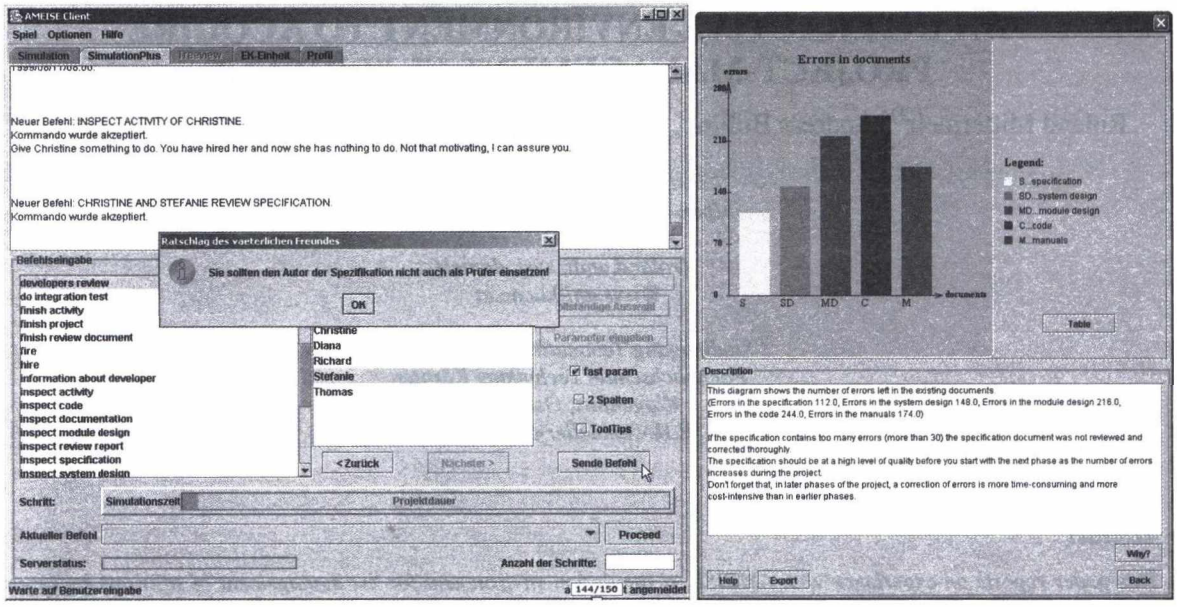


Fig. 1: The AMEISE System contains several support and assessment components.

In this paper we present AMEISE, a simulation environment that allows students to gain first experience in managing a software engineering project (with the focus on software quality). AMEISE, *A Media Education Initiative for Software Engineering*[1], has been developed by a consortium of three Austrian institutions of higher education: the Alps-Adriatic University Klagenfurt, the Carinthia Tech Institute, and Linz University[2]. AMEISE uses a simulation engine called SESAM, *Software Engineering Simulated by Animated Models*[3], developed under the direction of Jochen Ludewig at the University of Stuttgart [DL 00]. The AMEISE team adopted the educational model proposed by the team at Stuttgart and also used the SESAM system as initial prototype that was evolved in a series of iterations to the currently available AMEISE-system [MHBJN 03, Mitt 06] and it is still subject to further evolution.

The paper is structured as follows: Section 2 discusses the educational model behind AMEISE and the related extensions made to the SESAM system. Section 3 addresses the specific challenges related to an interface providing student-managers with comprehensive information about the state of their project. Section 4 presents the experience gained during AMEISE courses – be it in classroom situations or in blended-learning settings with simulation runs totally unattended by tutors. The paper closes with an outlook, planned further improvements of the system, and upcoming tests of the didactical concepts.

2. AMEISE AND EDUCATIONAL CHALLENGES

The approach AMEISE follows is distinct from the mainstream of eLearning approaches with pre-canned (multimodal and multi-medial) content in so far, as the systems knowledge base is just a complex set of rules that are not directly accessible to students. The content to learn from is "produced" interactively by each

[1]) AMEISE web-site: http://ameise.uni-klu.ac.at

[2]) Development of AMEISE has been supported by the Austrian Ministry of Science and Education, bm:bwk, in their eLearning support initiative „New Media in Tertiary Education", NML. Web-site: www.nml.at.

[3]) SESAM web-site: http://www.iste.uni-stuttgart.de/se/research/sesam/

student's assignments of work to various members of her or his development team. Thus, the "content" that serves as basis for learning is created on the fly by the student's activities acting as technical manager. The challenge for the system (and its developers) is to readily analyze this content and present it in the most suitable form to both, the student who produced it as well as the tutor, who should give further in-depth interpretations and hints.

While the system's interface grew from an initial pseudo-natural language textual interface to something closer to the state of the art (see Fig. 1, left side), we would not claim it to be multi-medial in a technical sense. It is rather multimodal, as the blended learning approach foresees that the results of the simulation run is discussed in group as well as in private with the students of a given course. With the feedback thus obtained (see Fig. 1, right side) students are ready to return to the system and strive for managing a second, hopefully improved, software engineering project.

2.1. Educational Aim and Description of the Task to Solve

The basic educational model is founded on experimental learning. This necessarily implies also to some extent the characteristics of learning from one's own failures. But to make sure that this is not de-motivating, our approach as well as the approach proposed by the developers of the SESAM-system, which has a parental role to AMEISE, recommend that students perform at least two simulation runs. In both, they should strive to obtain their best possible performance. However, usually, in the first one, a number of issues occur that give rise for discussion and improvement. In the second run, students can benefit from the experience gained. Thus, in general, they achieve more of the criteria marking success of the project or at least improve on the respective success dimension.

Concerning success dimensions, one should mention that steering a software development project to success implies to reach several, to some extent contradictory success criteria under a set of defined constraints. To mention just some constraints, projects use employees having different qualifications in different technical (and social) aspects. These (potential) employees have different salary expectations. Their qualifications result in different speed and, being (simulated) humans, to different error rates. Another important constraint results from the technical nature of software projects. Irrespective of the approach followed, there is a logical order in which some activities are to be performed. However, there is also a lot of freedom of scheduling or even skipping certain activities. Hence, student-managers have to decide, not only whom to hire (and release, if no longer needed) but also when to ask a person to perform which task. At this point, it should be mentioned that the virtual employees are keen in doing exactly what they are asked to do (unless already otherwise busy), but they are in no way motivated to overrule their managers orders if those don't make sense. This might result in their attempting an infeasible task and to give up after a day of unsuccessful attempts brief termination message. Before receiving a new order, they will lean back and goof.

Such situations, obviously, bear negatively on the student-managers performance, since projects have obviously budgetary constraints and an agreed upon fixed delivery date. At this date, the project (both software and documentation) should be handed over to the customer at a specified minimum level of quality (degree of completeness and acceptable error rate). Thus, the key criteria of success is whether the project was d*elivered on time*, *within budget*, how *complete* were the delivered *software* and how complete was the delivered *documentation* as well as how many *errors* were found in the delivered *software* and in the delivered *documentation* respectively.

To arrive at these results, the simulation model rests on an empirically rich set of rules that are invoked depending on the commands the student-manager issues on a particular day of work. The model we are using focuses besides on general project management issues particularly on quality aspects. Students should learn among other things that quality cannot be "tested into a product". Quality has to be taken care of right from the beginning of the project and they have to experience that such dull things as technical reviews – hardly ever practiced in educational settings – are valuable in real and sizable projects.

Students also have to experience the merits of planning. A person basically works only on a single task per time instance. Split attention (if at all possible) causes less productivity. Teamwork is necessary to meet time-

and quality-constraints. But teamwork causes communication and communication causes overhead. Thus, one should not be too liberal in striving towards a million-monkey approach. Scheduling and planning is also important when assigning certain tasks to be performed on a semi-finished piece of work to team members. A colleague of mine will be more productive in identifying faults, I have committed. But I should be more productive in making the necessary corrections. A third person might even introduce more new faults by correcting those previously identified.

Considering these and similar constraints, one can see that even with a relatively simple project (200 adjusted function points of effort to be expended over a development period of nine months) causes those, who do not adequately plan, update their plan when they cannot follow it, and keeping track of the assignments they made throughout the development period will towards the end, when money and time become scarce, run quite into trouble. Thus, enough material is produced that can be discussed in the feedback period.

2.2. Educational Setting

We used the AMEISE simulation system mostly in the context of general software engineering courses. Hence, supplementary development goals of the AMEISE extensions to the original SESAM system were to allow its use in larger classes. Thus an evaluation support for the tutor became necessary and mechanisms for self-evaluation by students were also deemed desirable. Only in exceptional cases, a special course was defined where we used the system. These special courses (electives) served besides their educational purpose also the role of didactical test environment to explore the proper use of certain new features.

The standard setup of AMEISE simulations follows a general pattern. It usually consists of the following phases:

- *Preparation:* It is highly advisable that the particular simulation model in use is introduced some days ahead of the first simulation. The type and depth of this introduction varied over the different courses depending on project-management related aspects dealt with in the core of the respective course. Sometimes it was confined to just minimal verbal hints and handing over a short user's guide. But even in these cases, it takes about an academic hour (45 minutes) to present features and idiosyncrasies of the system and to make students adequately aware that a simulated project is just a simulation. Hence, in real projects, they would have to control even further aspects. Till getting in contact with the real system, students can familiarize themselves by means of a flash-animated web-tutorial.
  The participants of recent courses got an AMEISE handout with the most useful hints and/or a list of the virtual software developers including their hourly rates of wages and a rating of their working experience with several facets of software engineering. Giving them this list saves the time of the interviewing phase before hiring and taking respective notes that would otherwise be necessary (or at least strongly advisable).
- *Tool and Syntax Explanation:* The final contact with the AMEISE environment usually takes place right before the first simulation. Sometimes the participants can even get acquainted to the environment by running a rather simplified simulation using the so-called mini-QA model (reduced size of project, reduced number of available developers with identical qualifications). Allowing students to play around in the context of the real project they have to pursue would also be an option to familiarize them with system specifics. However, we never used the latter option.
- *Planning the first simulation:* The participants are encouraged to prepare their first simulation such that they plan the prospective allocation of virtual software developers to the respective phases and activities within the software development process. How much effort goes into this step is up to the respective student (or group of students). If not prepared ahead of time, planning is part of the respective simulation session itself.
- *First simulation:* Usually, the first simulation will be processed in the classroom in the presence of the instructor. Thus, any ad-hoc questions which may arise during the simulation run can be addressed on the spot.
  Technically, this need not be the case. However our experience shows that with students as well as with people having already industrial project experience, enough questions pop up to warrant the presence of a tutor in the class room.

- *Feedback for first simulation:* Basically, there are four different kinds of feedback possibilities which can also be arbitrarily combined:
    - onLine assessment in the presence of the tutor interpreting and explaining the results provided by the evaluation component
    - self-directed onLine assessment in the absence of the tutor
    - feedback from the tutor in a plenary session with group discussion and the possibility to compare interesting aspects.
    - delivery of a generated evaluation report for each participant.

  We recommend to include at least after the first simulation run the group discussion approach in the options of feedback approaches. It allows students not only to see effects of management decisions by their peers. It allows also the instructor to highlight the difference between effects due to the very nature of a software development task from effects that might be attributed to the specific way certain aspects are represented in the model driving the simulation.
- *Planning the second simulation*: For didactical reasons, a second simulation run is absolutely necessary. It will be a valuable opportunity for the participants to show that they can improve their simulation results. Because of the motivation to achieve better results, the preparation time will slightly increase in comparison to that of the first planning period.
- *Second simulation:* As initial difficulties with the AMEISE environment were already overcome during the first simulation run, the presence of the tutor is not necessarily required for the second run. Usually, the time spent on running the second simulation is less than the time used for the first one.
- *Feedback for second simulation:* The kinds of feedback possibilities basically correspond to those indicated before. If an in-class discussion takes place, not only the results of the second simulation run will be analysed and discussed but also the deviations with respect to the first simulation. The simulation environment currently takes the two simulation runs as distinct simulations though.

In an overall assessment of the approach, one has to see that students need not only feedback on their performance after the simulation runs but to feel comfortable with the system. The amount of information provided throughout their acting as project manager is also important. Here we don't discuss it as feedback but rather as usability properties of the simulation.

One should state already here that this is aside form the general positive appraisal of the system a constant source of critique. Students want more and "better" information throughout their work. However, in this respect we are rather stubborn. The manager should know what individual members of her or his team are doing at a given time. If not, they have to be asked and in this case, the system gives instantly an honest reply. When asked how far they have proceeded with their work, their answer is biased though by a random fuzzyfication. This is realistic to the extent that different humans would also differ in their optimism/pessimism in assessing the work still to be done. Finally, one should mention that the kind of information provided by the system is realistic indeed. Managers won't walk into an office and inspect a software artefact to assess when a person will be done and available again. They will ask and get some estimate.

Nevertheless, run-time feedback is an issue where we still explore new options to provide student managers with an even more realistic picture (see section 3.4).

3. 3. ASSESSMENT AND EVALUATION COMPONENTS

As didactical aims may vary, instructors may customize the system according to their specific didactical aims. According to our experience, it is rather advisable to use the system in a rather plain fashion during the first run. Only the evaluation components should be available. Whether the support components discussed below are made available is a matter of taste (and to a moderate extent also of the performance needed). Other features such as the group comparison or the rollback facility are not recommended for the initial runs,

since students might be lured in just playing around and loosing track of the key aim to get familiar with the basic challenges and mayor pitfalls of steering a software development project.

Another alternative to use right from the beginning would be to contract certain features out to an *external software house*. This allows for focusing just on specific development phases while certain tasks are simulated in a reasonable but neutralized manner.

The system has been designed for blended learning situations. Hence, one major difference to SESAM is the support for accompanying feedback during simulation runs as well as support for immediate feedback right after the end of a simulation. Thus students get their first feedback while their memory is still fresh. AMEISE allows for increased learning experience. For this, the SESAM simulator has been extended by several support components. Some of them are integrated into the user interface and are accessible by the students; some of them are for instructors only. Depending on their use they can be categorized as support components, self-assessment, and assessment tools. This section describes them briefly.

3.1. Support Components

As discussed earlier in this paper, project management implies to act on one's own authority. However, typically one is not really alone and there are colleagues around that might be asked for advice. Having friends around, one might also get advices when not explicitly asking for them. AMEISE meets these requirements by providing both, information sources that get active on demand and agents running in the background.

Providing good advice in situations depending on the students past activities is challenging. It requires the capability of recognizing the situation the student has manoeuvred him- or herself into at this very moment (the state) and furthermore to know about past and (possible) future effects of decisions. Technically this problem is solved by two agents, whose responsibility is to store relevant states and special markers (e.g. reaching of milestones) in the database.

The *State Manager* is the agent responsible for determining and storing state information. A state consists of entities and relations in the simulation engine. Hence, the State Manager directly observes internal SESAM variables and places respective markers in the AMEISE database[4]. Management of these markers depends on specific actions and command of the trainees. Thus the storage is done by an agent (called AMEISE *marking agent*) running in the background of the trainees' client. This agent recognizes (a pre-defined set of) activities and state changes. They are defined by the model-designer and stored as a set of SQL queries in the database. These queries can be simple (representing only queries to match specific user commands), but they might also contain complex relationships between entities in the state space [Nuss 03].
The agent's activities are triggered by state changes in the database. Every state change in the engine results in trying to find suitable SQL queries that are then to be executed. E.g., the set of queries might contain an entry that looks for the beginning of a specific project phase: the specification. Thus it prescribes to listen to a command called "<developer> write the specification", and to make sure that the state space then also contains a <developer> that is working on the system's specification. In exactly that case the marking agent then assigns a label entitled "BEGIN_SPECIFICATION" to this state in the database. The marking agent plays also a special role when comparing simulation runs with each other returning to previous states of the project.

The agents discussed so far are not visible to the user of the system. But there are two support components that are visible. The first one is called AMEISE *Friendly Peer*. It makes use of the information stored in the database. It recognizes markers, the state, and state changes. By chains of rules based on Boolean logic even complex situations and conditions can be covered [Jaeg 03]. These expressions (including relevant explanation text) are stored as SQL templates in the model-relevant part of the database. They are composed on the fly during the simulation run. When made available by the instructor, the friendly peer might, e.g.,

[4]) For a detailed discussion of the AMEISE architecture, see [MHBJN 03] or the AMEISE site at http://ameise-ni-klu.ac.at

detect that the student assigned the least qualified developer to the specification task. A small window then pops up and tells the student about the importance of the specification phase and recommends changing the developer.

In contrast to the friendly peer, *advisor* is a reactive support component. It approximates an experienced colleague who might be asked for guidance and advice whenever the student manager feels at loss. The advisor component makes use of the state information and markers kept in the portion of database that keeps record of each individual simulation run. When called, it provides the user a list of questions (predefined in the model specific portion of the database) that might be asked. As an example the student has the possibility to ask whether the choice of the developer writing the specification was a good one or not.

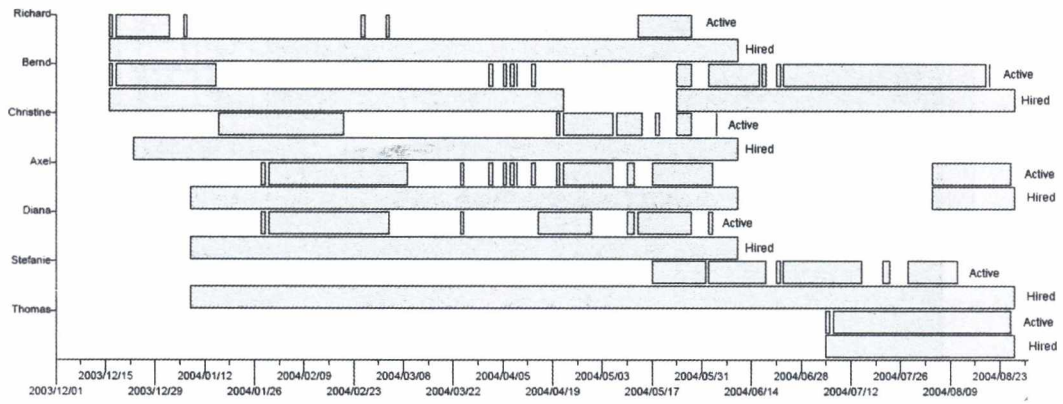


Fig. 2: The AMEISE self-assessment component provides several diagrams that can be generated by the user. Here the deployment of developers is presented (were the bars display the deployment and actual activities of several employees).

3.2. Self-Assessment Component

At the end of the simulation immediate feedback is important. The self-assessment component (also called AMEISE *evaluation component*) fulfils this requirement. When activated by the user at the end of a simulation[5] it generates diagrams and tables one would typically find in an AMEISE assessment report. Fig. 1 and 2 present two of such diagrams. On the right side of Fig. 1 the propagation of errors in documents across various phases is presented. Fig. 2 shows the timeline for each engaged developer while hired (lower bar) and while being active (upper bar). This particular project was not really successful as developers, while being on the project's payroll had lot of time where they were not active.

The evaluation component aggregates data stored in the student/simulation-run specific portion of database. As with the advisor component, the logic (including the explanation text) behind the diagrams is stored as SQL query templates in the model specific part of the database. The component itself is well-suited for a first assessment and provides motivating feedback to the student. However, the certain effects are far too complex for automatic evaluation. Therefore, the report cannot replace the detailed assessment report generated by the tutor of the simulation run and the ongoing discussions.

3.3. Assessment Support

Self-explanatory diagrams are necessary, but assessing a project trace needs more than a handful of Gantt charts. As every simulation run is different and might contain different situations of interest, it is up to the tutor identifying those parts of the simulation first and then aggregating the data for presenting them in a

[5]) The full power of the evaluation component can only be used after the simulation has been terminated. Limited intermediate evaluations are possible, though, at any time.

suitable and conveying manner. Here, AMEISE provides a tool called AORTA (short for *Analysis Of Relevant Tree Aspects*).

As mentioned above, AMEISE makes it possible to go several steps back in the simulation and to try different strategies. The game trace can then be seen as a tree of paths through the

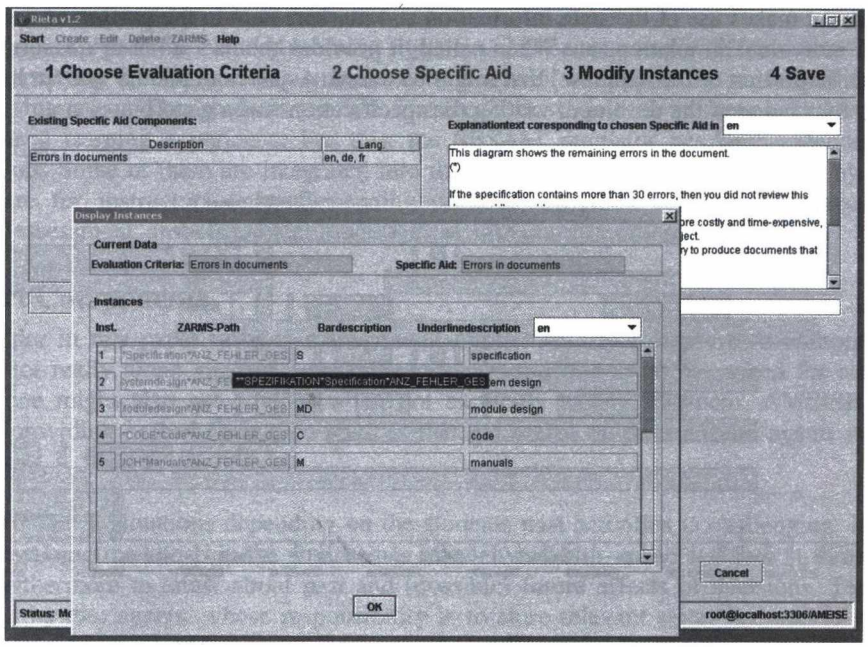


Fig. 3: The RIETA tool is used to add functionality to helper-components. Here, the diagram of Fig. 1 (right side) is constructed by adding descriptive text and providing attributes to be displayed in the diagram (e.g. the specification's number of errors).

simulation. AORTA focuses on this game tree and is able to display all diagrams and tables of the self-assessment component in respect to these paths. To sustain the tutor the component offers more functionality: it allows for tracing the complete simulation run, comparing several runs with each other, and exporting statistical data in comma separated format (for later external analysis). In addition to that it is also able to export all parts of the assessment report in PDF format, which eases the distribution of the result among the participants of the course.

In the above discussion, only the users' view onto the assessment has been described. At least as important is the backend: the tool that enables adding new queries (and thus functionality) to the supportive, model-dependent components. AMEISE provides a tool called *RIETA* (short for *Rule Insertion and Editing Tool for AMEISE*). It allows to first define so-called "specific aids", which are either references to attributes in the simulation state (e.g. actual costs of the project, or number of errors within a specific document), or the logical combination of and comparisons between them. Instances of these aids can then be combined in order to form diagrams or tables. Fig. 3 shows that these aids can be used to display the bar chart (errors in documents) and can be annotated with some explanatory text.

3.4. Challenges

Immediate Feedback and detailed assessment reports have turned out to be pedagogically valuable. But there are still some challenges. From the students' point of view the feedback is sufficient, but still ever-recurring effects (e.g. productivity loss) have to be explained on simulation basis by the tutor – every simulation run is different. The assessment report contains all necessary information, but the exploration of the relationship within the data needs guidance. At least some of the effort and time spent by discussing a single simulation

run can be saved by a more personalized assessment report containing references and context-dependent feedback.

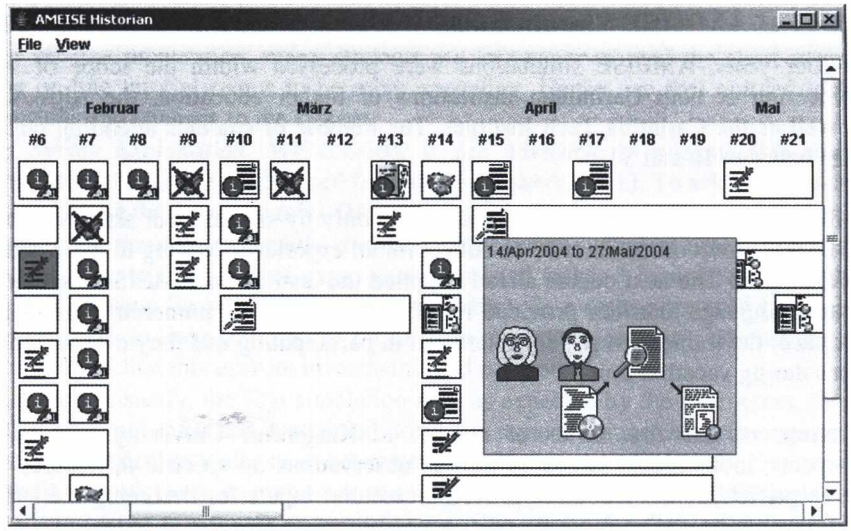


Fig. 4: The assessment of the approach is sustained by a framework presenting the project trace in an iconic manner.

From the tutors' point of view a fast detection of relevant effects (students should learn from) is more important. Scanning the report and looking for effects takes time as it has to be done by hand. Up to now the Historian tool is only able to mark un-successful commands of the user. However, the detection of relevant effects is much more difficult. A typical example: within an assessment report it does not make sense to state that it would have been wise to hire a more experienced developer when none was available at that time of the simulation. Knowledge about the specific context at this point of the simulation is required, knowledge that first has to be stored in order to be incorporated into report-generating tools.

Both requirements are part of ongoing extensions. The assessment report is going to be enriched by references to related data (within the report itself and to related literature), and one thesis is concerned with the identification of the necessary context. In a first step the context information is then to be used to personalize the assessment report. In a second step a focus mechanism will be integrated into AORTA and the context information will be used to ease the highlighting of interesting areas in a simulation trace.

Whereas AORTA is well suited for the generation of reports, up to now the identification of relevant decisions is left to the tutor. To identify critical regions in the trace, tutors are typically looking for failed activities and analyse the trace by looking at the plausibility of the sequence of commands. Both tasks are time consuming. To ease this situation, a component called *Historian* has been introduced. It displays the simulation trace in a graphical/iconic manner, and is able to mark failed activities [Putz 05]. Fig. 4 displays such a typical game trace, and the activity marked in red tells the tutor that it has not been successful.

4. EXPERIENCE AND EFFECTS

As already stated, AMEISE uses the simulation engine of SESAM as provided and still maintained by their original developers [DL 00]. Hence, test simulations could be carried out right from the beginning of the AMEISE project. This allows us to continuously gain experience (like stated in section 3.4) which in turn influenced the overall progress of AMEISE developments. Reporting in detail on the project management results achieved by the students would go far beyond the scope of this paper. Instead, after giving an overview of different course settings and stating some critical aspects related to the standard setup of

AMEISE simulations (cf. section 2.2), this section summarizes some noteworthy experience gained and effects observed during various AMEISE courses.

4.1. Settings of AMEISE Simulations

Over the last four years, AMEISE simulations were processed within the scope of several courses on Software Engineering at both Carinthian institutions of higher education, the Alps-Adreatic University Klagenfurt as well as the Carinthia Tech Institute. The number of students attending our AMEISE courses was in the range between 10 and 54.

The first and foremost course in 2002 was attended not only by students but also by lecturers. This course still used the SESAM environment in order to collect initial experience leading to more detailed requirements on the AMEISE project. The next course already applied the distributed AMEISE architecture but still used the pseudo natural language interface provided by SESAM. In spite of numerous complaints about the text-based user-interface, the trainees were very interested in participating and they even volunteered for running their simulations during vacation time.

The general nature of this first series of courses at Klagenfurt University was rather experimental. Performance aspects, motivational issues as well as observations on specific difficulties influenced further developments. In particular, these courses provided essential inputs for designing the support components. The urgency of developing a new, more supportive user interface (see Fig. 1 left-hand side) was recognized, too.

Since 2003, AMEISE has been used in courses on a regular basis. Generally, the standard setup as described above will be applied. In some courses it is up to the students to decide about a possible second simulation run. However, in most cases the students are eager to show that they can improve their results or even beat the results of their colleagues.

At Klagenfurt University, AMEISE perfectly fits the purpose of the compulsory course "System Development Process" in the graduate curriculum of "Applied Informatics". Having gained knowledge from preceding courses on Software Engineering and experience from the internship semester in industry, students attending this course will learn about improving the CMM maturity level of software companies [Hump 98] in order to be able to discus Humphrey's recommendations in the light of their own industrial experience and map them onto small and medium sized companies. The number of course participants usually varies between ten and twenty persons. Undergraduate students have the possibility to use AMEISE in the two courses "Management of Software Projects" and "IT Project Management and Change", the latter is part of the curriculum "Information Management".

At the Carinthia Tech Institute, former undergraduate courses on Software Engineering included the development of a small software system in order to be able to get a deeper understanding of and experience in the software development process [Hoch 02]. The excessive effort for both the students and the instructors was enough of a reason to abandon this project-centred educational model in favour of AMEISE simulations. Hence, AMEISE is used in the undergraduate courses "Software Engineering" of the curricula "Telematics / Network Engineering" and "Medical Information Technology".

A completely different setting was an in-house AMEISE course at a software-developing company. A quality manager and a project manager initiated an AMEISE course for their managers of software projects. Some of the eleven participants already possessed some years of experience in project management. The course started with a half-day lecture on quantitative aspects in project management including Function Point Analysis and COCOMO (the participants already knew about FPA and some of them also about COCOMO II). The subsequent part on AMEISE again followed the general pattern as given above.

Besides this, we made AMEISE also available to other educational institutions[6]. Some used similar approaches as recommended in Sec. 2.2, others departed from these recommendations. We will discus our experience based on the standard setup described above.

- *Preparation:* It is crucial that this phase is seriously considered. However, in spite of insisting on the respective off-line preparation, many students are less eager to read the short manual we pass to them. The on-line tutorial seems to be more appealing. Hence, we feel obliged to keep it current with extensions and changes to the system.
- *Tool and Syntax Explanation:* We consider it still advisable to relegate this phase to the first simulation setting (but keep it separated from this by a short break). To ask students doing it on their own, e.g. by having them run the mini-QA-model, might not be the best use of students time.
- *Planning the first simulation:* Usually students do not invest sufficient time and effort into this task. This is mostly due to their lack of experience in related tasks. A typical students planning time is less than half an hour (in some cases, it is negligible). In a course we held for practitioners, they used between half an hour (minimum) up to little more than an hour. The results of their (first!) simulation showed that this upfront investment paid off well.
- *First simulation:* Usually, the first simulation runs as expected by the instructors. Plans that existed initially that all simulations are unattended done in a remote manner have been given up. There are still too many technical as well as subject-related questions to be addressed.
  In order to fit all students to workstations in a single room, as well as for didactical reasons, we have them usually working in pairs or groups of three. This allows for discussions within the team. We consider these didactically valuable.
- *Feedback:* As a standard feedback procedure, we use all of the following options in our AMEISE courses:
  - immediate onLine assessment in the presence of the tutor interpreting and explaining the results provided by the evaluation component after the first run as well as self-contained onLine assessment in the absence of the tutor after the second run.
  - feedback from the tutor in a plenary session in depth discussion after each simulation.
  - each participant will get a generated evaluation report describing the simulation results.

  Among these feedback possibilities, we consider the discussions as most valuable. The delivery of the written report thereafter is of course a necessity. Experience gained in simulation runs (approx. 12 groups) shows that it takes about 1 hour to produce the assessment report. It then takes about 3 hours to scan the 12 reports, and about 4 hours to produce some PowerPoint slides summarizing the main results for later use in the lecture.
- *Second simulation:* We always offer this option and most students benefit from this option. They do it even, if the schedule permits only to do it during vacation time.
- *Further simulations:* As the system is used in several courses, possibilities exist that some students in the compulsory course "System Development Process" had already worked with AMEISE in other courses before. We ask those, to set themselves a specific further goal such as "use minimal number of developers", "reach some key parameters exactly", "minimize some key parameters", ... This works satisfactory, as these additional goals are not trivially to obtain and hence, a new challenge arrives. But we do not encourage a fourth simulation, since this seems to be too much of training for the tool instead of gaining project management experience.
  On the other hand, if the number of such students is large in a course, this has overall negative effects, since certain rumours of how to trick the system become student-gossip. This leads occasionally to strange behaviour not intended by the instructors.

For sake of completeness we would also mention that we held so far three AMEISE tutorials for instructors who wanted to use AMEISE with their students. They also comprise the execution of simulation runs. Here, the didactical aims are totally different from those of regular students. Hence, we don't discus them here any further.

---

[6] AMEISE has already been put into action at Ecole Nationale d'Ingénieurs de Tunis (several times since winter term 2004), at RWTH Aachen (2 times since summer term 2005), University of Applied Sciences Kufstein Tirol winter term 2005), and University of Heidelberg (summer term 2006).

4.2. Experience gained so far

The most important experience was the high degree of the students' acceptance of AMEISE as integral part of the courses on Software Engineering. In addition to the rather theoretical characteristics of knowledge in software project management offered by traditional courses, the big benefit of AMEISE bases on the hands-on experience gained through AMEISE simulations. Hence, students can particularly learn that is not impossible but rather cost-intensive to add quality to a software product in cases when intermediate products already were of poor quality.

A further result was the experience that even with a rather small project (200 AFPs) carried out with a small gaggle of software developers it can easily happen that the student project managers lose control if they do not plan their projects carefully and update their plans continuously. It certainly might be argued against that in real world projects with real colleagues it will not be likely that one forgets whether Richard or Christine was concerned with a particular activity three months ago. The new version of the SESAM user interface and the intended AMEISE "next generation" user interface try to overcome these deficiencies by using graphical elements and ideograms.

The planning dimension is certainly the most significant difference between the students' and the practitioners' simulation runs. While most of the students started with their simulations soon after a brief interview phase in which they learned more about costs and qualification profile of the various software developers, the experienced project managers spent more than one hour on planning and preparing their simulation runs. Although their elaborated plans did not hold till the very end of their simulation, the results of their first simulation runs nevertheless correspond on the average to the results of the students' second simulation runs. The practitioners' second simulation runs even resulted in excellent cost and time outcome as well as in extremely high output quality.

Generally, it was noticed that the preparation phase is crucial to the later results of the simulation runs. The Carinthia Tech Institute students who attended the first AMEISE course spent also a substantial effort on the preparation of their first simulation run. One reason was certainly also the fact that they carried out their simulations at the Klagenfurt University which was motivation enough to strive hard to compare well with university students. Hence, their first simulation runs yielded results above average. This effect could not be observed anymore when the simulations where carried out in-house.

The new AMEISE user interface providing a model-specific command-input list with parameter selection was beneficial to the total duration of the simulation runs. The interaction with the system became notably easier. Thus, the trainees had not to worry about intricacies of the system but could better concentrate on their actual project management tasks. The time spent for a simulation run using the text-based interface exceeded 4 hours on average. With the more usable AMEISE interface the average simulation time was reduced to slightly more than three hours. As usually signs of fatigue will arise between the third and fourth hour of work, a noticeable improvement of the overall results could be achieved.

We would like to stress an experience we had with French speaking instructors that should be mentioned: Although we made any effort, to provide them with a French user interface (English and German were available right from the beginning), we figured not only that our translation was in certain corners incomplete and that the (compiler of the) core of the system had problems in coping with French accents. We have also seen that a few commands had unforeseen interactions between technical and colloquial use of terms. This led to an excessive number of problems, both with the instructors and later with their students. Eventually, it caused us to re-adjust English and German commands to keep the translation consistent while avoiding the above problems.

Another effect we noted already when comparing simulation results from Stuttgart and Klagenfurt, but much more from Klagenfurt and Tunis that the embedding of such a system depends heavily on the overall general as well as university specific culture.

The simulation runs were either carried out by single persons or by a team of two. These different settings did not yield any observable discrepancies in the results. However, in cases of team-work we could observe that upcoming decisions were always discussed between the members. This prevented from quick-click effects which might be possible especially with the AMEISE point-and-click interface. Brief discussions between the group members also helped in bridging the time gap caused by performance problems during early simulation runs.

Last but not least, we got very positive feedback on the two support components, advisor and friendly peer. For both components, the trainees supplied us with valuable input on additional critical simulation situations which can be supported by these components, too. Initial performance problems could be solved so far. While some early simulation runs had to be carried out without using the friendly peer in favour of achieving acceptable performance, recent simulations can fully benefit from all the helper components.

In general, we may report that the students were quite excited about the experience gained with AMEISE. Asked about the difference between having been told most of the things that were discussed in the feedback sessions in earlier software engineering classes and the discussions on the basis of the simulation runs, the standard reply was: "Now, this knowledge is founded on (my own) experience!" or: "But now, I not only heard you telling this, now I felt it!"

5. OUTLOOK

Since we use AMEISE not just as eLearning environment but also as sizable legacy system (total approximating currently 55.000 LOC Java code plus roughly 70.000 LOC of code in other languages), it is a constantly evolving system.

The part that experienced the highest degree of evolution is certainly the user interface. At the time we started our development, SESAM had a text based pseudo-natural language command interface for accepting the student-managers orders and a text based protocol reporting responses of the individual team members or of the system on their reaction to these commands or completion of tasks assigned. Our initial efforts concerned the input side. We offered the menu-based command interface shown in Fig. 1 as an option to the pseudo natural language version. However, from the very beginning, ideas to produce an even more innovative interface were around. The difficulty we met, however, was that this interface had to cope with too many dimensions at a time. To illustrate the problem, consider the interlinking of developer & task at a given point in time. The desire to show the progressing of tasks, i.e. the emergence of an artefact follows from this requirement, adding another dimension. This, of course has to be seen in the context of other artefacts already available and in some situations even in the context of artefacts yet to be developed. The *Historian*, proposed in the context of a diploma thesis [Putz 05] is a conceptual answer for this problem. Its prototype, discussed in section 3.4 is ready for integration into the system. It still needs to survive the real time test in a class-room situation. The same applies for a sliding time wall currently developed as diploma thesis, that provides a compromise between a high resolution view on the current situation and a wide angle perspective on aspects dealt with before (or after) the day(s) currently under focus.

Ongoing research is also needed in the area of focusing on relevant decisions and parts of the simulation trace. To explain effects within the simulation, instructors typically have to know about the rules stored in the SESAM model (that are executed by the simulation engine under specific circumstances). These rules describe positive and negative effects taking place when proceeding in the simulation. To scan and know about hundreds of rules is not feasible, and one diploma thesis is currently developing mechanism to automatically deduce the relevant rules that lead to a very specific situation in the simulation run. With this a more crisp assessment report and extensive feedback can be provided.

In this list of further extensions, we like to mention also one that was demanded by students (and even more violently by practitioners) was a support tool informing student-managers who has been working on a given artefact already. This information is helpful when assigning correction tasks as well as when assigning QA-

tasks, notably reviews. We rejected this demand so far with the argument, that it is the managers' duty to keep track of the project and providing too much system support would run against one of the individual sub-purposes. Of course, memorizing all these things is hardly possible. But one might/should take notes. Right now, we are on the brink of giving up to the argument of "Not I am doing it, my secretary does it". Thus, eventually, there will be an administrative assistant as support tool. But in contrast to other support tools, this *administrative assistant* has to be paid a salary and therefore the student manager has to decide whether this service will be worth the money needed.

A further item to be mentioned is that, though students are free to issue basically any command at any time in the project, the empirical data the project is based from is from linear (waterfall) document driven models. It would be interesting to establish a new model according to agile processes. An initial attempt in this direction did not yield the desired effect due to lack of validated data. Currently, attempts are under way to overcome this bottleneck.

Aside from these technical extensions, quite a number of didactical experiments are in the offing. These have to be done with care, since they must not be conducted in conflict with the educational goals we are pursuing with AMEISE. Though we use the system in several courses, there is still only limited room for carefully planned experience. Hence, we are happy that institutions who acquired after taking part in an AMEISE-tutorial an AMEISE licence become development partners at least in so far as the otherwise free licence requires that we obtain feedback on their use of the system and obtain a report on the experience gained. Thus, an open-source community, exchanging software and experience is slowly developing.

6. SUMMARY

The paper reports on AMEISE, an educational simulation environment to allow students to acquire project management experience in a non-threatening situation. We consider it important to stress that in order to obtain the educational merits, adequate preparation of students and proper embedding of the simulation environment in an advanced software engineering course is mandatory.

REFERENCES

[Hump 98] Watts S. Humphrey: *Managing the Software Process*; Addison-Wesley Publ., SEI Series in SE, 1989.

[DL 00] Anke Drappa und Jochen Ludewig: *Simulation in Software Engineering Training*; in Proceedings, 23rd International Conference on Software Engineering, IEEE-CS and ACM, May 2001: 199 - 208.

[Hoch 02] Elke Hochmüller: Project-centered Software Engineering Education, Technical Report, Carintha Tech Institute, Klagenfurt, Austria, Oct. 2002.

[MHBJN 03] R.T. Mittermeir, E. Hochmüller, A. Bollin, S. Jäger, M. Nusser: AMEISE – A Media Education Initiative for Software Engineering: Concepts, the Environment and Initial Experiences; in Auer (ed): Proceedings International Workshop ICL – Interactive Computer Aided Learning, Villach, Sept. 2003, ISBN 3-89958-029-X.

[MN 03] Markus Nusser. (In German) Projektunterstützung durch Verlaufsanalysen und Agenten. Univ. Klagenfurt, Institut für Informatik-Systeme, 2003.

[PP 05] Peter Putzer. Trace Representation of Simulated Software Development Processes. Univ. Klagenfurt, Institut für Informatik-Systeme, 2005.

[SJ 03] Susanne Jäger. (In German) Entwicklung eines elektronischen Tutors zur Analyse von Projektverläufen. Univ. Klagenfurt, Institut für Informatik-Systeme, 2003.

[Mitt 06] R. T. Mittermeir: Facets of Software Evolution; in: Madhavji, N.H., Lehman, M.M., Ramil, J.F. and Perry, D.W.: Software Evolution, Wiley 2006.

EXPERIENCE WITH EDUCATION OF INFORMATICS AT THE FACULTY OF ECONOMICS, TECHNICAL UNIVERSITY OF KOŠICE, USING LMS MOODLE

Dáša Lacková, Libuša Révészová and Dana Šiškovičová
Novitech Partner Ltd. Moyzesova 58, 040 01 Košice, Slovak Republic
dana.siskovicova@tuke.sk

Abstract

Basic aim of presented paper is to demonstrate the way of using ICT in education process at the Faculty of Economics Technical University of Košice, namely the subject Informatics. The historical background, subject content, student's feedback and planned improvement of using multimedia during the education process are described briefly.

Keywords

eLearning, ICT, Informatics, LMS, multimedia, Captivate

1. INTRODUCTION

The present can be characterized by radical changes in field of education. Many of institutions (not only educational) research the new forms and methods of wide possibilities of usage the ICT. One of them is the concept of eLearning. The availability of eLearning has made it possible for increasingly varied categories of learners to participate in tertiary education. Today's students' population can benefit from valuable learning experiences by interacting with their peers and tutors online, without inconvenience of commuting to a physical location.
A variety of teaching and learning styles can be employed when using the Internet as a medium. Most universities have adopted eLearning to support and to improve existing educational activities for their students. One of them is also Technical University of Košice, which has many activities oriented on using ICT during education process in each of its faculties.

2. USING ICT AND ELEARNING AT THE FACULTY OF ECONOMICS

The Faculty of Economics (FE) was established on July 8th 1992 in response to market demand, within the framework of transformation of the Technical University to a "full-fledged" university covering also non-technical disciplines. Mission of the Faculty of Economics is:

- to develop knowledge and provide high-quality, professional education in the field of economic sciences,
- to contribute to the development of society and shaping of the regional economy through the provision of research and consultancy activity, continuing education and services for the local and regional communities.

The Faculty of Economics offers a degree program in the branches "Finance, Banking and Investment", "Regional Development and Public Finance" and its graduates contribute to the professionalization of the financial and banking sector of the Slovak economy. Degrees acquired from the FE are relevant for employment in banks, investment, and insurance companies, and in the entrepreneurial sector as well.
The aim of the faculty is to prepare our students for practical live as good as possible. The basic requirement of students' whole around the world is a lot of practical examples. The best way is possibility of improvement of their skills by simulation exercises. Many of offered subjects have implemented these trainings (e.g. Risk software in subject Risk and Uncertainty, Pohoda in subject Counting). The first step for the students that they have to accomplish is to complete the subjects Informatics I and II, where they have to demonstrate their practical skills using ICT. Then they can enroll to economic subject with ICT support.

With the increase of ICT we tried to motivate our students to use ICT as much as possible. Firstly we were providing different supporting materials and information on web pages of our departments (Fig.1.). These pages contain only basic information about teachers, content of subjects, brief description of expected assignments, etc. This "static" way had some disadvantages: impossibility of on-line communication, on-line testing or handling of students' projects.

Because today's students want to have an interactive environment, the existing web pages of subjects provided by responsible departments had become unsatisfactory. Appropriate environment enables them

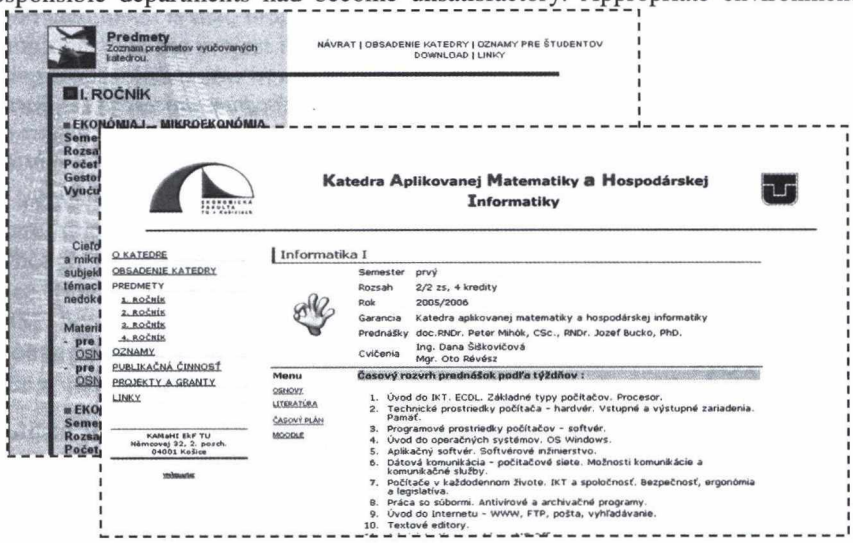


Fig.1 Examples of subjects' web pages provided by departments.

interaction with their peers and tutors online, without the inconvenience of commuting to a physical location and adhering to a rigid timetable to learn, or provides study material online every time.
Effort of fulfilling mentioned requirements flew into the introduction of LMS and its usage in the Faculty of Economics, Technical University of Košice. There are various accessible LMSs. We chose LMS Moodle [5] due to requirements of today's web development (portal structured webs) and easy navigation (Fig.2). It provideś the possibility of exchanging private information between teacher and students like evaluation of assignments, feedback from students, etc. as well.
We started to use Moodle in 2004 as the supporting environment for teaching subjects Informatics and Spreadsheets. After obtaining the positive feedback from students we decided to use it for supporting the

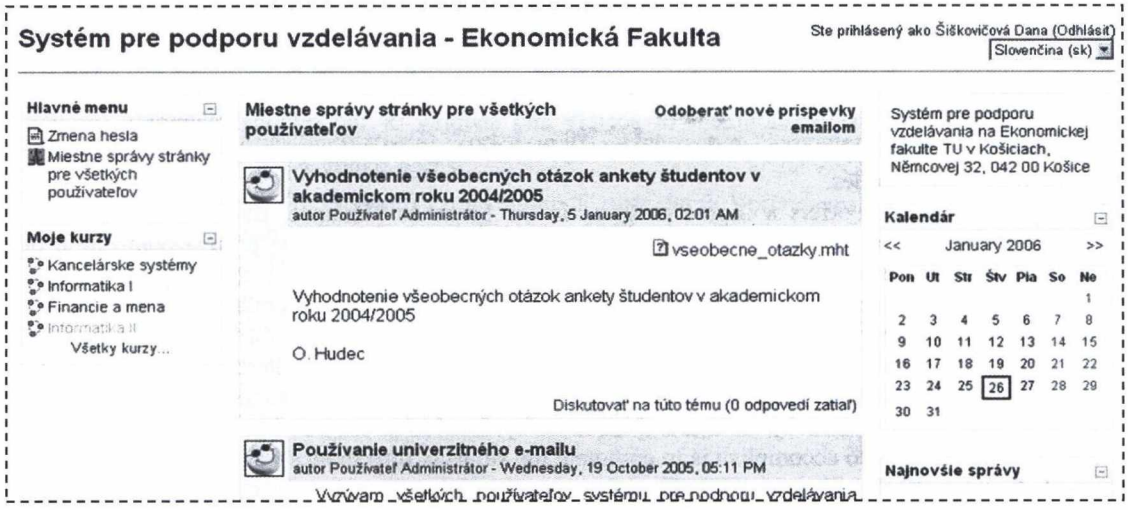


Fig.2 LMS environment - Moodle

whole of education process at the faculty (like Microeconomics, Macroeconomics, Economic Analysis, Risk and Uncertainty, Counting, Economic Information System…).

3. SUBJECT: INFORMATICS

A. Content

The aim of the faculty is education of professionals in economics with additional computer skills. That is why the content of the subject Informatics is mainly concentrated on practical knowledge and skills. For the purpose of finding out students' current IT skills and knowledge, we interview them at the beginning of the first year. In view of results of questionnaires and requirements of subjects thought in higher years of our faculty, within which ICT are used, we fit the content of subject Informatics every year. At the present time the course Informatics is divided into two parts – Informatics I and Informatics II and its content comes out from ECDL requests.

The first one is concerning on ECDL modules 1, 2, 3 and 7, concretely:

- **Module 1** – Concepts of Information Technology (IT, types of computer, main parts of PC, computer performance, hardware, software, information networks, health and safety, environment, security, copyright and law)
- **Module 2** – Using the Computer and Managing Files (computer environment, desktop, managing files, viruses)
- **Module 3** – Word Processing (basics of using application MS Word, main operations, formatting, objects)
- **Module 7** – Information and Communication (the Internet, web navigation, web searching, electronic mail, messaging)

The second one is related to ECDL modules 4, 5 and 6, concretely:

- **Module 4** – Spreadsheets (basics of using application MS Excel, cells and operations with them, worksheets, formulas and functions, formatting and charts)
- **Module 5** – Database (database concepts, basics of using application MS Access, tables, forms, retrieving information)
- **Module 6** – Presentation (basics of using application MS PowerPoint, developing and presentation, text and images, charts/graphs, drawn objects, slide show effects and preparing outputs)

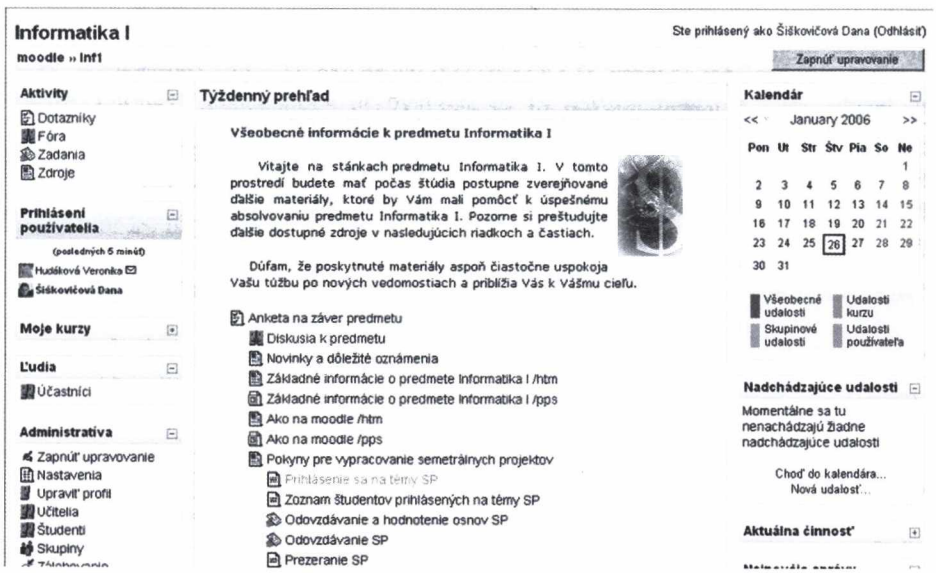


Fig.3 Course Informatics I in Moodle's environment.

Due to special requirements of our students there are some additions in mentioned subjects:

- web pages design including basics of HTML, W3C convention, etc.. Students gain real practice with pages design and locating them on web server (within Informatics I),
- information systems (IS) including knowledge of architecture of IS, processes in life cycle of the IS, etc. Outputs of the subject are real "mini ISs" – their basic design, definition of users' requirements, database design, presentation and documentation of created IS (within Informatics II).

B. Subjects outcomes

Informatics I

As was mentioned, subject Informatics I is concerned on ECDL modules 1,2,3 and 7. Skills gained during study of the subject are evaluated using three different ways:

- test of practical skills by which students have to demonstrate their basic skills in working with OS Windows, the Internet and its services (like web, e-mail, file transfers,...), basic applications in accessories of Windows and word processor,
- test of theoretical skills – used for verifying students' knowledge of terms and general view in the field of IT,
- semestral project – the outcome created by students themselves. The main objective of the semestral project is to design web page representing found information about chosen theme. Students have the possibility of choice from two basic fields:
  - informatics (simply interpretation of basic terms from IT, possibilities of using PC in our lives, etc.),
  - economy (list of economical schools and trends, simply interpretation of basic terms from economy, important persons of economy, etc.).

The elaboration strategy of semestral project can be briefly resumed into following steps:

1. Formulation of basic concept (MS Word document created using template) – brief description of topic which student wants to present on the final web page and used references.
2. Check of concepts by teacher. During this period we closely cooperate with our colleagues from economic departments to provide maximum relevancy of published information on the students' web pages.
3. Final web page design and its uploading to the server through Moodle.

After submitting the project through Moodle, each submitted project is evaluated. During evaluation we take in four basic pieces of web page design:

- graphical design – coloring of the page from users' point of view, spreading information within window area, lucidity, etc.
- navigation – comprehensiveness of designed web page, easy orientation for new user between other pages of web-site, possibility of easy transition to other pages, etc.
- used technology – the main criteria are functionality of hyperlinks, correct names of files, visibility of images or other used graphics like themes, buttons, other components, etc.
- information content – relevancy of used information (considering to basic concepts and teacher's notes).

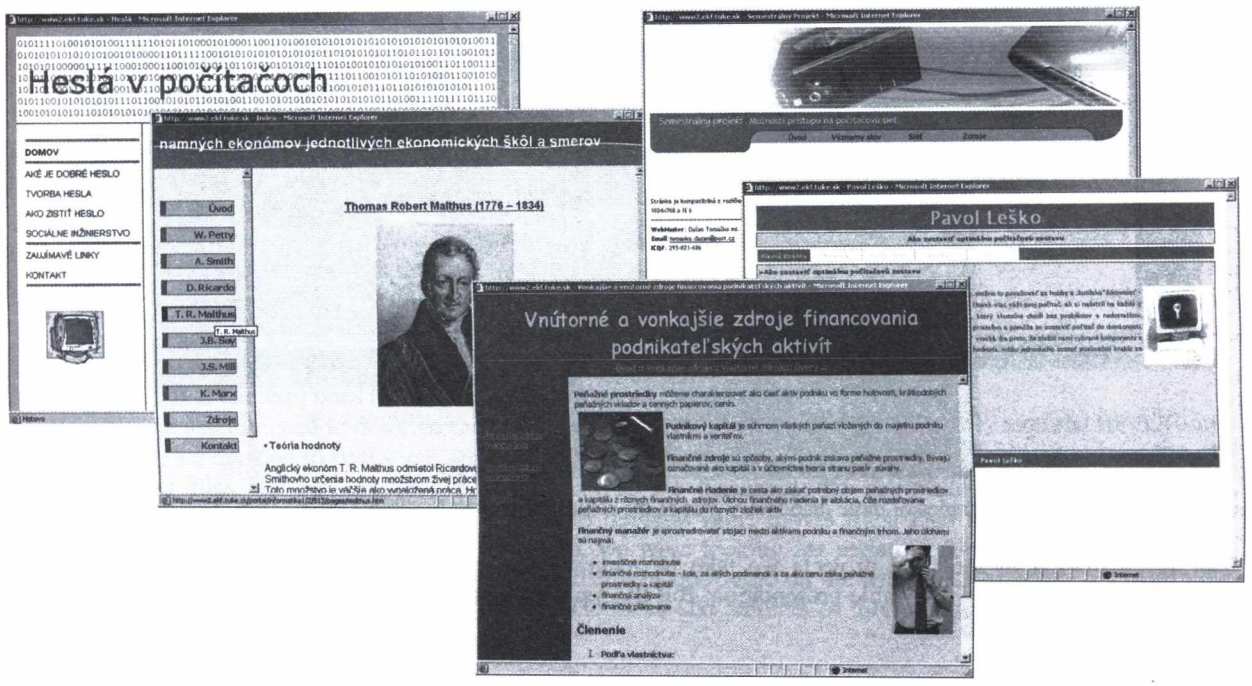


Fig.4 Examples of students' web pages (output of semestral project)

Informatics II

Subject Informatics II covers the missing ECDL modules 4, 5 and 6. Equally as it is in Informatics I students have to take the practical test and the theoretical one at the end of semester. Additionally students have to recompose information from the web pages, which they have designed within Informatics I, into "mini IS". "Mini IS" is divided into three parts:

- documentation – the doc file which includes title of project, brief and exact description of problem, specification of main aims (answers to the questions: What? Why?), definition of all groups of users (answers to the questions: Who? When? How?), schedule of most important parts and functions of new IS, connections and relations to another existing IS around, required infrastructure (hardware, software, intranet), conceptions of methods of IS development (CASE, object orientation,...), requirements of safety and reliability of IS, presumed dates of realization and approximation of costs (answers to the questions: How long? How much?), way of IS implementation (prototype, probative running,...), requirements of documentation and training and prospect of IS, expectation, expansion of IS, maintenance, service.
- design of database architecture – the mdb file which includes data categorized into tables (3 – 5 tables), relations between tables and functional queries (to obtain required specific information expected by the end user (which have been defined in documentation)).
- presentation – the ppt file which brings out the main ideas of "mini IS" and it's structure.

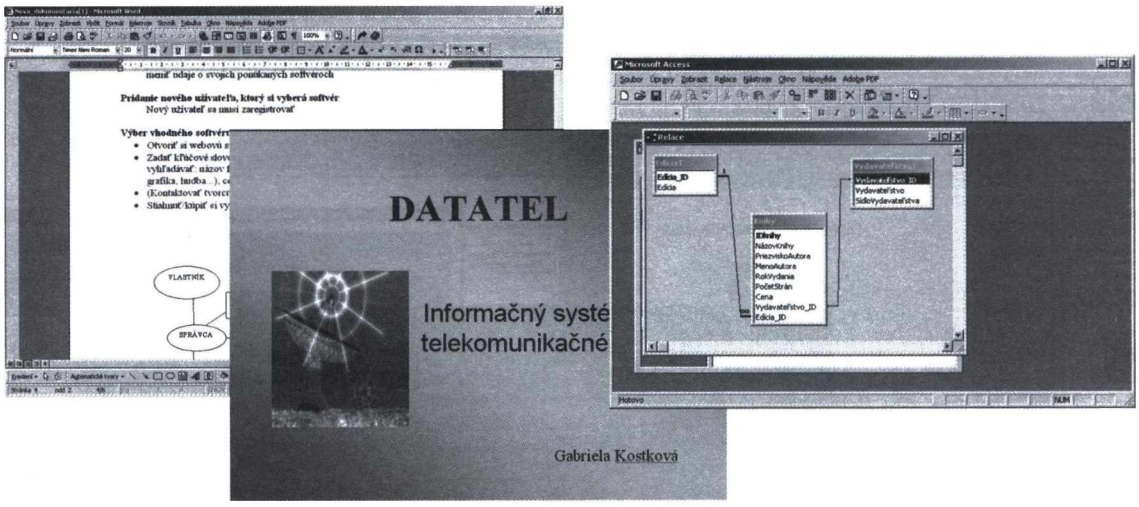


Fig.5 Examples of three mandatory part of the semestral project in subject Informatics II

During evaluation of the final projects we take in accuracy and distinctness of documentation, rational structure of database and enjoyable presentation. But the most important thing we take in is the functionality of the submitted "mini IS".

C. Students feedback

Our effort is to oblige students' requirements of the content of the subject. That is why we are interested in students' satisfaction with the subject as the unit and we decided to ask them to fill out another questionnaire (feedback).

The feedback questionnaire consisted of 22 questions concerning different parts of the subject: the content of subject, the way and quality of lectures and classes, the quality of study materials, communication between teacher and students, the evaluation methods of different parts of the subject and the quality of used LMS Moodle.

Generally students were satisfied with the content of the subject and available study materials. Of course they had some remarks and suggestions, but in this paper we do not want to deal with the detailed analysis of results of students' feedback. We just want to highlight one from frequent remark – living up the study materials by some animations or interactive videos.

4. ELEARNING AND MULTIMEDIA

At the beginning of "Internet age" many of teachers supposed that it would be enough to publish their existing printed textbooks and preparations in electronic way. Following time period showed this idea is great misunderstanding. Students, which use Internet medium during their learning, need more.

One of the advantages of electronic study materials is that they can be created using not only static text, but graphical information too [1,2]. We have an opportunity to choose from static or "moving" graphics like animations and video. It can be supplemented by audio. Standard electronic materials can provide all components using one source without necessity of interrupting learning and going to another source (like video recorder) to watch enclosed video. They make the study process easier. Graphics can be divided into:

- static - raster or vector graphics, product of commonly used graphical editors (Adobe Photoshop, Adobe Illustrator, Corel Suite, Paint Shop Pro, Paint, ...), digital camera, scanner, etc. The file formats jpeg, gif or png are usually used formats for web publishing of electronic study materials. This type of graphics is used to illustrate textbooks or other study materials.
- dynamic – different video forms:
  - video programs – uninterrupted, ordinarily about 30 minutes long videos created by television or TV studios (e.g. well-known language courses),

- video segments – their basic purpose is to describe some field and to wait for some activity of the student (e.g. student can stop the video and find the answer to question or solve the problem given in the video). Its length is ordinarily about 6-7 minutes. Video segments are used as supporting material to text.
- interactive videos and animations – usually used in interaction with PC. Their power is in living up the study materials and demonstration of simple processes. Usage of this video form is rapidly increasing nowadays.

If you decide to integrate video into your study material, you can choose from two possibilities:

- to invest into video technologies for recording and processing the video results. Specialized software like Adobe Premiere is usable for this purpose.
- to buy specialized software using which we are able to create a dynamic sequence of static images (ordinarily gif animators - Zoner, Barbarosa, CoffeCup,...), to capture the sequence of steps done on the desktop (Corel Capture) or to create new movies from drawn objects or imported (from scanner, digital camera) ones (Macromedia Flash).

Because we usually need to describe the environment of applications, sequence of steps required for fulfilling some task; we decided to find some application, which can help us to develop simple animations resulting in small-sized files. In our previous contribution [3] we discussed the opportunities and advantages of Macromedia Flash, which is commonly used to develop short movie files. Due to our experiences with Macromedia products like Flash and RoboHelp, we were looking for some product compatible with final files developed by both of them. We found application RoboDemo on the website of Macromedia [4]. In the following we would like to present basic characteristics and features of this product, which we were using during development of our study materials.

5. MACROMEDIA CAPTIVATE AND ITS BASIC FEATURES USABLE DURING ELEARNING STUDY MATERIALS' DEVELOPMENT

A. Brief description of Macromedia Captivate

Macromedia Captivate (former RoboDemo) is professional software tool for quickly creating interactive demonstrations and simulations in a variety of formats including flash (swf) and exe. We chose it due to its very intuitive and user-friendly environment, which enables user to add customized text captions, audio (voice-overs, background music and sound effects), video, Flash animations, text animations, images, hyperlinks and more to final movie. This tool provides small file size and high resolution like a flash movie, what is ideal for online publishing.

Making of the study materials includes two phases – creation of the materials (recording or importing movies/partial sequences, their annotating and editing) and their following publishing online, on CD or integrating them with other possible application (Fig. 6).

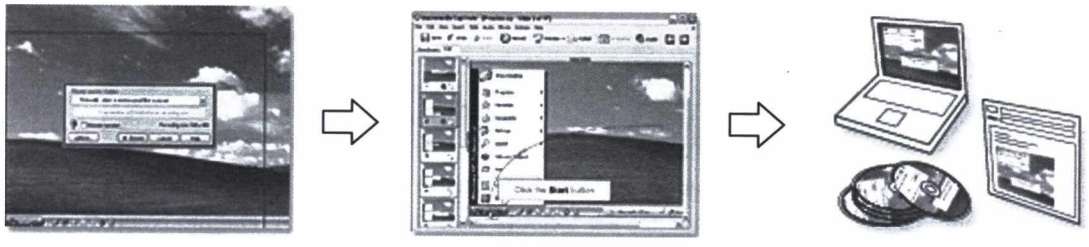


Fig. 6 Basic steps of movie development

Within these phases different Captivate's features are very useful. Let us emphasize the most important of them.

1. We can start to develop the movie in different ways. We can simulate required task and record on-screen actions using Captivate, which automatically inserts a text description of each recorded task

and scored interactions with instructional feedback just during recording. But Captivate enables us also possibility to import files (PowerPoint presentation, AVI format file), which can be used as the base for the new video. Then we can easily choose particular slides from PowerPoint presentation or frames from avi file.

2. After gaining the basis of our movie (the term "project" is used in Captivate), we can begin its modification to make it as efficient as possible. This modification can run on different levels – modification of the project as the whole, of the one particular slide/frame or modification of the individual objects placed in the slide (Fig. 7).
   - For the project as the whole we define basic options like quality of final movie (standard, jpeg, high quality), resolutions, speed of previewing frames (in frames per second), way of previewing (stopping of preview at the end of the movie, looping the movie, opening other movies,...), defining type and position of playback control panel.
   - Every screen is completely editable slide that can be removed, replaced or revised. This makes changing or updating movie easier. Furthermore we are able to add new slides according to our requirements – image, blank, question, PowerPoint and animation slides. If it is necessary, we can record other slides subsequently. Each slide can have specific properties: label, display time, transition, quality and background color.
   - The individual objects placed in the slides are used for the increase of the effectiveness of developed movie. Some objects types just top up information content of the slide (text caption, highlight box, image, rollover caption, rollover image, animation, text animation), the others provide special interactivity features (click box, text-entry field with multiple correct answer options, button) – including quizzing with scoring and branching.

Fig. 7 Example of the project edited in Captivate

Since we deal with movie, it is important to have possibility to edit its timing additionally. Visual timeline is used for this purpose. It enables to control the timing of each slide element by dragging and dropping in the timeline. Timeline is organized in two layers as it is in Macromedia Flash. Drag and drop layers help to change the preview order of the objects in the slide. Implemented changes can be seen instantly by playing the slide without previous compiling or generating preview. But for the checking the final movie we need to generate preview.

In the case we are satisfied with the final movie, we can start to deal with the last phase – publishing the movie. There are different ways of publishing the movie. It can be embedded into swf file, exe file or flash file. Captivate movies can be added to:

- Macromedia RoboHelp HTML on-line help system to show and tell users important information,
- Macromedia Flash to have possibility to additionally enhancing, scripting or integration the movie into a larger project,
- Macromedia Breeze to have possibility of quick and easy participation in on-line meetings, presentations and training over Internet,
- Microsoft PowerPoint to insert movie into PowerPoint presentations.

B. Our experiences

Considering our activities performed at our faculty in the field of eLearning and distance learning and our effort to prepare as better study materials as possible, we decided to extend existing study materials with above-mentioned multimedia. Previously we were using only PowerPoint presentations and their export to html format for living up the materials. Many people in Slovakia have traditional modem access (with speed about 56kB) to Internet. That is why students had problems with on-line viewing these files (due to their size or troubles with viewing objects blocked by firewall, browser,...). These problems and simplicity of editing and developing animations using Captivate motivated us to import mentioned PowerPoint presentations into Captivate movies. Moreover this software tool enables final movie bandwidth analysis. This key feature analyses individual slides' demand on bandwidth and consecutively displays required bandwidth across time for different modem and broadband speeds in the graphs (Fig. 8).

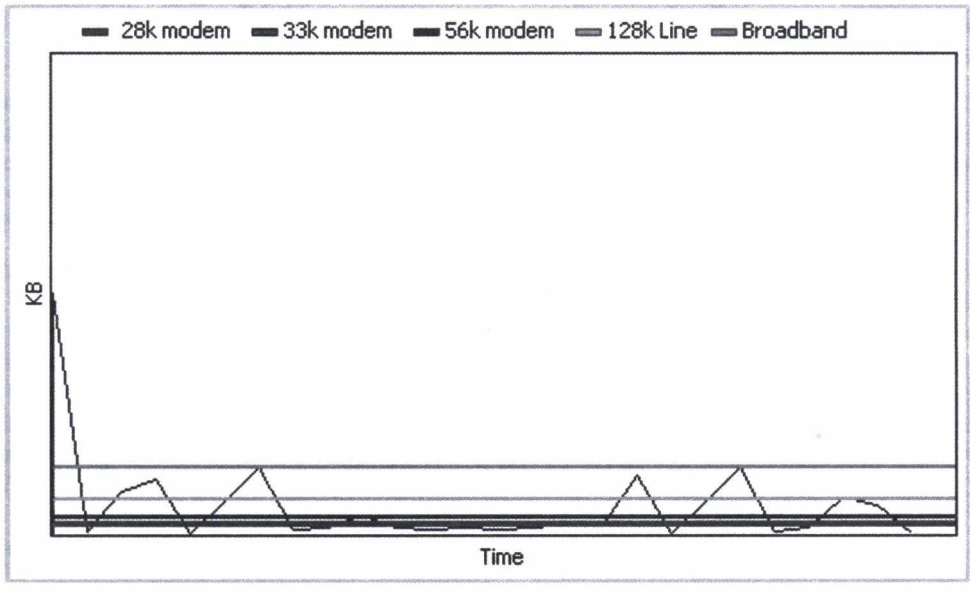


Fig. 8 Illustration of output from the Bandwidth Monitor feature

Thanks to obtained outputs from the Bandwidth Monitor we can modify the most demanding slides to fulfill requirements of transmission speed. As for viewing final movies students take advantage of the ubiquitous Flash Player. So no plug-in downloads are necessary because over 98% of Internet users can already view flash files. Therefore we are able to publish movies, which does not cause problems within their on-line viewing.

Last but not least Captivate's feature is that it is SCORM certified product. That is why it is easy to integrate the final movie into various LMSs.

6. CONCLUSIONS

In this paper we tried to explain briefly the way of teaching the subjects Informatics I and Informatics II using LMS Moodle. Mentioned LMS is often used for providing study materials, communication between students and teachers, publishing evaluation of assignments, etc. As results from realized feedback, the subject does not fulfill students' expectations on 100%, but (and it is interesting) 87,2% of student would recommend the subject to their younger colleagues.
Our effort is improving the subject's content as much as possible in the future. That is why we started to use Macromedia Captivate for developing more efficient eLearning study materials. The key features of this tool, which we appreciated during creation study materials the most, we have described.

REFERENCES

[1] Mikuláš Huba, Iveta Orbánová: "Pružné vzdelávanie", STU Bratislava, 2001, ISBN 80-227-1335-X (in Slovak).
[2] Mikuláš Huba, Katarína Žáková, Pavol Bisták: "WWW a vzdelávanie", Vydavateľstvo STU, Bratislava, 2003, ISBN 80-227-1999-4 (in Slovak).
[3] Dana Šiškovičová: "Opportunities of using the Macromedia Flash in creation the study materials", Proceedings of Technology-Enhanced Learning with Ubiquitous Applications of Integrated Web, Digital TV and Mobile Technologies: HUBUSKA Open Workshop: 6th eLearning Forum, Budapest, Hungary, 9-10 June 2005, p. 61-66, ISBN 963 311 361 X.
[4] www.macromedia.com
[5] www.moodle.org

DIGITAL WATERMARKING IN TRANSFORM DOMAIN

Rastislav Hovančák
Department of Electronics and Multimedia Telecommunications,
Technical University of Košice,
Park Komenského 13, 041 20 Košice, Slovak Republic
Phone: +421-55-602 2808; E-mail: Rastislav.Hovancak@tuke.sk

Abstract

Digital watermarking is a technique for hiding information in digital media. Hiding information is called digital watermark. The watermark can be latter detected or extracted for purpose of owner or author identification. In this paper we present four different digital watermarking methods in transform domain. Some similarities and differences between proposed watermarking methods are discussed.

Keywords

watermarking, transform domain, hiding information, attacks

1. INTRODUCTION

Nowadays, daily communications of all kinds over the Internet have become incredibly popular. However, message transmissions over the Internet still have to face all kinds of security problems. Easy access to digital media and the increasingly powerful tools available for editing digital media have made copyright protection and authentication a very important issue. While cryptography is about protecting the content of messages, watermarking is technique for hiding additional information in original data.

Digital watermarking is process of embedding additional information directly into the digital data, also called original data, by making small modifications to them. The additional information is called watermark. In watermarking, object of communication are original data and the watermark only references these original data. As watermark can be used image, voice sample, video etc. The watermark can latter be detected or extracted for purpose of owner or author identification and integrity verification of tested data.

Watermarks and watermarking techniques can be divided into various categories and in various ways. The basic and most common used partitioning of watermarking is the spatial domain, frequency domain and parametric domain watermarking. Another way for categorization of watermarking methods is based on the condition whether or not they use the original data for extraction of watermarks from tested (watermarked) data. And besides that we can distinguish visible and invisible watermarks. In this paper we present methods of invisible watermarking techniques operate in transform domain for still images. Digital watermarks have form of binary images.

2. GENERIC EMBEDDING AND EXTRACTING SCHEME

Figure 1 shows generic embedding and extracting scheme. The inputs to the embedding scheme are watermark, the original data and an optional public or secret key K. The output of the embedding scheme is watermarked image. Inputs to the generic extracting scheme are the tested data, the secret or public key and depending on the method, the original data and/or the original watermark. The output is either the extracted watermark or some kind of confidence measure indicating that the watermark was detected.

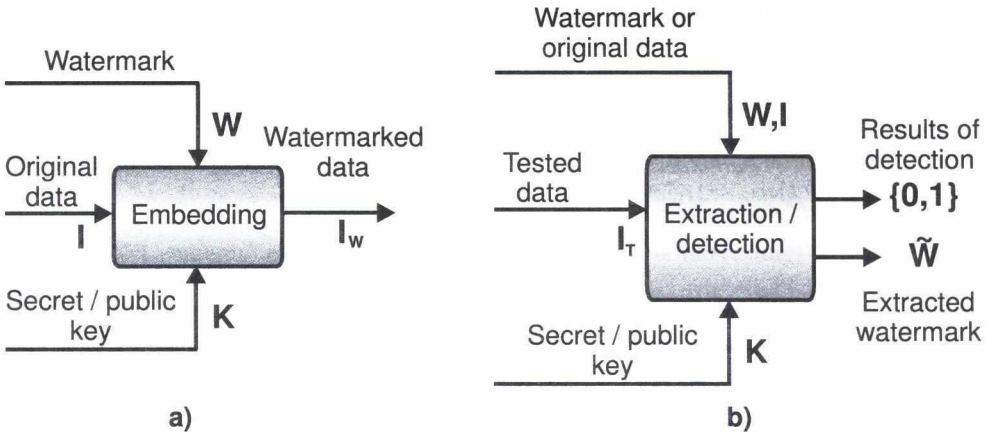


Figure 1 Generic watermarking scheme

Three types of watermarking systems can be identified:

- private watermarking
- semiprivate watermarking
- public watermarking

3. REQUIREMENTS

Watermarking systems have to guarantee these requirements [1]:

- **robustness** – The embedded information is said to be robust if its presence can be reliably detected after the image has been modified but not destroyed beyond recognition.
- **undetectability** – Embedded information is undetectable if the image with the embedded message is consistent with a model of the source from which images are drawn.
- **perceptual transparency** – It is based on the properties of the human visual system or the human audio system. The embedded information is imperceptible if an average human subject is unable to distinguish between carriers that do contain hidden information and those that do not.
- **security** – The embedding algorithm is said to be secure if the embedded information cannot be removed beyond reliable detection by targeted attacks based on a full knowledge of the embedding algorithm and the detector, and the knowledge of at least one carrier with hidden message.

The above requirements are mutually competitive and cannot be clearly optimized at the same time. If we want to hide a large message inside an image, we cannot require at the same time absolute undetectability and large robustness. A reasonable compromise is always a necessity. On the other hand, if robustness to large distortion is an issue, the message that can be reliably hidden cannot be too long. This observation is schematically shown in the Figure 2.

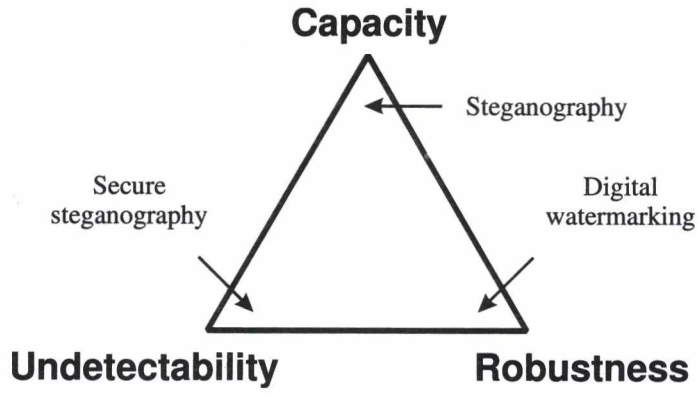


Figure 2 Conflicting requirements

4. APPLICATIONS

We can distinguish following applications of digital watermarking [2][6]:

- **copyright protection** - The watermark embedded into a multimedia content unambiguously identifies the true author. The embedded watermark must be robust against nonintentional processing such as compression and secure against intentional attacks that try to destroy or remove the embedded watermark.
- **Fingerprinting** - Digital fingerprinting is a technology for enforcing digital rights policies whereby unique labels, know as digital fingerprints, are inserted into content prior to distribution. This technique can facilitate the tracing of illegal copies of multimedia content and identify users who use their content for unintended purposes.
- **copy control** - Copy control uses the embedded watermark to carry information about the rules of usage and copying of multimedia content. In general these rules can simple express that
  - this content may not be copied,
  - this content may be copied, but no subsequent copies may be made of that copy.
- **authentication or integrity verification** - Multimedia authentication or integrity verification can be performed by using
  - checksum of multimedia content,
  - information digest, such as a message authentication code or a digital signature,
  - fragile (semifragile) watermarking

5. ATTACKS ON WATERMARKING SYSTEMS

Attacks on digital watermarking systems have two effects: either they reduce the effective channel capacity of fully disable the detection of the embedded watermark. Because it is not possible to enumerate all possible attacks, it is very difficult or even impossible to assess if given system is robust in the general sense (Figure 3) [3].

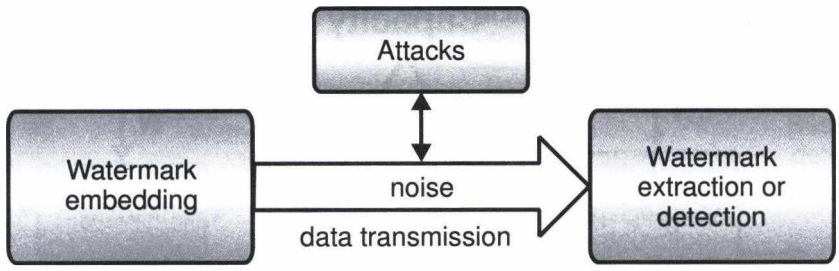


Figure 3 Attacks on watermarking systems

Five different groups of attacks can be identified: removal and interference attacks, geometrical attacks, cryptographic attacks, protocol attacks and other attacks (Figure 4).

5.1. Interference and removal attacks

The main idea of interference attacks consists of assuming that the watermark is additive noise relative to the original image. The interference attacks are those which further add noise to the watermarked image. This noise may have any of a number of different statistical distributions such as Gaussian or Laplacian. The removal attacks exploit the linear additive model in order to derive optimal estimators used for denoising and consequently removing of the watermark. In other cases both the removal attacks and the interference attacks can be combined such as in the denoising with perceptual remodulation attacks.

5.2. Geometrical attacks

In contras to the removal attacks, geometrical attacks intend not to remove the embedded watermark itself, but to distort it through spatial alterations of the stego data. The attacks are usually such that the watermark detector loses synchronization with the embedded information. The most well know integrated software versions of these attacks are Unzign and Stirmark. Unzign introduces local pixel jittering and is very efficient in attacking spatial domain watermarking schemes. Stirmark introduces both global geometrical and local distortions. The global distortions are rotation, scaling, change of aspect ratio, translation and shearing that belong to the class of general affined transformations. The line/column removal and cropping/ translation are also integrated in Stirmark. Most resent watermarking methods survive after these attacks due to the usage of special synchronization technique. If robustness to global affined transformations is a solved problem, the local random alterations integrated in Stirmark still remains an open problem almost for all techniques. The so called random bending attack exploits the fact that the human visual system is not sensitive against shifts and local affined modifications. Therefore, pixels are locally shifted, scaled and rotated without significant visual distortions.

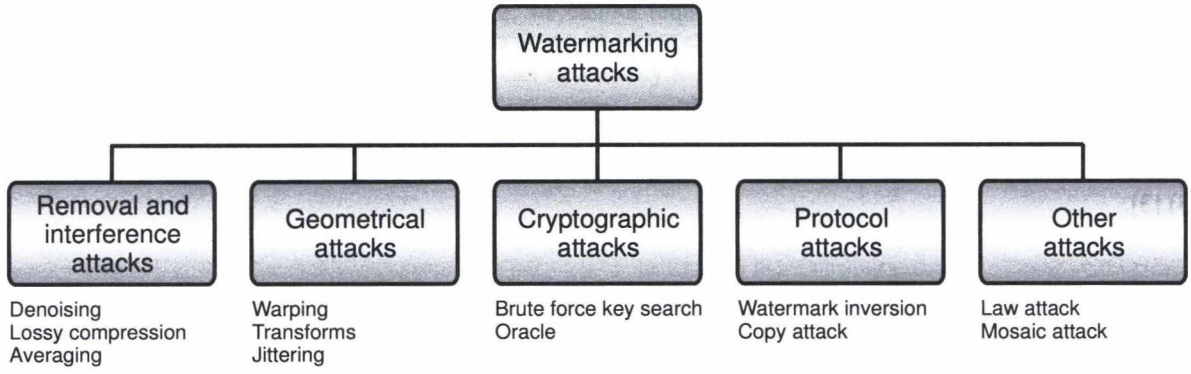


Figure 4 Groups of the attacks

5.3. Cryptographic attacks

Cryptographic attacks are very similar to the attacks used in cryptography. There are the brute force attacks, which aim at finding secret information through an exhaustive search. Since many watermarking schemes use a secret key it is very important to use keys with a secure length. Another attack in this category is the so called Oracle attack which can be used to create a non-watermarked image when a watermark detector device is available.

5.4. Protocol attacks

The protocol attacks aim at attacking the concept of the watermarking application. The first protocol attack was proposed by Craver et al. They introduce the framework of invertible watermark and show that for copyright protection applications watermarks need to be non-invertible. The idea of inversion consists of the fact that an attacker who has a copy of the stego data can claim that the data contains also the attacker's watermark by subtracting his own watermark. This can create a situation of ambiguity with respect to the real ownership of the data. The requirement of non-invertability on the watermarking technology implies that it should not be possible to extract a watermark from non-watermarked image. As a solution to this problem, the authors propose to make watermarks signal-dependent by using a one-way function.
The copy attack also belongs to the group of the protocol attacks. In this case, the goal is not to destroy the watermark or impair its detection, but to estimate a watermark from watermarked data and copy it to some other data, called target data.

5.5. Other attacks

The Mosaic attack consists of chopping an image up into a number of smaller subimages, which are embedded in a suitable sequence in a web page. Common web browsers render juxtaposed subimages stuck together, so they appear identical to the original image. This attack appears to be quite general, all marking schemes require the marked image to have some minimal size. Thus by splitting an image into sufficiently small pieces, the mark detector will be confused.

5.6. Classification of the attackers

First classification divides attackers into three groups: active attackers, passive attackers and malicious attackers. The most dangerous are active and malicious attackers.

According to other classification there are privacy attackers, professional attackers and private hackers. Privacy attackers are probably not very qualified. Professional attackers intend to make a commercial profit from their counterfeit. Therefore they may invest in powerful technologies to succeed in their piracy, and they will certainly have qualified competencies to break the watermarks. Private hackers are usually not interested in any financial benefit from their act, since their more frequent motivation is the performance of breaking a system which is supposed to be secreted. They are extremely well qualified and may use very powerful techniques. Hackers are not financially greedy, but as they are looking for the widest recognition, they may propagate their discovery as a mean of advertisement.

6. PROPOSED METHODS

In digital watermarking can be used many different approaches and techniques for watermarks embedding. Very important techniques are digital watermarking based on two-dimensional discrete cosine transformation (2D DCT) and two-dimensional discrete wavelet transformation (2D DWT).

In field of digital watermarking two basic methods are used: method using original image for extraction of watermark from image under test and method without using original image for extraction of watermark from image under test.

We propose comparison of four methods in transform domain [4][5]:

- DCT method using original image for extraction (DCT_M1)
- DCT method without using original image for extraction (DCT_M2)

- DWT method using original image for extraction (DWT_M1)
- DWT method without using original image for extraction (DWT_M2)

6.1. Embedding process

Figure 5 shows embedding process, which all methods use. As we can see embedding process contains of two-dimensional discrete transform of original image, permutation of watermark, modification of transformation coefficients and finally inverse 2D transform of modified coefficients. Inputs of embedding process are original image I of size (N_1xN_2) and watermark W of size (M_1xM_2). Outputs of this process are watermarked image I_W with size (N_1xN_2) which is different from original image and secret key. Secret key contains information about changed transformation coefficients, permutation random vector, size of watermark etc. Before modification of coefficients, two-dimensional pseudorandom permutation of watermark is used to disperse its spatial relationship. Permutation is used for increasing security for all watermarking method.

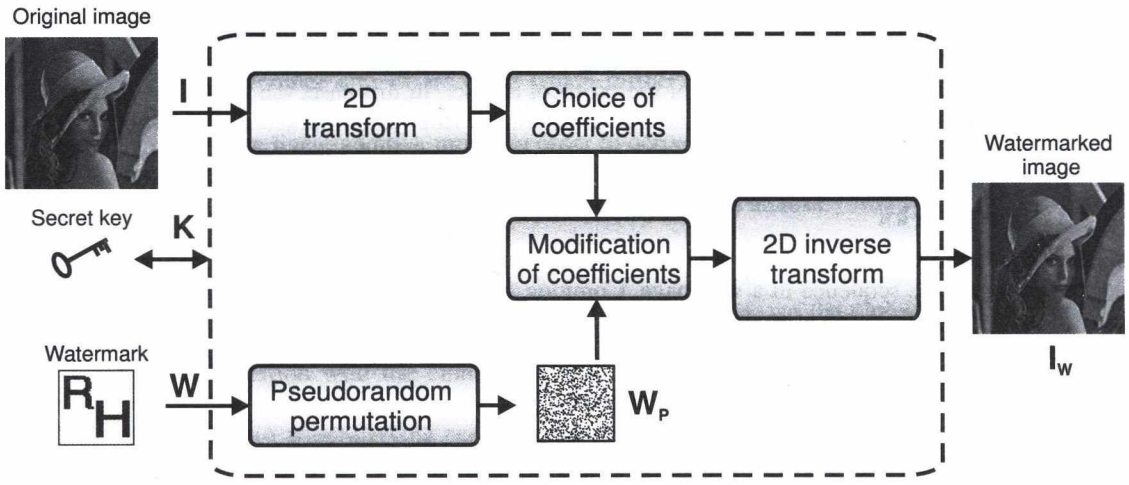


Figure 5 Embedding process for all methods

6.1.1. Embedding using discrete cosine transform

The main arguments for using DCT in watermarking are the following. Embedding rules operating in the DCT domain are often more robust to JPEG and MPEG compression, thus the watermark designer can prevent JPEG/MPEG attacks more easily. Watermarking in the DCT domain offers the possibility of directly realizing the embedding operator in the compressed domain in order to minimize the computation time. The original image I is divided into blocks of 8*8, and each block is DCT transformed independently.

The middle-frequency coefficients selected from the image are extracted. Figure 6 shows the examples of the coefficients' extraction for methods DCT_M1 and DCT_M2.

DC	1	5	6	14	15	27	28
2	4	7	13	16	26	29	42
3	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63

a) DCT_M1

DC			6	14	15	27	28
2	4	7	13	16	26	29	42
	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63

b) DCT_M2

Figure 6 Choice of the DCT coefficients

6.1.2. Embedding using wavelet transform

Wavelets are a key technique in the source compression standard JPEG-2000. In several publications, this technique has been applied to image watermarking. The positive arguments closely resemble those for advocating DCT for JPEG. The multiresolution aspect of wavelets is helpful in managing a good distribution of the watermark in the cover in terms of robustness versus visibility. The wavelet transform consists in a multiscale spatial-frequency decomposition of image. In our approach we include watermarks into the third level details coefficients. **Figure 7** shows the examples of the coefficients' extraction for methods DWT_M1 and DWT_M2.

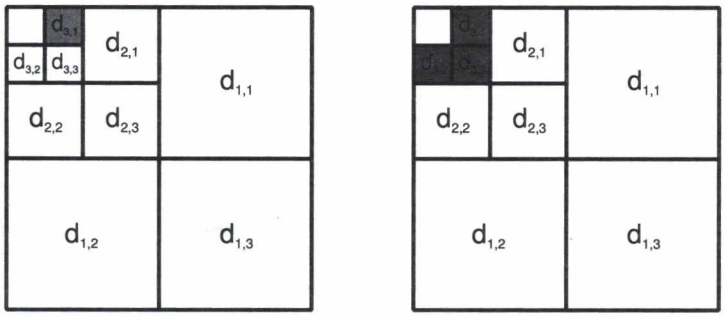


Figure 7 Choice of the DWT coefficients

6.1.3. Modification of transformation coefficients in DCT_M1 an DWT_M1

We denote watermark's bit as $w(x)$ and $i(x)$ as transformation coefficient. Watermarked coefficient $i_w(x)$ can be expressed in the form:

$$i_w(x) = i(x) + \alpha \quad if \quad w(x) = 1 \tag{1}$$

$$i_w(x) = i(x) - \alpha \quad if \quad w(x) = 0 \tag{2}$$

where α is a real number.

6.1.4. Modification of transformation coefficients in DCT_M2 an DWT_M2

Methods DCT_M2 and DWT_M2 have main advantage opposite to methods DCT_M1 and DCT_M2 that they don't need original image in extraction process. As is shown in Figure 6, three transformation coefficients for one watermark's bit are used. Extracted coefficients are compared each other (L-low, M-medium, H-high). It stands to reason, there exist 3^3=27 combinations, but some combinations are not used. All combinations, which represent embedding of watermark bit 0 or 1 are in Table 1. In these methods, embedding of watermark means that arbitrary combination to required combination is changed. If extracted combination doesn't corresponds with combination from Table 1 we have to change this combination.

Table 1 Combinations represents embedded bit 0 and 1

Bit of watermark	Combinations		
	C_1	C_2	C_3
0	L	M	H
	M	L	H
	L	L	H
1	M	H	L
	H	M	L
	H	H	L

Embedding process for bit "0" of watermark is described in Tab.2. Firstly extracted combination with combinations represent bit "0" of watermark is compared, and from this comparison type (sequence) of choice of coefficients is obtained. Finally extracted coefficients are changed using constant "α".

Table 2 Modification of transformation coefficient, when bit of watermark is 0

"0"	Change of combination	Type of choice	C_{W1}	C_{W2}	C_{W3}
1.	LMH	1	C_1-α	C_2	C_3+α
2.	MLM	1	C_1	C_2-α	C_3+α
3.	LLH	1	C_1-α	C_2-α	C_3+α
4.	MHL⇒MLH	2	C_1	C_2+α	C_3-α
5.	HML⇒LMH	3	C_1+α	C_2	C_3-α
6.	HHL⇒HML⇒LMH	3	C_1+α	C_2	C_3-α
7.	MMM⇒LMH	1	C_1-α	C_2	C_3+α
8.	LHL⇒LLH	2	C_1-α	C_2+α	C_3-α
9.	LHH⇒LMH	1	C_1-α	C_2	C_3+α
10.	LHM⇒LMH	2	C_1-α	C_2+α	C_3
11.	HLL⇒LLH	3	C_1+α	C_2-α	C_3-α
12.	HLH⇒MLH	1	C_1	C_2-α	C_3+α
13.	HLM⇒MLH	3	C_1+α	C_2-α	C_3

6.2. Extraction of the watermark

Extracting process for all methods is shown in Figure 8.

6.2.1. Extraction in methods DCT_M1 and DWT_M1

After transformation of image under test and original image we compute differences between coefficients as follows:

$$if \quad i_t(x) - i(x) \geq 0 \quad \Rightarrow \quad \tilde{w}(x) = 1 \tag{3}$$

$$if \quad i_t(x) - i(x) < 0 \quad \Rightarrow \quad \tilde{w}(x) = 0 \tag{4}$$

where $i_t(x)$ is tested coefficient, $i(x)$ is original coefficient and $\tilde{w}(x)$ is bit of the extracted watermark.

6.2.2. Extraction in methods DCT_M2 and DWT_M2

For extraction of watermark, firstly DCT or DWT coefficients using information about type of choice are extracted. Result of this operation are combinations (LMH, etc.), which are compared with combinations from Table 1. If extracted combination is identical with one of first three combinations (for example MLH), then extracted bit of watermark is "0". Extracting process for bit "1" is same. If extracted combination is not in Table 1, then extraction of the bit is failed.

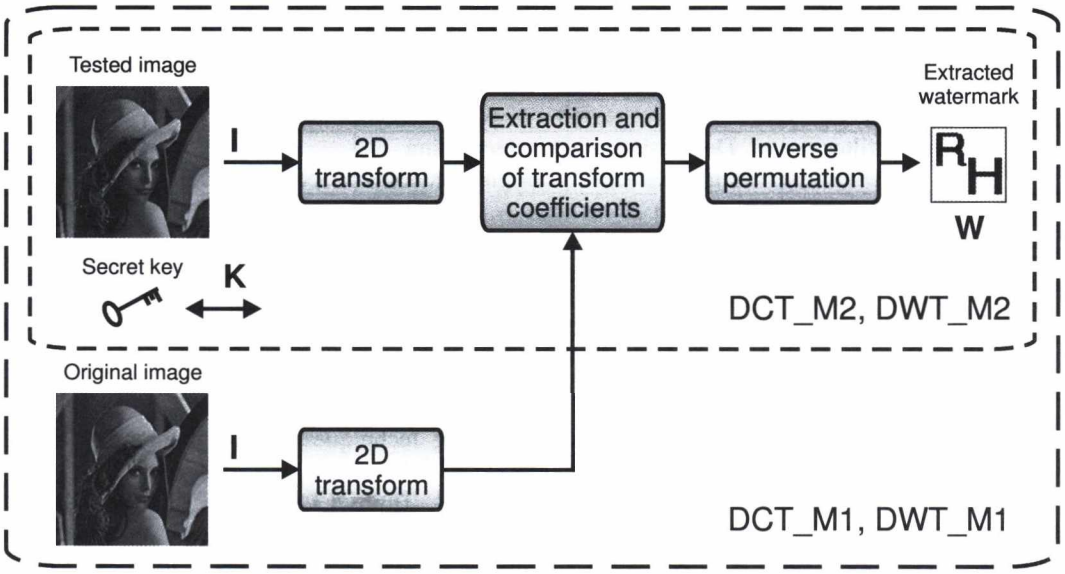


Figure 8 Extraction of the watermark

7. EXPERIMENTAL RESULTS

Verification of proposed methods has been performed on gray scale image "Lena" (256x256) by using digital watermark in form of binary image (32x32). In both DCT methods watermark was embedded into the middle frequency coefficients (Figure 6). In DWT methods, we use coefficients of detail subbands on third level wavelet decomposition (Figure 7).

Experimental results (Table 3) show, that watermarking techniques in transform domain have good robustness to most geometrical attacks except rotation and Gaussian noise.

8. CONCLUSION

In this paper was presented a comparison of two methods for invisible embedding digital watermark into the still grayscale images. Experimental results show, that methods are good candidates for copyright protection.

Our future research is oriented on color images, HVS models and combination of proposed methods. Better robustness is obtained by using combination of proposed methods i.e. using multiembedding of watermarks or consecutive multiembedding also called hybrid watermarking.

Table 3 Experimental results

Method / Attack	Agreement of extracted watermarks [%]				
	PSNR of waterm. image I_W	without attack	Cropping ¼ of image	JPEG Q 50	JPEG2000 0,5 bit/pixel
DCT_M1	45,7 dB	100	87,9	100	70
DCT_M2	47,1 dB	100	87,5	95,8	88,2
DWT_M1	48,1 dB	100	87,8	99,1	69,2
DWT_M2	50,5 dB	100	70,1	90,3	80,6
	"Gaussian" noise $\mu = 0, \sigma = 0.005$	"salt&pepper" noise 0.025	Rotation 1°	Resize	Wiener filter 3x3

DCT_M1	68,3	74,8	63,8	92,8	93,5
DCT_M2	79,6	80,8	78,9	93,1	98,5
DWT_M1	68,3	68,9	66,6	92,5	95,7
DWT_M2	69,53	67,7	65,9	90,3	96,4
	Median filter 3x3	Decreasing brightness -30	Increasing brightness +30	Gama corection 0,5	Stirmark
DCT_M1	90,4	100	100	87,3	61,4
DCT_M2	97,9	100	100	99,6	62,5
DWT_M1	91,0	100	100	82,3	62,8
DWT_M2	93,7	100	100	97,4	71,4

REFERENCES

[1] Katzenbeisser, S., Petitcolas, F.A.P,: Information hiding techniques for Steganography and Digital Watermarking, Artec House,INC., 2000, ISBN 1-58053-035-4

[2] Fridrich, J.: Applications of data hiding in digital images. Tutorial for the ISPACS´98 Conference in Melbourne, Australia, November 4-6, 1998

[3] Petitcolas, F.A.P., Anderson, R.J., Kuhn, M.G.: Attacks on copyright marking systems. Proceedings of the Second International Workshop in Information Hiding, pp. 218-238, Portland, April 14-17, 1998

[4] Hovančák, R., Levický, D.: Comparison of Two Watermarking Methods using Transformation. 12th International Czech-Slovak Scientific Conference Radioelektronika 2002, Bratislava, pp.150-153 , 2002

[5] Hovančák, R., Levický, D.: Comparison of Watermarking Methods using DCT Transformation. 13th International Czech-Slovak Scientific Conference Radioelektronika 2004, Brno, Czech Repulic, pp.403-406 , May 6-7, 2003

[6] LEVICKÝ, Dušan et al. : Digital right management. In: RTT 2005 : Research in Telecommunication Technology 2005 : 6th International Conference, September 12-14, 2005, Hradec nad Moravicí, Czech Republic. Ostrava : VŠB-Technical University of Ostrava, 2005. 4 s.. ISBN 80-248-0897-8.

EXPERIENCES WITH THE REUSABILITY AND VIZUALIZATION IN TEACHING JAVA

Geletka Peter
Kubov Jozef
Samuelis Ladislav Ing. PhD.
Ladislav.Samuelis@tuke.sk

Department of Computers and Informatics, Slovakia
Faculty of Electrical Engineering and Informatics
Technical University of Košice
Letná 9, 042 00, Košice,
Slovakia

Abstract

The aim of this contribution is to describe the progress achieved since the last presentation of the results in the frame of the HUBUSKA project (Second Open Workshop, Budapest, September 2005). The original objective was to demonstrate by visualization the software development steps in creating a case study devoted to the simulation of a simple automatic teller machine. This kernel was later widened by the application of the Jini technology in order to provide robustness to the application. As development progressed further quality checking functionalities were added in order to help grading students' assignments. To sum up, students use the system for obtaining practical experiences of programming in Java language by studying case studies. These are created from the students' assignments, which were submitted earlier. In the next step tutor selects (rates and chooses the best ones) suitable assignments for the inclusion into the pool of case-studies for the reuse in further academic years. Students are free to improve and add new functionalities to the already available case studies. The aim is to let students to observe the previous case studies and to enhance them by their own ideas.

Keywords

E-learning, three-layer architecture, reusability, visualization, pool of case-studies

1. TOOLS AND TECHNOLOGIES

System is built on Java servlets technology [1], which is an efficient and powerful solution for creating dynamic content for the web. The applied Jini technology [2] specifies a way for clients to find the required services on the network and to work together in order to get a task accomplished. This technology allows each service, as well as the entire pool of services to adapt to the changes in the allocation of the network nodes. For storing information about users of the system and students' assignments, the application uses open-source object-relational Database Management System PostgreSQL [3]. Web application is running on the application server Tomcat [4]. On the basis of the clients' requests, with the support of servlets and Jini services, the application generates the answer page.

2. DISTRIBUTING THE ASSIGNMENTS

System distinguishes two types of users in the system – *tutors* and *students*. At the beginning of the course teacher assigns to students suitable assignments from the pool of assignments. Tutor has various types of assignments at disposal. The already available pool of assignments enables to assign case studies for example. As case studies are online analyzed, for beginners it is easier to follow the source code and to adopt the existing code to the intended algorithm. The available code forces students to reuse fragments code and to experiment with partial solutions. Students are not limited, how the application should look like or which Java components to use. Students are required to understand the behavior of the components and then reuse them in robust applications. The understanding and the reusability of the components is stressed. This

approach is also a good experience for students when they see the fragments of the previous applications of students and tutors.

For example: an individual student or a group of students has to design and implement application that simulates cash dispenser. They know its functionalities; they know how it should react on users' requests, who are the users of the application, and the like. It is not prescribed algorithmically how to reach such behavior of the application or what technologies to use. Some students can choose to store information about the users or accounts in the database, but some other students can store it into the XML file or into the standard text file. It is fully in the consideration of the student, which approach student assumes as the most advantageous.

3. SUBMITTING AND COLLECTING THE ASSIGNMENTS

After completing the assignments, they are submitted and then archived in the database. Since these assignments will be later evaluated and processed automatically, they have to be submitted in the prescribed form. The *.jar* (Java archive) file, which is submitted and holds the name of the submitting student must have following structure:

classes – this directory contains the compiled form of all classes of the assignment; they are used to evaluate and run the assignment.
doc – directory contains documentation files
javadoc – subdirectory contains generated *.html* files of the assignments
META-INF – directory contains the *manifest* file.
src – directory contains source codes of the assignment. This can be archived into the pool of the case studies.
readme.xml – this file contains information about the annotation of the assignment

During the submission procedure assignment is checked against the prescribed form. In the case that the structure of *.jar* files is not correct formally, the assignment is not accepted and student receives information which parts of the assignment are missing. If the assignment was complete, it is accepted and its information as well as the file itself is stored into the database.

4. ASSESSMENT AND EVALUATION

Evaluation of the assignment is dual. Tutor evaluates first the assignment. Tutor is able to assess the assignment based on running, considering the used technologies, view and analyze the documentation, or check in what degree are individual assignments similar. To make it easier for the teacher, there is a second way of evaluation. For this purpose was used the open source package **JDepend** [5], which traverses Java class file directories and generates design quality metrics for each Java package. This approach is automatic and provides interesting data for tutors. The final assessment is done by tutor in any case.

On the basis of the **JDepend** package, the teacher has at disposal the following information about the quality of the source code:

- **Number of Classes and Interfaces** – The number of actual and abstract classes (and interfaces) in the package is an indicator of the extensibility of the package.
- **Afferent Couplings (Ca)** – The number of other packages that depend upon classes within the package is an indicator of the package's responsibility.
- **Efferent Couplings (Ce)** – The number of other packages that the classes in the package depend upon is an indicator of the package's independence.
- **Abstractness (A)** – The ratio of the number of abstract classes (and interfaces) in the analyzed package to the total number of classes in the analyzed package. The range for this metric is 0 to 1,

with A=0 indicating a completely concrete package and A=1 indicates a completely abstract package.

- **Instability (I)** – The ratio of efferent coupling (Ce) to total coupling (Ce + Ca) such that I = Ce / (Ce + Ca). This metric is an indicator of the package's resilience to change. The range for this metric is 0 to 1, with I=0 indicating a completely stable package and I=1 indicates a completely instable package.
- **Distance from the Main Sequence (D)** – The perpendicular distance of a package from the idealized line A + I = 1. This metric is an indicator of the package's balance between abstractness and stability. A package squarely on the main sequence is optimally balanced with respect to its abstractness and stability. Ideal packages are either completely abstract and stable (x=0, y=1) or completely concrete and instable (x=1, y=0). The range for this metric is 0 to 1, with D=0 indicating a package that is coincident with the main sequence and D=1 indicating a package that is as far from the main sequence as possible.
- **Package Dependency Cycles** - Package dependency cycles are reported along with the hierarchical paths of packages participating in package dependency cycles.

These features are analyzed automatically in every assignment. Analysis of object-oriented programs involves more much more features and we constrain the analysis only to the above mentioned points.

5. VISUALIZATION OF THE SOURCE CODE

To secure that student will be able to understand and reuse the information that case study offer, source code must be in understandable form. It means that system will show student only that part of source code, which he wants to see. Imagine that student is interested only in the implementation of one concrete method. System is able to choose class in which is the required method and show him its source code. This enables that student has not to bother with searching desired method in whole source code.

Applying the XML technology in the supporting of the visualization is a very powerful and comfortable tool. The situation is as follows. Case studies are stored in database in one *.jar file. Content of this file is described in the previous chapter of this paper. File is than loaded from database and stored on the hard disk. Than the Jini service unpacks *.jar file to separate parts. The source code from the .jar file is converted into XML code (*.xml files). Transformation is provided by package named [6] JAVA2XML. This is an open source Java program that allows conversion of Java source files to XML files. It marks all variables, literals, arguments, method declarations, and so on with XML tags. This allows some interesting new possibilities in the field of storage, manipulation and presentation of Java source files. Why is it necessary? Because we know that students do not respect always principles of correct code writing (e.g. code is not well formatted, contains comments that not according to the javadoc standard, etc.). By conversion into the XML format these imprecisions are eliminated. Source code in .xml file is clean pure code and contains only Java declarations and Java statements.

When source code is in the XML form, it is desirable to convert it into human understandable form. First system generates from *.xml files the package and class tree. Student chooses class that interests him/her and system shows source code of the desired class. But system shows only declaration of a class and the signatures of constructors and methods. Than student can browse through single methods (see Figure 1). Of course students have also the possibility to launch the case studies and to observe the behavior.

6. VISUALIZATION OF THE DOCUMENTATION

As it was already mentioned, by conversion into the XML form all comments are eliminated. So it is necessary to represent them in another way, because non-commented code is not easily understandable. If student clicks on the method identifier, constructor, field or class name, comment in *javadoc* form will

appear. The next Figure 1. shows the graphical user interface (GUI) of the system. This interface is available for students.

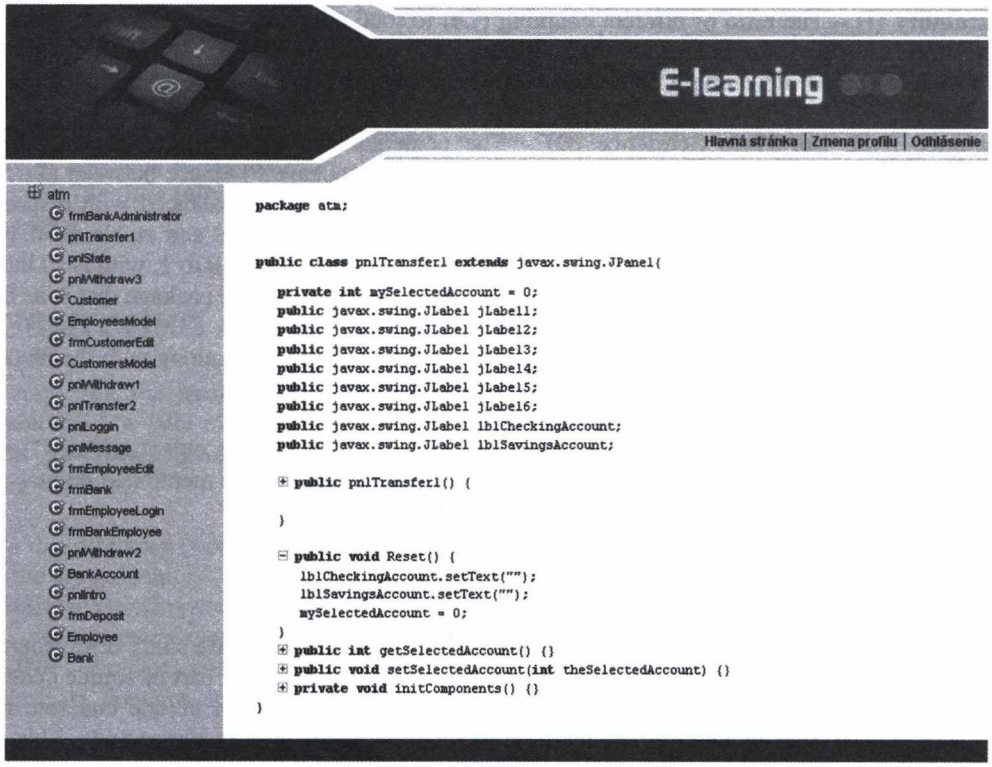


Figure 1. The student's graphical interface

7. CONCLUSION

Both modules together create a compact system. Module for archiving serves for storing case studies in the database and enables teacher easier to evaluate assignments. This helps teacher to decide, whether given assignment will be published in the pool of case studies or not. (Pedagogical issues are dealt in [7].) From the students' point of view the pool of the case studies are fine guides to complete his/her own assignment. He/she can draw upon other students' experience. To sum up, after each incremental actualization, students have at disposal more advanced case studies.

REFERENCES

[1] http://java.sun.com/products/servlet
[2] http://www.sun.com/software/jini
[3] http://tomcat.apache.org/
[4] http://www.postgresql.org
[5] http://www.clarkware.com/software/JDepend.html
[6] http://www.stormpages.com/java2xml/
[7] Pšenáková, I.: Internet v dištančnom vzdelávaní. In: Sborník příspěvků z mezinárodní konference modernizace vysokoškolské výuky technických předmětů. 1. vyd. Hradec Králové, Gaudeamus, 2000, pp. 171-173, ISBN 80-7041-723-4.

The project is supported by the following grants:

- Technologies for Agent-based and Component-based Distributed Systems Lifecycle Support, Scientific grant agency project (VEGA) No. 1/2176/05

- Mathematical Theory of Programming and its Application in the Methods of Stochastic Programming, Scientific grant agency project (VEGA) No. 1/2181/05
- Evaluation of Operational Parameters in Broadband Communicational Infrastructures: Research of Supporting Platforms, Scientific grant agency project (VEGA) No. 1/2175/05.

APPROACHES AND SOLUTIONS FOR PERSONALIZATION IN ELEARNING SYSTEMS

Desislava Paneva
Institute of Mathematics and Informatics, BAS, Bulgaria
8, Acad. G. Bonchev str.
dessi@cc.bas.bg

Mariya Monova-Zheleva, Yanislav Zhelev
Burgas Free University, Bulgaria
62,San Stefano str.
mariaj@bfu.bg
jelev@bfu.bg

Abstract

This paper presents approaches and solutions for eLearning personalization, which are implemented in some contemporary Learning Management Systems (LMSs). The importance of the appropriate user model construction and its relation towards support of the different types of personalization are briefly described in section 1. In the section 2 of the paper is described a solution for role- and competency-based learning customization that uses Web Services approach. A method for development and design of adaptive learning content and on the processes of customization in eLearning in terms of learning strategy system support is represented in section 3. A key requirement of LMSs is personalization of learners' access to learning objects by providing results tailored to the individual or group of learners as the response to search queries. Some aspects of personalization can also take place even before a query is submitted for evaluation. These issues are included in section 4. The last part presents how personalization techniques are implemented in Learning Grid-driven applications.

Keywords

Adaptive Learning Content, Customized learning, Grid technologies, Learning Management Systems, Ontology, Personalization, Web Services.

1. INTRODUCTION

Contemporary eLearning systems support the creation, storage and presentation (often via a web browser) of learning materials in a structured way. eLearning materials are developed as appropriately indexed learning objects (LOs) at the 'course' level of granularity. LOs are generally available and searchable via the Web and are collected in learning object repository (LOR). The 'courses' deal with by LMSs are an aggregation of the different kinds of objects found in LOR. The possibilities for adaptation, personalization, modality, record-keeping on student's performance, and usage statistics for the system as a whole are key components of the LMS functionality [1].

Learning personalization includes how to find and filter the learning information that fits the user's preferences and needs, how to represent it and how to give the user tools to reconfiguration the systems, in consequence, reconfiguration system could be part of personalized environment in some systems.

For instance, WebCT can provide personalised learning paths for users, as access to objects can be conditioned on a wide range of personal data. There would seem to be a lot more scope for individual personalization than this, as the WebCT system records quite a lot of information on the behaviour of the users (which documents they visit, how long they spend viewing them, test results and grades. At present this information seems to be used solely for the production of reports, but could be used towards providing a truly personalised educational experience without the need for large investments of course designers' time.

Xtensis claims to be "a revolution in the management and delivery of e-learning" as it is specifically designed to handle LOs and their (IMS and SCORM-compliant) metadata. It is usable 'out of the box' as a learning management system, but can be configured to reflect the structure of an organisation and is more an architecture than a single product. It is used as the content management system for several UK-based LOR

projects, including the National Learning Network (http://www.nln.ac.uk/), the Seeveaz Key Skills repository and Iconex (http://www.iconex.hull.ac.uk/).
Moodle (Modular Object-Oriented Dynamic Learning Environment) is an open source software package for producing internet-based courses and web sites. Depending on the difficulty level of the learning content and the general level of knowledge and skills of the learners, learning process can be realized dynamically. The exemplary scenario is the following. The capacity for personalization of the environment is subject to be improved further in the next main release of the product.
The user modelling is the process of constructing (often computer-based) users models, while the user model means all the information collected about a user that logs to a web site, in order to take into account her needs, wishes, and interests. Every LMS has its techniques to modelling his users so as to construct the user model or profile. The process of user modelling covers user model as an output, as well as background knowledge, and the user behaviour. Construction of an effective user model and tracking of its continuous changes are a real challenge in contemporary LMS. The quality and expressive power of the user model is crucial in respect to the implementation of intelligent support for different adaptive teaching strategies and their switching during the personalization, of [2] [3] [4]:

- the learning content, based on learner's preferences, educational background and experience, learning content tailored to individual learning style of the user;
- the representation manner and the form of the learning content (for example, learning content in the form of the adaptive learning sequences of learning objects).
- a combination of the previous two types.

Personalization in current Learning Management Systems tends to be concerned with remembering which courses the user is allowed to view and how they like their pages to be presented. In some cases users (learners, teacher and administrators) are able to edit their own profile; to maintain their personal calendar (monthly and weekly) which keeps track of their event transactions; to subscribe to forums, etc. Observing the educational process as a whole, learners are very rarely allowed to get access to learning objects which are conditioned on a wide range of personal data including achievement, date/time and class code.
The following approaches (based on user model) can be used to apply the learning personalization:

- Personalization, controlled by the learner – It requires direct input of the learner's needs and preferences by filling question forms or by choosing options and alternatives.
- Personalization, based upon an existing user profile and meta-descriptions of the information content - In this case, the learners' preferences are stored in their profile.
- Personalization via searching for a correlation between the learners - Correlation is through the values of the attributes, describing the learner's profile. If there is a strong correlation, there is a possibility that the content for a given profile is suitable for applying to its close (adjacent) profiles.

User modelling is a complex and sophisticated one. Application of the semantic web approach to representing student model based on multiple student data with respect to the most important and well-developed learner model standards will help and decrease the difficulty of the process. Ontology is a key Semantic Web technology for marking up electronic resources. Ontologies typically consist of definitions of concepts relevant for the domain, their relations, and axioms about these concepts and relationships.

2. CUSTOMIZED LEARNING AND WEB SERVICES APPROACH

Customized learning, presenting just the right material to the learner on demand, can be described using data representations from learning technology standards (learner profiles, competency definitions, sequencing rules, learning objects). William Blackmon and Daniel Rehak offer a web services-based methodology for customization by profile, specifically one of eliminating LOs from a course [7] because either:
a. Learner's current role does not require the learning objective taught by the LO, or
b. Learner's profile indicates the learner has already achieved the objective taught by a LO.
The learning content and data used in customization are represented in a set of standards-based data models. These are used in a content authoring and delivery process that customizes the activities delivered to the learner based on the learner's role and competencies [9] [10].
Content and learning activity customization uses six sets of data elements (with data representations taken from current learning technology standards):

- *Learning Objects* -- the collection of content and learning resources maintained in a content repository.
- *Content Structure* -- the organization of learning objects in a tree or hierarchical structure.
- *Roles* -- definitions of the job roles of a learner.
- *Competency Definitions* -- definitions of the skills and knowledge acquired by a learner.
- *Learner Information Package* -- the collection of stored profile information about a learner.
- *Sequencing* -- rules used to select content and sequence the learner through a content structure.

The major steps for a customized course preparation and delivering are[1]:

- *Create Course and Content Description* -- describe the course (content structure and set of LOs) and behaviour rules used to express the progression of the learner through the content:
    - Associate role and competency definitions with each learning object by mapping a sequencing objective id (used to label the objective) to a competency definition id or to a role id.
    - Specify the conditional rules used to customize the course by eliminating learning objects from the activity sequence.
- *Establish Learner Profiles* -- specify the role of the learner (which in turn may yield a set of competencies required to perform the role), and contain data on the learner's record relative to each of the specified competencies.
- *Register Learners* -- register the learner for the course.
- *Deliver Course* -- deliver the course, matching the course description to the learner's profile to select content. As the learner completes instruction, the profile may be updated to include mastery of subject matter. Delivery and customization continues until all required activities have been completed.

The customization process has been implemented through a set of web services. Rather than building large, closed systems, the focus is on flexible architectures that provide interoperability of components and learning content, and that rely on open standards for information exchange and component integration. The overall web services architecture for learning is divided into layered services. The layers from top to bottom in this services stack are:

- *User Agents* -- provide interfaces between users (both end user applications and program agents) and the learning services. Agents provide the major elements of learning technology systems: authoring of content, management of learning, and actual delivery of instruction to learners.
- *Learning Services* -- collection of (many small, simple) data models and independent behaviours. Service components are characterized as providing a single function that implements a particular behaviour. Each service is identifiable, discoverable, (de)referenceable, and interoperable. They include built-in security and rights management, and assume an unreliable underlying network. Services are grouped into logical collections, where upper-level services rely on the support from the lower-level services:
    - *Tool Layer* - Tools provide high-level, integrated server applications. Accessed via known, published interfaces, they provide the public interface to the learning tools (tutors, simulators, assessment engines, collaboration tools, registration tools, etc.). User agents and end user applications are built using collections of tool services.
    - *Common Applications Layer* - These are services that provide the commonly used learning functions and application support behaviours used by tools and agents (sequencing, managing learner profiles, learner tracking, content management, competency management, etc.).
    - *Basic Services Layer* - Basic services provide core features and functionality that are not necessarily specific to learning, but which may need to be adapted for learning (storage management, workflow, rights management, authentication, query/data interfaces, etc.).

All services are built on and use a common infrastructure model. The infrastructure layer relies on basic Internet technologies (e.g., HTTP, TCP/IP) to connect service components over the network. The services themselves are implemented using web services bindings. Messaging is done with SOAP; service descriptions are catalogued with UDDI, and described in WSDL - all are XML representations [6]. Overall

[1] Assuming there is a globally defined set of learner job roles and competency definitions

service coordination is expressed in a workflow or choreography language. These standard technologies permit the upper-level services to be implemented in a platform-neutral manner, and provide interoperability across different implementations of the actual learning services.

3. METHODOLOGY FOR DEVELOPMENT AND DESIGN OF ADAPTIVE LEARNING CONTENT

The term "adaptive learning" means the capability to modify any individual student's learning experience as a function of information obtained through their performance on situated tasks or assessments. With the integration of the IMS Simple Sequencing (IMS SS) Specification [10], SCORM [11] allows the learning strategies to be translated into sequencing rules and actions, which are associated with the activities a learning experience consists of. The sequencing rules are based on learner's progress and performance and affect the availability of the learner is allowed to experience.

All learning activities can be associated with sequencing information defined by the content author. In run time, each activity experienced by the learner is associated with tracking status data, which may affect the overall sequencing process. This means that learners with difficulties in satisfying the learning objective should be able to experience additional activities (or repeat some of the activities) to improve their knowledge level and skills. Some restrictions concerning number of attempts and/or period of time for any activity could be set by the content author.

The process of defining a specific sequence of learning activities begins with the creation of a learning strategy for the achievement of the determined pedagogical aim/s. Learning strategy specifies types of learning activities and their logical organization (the activity tree) as well as the prerequisites and expected results for each activity. The rules for managing the instructional flow are the other important part of the strategy. Describing the rules by means of IMS SS elements and attributes the content author transforms the sequencing strategy into strategy for the activity tree traversal management. The author establishes an aggregation of learning objects associating leafs of the activity tree with appropriate Sharable Content Objects (SCOs). The outcome of this process is a content package. The imsmanifest.xml file of the package describes SCOs organization and their sequencing. The implementation of adaptive learning in an eLearning environment could be promoted and facilitated by providing of sequencing templates for the development and design of instructional flows. The following figure describes the main stages of the sequencing loop.

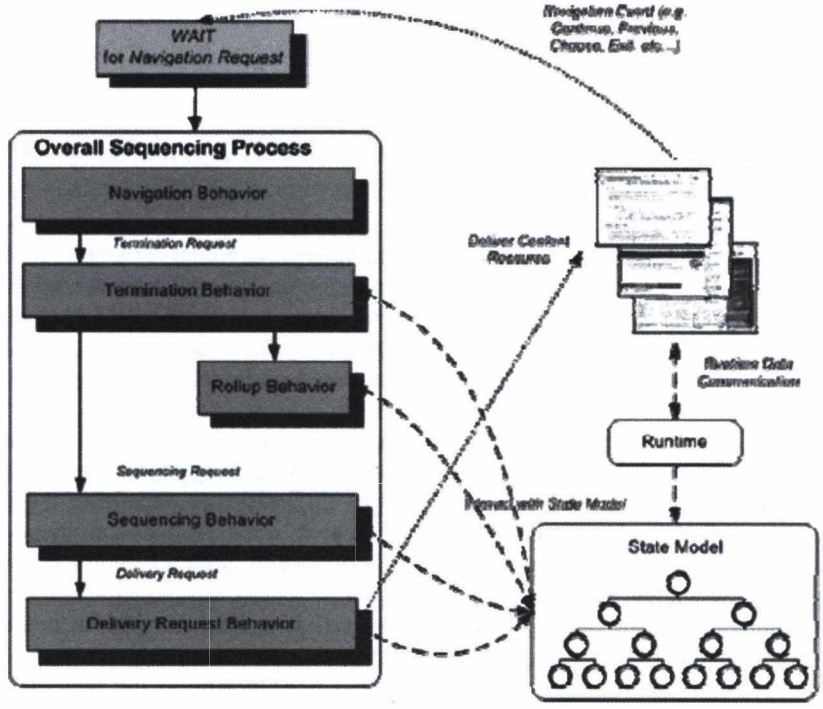


Figure 9: Main stages of the sequencing loop

The sequencing template describes the conceptual organization of the learning content as a sequence of template pages and provides the learning strategy implementation translating it into sequencing strategy. Such sequencing template can be used in different knowledge domains from different instructors who want to follow the described in the package content organization and the implemented learning strategy. In this case, instructor is responsible only to identify (or create) and then to incorporate the relevant multimedia content in each of the template pages accordingly the subject matter of the course taking into consideration the concrete learning objectives and context [12].

Simple Sequencing recognizes only the role of the learner and does not define sequencing capabilities that utilize or are dependent on other actors. The specification includes a limited number of widely used sequencing behaviours. In particular, IMS SS does not address, but does not necessarily preclude, artificial intelligence-based sequencing, schedule-based sequencing, collaborative learning, or sequencing requiring data from multiple parallel learning activities.

The main advantage of the Simple Sequencing approach is that the sequencing rules are described outside the learning objects' content. In this way, the instructional designer can change the rules (i.e. the learning strategy) without any changes in the content or its organization. Nesting manifests of the developed sample packages the content author can developed more complex strategies and content structures.

4. TAILORING TO INDIVIDUAL LEARNING STYLES

In the contemporary learning environments personalization techniques of learners' access to learning objects have to provide results tailored to the individual or group of learners and their learning styles as the response to search queries. When users search for LOs the results returned to them will depend on who they are as well as their query, since different LOs may be more appropriate for different learners. Personalization will have an effect on search results returned from a keyword-based query at three different levels [8]:

- *Filtering* of the returned LOs - excluding those LOs deemed unsuitable for the learner, even though they satisfied the original query;

- *Ranking* of the returned LOs - the 'best' LO for one user may be different from the 'best' LO for another, but personalized ranking means that they can both have the most suitable LO for them returned at the top of their search results;
- *Presentation* of results - users will have different preferences for the display of their search results (e.g. display results as trails or as a simple list, display 10 results per page or 50 results per page).

Some aspects of personalization can also take place even before a query is submitted for evaluation: personalized queries can be constructed using information stored in the profile, by re-formulating or annotating the user's original query to reflect elements of their profile. The user profile has to contain information about preferences, aims, and educational history that can be used by the system. This is the first stage of filtering.

Keyword-based query is not the only way that users can locate LOs – the schema of the LO descriptions can also be browsed to find relevant LOs, providing facilities such as 'browse by author' and 'browse by subject'. Personalization of the browsing process can occur at two levels:

- Allowing users to restrict the information they see to only those attributes of interest to them, organised in their preferred manner.
- LMS can use knowledge of a user's preferences (either those explicitly supplied by the user or those learned by the system itself) to recommend individual LOs or categories of LOs to the user as they are browsing.

4.1. Filtering and ranking search results

The query service will return a set of LO descriptions - all those LOs that satisfy the user's query. The user wants to be able to find exactly the right LO quickly, without having to browse too many of the results, so rather than present the results exactly as they are returned by the query service some processing is done first.

If a profile of the user is not available (or the user has personalization turned off) then all that can be done at this stage is some rudimentary ranking of the result set, possibly using standard ranking techniques from information retrieval and web search.

However, we anticipate that usually some minimal profile will be available to the system, as users should supply at least some minimum information into their profile when first registering. In this case the ranking of LOs will involve personalization. This means that the system can attempt to show the user only those results likely to be most relevant to them personally, as well as relevant to the query in general.

The first step in this processing is to filter the results - remove all those LOs that we are certain will be of no use to the user. At this stage, for example, any LOs in languages that the user does not understand can be eliminated, as can those not meeting accessibility requirements, those at a far too high or low level of difficulty and possibly those covering only material that the learner is already completely familiar with.

Next, the remaining set of LO descriptions must be ranked in order of relevance to the user. Whereas filtering can be done with just the user profile, ranking a set of results should take the original query into consideration too (i.e. relevance must be judged against the combination of user profile and query, not just the profile).

The best algorithm to use for this ranking is still an open question, but it will take into consideration:

- Relevance of the LO to the query;
- How well the LO caters for the user's accessibility requirements;
- Whether the user has the prerequisite knowledge and experience;
- Matching between the user's goals and the learning objectives of the LO;
- If the user's learning styles are those catered for by the LO;
- If the user is likely to prefer it for other reasons (it is by a preferred author, say);
- The user's most recent activity.

The clear individual semantics of each section of the user profile allows focussed matching against relevant sections of the LO descriptions. For a LO to be a „good" LO for the user, the greatest possible number of different elements will match to some degree. Clearly, though, some factors are more important than others to the user and a good algorithm for combining them will reflect this. For example:

- If LO X caters for one of the user's learning styles but is not very relevant to the original query then other, more relevant LOs should be ranked higher even if their descriptions don't list one of the user's learning styles;

- If LO Y has a learning outcome that matches one of the user's goals but is far too difficult for the user to tackle (they have none of the prerequisite knowledge, say) then again other LOs (closer to the user's level) should be ranked higher.

With so many factors to take into consideration, discovery of which algorithms work better or worse for which groups of users requires much further work and testing, and is beyond the scope of this project. It may be that the ranking algorithm itself needs to adapt to the individual, and will differ from user to user (an additional section could be added to the user profile to store information about parameters used by the ranking algorithm).

4.2. Support for browsing as a trail

As the user is browsing LOs the trails and adaptation service can actively recommend the next LO to look at, effectively generating trails of length two (i.e. a trail consisting of the current LO and a suggestion for the next one) at every stage of the user's browsing, based on the user profile.
The recommendations can be derived in several ways:

- from the semantic relationships between the current LO and other LOs in the LMS repository;
- from the user's profile plus LO metadata - perhaps suggesting LOs that cover more advanced material on the same topic, and also suit the user's preferences (learning style, accessibility, etc.);
- through a process of collaborative filtering, suggesting as the next step a LO that other similar users browsed after seeing the current LO (where similar users can be identified by having similar preferences or similar histories of LO access).

5. PERSONALISATION TECHNIQUES IN LEARNING GRID-DRIVEN APPLICATIONS

The philosophy and the approach behind Grid technologies [9] show the right characteristics for achieving an effective learning. Indeed, they allow to access and integrate the different technologies, resources and contents that are required in order to realise new paradigms in eLearning. They are the most promising approach to realise an infrastructure that will allow learning process actors to collaborate, to take part in realistic simulations, to use and share personaliselly high quality learning data and to innovate solutions of learning and training. Grid will be able to support learning processes allowing each learner to use, in a transparent and collaborative manner, the resources already existing on-line, by facilitating and managing dynamic conversations with other human and artificial actors available on the grid, etc.
The SeLeNe (Self eLearning Networks, http://www.dsc.bbk.ac.uk/selene/) project was funded as an EU FP5 Accompanying Measure (IST-2001-39045) running from 1st November 2002 to 31st January 2004. This project was part of action line V.1.9 CPA9 of the IST 2002 Work Programme, contributing to the objectives of Information and Knowledge Grids by allowing access to widespread information and knowledge, with eLearning as the test-bed application. The developers conducted a feasibility study into using Semantic Web technology for syndicating knowledge-intensive resources (such as learning objects) and for creating personalized views over such a Knowledge Grid. A self e-learning network consists of web-based learning LOs that have been made available to the network by its users, along with metadata descriptions of these learning objects and of the network's users. The architecture of the network is distributed and service-oriented. The personalization facilities include: querying learning object descriptions to return results tailored towards users' individual goals and preferences; the ability to define views over the learning object metadata; facilities for defining new composite learning objects; and facilities for subscribing to personalised event and change notification services.
ELeGI (European Learning Grid Infrastructure, http://www.elegi.org/) is an EU-funded Integrated Project that aims at facilitating the emergence of a European GRID infrastructure for eLearning and stimulating research of technologies to enhance and promote effective human learning. The project is supported by the European Community under the Innovation Society Technologies (IST) programme of the 6th Framework Programme for RTD - project ELeGI, contract IST-002205.
ELeGI promotes and supports a learning paradigm shift focused on knowledge construction using experientials based and collaborative learning approaches in a contextualized, personalized and ubiquitous

way. This new paradigm is based on a learner centred approach, to replace the classical, content centred approach to learning. The learner plays an active and central role in the learning process. Rather than stressing the memorization of information, learning activities are aimed at aiding the learner in the construction of an autonomous, functional base of knowledge and skills. In keeping the learner at the centre of the learning process, personalisation/individualisation (creating and adapting learning paths according to learner's previous knowledge, preferences, skills, preferred learning style), and collaboration (with other students, teachers, tutors, or experts) become relevant aspects to be supported by technologies through the creation of the appropriate context. Considering humans at the centre, learning is clearly a social, constructive phenomenon. It occurs as a side effect of realistic simulations, interactions, conversations, collaborations and enhanced presence in dynamic Virtual Communities.

Project DILIGENT (Digital Library Infrastructure on Grid Enable Technology, http://www.diligentproject.org/) [5] is an integrated project funded in part by the European Commission FP6 IST Programme. DILIGENT is aimed at the creation of virtual digital libraries on the basis of grid-based infrastructure so that the integration of metadata, personalization services, semantic annotation, and on-demand availability of information collection and extraction to be supported. Such new decentralized and service-oriented architecture for digital library assures a better and adaptive tailoring of the content and service offer to the needs of the relevant community as well as to the current service and content offer, and a more systematic exploration of existing resources.

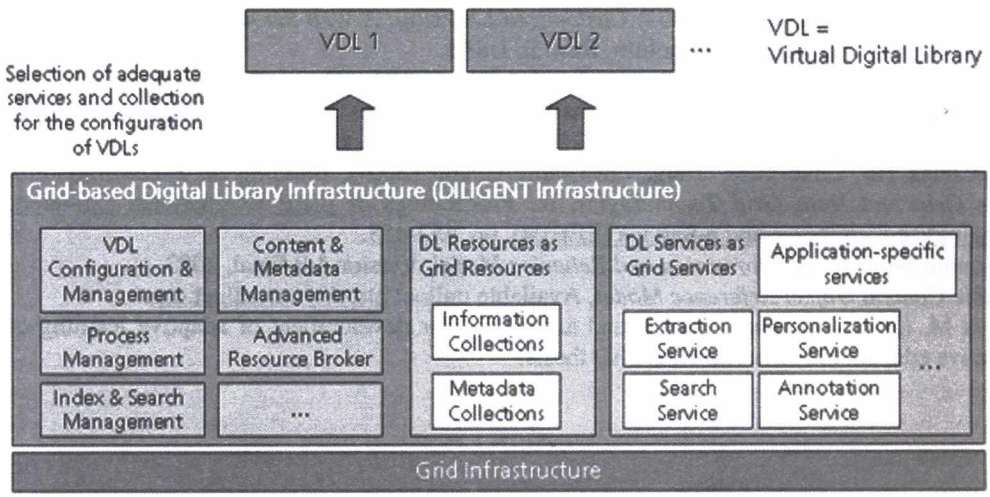


Figure 3: Grid-based digital library infrastructure

6. CONCLUSIONS

Important requirement for each modern LMS is the learning adaptation to be assured for each learner in respect to her/his necessities, preferences, needs, performance, and progress. The achievement of interoperability and content reusability in the existing diversity of software and hardware platforms is a real challenge. One big limitation of the web-based interaction is the smaller communication bandwidth than traditional face-to-face interaction. The term bandwidth represents the amount of information that can be transferred in a unit of time through any means possible. In the face-to-face communication mode, if a verbal instruction is not understood, the clue can be available to the counterpart through gestures, group dynamics and other such means, but the clues in the web-based mode are not always so clear and in many cases not available at all. Therefore, tailoring the information to the right-level for the receiver to understand and integration of different appropriate methods for learning adaptation are crucial factors for the success of any LMS.

REFERENCES

[1] Pavlov R., Dochev D. (2004), *New Information Technologies and Interactive Environments for Vocational and Life-long Learning*, Analytical study, ICT Development Agency, Sofia.

[2] Zheleva M. (2005), *Design and development of Intended Instructional Flows in Web-based Learning Environments*, In: I. Simonics, R. Pavlov, T. Urbanova (Eds.) "Technology-enhanced Learning with Ubiquitous Applications of Integrated Web, Digital TV and Mobile Technologies", Proceedings of the HUBISKA Open Workshop, 6th eLearning Forum, 9-10 June 2005, Budapest, pp. 49-60.

[3] Graziano A., Russo S., Vecchio V. (2003), *Metadata-based Distributed Architecture for Personalized Information Access,* Proceedings of the European Distance and E-Learning Network /EDEN/ Annual Conference "Integrating Quality Cultures in Flexible, Distance and eLearning", June15-18, Rhodes, Greece, pp. 66-72.

[4] Gibson, D., Knapp, M., & Kurowski, B. (2002) *Building responsive dissemination systems for education with the semantic web: Using the new open-source "liber" application*, Proceedings of the EdMedia 2002 conference, Montreal, Quebec.

[5] Paneva D., Pavlova-Draganova L., Draganov L. (2005), *Digital Libraries for Presentation and Preservation of East-Christian Heritage,* Proceedings of the HUBISKA Second Open Workshop, 14 September 2005, Budapest, pp. 75-83.

[6] Samtani G., Sadhwani D. (2002), *Web services and application framework working together*, In: Journal "Web Services Architect", March, Available online: http://www.webservicesarchitect.com/content/articles/samtani04.asp

[7] Blackmon W. H., Rehak D. R, (2003) *Customized Learning: A Web Services Approach*, Proceedings of the EdMedia 2003 conference, Honolulu, Hawaii, USA.

[8] Paneva D., *Some Approaches for Personalization in Learning Management Systems*, In D. Dochev, R. Pavlov (Eds.) "e-Learning solutions – On the Way to Ubiquitous Applications", Proceedings of the Joint KNOSOS-CHIRON Open Workshop, 26-27 May 2005, Sandanski, Bulgaria, pp. 65-74.

[9] Hsing-Chuan Ho, Chao-Tung Yang, Chi-Chung Chang (2004), *Building an E-Learning Platform by Access Grid and Data Grid Technologies*, In: Proceedings of IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE'04), pp. 452-455.

[10] *IMS Simple Sequencing Information and Behavior Model,* Version 1.0 Final, 2003.

[11] *Sharable Content Object Reference Model*, Available online: http://www.adlnet.org.

[12] Zheleva M. (2005), Methods, models and algorithms for development of adaptive learning content in interactive educational environments, PhD thesis.

EU FP6 SSA Project INCO-CT-2003-003401

Methods and tools for development of semantic enabled systems and services for multimedia content, interoperability and reusability

ISBN 3-9500593-7-7