Media and Service Integration for Professional E-Learning

Klaus P. Jantke	Martin Memmel	Oleg Rostanin	Bernd Rudolf
DFKI GmbH	DFKI GmbH	DFKI GmbH	micronomics GmbH
66123 Saarbrücken	67663 Kaiserslautern	67663 Kaiserslautern	10969 Berlin
Germany	Germany	Germany	Germany
jantke@dfki.de	memmel@dfki.de	rostanin@dfki.de	b.rudolf@micronomics.de

Abstract: When academic e-learning projects are abandoning the ivory tower of higher education, this is not only changing their audience. The needs of corporate, government and healthcare result in high demands on those e-learning system's and service's contents, structure, appearance and performance. There are key technologies supporting the step from academia into the wild. The case study of the DaMiT system's transformation from an academic project into an industry service is taken as a basis for discussions of particular technologies that allow for a new quality of media integration including high resolution offline video and online learner modeling as well as a new quality of service integration including adaptivity to context, learner progress and learning goals.

The Domain, the Approach, the System, the Service – Introduction and Overview

The present introductory chapter is intended to provide a brief, but sufficiently comprehensive overview of the project on which the present paper is based upon.

All teachers know it from their daily work at schools, all professors know it from their regular lectures and seminars, and even most of us in e-learning did recognize it—teaching and studying does depend very much on the domain. There are not so many universal concepts all of us can adopt without any necessity for adaptation. On the one hand, the authors would like to present their previous work as lucidly as possible, and for this purpose, peculiarities of the work have to be outlined in some detail. On the other hand, the main intention of a paper presentation at a leading conference like this, naturally, is to deliver general insights, to publish generic concepts, to report experience valuable to others in the area, and to stimulate an engaged discussion. So, we need to find and to exhibit the general in the particular.

DaMiT was a research and development project sponsored by the German Federal Ministry for Education and Research (BMBF) for 30 months. The project's practical goal was to provide an e-learning system for the domain of knowledge discovery & data mining; DaMiT abbreviates "Data Mining Tutor". This system is ready now and available under `http://damit.dfki.de` under the link "Lernsystem". It has been entering the market on the fair WEITERBILDUNG 2004, May 4/5 2004, Frankfurt/M., Germany (see http://www.hrforum.de and use the keyword search for terms like 'Data Mining' or 'e-Learning', for example). First commercial courses are booked.

Data mining is the subject of the DaMiT system. Many vendors and even some scientists claim that data mining deals with the discovery of formerly unknown, but extremely valuable knowledge in large and, perhaps, distributed databases. It is described as a search for "golden nuggets" and, as such, it is suggested that the nuggets were already there, you just had to find an appropriate tool and to liberate them. There is an obvious motivation behind such a view of data mining: selling technologies and/or tools as silver bullets. But this approach does not work, in general. In contrast to the view of data mining sketched above which is glossing over the difficulties users are frequently facing, data mining is more realistically seen as a creative process of building mathematical models over given data. The crux is that those models are expected to have predictive power over future data that are widely unknown when the models are constructed.

In its right perspective, data mining is an interactive process of creating mathematical models of given data such that these models demonstrate a valuable predictive power over other data not available at model generation time. The criteria of success for models under construction are not available at construction time. Thus, data mining is somehow inductive by nature. There is a well-known problem in data mining denoted by the topical term "overfitting".

Models that fit given data too closely have a tendency of delivering insufficient forecast on future data. Therefore, one has to generate models fitting the given data well, but not too well, to avoid overfitting. This is obviously a more ore less vague goal—data mining is not a thoroughly mathematical discipline. From the authors' point of view, data mining is both a science and an art. Thirty years ago, Donald Knuth said in his famous Turing Award Lecture: *Science is knowledge which we understand so well that we can teach it to a computer; and if we don't fully understand something, it is an art to deal with it. ... we should continually be striving to transform every art into a science: in the process, we advance the art* (cf. ACM 1987, pp. 36/37). This perfectly fits the situation of data mining and clarifies the problems we are facing when dealing with e-learning in this domain.

DaMiT is an Internet-based intelligent tutoring system to study the science and to experience the art of data mining. Explorative learning and learning by doing are inevitable, especially concerning e-learning in business environments (cf. Jantke, Lange et al. 2004a, Jantke, Lange et al. 2004b), though in-depth studies of the mathematical background, like probability theory, statistics, discrete mathematics and algorithmics, are substantially supported as well.

For the purpose of the present introductory chapter, a glance at a few features of the DaMiT system should suffice. Take the interactive applets for *explorative learning*, for instance: Students using those applets in a blended learning environment may get introductions of how to use those applets and which phenomena to be investigated during a lecture or a seminar. Even without such an introduction, the communication environments of universities usually provide enough opportunities for students to discuss their needs in using applets of a system like DaMiT. Students are rarely left to their own devices. In contrast, using the DaMiT system in an enterprise, governmental institution or healthcare working environment, the focus is usually much more narrow and there is much less time available than during a university course. Applets of the DaMiT system have to be presented differently. The system has to be highly adaptive (cf. Jantke, Rostanin et al. 2004). For the DaMiT system, the authors have developed a series of video films explaining an appropriate usage of certain applets by means of truly interesting case studies. These videos have proven to be very helpful, but there is a serious difficulty: Some of the videos are of considerable size. Let us briefly touch *learning by doing*, as a second instance: On the one hand, to strongly motivate students and practitioners, the DaMiT system does offer complex case studies and contains exercises with realistic data. These comprehensive exercises, on the other hand, provide an ideal playground for really doing data mining according to the aim of experiencing the art. Doing data mining does include the access to huge data sets and the availability of appropriate tools. The system is providing all this; however, it still faces bandwidth problems as mentioned before. Furthermore, pedagogically and methodologically well-organized learning by doing needs scenarios of evaluating the learners' results and providing feedback. This, in turn, includes the need for upload services and standards, which allow for a widely automatic processing of learners' submissions. We do arrive at XML standards for data mining.

To sum up briefly, the peculiarities of the domain of data mining are bringing with them several implications. First of all, the attractiveness of the domain, in general, and of the DaMiT system, in particular, does enable the step from the ivory tower of academia into the wild world of business, governmental institutions, and healthcare. Second, this step is raising a number of requirements concerning the content, the appearance and the performance of the system including IT security issues. To meet these requirements, the DaMiT system does offer a service of a sashaying integration of online and offline functionalities.

Last but not least, the DaMiT system and service has a few peculiarities, which, perhaps, distinguish this approach from most other current e-learning services in the Internet. *Pars pro toto*, two shall be briefly mentioned here. First, there is a lot of valuable content of the DaMiT system that may be of commercial interest to a wider audience. Though all this content is freely available to registered students, there is some barrier protecting this content from unconstraint anonymous users. The DaMiT system is equipped with some micro payment system. Several pages are inaccessible without payment. This is adding another good reason for booking a course with the DaMiT system or getting registered within regularly scheduled e-learning activities. A second peculiarity is DaMiT's high security standard. Registered learners are getting software certificates. DaMiT is running a Public Key Infrastructure, and students' registration has to go through DaMiT's registration authority according to the standards of IT security. This allows for a secure login and for encrypted communication. Furthermore, every learner's certificate enables the learner to digitally sign documents and to load up those documents in a secure way. The technology of the DaMiT system is setting the stage for a new generation of curricular implementations of e-learning. By digital signatures, non-repudiation as a key requirement of using students' uploads for assigning credit points and the like is given.

The Aim of the Paper and of its Related Conference Presentation

In (Self 1992), the author has been critically reviewing the state of publications about e-learning ending up with a classification of papers that nowadays, a dozen years later, seems still relevant. We are repeating this categorization (citations from (Self 1992) are in italic) and try to explicate how the present work is intended to overcome the weaknesses mentioned, i.e., to escape from the categories John Self has identified:

- *descriptions of partially implemented systems not susceptible to any kind of empirical evaluation* – our underlying system DaMiT is in daily use and is currently entering the market of real-world corporate training;
- *discussions of principles or 'slogans' which are so vague that they can be bandied back and forth forever* – our requirement of integrating media types and services may sound like that, but details down to the implementation shall exhibit that this is now reality and under continuous investigation and evaluation;
- *ridiculously ambitious research proposals which are quite incapable of implementation in the foreseeable future* – hm, well, our work on learning goal induction is surely quite ambitious, but first steps do work.
- *empirical studies which do not reveal which aspects of designs are responsible for the success or failure of implemented systems* – this, at least, does not apply, because the present paper does not systematically report about our system's application (what we tell about the DaMiT system in use is intended only to illustrate the key conceptual and technical issues under investigation).

Being aware of these problems, the authors are aiming at convincingly demonstrating their approach to e-learning, their implementation in the DaMiT system, and the implementation's use. Emphasis is put on own solutions to deeply dovetail online and offline content and services. The presentation leads from motivations through derived functionalities to technologies for implementing these functionalities to meet the motivating requirements.

Intelligent Tutoring Systems – an Artificial Intelligence Perspective

Recently, some authors in e-learning advance from the investigation of e-learning tools to e-learning assistants, as nicely expressed in (Santos et al. 2002): "ITSs are a class of Artificial Intelligence systems that act as assistants in the teaching-learning process." In the authors' opinion, this is just an instance of the urgently needed transformation of current IT systems from more and more overwhelmingly complex tools to assistants proactively supporting the human user. Especially human learning, in general, is an ambitious task that may easily become lengthy, exhausting, boring or even frustrating. Humans do depend on assistance in learning which, at least nowadays, is mostly provided by other humans like, e.g., teachers, classmates, friends or parents. The ultimate goal of Artificial Intelligence research is to develop e-learning systems towards becoming those assistants.

Currently, machine intelligence is still remarkably limited. In (Lelouche 1999), the author is correctly bemoaning that an e-learning system "should know the taught material". He continues: "This capacity is commonly demanded from human teachers, and strangely enough most present computerized systems, and all commercial systems, do not exhibit it." There are, in fact, many more features of intelligence missing in today's e-learning systems in use.

The DaMiT approach is not yet far-reaching enough to resolve all these deficiencies, but it is intended to provide a little contribution to our overall endeavor towards computerized teaching assistants (Bull et al. 2003). A key functionality of the system's intelligence the authors are aiming at is *adaptivity* to contexts and learner needs. The e-learning system's adaptivity shows in the selection of content and in its presentation on the client computer, in the way media types are combined and offered, and in the collection of services available to a particular student.

Adaptivity includes the dynamic generation of the material according to dynamically changing needs. It is strongly motivated by didactic reasons, but also ultimately required in business applications, because of different roles of users and different positions in a management hierarchy imply varying learning needs and different ways of dealing with an e-learning system's content. Adaptivity even influences the different appearance of video material. Sometimes, it is desirable to have a long and detailed film that contains in-depth explanations and is equipped with a variety of links into up-to-date server-based data. In contrast, sometimes a very condensed version of the same film is needed. Adaptivity means to generate those variants from learning objects that are movie building blocks on demand. All this is available in DaMiT.

The present paper is leading from the motivation to technologies like metadata concepts, presentation generation functions and technologies to dovetail online content and offline content of large volume prepared on DVD.

Didactics and Curricular Implementation Requirements – Media and Service Integration

The authors do not believe in simply distributing knowledge over networks, i.e., encoding at the one end, pumping through the network, and decoding at the other end. In contrast, they believe that in complex domains knowledge results from creative processes and is constructed to some extend through the learner's activities. This does apply, particularly, when learners are facing problems of *studying a science and experiencing an art*. In DaMiT, therefore, *learning by doing* and *explorative learning* play a crucial role. Under the offers for explorative learning, those are particularly exciting in which the learner can turn the learning scenario upside down by posing problems to the DaMiT system and watching the system in it's attempts to solve the task. The learner can modify the tasks to make them more difficult or easier to the system. As a side effect, the learner acquires a keen of feeling for the type of problems under consideration and learns many things implicitly that are not easily expressed in a declarative way.

Let us have a little closer look at such a case of explorative learning. There is an applet for learning decision trees over regular patterns—an exciting practical area that has found fruitful applications in the pharmaceutical industries, hence, also a case from the DaMiT system for problem-driven learning with cutting-edge technologies. Because it is not easy to formulate interesting problems that really drive the system into difficulties and that allow to explore interesting phenomena, a video is offered (further remarks on offline videos will follow) exemplifying the use of the applet. To see all the important details (see Fig. 1), the video must be of high resolution and full-screen. Furthermore, the video is equipped with several interaction features that allow, for instance, to stop the video and to go into the online e-learning system for background information. There is an obvious need for dovetailing offline and online content. This media integration can be provided dynamically with respect to contexts and the learner's current progress in learning. Thus, integrated services can be either offered or dropped adaptively.

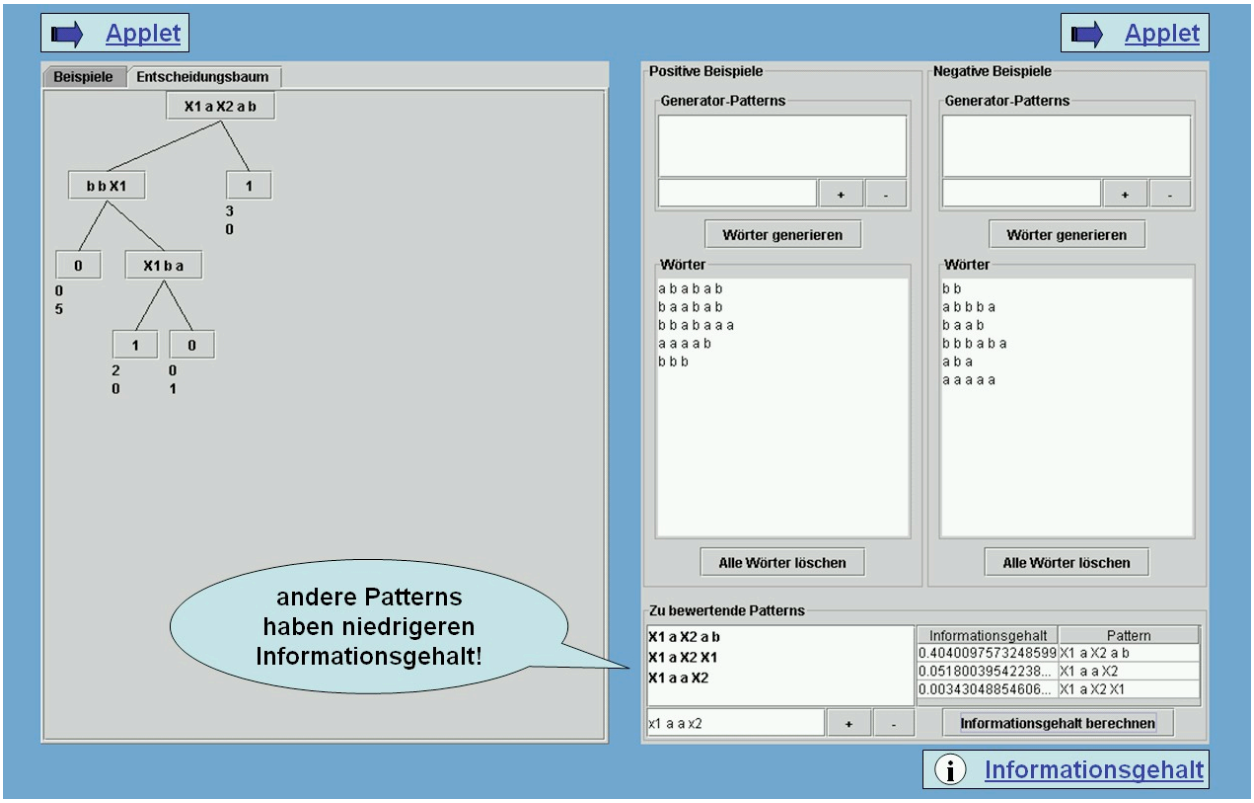

Figure 1: Snapshot of an introductory video of an exploration applet; the video is overlaid with interaction widgets

According to a student's role, to current learning goals, to the student's state in the learning process and to other dynamic parameters, different opportunities may be offered. A variety of system appearances and media integrations do reflect different service combinations that may be provided. Curricular requirements, e.g., may require secure and non-repudiatable uploads. But this does only apply to those students who own a digital certificate.

Quality and Variety Requirements – Media and Service Integration

Training in the corporate, in governmental institutions and in healthcare has to meet high quality expectations, an issue frequently underestimated by academic researches who are proud of putting most emphasis on deep content and considering the appearance of their work only of secondary importance. Offers to the business world have to be attractive and appealing meeting the audience's standards they are used to. The authors have some experience in bringing high-resolution video to work in enterprise training using the technology of the DVDconnector®.

Beyond the need for an attractive appearance, there are more substantial necessities resulting from disciplinary requirements. In demonstrations of software systems, interfaces should appear in a realistic size. Text on buttons, in pull down menus and the like must be readable. Objects handled by humans appearing in a video have to have a size sufficient to identify them and, when necessary, to distinguish them from each other. Documents and document structures known from the daily work should be recognizable when occurring in a video. The video screenshots seen in the two snapshots of the following figure are taken from a full screen video in use for management training within a leading German retailer enterprise.

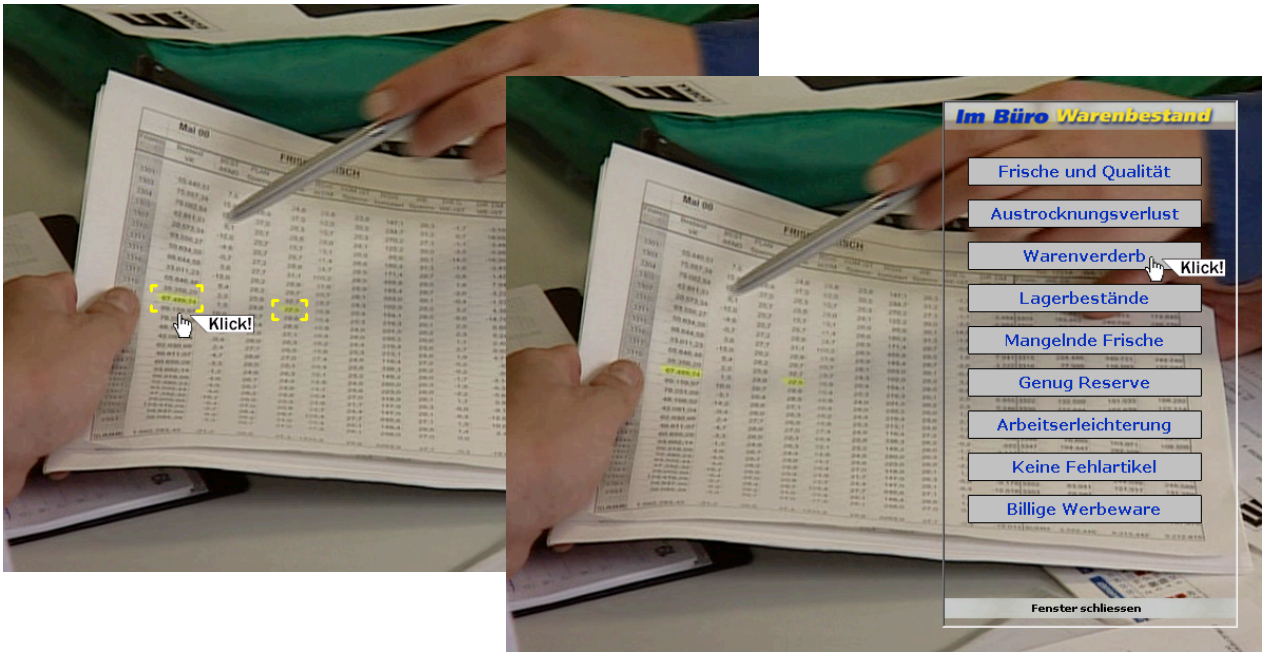

Figure 2: The offline video dovetailed with online content via hotspots and links in a transparent HTML page

The DVDconnector technology allows for dovetailing offline and online content. In the case under investigation, large size high-resolution videos (and much more material) are delivered on DVD. (Fig. 2) is displaying two variants of dovetailing the offline material with a server-based e-learning system. In the left screenshot, hotspots are placed over critical figures. Learners can click the hotspots to get into the online material of, e.g., recent information about those figures. Under a certain learning context, a transparent HTML page as shown in the right hand side screenshot is overlaid to the video. Clicking a link leads the learner to an in-depth discussion of particular problems involving an extra human-machine interaction to collect information about the current learning state.

On the one hand, this media integration's necessity derives from the enterprise's requirements. On the other hand, it allows for the integration of services by providing system functionality like tracking the learner's progress, e.g., that were not available when simply watching a video, otherwise.

Last but not least, a media integration as sketched above provides another service—adaptivity of offline material to learners according to information recorded and processed online. With respect to learner roles, learning goals and other content of the learner model, the appearance of video material can be adapted dynamically. E.g., videos may be assembled from different building blocks and overlays may be shown or hidden according to changing needs.

Concepts, Architectures and Technologies

Content structuring and annotation

If one takes a look at learning material, one will always find a lot of relations, links and cross references. Some content is prerequisite to understand other material, examples are useful for different purposes, some keywords are explained in different sections and so on. In a book, the learning material is or should be, at least, arranged in a reasonable way—sequentially, of course—to enable the reader to understand it. But if one is interested in something not presented on the current page, the book's index, a bookmark—or even other sources—have to be used to find the respective information.

In contrast to a book, e-learning gives the great opportunity to use a huge multimedia repository—whether it's online or offline—and to adapt to the user's individual needs and preferences. The learning material no longer has to be presented in a static way. If there's more than one way to do it, there's no more need for restriction to one way. The system can try to propose next steps, and presentation style or difficulty can be chosen by the user or be automatically selected by the system. To generalize: All relations and links existing within the learning material can be used to help the user and to create or offer individual scenarios.

To realize this opportunity, the content must be structured into relatively small fragments—called learning elements—which can then be combined into bigger learning objects in the preferred way. The learning elements and objects have to be annotated with adequate metadata to provide information about relations to other objects, technical prerequisites, presentation style, the interaction between on- and offline content and so on. The metadata concept used in DaMiT is a derivative of the IMS Metadata Standard (see http://www.imsproject.org); some parts were left out, and some substantial extensions turned out to be necessary.

```
- <unit multi="embedded,illustrated">
    <general name="mot_alfie" description="A Motivation Video for Information Extraction and Algorithmic Learning"
      keyword="information_extraction,wrapper_induction,aefs, formal_language_learning" />
    <lifecycle author="SB_KJ" status="final" />
    <classification type="dvd_illustration" />
  - <contentextern variant="embedded">
      <general name="chapter1" />
      ...
    </contentextern>
    ...
  </unit>
```

Figure 3: Metadata of some video which may be dynamically generated upon 15 learning elements named chapters

Roles and rights

Every user of an e-learning-System has individual needs, motivations and aims. For example, a user in medical healthcare surely needs different information than a user in a governmental institution. Furthermore, not only learners deal with the system: There are authors, tutors, administrators etc. Thus, a professional e-learning system must be able to manage different types of users with different interests. In the DaMiT system, the concept of *roles* and *groups* is used. Each role means different rights and functionality, whereas the group determines which content will be offered, which calendar information and messages will be shown.

In DaMiT, there are five different roles: **anonymous learner** (guest login with reduced functionality), **standard learner**, **teacher**, **content provider** (for authors of learning objects) and **administrator**.
Each learner is member of at least one group, and each teacher is responsible for a group. He decides which users belong to the group, which content will be made available, he manages group calendars, and so on.

Payment

There are different kinds of payment one can think of in e-learning systems, especially in non-academic scenarios. A lot of effort has to be put in the development and maintenance of the system itself as well as in the content provided by the system. So it's only natural that there is non-free content for which some users have to pay. For this purpose, DaMiT is equipped with some micro payment system, allowing the administrator or the teacher to fix the cost of the content provided by the users of his group.
An author can only decide whether his learning element shall always be available for free or not. A teacher who wants to offer such a learning element to his group will always have to contact the author to clarify the payment

issues. If the content is free for members of a group or if the user already paid, then the content will be presented to the user without any delay. Else, the payment component will be activated and the user can choose whether to buy the element or not.

Information Model

The authors strictly separate two levels of the information model—the lowest level containing the smallest pieces of the learning information (learning objects) and the level of views that bind these pieces of information into autonomously sensible units. The units in their term are combined into learning objects that determine the structure of the learning module. These views are user-dependent and serve as the base of the system adaptivity.

There is introduced the sub-layer of the XML views over the sub-layer of the database views. XML views are generated by stored procedures. They allow to abstract from the database logic and structure. Thus, the application logic deals only with XML structures and does not depend on the database (Rostanin 2004). The bottleneck inherent in such an approach is the problem of portability between databases because, in this case, the database logic must be rewritten completely (stored procedure syntax is different in different databases). However, there are the following advantages: *effectiveness*—stored procedures are effective for accomplishing the complex database tasks; one can also use specific database features, so that problems be solved in a way that is optimal for the particular database; *abstraction* from the lower level database structures—this allows to develop the database and application logic separately.

Chosen Application Architecture

In order to satisfy the requirements on the system's scalability, the standard J2EE MVC Model 2 architecture has been adapted and extended (Rostanin 2004) by introducing the client abstraction components: a *client specific request adapter* parses the incoming HTTP request and redirects it to the controller component as standardized interfaces; *controller component* makes the necessary calls to the business logic and after that gives the control to the output engine that interprets the model data and generates output specific to the given client type. Accordingly, using this approach, one can use the same business logic for serving different client types (from WAP and HTML to Flash and Java application).

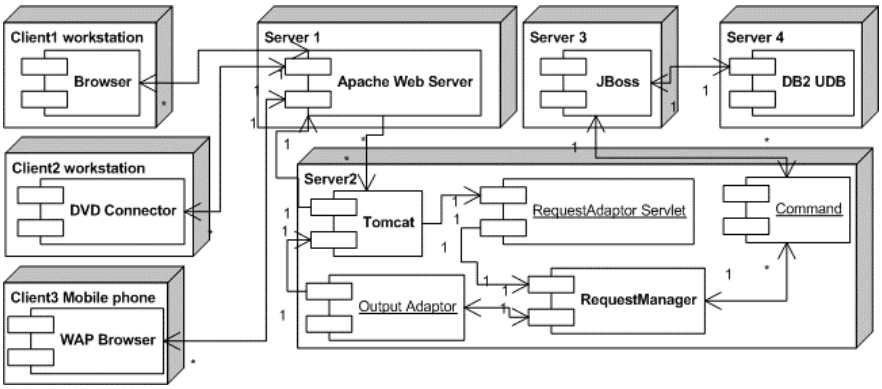

Figure 4: The component architecture of the system DaMiT stripped down to the essentials

Adaptive Content Generation

In (Jantke, Rostanin et al. 2004), there has been developed a generalized model by introducing the notion of a *learning goal*, so that the layer of the views representing the content can be dynamically adapted to the current learning goal (s) in order to satisfy the learner's needs. To realize this concept, the stored procedures responsible for the content structure generation have been parameterized. It accordingly delivers the XML document describing the learner-adapted lesson's structure. After the lesson structure was generated the application parses this XML document and creates its object representation in the application memory. Content itself is not loaded at once, but subsequently on demand.

The proposed approach has two main disadvantages: First, it does not allow to change the lesson's structure on the fly. Second, it uses much operative memory on the server. These problems will be the subject of the future work.

Summary and Conclusions

Finally, it is time to ask what the authors have been delivering with the present paper and how the related conference presentation can be expected to serve the audience.

- The authors have presented a case of an e-learning system which has found its way from the ivory tower of academia to the business world—a key problem of current projects of universities worldwide. We are not yet able to present a process model of how to go that way, in general, but however to generalize, we need experience.
- There has been focused the need for high dimensional media. Interesting cases of dovetailing high-resolution videos, i.e., offline content, with dynamic online content are discussed. The oral presentation of this contribution may include a life demonstration to convincingly clarify the added value of integrating media and services.
- The technologies underlying the described media integration are explained in some detail, from the database and system architecture over the metadata concepts and their syntactic appearance to the DVDconnector technology.

There are many directions of future work, of which one should be mentioned here: extending the "intelligence" of the e-learning system by implementing learnability of a wider reach. Currently, the DaMiT system is able to deal with deeply structured (sets of) learning goals (Jantke, Rostanin et al. 2004), and there are a few approaches of understanding the learner by inferring the learner's current goals. More sophisticated ideas should allow the system to hypothesize the learner's needs and desires from the learner's behavior towards a higher degree of adaptivity.

References

ACM (1987). ACM Turing Award Lectures. The First Twenty Years 1966-1985. ACM Press Anthology Series, ACM Press.

Bull, S., Greer, J., McCalla, G. (2003). The Caring Personal Agent. International Journal of Artificial Intelligence in Education, 13, 21-34.

Henze, N., Nejdl, W. (2001). Adaptation in Open Corpus Hypermedia. International Journal of Artificial Intelligence in Education, 12, 325-350.

Jantke, K., Lange, S., Grieser, G., Grigoriev, P., Thalheim, B., Tschiedel, B. (2004a). Learning by Doing and Learning when Doing. Dovetailing E-Learning and Decision Support with a Data Mining Tutor. Sixth International Conference on Enterprise Information Systems ICEIS 2004, Porto, Portugal, Proceedings, Vol. 1, 238-241.

Jantke, K., Lange, S., Grieser, G., Grigoriev, P., Thalheim, B., Tschiedel, B. (2004b). Work-Integrated E-Learning on Demand – The DaMiT Approach. Proc. 49th IWK, Ilmenau, Germany.

Jantke, K., Rostanin, O., Tschiedel. B. (2004). Adaptivity to Dynamic Learning Goals – Database & Generation Technologies. Submitted

Lelouche, R. (1999). Intelligent tutoring systems from birth to now. Künstliche Intelligenz 13 (1999) 4, 5-11.

Rostanin, O. (2004). An Alternative Approach to Building Web-Applications. Sixth International Conference on Enterprise Information Systems ICEIS 2004, Porto, Portugal, Proceedings, Vol. 1, 119-124.

Self, J. (1992). Computational Mathetics: the missing link in Intelligent Tutoring Systems research? Directions in Intelligent Tutoring Systems, NATO ASI Series F, Vol. 91, 36-56.

Self, J. (1999). The defining characteristics of intelligent tutoring systems research: ITSs care, precisely. International Journal of Artificial Intelligence in Education, 10, 350-364.

Santos, C., Frozza, R., Dhamer, A., Gaspary, L. (2002). DÓRIS – Pedagogical Agent in Intelligent Tutoring Systems. Sixth International Conference on Intelligent Tutoring Systems ITS 2002, Springer-Verlag, LNCS 2363, 91-104.