MILESS – A Learning and Teaching Server for Multi–Media Documents

Holger Gollan, Frank Lützenkirchen, and Dieter Nastoll

Computer Center, Essen University, Schützenbahn 70, 45117 Essen, Germany

Abstract. MILESS [7] is a joint project between the Computer Center and the Central Library of Essen University, together with the two pilot departments of linguistics and physics. The main purpose is to provide students and faculty of Essen University with a library server that supports several different functions that are needed within a digital library. Based on the IBM DB2 Digital Library product [4], MILESS can store and retrieve digital documents in any given format; moreover, searching is possible in a very elaborate way, and access control is supported as well.

In this article, we will first discuss why there is a growing need for digital library servers, followed by a description on how MILESS is build on top of the IBM DB2 Digital Library. We will describe the software techniques that are used to build the system, and we give a test case for the use of MILESS when referencing different articles within a mathematical journal.

1 The Need for a Digital Library

With the evolving web technologies, the amount of digital data that is accessible via the internet is enlarging in a dramatic way. Usually, these data will appear on certain websites, either personal or institutional. In addition, there might be commercial places on the web that hold lots of information in different digital formats. But this huge set of information leads to several problems.

- It is sometimes hard to find.
- It has no systematic order.
- It might vanish without further notice.

On the other hand, classical libraries have to find new ways to enable their customers to work with this new material in addition to the wellknown books and journals on the shelfs. Moreover, university students and faculty want to use digital and esp. multimedia material in learning, teaching and research. To face these problems and to meet these needs, a digital version of the classical library services is needed. It should support the use of such material by providing a reliable, permanent, and systematically ordered access to it.

2   MILESS and the IBM DB2 Digital Library

In late 1997, the Computer Center and the Central Library of Essen University started the MILESS project, which was funded by the local state ministry and the university. The idea was to install a digital library server that could solve the problems mentioned in the previous section. While the Computer Center brought in its knowhow in information technology and software development, the Central Library started to redefine the classical library techniques and services for the new types of digital and multimedia objects. In addition, two pilot departments (linguistics and physics) started to fill the digital library server with appropriate material.

To store and archive the digital documents, MILESS uses the IBM DB2 Digital Library product [4]. Its main parts consist of one library server and several object servers, where the object servers are responsible for the actual storage of the documents, but access is only possible via the library server that controls and manages the documents that are put into the digital library. Using this control mechanism, it is impossible to delete any documents in the object servers without notification of the library server, hence there can be no dead links within the system. The library server itself is running on top of a DB2 database, and the object servers can be connected to an ADSTAR Distributed Storage Manager (ADSM) [1] that handles the storage and archiving problems. E.g., documents can be archived when they haven't been used for a longer time.

The IBM DB2 Digital Library product offers a lot of features including several services that are useful for a digital library server. It handles storage and management of the documents via the object servers, and it enables access control via a rights management. Moreover, it has sophisticated search techniques, e.g. text mining and Query By Image Content (QBIC) to enable the user to find what he (she) is looking for within the stored digital documents.

While the IBM DB2 Digital Library product does an excellent job when it comes to the storage and retrieving problem, it is of no great help for the implementation of the MILESS data model. Since we wanted to have great flexibility in this respect, we adopted the Dublin Core [2] standard for the description of electronic resources, adding some additional features like contact information for the creators and contributors of the documents. Thus a document within MILESS can e.g. have several titles, several creators and contributors, several types and formats, etc. In particular, documents in MILESS can have several derivates in different formats; no standard format is required. Moreover, MILESS can handle hierarchical classifications that are widely used in science to help capturing the subjects of a document in a standardized way. This very complex and yet flexible data model for the metadata of electronic documents enables the librarians in the project to extend their classical library services to the digital material within the MILESS system.

Since the IBM DB2 Digital Library product can not handle such complex data models, additional software had to be written to enable MILESS to work with the Dublin Core standard. We will take a closer look on the new software in the next section. The following figure illustrates the different parts of the IBM DB2 Digital Library product.
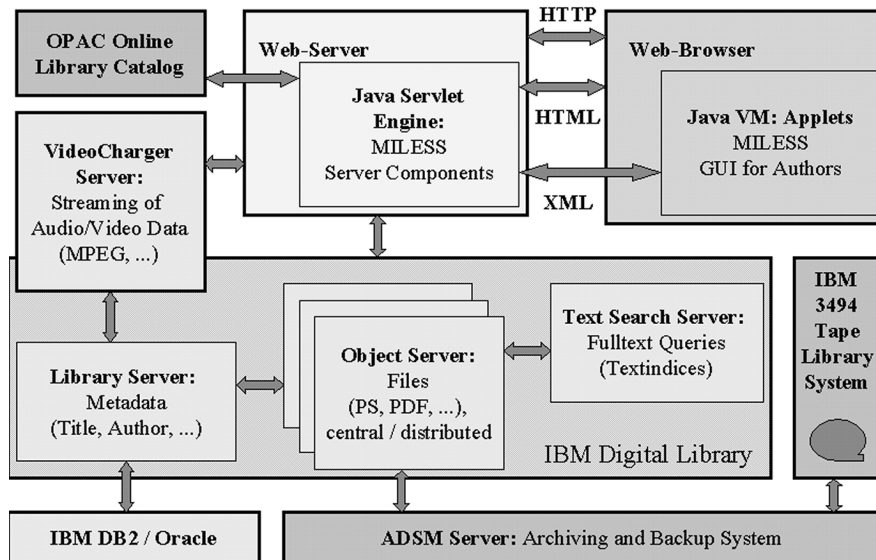

3 MILESS – A Closer Look

Besides the need for special software because of the complex data model, the additional MILESS software is divided into different layers, dependent on their functionality. To be platform independent, a design decision was made to use JAVA as the programming language, made possible by a JAVA API for the IBM DB2 Digital Library that can be used to connect JAVA code with the Digital Library product. Thus the bottom layer of the system is given by this programming interface that connects the Digital Library product with the outside world. This API is used by a so–called Persistency Layer that is responsible for the storing and retrieving of documents.

On top of that there is a collection of JAVA classes that implement the funtionalities for documents, legal entities (creators and contributors), etc.

Another part of the inner system is using JAVA servlets. MILESS is running inside a web server that is capable of using servlets, and these servlets are used for the connection and communication between the user and the system, e.g.

- A *DocumentServlet* is used to present the metadata of a document on a page within the web server.
- The *DerivateServlet* is needed to access a certain format of a specified document.
- The *SearchServlet* takes the user queries, connects to the Digital Library product to do the search, and presents the results on a page within the web server.

The following figure illustrates the different software layers of the MILESS system.
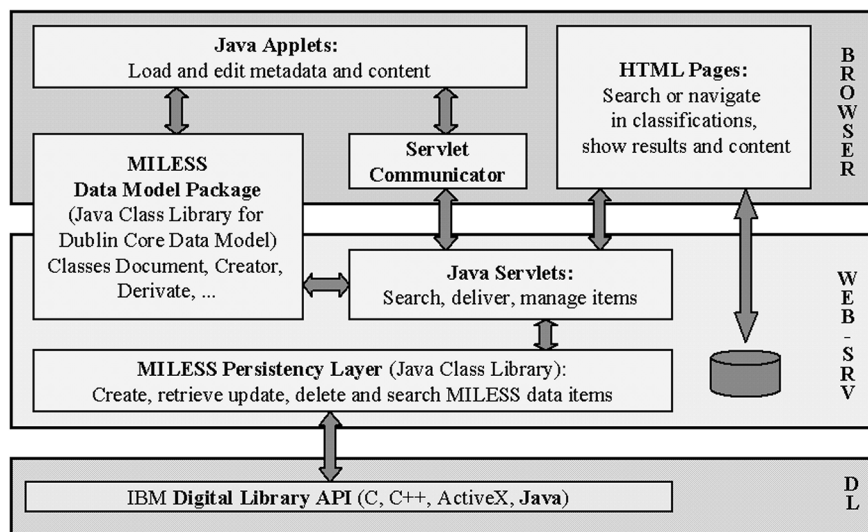
Besides the inner parts of the system that run on the server side, there are other parts that run on the side of the user client. Basically, the only thing the user needs is a web browser that connects him to the MILESS homepage at http://miless.uni-essen.de. From here, the user has access to the full functionality of the system; e.g. the search facilities can be reached just via normal HTML–pages. In addition, there is the possibility for an author to create or change documents inside the MILESS system. To do this, he (she) can call a graphical user interface (GUI) that runs as a JAVA applet inside the web browser. This GUI can be used to create new documents for the system, or to change already existing documents or personal data of creators and contributors. Moreover it helps navigating through hierarchical classifications to find the correct subjects for the documents. The communication between this GUI on the client side and the inner part of the system on the server side is handled via a Servlet Communicator, the data exchange is done via XML [3]. In the near future contributors can use XML directly to put material

into the system. Thus the application on the client side doesn't have to know anything about the internal representations on the server side; especially the IBM DB2 Digital Library product and its internal structure is totally invisible to the outside world.

4   MILESS – From the user point of view

There are a lot of possible scenarios for a user of the MILESS system. Most people will use it just the way they use a classical library, but with enhanced features. One of these is the search facility. By making use of the search techniques of the underlying IBM DB2 Digital Library product, the user can not only search for certain authors or words and phrases within the title or the keywords, but it is also possible to search for words and phrases within the texts of the documents. Moreover, because of the way hierarchical classifications are implemented inside MILESS, he (she) can navigate through the hierarchy of such classifications, looking e.g. at all documents at a specific level.

Because of the capability of the system to handle documents in any given format, the user might have a problem with the actual format of a retrieved document. To overcome this difficulty, MILESS has a plug–in collection that could help the user and his browser to understand a strange format. Moreover, we are collecting different converters that could be used to create new formats from existing ones when putting new material into the system.

Creating new material is another scenario for the use of the MILESS system. With the help of the user GUI, anybody can create new documents in the system by providing the necessary metadata within the GUI and uploading the data files of the document into the system. This can be used e.g. by lecturers to put their lectures and exercises into the system, enabling students to work with this material online whenever they want.

Another scenario sees a lecturer preparing his next talk and searching the system for certain multimedia material he can use in the class. Such material can be JAVA applets or simulations/animations, audio-/video material, etc. To access video material, a video server [5] is included in the system that uses streaming techniques to deliver multimedia material in real time to multiple users.

Yet another use of MILESS is the linking between different articles of a mathematical journal; a first test case for this will be presented in the next section.

5   MILESS and the "Archiv der Mathematik" – A Test Case

This final section presents a collaboration with the Institute for Experimental Mathematics at Essen University, where six old volumes of the mathemat-

ical journal "Archiv der Mathematik" will be retrodigitized to make them available on the web in digital form (see [6]).

One small piece in this project is the automatic extraction of the bibliographic data like title, authors, author address, journal name, volume number, etc. These data, stored in an XML–format [3], can be used to fill the metadata fields of the Dublin Core standard automatically and to put the retrodigitized articles into the MILESS system, with restricted and controlled access because of copyright issues.

Another piece in this project is the automatic recognition of the cited references at the end of any article. Using Optical Character Recognition (OCR) and heuristics, an HTML–page is produced that contains the references and tries to link, where possible, to an online copy of the cited article. To do this, some standardization is needed to produce a correct link. To install a prototype for such a referencing functionality, we have put two articles of the "Archiv der Mathematik" into the MILESS system, namely

- William Crawley–Boevey, *Tameness of biserial algebras*, Arch. Math. **65**, 399–407.
- Christof Geiss, *On degenerations of tame and wild algebras*, Arch. Math. **64**, 11–16.

where the second one is a cited reference in the first article. After the automatic recognition of the references of the first article, the link to the second article will be produced automatically as

http://miless.uni-essen.de/iem/Archiv_der_Mathematik/64/11

where the volume number and the first page are used to uniquely identify the cited article. Upon this request the MILESS system starts an internal search to retrieve the referenced article. With this feature the reader can view the first article, realizing the citation, and following the link by just clicking on it. This can easily be extended to referenced articles being published in other journals, once these journals are available online and a unique way to reach their articles is realizable just from the bibliographic data as in the example above. An extension of the prototype in this direction is planned for the near future. Such an extension can lead to a distributed library for scientific journals, not necessarily restricted to mathematics, adding new features and functionalities for the user, but creating some demands on the underlying networks as well.

6   Acknowledgement

References

1. ADSTAR Distributed Storage Management,
   http://www.storage.ibm.com/software/adsm/
2. The Dublin Core Standard, http://purl.oclc.org/dc/
3. Extensible Markup Language, http://www.w3.org/TR/REC-xml
4. IBM DB2 Digital Library, http://www.software.ibm.com/is/dig-lib/
5. IBM DB2 Digital Library Video Charger,
   http://www.software.ibm.com/data/videocharger/
6. G. O. Michler, "A Prototype of a Combined Digital and Retrodigitized Search-
   able Mathematical Journal", Preprint.
7. MILESS – Multimedialer Lehr- und Lernserver Essen, http://miless.uni-essen.de