dLibra
Digital Library Framework

Paweł Gruszczyński Cezary Mazurek Stanisław Osiński
Andrzej Swędrzyński

Poznań Supercomputing and Networking Center,
ul. Noskowskiego 10, 61–704 Poznań, POLAND

phone: +48 61 858 20 30, fax: +48 61 852 59 54
e-mail: {grucha,mazurek,stachoo,kokosz,szuber}@man.poznan.pl

Abstract

The aim of this paper is to introduce *dLibra* - a Digital Library Framework developed by Poznan Supercomputing and Networking Center (PSNC). *dLibra* is a modern content delivery and management tool for New Generation Networks. We present the system with respect to its functionality, internal architecture and technologies employed.

Keywords: digital library, electronic publishing

1 Introduction

The *dLibra* Digital Library Framework has been developed by Poznan Supercomputing and Networking Centre since 1999 [1][2]. The software delivers tools for setting up, accessing and managing a library of digital documents of various types. With a hierarchical directory structure, advanced document versioning support, comprehensive right management and flexible searching capabilities the *dLibra* framework can provide a wide variety of multimedia services such as multimedia books, distance education or media on demand.

2 System functionality

dLibra facilitates all phases of a digital publishing process by supporting three basic groups of users: readers, writers and publishers. Using a web-based interface the readers can easily browse the library and view selected publications. A search engine enables them to issue a query regarding various multilingual metadata attributes (e.g. using Dublin Core attribute scheme [3]) such as the publication author, title, description, keywords, creation date and many more. Full-text searching of the library content is also supported. The writers are delivered intuitive GUI-based tools for placing new publications in the library and retrieving publications or some of their components for further editing. An advanced versioning system supports managing subsequent revisions of publication objects as well as branching. Finally, the publishers

receive tools for managing the whole library structure, in particular, putting out and hiding publications, managing access rights and other library resources.

The whole library content is organized in a hierarchical structure of entities. A *directory* is an entity that groups any number of other items - subdirectories or publications. A *publication* is a unit of information (e.g. an article or a book) that consists of one or more basic objects of various types (e.g. HTML, image, sound). Every entity in the library – from the root directory down to a single publication object can be described by means of user-defined attribute schemes (e.g. Dublin Core). The values of attributes can be defined in several user-defined languages and are considered while searching the library.

dLibra provides support for both publication- and object-level versioning. Object-level versioning enables the writers to introduce new information to the already existing publication components, whereas publication-level versioning makes it possible to combine the new material into a new publication *edition*. Every publication can have an unlimited number of editions comprised of different versions of publication files or even different files.

Access management in the *dLibra* library is based on a system of users and groups. A user can be made a member of any number of groups, which may be considered as an assignment of a specific role (the user inherits all the rights granted to the groups of which he is a member). Rights can be granted on a library, directory or publication basis. On each level, several access sub-levels have been defined to enable a precise definition of user roles.

3 Internal architecture and technologies employed

The *dLibra* Digital Library Framework is implemented as a client-server system (Figure 1). On the server side there are a number of independent modules connected via network interfaces. Being responsible for data storage the Content module provides means for reading and writing particular versions of publication objects. The Metadata module manages the logical library structure - directory hierarchy, publication editions and attributes of all library objects. The User module supports access control and right management by storing user and group information along with access rights for all library objects. The Event module provides a versatile communication mechanism for the other server-side components. Using the mechanism, new functionality such as accounting or subscription services, can be easily implemented. The already implemented Search module uses the event service to update the library content indexes. All available services are managed by the System Services module and are made available to client applications by the User Interface module. Finally, the Presentation module serves a web application providing access to the library resources. On the client side GUI applications are delivered that support publication creating and library management processes.

All modules are implemented using Java 2 technology, in particular RMI (Remote Method Invocation), JDBC (Java Database Connectivity), JSP/Servlet, JavaSwing technologies. Currently, the data storage module utilizes the Oracle database system. Nonetheless, because of the use of JDBC and SQL 92 standards *dLibra* can be easily ported to work with any other RDBMS.

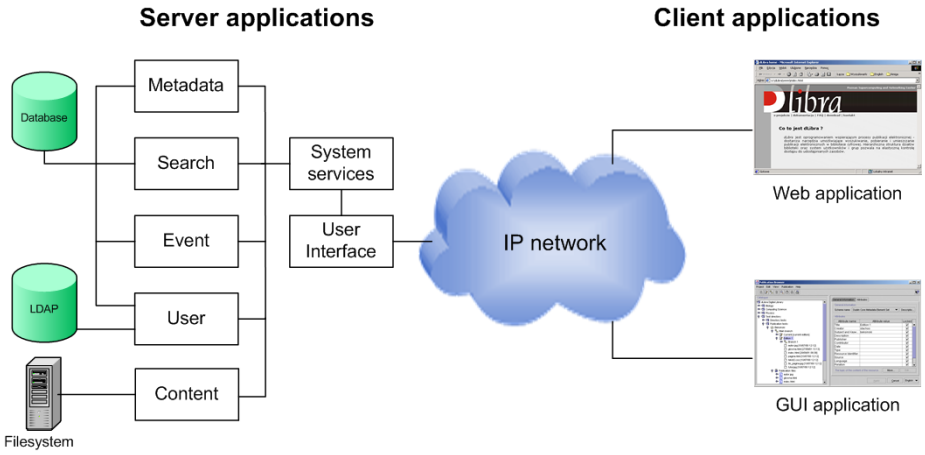

Figure 1: *dLibra* internal architecture

4 Further Work

In this paper we have briefly discussed features of the *dLibra* Digital Library Framework. In May 2002 *dLibra* was successfully deployed in PSNC to support publication, storage and access to company internal documents such as articles or reports. At present, in cooperation with the Poznan Foundation of Scientific Libraries, *dLibra* is being used to build a regional digital library (Digital Library of Wielkopolska) which aims at delivering scientific content to the local scholarly community. However, the design and development work is carried on to implement new functionality connected with group work, subscription services and accounting.

References

[1] Cezary Mazurek and Sebastian Szuber. *Development of Digital Libraries at Poznań Supercomputing and Networking Center*. ISThmus 2000. Research and Development for the Information Society, Poznań (Poland), April 2000.

[2] Cezary Mazurek, Maciej Stroiński, and Sebastian Szuber. *Digital Library for Multimedia Content Management*. ERCIM 9th DELOS Workshop, Digital Libraries for Distance Learning, Brno, April 1999.

[3] Dublin core metadata element set, version 1.1: Reference description. Technical report.

About the authors

Paweł Gruszczyński, born in 1978, received his Bachelor's Degree in Computer Science at the Poznan University of Technology in 2000. Since 1999 he has been working for the Poznan Supercomputing and Networking Center. His interests concern digital content management and distributed computing. Since 1999 a member of dLibra Digital Library Framework developer team.

Cezary Mazurek, born in 1969, received his Master's Degree in Computer Science at the Poznan University of Technology in 1993. He worked for Poznan University of Technology between 1993 and 1994 and from 1993 for the Poznan Supercomputing and Networking Center (PSNC). He is currently the head of Network Services Department at PSNC. Since 1994 he has been a project leader of the international software project on designing and development of network testers' GUI, which is led at Poznan University of Technology. His interests focus on advanced network services' design and development, digital multimedia libraries, distance education systems and portal access to GRID services. He is a co-author of an offer concerning advanced services for scientific users in the national POL-34/155 ATM network. The offer was presented during the POLMAN'99 conference and was rewarded by the National Committee for Scientific Research. He is an author of many publications presented on national and international conferences. Since 1998 he has been organising an annual presentation of advanced network services and applications developed by PSNC during European Conference and Exhibition - Information Society Technologies.

Currently he is leading the development of services based on Internet technologies (e.g. Digital Library Framework: dLibra, Polish Educational Portal in co-operation with Interkl@sa, Multimedia City Guide in co-operation with Poznan City). Since December 2001 he has been a leader of the project "Creating an access environment for GRID computational services performed by cluster of SUNs", co-founded by the National Committee for Scientific Research and SUN Microsystems.

Stanisław Osiński, born in 1979, received his Bachelor's Degree in Computer Science at the Poznan University of Technology in 2001. Since 2000 he has been working for the Poznan Supercomputing and Networking Center. His interests concern digital content delivery and management; since 2000 a member of dLibra Digital Library Framework developer team.

Andrzej Swędrzyński, born in 1978, received his Bachelor's Degree in Computer Science at the Poznan University of Technology in 2000. Since 1999 he has been working for the Poznan Supercomputing and Networking Center. His interests concern digital content delivery and management, distributed computing and software engineering. Since 1999 a member of dLibra Digital Library Framework developer team.