

**Pavel Pavlov, Maria Nisheva-Pavlova,
Anton Iliev, Klimentina Rousseva,
Nadezhda Apostolova**
Faculty of Mathematics and Informatics
Sofia University “St. Kliment Ohridski”

AUTHORING TOOLS FOR AN ACADEMIC DIGITAL LIBRARY

Abstract: The paper presents some aspects of an ongoing project aimed at the development of a methodology and proper software tools for building academic digital libraries. An ontology-based approach is suggested in order to standardize the semantic annotation of the library resources and to facilitate the implementation of the functionality of the search engine. The emphasis of the discussion falls on the analysis of the requirements of the main types of users of academic digital libraries and the corresponding authoring tools implemented in a particular academic digital library being under development at the Faculty of Mathematics and Informatics, Sofia University. These tools are oriented to the teaching and research staff members and the library administrators who may create and edit catalogue descriptions and upload library resources in the proper repositories. Some solutions of the variety of problems concerning the development of adequate mechanisms for semantics oriented search in multilingual academic digital libraries are discussed in brief.

Keywords: Electronic Publishing, Digital Library, User Interface, Authoring Tools, Metadata, Semantic Annotation, Ontology.

1. Introduction

During the last two decades libraries are rapidly changing to meet the increasing requirements of their users and the new opportunities given by the contemporary information and communication technologies. Their new generation has been developed as well equipped and interconnected *digital libraries*. The term *digital library* refers [1] to a range of systems, from digital object and metadata repositories, reference-linking systems, archives, and content management systems to complex systems that integrate advanced digital library services and support for research and practice communities. A digital library may offer many technology-enabled functions and services that support users, both as information producers and as information consumers. It integrates access to materials with access to tools for processing materials and supports individual and community information spaces through functionality for selection, annotation, authoring/contribution, and collaboration.

To this effect, many institutions are actively involved in building appropriate repositories of the institution's books, papers, theses, and other works which can be digitized or were “born digital”. In particular, universities and other academic institutions participate successfully in lots of projects directed to the development of suitable *academic digital libraries*. Academic digital libraries are committed to maintaining valuable collections of scholarly information. By their means, essential information resources should remain available and acces-

sible into the future – a serious challenge in the cases of digital resources that are increasingly transient and at risk.

In addition to its concrete institutional function, an academic digital library could play a significant role to:

- emphasize effective knowledge management and discovery as a key to innovation in learning and research;
- encourage the widest possible distribution of scientific, educational, and cultural information at the lowest possible cost;
- promote the needs and interests of diverse communities of users;
- inspire the creative application of technology in the service of scholarship and learning.

The paper presents some results of the work on a project, directed in particular to the development of a methodology and proper software tools for building academic digital libraries. A special attention has been paid to the elaboration of tools for semantics oriented search in multilingual academic digital libraries. The study and the practical experiments are oriented to the development of DigLib-CI – a digital library with research and learning materials (articles, dissertations, monographs, lecture notes, textbooks, presentations, example program sources, data sets, quizzes, manuals etc.) created at the Department of Computer Informatics of the Faculty of Mathematics and Informatics (FMI), Sofia University, or especially selected among the scholarly materials freely available on the Web.

The discussed activity has the following main objectives:

- to explore the architectural principles, user interfaces and search engines of the most popular academic digital libraries;
- to study the various aspects of creation and integration of appropriate ontologies oriented to the contents of multilingual academic digital libraries;
- to define suitable metadata to accompany various types of scholarly and learning materials, taking into account the internationally approved classification schemes, the Bulgarian national traditions and FMI experience;
- to develop a proper framework for application of advanced information technologies and particularly Semantic Web technologies in building tools for semantics oriented search in multilingual digital libraries.

2. Architecture of DigLib-CI

DigLib-CI follows the architectural principles of a typical academic digital library. It is designed in order to provide open access to various kinds of scholarly and instructional content in a number of subfields of Computer Science and Information Systems, developed or available at FMI. Its implementation is based on the use of proper Semantic web methods and technologies. The functional structure of DigLib-CI [2] is shown in Figure 1.

The content repositories treasure various research and learning materials in the areas of Computer Science and Information Systems, written in Bulgarian or in English. They contain library resources of different types: books, dissertations, periodicals and single articles, manuals, lecture notes, presentations, source code of computer programs, data sets, tests, quizzes, etc. These resources are available in various digital formats: pdf, html, plain text, doc, ppt, jpeg etc.

The metadata catalogues are designed in order to facilitate the identification of the needed research or learning materials by the search engine. They contain descriptive metadata stored in XML format and support the reusability of all library resources and facilitate their interoperability.

The subject ontologies describe the main concepts of the areas of Computer Science and Information Systems, with their properties and the various relationships between them. These ontologies play a significant role in the implementation of the full functionality of the search engine.

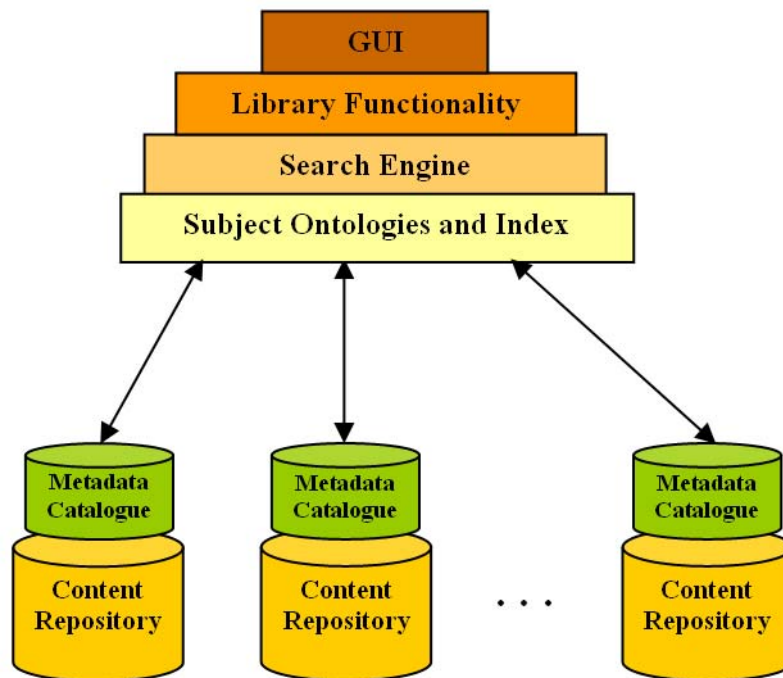


Figure 1: Functional Model of DigLib-CI

Two subject ontologies are included in the current version of DigLib-CI. The Computer Science ontology is based on the Computer Science Curriculum 2008 of ACM and IEEE/CS [3]. It defines the atomic knowledge units for the University courses and available research materials in the field of Computer Science and makes them sharable and reusable. This ontology includes more than 300 concepts with their properties and relationships. The Information Systems ontology has been under development using the Model Curriculum and Guidelines for Undergraduate Degree Programs in Information Systems of ACM, AIS and AITP [4]. The subject ontologies are designed in order to play the role of information sources describing the hierarchy and the other relationships between the main concepts in the discussed domains. They provide the search engine with a complete viewpoint to the conceptual structure of the areas of Computer Science and Information Systems.

The purpose of the search engine is to provide adequate access to the entire set of resources (learning materials, research materials, catalogue metadata, subject ontologies) stored in DigLib-CI.

The library functionality and the user interface of DigLib-CI are designed in accordance with the expected needs and requirements of the basic types of users of the library. The interface module provides adequate online access to the requested library resources and supporting software tools.

3. Attributes of the Library Resource Descriptions

The content repositories of DigLib-CI contain materials of the following 14 types:

- books (monographs, textbooks, manuals, etc.);
- periodicals;
- papers (single articles);
- master's theses;
- dissertations;
- synopses of dissertations;
- lecture notes;
- presentations of lectures;
- presentations of seminars;
- descriptions of course projects;
- source code of educational software;
- instructional data sets;
- tests and quizzes;
- unofficial materials.

In principle, the organization of the repositories of DigLib-CI is subject-oriented, therefore they are expected to contain research and learning materials of different types. Each repository has its own catalogue which includes proper metadata for the particular library resources.

The library catalogues contain descriptions of the resources stored in the corresponding repositories. These descriptions consist of metadata (proper attributes with their values) which support the identification of the requested resources by the search engine. The metadata are stored in XML format and comply with the IEEE Standard for Learning Object Metadata [5].

Most kinds of research and learning materials have the following attributes: type of the material; author; title of the material; language(s) (human and/or programming one(s)); digital format; location; version; date of creation; completion status; restrictions on use; subject(s); semantic annotation – list of concepts from the relevant subject ontology and proper keywords describing the Computer Science or Information Systems subfields and/or concepts covered or treated by the material. Learning materials have been characterized also by their educational level and the principal types of users for which the corresponding material was designed. Officially published research materials (articles, periodicals etc.) and books are supplied with the appropriate bibliographic metadata.

All resource descriptions (i.e., all catalogue entries) consist of two equivalent parts in which the element values are texts in Bulgarian or English respectively. The search engine examines the corresponding parts of the descriptions according to the language of the user's query.

Figure 2 illustrates a selected part of the description of a paper (single article) in XML format.

The element <bib_data> in the descriptions of officially published research or learning materials contains the respective bibliographic information (place of publication, e.g. periodical or conference proceedings; editor(s); publisher/publishing house; ISBN and/or ISSN of the book/periodical; year of publication etc.).

The elements <ontologyRefs> and <keywords> of the resource descriptions play the role of semantic annotations of the corresponding library materials. The values of the child elements of <ontologyRefs> are concepts of the appropriate subject ontologies (names of classes in these subject ontologies) which present most precisely the content of the respective document.

The concepts of the subject ontologies are too general from the point of view of the expectations of the typical users of DigLib-CI. Because of that, one can include in the resource descriptions additional lists of keywords which describe the content of the corresponding documents at an adequate level of abstraction. These keywords are set as values of the child elements of the <keywords> resource description elements.

The names of the relevant subject areas and the names of files containing the suitable subject ontologies have been assigned as values of the child elements of the catalogue description elements <subjects> and <ontologies> respectively.

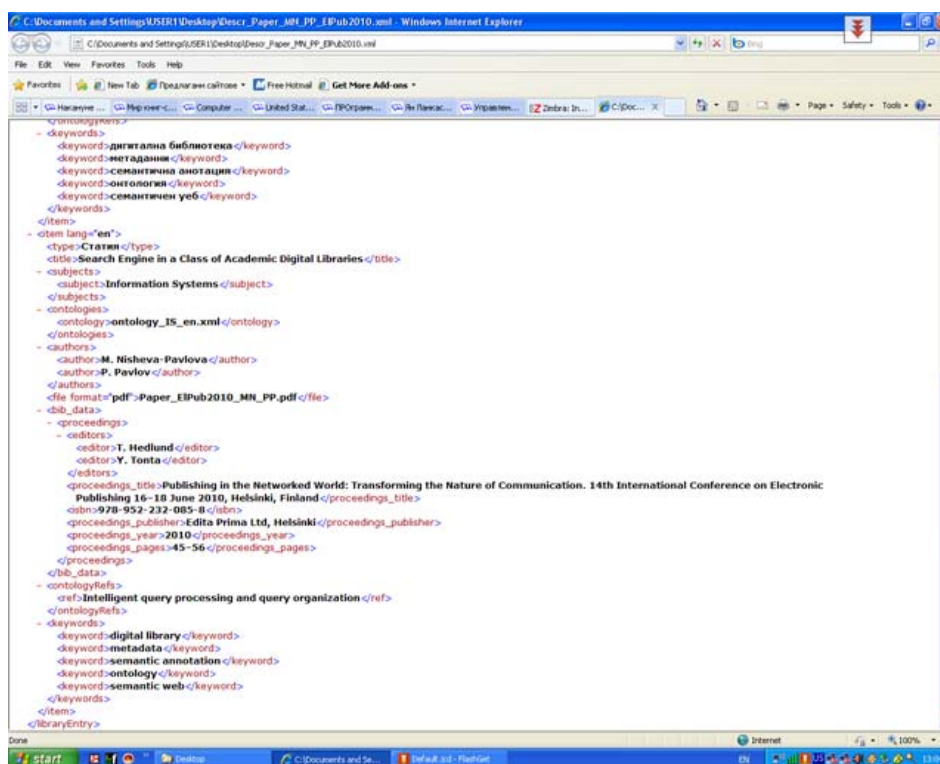


Figure 2: Part of the catalogue description of an article

4. Types of Users of DigLib-CI

The library functionality and the user interface of DigLib-CI are designed in accordance with the expected requirements of the basic types of users of the library:

- FMI students – they may read/download textbooks, open lecture notes and presentations from all public sections of the library as well as all manner of other kinds of materials (monographs, dissertations, articles, periodicals, degree theses, lecture notes, presentations, exercises, programs, data sets, quizzes, tests etc.) from fixed public library sections;
- FMI lecturers and researchers – in addition to the students' access rights, they may upload materials to fixed public sections as well as create and update private sections and use materials in some of them;
- librarians (library administrators) – they have full access to all public resources of the library (may download and upload materials destined for all public sections of the library);
- interested citizen (guests) – they may read and download public materials of fixed types (e.g., dissertations, textbooks, open lecture notes and presentations).

The interface module of DigLib-CI provides the particular types of users with adequate online access to the necessary library resources (research and learning materials from the relevant library sections and their catalogue descriptions) and supporting software tools (authoring and administrator's tools, communication tool, tool for ontology visualization, etc.).

5. Levels of User Interface

The current version of the user interface allows one to formulate queries in Bulgarian or English. All types of users of DigLib-CI have access to the standard input interface which gives up convenient means for entering, editing and submitting queries for various kinds of document search and retrieval (see for example Figure 3).

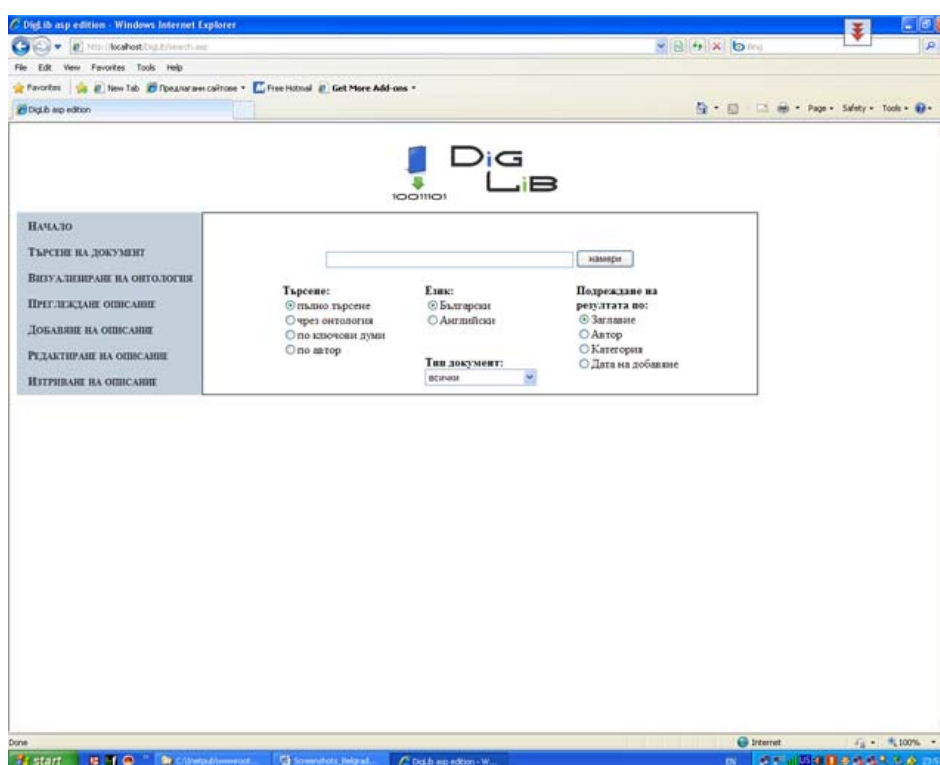


Figure 3: User interface of DigLib-CI (search form)

FMI lecturers and researchers as well as the library administrators may play the role of authors of library resources and catalogue descriptions, thus they have access to the author's part of the user interface. This part of the user interface provides for the authorized persons full access to the appropriate forms and enables them to enter (Figure 4) and edit (Figure 5) catalogue descriptions of all types of library resources.

More precisely, the forms for entering catalogue metadata (i.e., for creation of library resource descriptions) give the authorized user an access to sets of properly labeled empty fields in order to put in them the values of the corresponding description elements. The author (or the administrator) may enter the values of some of the elements or pick out the values of others from previously drawn lists. A tool for ontology visualization has been under development in order to assist the creation of semantic annotations of the available library resources. It will give the opportunity to visualize in a proper way the available subject ontologies and to pick out the necessary concepts in them as values of the child elements of the element <ontologyRefs> in the corresponding catalogue description.

The forms for editing catalogue metadata give the author an access to the particular fields containing the current values of the corresponding elements of the desired catalogue descriptions (in Bulgarian and in English language) in order to change them if necessary.

Figure 4: User interface of DigLib-CI (author's view – form for entering catalogue metadata of periodicals)

Figure 5: User interface of DigLib-CI (author's view – form for editing catalogue metadata of articles)

6. Conclusion

The most significant results of the discussed project activities obtained so far may be summarized as follows:

- An approach to building academic digital libraries was proposed. As a main feature of the suggested approach we could specify the provision of facilities for flexible semantics-oriented access to the library resources for users with various professional profiles and language skills;
- A prototype of DigLib-CI – an academic digital library with research and learning materials in the areas of Computer Science and Information Systems, was developed.

The current version of DigLib-CI integrates many presentation formats as well as various browsing, reading and authoring/creating tools. Our further plans envisage the implementation of some advanced digital library functionalities like tools for reviewing, maintenance of discussion forums and other forms of collaboration.

Acknowledgements. This work has been funded by the Sofia University SRF within the “Methods and Tools Supporting the Lifecycle of Rich Digital Content” Project, Contract No. 168/2010.

References

- [1] Soergel, D. *Digital Libraries and Knowledge Organization*. In: S. Kruk, B. McDaniel (Eds.), *Semantic Digital Libraries*. Springer, 2009, ISBN 978-3-540-85433-3, pp. 9–30.
- [2] Nisheva-Pavlova, M., P. Pavlov. *Search Engine in a Class of Academic Digital Libraries*. In: T. Hedlund, Y. Tonta (Eds.), “*Publishing in the Networked World: Transforming the Nature of Communication*”. 14th International Conference on Electronic Publishing 16–18 June 2010, Helsinki, Finland”, Edita Prima Ltd, Helsinki, 2010, ISBN 978-952-232-085-8, pp. 45–56.
- [3] ACM, IEEE Computer Society. *Computer Science Curriculum 2008 (An Interim Revision of CS 2001)*, <http://www.acm.org/education/curricula/ComputerScience2008.pdf>, last accessed on August 16, 2010.
- [4] ACM, AIS, AITP. *IS 2002 (Model Curriculum and Guidelines for Undergraduate Degree Programs in Information Systems)*, <http://www.acm.org/education/is2002.pdf>, last accessed on August 16, 2010.
- [5] *IEEE Standard for Learning Object Metadata*, <http://ltsc.ieee.org/wg12/20020612-Final-LOM-Draft.html>, last accessed on August 16, 2010.

pavlovp@fmi.uni-sofia.bg

marian@fmi.uni-sofia.bg

anton_ii@abv.bg

klimentina.ruseva@gmail.com