Преглед НЦД 13 (2008), 17–21

Pavel I. Pavlov and
Maria M. Nisheva-Pavlova
(Faculty of Mathematics and Informatics,
Sofia University “St. Kliment Ohridski”)

SOME IT ASPECTS OF BUILDING DIGITAL LIBRARIES WITH LEARNING MATERIALS

Abstract: The paper presents some initial results of a project directed to the development of a methodology and corresponding software tools for building academic digital libraries. A particular functional model of academic digital library has been discussed. The emphasis falls on some solutions of the large set of problems concerning the development of proper mechanisms for semantics oriented search in multilingual digital libraries. A model and a prototype of an academic digital library providing different categories of users with proper access to learning materials in the fields of procedural, object oriented and functional programming will be the final result of the project. The paper discusses the requirements of the basic types of users of such digital library and suggests some relevant solutions.

Keywords: Digital Library, Electronic Publishing, Metadata, Semantic Annotation, Ontology, Semantic Web

1. Introduction

Recently Computer Science and information technologies play an important role in numerous successful projects directed to digital preservation of collections of handwritten, typewritten and printed archival documents, photographs etc. which are considered as significant scientific or cultural heritage. At the same time, the number of “digitally born” documents, papers, books etc. that are of great importance for the development of civic activities and industry, is increasing extremely fast.

The existence of a large amount of web available documents of various types troubles the access to them and therefore enjoins the necessity of their organization in the form of the so-called digital libraries. During the last decade a series of promising results were reached in the area of development of digital libraries and tools providing adequate access to their content. As a leading example in this direction we could mention the European Digital Library Project funded by the European Commission under the eContentplus Programme and coordinated by the German National Library [10].

At the same time, the developers of digital libraries and means for access to their content are still faced with a number of challenges. One of these challenges is the execution rate of the user queries. Another challenge is the provision of most precise and rich in content answers of the user questions and queries. The next serious challenge is the necessity of development of search methods and techniques which will be appropriate for sets of materials containing documents of different types and multiform content, available in various electronic formats.

The paper discusses some initial results of a work in progress in the scope of a project funded by the Sofia University SRF which is directed to the development of a methodology

and corresponding software tools for building academic digital libraries. A special attention has been paid to the elaboration of means for semantics oriented search in multilingual digital libraries. The study and the practical experiments are oriented to the development of ProgDL – a digital library with learning materials supporting the courses on procedural programming, object oriented programming and functional programming directed to Computer Science students at the Faculty of Mathematics and Informatics (FMI) of Sofia University. The main idea here is that one can elaborate proper tools for adequate access to the components of an academic digital library using minimal amount of resources, with the assistance of an appropriate set of metadata and suitable ontologies. This work is based on some former results in the development of software tools for semantics oriented access to digitized collections of manuscripts and digitized archival collections.

2. Scope and Architecture of ProgDL

ProgDL is a typical institutional digital library. It has been under development at the Department of Computer Informatics in order to provide open access to various kinds of instructional content in a wide range of subfields of Computer Programming. The library is intended primarily for Computer Science and Information Systems BSc students at FMI. Its functional model is shown in Figure 1.

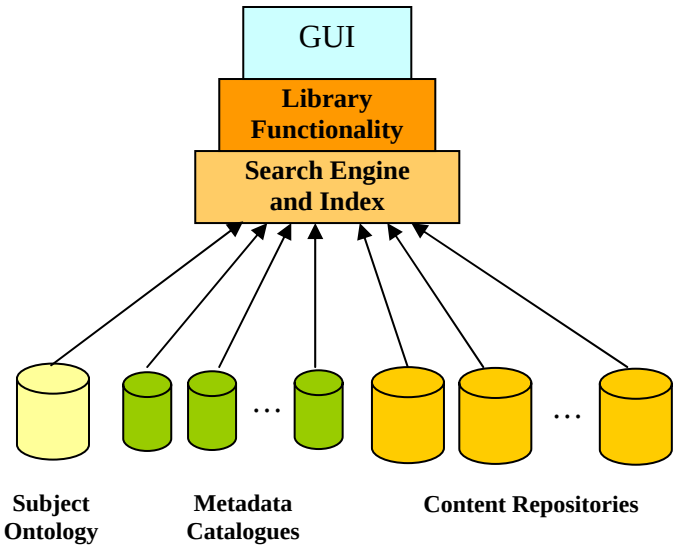

Figure1. Functional Model of ProgDL

The content repositories include learning materials (textbooks, papers, lecture notes, presentations, exercises, programs, data sets, tests, quizzes etc.) in the fields of procedural, object oriented and functional programming available in various digital formats: pdf, html, plain text, doc, ppt, jpeg etc. The main part of these materials is developed by faculty members, the others are carefully selected among the learning materials freely available on the Web. The content repositories are stored in a small number of locations. The learning materials in them are written in the Bulgarian, English or Russian languages.

The metadata catalogues are designed in order to facilitate the identification of the needed learning materials by the search engine. They contain descriptive metadata that comply with the IEEE Standard for Learning Object Metadata [11] stored in XML format. Relevant attributes of learning materials to be described include: type of material; style(s) of programming studied in the material; author; title; format; location; educational level;

restrictions on use; semantic annotation – list of concepts from the subject ontology describing the Programming subfields and/or concepts covered or treated by the material. In this way metadata catalogues support the reusability of learning materials, aid discoverability, and facilitate their interoperability. The names of the metadata tags (attributes) are originally picked out in the Bulgarian language.

The subject ontology includes a large set of concepts studied in the University courses in procedural, object oriented and functional programming, with description of their properties and the relationships among them. This ontology is based on the computing curricula CC2001 of the ACM and IEEE/CS. Using the curriculum as a guideline, the ontology defines the atomic knowledge units for the mentioned set of programming courses and makes them sharable and reusable. It has been under development using Protégé/OWL.

The search engine provides access to the complete palette of information stored in ProgDL.

Digital libraries are online collections of information sources that should be managed by and made accessible to a community of users. In contrast with traditional libraries, digital libraries can immediately adopt innovations in technology providing users with improvements in electronic and audio book technology as well as presenting new forms of communication such as wikis and blogs. These considerations and the expected requirements of the basic types of users served as arguments for the design principles of the library functionality and the user interface of ProgDL. More precisely, the current version of the user interface is intended for four types of users:

- general citizen – they may read and download public materials of fixed types (textbooks, open lecture notes and presentations);
- FMI students – 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 (lecture notes, presentations, exercises, programs, data sets, quizzes, tests) from definite public library sections;
- FMI lecturers – in addition to the students' access rights, they may upload materials to definite 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.

3. Key Features of the Search Engine of ProgDL

The search engine is responsible for locating the user-desired resources stored in the content repositories of ProgDL. It accepts metadata queries on the properties of resources and returns a list of metadata descriptions and references pointing to the corresponding repository units.

The search engine of ProgDL uses as auxiliary structures an index of the metadata catalogues and a dictionary of synonyms of the subject ontology concepts. The user queries define restrictions on the values of certain metadata attributes of the required learning materials. In particular, they may contain conjunctions and disjunctions of "atomic" queries. For example, the user could be interested in the materials of type "textbook" or "lecture notes" concerning all styles of programming that cover the concept of "generic function" (or the concepts of "tail recursion" and "iterative process").

During the working cycle of the search engine the user query is augmented with regard to the concepts searched out in the semantic annotations of the required learning materials. The more specific concepts from the subject ontology and some synonyms of the main terms from the dictionary mentioned above are added to the original ones in the resulting query. A more detailed description of the used algorithm for processing of user

queries containing conjunctions or disjunctions with respect to the value of a specific metadata attribute can be found in [8]. Our further plans envisage the enrichment of the search engine capabilities with means for processing of user queries formulated in the English or Russian languages. A short English–Russian–Bulgarian terminological dictionary has been under development for the purpose and the names of the metadata tags (attributes) are meantime available in the English, Bulgarian and Russian languages.

The design principles of the search engine of ProgDL are based on the use of some well-known methodologies for application of Semantic web technology in building knowledge-based tools for access to digitized scientific and cultural heritage [6] as well as on our direct impressions of the access tools provided by several popular specialized academic digital libraries [1–3]. Some actual issues in the areas of personalizing digital library access with preference-based queries [4] and advanced search engine technologies successfully applied in digital libraries [5] are also taken into account.

The implementation of the project is based on the experience obtained in the development of tools for intelligent (semantics oriented) search in collections of digitized manuscripts [7,8] and the elaboration of software tools for semantics oriented access to heterogeneous multilingual collections of archival materials [9]. Some appropriate Semantic web methods and technologies and a suitable web-based content management system (MediaWiki [12], a publicly available wiki software) have been used for the purpose.

4. Summary and Future Work

The most significant results of the discussed project obtained to the moment could be formulated as follows:

- A functional model of an academic digital library was proposed. This model provides means for semantics oriented access to learning and research materials in different electronic formats;
- The requirements of the different types of users of such digital library are examined and the basic library functionalities of ProgDL are formulated on this basis.

The current activities are directed to the development of:

- a methodology for building academic digital libraries with learning materials;
- a prototype of ProgDL – an academic digital library with learning materials in the areas of procedural, object oriented and functional programming.

The implementation of the prototype of ProgDL will be realized by integration of some proper Semantic web methods with techniques and technologies for development of rich Internet applications.

Acknowledgements. This work has been partly funded by the Sofia University SRF under Contract No. 123/2008.

References

[1] The ACM Digital Library, http://portal.acm.org/dl.cfm, last accessed on March 18, 2008.
[2] IEEE Computer Society Digital Library, http://www.computer.org/portal/site/csdl/, last accessed on March 18, 2008.
[3] J. Rákosník, *DML-CZ: Czech Digital Mathematics Library*, ERCIM News 66 (2006), 46–47.
[4] P. Georgiadis, N. Spyratos, V. Christophides, C. Meghini, *Personalizing Digital Library Access with Preference-Based Queries*, ERCIM News 66 (2006), 34–35.
[5] J. Lervik, S. Brygfjeld, *Search Engine Technology Applied in Digital Libraries*, ERCIM News 66 (2006), 18–19.

[6] D. Bearman, J. Trant, *Issues in Structuring Knowledge and Services for Universal Access to Online Science and Culture*, Nobel Symposium "Virtual Museums and Public Understanding of Science and Culture", May 26–29, 2002, Stockholm, Sweden; available at http://nobelprize.org/nobelfoundation/-publications/symposia/ns120-lectures/bearman.pdf, last accessed on March 18, 2008.

[7] M. Nisheva-Pavlova, P. Pavlov, *Tools for Intelligent Search in Collections of Digitized Manuscripts*; In: M. Dobreva, J. Engelen (Eds.), *From Author to Reader: Challenges for the Digital Content Chain, Proc. 2005 ELPUB Conf.*, Peeters Publishing, Leuven, 2005, pp. 145–150.

[8] P. Pavlov, M. Nisheva-Pavlova, *Knowledge-based Search in Collections of Digitized Manuscripts: First Results*, Proc. 10th ICCC Internat. Conf. Electronic Publishing (Bansko, 14–16 June 2006), FOI-Commerce, Sofia, 2006, pp. 27–35.

[9] M. Nisheva-Pavlova, P. Pavlov, N. Markov, M. Nedeva, *Digitisation and Access to Archival Collections: A Case Study of the Sofia Municipal Government (1878–1879)*, Proc. 11th Internat. Conf. Electronic Publishing (Vienna, Austria, 13–15 June 2007), pp. 277–284.

[10] The European Digital Library Project, http://www.edlproject.eu/, last accessed on April 28, 2008.

[11] The IEEE Standard for Learning Object Metadata, http://ltsc.ieee.org/wg12/index.html, last accessed on April 28, 2008.

[12] The MediaWiki Homepage, http://www.mediawiki.org/wiki/MediaWiki, last accessed on April 28, 2008.

pavlovp@fmi.uni-sofia.bg
marian@fmi.uni-sofia.bg