

BUILDING A DIGITAL LIBRARY WITH LEARNING MATERIALS

Maria Nisheva-Pavlova^{1,2}, Pavel Paolov¹.

¹Faculty of Mathematics and Informatics,
Sofia University "St. Kliment Ohridski"

5, James Bourchier Blvd., Sofia 1164, Bulgaria

²Institute of Mathematics and Informatics, Bulgarian Academy of Sciences

Acad. G. Bonchev Str., Bl. 8, 1113 Sofia, Bulgaria

e-mail: marian, pavlovp@fmi.uni-sofia.bg

Abstract

The paper discusses some aspects of a project aimed at the development of a methodology and proper software tools for building academic digital libraries. A particular functional model of academic digital library has been proposed and analyzed. The emphasis of the presentation falls on some solutions of the large set of problems concerning the development of adequate mechanisms for semantics oriented search in multilingual 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. A proper subject ontology covering the area of Computer Science has been developed for this purpose. The paper discusses the requirements of the basic types of users of an academic digital library and suggests some relevant solutions.

Keywords: Digital Library; Metadata; Semantic Annotation; Ontology; Semantic Web.

1. Introduction

Recently computers and networks have changed the ways in which people retrieve information and communicate with each other. In particular, in many areas digital libraries are gradually taking the place of traditional ones because a professional or scholar is better served by sitting in front of a personal computer connected to a communications network than by going to a library. Information that was previously available only to professionals is now directly available to everyone. From a

personal computer, the user is able to consult materials that are stored on computers around the world.

A digital library brings the information to the user's desk, either at work or at home, making it easier to use and therefore increasing its usage. With a digital library on the desktop, a user might never need to visit a library building. The library is wherever there is a personal computer and a network connection.

At the same time, developers of digital libraries and tools 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 sufficiently precise and rich content answers for the user queries. The next considerable challenge is the need for the development of search methods and techniques which will be appropriate for repositories containing documents of different types and multiform content, available in various digital formats.

The paper aims to present an ongoing project, which is directed to developing a methodology and corresponding software tools for building academic digital libraries. Special attention has been given to the elaboration of semantics oriented searches in multilingual digital libraries. The study and practical experiments are oriented to the development of ProgDL – a digital library with learning materials supporting the courses on Procedural Programming, Object Oriented Programming, Functional Programming and Data Structures and Algorithms designed for Computer Science students at the Faculty of Mathematics and Informatics (FMI) of Sofia University.

The discussed project has the following main objectives:

- To explore the architectural principles of institutional digital libraries and academic digital libraries;
- To study the various aspects of the creation of appropriate ontologies oriented to the contents of multilingual digital libraries;
- To define suitable metadata to accompany various types of learning materials in the domain of Computer Programming, taking into account the internationally approved classification schemes and the FMI experience;
- To develop a proper framework for the application of advanced information technologies and particularly Semantic Web technologies in building tools for semantics oriented searches in multilingual digital libraries.

The implementation of the project is based on the former results of the authors in the development of software tools for semantics oriented access to digitized collections of manuscripts and digitized archival collections [1–3]. Some actual issues in the areas of personalizing digital library access with preference-based queries [4] and advanced search engine technologies [5] are also taken into account.

2. Related Work

During the last decade a series of promising results have been reached in the area of the 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 eContent-plus Programme and coordinated by the German National Library [6].

A large number of successful projects which provide tools for building academic digital libraries and in particular, building digital repositories with learning materials, already exists. Among the most popular representatives of this group are the projects DSpace, Open Educational Resources and Moodle.

DSpace (<http://www.dspace.org>) is an open source software platform for creating digital libraries that enables institutions to:

- Collect and organize the institution's intellectual production; describe and disseminate digital works, digital objects etc.; afford opportunities for full-text retrieval through different search options and network points;
- Distribute the institution's digital works over the Web through a search and retrieval system;
- Provide long-term preservation of digital work.

DSpace holds three types of metadata of archived content:

- *Descriptive Metadata*. Each *Item* has one qualified Dublin Core metadata record. The set of elements and qualifiers used by the MIT Libraries is the default configuration included in the DSpace source code. Other descriptive metadata of items may be held in serialized bitstreams. *Communities* and *Collections* have some simple descriptive metadata (a name, and some descriptive text), held in the DBMS.
- *Administrative Metadata* which include preservation metadata, provenance and authorization policy data. Most of them are held within DSpace's relation DBMS schema. Provenance metadata (text) are stored in the Dublin Core records. Additionally, some other administrative metadata (for example, bitstream byte sizes and MIME types) are replicated in the Dublin Core records so that they are easily accessible outside of DSpace.
- *Structural Metadata*. They include information about how to present an item or bitstreams within an item to an end-user, and the relationships between the constituent parts of the item. For example, consider a thesis consisting of a number of TIFF images, each depicting a single page of the thesis. Structural metadata would include the fact that each image is a single page and the ordering of the TIFF images/pages. Structural metadata in DSpace are currently fairly basic; within an item bitstreams can be arranged into separate bundles as mentioned above. A bundle may also optionally have a *primary bitstream*. This

is currently used by the HTML support to indicate which bitstream in the bundle is the first HTML file to be sent to the browser.

DSpace is advertised as the software of choice for academic, non-profit, and commercial organizations building open digital repositories. It preserves and enables easy and open access to all types of digital content including text, images, moving images, mpegs and data sets. With an ever-growing community of developers committed to continuously expanding and improving the software, each DSpace installation benefits from the next.

Open Educational Resources (OER, <http://www.oercommons.org>) is a popular project which provides open access to more than 20,000 learning materials at the primary, secondary and post-secondary level, covering 6 subject areas: Arts, Business, Humanities, Mathematics and Statistics, Science and Technology and Social Sciences.

OER content is made free to use or share, and in some cases to modify and share again, which is ensured through licensing, so both teachers and learners can really share what they know. OER focus not only on textbooks, but also on full courses, course materials, modules, journals, streaming videos, tests, software, and any other tools, materials, or techniques that are critical in the learning environment.

Moodle (<http://moodle.org>) is one of the most popular course management systems – a free, open source software package designed using sound pedagogical principles, to help educators create effective online learning communities. Moodle's features include: creation of courses, individual logins, student activity tracking, grading, file upload and download, arrangement of assignments and quizzes, wikis, email alerts, moderated blogs and forums, etc.

The focus of the Moodle project is always to give educators the best tools to manage and promote learning, but there are many ways to use Moodle, for example:

- Many institutions use it as their platform to conduct fully online courses, while some use it simply to augment face-to-face courses (known as blended learning);
- Many users like to use the special activity modules (such as Forums, Wikis, Databases and so on) to build richly collaborative communities of learning around their subject matter (in the social constructionist tradition). Others prefer to use Moodle as a way of delivering contents to students (such as standard SCORM packages) and assess learning using assignments or quizzes.

3. Contents and Functional Structure of ProgDL

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

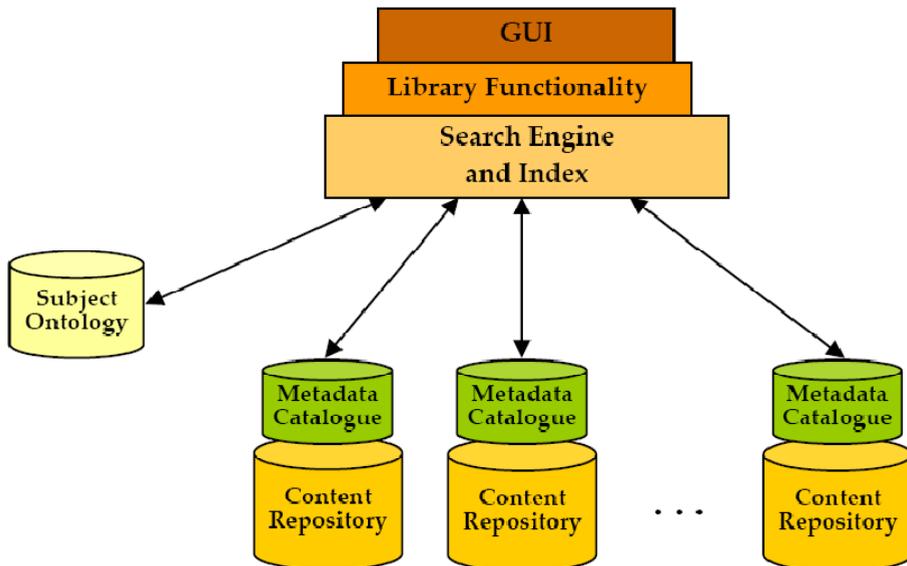


Figure 1: Functional model of ProgDL.

The content repositories include learning materials of different types (textbooks, papers, lecture notes, presentations, exercises, programs, data sets, tests, quizzes etc.) in the fields of data structures and algorithms, procedural, object oriented and functional programming. These library resources are available in various digital formats: pdf, html, plain text, doc, ppt, jpeg etc. Most of them are developed by faculty members, the others are especially selected among the learning materials freely available on the Web. The content repositories are stored in a small number of locations. The learning materials are written in the Bulgarian or English language.

The metadata catalogues are destined to facilitate the identification of the needed learning materials by the search engine. They contain descriptive metadata stored in XML format that comply with the IEEE Standard for Learning Object Metadata [7]. Typical examples of relevant attributes of learning materials are: type of the material; style(s) of programming studied in the material; author; title of the

material; language(s) (human and/or programming one(s)); digital format; location; version; date of creation; completion status; educational level; principal users for which the material was designed; restrictions on use; semantic annotation – list of concepts from the subject ontology describing the Computer Programming sub-fields and/or concepts covered or treated by the material. In this way the metadata catalogues support the reusability of all learning materials and facilitate their interoperability.

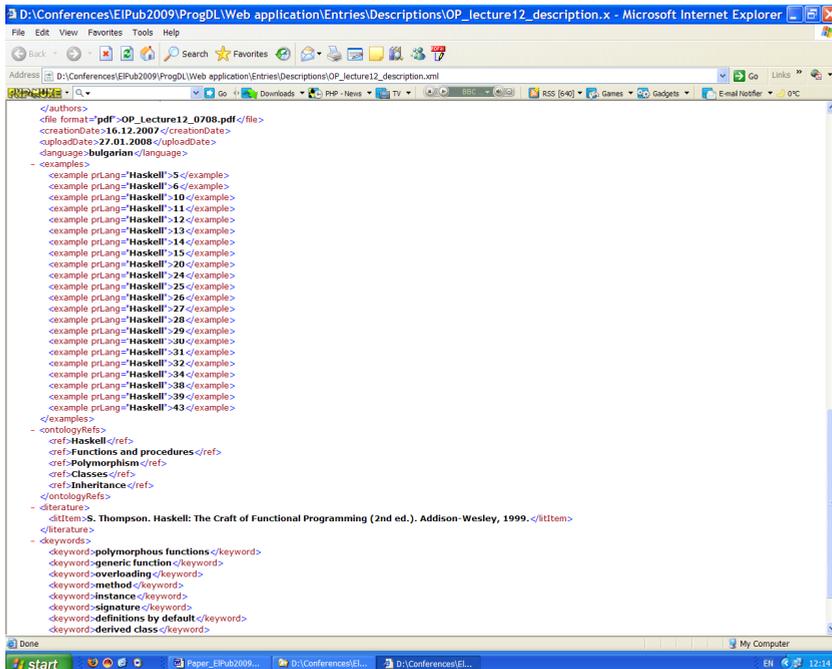


Figure 2: Part of a resource description.

Each catalogue entry (i.e., each resource description) consists of two equivalent parts in which the element values are texts in Bulgarian or English language respectively. The search engine examines the corresponding parts of the descriptions according to the language of the user query.

Figure 2 shows a selected part of a resource description in XML format.

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

The concepts of the subject ontology are too general from the point of view of the expectations of the typical users of ProgDL. For this reason one can include

additional lists of keywords in the resource descriptions which describe the content of the corresponding documents at the necessary level of abstraction. These keywords are set as values of the child elements of the <keywords> resource description elements.

The next sections briefly describe the rest of the components of ProgDL.

4. Subject Ontology

The subject ontology includes a large set of concepts studied in the University courses in procedural, object oriented and functional programming and data structures and algorithms, with description of their properties and the different kinds of relationships among them. This ontology is based on the Computer Science Curriculum 2008 of ACM and IEEE/CS [8]. 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. Its current version includes approximately 300 concepts with their relationships. Figure 3 shows a small part of the subject ontology visualized by the OWLViz tab of Protégé/OWL.

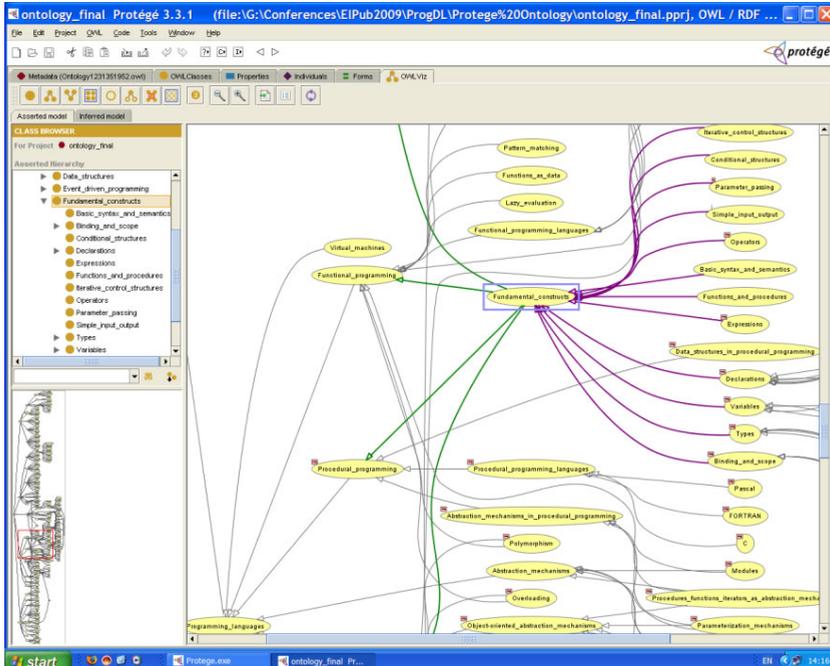


Figure 3: Part of the subject ontology.

The subject ontology is designed in order to play the role of an information source describing the hierarchy and the other relationships between the main concepts in the discussed domain. A dictionary of synonyms has also been under development with the purpose of providing the search engine with other viewpoints to the conceptual structure of the domain of Computer Programming, Data Structures and Algorithms.

5. Types of Users and User Interface

The library functionality and the user interface of ProgDL are designed in accordance with the expected requirements of the basic types of users of the library. The interface module provides adequate online access to the corresponding library resources (Figure 4).

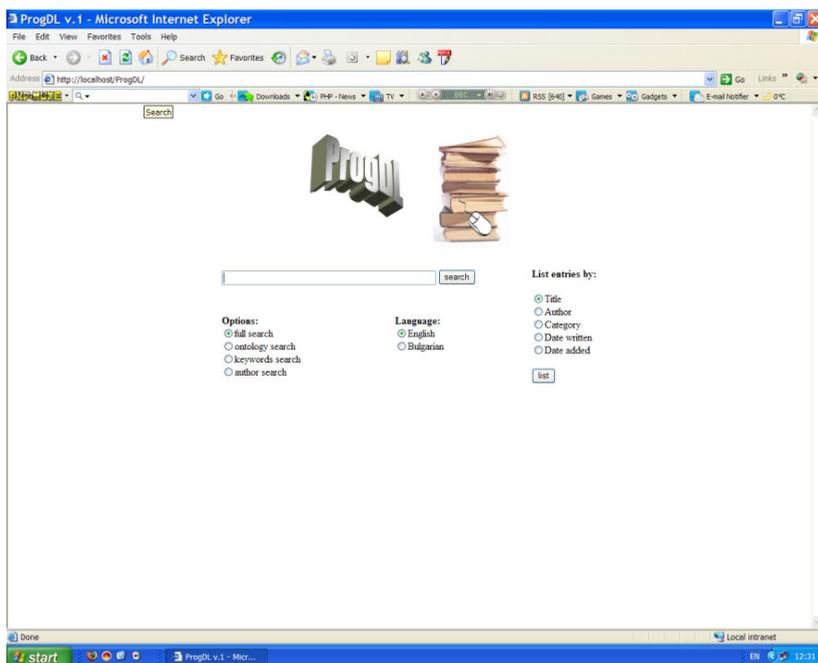


Figure 4: User interface of ProgDL (students' point of view).

The current version of the user interface is intended for four types of users:

- FMI students – they may read/download textbooks, open lecture notes and presentations from all public sections of the library as well as all other kinds of materials (lecture notes, presentations, exercises, programs, data sets, quizzes, tests) from fixed public library sections;

- FMI lecturers – 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);
- general citizen – they may read and download public materials of fixed types (textbooks, open lecture notes and presentations).

6. Working Principles of the Search Engine

The purpose of the search engine is to provide adequate access to the complete palette of resources stored in ProgDL.

The search engine maintains several types of search and document retrieval within ProgDL. The user queries define restrictions on the values of certain metadata attributes of the required learning materials. Generally the search mechanism may be formulated as follows: the document descriptions included in all permissible user sections of the library are examined one by one; the descriptions which have a specific element (determined by the type of the user query) with a value matching the user query are marked in order to form the search result. The matching process is successful if the value of the element or the value of one of its child elements is equal to the user query. The documents pointed by the marked descriptions are retrieved and the user is given access to these documents and their catalogue descriptions.

The current implementation of the search engine supports four types of search and document retrieval:

- full search - search and retrieval of all available learning documents, ordered by title, by author, by category, by date of creation or by date of insertion in the library;
- author search (search and retrieval of the documents created by a given author)
 - the search is performed on the value of the element <authors>;
- ontology search – the search is performed on the value of the element <ontologyRefs>;
- keyword search – the search is performed on the value of the element <keywords>.

During the ontology search the user query is augmented with regard to the concepts searched in the semantic annotations of the required learning materials. The more specific concepts from the subject ontology are added to the original one in

the resulting query. Then the search engine retrieves all library documents containing at least one component of the augmented query in their descriptions as the value of a child element of <ontologyRefs>. In this way the ontology search enables one to find documents described by ontology concepts which are semantically related to the concept defining the user query.

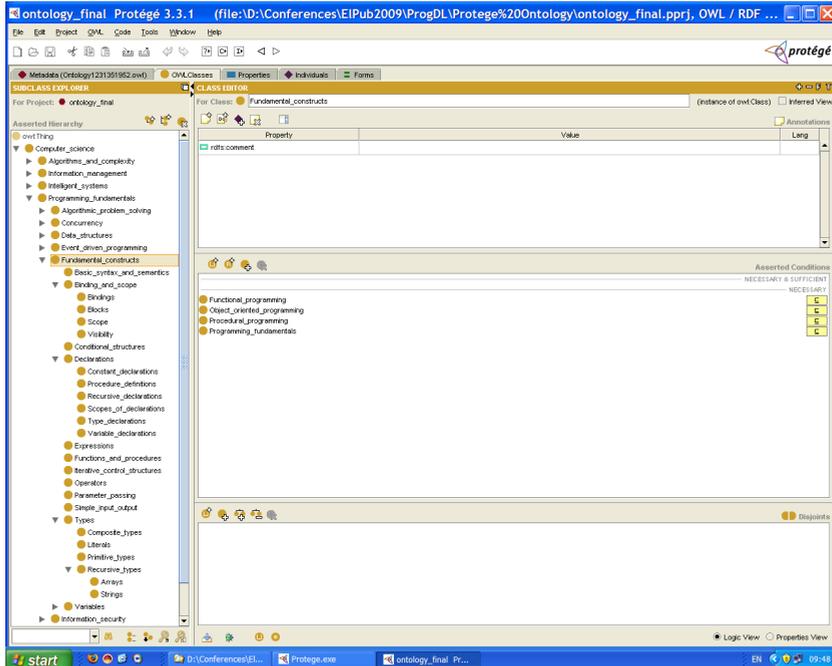


Figure 5: Part of the concept hierarchy.

Example. If the user defines a request (a query) for ontology search concerning the concept “fundamental constructs”, first an extension of this request will be made with all ontological concepts which are special cases of the concept given by the user. For this purpose, breadth-first search in the graph that represents the ontology will be performed, starting from the concept chosen by the user.

In this case an extended request (an augmented query) is generated. It includes the concepts “fundamental constructs”, “basic syntax and semantics”, “binding and scope”, “conditional structures”, “declarations”, “expressions”, “functions and procedures”, ... , “variables”, “bindings”, “blocks”, ... , “simple variables” (see Figure 5).

After that a consecutive search in the catalogue descriptions follows. In this search all documents with descriptions that are juxtaposed with at least one element of the extended request are extracted. In the current implementation each document appears as many times in the result list as many elements of the aug-

mented query are juxtaposed with its description (which means that the element <ontologyRefs> of the description includes a sub-element that has value, coincident with an element of the augmented query).



Figure 6: Some search results for the query "composite types".

Figure 6 shows a screenshot displaying part of the ontology search results for the query "fundamental constructs".

Till now, we have not had a suitable dictionary available of synonyms of the concepts in the area of Computer Programming, neither in Bulgarian nor in English, but our idea is to provide the possibility for a two-stage augmentation of the user query. At the first stage the request for ontology search will be extended with the more specific concepts (its successors) from the subject ontology. At the second stage the synonyms found in the dictionary will be added to the main (given by the user) concept and its successors.

7. Conclusions

The most considerable results obtained of this project up to date may be summarized as follows:

- A functional model of an academic digital library was proposed. This model provides tools for semantics oriented access to learning and research materials in different digital formats;
- A prototype of ProgDL – an academic digital library was developed with learning materials in the areas of data structures and algorithms, procedural, object oriented and functional programming.

Compared to some well-known large scale initiatives like those discussed in Section 2, our project is on a significantly smaller scale. However, differently from all of the other projects, ours investigates the use of a subject ontology which makes it possible to provide more flexible, semantics-oriented access to the library resources for users with different profiles and language skills. The application of free license software is an additional advantage of the suggested approach which can be utilized by developers of institutional digital libraries.

The implementation of the project will help to enhance the exchange of teaching innovation, and thus will improve the overall teaching quality in Computer Science at FMI. It will also increase the students' learning experience and their graduation rates.

Acknowledgements

This work has been partly funded by the Sofia University SRF within the "Methods and Tools for Building Academic Digital Libraries" Project. The authors are thankful to their students Nadejda Apostolova, Anton Iliev and Klimentina Rousseva for their contribution to the implementation of the prototype of ProgDL.

References

- [1] NISHEVA-PAVLOVA, M. and PAVLOV, P. Tools for Intelligent Search in Collections of Digitized Manuscripts. In: M. Dobрева, J. Engelen (Eds.), "From Author to Reader: Challenges for the Digital Content Chain. Proceedings of the 2005 ELPUB Conference". Peeters Publishing Leuven, 2005, p. 145-150.
- [2] PAVLOV, P. and NISHEVA-PAVLOVA, M. Knowledge-based Search in Collections of Digitized Manuscripts: First Results. Proceedings of the 10th ICC International Conference on Electronic Publishing (Bansko, 14-16 June 2006), FOI-COMMERCE, Sofia, 2006, p. 27-35.
- [3] NISHEVA-PAVLOVA, M., PAVLOV, P. and MARKOV, N., NEDEVA, M. Digitisation and Access to Archival Collections: A Case Study of the Sofia Municipal Government (1878-1879). Proceedings of the 11th International

- Conference on Electronic Publishing (Vienna, Austria, 13-15 June 2007), ISBN 978-3-85437-292-9, 2007, p. 277-284.
- [4] GEORGIADIS, P., SPYRATOS, N., CHRISTOPHIDES, V. and MEGHINI, C. Personalizing Digital Library Access with Preference-Based Queries. *ERCIM News*, 66, July 2006, p. 34-35.
 - [5] LERVIK, J. and BRYGFJELD, S. Search Engine Technology Applied in Digital Libraries. *ERCIM News*, 66, July 2006, p. 18-19.
 - [6] The European Digital Library Project. Available at <http://www.edlproject.eu/> (March 2009).
 - [7] The IEEE Standard for Learning Object Metadata. Available at <http://ltsc.ieee.org/wg12/20020612-Final-LOM-Draft.html> (March 2009).
 - [8] ACM Education Council, IEEE Computer Society Educational Activities Board. *Computer Science Curriculum 2008: An Interim Revision of CS 2001*. Available at <http://www.acm.org/education/curricula/ComputerScience2008.pdf> (March 2009).

June 2009
Printed on demand
by "*Nuova Cultura*"
www.nuovacultura.it

Book orders: ordini@nuovacultura.it