IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITB12886

This paper appears in the book, *Emerging Trends and Challenges in Information Technology Management, Volume 1 and Volume 2* edited by Mehdi Khosrow-Pour © 2006, Idea Group Inc.

Integrated Access to Learning Objects Repositories and Digital Libraries

Geórgia R. R. Gomes & Rubens N. Melo
Computer Science Dept, Pontifical Catholic University of Rio de Janeiro, Rua Marques de Sao Vicente 225, Gavea, Rio de Janeiro - RJ, Brazil 22453-900, T: +55-21-3114-1500 x.:3603, F: +55-21-3114-1848, {georgia, rubens}@inf.puc-rio.br

Sean W. M. Siqueira, Informatics Institute, Federal University of Goias, IMF 1, Campus Samambaia, Goiania - GO, Brazil 74001-970, T: +55-62-3521-1505, F: +55-62-3521-1182, sean@inf.ufg.br

Maria Helena L. B. Braz, DECivil, Technical University of Lisbon, Av. Rovisco Pais, Lisbon, Portugal, 1049-001, T: +351-21-8418329, F: +351-21-8418344, mhb@civil.ist.utl.pt

ABSTRACT

Nowadays, Digital Libraries (DLs) are important information resources and have been used to support education. As e-learning has gained attention with the development of the Information and Communication Technologies, DLs are being considered as great enablers of good e-learning environments. However, although DLs and e-learning systems have some common services, they also have some different characteristics and deal with different types of documents. Usually DLs deal with bigger documents while e-learning systems use documents that are developed according to object oriented principles aiming at providing content reuse. Therefore, e-learning systems usually deal with documents called Learning Objects (LOs). Besides the differences on the documents, the metadata structures that are used to describe these documents are also different, which makes an open problem their search and retrieval in an integrated way. In this paper we describe a general architecture for providing integrated access to distributed and heterogeneous LO repositories and DLs. This architecture is based on information integration techniques: mediators and wrappers. A prototype has been developed considering the integration of metadata for LOs with DL metadata and it has been implemented through web-services and ontologies, which provide more flexibility on the implementation of the integration components while supporting semantic treatment through a more formal representation.

1 – INTRODUCTION

Digital Libraries (DLs) have been developed worldwide to support education. There are several concepts of a DL in the literature [1], [2], [3], but there is a consensus agreement that DLs are complex. In [4], some abstrations are presented to model DLs in order to make their development easier. Considering the DL documents, it is possible to notice that they are complex objects that are developed to be a consistent and complete unit of knowledge. These documents are usually described by standard metadata structures such as Dublin Core (DC) [5] and Machine Readable Catalog (MARC) [6], [7].

By the other side, e-learning has gained attention with the development of the Information and Communication Technologies. Education, training and learning can be understood through different aspects such as content, pedagogy and technology. In this paper, the term e-learning is used in the context of web-based education, i.e., the technology is based on the web. As developing high quality web-based learning content is time-consuming and expensive, content reuse approaches such as the development of learning objects (LOs) are essential. LOs are defined in this paper as reusable digital entities that may be used for learning, education or training. They are usually stored in Learning Objects Repositories, referenced through metadata and accessed through Learning Content Management Systems (LCMSs) while the e-learning system is called Learning Management Systems (LMSs). There are several metadata standards proposals for describing LOs, such as Dublin Core, Ariadne [8] and IEEE LOM [9].

It is straightforward that DLs represent good sources of information for complementing learning resources. Therefore, LMSs and DL systems should be combined, which means their services and repositories should be integrated. It could be accomplished in two different ways:

a) the services and repositories should be planned and developed (since the beginning) to result in only one system; or
b) there should be an integration mechanism for making the different repositories to be accessed as one, while the services should be generalized to satisfy both requirements.

In this paper we describe an architecture for providing integrated access to distributed LO repositories and DLs. This architecture is based on information integration techniques, in special mediators and wrappers. A prototype is under development and considers the integration of metadata for LOs with DL metadata. It has been implemented through web-services and ontologies.

The remainder of this paper is organized as follows: in section 2 we present the general architecture to integrate LO repositories and DLs; in section 3 we describe the prototype implementation and the case study; in section 4 we describe how to deal with the heterogeneity and in section 5 we present some final remarks.

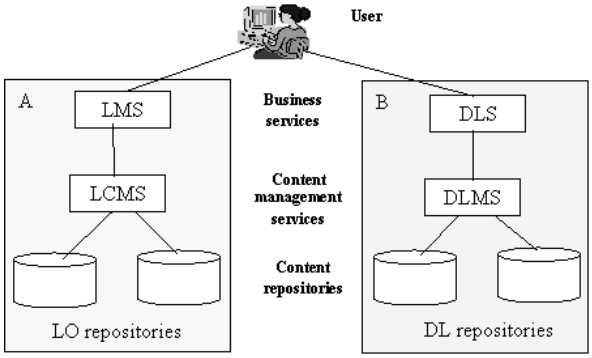
2 – THE PROPOSED ARCHITECTURE FOR INTEGRATING LO REPOSITORIES AND DLS

Figure 1 shows the e-learning (A) and the DL (B) environments. The e-learning environment (A) has the LO repositories, which are accessed through the LCMSs. The LMS provide the functionality and interaction with the end users. Notice, however, that LCMSs usually access/manage only one LO repository. Global repositories of instructional material are common on the Internet, such as Campus Alberta Repository of Educational Objects (CAREO) [10] and Multimedia Educational Resource for Learning and Online Teaching (MERLOT) [11].

An interesting centralization approach is the EducaNext/Universal Project [12], which deals with brokers. Moura et al [13] presented an approach, in which LO metadata from distributed and heterogeneous sources are replicated throughout the e-learning community. One example of a successful P2P implementation in e-learning is the Edutella project [14]. Moura et al [15] describe a mediators and wrappers architecture, which is implemented through web-services and ontologies for providing integrated access to heterogeneous LO repositories.

The DL environment (B) has an application or system (DLS) that provides the interaction with the end user. The DLS uses a management system to access/manage the repositories of digital documents. It is also

Figure 1. E-Learning Vs. DL Environments



important to notice that each DL usually focuses a specific repository. Some approaches for data integrating in traditional libraries can also be used for DLs, for example the one described in [16].

E-Learning and DL environments can be seen in three levels (Figure 1): the content repositories, which represent the storage unit for the digital documents; the content management services, which are responsible for providing query capabilities as well as metadata management, indexing, cataloging, documents retrieval etc.; and business services, which are related to the specific environment, such as workgroup, communication services etc. Usually the user has two different environments that are not integrated.

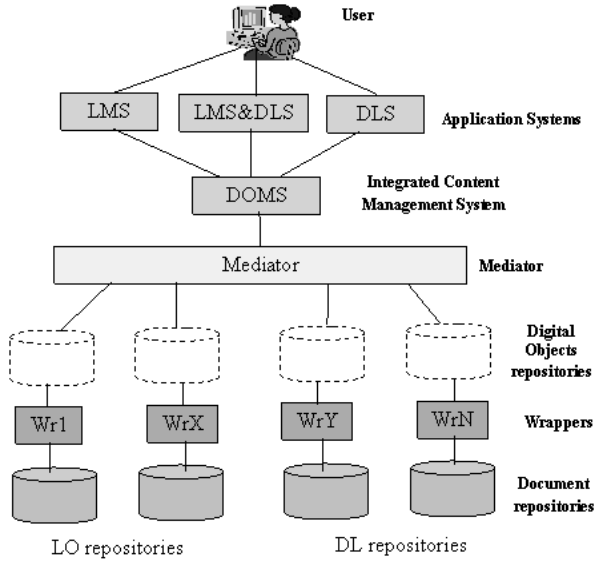
In the proposed architecture we use an approach based on information integration techniques. Data integration has been researched for several years in the database area. One of the possible and consolidated ways to treat information integration is the use of mediators and wrappers [17], [18]. The use of wrappers to encapsulate data sources allows the sources to be maintained and evolve with some independence. The wrappers can be seen as a communication mechanism to provide access to the data sources independently of their format and implementation. The mediators are used to provide the uniform and integrated access to the information through the wrappers. A mediator has a set of articulations that represent the relations between their terms and the terms of their sources. The mediator uses these articulations to define the queries to be submitted to the data sources from a global query. Also, the mediator should combine the answers of each data source and return a consolidated result to the global application.

Therefore, the proposed architecture (Figure 2) for providing integrated access to LO repositories and DLs emphasizes the different documents structures. According to our approach, the LOs and digital documents of the DLs must be restructured and re-organized so that they have the same semantics and structural guidance. It is accomplished by the use of wrappers and a common ontology, generating the Digital Objects (DOs) repositories, which can be virtual or materialized. After that, a mediator is used to integrate the DOs, which are managed by a DOMS (Digital Objects Management System). Then, the user can access these objects through LMS, DLS or even an integrated system that has the services of both systems.

3 – THE PROTOTYPE AND THE CASE STUDY

A prototype has been developed at our research lab considering the proposed architecture. However, this prototype considered only metadata integration. The initial prototype considers wrappers from two sites: TecBD research lab (with LOM metadata) and an extract of PUC-Rio's DL (with MARC metadata) as well as the mediation components. The prototype has been implemented through web-services and ontologies and has partially reused the work developed in a previous work [13] that aimed at integrating LO repositories.

Figure 2. An Overview of the Proposed Architecture



Web-Services are defined by the W3C (World Wide Consortium) as a software application or component that is identified by a URI (Universal Resource Identifier), whose interfaces and connections are capable of being defined, described and discovered in XML and supports interactions directly with other software applications using XML coded messages via Internet-based protocols. They provide interoperability among software components because they are based on standard mechanisms and protocols. The use of ontologies to describe the implicit knowledge of data sources is interesting to solve problems of semantic heterogeneity. Ontology is usually defined as the explicit specification of a concept [17]. The construction of common ontologies has been proposed as a promising approach to the interoperability of systems.

Figure 3 shows an overview of the prototype for the case study. In the prototype, the Application Layer allows the client applications (with LMS and DLS services) to access the Mediation Layer, which offers the integration services. This layer accesses an integrated representation of data sources, using a global query language. In order to allow an integrated query over the DL and LO repositories, it is important to add these services on the applications that interact with the user or enable the user queries directly activate the integration services in order to get an integrated answer.

The Mediation Layer gets the applications' global queries, interprets and translates them, so that it is possible to access a uniform and transparent view of the data sources (i.e., an integrated view of the LOs and DL repositories). Therefore, this layer receives the global queries from the Application Layer and makes the homogenization of heterogeneous data sources. The global queries are validated according to a Global Ontology (common schema). Then the mediator identifies which sources should be accessed, transforming the user queries in sub-queries, sending them through a standard codification to the corresponding wrapper (Wr) of each data source. The Global Ontology represents an integrated schema, with the concepts of both standards, while the mediator has the knowledge of the available data sources (DLs and LO repositories).

The Data Access Layer (containing wrappers and local ontologies) receives the sub-queries from the mediator and accesses the local ontologies, which contains the mappings of the global scheme (integrated scheme) to the local data source scheme (LOM or MARC). Then, the wrappers transform the sub-queries into corresponding queries according to the query language of each data source. Therefore, in order

to access the DL or LO repository, its corresponding wrapper should get the respective sub-query and access the local ontology that maps the global schema into the local schema and translates the sub-query to the adequate query to the DL or LO repository.

The result of a query in the local data source is sent back to corresponding wrapper that translates the answer according to the global language (through the local ontology). Then, the wrappers send the results to the mediator that centralizes all the answers. The mediator executes the necessary operations and finally sends the integrated answer to the user. Therefore, information about the learning objects and digital documents that correspond to the global query are presented to the user that can access/download them.

Therefore, in each site (our research lab and our library) we implemented the translation and data access services. In the research lab, we used .Net as development platform, C# as programming language and DB4Objects to store LOM metadata. In the library, we used Java as development language and SQL Server to store the MARC metadata. The JDBC and DOM APIs were used, respectively, in the access to the DBMS and for manipulating XML documents.

In the development of the mediation components we used Jena for storing ontologies and taxonomies in OWL and RDF descriptions as XML. The mediation and translation services were described as web-services using WSDL, allowing the interoperability among the applications that were developed in different platforms and programming languages.

4 – DEALING WITH THE HETEROGENEITY

In order to treat structural and semantic heterogeneity, we used mappings supported by ontologies (Figure 4). First, to each metadata standard that is used in the case study (LOM and MARC), there is an equivalent description in RDF containing its respective structure. Similarly we created a common ontology from each standard, containing a generalization of concepts such as author, title and keyword. When considering the metadata represented in the different standards for educational material and for digital bibliographic resources, it is important to have a "semantic understanding" among the concepts expressed by these standards. Our global/common scheme takes into account concepts from bibliographic and e- learning metadata standards. LOM and MARC were the base for the common scheme (and were chosen for the case study) because they are the most used and cover great amount of the metadata represented in other standards.

In addition to considering metadata elements, it was also important to take into account the reference values. Therefore, complex elements such as subject/keyword were linked to taxonomies and only valid values were mapped. The concept keyword has a taxonomy defining its possible values. This taxonomy is based on the Library of Congress and Brazilian National Library standard values. Then, we defined mappings from the common ontology to the schema of the metadata standards. The descriptions in RDF, as well as the mapping ontologies are stored in the ontology repository.

The ontologies were represented in OWL - Web Ontology Language [19]. The OWL aims at providing a language that can be used to describe classes and relationships among them that are inherent to web documents and applications. This language can be used to formalize a domain through the classes' definition and their properties; to define individuals and to assert properties about them and to provide logical reasoning about these classes and individuals according to the degree that is allowed for the formal semantic of OWL.

5 – CONCLUSION

The architecture presented in this paper aims at providing the integration of LO repositories and DLs. This architecture provides to the users a transparent and integrated view of the learning objects and digital documents that are stored in the data sources. This integration is independent of data model, query language, operational system and localization. A prototype of the proposed architecture has been implemented to integrate digital documents of PUC-Rio's DL and LOs from TecBD lab, considering MARC and LOM metadata standards.

There are other works in the specialized literature presenting the use of DLs for web-based learning, but they have different approaches. Ilumina [20] and DILLEO [21] have a DL of LOs. A tentative mapping of LOM and MARC standards can also be found at [22]. LEBONED Project [23] (Learning Environment Based on Non Digital Educational Libraries) considers an architecture to integrate DL into a Learning Management

Figure 3. An Overview of the Prototype for the Case Study

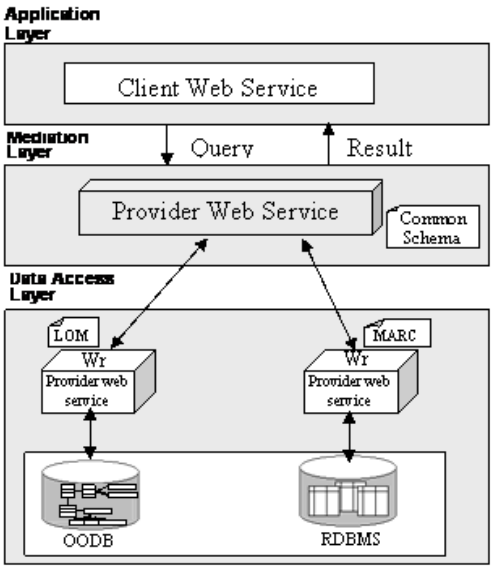
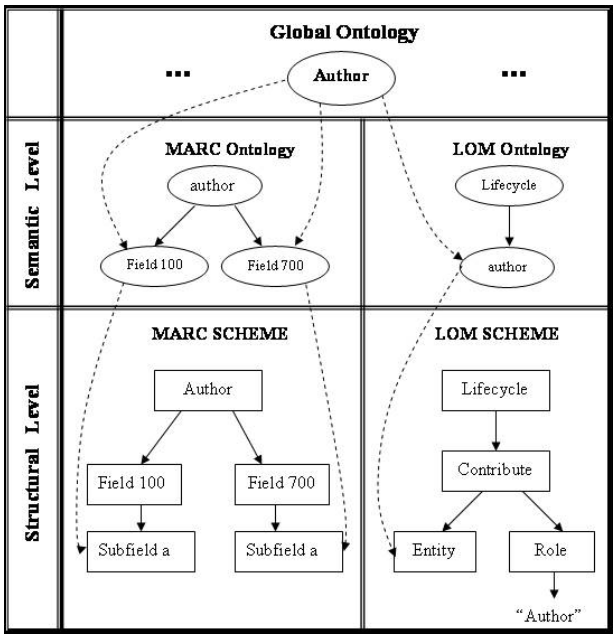


Figure 4. Concept Mappings



System (LMS). They use an extension of SCORM standard [24] (which is based on IEEE LOM) in order to support the METS [25] standard (which is based on MARC). LEBONED uses wrappers to export data from the DLs to the extended SCORM repository.

The architecture presented in this paper is more general than that used in LEBONED. Although we are still working on a digital object model, which would incorporate concepts from DL documents and LOs, in this paper we also presented a prototype that has been implemented through web-services and uses ontologies to provide mappings among DL and e-learning standards. The use of web services and ontologies as considered in this work allows easier and more flexible implementation of the components of the architecture, the definition of better mapping rules, as well as a better semantic orientation through the representation of the respective metadata schemes. Therefore, integration processes are more reliable and the corresponding semantics are better represented.

Although we used LOM and MARC in the case study, any other standard for describing digital documents or learning objects could be considered. We defined a global ontology, with an integrated view of the LO repositories and DLs. It is an integrated virtual database with the digital documents and LOs of all data sources, but we intend to have a materialized database with de common semantics of these documents.

Data sources can be added to the integration environment and then a series of conflicts (heterogeneity) needs to be treated. The proposed approach, through the use of ontology and mediators, has more flexibility and versatility when a new data source is added. If a new data source uses a data model that already exists in a local scheme, it can use the same wrapper of this other local source, thus allowing code reuse. As future works it is important to define wrappers to translate the documents and LOs (instead of the metadata standards) and enrich the use of taxonomies, for example treating values that were not initially considered in the ase study. It is also important to generate educational values semi-automatically for bibliographic resources.

ACKNOWLEDGMENTS

The authors would like to thank professionals and researchers that have helped us in our work, specially the database technology group from PUC-Rio (TecBD), the e-learning technology research group from ICIST/DECivil and the Pergamum development group. This work has been partially supported by FCT Portugal – Foundation for Science and Technology, through the Multi-annual and Programmatic funds of ICIST.

REFERENCES

[1] Levy, D. M. And Marshall, C. C. Going digital: a look at assumptions underlying digital libraries. Commun. ACM 38, 8, 77–84, 1995.

[2] Borgman, C. L. What are digital libraries? competing visions. Inf. Proc. Manage. 35, 3 (Jan.), 227–243. 1999

[3] Kochtanek, T. And Hein, K. K. Delphi study of digital libraries. Inf. Proc. Manage. 35, 3, 245–254. 1999

[4] Gonçalves, M. A., Fox, E. A., Watson, L. T., and Kipp, N. A. Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries, ACM Transactions on Information Systems, Vol. 22, No. 2, April 2004.

[5] Dublin Core Metadata Initiative. http://dublincore.org/

[6] Library of Congress. Network Development and MARC Standards Office. MARC21 format for bibliographic data: including guidelines for content designation. 2000 ed.

[7] Library of Congress, Cataloger's Desktop [CD-ROM].n.2, 2004

[8] ARIADNE Foundation for the European Knowledge Pool. http://www.ariadne-eu.org/

[9] Learning Technology Standards Committee of the IEEE: LOM Draft Standard 1484.12.1-2002. http://ltsc.ieee.org/wg12/20020612-Final-LOM-Draft.htm.

[10] CAREO "CAREO: Campus of Alberta Repository of Educational Objects", Last Access: December, 2005. URL: http://www.careo.org

[11] MERLOT "Multimedia Educational Resource for Learning and Online Teaching", Last Access: December, 2005. URL: http://www.merlot.org

[12] Law, E., Maillet, K., Quemada, J., Simon, B.. "EducaNext: A Service for Knowledge Sharing", In:*Proceedings of the Ariadne Conference*, Leuven, Belgium, 2003.

[13] Moura, S. L., Coutinho F., Siqueira S. W. M., Melo, R. N., Nunes, V. Jr.:Integrating Repositories of Learning Objects Using Web-services to Implement Mediators and Wappers, In Proceedings of the International Conference on Next generation Web Services Practices (NWeSP'05), Seoul, Korea.

[14] Wolf., B. "Peer-to-Peer Networking for Distributed Learning Repositories V1.0." Institut für Rechnergestützte Wissensverarbeitung. Last Access: December, 2005. URL: http://www.kbs.unihannover.de/Arbeiten/Diplomarbeiten/02/Edutella.pdf.

[15] Moura, L. S., Silva, F. J. C., Siqueira, S. W. M. and Melo, N. R., Integrating Repositories of Learning Objects Using Web-Services and Ontologies, International Journal on Web Services Practices (IJWSP), Seoul, v. 1, n.1, 2006 (to be published)

[16] Gomes, G. R. R., Melo, R. N. and Cortes, S. C., "An Informatics Architecture for Integrating Library Systems in the Internet", In: Proceedings of the National Seminar on Universitary Library, Florianópolis, Brazil. SNBU2000. Florianópolis: CliCData Multimídia, 2000 (in Portuguese).

[17] Gruber, T.: Towards principles for the design of ontologies used for knowledge sharing. International Journal of Human and Computer Studies, v43, n.5/6, pp. 907-928, 1994.

[18] Tzitzikas, Y., Spyratos, N., Constantopoulos, P.: Query Translation for Mediators over Ontology-based Information Sources. In Proceedings of the Second Hellenic Conference on AI: Methods and Applications of Artificial Intelligence 2002 pp.423 - 436 , Lecture Notes In Computer Science, Springer-Verlag London, UK.

[19] Web Ontology Language (OWL) Reference Version 1.0. W3C Working Draft 21 February 2003. http://www.w3.org/TR/2003/WD-owl-ref-20030221/

[20] Ilumina, http://www.ilumina-dlib.org/browse.asp

[21] Mikulecký, S.,Maruna, Z., Smr´ka, I.; DILLEO – Digital Library of Learning Object. https://dilleo.osu.cz/dilleo_v2/About.aspx

[22] Qin, J.: Crosswalk of MARC 21 and IEEE LOM. 2002 http://web.syr.edu/~jqin/LO/Metametadata_crosswalk.html#extended

[23] Oldenettel, F., Malachinski, M., Reil D.: Integrating Digital Libraries into Learning Environments: The LEBONED Approach. In Proceedings of the 3rd ACM/IEEE-CS: Join Conference on Digital Libraries 2003. pp. 280- 290

[24] Advanced Distributed Learning, SCORM 2004 2nd Edition Overview, http://www.adlnet.org/scorm/history/2004/documents.cfm

[25] Library of Congress, METS: An Overview & Tutorial. http://www.loc.gov/standards/mets/METSOverview.v2.html