

The Construction of a Spanish Educational Repository: Advances and Future Directions of the DOTEINE Project

Miguel Ángel Marzal¹, Javier Calzada-Prado¹,
M^a Jesús Colmenero Ruiz¹, and Pilar Beltrán Orenes²

¹ Library & Information Science Department, Carlos III University of Madrid, Spain
{mmarzal, fcalzada, mcolmene} @bib.uc3m.es

² Library & Information Science Department, Complutense University of Madrid, Spain
mpbeltran@ccinf.ucm.es

Abstract. This paper presents the advances and future directions of the DOTEINE project, whose main goal is to develop advanced instruments for document organization and retrieval that can be used to support learning in educational digital libraries. It exposes the development of an educational repository as well as an educational metadata application profile and its corresponding vocabularies to describe the learning resources gathered in it. Finally, it discusses the possibilities offered by the tools used by the project to transform semantic structures into navigational Topic Maps and Ontologies and their application to educational digital libraries.

Keywords: Educational Repositories; Digital Libraries; Metadata; Application Profiles; Controlled Vocabularies; Thesauri, Topic Maps; Ontologies.

1 Introduction

The DOTEINE¹ project is a three-year CICYT² funded project whose main goal is to develop advanced instruments for document organization and retrieval that can be used to support learning from educational digital libraries in Learning Resources Centers (LRC) and Learning and Research Resources Centers (LRRC). An annex project, the IACORIE³ project, has supported the ongoing work of the IACORIE project undertaking the development of some of its phases. Both projects had the following specific goals:

- To analyse the possibilities and difficulties of the currently available tools for storing, representing, organizing and retrieving educational resources (learning objects as well as informative digital documents with a potential educational value) in institutional settings.
- To develop advanced instruments in Spanish for educational information organization and retrieval based on semantic structures (thesauri, ontologies, concept maps and Topic Maps).
- To explore the integration of such instruments into LRC and LRRC and the evaluation of their educational impact.

This contribution presents the main advances and future directions of the ongoing work.

2 Advances: Development of the Repository and Thesauri

2.1 Development of an Educational Metadata Application Profile

After defining the educational domains we were going to work with (Biology and Language & Literature for Primary and Secondary education, Law for the undergraduate level and Content Analysis for the graduate level), we designed a descriptive model for the representation of documents belonging to such domains that will be gathered in the repository. Once we analysed the main educational digital library projects, we found that there are two predominant standards in the educational domain: the IEEE *Learning Object Metadata* [1] and *Dublin Core* [2]. After analysing the pros and cons of both of them and observing the current draft of the ISO/IEC JTC1/SC36 standard, we decided to adopt an application profile that combined the

¹ *Documentación y tecnologías de la información para Educación: herramientas para la alfabetización informacional y organización de recursos didácticos*

² Spanish Inter-Ministerial Commission for Science and Technology.

³ *Instrumentos de Análisis de Contenido para la organización de recursos informativos en Educación: centros de recursos de aprendizaje para la Sociedad del Conocimiento*

simplicity of DC and the semantic richness of LOM. Such application profile was called MIMETA (IACORIE Metadata Model). We distinguish two element groups in it: a group of elements used to describe the general characteristics of the resource and another group of elements used to describe the educational content of the resources. The elements belonging to the first group are described in Table 1.

Table 1. Main elements of the MIMETA application profile.

Element	Description	Type
<identifier>	Code number	Mandatory, non-repeatable
<author>	Author/s	Mandatory, repeatable
<author_type>	Type of authorship	Optional, repeatable
<institution>	Institutional authorship	Optional, repeatable
<title>	Title of the resource	Mandatory, non-repeatable
<date>	Publication or release date	Mandatory, non-repeatable
<language>	Language/s of the resource	Mandatory, repeatable
<description>	Description of the resource	Mandatory, non-repeatable
<format>	Format of the resource	Mandatory, repeatable
<type>	Resource type	Mandatory, repeatable
<rights>	Property rights involved	Mandatory, non-repeatable
<subject>	Subject matter, controlled	Mandatory, repeatable
<keywords>	Keywords, uncontrolled	Mandatory, repeatable
<relation>	Relation considered: "Is part of"	Optional, repeatable

The second group of elements received a different treatment, given its relevance for document retrieval and reuse and the existing differences between the educational models described by the standards and the Spanish educational model. First, we made a detailed quantitative analysis of the educational elements used in several educational repositories (ARIADNE, EdNA Online, EduSource, *The Gateway to Educational Materials* GEM, MERLOT and *The Learning Federation* SOCCI), discovering five semantic categories among them which describe: a) characteristics related to the kind of user the resource is particularly suitable for; b) characteristics related to the educational context or level in which resource might be used; c) pedagogical characteristics of the resource; d) the kind of resource it is described; and e) the time that the user will need to use or accomplish the educational goals associated to the resource. After this, we contrasted the semantic categories with our own descriptive needs and defined six educational elements, which are described in Table 2.

Table 2. Educational elements of the MIMETA application profile.

Element	Description	Type
<user_type>	Kind of potential user	Mandatory, repeatable
<context>	Educational context/level	Mandatory, repeatable
<difficulty>	Degree of difficulty	Mandatory, non-repeatable
<prerequisites>	Previous knowledge	Optional, repeatable
<interactivity_type>	Kind of interaction of the resource	Mandatory, non-repeatable
<pedagogical>	Pedagogical methods and goals	Optional, repeatable

2.2 Development of the Repository

The technical implementation of our repository has been developed by the CRISOL⁴ group of the Carlos III University of Madrid. They have developed a SQL database that follows the MIMETA model, providing us with a tool to assign external metadata to resources. The user interface has been developed with JavaServer Pages and offers different priorities of access, depending on the user profile. The repository has been conceived to be open access to the targeted communities, but we still have to decide the convenience and implications of allowing all users to contribute resources.

2.3 Development of the Controlled Vocabularies

Once we had developed the repository and a metadata application profile, we focused on the vocabularies we would need to describe educational resources. After accepting most of the vocabularies and value namespaces recommended by the reference standards, we decided to develop the vocabularies corresponding with the <subject> and <type> elements.

To construct the subject vocabulary, we made use of an automatic indexing and thesauri construction tool developed in a precedent research project of our university. First, we designed a method to select a group of documents representative of each educational domain covered. Such method followed the principles of the Linguistic Corpora methods and comprised the following stages:

- 1) Selection of controlled vocabularies in the chosen domains to provide the indexing tool with a basic and specific conceptual structure.
- 2) Selection of intermediate linguistic resources like glossaries and dictionaries to complete the conceptual structure.
- 3) Selection of natural language resources (articles and learning objects) representative of the chosen domains and freely available on the Internet.

These terminology instruments and documents along with the results of a categorization of verbal structures belonging to the educational domain, were used then to feed the indexing tool. Once the resulting terms to each domain were refined and organized according to the ISO standard [3], we had four thesauri: Biology Thesaurus, Language & Literature Thesaurus, Law Thesaurus and Content Analysis Thesaurus⁵. All of them have been made available online through *Tematres*⁶ open access thesauri management software, adapted by our development team from CRISOL to our specific needs. Figure 1 shows a partial view of the hierarchy of the Biology Thesaurus, presenting the classical relationships between terms: equivalence, hierarchical and associative.

⁴ Centro de Recursos de Información y Software Libres. Available at: <http://crisol.uc3m.es/>

⁵ Available online at: <http://doteine.uc3m.es/tesauros/index.htm>

⁶ Developed by Diego Ferreyra. Available for download at: <http://www.r020.com.ar/tematres/>

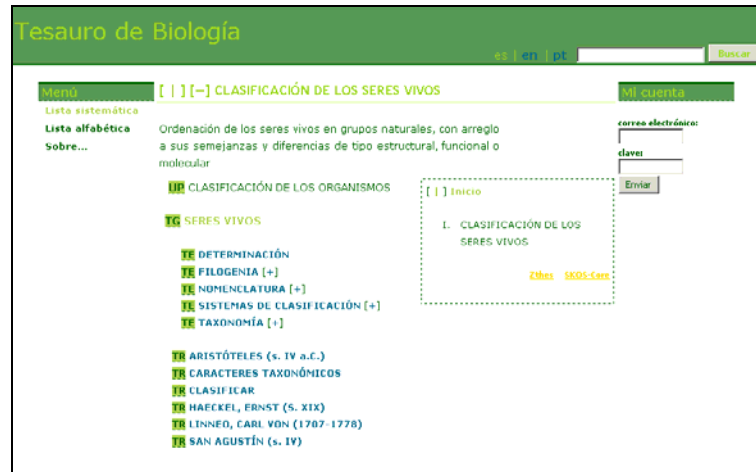


Fig. 1. Biology Thesaurus. Partial view of its hierarchy.

The document type controlled vocabulary was designed to support a specific retrieval need in educational communities: the ability to retrieve documents not only by their content matter but also by their type and function. This is specially significant for teachers as they need to retrieve learning objects and learning object components in order to reuse them in their own materials. The vocabulary, designed in a faceted way, describes the existing variety of learning materials, allowing flexibility and semantic richness for their description. Its basic structure is shown below:

EDUCATIONAL DOCUMENTS

By content

INFORMATIVE

EDUCATIONAL

By aggregation level

UNITS OF LEARNING

COURSES

MODULES

LESSONS

LEARNING OBJECTS

COMPONENTS

By function

PRESENTATIONS

INTRODUCTIONS

LEARNING OBJECTIVES

EXPOSITIONS

THEORETICAL CONTENTS

DEFINITIONS

GLOSSARIES

ILLUSTRATIONS

EXAMPLES

DEMONSTRATIONS

SIMULATIONS

EVALUATIONS

EXERCISES

TESTS

SELF-EVALUATIONS

3 Future Directions: Development of Advanced Instruments for Document Organization and Retrieval

Subject vocabularies have been conceived by the DOTEINE project not only to be useful for retrieval and reuse purposes, but also as a learning device. Semantic representations like concept maps have traditionally supported learning in educational settings by fixing mental models and discovering new concepts and relationships between them [4], [5]. Similarly, the implementation of richer, user-tailored semantic relationships within thesauri will improve discovery learning and retrieval in educational digital libraries⁷. Topic Maps and ontologies seem to be promising in various ways: defining richer semantic relationships, visualizing knowledge structures and relating concepts to their corresponding resources, allowing organization and retrieval in a way that cannot be only interpreted by humans but also by machines.

3.1 Transforming Vocabularies into Topic Maps

Since *Tematres* allows to export our thesauri in Zthes format (see example below), which can be transformed into XTM, we are ready to explore various possibilities:

- Visualizing them in an almost direct way via XSLT stylesheets.
- Using them by means of a Web application generated with XTM documents.
- Adopting an open-access Topic Map processor to allow its visualization through graphs, converting it into a navigational interface.

Example of a Zthes exportation of a descriptor in the Biology Thesaurus

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
  <!DOCTYPE Zthes (View Source for full doctype...)>
= <zthes>
= <term>
```

⁷ As already approached by experiences like Aquabrowser. Aquabrowser demonstration of the Queens Library Online Catalogue available at: <http://aqua.queenslibrary.org/>

```

<termId>2</termId>
<termName>CLASIFICACIÓN DE LOS SERES VIVOS</termName>
<termType>PT</termType>
<termNote>Ordenación de los seres vivos en grupos naturales,
con arreglo a sus semejanzas y diferencias de tipo
estructural, funcional o molecular</termNote>
= <relation>
<relationType>NT</relationType>
<termId>40</termId>
<termName>DETERMINACIÓN</termName>
<termType>PT</termType>
</relation>
= <relation>
<relationType>NT</relationType>
<termId>6</termId>
<termName>FILOGENIA</termName>
<termType>PT</termType>
</relation>
= <relation>
<relationType>NT</relationType>
<termId>11</termId>
<termName>NOMENCLATURA</termName>
<termType>PT</termType>
</relation>
= <relation>
<relationType>NT</relationType>
<termId>30</termId>
<termName>SISTEMAS DE CLASIFICACIÓN</termName>
<termType>PT</termType>
</relation>
= <relation>
<relationType>NT</relationType>
<termId>4</termId>
<termName>TAXONOMÍA</termName>
<termType>PT</termType>
</relation>
</term>
</Zthes>

```

3.2 Development of educational domain ontologies

We consider that ontologies can contribute to our purposes in two ways. Transforming the MIMETA application profile into an ontology could provide us with a conceptual framework to describe educational resources independently of their subject domain. Complementarily, transforming our controlled vocabularies into domain ontologies would prepare them for the Semantic Web. Since *Tematres* allows to export our thesauri in SKOS-Core format (see example below), we are ready to test its conversion to ontologies and integrate them in digital libraries⁸.

Example of a SKOS-Core exportation of a descriptor in the Biology Thesaurus

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
<!DOCTYPE rdfs (View Source for full doctype...)>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"

```

⁸ As suggested by the European SWAT project, available at <http://www.w3.org/2001/sw/Europe/showcase/digital-libraries.html>

```

xmlns:skos="http://www.w3.org/2004/02/skos/core#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
<skos:ConceptScheme
rdf:about="http://doteine.uc3m.es/tesauros/biologiaindex.php">
<dc:title>Tesouro de Biología</dc:title>
<dc:creator>Grupo Doteine</dc:creator>
<dc:subject>BIOLOGÍA; BACHILLERATO; TESAURUS DE EDUCACIÓN</dc:subject>
<dc:description>Tesouro para los materiales curriculares de las
asignaturas de Biología de nivel de bachillerato del sistema educativo
español (17-18 años). Desarrollado por Grupo DOTEINE</dc:description>
<dc:publisher>Grupo Doteine</dc:publisher>
<dc:date>2005-10-09</dc:date>
<dc:language>es</dc:language>
</skos:ConceptScheme>
<skos:Concept
rdf:about="http://doteine.uc3m.es/tesauros/biologiaindex.php?tema=2">
<skos:prefLabel>CLASIFICACIÓN DE LOS SERES VIVOS</skos:prefLabel>
<skos:altLabel>CLASIFICACIÓN DE LOS ORGANISMOS</skos:altLabel>
<skos:scopeNote />
<skos:inScheme rdf:resource="http://doteine.uc3m.es/tesauros/biologia"
/>
<skos:related
rdf:resource="http://doteine.uc3m.es/tesauros/biologiaindex.php?tema=8"
/>
<skos:related
rdf:resource="http://doteine.uc3m.es/tesauros/biologiaindex.php?tema=25" />
<skos:related
rdf:resource="http://doteine.uc3m.es/tesauros/biologiaindex.php?tema=28" />
<skos:related
rdf:resource="http://doteine.uc3m.es/tesauros/biologiaindex.php?tema=23" />
<skos:related
rdf:resource="http://doteine.uc3m.es/tesauros/biologiaindex.php?tema=10" />
<skos:related
rdf:resource="http://doteine.uc3m.es/tesauros/biologiaindex.php?tema=9" />
<skos:broader
rdf:resource="http://doteine.uc3m.es/tesauros/biologiaindex.php?tema=1" />
<skos:narrower
rdf:resource="http://doteine.uc3m.es/tesauros/biologiaindex.php?tema=40" />
<skos:narrower
rdf:resource="http://doteine.uc3m.es/tesauros/biologiaindex.php?tema=6" />
<skos:narrower
rdf:resource="http://doteine.uc3m.es/tesauros/biologiaindex.php?tema=11" />
<skos:narrower
rdf:resource="http://doteine.uc3m.es/tesauros/biologiaindex.php?tema=30" />
<skos:narrower
rdf:resource="http://doteine.uc3m.es/tesauros/biologiaindex.php?tema=4" />
</skos:Concept>
</rdf:RDF>

```

4 Discussion

The results of the DOTEINE project have been relevant until now, but several research questions are still waiting to be answered, many of which have already been exposed in the precedent sections. Discussions held during the LODL Workshop will definitely provide us with a great feedback to enrich our work.

Acknowledgments. The DOTEINE project is being financed by the Spanish Inter-ministerial Commission of Science and Technology (ref. BSO2003-04895). The IACORIE project has been financed by the Regional Government of Madrid (ref. 06/HSE/0165/2004).

References

1. IEEE 1484.12.1-2002: Draft Standard for Learning Object Metadata (2002)
2. ISO 15836:2003: Information and documentation – The Dublin Core metadata element set (2003)
3. ISO 2788:1986: Documentation-Guidelines for the establishment and development of monolingual thesauri (1986)
4. Safayeni, F., Derbetseva, N., Cañas, A.J.: Concept Maps: A Theoretical Note on Concepts and the Need for Cyclic Concept Maps (2003).
5. Fisher, K.M.: Overview of Knowledge Mapping. In: Fisher, K.; Wandersee, J.H.; Moody, D.E.: Mapping Biology Knowledge. New York: Kluwer Academic (2002)