

Use of Ontologies to Annotate and Retrieve Educational Contents: the AquaRing Approach

Stefano Bianchi¹, Christian Mastrodonato¹, Gianni Vercelli², and Giuliano Vivanet²

¹Softeco Sismat S.p.A. {stefano.bianchi, christian.mastrodonato}@softeco.it; ²DISA – Università di Genova, {gianni.vercelli, giuliano.vivanet}@unige.it

Keywords: semantic web, learning resources, ontologies, content management, digital libraries.

Acknowledgements

This paper has been partially funded by the Aquaring Project under the EC eContentPlus Programme. The authors would like to thank Serena Alvino and Luigi Sarti from ITD-CNR and all partners involved in the AquaRing projects for their precious suggestions during the design phase of the Aquaring educational ontology.

Abstract

The paper introduces the use of Semantic Web services within AquaRing, an EC-funded project concerning aquatic environments and their resources, and an ontology used to support educational contents annotation and retrieval. The aim of the project is to improve access to the vast amount of digital content concerning the aquatic environment and its resources, as well as to support enhanced education and informal learning in this specific domain. In order to achieve these goals a semantic web based infrastructure has been designed, implemented and tested and an educational ontology has been developed. In this paper, we start from a short description of the AquaRing project and then describe the educational ontology and its use to annotate and retrieve learning contents within the AquaRing architecture.

1 Introduction

The rapid development of the World Wide Web in the last few years has brought great opportunities in the way educational resources can be made available to teachers and learners. The number of online contents available is rapidly growing but, on the other hand, some problems emerge as a result of this proliferation of materials, such as the increasingly difficult management and accessibility of resources. Keyword-based search engines are the main tools for content retrieval today, but there are some problems associated with their use, such as high recall, but low precision (a lot of irrelevant documents are retrieved); low or no recall (relevant pages are not retrieved); results highly sensitive to vocabulary (relevant documents use different terminology from the original query); results are single Web pages (if we need information that is spread over various documents, we must manually extract partial data from single web pages); etc. (Antoniou and van Harmelen, 2008).

The use of Information and Communication Technology (ICT) in learning contexts has become so pervasive that new models are needed for the process of content management, based on environments and tools enabling users to build, represent and share their knowledge. Moreover, another key point in the evolution of Learning Management Systems (LMS) is in providing flexibility and personalization of contents and services. From a technical point of view, semantic technologies can support both developers and users in achieving such goals (Adorni *et al.*, 2007). Semantic Web (Berners Lee *et al.*, 2001) is an evolving extension of the WWW that allows expressing information in a machine-interpretable form and it is expected to revolutionize scientific publishing and sharing of data on the Internet.

There are several knowledge representation models, technologies and languages, such as Resource Description Framework (RDF), Web Ontology Language (OWL) and other XML-based languages that allow description of resources in a standardized way, enhancing information reusability and interoperability. Ontologies are the solution adopted in the context of the EC-funded AquaRing project (eContentPlus Programme) whose aim is to improve access to the vast amount of digital contents concerning the aquatic environment and its resources, as well as to support enhanced education and informal learning in this specific domain. In order to achieve these goals a semantic web based infrastructure has been designed, implemented and tested and an educational ontology has been developed.

2 The AquaRing project

AquaRing, an acronym for “Accessible and Qualified Use of Available digital Resources about the aquatic world In National Gatherings”, is a European project that addresses the sector of marine and aquatic sciences

and the large cultural heritage of knowledge and information on aquatic environment and resources available from European aquaria, science centers and natural history museums. The project aims at enhancing access to and use of the vast amount of digital content (such as images, documents, movies, slide-shows, etc.) concerning the various aspects of the aquatic environment; to improve the accessibility and interoperability of cultural/scientific heritage in a standardized way across the different involved organizations and to support enhanced education and informal learning in this specific domain. In order to achieve these goals, the project investigates and defines suitable common data and metadata schemes as well as domain semantic descriptions to design an open semantic web based infrastructure and to develop a virtual global knowledge space on aquatic heritage that visitor will be able to explore according to their own interests and needs (AquaRing, 2008).

The AquaRing semantic layer is based on some main controlled vocabularies (a restricted list of terms used for indexing in contrast to natural language vocabularies), thesauri (a more structured kind of controlled vocabulary which provides information about each term and its relationships to other terms within the same thesaurus, such as synonyms, narrower, broader and related terms) and ontologies, dealing with aquatic environments and their resources, adopted for a proper semantic annotation of contents and several simple top level ontologies related to some user-oriented issues.

Currently, after a deep analysis on existing ontologies about the AquaRing subject matters, seven ontologies covering different complementary aspects of the domain considered are used for content annotation: Biological Species, Fishing Areas, Land Areas and Vessels Types, all from the Food and Agriculture Organization of United States (FAO); Aquatic Sciences and Fisheries Abstracts (ASFA), from FAO thesauri; Habitats Classification, from European Nature Information System (EUNIS); and EDUcational (developed by Aquarium of Genoa, DISA University of Genoa and Softeco Sismat SPA).

After the agreement on the adoption of these ontologies, the AquaRing consortium considered that the whole AquaRing knowledge domain was not totally covered and decided to adopt an approach based on ontology extension by means of controlled free tags and an ontology learning process (conceived for integrating and merging ontologies by creating relationships among terms of different ontologies). Through the latter process, an “AquaRing ontology” has been created, which merges the ontologies used for annotation, integrate the free tags and will be used to provide semantic services to the AquaRing portal visitors (González Rodríguez, 2008).

In order to offer personalized services, a preliminary analysis has been

performed by means of a user survey form and proper questionnaires with the aim of identifying the target users' needs. Consistently with the results, the following main categories of AquaRing users have been identified: general individual visitors, teachers (that have been suggested as a priority target by many project partners); sea museums, science centers, aquaria and zoos; children; and media. Then, the needs of each target user category have been identified in terms of document types, relevant topics, aims, services and graphic interface. With reference to the learning purposes of AquaRing, we have focused our attention on the main teachers' interests that, according to this analysis, concern (Torrighiani, Valettini, 2007):

- instructional resources for their lessons;
- pre-constructed learning paths;
- news about activities and projects developed by aquaria and similar organizations;
- to have the opportunity to discuss with colleagues and experts about the AquaRing subject matters;
- to have the opportunity to prepare visits with their students.

Regarding the type of digital resources, teachers are very interested in multimedia files, images and documents like papers, essays, project reports and conference proceedings. To satisfy not only teachers', but also students' and parents' needs, an educational area has been designed including digital resources that are intended mainly for teaching purposes, such as bibliographies, drawing books, educational games, exercises, glossaries, lecture, lesson plans, simulations, etc. These resources have been annotated and made browse able in a semantic way by means of an educational ontology specifically designed for AquaRing's goals.

3 The AquaRing educational ontology

The term "ontology" comes from the field of philosophy that is concerned with the study of being or existence (Gruber, 2008). In the context of computer and information science, an ontology may be defined as "a formal and explicit specification of a shared conceptualization" (Studer *et al.*, 1998). In the context of Semantic Web, Hendler defines an ontology as "a set of knowledge terms, including the vocabulary, the semantic interconnections, and some simple rules of inference and logic for some particular topic" (Hendler, 2001). Ontologies are typically specified in languages that allow abstraction away from data structures and implementation strategies (Gruber, 2008). OWL is the language developed by W3C for representing ontologies on the Web in an XML-based syntax.

In the context of AquaRing, standards were adopted to formalize content

annotation and metadata scheme by means of Qualified Dublin Core Metadata Element Set (DC, 1998) and to write the domain ontologies using OWL. As for AquaRing's purposes, the Dublin Core Subject element is used to semantically annotate contents by means of ontology terms describing the meaning of the information included and to host the hierarchical free tags selected by content providers when a gap in the formalized AquaRing knowledge domain is discovered during content annotation.

As no appropriate ontology was available to cover the educational field of the AquaRing knowledge domain, an ontology has been expressly designed in order to index pedagogical resources and to organize and retrieve learning materials in a more efficient and meaningful way. The AquaRing educational ontology development has been based on the following steps:

- a) Context analysis. The ontology development has begun with a preliminary study of the domain of interest. During this early stage, we have focused our attention on users' needs and on AquaRing educational services purposes. It is worth noting that educational ontology has to integrate (and not to overlap) data covered by the Dublin Core metadata scheme, such as Author, Subject, Format, Size and so on; this is the reason why some basic descriptive elements are not included therein. In the course of context analysis, a survey of the existing ontologies (and the possibilities of reusing them) has been carried out; but although many studies on ontologies and controlled vocabularies for education are available, no one has been considered suited enough to be seamlessly integrated in the semantic framework.
- b) Vocabulary definition. In order to define the ontology vocabulary, we have carefully analyzed different reference works on indexing of learning objects, namely:
 - 1484.12.1: IEEE Standard for Learning Object Metadata (LOM): a standard that specifies the syntax and semantics of Learning Object Metadata, defined as the attributes required to describe a Learning Object (IEEE, 2002);
 - Dublin Core Education Application Profile. Draft version 0.3 (DC-Ed AP): defines meta-data elements for use in describing properties of resources related to their use in teaching and learning (DCMI, 2006);
 - Dublin Core Meta-data Element Set. Version 1.1 (DC): a standard for cross-domain information resource description; maintained by an international organization, the DCMI (DC, 1998);
 - EUN Learning Resource Exchange Meta-data Application Profile. Version 3.0 (LRE AP): an application profile that uses the IEEE LOM standard for expressing meta-data about learning resource (EUN, 2007);

- Pedagogy Oriented Educational Model (POEM): an application profile with an emphasis on pedagogical features developed by the Institute of Applied Mathematics and Information Technology and the Institute for Educational Technology of the Italian National Research Council (Alvino *et al.*, 2008).

Then we have performed a comparison of this vocabulary with the results of previous context analysis and users' needs survey. At the same time, a preliminary investigation on educational resources owned by content providers has been carried out, in order to adapt the vocabulary to the specific educational AquaRing contents.

- c) Terms classification. After a brainstorming session with learning resources experts (with the aim of validating the vocabulary), the terms have been organized into a primitive taxonomy, using only hierarchical relationships. The preliminary output of this work has been a draft model based on two main categories - Audience and Educational features-, the former intended to describe the hypothetical learning resource users; the latter intended to define the key educational characteristics of learning contents.
- d) Definition of Classes and Properties. After a new brainstorming session with project partners, a first draft ontology has been designed, identifying key concepts, sub-concepts and their relationships. In this early stage, a combination of top-down and bottom-up approach has been used: firstly the more salient concepts have been identified, and then they have been generalized and specialized appropriately. The resulting ontology has been expressed in OWL DL, using Stanford University Protégé Ontology Editor (v3.3.1).

The final result (see Figure 1) has been an ontological model centered on the concept of Resource and based on five classes:

1. CONTEXT: the principal environment within which the learning resource is intended to be used (Individuals: PreSchool, PrimaryEducation, FirstGradeSecondaryEducation, SecondGradeSecondaryEducation, HigherEducation, LifelongEducation, SpecialEducation, TrainersTraining, VocationalEducation, OtherEducationalContext);
2. OBJECTIVE: the cognitive learning outcomes based on Bloom's taxonomy (Bloom, 1956) (Individuals: Knowledge, Comprehension, Application, Analysis, Synthesis, Evaluation);
3. RESOURCE: the learning resource (the Resource individuals will be the specific learning resources that will be added to the repository);
4. RESOURCE_FEATURE: the key educational characteristics of learning resource (no individuals);
5. USER: principal user(s) for which the learning resource is designed (no individuals).

RESOURCE_FEATURE has five subclasses:

- DIFFICULTY_LEVEL: how hard it is to work with the learning object for the typical intended target audience (Individuals: VeryEasy, Easy, Medium, Difficult, VeryDifficult);
- FRUITION_MODE: principal fruition mode for which the resource is designed (Individuals: PresenceFruitionMode, DistanceFruitionMode, BlendedFruitionMode);
- FRUITION_TIME: approximate time it takes to work with the resource for the typical intended target audience (Individuals: ShortFruitionTime, MediumFruitionTime, LongFruitionTime);
- INTERACTIVITY_MODE: predominant mode of learning supported by the learning resource (Individuals: ExpositiveInteractivityMode, ActiveInteractivityMode, MixedInteractivityMode);
- TYPE: specific kind of educational resource (40 learning resource types have been identified, e.g. CaseStudy, E-Book, EducationalGame, LecturePresentation, LessonPlan, Tutorial, etc.).

USER has two subclasses:

- LEARNER: one who works with a learning resource in order to learn something (Individuals: GenericLearner);
- MEDIATOR: one that mediates access to the resource and for whom the resource is intended (Individuals: Counsellor, GenericMediator, InstructionalDesigner, Parent, Teacher).

With reference to the element User in this ontology, it is important to note that it does not correspond with the five target users previously mentioned, because it refers only to users of educational area that we have identified in learners and mediators.

Eight relationships have been identified among these classes:

- hasContext with Resource as domain and Context as range;
- hasObjective with Resource as domain and Objective as range;
- hasDifficultyLevel with Resource as domain and DifficultyLevel as range;
- hasFruitionTime with Resource as domain and FruitionTime as range;
- hasInteractivityMode with Resource as domain and InteractivityMode as range;
- hasType with Resource as domain and Type as range;
- hasFruitionMode with Resource as domain and FruitionMode as range;
- has User with Resource as domain and User as range.

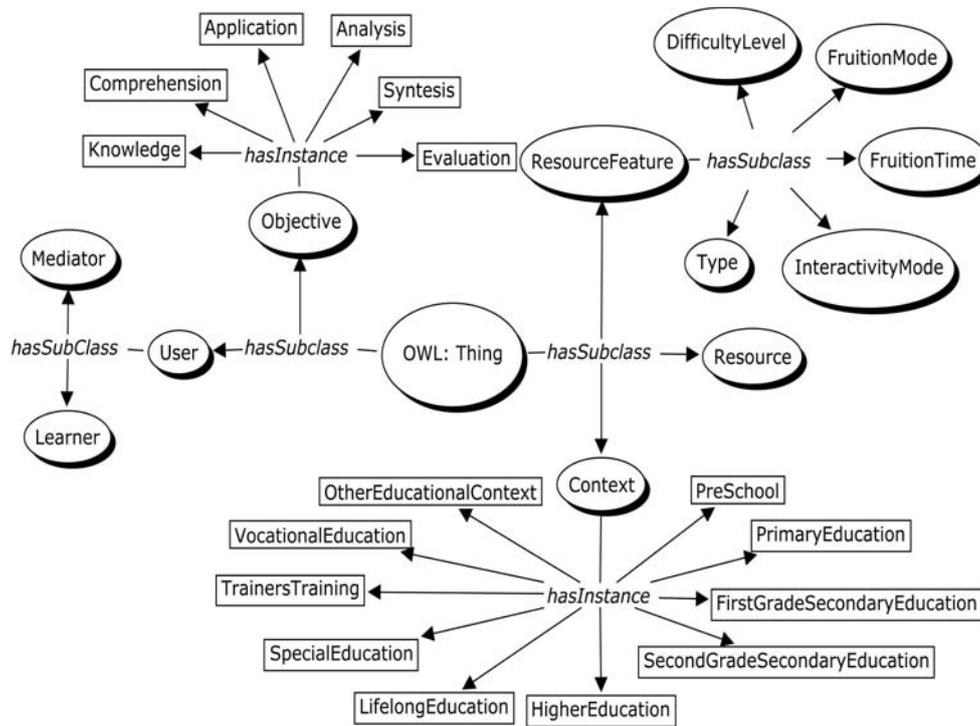


Figure 1 The structure of the AquaRing Educational Ontology.

As the other ontologies selected for annotation focus specifically on the aquatic domain, the educational ontology has been designed and used to enrich the description of the contents included in the AquaRing knowledge base. It is interesting to highlight that, whereas other ontologies refer directly to what a content is about, the educational ontology refers to educational purposes of contents. From this point of view, this ontology can support content-type oriented search: for example, a teacher, searching for information on “hyppocampus”, may restrict the search algorithms by specifying that only Resources for PrimaryEducation students with an Objective focused on plain Knowledge must be extracted from the knowledge base. It is also worth noticing how the design of this educational ontology is almost seamlessly applicable to other scenarios, as it does not contain any specific restriction on the type of resources that can be described using the concepts specified herein.

An interesting issue for further research activities currently under investigation is the use of this ontology to provide educational oriented services to the users (an interesting example could consider the semi-automatic creation of learning paths inside the annotated resources following educational and scientific annotations).

4 Conclusion

In this paper, we have presented AquaRing, an EC-funded project concerning aquatic environments and their resources, and an ontology used to support educational contents annotation and retrieval. The AquaRing project will be completed by March 2009; by that time, a first quantitative and qualitative assessment work of the approach here proposed will be performed through ad hoc test cases. The results of this analysis will be evaluated with the aim of improving the semantic services provided by AquaRing web portal and releasing an updated version of the afore described ontological model. In fact, every ontology design process is necessarily an iterative process (Noy, McGuinness, 2001). Thus, also our model has to be considered a work in progress.

At the same time, we are exploring two different research hypotheses with the aim of further developing our approach to educational content annotation and representation. In a first scenario, we are interested in methods and techniques that allow reducing the effort necessary for the knowledge acquisition process (this is a very important issue, because every ontology development process requires much time and many resources). In a second scenario, we are interested in methodologies that allow formalizing the knowledge acquisition process and, as a consequence, the ontology-driven conceptual analysis. In regard to this issue, one of the most interesting approach we are studying is OntoClean (Guarino, Welty, 2004), a formal methodology that applies domain-independent notions, used for ontological analysis in philosophy to analyze conceptual modelling process.

BIBLIOGRAPHY

- Adorni G., Coccoli M., Vercelli G., Vivianet G. (2007), *Topic Maps e XTM per l'e-learning*, Journal of e-Learning and Knowledge Society (3).
- Alvino S., Forcheri P., Ierardi M., Sarti L. (2008), *Describing learning features of reusable resources: a proposal*, Int. Journal of Social Science , 2 (3), 156-162.
- Antoniou G., van Harmelen F. (2008), *A semantic web primer - 2nd ed.* Cambridge, Massachusetts, MIT Press.
- AquaRing Website, URL: <http://www.aquaringweb.eu/> (accessed on 1st March).
- Berners Lee T., Hendler J., Lassila O. (2001), *The Semantic Web*, Scientific American 284, 34-43.
- Bloom B. (1956), *Taxonomy of educational objectives – Handbook I*, New York, David McKay Company Inc.
- D. C. (1998), *Dublin Core Metadata Element Set - Version 1.1*, Retrieved September 1, 2008 from <http://dublincore.org/documents/dces/> (accessed on 1st February).

- DCMI (2006), *DCMI Education Community*, Retrieved September 1, 2008 from <http://projects.ischool.washington.edu/sasutton/dcmi/DC-EdAP-7-18-06.html> (accessed on 1st February).
- EUN. (2007), *The EUN Learning Resource Exchange Metadata Application Profile*, retrieved September 1, 2008 from <http://fire.eun.org/LRE-AP-3.0.pdf> (accessed on 1st February).
- González Rodríguez M. (2008), *Formalised AquaRing Domain Ontologies*, Aquaring Project: Deliverable number 3.2.
- Gruber T. (2008), *Ontology*, on L. L. (Eds.), *Encyclopedia of Database Systems*.
- Guarino N., Welty C. (2004), *An Overview of OntoClean*, on S. Staab, & R. Studer, *The Handbook on Ontologies*, Berlin, Springer-Verlag, pp. 151-172.
- Hendler J. (2001), *Agents and the Semantic Web*, *IEEE Intelligent Systems*, 2 (16), 30-37.
- IEEE (2002), *IEEE Standard for Learning Object Metadata*, retrieved September 1, 2008 from IEEE Learning Technology Standards Committee: <http://ltsc.ieee.org/wg12/>(accessed on 1st February).
- Noy N., McGuinness D. (2001), *Ontology Development 101: A Guide to Creating Your First Ontology*, retrieved September 1, 2008 from <http://www-ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html> (accessed on 1st February).
- Studer R., Benjamins, V., Fensel D. (1998), *Knowledge Engineering: Principles and Methods*, *IEEE Transactions on Data and Knowledge Engineering*, 25 (1,2), p. 161-197.
- Torrigiani C., Valetini B. (2007), *AquaRing User Segments, Profiles and Needs*, AquaRing Project: Deliverable number D2.1.