BROAD Project: Semantic Search and Application of Learning Objects

Thiago Nery Teixeira[1], **Fernanda Campos**[1]

[1]Departamento de Ciência da Computação (DCC)
Instituto de Ciências Exatas (ICE) – Universidade Federal de Juiz de Fora (UFJF)
Rua José Lourenço Kelmer, s/n, São Pedro – 36.036-900 – Juiz de Fora – MG – Brasil

{thiago.nery,fernanda.campos}@ufjf.edu.br

Research Area: Software Engineering.

Admission to the program: march/2011

Scheduled for completion: february/2013

Resumo. *As Web-based learning grows; we assist a proliferation of data, learning objects (LO) and tools. This imposes new challenges as, for example, how to organize these resources, how to share and reuse successful learning objects, and how to provide interoperability among data and tools. A key point is how to find artifacts that fit user´s need. This paper presents the current stage of BROAD Project, detailing the network ontology. The architecture evolution is also presented and argued considering technology adoption. Some LO applications are highlighted to illustrate the using scenarios. Some related works are discussed.*

Keywords: Learning Object, Metadata Standards, Ontology, Semantic Search, LO Composition.

1. Introduction

An important challenge is educational activities is the organization of content. The ability to automatic or semi-automatic recommendation of different resources, customized to each student, based on individual need, interest and skills is a trail that involves researchers from Computer Science and Computer in Education domains.

There are many resources available on the Internet waiting to be reused since the growth of distance learning activities has been accompanied by a proliferation of data, learning objects (LO) and tools. This new setting imposes new challenges, such as how to locate these resources, how to organize them, how to share and reuse successful learning objects and how to provide interoperability among data and tools from different platforms, as they are used by people with different backgrounds and needs.

Computers in Education community is looking for an infrastructure that enables the design, reuse, annotation, validation and sharing of computing artifacts, among them, the learning objects. Learning objects are "small, reusable chunks of instructional media" [Wiley 2000]. They are "highly interoperable and reusable modular building blocks for e-learning content based on widely shared specifications or already accredited standards" [Clyde 2004].

LO localization and retrieval require appropriate register using strong metadata model able to describe it fully and correctly, in order to facilitate the use. In LO domain, metadata describe relevant characteristics used for their cataloging in repositories, which can be retrieved later through search engines or used by learning management systems for composing learning units. The reusability of LO shows their potential to accelerate distance courses and classes composition, thus help to increase the supply of Education.

One aspect that may slow the spread use of learning objects is the difficulty of locating and retrieving them in accordance to end users' needs. The task of manually reviewing LO is colossal because of the growing quantity of educational resources. In addition to this, learning object repositories, developed to retrieve LOs in a useful way, returns a large number of hits. Users are faced with the problem of deciding which resources they want. Therefore, we expect to retrieve LOs using an automatic method, or at least a semi-automatic one. According to [J. Sanz-Rodríguez and Sánchez-Alonso 2010] and [Ochoa and Duval 2008], it is common that most of search engines rank LO by giving an example of LO, by setting search fields or by giving a user profile. Open issues are how to retrieve automatic or semi-automatic learning objects in an efficient, reliable and easy way? How to use these LO in a composition process?

In the last two years, we are working in the BROAD Project and so far we have achieved the following steps: metadata set for LO registration; LO repository (a web interface) to register and retrieve LO; ontology for the LO domain using our QDAontotoly approach; domain ontologies in Biology, Mathematics, Software Engineering; set of rules in SWRL to allow inferences on the concepts and relationships described in the ontology; framework to promote LO semantic search and retrieval; tool to create personalized composed content using workflow technology.

This paper aims to present the current stage of BROAD Project, detailing the ontology network. The architecture evolution is presented and argued considering technology adoption: BROAD-WS (Webservice), BROAD-SS (Semantic Search), BROAD-WP

(Workflow) and BROAD-DP, our ongoing research. We also briefly discuss some related works and finally, we offer the final remarks and the forward steps.

2. Theoretical Background

Traditionally, the Web has been seen as a distributed source of information. Considering the growth of Web both in size and in diversity, there is an increasing need to automate those aspects related to discovery, execution, automatic selection and composition of artifacts. The use of Semantic Web technologies like semantic Web services and ontologies can assist in the construction of tools that allow this automation.

Among techniques and fundamentals needed to achieve goals of the Semantic Web, ontology has a key role. Ontology is a formal and explicit specification of a shared conceptualization [Studer 1998]. A language for defining ontology is OWL (Web Ontology Language) [OWL 2004], which is ontology representation language that is being currently recommended by the W3C as a standard for semantic Web.

Workflow specifies a process or a software artifact to be executed (e.g. a web service). It provides the connection between steps of the experiment modeled according to the flow of data and their dependencies. The computational representation of a workflow contains several important information, that can be used in result analysis and verification process, including tasks that were executed and storage resources used in distributed environments [Gil 2007].

E-Learning is a comprehensive domain that encompasses forms of technology-enhanced learning such as online or Web-based learning. The fast grow of new information technologies has increased the interest in educational resources, or artifacts, on the Web. These artifacts are often described as Learning Object (LO) which focuses on reusability and automation. LO is usually identified by a set of metadata descriptors, such as LOM (Learning Object Metadata) and SCORM (Sharable Content Object Reference Model).

3. Related Work

This section contains a brief description of some works that are related to this paper. We highlight them as they deal with LO repositories, semantic queries, ontologies or propose learning object composition.

Some international repositories are well known and used, including: [CAREO 2010], [EDNA 2010] and [MERLOT 2010]. In Brazil, there are two important repositories: BIOE (Bank for International Educational Objects, in Portuguese, Banco Internacional de Objetos Educacionais) and FEB (Federation of Repositories Education Brazil, in Portuguese Federação de Repositórios Educa Brasil). BIOE is a repository created in 2008 by the Education National Secretary, in partnership with the Science and Technology Board, the Latin American Network of Educational Portals and the Organization of Ibero-American Countries. It aims to maintain and share digital educational resources freely accessible and in different formats, integrated into the Education National Secretary' Teacher Portal. [BIOE 2010]. FEB (Federation of Repositories Educa Brazil) is a Working Group of the Brazilian National Research and Education Network. It aims to organize various repositories in a hierarchical system, called federation which centralizes the information contained in the repositories into a single portal. FED uses the Brazilian metadata model OBAA [FEB 2010].

Available researches on the reuse of learning objects adopt IEEE specification. ALOA [Chatti 2009] is a Web Service driven framework for IEEE LOM compliant automatic learning object metadata generation. ALOA can be used in two different ways: ALOA provides a public API as Web Service that can be used by third party applications. They do not adopt ontology.

Reference [Moura 2005] proposes the integration of LO repositories of the Partnership in Global Learning (PGL members) in order to share LO. The solution uses web-services and considers distribution, heterogeneity of data sources and LO queries optimization. They propose mediators and wrappers architecture, implemented as web-services, providing a common understanding among PGL members.

[Menezes 2011] proposed a Learning Object composition process on the basis of structural theory of discourse whose form of content is the domain ontology expressed by the learning objects composition. The composition processed is based on the definition of a learning path from the domain ontology. Each concept in the learning path is mapped to a sequence of learning objects in a repository. The composition outcome is presented to the user in a way that is device independent relying on the emerging web technologies. Our proposal deals with semantic composition, based on educational metadata.

According to [J. Arteaga and Rodriguez 2010], the current state-of-the-art in semantic search in LO repositories show that advances in research and development in the area should include: Web services; agent approach; semantic search, federation membership; and the use of ontology. The main difference of these projects to ours is that BROAD Project deals with semantic annotations in order to improve the LO search and adopts the concept of Web service approach. Also, we work with a distributed approach, considering that we can search in distributed repositories. BROADWS only needs to store a semantic description about the repository and the domains of its stored LO.

4. BROAD Ontology Network

NEnC group has a long experience in developing domain ontologies. For BROAD project we built a LO ontology [F. Campos and Santos 2011], some domain ontologies, as CeLO Human Disease [D. Palazzi and Coimbra 2010], we are also working on an upper ontology about knowledge areas. Those ontologies integrate our Ontology Network. LO ontology main goal is to model the general context of e-Learning with emphasis on artifacts and learning objects, allowing users, such as research groups, teachers, students, etc., to share artifacts for planning, giving classes, studying or any other e-learning activity.

According to [Pernas et al. 2011] ontologies cannot be treated as standalone artifacts, because they are related to each other in ways that might affect their meaning and are inherently distributed in a network of interlinked semantic resources. In this context an ontology network can be seen as a collection of ontologies related together via a variety of relationships, and meta-relationships.

Our ontology network BROADNet of learning objects, is used in the semantic suport to applications. Ontology is generally not static and from time to time, it needs expansion to include new concepts, properties, and relationships. The ontology network must ensure interoperability with other domain ontologies, be used in applications that require computational semantic navigation for registration and search of learning objects and be useful in defining metadata for learning objects and assessment.

In terms of the general classification [Guarino, 1997 in D. Palazzi and Coimbra 2010] the ontologies that compose BROADNet can be classified as: learning objects and metadata are general ontologies, the Celo Human Disease ontology is a domain ontology and the kwnoledge ontology is a top-level ontology. Considering the classification of ontologies for e-learning environments [Antoniou and van Harmelen 2008], they can be classified as follows: learning objects and metadata are pedagogical ontologies, the Celo Human Disease is a content ontology and the kwnoledge ontology is a structural one, because it defines the hierarchical navigation (Figure 1).

In the proposed network the LO ontology seeks to model the general context of e-Learning with emphasis on artifacts and learning objects, allowing users, such as research groups, teachers, students, etc., to share artifacts for planning, giving classes, studying or any other e-learning activity. LO ontology is composed of more than 150 terms. Knowledge Areas Ontology is still under construction.

BROADNet is characterized by some specific relationships, among them we highlight [Suarez-Figueroa et al. 2011]: dependency (an ontology refers to definitions included in another ontology), alignment (correspondence by declaring which of their entities should be considered as being the same, or as being more general than the others) and modularization (an ontology is independent but is a sub-domain of another ontology). The formalization of these relationships includes [Allocca et al. 2009] [Diaz et al. 2011]: includedIn, similarTo, isAlignedTo, disagreesWith, agreesWith, isTheSchemaFor and usesSymbolsOf.

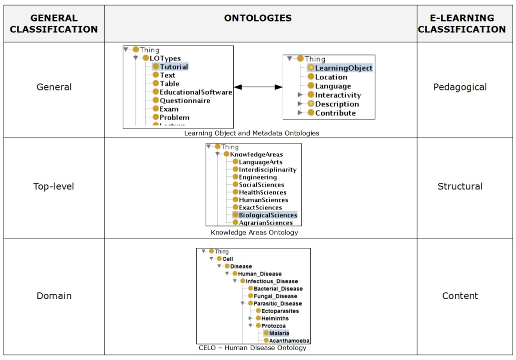

Figure 1. BROADNet: ontology network

5. BROAD Architecture

The main idea of the BROAD Project is that each new version of the architecture goes one step further than the previous one, considering technology adoption. A development process was adopted for BROAD Project and its main steps are: conception, construction and evaluation. For each version we will describe the actual stage, considering these steps.

5.1. BROAD-WS

The primary purpose was to build architecture of a semantic repository of learning objects. In order to achieve this goal, some secondary goals need to be addressed. Among them we can mention: i) create a semantic data repository, responsible for storing the ontology itself and semantic annotations, based on the ontology, of all e-learning artifacts and learning objects ii) promote the reuse of knowledge through storage, search and retrieval of e-learning artifacts and learning objects and iii) represent the domain through a logical language, with inference capability and use of rules. The architecture also uses concepts of semantic web services.

The architecture considers three different tiers: (i) BROAD-WS (Middle Tier): implemented as a RESTFull web service, it can be installed in different sites, allowing the distribution of repositories, and it is independent of the interface that will access services, allowing its integration with existing tools and frameworks; (ii) Backend Tier: it is the service tier, used by the BROAD-WS to access the relational database (BROADDB) where the ontology (classes and individuals – the LOs) is stored; (iii) Client Tier: implements the user interface and it can be developed in any language or framework as a REST client.

The architecture offers four services to its users: (i)Registry (Form): Get the LO data from a HTTP POST, possibly through a HTML Form. The data is stored at a relational database at BROADDB; (ii) Query (SQL): The user makes a SQL query and receives, as a result, LOs which match the query; (iii) Registry (ontology): Based on the URI given by the user, the LO, expressed as an OWL model, is recovered and, using inference mechanisms, the model is stored in a database; (iv) Query (SPARQL): The user makes a SPARQL query and receives, as a result, individuals (LOs) from ontology stored at database which match the query.

The BROAD-WS services are distributed in three layers: (i) Client Manager: Responsible for users interaction, implementing the Facade design pattern. Its purpose is to provide a single interface with clients of system, so customers do not have access to the internal structure of the BROADWS; (ii) Storage Manager: Responsible for ontology storage/recovery in the database, and queries carried out by the user, encapsulating the access to database; (iii) Ontology Manager: Responsible for inference on the models, as well as for providing an API to access BROAD ontology.

The use of the complete BROAD-WS infrastructure depends on the development of end users applications, using graphical interfaces and encapsulating its services.

5.2. BROAD-SS

The main goal of BROAD-SS is the search, recovery and invocation (if it is a web service or application) of LO generated by people evolved with e-Learning activities. This main goal can be decomposed into a set of secondary ones: i) existence of a knowledge base that represents the knowledge needed to support the management of LO through the use of ontology, ii) to support different roles that the "end user" of the system can take: either as a consumer or as a provider of LO, while allowing this user to have the ability to configure his own profile, with his main preferences. As a provider of LO, the user has to store a semantic annotation, based on ontology, of the LO repository and of the available LO objects of this repository.

5.3. BROAD-WF

As students learn in different ways, only one teaching process or delivering the same content does not aware all student's needs, we are trying to achieve means to personalized educational contents. In this study we discuss the potential of learning style modeling for improving the e-Learning experience.

According to [Menezes 2011] "learning objects composition can be defined as an aggregation of learning objects to be used in a learning environment in the learning and teaching process. Since this process aims to fulfill a set of learning objectives it is expected that as the learner interacts with a learning environment, he/she will be involved in a learning experience".

Our next goal is to implement a semantic learning object composition, based on domain ontology and educational metadata. The main idea is to generate personalized workflow, where the LOs found in distributed repositories constitute a content sequence. The recommendation system model is based on media data, domain context and student expertise. The content workflow will enable the automation of the process as the student's interests and preferences will be detect during the query, considering the metadata, and in accordance to teacher´s plan.

Most studies on the automatic composition only consider the input and output of available LO, regardless of their semantics. In addition to processing complexity, we try to meet user's need during search process. The proposed solution allows students to choose from a workflow sequence of domain context the LOs that best fit his/her profile.

The proposed composition model includes: (i) Step 1 (user) - selection of knowledge area and domain terms from the network ontology; (ii) Step 2 (user) - selection of metadata descriptor; (iii) Step 3 (BROAD-SS) - semantic search; (iv) Step 4 (BROAD –WF) – content workflow generation and composition process; (v) Step 5 (user) – LO selection from the composition process, generation of a personalized workflow.

5.4. BROAD-DP

Currently the BROAD architecture is being revised to include concepts of provenance data. The provenance of data, also known as lineage or genealogy, describes the origins of a particular data and also the process that produced it. Thus it has an important role in establishing quality, validity and relevance about the produced information. The main intention is that the provenance tracking assists the search and recovery of learning objects. It will also be used to create a user profile, based on the types of learning objects that he chosed after searching. The purpose is to provide the user, in a future search, results directly related to his profile. Figure 2 gives an overview of the proposed architecture BROAD-DP (BROAD Data Provenance).

6. Final Remarks and Future Works

Web has available several resources to be reused. It is not different in the educational context, where data, learning objects (LO) and tools are ready to be recycled. Finding LO is still a bottleneck for wide dissemination of its use. So, it is essential to develop infrastructures that enable the design, reuse, annotation, validation and sharing of computing artifacts, among them, learning objects. Our contribution is to propose BROAD

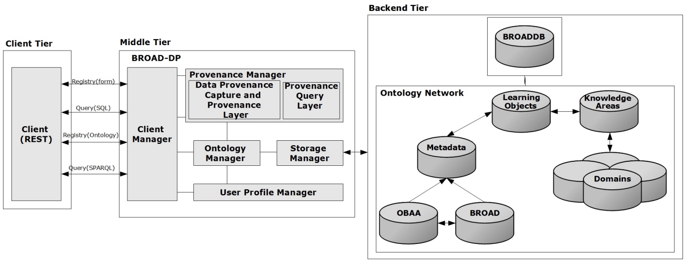

Figure 2. BROAD Architecture overview

architecture, based on ontology, semantic web services and workflow to search, retrieve and use of learning objects in content composition.

Our work is still a work in progress, but we have completed important steps, such as: LO metadata set; a repository that allows register and semantic retrieval of LO according to ontological models; the automatic composition of educational contents in the Web and the definition of a semiautomatic composition model based on workflow.

The contributions of this study are relevant because they show that it is possible to apply semantic to LO query in addition to show viability of creating semantic-based LO composition. Our next steps include the tracking of educational provenance data. The use of provenance aims to improve the trustworthiness of a user in the use of a learning object. It can also help to trace a user profile based on search and choices made.

Although ontologies are used in semantic solution it is still restricted because of the limitations found in persistence tools that are actually available, such as loss of performance in the conversion process of ontologies to relational databases, limitations related to the horizontal scalability, limiting the use of the same ontology between nodes in a geographically distributed network or the limitations related to querying and reasoning capabilities of the database.

This paper is based on [Campos et al. 2012] and [Teixeira et al. 2012].

References

Allocca, C., Daquin, M., and Motta, E. (2009). Towards a formalization of ontology relations. *Internationa Conference on Knowledge Engineering and Ontology Development*.

Antoniou, G. and van Harmelen, F. (2008). *A Semantic Web Primer*. MIT Press, Cambridge, MA, USA.

BIOE (2010). Banco internacional de objetos educacionais.

Campos, F., David, J. M. N., Braga, R., Teixeira, T. N., and Santos, N. (2012). Rede de ontologias: apoio semântico a linha de produtos de objetos de aprendizagem. *Anais do SBIE 2012*.

CAREO (2010). Campus alberta repository of educational objects.

Chatti, M. A. (2009). Aloa: A web service driven framework for automatic learning objects annotations. *Times of Convergence: technologies across learning contexts*.

Clyde, L. A. (2004). Digital learning objects. *InfoTech*, 31.

D. Palazzi, F. Campos, R. B. and Coimbra, E. (2010). Desenvolvimento e integração de ontologias: um estudo de caso em doenças humanas. *X Workshop de Informática Médica*.

Diaz, A., Mota, R., and Rohrer, E. (2011). Making ontology relationships explicit in a ontology network. *V Alberto Mendelzon International Workshop on Foundations of Data Management*.

EDNA (2010). Education network australia.

F. Campos, R. B. and Santos, N. (2011). Expanding access to distance learning using learning objects. *Education in a technological world: communicating current and emerging research and technological efforts*.

FEB (2010). Federação de repositórios educa brasil.

Gil, Y. (2007). Examining the challenges of scientific workflows. *Computer*, 40.

J. Arteaga, E. Moreno, C. Z. R. S. and Rodriguez, F. (2010). Use of agents to realize a federal searching of learning objects. *Trends in practical applications of agents and multiagent systems*.

J. Sanz-Rodríguez, A. D. and Sánchez-Alonso, S. (2010). Ranking learning objects through integration of different quality indicators. *IEEE Transactions on Learning Technologies*.

Menezes, H. J. (2011). A semantic learning object composition: an experiment based on theory of discourse. *Iadis International Conference WWW/Internet 2011*.

MERLOT (2010). Mutimedia educational resource for learning and online teaching.

Moura, S. L. (2005). Integrating repositories of learning objects using web-services and ontologies. *International Journal of Web Services Practices*.

Ochoa, X. and Duval, E. (2008). Relevance ranking metrics for learning objects. *IEEE Transactions on Learning Technologies*.

OWL (2004). Ontology web language.

Pernas, A. M., Diaz, A., and Motz, R. (2011). Situations and ontology networks to define adaptive actions in e-learning systems. *IADIS International Conference*.

Studer, R. (1998). Knowledge engineering: Principles and methods. *Data and Knowledge Engineering*.

Suarez-Figueroa, M. C., García-Castro, R., Terrazas, B. V., and Gómez-Pérez, A. (2011). Essentials in ontology engineering: Methodologies, languages, and tools. *Ontological Engineering State of the Art*.

Teixeira, T. N., Campos, F., Braga, R., Santos, N., and Mattos, E. (2012). Broad project: Semantic search and application of learning objects. *IEEE Technology and Engineering Education (ITEE)*, 7.

Wiley, D. A. (2000). Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. *The Instructional Use of Learning Objects: Online Version*.