An Ontology-Based Digital Library on the e-Learning Domain

¹Computer Science Department -Universidade do Estado do Rio de Janeiro - Brasil Rua São Francisco Xavier, 524, 6º andar, Bloco B, Prédio João Lira Filho Rio de Janeiro – RJ – Brasil – CEP 20550-013

 2 Computer Science Department – Research Group in Software Quality (NPQS) Universidade Federal de Juiz de Fora - Juiz de Fora, MG - 36036330 neide@ime.uerj.br, fcampos@dcc.ufjf.br, regina@acessa.com.br

Abstract: The technological changes stress professional requirements and demand the mastery of new skills. The changeable context also stresses teachers' job. In this context, the delivering of Web-based continuing education can be useful, since it is an important information space. Moreover, tools for data retrieval at the Internet work at the syntactic level, disregarding the semantic aspects. Solutions point out to Web Semantic technologies. Our contribution is to develop a digital library for the e-learning specific domain using ontology-based semantic querying.

1. Introduction

Nowadays, social, economical, cultural and technological changes deeply stress the professional profiles. As a consequence, everyone needs to continually improve his/her professional skills by means of different kinds of continuing learning or lifelong learning. The changeable context also stresses teachers' job. It is consensual that teachers' new skills are related to how to plan, build and manager remote courses and how to monitor and assess online students by using e-learning facilities. Networking technologies can be useful helping teachers to improve their skills anywhere: home, office or school.

Lifelong learning makes use of a set of educational activities and resources, such as virtual or face-to-face conferences, seminars, short courses, meetings, workshops, and so on. Currently, portals at the Web are becoming widely adopted to promote lifelong learning. However, the Web was designed as an information space, with the goal that it should be useful not only for human-human communication, but also that machines would be able to participate and help (in www1). Therefore, computers on the Web have been basically playing the role of redirecting and delivering information. They cannot understand the Web page contents organized in natural language for human use and understanding. It does not allow information retrieval in a machine processable form and people have the cognitive overhead of extracting and interpreting information.

The Web had grown up as a business space and became an important repository of all kinds of information. As a result, searching information is a hard and slow process. Current search engines were developed to work only as productivity tools helping users to retrieve data in the Web databases by means of information retrieval at syntactic level, often disregarding the semantic aspects. Words and sentences have different meanings depending on the context of use. The Web, however, is organized as an information network without any hierarchy or navigational context. Users often feel lost and cognitively overloaded, and not always get the information they are searching for. Solutions point out to Semantic Web (Berners-Lee, Handler & Lassila, 2001), which key point is the notion of ontology*.* Ontology specifies a shared understanding of a domain of interest and contains a set of concepts, together with its definitions and interrelationships, and possibly encodes a logical layer for inference and reasoning.

Aiming to help lifelong learning initiatives and improve information search at the Web, we are developing a digital library for the e-learning specific domain. Our digital library extends Web portal's functionalities, providing technical information and communication and collaboration spaces and hosts a wide variety of information (technical papers, Web sites of systems and tools, Web sites of Brazilian and international experience on e-learning, and some e-learning software artifacts to be used to create tailored e-learning applications). It provides a set of services for capturing, cataloging, storing, searching, protecting, and retrieving information, based on ontology-based semantic queries. The digital library main service is a search engine that retrieves information by tracing the domain vocabulary met at an ontology (Santos, Campos and Braga, 2005). Next sections briefly argue domain ontology and e-learning ontology, describe the digital library and offer the conclusions and the future works.

2. Domain Ontology and e-Learning Ontology

2.1. Ontology: A Brief Overview

The role of an ontology is to reflect a community's consensus on a useful way to conceptualize a particular domain. Ontology is a useful support for search engine design since it provides means for machine-based knowledge sharing and reuse between applications. It is generally considered to provide definitions for the vocabulary used to represent knowledge. It can be seen as a schema that provides precise and complete models of particular domains. In the process of building ontology, each term must be defined by means of a formal and informal description, as well as the specification of the relationships among the terms, shaping a semantic net.

Ontologies are beginning to be used in the context of digital libraries for many different purposes. They can assist the extraction of concepts from unstructured textual documents (Embley, Campbell, Smith and Liddle, 1998), by serving as a source of knowledge about the particular topic. They can also assist in managing document descriptions in large digital libraries (Weinstein and Alloway, 1997).

Building ontology means different things to different practitioners. The distinction of how one carry out describing something reflects a progression in ontology from: simple lexicons or controlled vocabularies to categorically organized thesauri; to hierarchies called taxonomies where terms are given distinguishing properties; to fullblown ontologies where these properties can define new concepts and where concepts have named relationships with other concepts.

A typical Web ontology is composed of taxonomy and a set of inference rules (Berners-Lee et al, 2001). Taxonomy describes classes of objects and the relationships among them. If it is properly described, it preserves the specific meaning of terms and current expressions in a certain knowledge domain. But taxonomy is not enough to put semantic at the Web. It happens because taxonomy information is expressed under stanched and self-constrained categories, subcategories, classes and subclasses. It restrains relationship of information not explicitly related. Our work assumes Berners-Lee vision of Web ontology, i.e., we have build a taxonomy and some inference rules about the domain. Our search engine works by sweeping the ontological terms and their relationships.

On the e-learning domain, we also adopted the notion of domain ontology (Guarino, 1998) to organize the common vocabulary. We adopted the notion of feature model (Cohen, 1994) that specifies the domain vocabulary and shapes a semantic net of terms. The model captures the general features in specific domains and allows the insertion of new terms in as domains grow up. We used Domain Engineering (DE) techniques to specify the e-learning taxonomy, and we used the domain engineering method called Odyssey-DE (Braga et alii, 1999).

Building ontology implies in acquiring the domain knowledge and collecting appropriate information resources that will define, with consensus and consistency, the terms used formally to describe the domain of study. Several works proposed the organization of the e-learning vocabulary. The Learning Technology Standards Committee of the IEEE Computer Society (2001) described different types of learning technology systems (LTS), with a wide range of tools, environments and resource management systems for educational purposes. For the Committee, terms for LTSs belong to two groups: LTSs defined through their educational functions (intelligent tutoring system, learning environment) and LTSs defined through their technical features (distributed learning technology system, web-based learning technology system). Our ontology embodies the main technical features of the e-learning domain.

Paulsen (2002) categorizes the e-learning domain and describes a basic vocabulary. His categorization includes: (a) Online Education, E- learning, Online Education Systems, Integrated Online Education Systems, and Standards Specifications; (b) Content Creation Tools, Authoring Tools, Assessment Tools, Learning Content Management Systems, and Learning Objects; (c) Learning Management System, Learning Platform, Digital Learning Environment, and Learning Service Provider; (d) Student Management System, Enterprise Resource Planning System, Human Resource Information System, Knowledge Management System, and Competency Management System; and (e) Accounting System. A great part of our ontology terms was defined from Paulen's categorization.

Other contribution is the e-Learning Cybrary (www1). It is an ontology-based collection of annotated links to e-learning sites, news, documents, portals, and other resources available on the Web. In the Cybrary point of view, the area covers Shared learning, Communities of practice, Learning networks, Peer-to-peer e-learning, Learning portals, Collaborative workspace, Digital university/online campus and, Digital classroom. According to e-Learning Cybrary, the main technologies employed

to support e-learning tasks are: Synchronous communication technologies, Asynchronous communication technologies, Authoring tools, Simulation & Computer-Based Training, Visualization Tools, Electronic Performance, Intelligent agents, Integrated platforms, Learning Management Systems. Despite the reported works, the e-learning domain is diffuse without consolidated and worldwide vocabulary.

Our proposal sketches very similar vocabulary to those proposed in other works. Therefore, our approach is distinctively oriented to the Domain Engineering process, which main purpose is to develop reusable software artifacts and information. As described before, we adopted the notion of domain ontology (Guarino, 1998), which requires the employment of features model that specifies the domain vocabulary and shapes a semantic net of terms. It allows the detailing of the general characteristics – the features – of domain information. Similarities and differences are expounded by means of a hierarchy of features. In previous works, we extended the definition of features by introducing deeper detailing into the ontological terms of the domain (Braga, Werner & Mattoso, 1999). Hence, each ontological term is specified as a domain feature and a detailed description of it (i.e., *Domain Patterns*), and relationships with other ontological terms (features), at different levels of abstraction, are created. Our work also provides some ontological rules in order to discover new ontological relationships. This set of rules allows the creation of synonym, hypernym and hyponym relationships that domain engineer did not discovered by his own. As a result, the features are expressed with more precise elements in Sub-domains (divided into Sectors and Sub sectors) and in Functionalities (divided into Sub Functionalities and Specific Functionalities). Subdomains embrace the high-level domain characteristics and Functionalities contain the most general domain functionalities (see Tables 2 and 3).

2.2. The e-Learning Ontology

Building an ontology implies in defining the domain boundaries and the basic vocabulary. We classified the e-learning domain in two comprehensive sub domains: Web-based instruction (WBI) and Computer-supported cooperative learning (CSCL). We highlight that the sub domains are complementary and in many cases interdependent approaches. However, they have an important difference: WBI incorporates the main features of web-based learning technology area, but it does not emphasize the cooperative learning activities. CSCL focuses basically the cooperative side of the educational process, and does not highlight the instruction delivery.

WBI concerns basically the use of the Web platform to delivery educational material to provide students support and to offer access to a large range of communication tools. The instruction can be as simple as using the system as a "page turner" for information, or as complex as an integrated system which logs learner inputs and responses, providing interactions with video, animation, imagery, forms, examinations, or software. CSCL supports and promotes shared experience, peer exchange, interaction, shared purpose, common practices of interaction and communication. Its purpose is to scaffold or support student's effectively in learning together. It is based on the promise that computer supported systems can support and facilitate group process and group dynamics in ways that are not achievable by face-to-face, but they are not designed to replace face-to-face communication. CSCL are systems typically tailored for using by multiple learners working across networked machines.

The digital library users can search information using any ontological term, because the features model shapes a network of nodes (vocabulary) and links (relationships among basic vocabulary) that is useful to explicit available and new ontological relationships. The ontological relationships will only appear by means of a set of inference rules. In the particular case of our ontology, some inference rules were defined and expressed by the following set of propositions.

In order to discover new ontological relationships, we defined a complementary set of rules to support the discovery process of new relationships into the same ontology or into related ontology. Oliveira et al. (2004) defined the following set of basic rules:

- **Contained** (**Name** (A), **Description** (B)) ♠ **Synonym** (A, B): If term A is contained in the definition of term B then term A can be a synonym of term B;
- **Synonym (A, B)** ♠ **Synonym (B, A)**: If term A has term B as synonym then term B has term A as synonym;
- **Hyponym (A, B) ^ Hyponym (B, C)** ♠ **Hyponym (A, C)**: If term A has term B as hyponym (more specialized term) and B has term C as hyponym then term A has term C as hyponym.

The ontology was developed in three stages: (a) textual description of the domain terms (Santos, Campos & Braga, 2004); (b) specification of the domain rules and the complementary rules above); and, (c) specification of the relationships among the terms. Representing the ontological terms of the e-learning domain demands to specify a model for the domain. The model gathers together terms and functionalities commonly used by the community. The features model and the analysis of the different kinds of elearning applications allow us to identify the common functionalities and to start the process of modeling the ontology. Table 1 presents a partial representation the ontology sub domains, sectors and sub sectors.

Each term of the sectors and sub sectors were described and related to others related terms. The main functionalities, sub functionalities and specific functionalities of elearning systems were also described. For example, the sub sector **Virtual classroom** was defined as multi user environments in where online resources are used to facilitate the learning process among students, between students and instructors, and between a class and wider academic and non-academic community. It also can be defined as a teaching and learning environment located within a WBI system.

Virtual classroom sub sector can be described under software perspective. As summarized in Table 2, Virtual Classroom sub sector encompasses many functionalities, sub functionalities and specific functionalities, converted in computational tools that support the work of its three kinds of users: the administrator, the instructor (or teacher or tutor) and the student.

Administrator tools congregate, far example, the functionality Management system tools that in its turn include one or more than one specific functionality. These specific functionalities are known by terms as Authorization tools, Security access, Remote access tools, Crash recovery tools, etc. Other functionality is Course management tools that includes a set of tools or software facilities, designed by terms, as Student support tools, Instructor support tools, Online registration, Guest account creation, System installing tools, Course fault retrieval tools, Resource monitoring.

Instructor tools include a set of functionality able to provide support to Course development by means of many software functionality, designed by terms such as Syllabus facilities, Course materials editing tools, Multimedia features, Automated glossary tool, Automated index tool, Importing/exporting capabilities, Links to supplemental resources capabilities, Presenting information tools, Previewing courses resources; and, Students assessing by means of tools designed by terms as Assessing tools, Self-evaluation, Peer review, Instruments of assessment, with specific functionalities, such as Online testing, Essays, *Portfolios,* Reports, and so on.

Virtual Classroom systems also offer a set of student tools, such as Productivity tools that gather together many software functionalities, designed by terms as Bookmarks, Calendar, Progress Review, Orientation/Help, Search tool for course content, Student involvement tools, Self assessment, Student community building, Student portfolios, Help online, Online testing, File download and File upload.

 $\mathfrak{g}_{\mathfrak{p}}$ domain: Web-based instruction

All terms of the ontology work as user guides, orienting searches at the digital library.

3. Digital Library on E-learning Domain

Digital library is a web-based electronic storage and access environment for information stored in the digital format either locally in a library, or in a group of networked libraries, or at a remote location (Cleveland, 1998). It also means an integrated set of services for capturing, cataloging, storing, searching, protecting, and retrieving information. It comprises digital collections, services and infrastructure to support lifelong learning, research, scholarly communication and preservation.

In agreement with Cleveland' point-of-view, the understanding about digital library is quite different according to its specific users:

- for computer scientists and software developers, collections of computer algorithms or software programs are digital libraries.
- for database vendors or commercial document suppliers, their databases and electronic document delivery services and digital libraries.
- for large corporations, a digital library is the document management systems that control their business documents in electronic form.
- for a publisher, it may be an online version of a catalogue.
- and for at least one very large software company, a digital library is the collection of whatever it can buy the rights to, and then charge people for using.

For many people the Web is not a digital library, because digital libraries are libraries with the same purposes, functions, and goals as traditional libraries - collection development and management, subject analysis, index creation, provision of access, reference work, and preservation. Due to its inherent complexity, current tendency in building digital libraries is to move forward in small, manageable, evolutionary steps, rather than in a rapid revolutionary manner. We are following this tendency.

The search engine is an important point for digital library. Traditionally, available approaches for search at the Web are based on the use of keywords. These approaches often work searching information at the syntactic level, returning to the user a large amount of unwished results. It happens because it is lacking descriptive information about information (metadata). In addition, it is also lacking the semantic about different vocabularies used to describe similar information into a same context. These issues decrease the quality and accuracy of many available search engines.

Digital libraries commonly use more sophisticated approaches. DLNET (Pushpagiri & Rahman, 2002), a digital library for lifelong learning on engineering and technology domains, for example, adopts other two approaches: an advanced search and a taxonomy-based browse search. A three-level taxonomy has been designed to subclassify engineering field into two subgroups. Our approach is a bit different since it provides ontological search to guide semantic querying, by using domain ontology built from the features model basis. It is potentially more precise because the search goes deeper searching information in more detailed level of specificities. Figure 1 shows a partial view of the ontological search graphical interface of our digital library for the sector Integrated and Distributed Learning Environments (IDLE).

Fig. 1. The Ontological Search Graphical Interface Sector Integrated and Distributed Learning Environments

In order to provide the ontological search, each searchable information must be classified according to the domain terms. Successful matches require correctness in this classification of the information, at the moment of its storage into the library database and it must follow closely the ontology terms.

The library has three kinds of users to portray different information meanings and point of view:

- **Library information providers**: they insert and classify information according to ontology terms;
- **Domain experts**: they examine and classify information. They evaluate the relationship between stored information and user needs. They also evaluate the coherence between the information classification and the ontology terms. These users have access to an online Web feature that allows them to assign values to the straight of the semantic relationship of an information and the domain vocabulary;
- **Library end users**: they utilize the search engine to find information that fulfills their needs. They use our ontology to select the terms related to their search, according to the interest features.

Word sense disambiguation in semantic analysis involves picking the correct set of senses in order to maximize the global needs. It is so crucial for the approach success that we ask the expert users to explicit the straight of the relationship between the information to be inserted into the library and the ontology terms. It is a value that the experts and other library users assign to their classification of the information regarding the ontology terms. When experts and users classify information, they assign the relationship level of the information with ontology terms, following the ontology organization, i.e.: sub domain, sector, sub sector, functionalities, sub functionalities and specific functionalities. There are 5 relationship levels: Tightly related (score 4), Partially related (3), Related (2), Loosely Related (1) and Not related (score 0).

The semantic issue also involves different vocabularies describing similar information in similar domains or contexts. This is a complex semantic problem that the digital library tries to solve using a set of complementary rules (section 2), to link *inter* and *intra* domains terms, providing the necessary mapping among terms of the same domain and of different domains. In addition, it updates domain information, reinforcing the straightness or weakness of terms relationship. The library works with three kinds of relationships: synonym (same semantic terms), hypernym (more generic terms) and hyponym (more specific terms). In this way, the user's search can be rewritten with the same semantic, replacing the searched term by related terms in the same ontology or in different ontologies using proposed mapping (Oliveira et al, 2004). A query using unknown terms in a specific ontology can be translated into terms of related ontologies, resulting in more satisfactory matches for users. The mapping between ontologies works under a set of rules and an inference machine.

3.1. Search Models in the Digital Library

At the digital library, users can query using keywords or can use the ontological search:

 Keyword search: it aims to identify the ontology terms related to the words typed by users at the search engine interface. The engine searches the term looking at the synonym database and if it does not find any similarity, it searches for synonyms, hypernyms and/or hyponyms in the same ontology and also in related ontologies. In this manner, the search universe will be enlarged, but although the search is more precise because it will involve different vocabulary for the same domain. Matches are presented in decreasing level according to the relationship level between term and available information.

 Ontological search: users also can request an advanced search. For that purpose, the digital library presents a graphical interface with the ontological terms (Figure 1). The users browse on the digital library, searching for information, through the different hierarchical levels of the ontology, choosing at each level, the interested terms. They must select one or more sub domain, one or more sectors and one or more sub sectors (see table 1). They also can search terms for e-learning functionalities. The search engine show up first all functionalities and later all functionalities related to the chosen functionalities. At the end of this process, the search engine presents the list of matches related to the terms as information links.

3.2. Implementation issues

The digital library was implemented using a Component Based Software Engineering (CBSE) approach, named UML Components (Cheesman & Daniels, 2001), customized in order to fulfill some specific needs. CBSE exploits interrelations between preexisting components and reuse of components that have been exhaustively tested to reduce complexity and costs of software development (Braga, Werner and Mattoso, 1999). We also used EJB (Enterprise Java Beans) technology to implement the Digital Library. EJB technology enables rapid and simplified development of distributed, transactional and portable applications based on Java technology.

It follows the standard interface of Web pages, with common icons and highlighted links. Even so, future works intend to test system interface usability, by means of laboratory tests. In the computational point-of-view, over time, we expect to improve the search engine accuracy forming a knowledge base that will compare and accumulate the needs fulfillment of end users and the provided information, making the searches progressively more accurate.

4. Conclusions and Future Works

Today, lifelong learning is mandatory for everyone because the speed of knowledge production changes marketplaces and stresses professional skills. Current programs involved in lifelong learning provide a set of educational activities and resources at the Web. It can be a powerful support for helping these initiatives, but finding information in Web databases remains a hard task. Its approach to information retrieval basically considers the syntactic aspects involved in the searches, often neglecting the semantic aspects. Our contribution was to develop a digital library to help the encouragement of lifelong learning activities. We started this process, including some e-learning information. The expectation, moreover, is that our information providers continually insert massive information following the ontology terms and experts validate provider's classification, in order to deliver the digital library to its end users – the teachers - soon. In long term, our intention is to extend the digital library access to all Brazilian teachers.

Acknowledges: The Brazilian research boards FAPEMIG and CNPq, UFJF and OdysseyShare Project (COPPE/Sistemas-UFRJ).

References

- Berners-Lee, T.; Handler, J.; Lassila, O. (2001)*. The semantic Web. Scientific American*, May 2001
- Braga, R.; Werner, C.; Mattoso, M. (1999). Odyssey: A Reuse Environment based on Domain Models. *Proceedings of IEEE Symposium on Application-Specific Systems and Software Engineering Technology (ASSET'99),* Richardson, Texas, March.
- Cheesman, J., Daniels, J. (2001), *UML Components A simple Process for Specifying Component-Based Software*. Addison Wesley Longman.
- Cleveland, G. (1998). Digital Libraries: Definitions, Issues and Challenges. In http://www.ifla.org/VI/5/op/udtop8/udtop8.htm
- Cohen, S. (1994) Feature-Oriented Domain Analysis: Domain Modeling; Tutorial Notes; *3rd International Conference on Software Reuse;* Rio de Janeiro, November
- Embley, D.W., Campbell, D.M., Smith, R.D. and Liddle, S.W. (1998). Ontology-Based Extraction and Structuring of Information from Data-Rich Unstructured Documents. In *Proc. CIKM'98: Conference on Information andKnowledge Managemen*t, Bethesda, MD, USA, ACM Press: New York, pp. 52-59
- Guarino, N. (1998). Formal Ontology in Information Systems*. Proceedings of FOIS'98*, Trento, Italy, June. Amsterdam, IOS Press, pp. 3-15.
- Guarino, N. (2002) *Ontology-Driven Conceptual Modelling: Advanced Concept*. ISBN:3-540-44277-4.
- Oliveira, A. M; ; Braga, R.; Campos, F.; Mattoso, M. (2004). GOS: Especificação de um Mecanismo de Busca e Recuperação de Componentes. *Proceedings 7º Workshop Iberoamericano de Ingeniería de Requisitos y Ambientes Software Ideas 2004*, Arequipa, Peru.
- Paulsen, M. F. (2002). Online Education Systems: Discussion and Definition of Terms. NKI *Distance Education*, July. Available in *http://home.nettskolen.com/~morten*
- Pushpagiri, V. P.; Rahman, S. (2002). DLNET: A Digital Library Architecture for Lifelong Learning, *2nd IEEE International Conference on Advanced Learning Technologies*, Kazan, Russia, Sept.
- Santos, N; Campos, F. C. A; Braga-Villela, R. M. (2005). A Digital Library for Lifelong Education on E-learning Domain. *Proceedings of E-Learn2005*. Vancouver, October (to appear).
- Santos, N; Campos, F. C. A; Braga-Villela, R. M. (2004). Ontology for Web-Based Learning Systems Domain. *Cadernos do IME – Série Informática*. Universidade do Estado do Rio de Janeiro. Rio de Janeiro. Brasil (also available in http://www.ime.uerj.br/cadernos/cadinf/.
- The Learning Technology Standards Committee of the IEEE Computer Society (2001). Available in http://jtc1sc36.org/doc/36N0175.pdf at July 4th 2005)
- Weinstein, P. and Alloway, G. (1997). Seed Ontologies: Growing Digital Libraries as Distributed, Intelligent Systems. In *Proceedings of the Second ACM Digital Library Conferenc*e, Philadelphia, PA, USA (July), 1997, ACM Press
- Werner, C. M. L.; Mangan, M.; Murta, L. G. P.; Pinheiro, R; Mattoso, M.; Braga, R.; Borges, M. (2003). OdysseyShare: an Environment for Collaborative Componentbased Development. In: *Proceedings IEEE International Conference on Information Reuse and Integration*. Las Vegas. v. 1. p. 61-68.

www1 -e-Learning Cybrary: http://www.co-i-l.com/elearning/about/