The Use of Ontologies in Web-based Learning

Marvin Tan and Angela Goh
School of Computer Engineering, Nanyang Technological University,
Blk N4, Nanyang Avenue,
Singapore 639798
{ps7726426d, asesgoh}@ntu.edu.sg

Abstract

This paper proposes the use of domain-specific ontologies to provide semantics-based classification, navigation and query for learning repositories. Ontologies are used with metadata to classify resources into specific domains. Concepts from the ontologies are associated with these resources, thereby facilitating navigation of the e-learning material via a conceptual map. We also propose semantics support for query processing. By selecting a specific domain, the corresponding ontology will enable users to formulate conceptual-based multi-criteria queries, leading to more relevant and precise results.

1 Introduction

The advent of the Internet has changed computer-based learning tremendously. Learners have increasingly turned to web-based learning in order to access learning resources across the World Wide Web. In the past, learning repositories provided storage for large volumes of information, with index-and-search capabilities to facilitate self-paced learning. The resources available from these libraries contain links to other web resources holding the actual content. They are often catalogued using proprietary taxonomies or classification schemes and the accompanying metadata are usually proprietary. Each Learning Repository (LR) may have its own classification taxonomy. As a result, different digital libraries were unable to exchange resources with one another efficiently, reducing the effective reach of these libraries. Users also had to cope with the differences between metadata formats and terminologies employed, making the resource identification process more laborious. The introduction of metadata standards in the learning domain, such as the IEEE/IMS LOM (Learning Object Metadata) standard [IMS Global., 2001], has corrected much of these shortcomings. The metadata standards are meant to support the exchange of learning resources between repositories. Theoretically, users are able to identify these resources via a common set of description terminologies. In practice, however, although the IMS Metadata standard provides elements for classification of resources, these elements are not mandatory and the elements basically accept any kind of text input. This can be a problem, as users may find it difficult to relate learning resources from different repositories. This is because the 'meta-data content' created are neither represented in machine-readable semantics nor created using standardized semantics. Ontologies may help to achieve semantic interoperability. An ontology is defined as "an explicit specification of a conceptualization and provides an agreement about a shared conceptualization of a given domain of interest" [Gruber, 1993]. With proper semantic support in the form of standardized ontologies, the sharing of learning resources will be more effective. This paper proposes a framework that harnesses the benefits of ontologies and machine-readable semantics. The objectives are summarized as follows:

- To provide semantics-based classification of learning resources based on metadata
- To provide visualization of conceptual maps for domain ontology and the lists of associated resources.
- To provide more semantically precise query capabilities for identification and retrieval of resources using domain-specific ontologies.

2 Overview of Proposed Framework

Figure 1 shows a general overview of an Ontology-based Framework and its interaction with both human users and different data sources. Details of each component will be elaborated on in the following sections.

2.1 Ontology-Based Classifier (OBC)

The first component is the **Ontology-Based Classifier (OBC)**. Learning resources are classified into different domains with the use of an ontology. For the purpose of testing, the Upper Cyc Ontology [Cyc, 2002] was used. This can be replaced by other suitable ontologies such as SUMO (Suggested Upper Merged Ontology) [SUMO, 2004]. By using the metadata available from the resource, the OBC will match the key terms with concepts within the ontology to determine the best match domain as well as the other possible domains the resource may belong to.

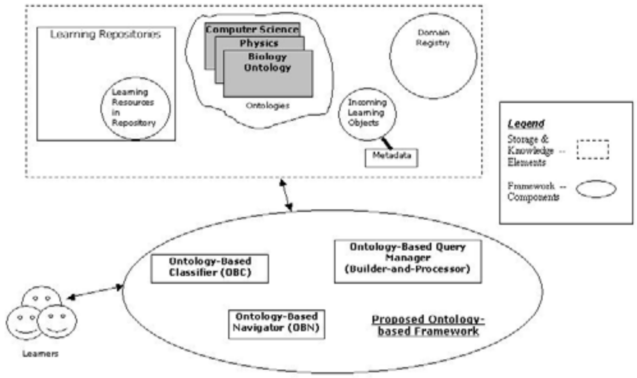

Fig. 1. Overview of Proposed Ontology-Based Framework

We will assume that the metadata taken from the resource is adequate and semantically-correct. While this assumption may not be valid for some resources, most of the resources encountered should have basic metadata that will describe the content of the resources adequately. We do not expect any system to be able to operate properly with inaccurate or incorrect metadata. An overview of the OBC is shown in figure 2.

The classification process is as follows:

1. The elements, 'title', 'keyword', 'description' and 'catalog' are retrieved from the XML metadata document that accompanies the learning object.
2. NLP pre-processing techniques such as "stopping" and "stemming" are performed on the text blocks (e.g. description). For each remaining term, the frequency is obtained and a histogram-like data structure is obtained.
3. Each word obtained is sent to an interface to discover if the concept represented by the word is found in any of the available ontologies. If found, the concept and the location of the concept is stored with the original word.
4. The concepts are then associated to the particular learning resource object and are used to classify the resource. As the concepts may be of different levels, top-level concepts will form the basic categories of this resource. Currently, we assume that every concept carries the same weight, as the ontologies we are employing do not allow us to impose weights directly. In addition, ontologies are represented in graph-like forms that do not show any hierarchy. Hence, it will be difficult to introduce weights without affecting the ontologies.

Using an IMS record taken from [Li, 2003], an example of this classification process is follows: Some of the critical terms obtained at the end of the preprocessing include microbes, organism, biosphere, entities, fossil and cycle. The terms, microbes, organism, biosphere, fossil are found in the Biology domain, while entities and cycle are found in other general ontologies. Hence, the resource is clearly in the Biology domain.

Learning resources can also be classified across domains. Concepts may also belong to several domains depending on the context. It has been mentioned that weights are not imposed at concept level but terms that belong solely to a particular domain are more indicative of the domain. Similarly, terms belonging to many domains are less authoritative. Hence, we will assign weights to concepts according to the number of domains they belong to. The significance of the concept in the classification process is inversely related to the number of domains the concept can be found under. In addition, we will consolidate the frequency of the terms or concepts under multiple domains. The domain that appears most frequently could be the common domain that these concepts belong to. In our framework, we maintain a list of resources associated with each concept within the domain-specific ontology and each list will be updated during the classification process.

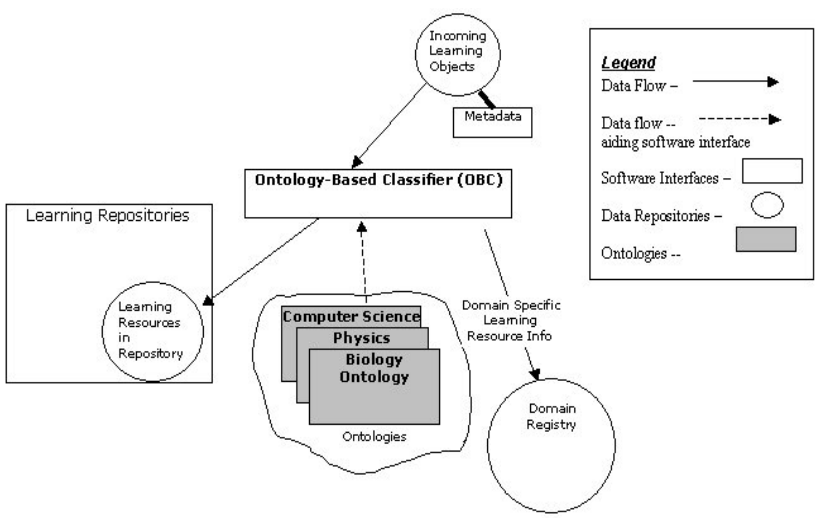

Fig. 2. Overview of Ontology-Based Classifier

2.2 Ontology-Based Navigator (OBN)

The second component is the **Ontology-Based Navigator (OBN)**. Currently most LRs allow the browsing of resources through the navigation of the classification hierarchies. We propose to navigate resources using a conceptual map view. By displaying the concepts from the ontology in a graphical form, users may navigate the concepts and view the list of resources associated with each concept. Figure 3 shows how the different elements interact with the OBN.

Navigation Interfaces for LRs are extremely important as they provide the first level of support for users looking for specific information. Many LRs provide two basic methods of accessing available learning resources. The first method is to provide a search interface that allows users to look for resources based on certain criteria. The query processor will then present a list of possible results. Details of the resource querying will be given in the next section. The second method is to provide browsing based on topics and keywords. Usually, the user is presented with a list of main topics or category. The user selects a particular category and the system presents another list of sub-topics under the category selected. In this way, the user is able to 'specialize' a request for a certain class of materials.

Keyword-based browsing refers to the presentation of a list of indexed keywords to the user. When the user selects a particular keyword, he or she is presented with the list of resources containing the selected keyword. Both forms of browsing have their merits in providing a more structured and organized approach to filter learning material. However, they narrow the learning experience as users tend to zoom into an specific area of interest without getting to learn about related concepts and ideas. There is little room for adaptive learning.

The user will be better rewarded if he or she is able to see the related concepts of a category or concept selected. We propose the following solution, which caters to different learning experiences. The classical system of hierarchical taxonomies is provided for users who are regular users of the LR and are only interested in particular resources. For general users, a 'concept map and index' interface is provided. On the same interface, an index list of all the top-level concepts are presented to the user. Selection of any concept results in the visualization of all directly related concepts and cursor support is provided to allow the users to navigate to indirectly-related areas of the conceptual graph. In this way, users can process related knowledge before zooming into the specific resources. To prevent information overload, the 'concept map' will be designed in a similar way as 'online street maps', where basic operations such as zoom and directional navigation are provided. Only limited information will be available to the user at any point of time, improving the assimilation of the information and any decision-making process. A typical navigational model of LRs is shown in figure 4(a) while our enhancement is presented in figure 4(b).

2.3 Ontology-Based Query Manager (OBQM)

The third component is the **Ontology-Based Query Builder-and-Processor (OBQM)**, which allows users to build ad-hoc queries based on the concepts provided within a domain-specific ontology. The criteria for querying may be formulated using the concepts from the ontology. Search engines today are built on top of huge networks of indexed keywords with neither semantics nor contextual information.

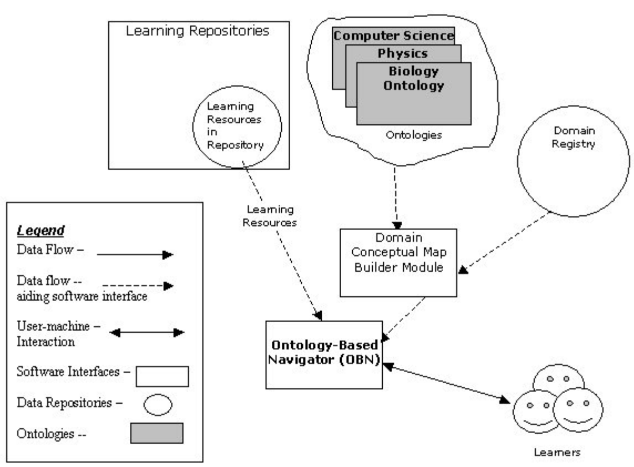

Fig. 3. Overview of Ontology-Based Navigator

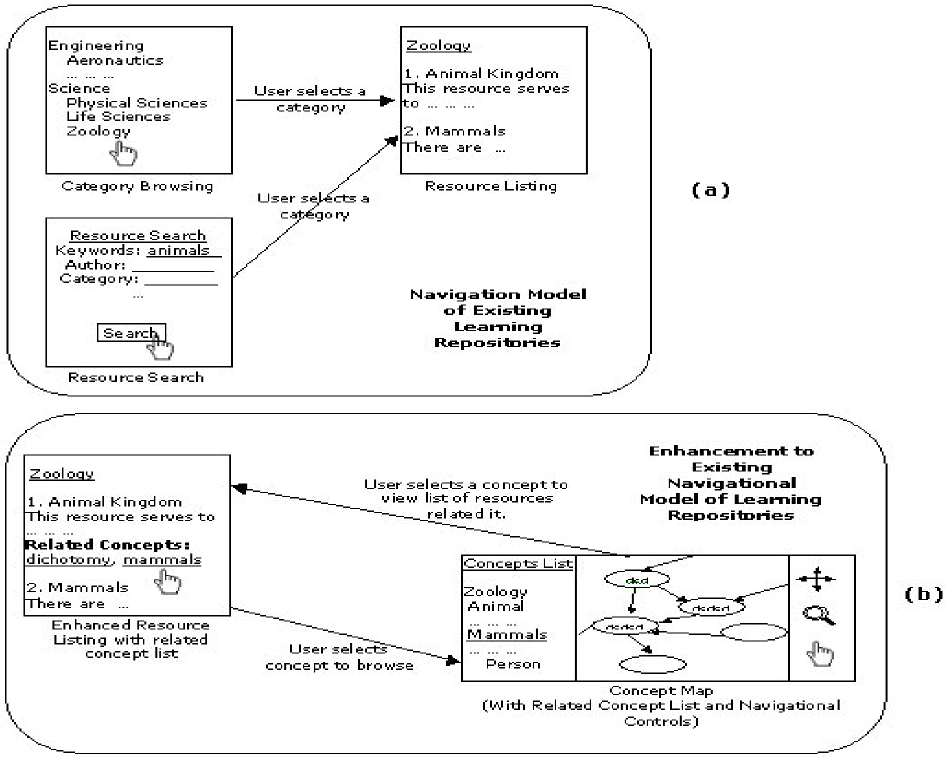

Fig.4. (a) Existing Navigational Model of LRs and (b) our Proposed Enhancement

An example search was performed at www.google.com for the keyword, "arm". We had expected articles and information on a part of human anatomy such as the bones and ligaments making up the arm. Instead, the first few results were on Microprocessors and the remaining results included brand names, club names, firearms and an acronym for 'Atmospheric Radiation Management'. Within the top 10 results, there was not a single link to an article on the 'arm' anatomy. When we refined our search to human arm, the results were better, returning articles related to the anatomy, animation or news containing the words human arm. However, the number of results returned was 1.5 million articles, which makes it impossible for a user to filter through individually. If the user was only strictly interested in human arm anatomy, then the top 10 results contained only 4 totally relevant results (2 of which were from the same site).

The problem of irrelevance remains in current search engine technologies that rely simply on keywords. In the context of Learning Repositories, this problem could be better managed. As each individual resource can be classified within specific domains, we could use the domains as primary filters. Users will be able to choose the domain ontology for the search to be performed on.

In addition, we also propose the use of domain-specific concepts to refine these queries. For example, when a user selects the Biology Ontology, a list of related concepts will be made available. Logical operators such as 'AND', 'OR' or 'NOT' will also be made available for the user to create combinations of these concepts. In this way, we will improve the quality of the results returned during the query process, using ad-hoc query builders of different complexities for different types of users. Novice users may adopt the keyword-domain search while more experienced users may use the available concepts to formulate more complex queries for better precision. The results will be presented in a ranked list. However, we propose the inclusion of the option to invoke the OBN. This allows users to obtain an overview of the concepts related to the results. Figure 5 presents an overview of the OBQM.

3 Domain Registry

As seen in figures 2, 3 and 5, a Domain Registry (DR) is present to keep additional information, which will enable our framework to function independently of the existing repositories. This Domain Registry is a novel idea of placing critical index information of learning resources from multiple learning repositories. To have the DR work successfully, we make an assumption that most learning repositories have a script or interface that exposes the record of a learning resource to Internet users. For web-based LRs, such interfaces are usually web services or dynamic web pages that can display learning resource details. The Domain registry will act as an independent information store, which allows users to access learning resources from different LRs through external interfaces.

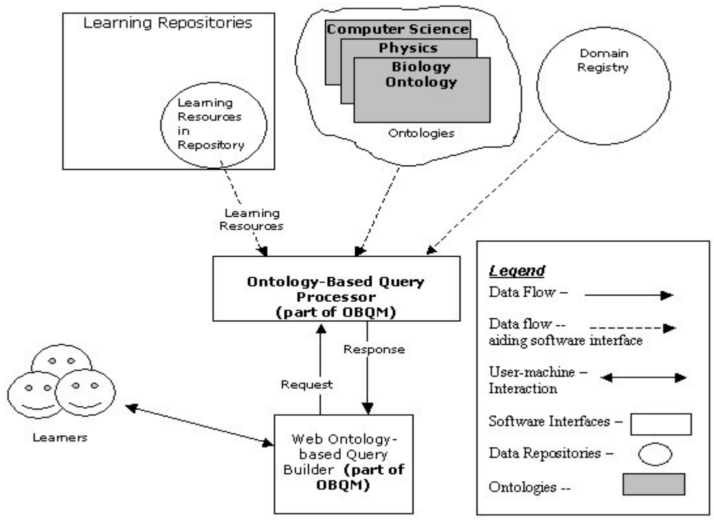

Fig. 5. Overview of Ontology-Based Query Manager

Each LR is registered at the DR by administrators who are required to provide a script or an interface URL and the necessary key parameters to identify or access specific learning resources. At the same time, the DR should also provide single-sign-on for users wishing to access several repositories through a common framework.

Hence, the DR plays a key role in establishing transparent loosely-coupled integration between different learning repositories. The registration of these repositories must be made relatively simple to encourage use of this service and as part of an open-source initiative, free access to the resources should be made available. However, the creation of such registries may encounter problems. One of the most critical is scalability. One of the objectives of this project is to provide ontological support for learning resources across the World Wide Web. Hence, we require the registry to be accessible on the Internet. Over the years, different communities have recognized that centralized registries are usually not scalable and not feasible unless the registry is monitored regularly and information is replicated across servers. Hence, we would like to examine peer-to-peer frameworks [Neijdl *et al.*, 2002] to implement distributed domain registries. The availability and reliability of such registries will be critical to the success of our proposed framework.

4 Related Work

In the past few years, there has been much interest in the roles of ontologies in the learning environment. One of the earliest research efforts on eLearning and the Semantic Web was discussed in [Stanjonovic *et al.*, 2001], where an eLearning scenario was presented. Ontology-based descriptions of content, context and structure of learning materials are employed to provide flexible and personalized access to the resources. The main difference between their scenario and our framework is the use of ontologies. The main ontology they adopt is a 'course ontology' that describes the context and structure of the learning resources. They also make use of the Ontobroker [Decker *et al.*, 1999] inference engine to answer queries and infer new facts from the existing information. However, the ontology used in [Stanjonovic *et al.*, 2001] is restricted to course description and the links with the learning resources are more structural than semantic.

The generation of flexible taxonomies for learning resources is also discussed in [Papatheodorou *et al.*, 2002]. Their methodologies are based on data mining techniques. The ontology discovery process consists of basic NLP (Natural Language Processing) and machine learning techniques such as pattern discovery. The main drawback in this process lies in the requirement of training data and its intractability. Since the level of accuracy cannot be guaranteed using machine learning techniques and large corpuses of training data are hard to come by, we prefer to employ heuristic techniques to achieve similar results. However, there was one interesting issue raised in the paper - the adaptive discovery of ontology-based taxonomies may result in the limitless depth of the hierarchical taxonomies and the display of the taxonomies may result in poor human-computer interaction design and information overload for the user. Similarly, the challenge of building and maintaining metadata using an ontology based on an extended dictionary was addressed by [Apted *et al.*, 2004].

Other related research in the support of learning using ontologies includes the 'Collaborative Courseware Authoring Support' [Dicheva *et al*, 2002] and 'Object-oriented collaborative course authoring environment'[Cristea and Okamoto, 2001]. [Dicheva *et al*, 2002] provides support for concept-based web courseware authoring with the use of a domain ontology. The ontology can be used by authors in queries or by the system to perform semi-automatic authoring activities. Courseware are then manually linked to the concepts by the authors. The system also provides domain engineering of the ontology. Hence, the ontology has to be built from scratch by the courseware authors. The two main problems are, firstly, the ontology is expected to grow exponentially and links to new concepts may have to be created continuously for earlier created courseware. Secondly, with the authors creating and maintaining the domain ontology, the resultant ontology becomes proprietary and unsynchronized with external ontologies.

"My English Teacher" [Cristea and Okamoto, 2001] focuses on an English upgrading course authoring environment. In this project, 'concept mapping' is used to support authoring where the concepts are built using keywords. 'Concept mapping' is the graphical technique of representing concepts and corresponding hierarchical relationships to present the domain model. It presents some ideas of resolving the display of concept maps or hierarchical views that are too wide or deep. At the same time, it also explores automatic concept mapping to content, which we will refer to in our project. However, the emphasis is on courseware authoring for a proprietary domain, whereas we are proposing a more generic framework for learning repositories

5 Conclusion

One of the key problems affecting the effectiveness of Learning Repositories is the lack of semantic support. Without this support, users are forced to spend more time to review irrelevant learning resources. At the same time, non-standard classification schemes also prevent efficient interoperability between these repositories.

Our proposed framework will remove most of these inadequacies by providing support in the form of commonly available domain-specific ontologies. By provided semantic support for the classification, navigation and query processing for Learning Repositories, users will benefit from a clearer conceptual view of resources as well as a more efficient learning experience through more precise retrieval of relevant information.

A key issue in our future work is to support Ontology growth. As research on Ontology engineering and management is still at an early stage, it is envisaged that considerable improvements to ontological support and tools will be seen in near future. In addition, the ontologies to be employed will constantly be updated. Hence, our framework must cater for the possible growth of the underlying ontologies. In our framework, local copies of the ontologies will be used. The local copies will be compared periodically with the master copies and the changes will be propagated to the rest of the system. In this way, we will be able to maintain the semantic accuracy of the repository and at the same time maintain conformance to the ontologies employed.

References

[Apted *et al.*, 2004] Trent Apted, Judy Kay and Andrew Lum, Supporting metadata creation with an ontology built from an extensible dictionary. In *Proceedings of 3rd Int, Conf on Adaptive Hypermedia & Adaptive Web-Based Systems* August 2004. Springer-Verlag.

[Cristea and Okamoto, 2001] Alexandra Cristea and T. Okamoto, Object-oriented collaborative course authoring environment supported by concept mapping in MyEnglishTeacher, Educational Technology & Society 4 (2), 104–115, 2001.

[Cyc., 2002] Cyc, http://www.cyc.com/cycdoc/vocab/ upperont-diagram.html, OpenCyc Selected Vocabulary and Upper Ontology, 2002.

[Decker *et al*, 1999] Stefan Decker, Michael Erdmann, Dieter Fensel and Rudi Studer. Ontobroker: Ontology based access to distributed and semi-structured information, In Database Semantics - Semantic Issues in Multimedia Systems, *IFIP TC2/WG2.6 Eighth Working Conference on Database Semantics (DS-8)*, Rotorua, New Zealand, pages 351–369, January 1999.

[Dicheva *et al*, 2002] Darina Dicheva, Lora Aroyo and Alexandra Cristea. Collaborative Courseware Authoring Support, In *Computers and Advanced Technology in Education*, Cancun, Mexico, pages 52–57, May 2002.

[Gruber, 1993] Tom Gruber. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In *Formal Ontology in Conceptual Analysis and Knowledge Representation*, N. Guarino, R. Poli (Eds.), Kluwer Academic Publishers, 1993.

[IMS Global., 2001] IMS Global, http://www.imsglobal. org/metadata/index.html, *IMS Learning Resource Meta-data Specification* Version 1.2, 2001.

[Li., 2003] Yaxin Li, Biological control: *A guide to natural enemies in North America*, sMETE.org, URL http://www.nysaes.cornell.edu/ent/biocontrol/, 2003.

[Neijdl *et al.*, 2002] Wolfgang Nejdl, Boris Wolf, Changtao Qu, Stefan Decker, Michael Sintek, A. Naeve, M. Nilsson, M. Palmer, T. Risch. Edutella. A P2P networking infrastructure based on rdf. *11th International World Wide Web Conference*, Honolulu, Hawaii, USA, May 2002.

[Papatheodorou *et al*, 2002] Christos Papatheodorou, Alexandra Vassiliou and Bernd Simon. Discovery of ontologies for learning resources using word-based clustering, *ED-MEDIA 2002*, Denver, USA, 2002.

[Stanjonovic *et al*, 2001] Ljiljana Stojanovic, Steffen Staab and Rudi Studer. elearning based on the semantic web. WebNet 2001 - *World Conference on the WWW and Internet*, Florida, USA, pages 23-27, 2001.

[SUMO, 2004] SUMO - Suggested Upper Merged Ontology. http://ontology.teknowledge.com/, 2004.