Ontology-based Information Integration in Virtual Learning Environment

Aijuan Dong and Honglin Li
Department of Computer Science, North Dakota State University, Fargo, ND 58105
aijuan.dong, honglin.li@ndsu.edu

Abstract

A good virtual learning environment should deliver relevant learning materials to learners at the most appropriate time and locations to facilitate learners’ acquisition of knowledge and skills. In this paper, we propose ontology-based information integration in virtual learning environment using ontology and web services. Relevant concepts extracted from domain ontology provide ontology-based browsing space that allows users to browse and select relevant terms of interest and increases the degree of relevancy. By using web services to integrate learning materials from heterogeneous public domain data sources, applications do not need to know the internal structure and working of public domain data sources, and reuse existing applications and resources. We use Gene Ontology, PubMed eUtils and Google Web APIs to demonstrate our idea. The implementation involves techniques in image and video processing, database management, programming, and multimedia learning materials presentation.

1. Introduction

Virtual learning environments, such as virtual conferences, virtual classrooms and virtual seminars, enhance learning experience by supporting learning at flexible time and locations. A good virtual learning environment should deliver relevant information to learners at the most appropriate time and locations to best advance the learners’ acquisition of knowledge and skills.

Historically, relevant information is manually integrated by teachers/presenters. While providing relevant information to learners, this approach involves lots of work. Research of automated retrieval of educational relevant material has been focused on computer simulation [4, 6]. The parameters of the simulation environment are used as input to the information retrieval. Rather than using simulation environments, paper [5] uses annotated video as an information retrieval interface.

With advances in computation capacity, network bandwidth, and information storage, the volume of learning materials increases dramatically in recent years. However, awash with huge and diverse learning materials, users still feel that retrieving and integrating educationally relevant learning materials in virtual learning environments is a challenging task. Two important issues that are often raised are: how to increase the degree of relevancy among retrieved learning materials and how to deal with heterogeneity of various data sources. Relevancy is usually domain dependent. Terms that are closely related in one domain might irrelevant in another domain. The heterogeneity of information sources exists in three aspects [9]: syntactic heterogeneity, structural heterogeneity and semantic heterogeneity. Despite efforts in addressing these two issues, they have not been sufficiently addressed, especially in multimedia domain, and remain as important and challenging problems.

Our major goals are: 1) to dynamically integrate domain-ontology into information retrieval and integration in an effort to deliver relevant information to learners; 2) to apply web services provided by public domain data sources so that heterogeneous data sources can be easily retrieved and integrated. The rest of paper is organized as follows: in section 2, we give an overview of the methodology; then ontology-based information integration in multimedia annotation and multimedia access is described in section 3 and 4 respectively; in section 5, we take Gene Ontology, PubMed and Google to illustrate our idea; section 6 concludes the paper and proposes some future research.

2. Methodology overview

To extract relevant learning materials internal and external to virtual learning environments and deliver them to learners, we propose ontology-based browsing space. Ontology represents key concepts and concept relationships in a machine-processable format and facilitates information sharing and exchange within a given domain. By semantically narrowing or widening queries based on semantic relationships embedded in domain ontology, relevant terms/concepts can be extracted. These terms/concepts provide an ontology-
based browsing space that allows users to browse and locate the ones that meet their particular retrieval and integration needs. Ontology-based browsing space enables semantic information retrieval and integration. Since relevancy information is extracted directly from domain ontology, in which domain knowledge is embedded, the degree of retrieval relevancy can be greatly improved. Moreover, users can select which ontology to interact with. The number of domain ontology that can be integrated is unlimited.

To address heterogeneity of public domain data sources, we propose the use of web services [1]. Web services have evolved as a result of building distributed applications. The biggest advantage of web services is interoperability. They change the way of application development and information sharing over heterogeneous information sources and are effective ways of addressing syntactic and structural heterogeneity. We argue web services are an effective mechanism to integrate external public domain data into virtual learning environments. By using web service to integrate heterogeneous public domain data into virtual learning environments, client programs do not need to know the internal working of the web service hosts, and reuse existing applications without reinventing them from scratch.

In our previous work [7], we proposed an ontology-based multimedia access platform that consists of two modules: integrated multimedia annotation module and ontology-based multimedia. In the following two sections, we give a detailed description of how to apply ontology-based information integration in these two modules.

3. Ontology-based information integration in multimedia annotation

The general process of integrated multimedia annotation generation is described in [7]. After the live presentation in virtual learning environment is over, captured individual multimedia object such as image, text, and video, is processed and annotation data, such as keywords, table of contents, and representative images, is extracted. Then the selected keywords are sent to the selected domain ontology to extract relevant terms. After that, both keywords and relevant terms are fed into the selected web service, which transforms the user query to SOAP requests, communicates with the web service host, and sends back the relevant documents to users/programs. Users/programs extract URL from the response and use it as annotation to corresponding public domain data source. At last, annotation data from both the presentation and external public domain data sources are integrated based on temporal, spatial, and semantic relationships among them. Moreover, relevant documents can be extracted at multiple levels, such as presentation level and presentation segment level.

4. Ontology-based information integration in multimedia access

Ontology-based multimedia access allows users to access multimedia learning materials based on domain ontology as illustrated in Figure 1. A user can start
with domain specific terminology. The system first retrieves relevant terms by querying domain ontology. Both query terms and relevant terms are then fed into both integrated multimedia annotation data repository and selected web services, which will then return relevant multimedia learning materials from both multimedia data collection of virtual learning environments and corresponding external public domain data source. Based on proposed architecture, one single search pulls out all the relevant material from sources both internal and external to virtual learning environments, which greatly simplifies and reduces work on users’ side.

The purpose of user management in Figure 1 is to manage user profiles. By knowing learning habits or research preferences of users, there is a better chance that a platform can present relevant information of more interest to learners.

5. Case study

Background

PubMed [8], developed by National Center for Biotechnology Information (NCBI), includes over 15 million citations for biomedical articles back to the 1950’s. Since the internals of databases involved are not made publicly available, the Entrez Programming Utilities (eUtils) are provided to retrieval information outside of the regular web query interface and is helpful for retrieving search results for use in another environment. In this paper, we focus on integrating PubMed literature database into virtual learning environments

Google Web APIs [3] is a web service that enables developers to access Google’s web search and to develop software. With Google Web APIs, applications can retrieve information from more than 8 billions web documents that are constantly refreshed. The results are returned as structured data. Developers can then write programs to transform the structured result to XML or other formats based on application needs.

Presentation data form Virtual Conference on Genomics and Bioinformatics (VCGB) [2] is used in this implementation.

Interaction diagram

In our implementation, the following sequence of steps describes general interaction among different system components as illustrated in Figure 2:

1) A user enters subject term(s), sends them to selected Gene Ontology.
2) Programs extract relevant terms from Gene Ontology, and display to user ontology-based browsing space.
3) The User selects relevant terms of interest and feeds both subject terms and relevant terms to controller servlet;
4) Controller servlet sends all terms to VLE servlet, PubMed servlet and Google servlet to extract relevant learning materials from VLEs multimedia data collection, PubMed data sources and Google web sources respectively:
   i. In VLE servlet, the servlet first searches presentation level annotation data. If there is a match, it searches video segments of that presentation. After that, the servlet searches the rest video segments with no matching presentation. At last it creates integrated multimedia learning material using SMIL, RealText and RealPix, links it to a HTML page and returns that HTML page to controller servlet.
ii. In PubMed servlet, the servlet first calls esearch utility of PubMed web service by sending out HTTP GET request with a URL containing all required parameters, esearch returns XML data that contains result set identifiers. PubMed servlet then extracts QueryKey and WebEnv from the XML data, sends out another HTTP GET request with a URL containing all required parameters. This request calls efetch utility and efetch returns document data to PubMed servlet. PubMed transforms the result using xsl and return them to controller servlet.

iii. In Google servlet, the servlet executes doSearch with Google license key and query parameters. Google Web APIs sends back query result in structured data format. Google servlet then converts structured data to XML and transforms XML data to HTML using xsl. Finally, it returns HTML results to controller servlet.

5) Controller servlet organizes the results from three servlets and sends back the final result as a HTML page to the user/program. Figure 3 is an example HTML page sent back to users.

For multimedia learning materials retrieved from virtual learning environments collection, a friendly and efficient browsing interface is provided. The detail of the interface can be found in [7]. Relevant documents are grouped together under related documents link in index pane. When clicked, relevant documents are displayed in browser pane. Users can pause or continue watching the presentation while browsing related documents.

6. Conclusions and future work

In this paper, we propose ontology-based information integration in virtual learning environments. To increase the degree of relevancy, we propose the use of domain ontology during information retrieval and integration. To address the problem of heterogeneity, we proposed the use of web services. We use PubMed eUtils and Google Web APIs to demonstrate our idea, the learning materials extracted from PubMed data sources and Google web sources are integrated with that of virtual learning environments and are delivered to learners’ at the most appropriate time and locations.

In the future, we will try to explore and incorporate more domain ontologies and web services into our test bed. We are also interested in applying ontology-based ranking algorithm in our later work.

References