

Study of E-learning Information Retrieval Model based on Ontology

R. Lakshmi Tulasi
QIS College of Engineering &
Technology
Ongole
India.

M. Srinivasa Rao, PhD.
Professor
School of Information
Technology
JNTU, Hyderabad

G.Rayana Gouda
Professor
Hubli

ABSTRACT

Research work in the field of E-learning is represented by a broad area of applications ranged from web learning to virtual courses. Web-based courses offer obvious advantages for learners by making access to educational resource very fast and relevance, at any time or place. The objective of this paper is to create an ontology based domain specific representation for e-learning domain. This paper analyzes the drawbacks of traditional keyword based search engines and proposes the need for semantic based intelligent information retrieval systems. This paper presents ontology based information retrieval (IR) for e-learning domain which is developed using Protégé tool. This discusses about the E-learning information retrieval model which consists Resource collection module, Semantic procession module, Resource collection module, resource integration system.

General Terms

Computer Science, Information Retrieval.

Keywords

E-learning, IR, Ontology, OWL, Semantic Web, Protégé tool.

1. INTRODUCTION

The World Wide Web has become a vast resource of unstructured/semi-structured information. The problem is that finding the information that the individual desires is often quite difficult, because of complexity in organization and quantity of the information stored. So we need some advanced techniques and tools that can handle semantically.

E-learning is not just concerned with providing easy access to learning resources, any time, anywhere, via a repository of learning resources, but is also concerned with supporting such features as the personal definition of learning goals, and the synchronous and asynchronous communication, and collaboration, between learners and between learners and instructors.

Various research works[1] in the past few years have identified a number of problems associated with the traditional web search engines which are based upon the keyword based index construction and Boolean logic based query matching. The major problems with content based approaches are poor precision and poor recall i.e out of the total no of results possible for a search query only a fraction is received(poor recall) and out of the results that are returned only a fraction of them are relevant(poor precision). The search results that are returned are so large in number and it is very difficult for the user to navigate through all the results

to find relevant information. It is difficult for the user to determine which query words give better results.

The semantic web is an evolving development of the WWW in which the meaning of information and services on the web is defined making it possible for the web to understand and satisfy the requests of people and machines to use the web content. The intelligent extraction system searches the information on the concept but not on the matched words. It can give the answers to users quickly and precisely.

Semantic search techniques for the web work on the basic assumption that the underlying information has been represented in a structured way which allows it to be given semantics. This semantic structure can be used by a context based search mechanism (semantic search) to return results that are not only lesser in number but are more accurate and more relevant to the user query compared with keyword based approaches.

The basic difference between the two search techniques is the keyword based search engine view the web as a collection of web documents in turn each document can be viewed as a collection of keywords by the search engine. But in other way a semantic search engine view the web as a collection of domains. Each domain is defined by no of concepts which are connected to each other with relationships and is defined by a number of attributes.

Ontology is often used to model domain-specific knowledge for a particular domain. Ontologies are explicit specifications of the conceptualization and corresponding vocabulary used to describe the domain [Gruber 1993]. "Conceptualization" refers to identifying the key components of the domain in terms of the domain in terms of which the domain can be fully defined." Explicit "means each concept should be uniquely completely and unambiguously defined in such a way that software could understand it. "Specification" refers to identifying the various properties that are used to define the particular concept and defining the relationships between them. The number axioms are also defined, based on which are inferences can be drawn.

The rest of the paper is organized as follows: In section 2 related works on E-learning are discussed. In section 3 different semantic web technologies are discussed. In section 4 the case study of E-learning domain and query results are shown with the help of description logic. In section 5

E-learning information retrieval model is discussed based on ontology.

2. RELATED WORK & MOTIVATION

Recently, several researchers studied the issue of Web based application. F. P. Rokou et al. [4] distinguished three basic levels in every web-based application: the Web character of the program, the pedagogical background, and the personalized management of the learning material. They defined a web-based program as an information system that contains a Web server, a network, a communication protocol like HTTP, and a browser in which data supplied by users act on the system's status and cause changes. The pedagogical background means the educational model that is used in combination with pedagogical goals set by the instructor. The personalized management of the e-learning materials means the set of rules and mechanisms that are used to select learning materials based on the student's characteristics, the educational objectives, the teaching model, and the available media. Many works have combined and integrated these three factors in e-learning systems, leading to several standardization projects.

Some projects have focused on determining the standard architecture and format for learning environments, such as IEEE Learning Technology Systems Architecture (LTSC), Instructional Management Systems (IMS), and Sharable Content Object Reference Model (SCORM). IMS and SCORM define and deliver XML-based interoperable specifications for exchanging and sequencing learning contents, i.e., learning objects, among many heterogeneous e-learning systems. They mainly focus on the standardization of learning and teaching methods as well as on the modeling of how the systems manage interoperating educational data relevant to the educational process [5].

IMS and SCORM have announced their content packaging model and sequencing model, respectively. The key technologies behind these models are the content package, activity tree, learning activities, sequencing rules, and navigation model. Their sequencing models define a method for representing the intended behavior of an authored learning experience, and their navigation models describe how the learner and system initiated navigation events can be triggered and processed.

Juan Quemada and Bernd Simon have also presented a model for educational activities and educational materials [6]. Their model for educational activities denotes educational events that identify the instructor(s) involved and take place in a virtual meeting according to a specific schedule. F. P. Rokou et al. [7] described the introduction of stereotypes to the pedagogical design of educational systems and appropriate modifications of the existing package diagrams of UML (Unified Modeling Language).

The IMS and SCORM models describe well the educational activities and system implementation, but not the educational contents knowledge in educational activities. Juan Quemada's and F. P. Rokou's models add more pedagogical background by emphasizing educational contents and sequences using the taxonomy of learning resources and stereotypes of teaching models. But the educational contents and their sequencing in these models are dependent on the system and lack standardization and reusability. Thus, we believe that if an educational contents frame of learning resources can be introduced into an e-learning system,

including ontology-based properties and hierarchical semantic associations, then this e-learning system will have the capabilities of providing adaptable and intelligent learning to learners. The hierarchical contents structure is able to show the entire educational contents, the available sequence of learning, and the structure of the educational concepts, such as the related super- or sub- concepts in the learning contents. For this purpose, ontology is introduced in our model. such as XML and RDF(S), so as to share and reuse it in any web-based knowledge system [8,9]. Thus, we have devised a model that provides the contents structure using an ontology for an adaptive and intelligent e-learning system.

Zhang jinghua proposed a Semantic Web Based Personalized Learning Service for programming course in e-learning. In this he used personalized learning service the basis of this service is resource base, ontology base, and strategy base. This is not suitable for all strategies [10].

3. SEMANTIC WEB TECHNOLOGIES

OWL (Web Ontology Language) is a W3C recommendation Web Ontology description language. OWL can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms. This representation of terms and their interrelationships is called ontology. OWL has more facilities for expressing meaning and semantics than XML, RDF, and RDF-S, and thus OWL goes beyond these languages in its ability to represent machine interpretable content on the Web. OWL is a revision of the DAML+OIL web ontology language incorporating lessons learned from the design and application of DAML+OIL. OWL has three increasingly expressive sublanguages: OWL Lite, OWL DL, and OWL Full.

Ontologies are created by ontology engineers with the help of domain experts. Ontology construction is an iterative process and involves the following steps:

- **Design:** Specifies the scope and purpose of the ontology. Also reveals the relationship among classes and subclasses.
- **Develop:** Decides whether construction of ontology has to be done from scratch or to reuse an existing ontology.
- **Integrate:** Combine the developed ontology with the already existing one.
- **Validate and Feedback:** The completeness of the constructed ontology is verified with the help of automated tools or by seeking the opinion of the experts.
- **Iterate:** Repeat the process and incorporate the changes given by the expert. Ontologies are commonly encoded using ontology languages.

Different ontology constructing methods are available like TOVE method, METHONTOLOGY method, Frame work method, KACTUS method, SENSUS method, DEF5 method. After that domain ontology creating 7 steps method developed by Stanford University School of Medicine is acceptable. Ontology creating tools are Ontolingua, Ontosaurus, Protégé, webODE, OntoEdit and so on. Protégé is the tool of choice for constructing ontologies which is open source and can be used to build frame based, as well as OWL ontologies. The OWL plug-in for Protégé allows the Ontology Engineer to load, save, edit and visualize ontologies in OWL and RDF.

Domain experts are responsible for identifying the set of concepts for their specific domain. These concepts are defined by their attributes (called data properties in Protégé) and relationships between various concepts of the domain (called object properties in Protégé). A set of rules called axioms are defined which allows a reasoner (Protégé provides a number of integrated reasoning tools like fact ++ and pallet etc) to draw inferences for extending the ontologies (future needs) and for extracting and modeling the *non explicit* domain knowledge.

4. CASE STUDY

4.1 Ontology for E-learning Domain

The main objective of this ontology is to create a knowledge base for E-learning. It provides relevant results based on domain specific knowledge and improves both the precision and recall.

4.1.1 Create new Ontology for E-learning Domain:

Start protégé .When welcome protégé dialog box appears, press 'Create New OWL ontology'. Enter 'E-learning. Owl' name in the place of default URI and save it in our PC.

4.1.2 Add a comment to ontology: Using 'Active Ontology' Tab in the 'Ontology Annotations' view adds a comment, "This is the ontology that describes various sources of E-learning.

4.1.3 Create subclasses of 'Thing': With 'Add subclass' button in the 'Classes' tab creates Access Type, Document,, GenericConcept, LearnersProfile etc as subclasses of 'Thing'. Repeat same process to create Facilitator, Learner as subclasses of Actor. Author, Manager, Presenter, and Teacher as subclasses of Facilitator. Concept, Course, Knowledge element as subclasses of Generic Concept as shown in Fig 2.

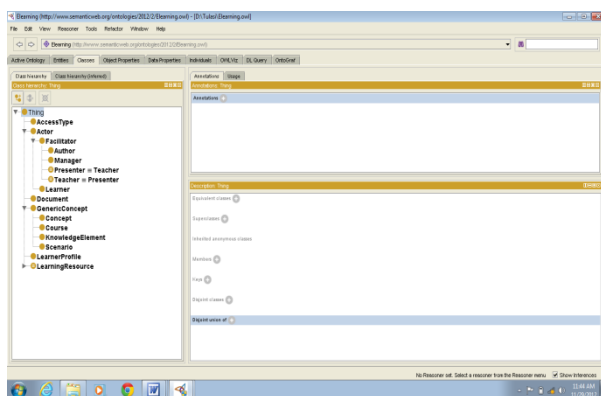
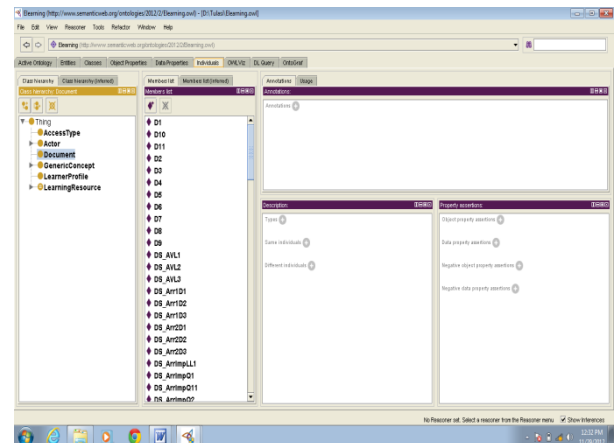


Fig 2: Named subclasses of 'Thing'

4.1.4: Add individuals to a class: Add individuals to a class using 'Add individual' button in 'individual members list view'. Ex: D1, D2, D3, D4, D5, D10, D11, D12 etc as individuals of Document as shown in Fig 3.

4.1.5 Add OWL Properties: With 'Add Object Property' button of the 'Object Properties' tab create Object Properties

like ConsistsOf Documents, ConsistutedOf, Creates, dependsOn etc. as shown in Fig 4.



Query1: Document and has Resource Type some (Proceeding or journal)

Result: Search engine displays list of documents those have resource type of proceeding or journal. Results are shown in Fig 6.

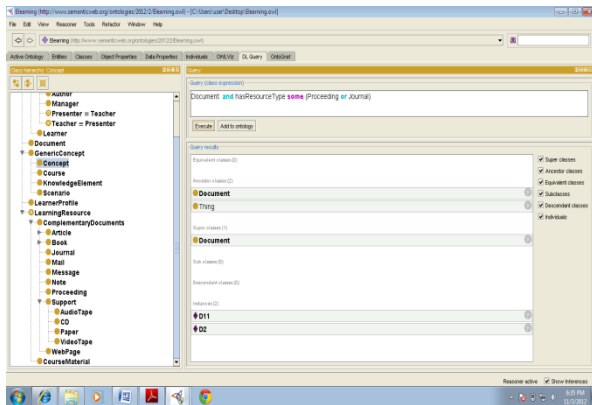


Fig 6: Query result for “Document that has resource type Proceeding or Journal”

4.1.8 Represent Ontology Graph which shows semantic relationships between classes and instances of E-learning Domain using ‘Onto Graf’ tab. It is shown in Fig7.

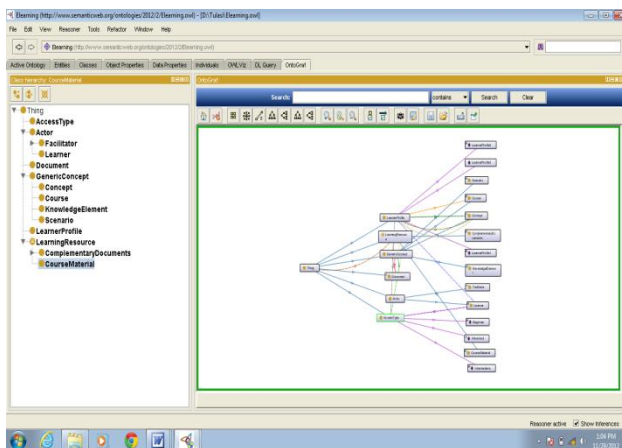


Fig 7: Ontology graph for E-learning domain

5. E-LEARNING INFORMATION RETRIEVAL MODEL BASED ON ONTOLOGY

This system is composed of 4 modules: ontology Management module, resource collection module, semantic procession module and retrieval module. Ontology management module plays the role of building ontology, storage ontology and ontology evolution; Resource collection module can collect information from the resource integration system and store in net for procession further; Semantic procession module can process the document from the resource collection module. It implements the semantic index for non structure and half structure data, then extracts metadata, encodes RDF triple, and stores into semantic metadata base.

Retrieval module receives user query request, retrieval preprocessor expands query request refer to corresponding ontology, building user retrieval model for retriever. Retriever implements reasoning based on some rules in semantic metadata library by ontology reasoner, and brings implicate information in semantic metadata library. Reasoner combines

primary triple and triple driven reasoning, matches them with retrieval formula, and gives answer for user. The whole system processing flow shows as following figure 8.

6. CONCLUSIONS

The conventional search engines view the web as a collection of documents and keywords. This causes poor precision and recall values. These drawbacks can be rectified with semantic web technology like ontology where data can be stored in a structured form. OWL which has more expressive power has been chosen to develop ontology. The proposed system is an effort to retrieve relevant documents in an E-learning domain. The proposed information retrieval model contains ontology management module, resource collection, semantic precision module, retrieval module. The proposed model can be helpful to create, store, evaluate, resource collection, and retrieval of documents.

7. REFERENCES

- [1]. Wang Liangshen, Hou Jie, Xie Zaiyu, wang Xiaochen, Qu Caiyue, Li Hui, Problems and Solutions of Web Search Engines, 2011, IEEE.
- [2]. Gerald Kowalski and Mark T May bury, *Information Storage and Retrieval Systems: theory and implementation*, 2nd ed., Kluwer Academic Publishers, Springer, 1997, ch.1, pp 6-10.
- [3]. W.B Frakes and R.A.Baeza-Yates, eds., *Information Retrieval: Data structures and Algorithms*, Prentice-Hall, 1992.
- [4]. F. P. Rokou et al., “Modeling web-based educational systems: process design teaching model,” *Educational Technology and Society*, Vol. 7, pp. 42-50, 2004.
- [5]. H. Adelsberger et al., “The Essen model: a step towards a standard learning process,” <http://citeseer.ist.psu.edu/515384.html>, 2003.
- [6]. J. Quemanda and B. Simon, “A use-case based model for learning resources in educational Mediators,” *Educational Technology and Society*, Vol. 6, pp.149-163, 2003.
- [7]. M. D. Merrill, “Knowledge objects and mental models,” <http://reusability.org/read>, 2003.
- [8]. Devedzic, V. Knowledge Modeling - State of the Art. *Integrated Computer-Aided Engineering*, Vol. 8, 257-281, 2001.
- [9]. C. Brewster et al., “Knowledge representation with ontologies: the present and future,” *IEEE Intelligent Systems*, Vol. 19, pp. 72-81, 2004.
- [10]. Junaidah Mohamed kassim and Mahathir Rahmany, *Introduction to semantic Search Engine*, 2009 International Conference on electrical Engineering and Informatics, August 2009, Selangor, Malaysia.
- [11]. C. Debruyne, P. De Leenheer, and R. Meersman, “Fact Type Reuse in the DOGMA Ontology framework,” in *Proc. On the Move to Meaningful Internet Systems: OTM 2009: Confederated International Conferences*, Vilamoura, Portugal, 2009, pp 1148-1163.

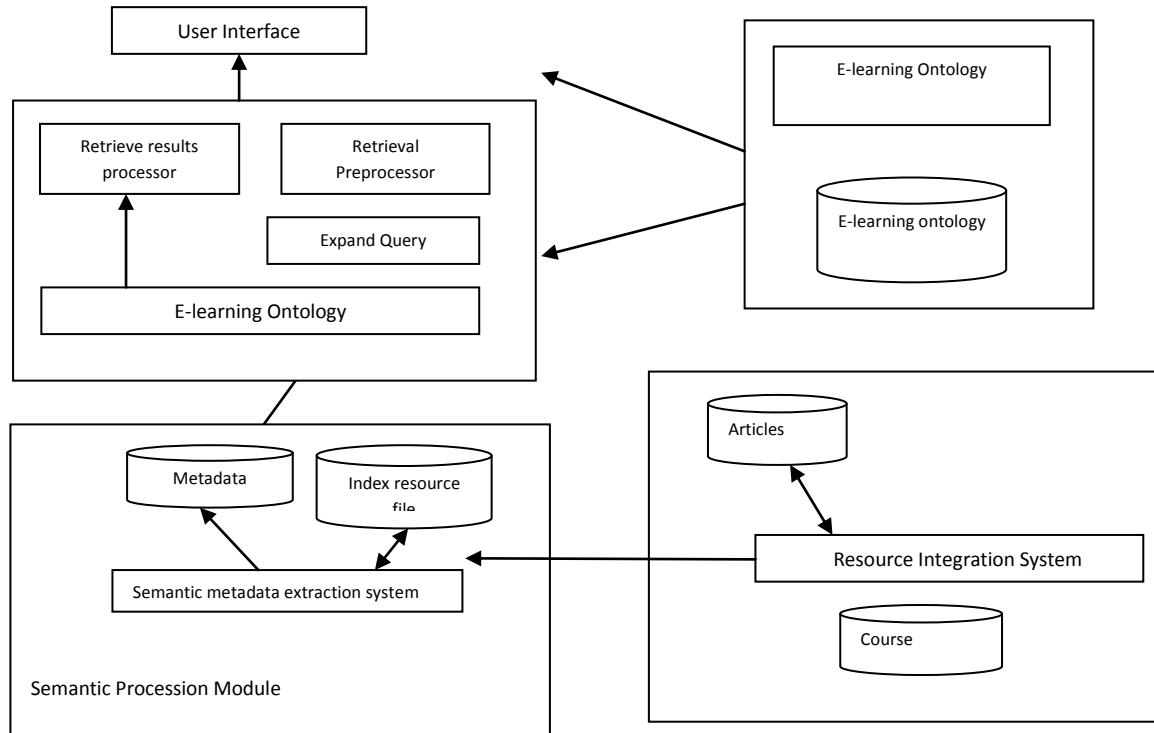


Fig 8: The E-learning information retrieval model based on ontology