A Novel Approach to Develop a Self-Organized Domain Specific Search Mechanism for Knowledge Acquisition using Ontology

B.Saleena
Research Scholar
School of Information Technology and Engg,
VIT University, Vellore, India

Dr.S.K.Srivatsa
Senior Professor
St.Joseph's College of Engineering,
Chennai, Tamilnadu, India

ABSTRACT

There is an impending need for an intelligent system to organize the large amount of information currently present in the World Wide Web (WWW). It has become humanely impossible to retrieve, sort and organize the surplus information in the WWW. Semantic web is an evolving extension of the World Wide Web in which web content is organized meaningfully in a structured format using web ontology language (OWL), thus permitting them to find, share and integrate information more easily. This paper presents a methodology to design a self organized system to retrieve information about a particular topic with the pre requisites needed to assimilate the topic that the user is interested in learning and the further readings related to the topic are also listed in the search. For this, a knowledge library for DBMS domain is created using ontology and knowledge management technologies and then a strategy is devised to group the relevant prerequisites and present it to the user in a single search. The effectiveness of our approach is demonstrated by implementing a prototype and presenting the search results of a particular keyword, its interrelated contents together with the prerequisites and further readings needed to assimilate the particular topic.

Keywords

Semantic web, Ontology, Knowledge management, E-learning

1. INTRODUCTION

E-Learning system utilizes the learning environments with modern communication mechanism available in the web to help users learn more effectively without the barriers of time and distance. In the currently available online search engines and educational materials there is no proper semantic relationship between the web resources. Finding the specific information and the pre requisites needed to assimilate the topic is difficult because of the lack of semantic description of learning resources. Users have to perform a lot of manual intervention tasks such as searching, grouping or sharing the contents across the web in order to assimilate the topic they are interested in. Therefore, acquiring and providing the suitable knowledge to meet learner's diverse learning needs is a key issue [6].

To overcome this problem, the web resources need to be efficiently organized semantically using ontology. The Semantic Web is a extension of world wide web (WWW) architecture that represent resources in standard format with formal semantics, which facilitates better searching and navigating through the web. Mechanisms that enable annotating, searching for, combining the semantic information, together with semantic language, constitute the basic ingredients of an e-learning framework [2].

Ontology is defined as a data model that represents a domain and is used to reason about the objects in that domain and the relations between them. Ontology will represent the particular meanings of terms as they apply to a domain. It is a formal representation of concepts of a knowledge domain. It is the key component needed to use the semantic web approach for searching repositories. Ontologies are a key technology emerging to facilitate Web information processing by supporting semantic structuring, annotation, indexing, and search. Ontologies allow organization of learning material around components of semantically annotated topics. This enables ontology-based course to do efficient semantic querying and navigation through the learning content [7].Current artificial intelligence in education systems has tried to incorporate semantic web resources to their design and architecture. The main idea behind this purpose is the attempt to represent information on the Web so that computers can understand and manipulate it, leading to more adaptable, personalized and intelligent learning environments [8].

Data is defined as raw facts (atomic data). Information is a collection of series of data. Knowledge is derived from information and context through analysis, which is based on thinking and understanding of a certain subject. The structure of knowledge is more complex and difficult to be represented and processed by the computer than data. Intelligence is the ability to acquire and apply knowledge and skills. At first data needs to be processed into information, then information is transformed into knowledge, and finally knowledge needs to be applied into intelligence [1].

The system is human understandable that can understand and customizes the content as per the individual learner's requirement. The system gives the complete handy information about the interrelated contents/prerequisites and further readings that the users should be aware of. It overcomes the problem of lot of manual intervention required to assimilate the topics that the users are interested in.

The purposes of this paper is to (a) construct a knowledge base for DBMS domain (b) and to devise a methodology to self organize the contents in the knowledge base and (d) report the

results of research on how the pre requisites and the further readings of the search content are provided to the user together with the results.

The paper is organized as follows. In Section 1 we present the problems of information retrieval for educational purposes from the commonly used search engines. Section 2 discusses the related research work going on in this field and Section 3 describes the architecture for self organized search mechanism and the Phases of Development and Section 4 concludes this paper and discusses our future work.

2. RELATED WORK

The evolution of E-learning systems over the years supports not only teaching and learning over the web but also help in effective knowledge management. Ontologies and Semantic Web technologies are used to integrate e-learning resources in the Web. These technologies help to manage the knowledge base in a more efficient way for quicker retrieval of relevant information. The recent research works going on in the area of semantic web technologies and e-learning are discussed below.

Farrell et al. [3] focus on assembling relevant learning objects into Web-based custom courses automatically. The system relies on a static ontology to support the assembly of learning objects. Gatevic et al. [4] propose an ontology-mapping-based framework that allows searching for learning resources using multiple ontologies. Their solution uses a mapping ontology from a Semantic Web initiative. Tane et al. [5] show how an ontology-based tool suite allows making the most of decentralized e-learning resources available.

Yanyan Li1, Mingkai Dong et al.[6] proposes a knowledge portal to effectively support e-learning, enabling flexible knowledge acquisition, knowledge refinement and maintenance, as well as knowledge retrieval and accessing based on Semantic Web technologies.

Jelena Jovanovic et al [11] have described how to use semantic Web technologies to improve the state-of-the-art in online learning environments and bridge the gap between students and teachers. The ontological framework described in their work helps formalize learning object context and shows how we can use semantic annotation to interrelate diverse learning artifacts.

Amal zouaq et al. [9] proposes a methodology to transform Textual Resources and learning Objects to concept maps and derive domain ontologies from Concept maps. A common domain model was provided to bridge the gap between e-learning and Intelligent Tutoring System.

Yufei Li et al. [10] has developed and implemented a prototype of Ontolook – a relation based search engine. The concept of cutting concept relation graph is used in the semantic search.

Existing works on Semantic web and ontology based e-learning tend to use ontologies and semantic web to organize the e-learning resources and improve the retrieval of resources in a meaningful way. The concept of retrieving pre requisites needed to understand a particular topic during a search is not concentrated in any of the papers discussed above. This paper proposes to develop a search mechanism to extract the

knowledge about the learning resources together with its pre requisites and further readings in a self organized way based on the user's query.

3. ARCHITECTURE FOR SELF ORGANIZED SEARCH MECHANISM

The idea of retrieving the pre requisites further readings required to assimilate a topic clearly is being highlighted in the paper. The effectiveness of the system is proved by developing a prototype based on a specific domain. This work is divided in to several phases like Interpretation of User's Request, Domain Ontology Creation, Creation of Knowledge base and Searching and Grouping of relevant information. The architecture diagram (Figure 1) for semantic framework prototype that is developed to retrieve the contents together with its pre requisites and further readings is shown below.
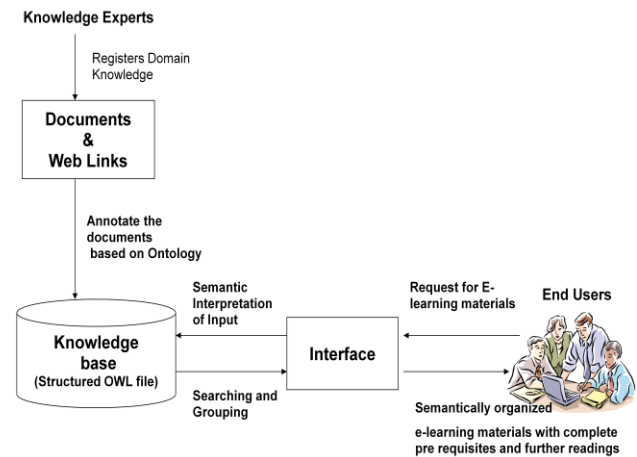
Figure 1: Architecture for Semantic Self Organized Domain Specific Search Mechanism

## 3.1 Interpretation of End User Requests

End user's request is placed in terms of keywords through a user interface. The keywords are matched with the semantic context of the topic. In other words, input is interpreted semantically according to the needs of user's requirement and is mapped into a particular class in OWL ontology. For example, if "normalization in database" is the user's keyword then the semantic interpretation of it leads into an OWL Class "Normalization" in OWL ontology [12].

3.2 Domain Ontology Creation

This phase deals with creating the domain ontology for various topics under the DBMS using Protégé. The following steps are used to create the Domain Ontology.

- Define classes and sub classes of the domain
- Arrange the classes in taxonomic hierarchy
- Establish the relationship between the classes in the domain
- Create a formal ontology framework
- Ontology testing and evaluation

The constructed ontology is stored as an OWL file (Has extension .owl). OWL Viz is the Protégé OWL plug-in which enables the class hierarchies in an OWL Ontology to be viewed and incrementally navigated, allowing comparison of the asserted and inferred class hierarchies. The domain concentrated in this paper is of pedagogic type. It deals with the different topics in DBMS. All the information under this topic are semantically related. An excerpt of the domain ontology (Figure 2) was created using Protégé 4.1.0 which is a free, open-source platform to construct domain models and knowledge-based applications with ontologies depicting the details of DBMS domain.
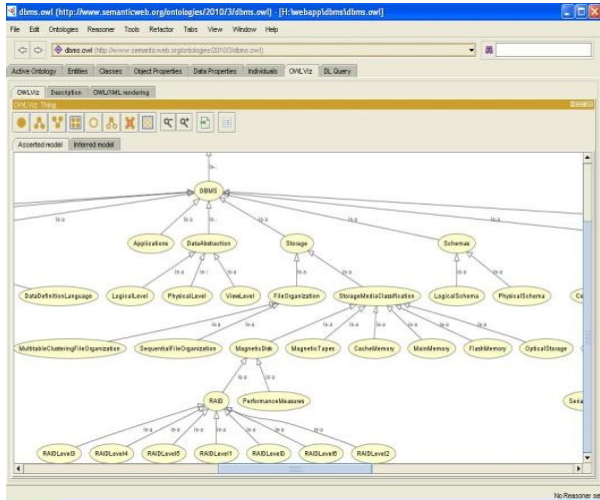
Figure 2: A Portion of DBMS Domain Ontology

3.3 Creating a Knowledge Base

The Process for creating a knowledge base (Figure 3) of interrelated resources is elucidated here. Knowledge is gathered from different sources like domain experts, structured or unstructured documents, web (HTML or XML) documents or digital libraries. The knowledge is registered in the form of text, URI's or URL contents. Registering the information in the form of URI and URL's provides a global naming scheme to information and will enhance reusability.
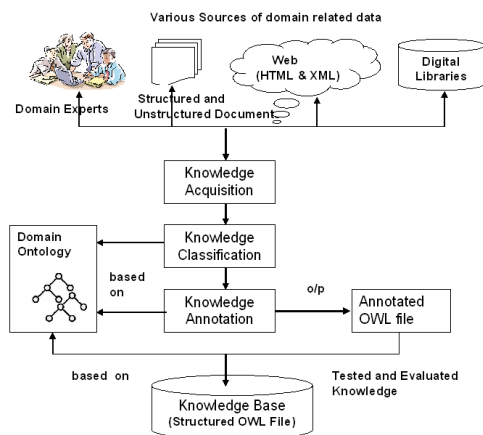
Figure 3: Process of Knowledge Acquisition

The acquired knowledge is classified and annotated based on the domain Ontology. The annotated links registered in the template are updated to the OWL file. Annotation properties in OWL were also used to add additional information (metadata) to classes, individuals and object/data type properties. The final annotated OWL file is tested and evaluated and it serves as a knowledge base for the system.

The following segment of code shows a part of annotated OWL file for the concept of Data Abstraction in the DBMS domain.

<Declaration>

    <Class IRI="#DataAbstraction"/>

</Declaration>

<AnnotationAssertion>

<AnnotationProperty

     abbreviatedIRI="rdfs:comment"/>

       <IRI>#DataAbstraction</IRI>

<Literal datatype IRI="&xsd;string">

 Three levels of data abstraction are

1. Physical level: how the data is stored physically and where it is stored in database.

2. Logical level: what information or data is stored in the database (like what the data type is or what is format of data?)

3. View level: end users work on view level. If any amendment is made it can be saved by other name.

</Literal>

</AnnotationAssertion>

3.4 Searching and Grouping of Required Information

OWL Ontology is represented in the form of hierarchical model in which each node represents a particular topic and the sub topics of the DBMS domain. The user requests for the information in terms of keywords through the user interface screen. The keywords are matched with the semantic context of the topic and all the contents related to that are retrieved. Knowledge retrieval is performed by searching the appropriate node which matches the keyword with the semantic context.

The actual definition of the topic together with its entire parent nodes that represents the complete prerequisites/interrelated contents are grouped together. The grouped information is presented to the user in a systematic way. The search results for the "Normalization" keyword are explained here. It retrieves the topics related to the key word with a sample definition and related web links. It also lists the prerequisites (Figure 4) and further readings (Figure 5) needed to assimilate the topic searched.
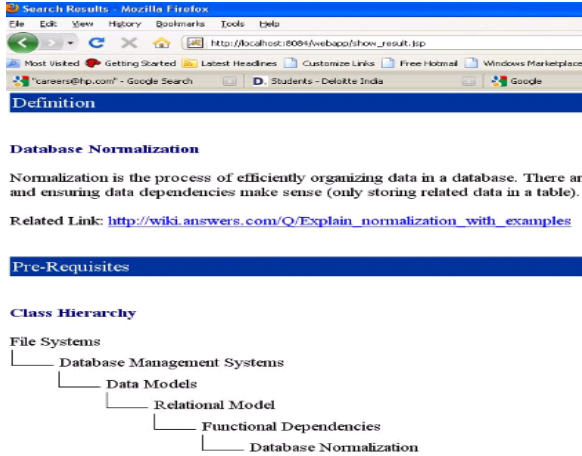
Figure 4: Sample Output of Pre requisites
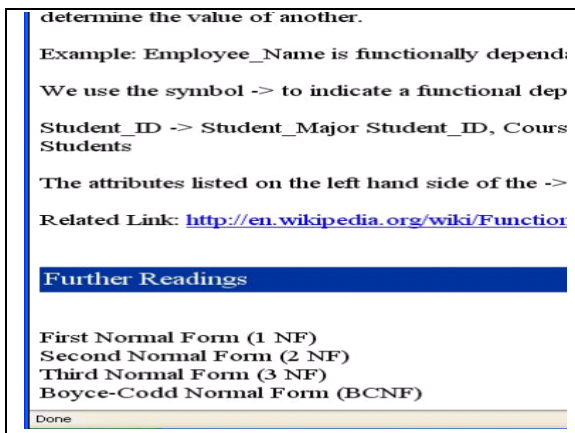
Figure 5: Sample Output of Further Readings

4. CONCLUSION & FUTURE ENHANCEMENTS

The core idea of this paper is to semantically model the heterogeneous domain specific learning resources in a coherent and meaningful way. The main objective of retrieving the interrelated contents , prerequisites and further readings needed to assimilate a particular topic as per user's interest is successfully achieved. Prerequisites and further readings are displayed in a hierarchical model so that users can understand the contents clearly. The system has been developed in Java2 with OWL API for Semantic Web. With the effective support of ontology-based knowledge retrieval, this E-Learning Framework aims to provide complete knowledge to meet learner's diverse learning needs in a user-friendly way. This E-Learning Framework has been developed only for DBMS domain. It can be extended to various domains, and a large set of functionalities can be added to the user interface screens to enhance the user friendliness. This system follows a semi automatic annotation mechanism to gather knowledge about the domain. Mechanisms to perform Automatic extraction of information can be performed.

5. REFERENCES

[1] Ruoman Zhao chuan Zhang, 2009.An Ontology–Based Knowledge Management Approach for E-Learning", IEEE International Conference on Management and Service science, pp.1-4, Sep 2009

[2] Wei Yang Xinmeng Chen, 2007 Applying Semantic web technologies to E-learning, International Conference on Wireless communications, Networking and Mobile Computing, 5415-5418

[3] R. Farrell, S.D. Liburd, and Thomas J.C.2004, Dynamic Assembly of Learning Objects, Proc. 13th Int'l World Wide Web Conf. (WWW '04), 162-169

[4] D.Gasevic and Hatala.M. 2006, Ontology Mappings to Improve Learning Resource Search," British J. Educational Technology, vol. 37, no. 3, 375-389.

[5] J. Tane, C. Schmitz, and Stumme G.2004 Semantic Resource Management for the Web: An e-Learning Application, Proc. 13th Int'l World Wide Web Conf. (WWW '04), 1-10

[6] Yanyan Li, Mingkai Dong, 2008,Towards a Knowledge Portal for E-Learning based on semantic Web, Eighth IEEE International Conference on Advanced Learning Technologies, 910-912

[7] Ling Jiang, Chengling Zhao,Haimei Wei ,2008, The development of Ontology Based course for computer Networks, IEEE International Conference on Computer science and Software Engineering ,.487-490

[8] Ig Ibert Bittencourt, Evandro Costa , Marlos Silva , Elvys Soares, 2009, A computational model for developing semantic web-based educational systems, Knowledge-Based Systems Vol 22 ,302-315

[9] Amal zouaq,Roger Nkambou,2008, Building Domain Ontologies from text for Educational Purposes,IEEE transactions on Learning Technologies ,Vol 1 ,No:1 , 49-62

[10] Yufei Li, Yuan Wang and Xiaotao huang , 2007, A Relation based Search Engine in Semantic web, IEEE Transactions on knowledge and Data Engineering vol 19 , No:2 , 273-282

[11] Jelena Jovanovic, Dragan Gasevic, Marek hatala, Christopher Brooks,2007,Using Semantic technologies to analyse learning content, IEEE computer Society, 45-53.

[12] Saleena.B., Srivatsa S.K. and Chenthil Kumar M. 2010, A Semantic approach to design an intelligent and self organized search engine for extracting information relating to educational resources, International Conference on Computer science and Information technology, Jan 2011(Proceedings to published by Springer LNCS –CCIS Series) ,In Press