

Ontology Based Semantic Information Retrieval Model for University Domain

Reshma P K¹, Lajish V L²

Department of Computer Science, University of Calicut, Kerala, India.

Abstract

Semantic information retrieval is becoming more demanding now a days as the data and information is growing day by day. This enhances the chances for getting irrelevant information also. Irrelevant information must be separated from relevant ones, especially from the world wide web or web which cannot fully utilize the information retrieval techniques. To overcome this limitation Tim Berner's Lee proposed a new web architecture known as semantic web which is an intelligent web. Ontology is the key concept used in the design of semantic web for exchange of information, use and reuse of knowledge and to capture knowledge about any domain of interest with the goal of integrating the machine understandable data on the current human-readable web. Web Ontology Language (OWL) is a semantic markup language for sharing ontologies on the web. There are different methods and tools for ontology development and SPARQL is the commonly used query language to extract information from the Ontology. In this paper, the development of a University Ontology is considered, with the help of Protégé 4.1 Editor.

Keywords: Semantic Web, Ontology, Query, Intelligent web, Protégé.

INTRODUCTION

Conventional search engines and information retrieval are widely used today and become a part of our day today life. But the main drawback is that, the search engines cannot think for the users and cannot understand the meaning of concepts and terms. The documents and the contents retrieved are based on keywords provided by the user in the query. This may not always provide the most relevant and useful contents related to the user query. Generic search engines usually provide the results based on mere keyword search and the semantics are not considered. If the user query can be processed in a semantic way, the results may be most relevant and apt.

A. Semantic Web

Semantic web refers to a web with a meaning. It describes things in a way that computers can understand. The next generation intelligent web called the semantic web offers users the ability to work on shared meaningful knowledge representations on the web. Semantic Web creates an application which will make web content meaningful to computers, that intends to support machine-processing capabilities that will automate web applications and services.

Thus semantic web can be understood by computers, does the searching, aggregating and combining the results without human interference. Software programs called agents will perform various tasks by communicating with other agents and seeking information from web resources[1]. The Semantic Web is a vision of having data on the web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications. It is a concept of how computers, people, and the web can work together more effectively than is possible now[2]. The semantic web architecture in relation to the current web can be shown as

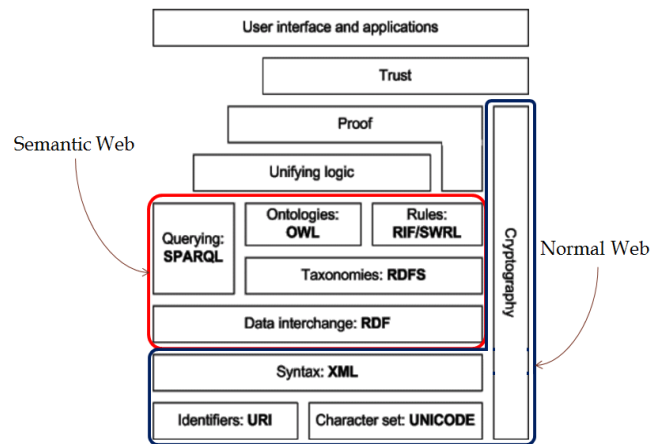


Figure 1: Semantic web architecture
 Source : <http://semanticsage.blogspot.in>

B. Ontology

Ontology provides a vocabulary comprising unambiguous definition for terms that can fundamentally serve as a formal support for communication among software agents. To compare conceptual information across two knowledge bases on web, a program must have a way to discover common meanings. Ontology is a solution to this problem. Ontology formally describes a list of terms which represent important concepts, such as classes of objects and the relationships between them in order to represent an area of knowledge[1]. Ontologies provide a formal semantics that can be employed to process and integrate information on the web. Gruber [3] describes ontology as an explicit specification of conceptualization. They define the semantics for different domains for the purpose of interactions on the web, and help in creating a knowledge base that will enable people to work on a particular domain [4]. Web Ontology Language(OWL),

recommendation from W3C, is widely used to construct a domain ontology. In order to perform useful automatic reasoning tasks on web data, it requires to go beyond the basic semantics of XML Schema and RDF Schema and there is a need of a more expressive and reasoning language which enhances the RDF with more vocabulary. W3C recommends Web Ontology Language (OWL), for developing ontologies.

Other important factors playing a key role towards the achievement of intelligent or efficient retrieval of information on web are XMLS, RDFS, URI, SPARQL etc. XMLS (XML Schema) extends the capabilities of XML where XML (Extensible Markup Language) is used for data exchange and to add meaning to data. RDFS (RDF schema) is to represent the web resource where RDF (Resource Description Framework) is for representing the knowledge resources on the web and uses the web identifier URI (Uniform Resource Identifier) to identify the resources. SPARQL (Standard Protocol for RDF Query language) is to extract information from RDF graphs for machine understandable representation. In the context of the Semantic Web, social networks like twitter assist people in finding common relationships and discussion forums for exchange of information and are playing a crucial role in information credibility and trustworthiness.

C. Protégé for Ontology Development

A number of ontology editors are available for developing an ontology such as Protégé, SWOOP, OntoEdit, Altova SemanticWorks etc. But Protégé is most widely used by researchers, professionals, programmers, and others [6]. Protégé is an open source freely available ontology editor and knowledge base framework [7]. In this paper, Section 2 highlights Ontology development in education domain. Section 3 illustrates the development of Ontology for the University of Calicut using protégé editor. It presents the query retrieval using query tab of protégé. Construction of a class hierarchy (also called taxonomy) for the University domain is illustrated with the help of code snippets and figures. Section 4 illustrates the result and gives an analysis. Section V he concludes the paper.

ARCHITECTURE OF THE PROPOSED SYSTEM

This is the top level architecture of the proposed system. A GUI is developed using python to get the query from the user. This query is then refined using the extracted domain related keywords to be used as SPARQL query. It can directly be used with the ontology designed from the university domain.

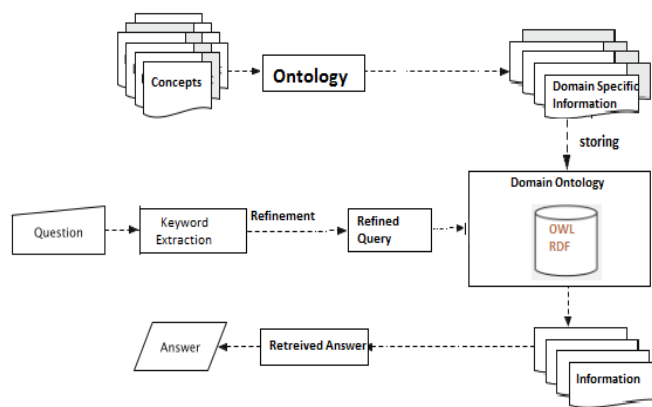


Figure 1: Architecture of the proposed system

ONTOLOGY DEVELOPMENT IN UNIVERSITY DOMAIN

The Protégé is the most widely used tool for building a particular domain ontology and has features like: customizable user interface and output file format, redefining the representational primitives, integration with other applications etc.

Building Ontologies is further divided into three steps: ontology capture, ontology coding and possible integration with existing Ontologies[8]. Ontology life cycle involves steps like specification, conceptualization, formalization, integration, implementation and maintenance[9]. Fundamental steps which are key in the development of an ontology using a tool are [10]:

Obtain domain knowledge: A deep insight into and a thorough knowledge of the respective domain is prerequisite to construction of any domain ontology.

Identify the key concepts: Concepts that represent the domain are identified and hence implemented by means of classes.

Build the taxonomy: the class hierarchy is created by creating the classes and their respective subclasses, and instances of classes.

Identify relationships between classes: Properties are used to represent relationship between classes.

Consistency checking: the constructed domain ontology must be checked for consistency using reasoners.

Implementation of ontology: involves deployment of ontology to enable machine-to-machine communication.

A domain ontology specifies the concepts, and the relationships between concepts, in a particular subject area rather than specifying only generic concepts, as found in an upper ontology such as SUMO (the Suggested Upper Merged Ontology). A domain ontology models the information known about a particular subject and therefore should closely match

the level of information found in a textbook on that subject[13].

ILLUSTRATING ONTOLOGY DEVELOPMENT WITH CASE STUDY : UNIVERSITY ONTOLOGY FOR UNIVERSITY OF CALICUT

University of Calicut, Kerala has been taken as an example for the Ontology Development using the protégé editor.

Step 1: Classes and Class hierarchy : The first step was to give the university related classes or concepts.

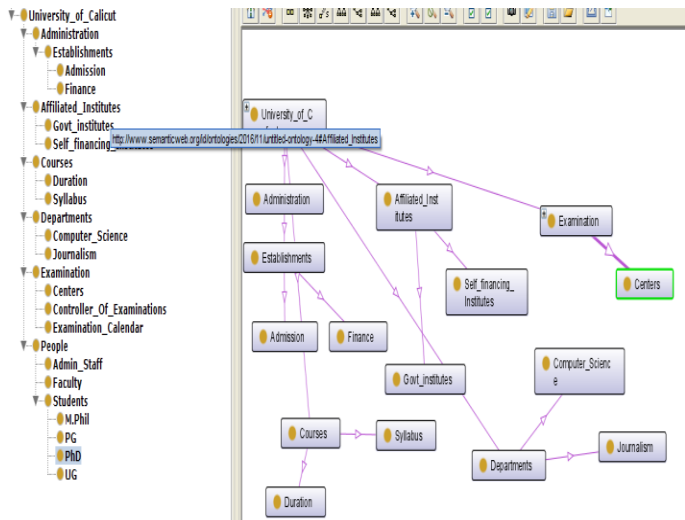


Figure 2. The Classes and Subclass of Ontology

Step 2: Object Properties of Ontology: Defining the object according to the relationship which we want to add among classes as given in the figure 3.

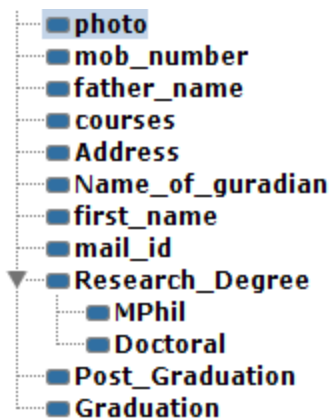


Figure 3. Properties of Objects

Step 3: The Instance of Ontology: To defining the instance first select the right class, and then create its instances for the class Use rdf: type to state its class, and one instance can be

long to many classes or many class belong to same in-stances, for example:

```
</owl: thing rdf: id=lusty100|>
<rdf: type rdf: resource=|#faculty|/>
<rdf: type rdf: resource=|#student|/> </owl: thing>
```

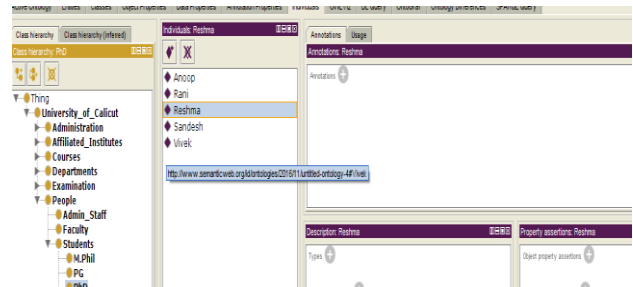


Figure 4. Individuals

Step 4: Query Retrieval Process:To build a correct ontology, reasoning is the most important part. Reasoner checks the consistency and finds the logic contradictions in the definitions. The querying can be done with DLQuery.

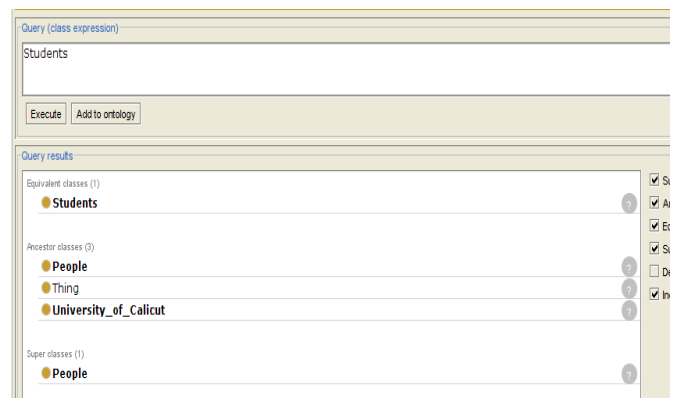


Figure 5. Testing the Ontology

RESULTS AND ANALYSIS

Precision and recall are used to measure the performance of the system. Precision is the probability that a retrieved document is relevant and Recall is the probability that a relevant document is retrieved in a search. These values are calculated as :

$$Precision = \frac{relevant \cap retrieved}{retrieved}$$

$$Recall = \frac{relevant \cap retrieved}{relevant}$$

When the proposed system was tested, a marked performance was observed for most of the queries as a result of semantic

analysis. Since there is no existing university ontology to be compared, it is compared with manual retrieval. The table 1 gives a brief description about the performance level.

Table 1. Performance of the proposed system

Sample Query	Proposed system	
	Precision	Recall
Teaching staff in Computer Science Dept, Calicut university	0.87	0.56
Teaching departments in Calicut university	0.79	0.55
Research areas in the computer science	0.83	0.45
Last date for application to PG course	0.73	0.6
Students enrolled for PhD in computer science in Calicut university	0.6	0.54

The above table depicts that the precision and recall value of our system. The more relative recall values of the system shows that this framework is more effective in retrieval of information.

CONCLUSIONS AND FUTURE WORK

The current web based on HTML describes how information is to be displayed on a web page for humans to read and it is not having the capacity of being analyzed by machines. In semantic web-based system, the concept of ontology is used to search results by contextual meaning of input query instead of keyword matching. Semantic web aims to bring the present web to a state where besides humans, machines also understand and perform various tasks which involves various complexities and challenges. Ontology is significant where Ontology development is the most fundamental step which is the objective of this paper and been illustrated with a university domain example. Various methodologies and challenges exist in the process. Ontology Web Language (OWL) is used to build ontologies. Ontology can be built in two ways: generic and domain specific. Ontology can be used in various applications such as geospatial imaging, military, medial image retrieval[14]. The Ontological approach of domain interpretation provides a shared and common forum that can be communicated across people and application systems. Ontology incorporated with SPARQL, ontoviz and ontograph automate the results of various application will facilitate image annotation and increase the efficiency and accuracy.

This paper demonstrated a university ontology with respect to the super and subclass relation. Moreover data properties and object properties are being highlighted and elaborated. Protégé is the tool used for ontology construction.

ACKNOWLEDGMENT

We would like to thank the staff and research scholars of the Dept. Computer Science, University of Calicut and faculty members of the Dept. Computer Science Mahatma Gandhi College, Iritty.

REFERENCES

- [1] Berners Lee, Godel, and Turing, "Thinking on the Web", Wiley, pp xv, pp xxv, xxvi, pp-108
- [2] Thomas B Passin, "Explorer's guide to the semantic web", Manning, pp-18
- [3] T. R. Gruber. "A translation approach to portable ontology specifications", *Knowledge Acquisition*, 5:199-220, 1993.
- [4] T.S. Dillon, E. Chang, P. Wongthongthom, "Ontology-Based Software Engineering—Software Engineering 2.0", 19th Australian Conference on Software Engineering
- [5] Jorge Cardoso, "The Semantic Web Vision: Where are We?" *IEEE Intelligent Systems*, September/October 2007, pp.22-26, 2007.
- [6] Protégé, <http://protege.stanford.edu>
- [7] John Davies, Rudi Studer, Paul Warren, "Semantic Web Technologies", trends and research in ontology-based systems, Wiley, pp 22-23.
- [8] Xiaomeng Su, "Semantic enrichment for ontology mapping", pp 29.
- [9] Michael Denny, "Ontology Building: A Survey of Editing Tools" November 06, 2002
- [10] Boyce, S., & Pahl, C." Developing Domain Ontologies for Course Content. *Educational Technology & Society*", 10 (3), 275-288.2007
- [11] Kumar Malik, Nupur Praksh, S.A.M Rizvi .Developing an University Ontology in Education Domain using Protégé for Semantic Web , 2010
- [12] Kumar Malik, Nupur Praksh, S.A.M Rizvi .Developing an University Ontology in Education Domain using Protégé for Semantic Web , 2010
- [13] T.Kanimozhi et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 2014, 763-769
- [14] Toby Segaran, Colin Evans & Jamie Taylor "Programming the Semantic Web", O'Reilly , March 2014
- [15] Dean Allemang, Jim Hendler, "Semantic Web for the Working Ontologist, Morgan Kaufmann, 2013