An Ontology Based Approach for Information Retrieval from Education Domain

Naveed Nazar Ali and Syed Zeeshan Arshad
SZABIST
Karachi, Pakistan.

Abstract: The search carried out by search engines to find any information on web is based on keywords and generally, there is no semantics involved. Therefore, there is a requirement of a system in which computer can perform this time-consuming work of searching the relevant information. The main objective of Semantic Web is to hand over most of the information to software agents. In this study, we designed the strong model of ontology using OWL-2. The domain area chosen for this study is Education-Domain. Initial study is based on understanding semantic web and its standards. Then the different methods of developing ontology were explored. Ontology is designed using one of those methods. This ontology is use for the development of Semantic Web using JENA Semantic web framework.

Keywords: SEMANTIC WEB, ONTOLOGY, OWL

1. INTRODUCTION

This independent study focuses on the development of "Education Domain Ontology" and then use this ontology for the retrieval of information from semantic web based portal.

1.1 Literature Review
There has been a significant research work carried out in the field of web technology. Moreover, this research work is now moving from search engine and information retrieval to the field of Semantic Web.

The work done by [1] discusses about the information retrieval technology. This paper addresses the issue of compatibility between semantic web and current web indexing and retrieval algorithms. In addition, to explore these issues they implemented three prototype systems and based on these prototypes, a framework is presented for the support of both inference-driven and search-driven processing.

Authors in [2] talked about the application of semantic web technologies in the field of computer-based education. It talked about interoperability between the current web and education environment to support the need of current e-learning education system. It also presented the case study, which uses ontology for this interoperation.

Nicola and Serigo[3], describe the challenge of accomplishing the "AAAL: Anytime, Anywhere, Anybody Learning" educational system using the Semantic Web. It described IWT as an effective solution, which are being used in many Italian educational institutes and enterprises.

Yang, Chen, & Shao, [6], discuss the retrieval of contents based on semantic web. In this paper, authors use the ontology of e-learning environment. Using this constructed ontology, retrieval system can find for relevant information.

The research work in the field of developing ontology for education is well described by [4]. In the first part, paper focuses on the development of "O4E Ontology", which is based on the development of ontology by gathering information about domain under consideration. In the second part, "O4E Web Portal" is developed where the created ontology can be published along with this people can share knowledge and contribute in the field of ontology.

2. BACKGROUND

The World Wide Web is considered as one of the biggest invention in the field of science and technology after computers. It has brought information from all around the world under one roof, which is easily accessible to everyone. We use search engines like google, yahoo, bing etc. to search for this information. However, information retrieved by this search engine is generally raw information, which needs the interaction of human being to understand it. Semantic Web is the web technology, which can perform this time-consuming work of finding, collecting and presenting the information.

2.1 Semantic Web
Semantic web is not the creation of new Web, but actually, it is the creation of meaningful web from the existing one. The word "meaningful web" means the web that can be understood by computer or software agents. The role of software agents here is to gather the data, infer it and then present most relevant information more efficiently [5].

2.2 Web Ontology Language (OWL)
OWL is a language for describing web ontology. It is the extension of existing web standard such as XML, RDF etc. The first recommendation of OWL was released in 2004 and its second recommendation by W3C was released in 2009. "OWL 2 ontologies is used with the information written in RDF, and OWL 2 ontologies themselves are mainly exchanged as RDF documents [13]".

2.3 OWL Types
There are three different type of OWL based on its application and implementation [7]. OWL Full, OWL DL and OWL Lite.

2.4 OWL 2 Structure:

In OWL 2, classes and property expressions used to construct class expressions and complex concepts. "OWL 2 supports various ways of describing classes: class identification, the intersection and union of two or more class descriptions, the complement of a class description, property restrictions, and the enumeration of individuals that form class instances [13]".
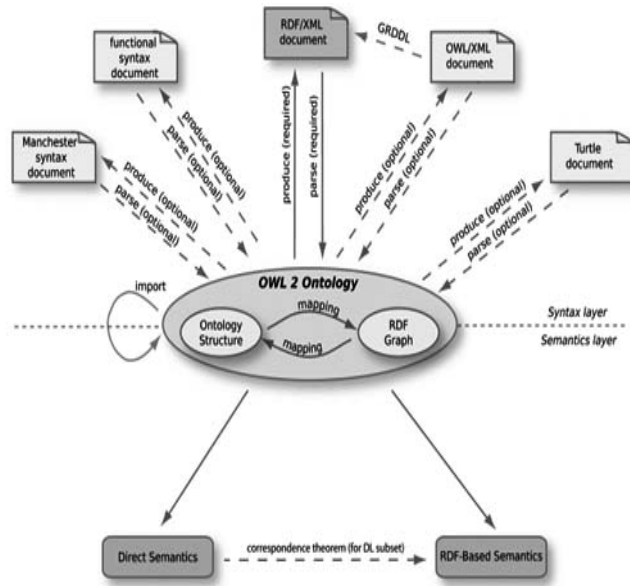

Figure 1: Structure of OWL-2

3. BUILDING AN ONTOLOGY

There have been different approaches or methods by which ontology can be developed. Some of the earlier work in this domain includes [8, 9, 10 and 11]. From the analysis of all these approaches, we can conclude that these methodologies were not mature enough, and significant efforts were required for unifying all these methods. A more recommended methodology with the name "Menthontology" was described in [15]. This method presented guidelines for building ontology from the beginning or by re-engineering process, although this method was mature but it was too complex. The approach presented by Dershing Luo in [12] was based on ERD. This approach is use to build ontology for education domain.

3.1 Entity Relationship Model:

"In designing ontology, there are five different ontology-engineering approaches as follows: inspiration, induction, deduction, synthesis and collaboration approaches." [12]

In the design of any Information System, data modeling is serve as the building block. The real-world objects can be easily mapped to this IS model using ERD (Rob & Coronel, 2000). There are three data models define by American National Standards Institute (ANSI) based on their degree of abstraction: internal, external and conceptual. In this research we are going to focus only on the conceptual model. Some useful semantic concepts are entity, property, relationship, and subtype.

3.2 Consideration to Education domain:

The ERD model given below describes different educational domain entities, their attributes and the relationship between different entities. The university data model stores details about university students, courses, the semester, faculty, publication of research paper and books, project and thesis supervisor etc. A student took a particular course, and what degree program each student is enrolled.

3.3   Identification of Entities, Relationship and Attributes:

After taking this analysis, we derived following entities and their relation:

3.3.1 Entities

Following are some of the identified entities which are identifies Program Offered, Department, Student, Education, Semester etc.

3.3.2 Relationship

Relationship between different entities are identified .Following are some of the identified relationships Student hasregister Program Offered, Student hasEducation Education, Student graduateStudent Project Supervisor etc.

3.3.3 Attributes

Attributes define characteristics of the entities. Following are some of the identified attributes which are Program Offered (programName, programCode), Department(departmentName, departmentCode), Student (studentName, studentRegisteration and etc.

3.4 Ontology Editor: Protégé 4.0

Protégé is an open source, ontology editor tool, which is chosen for this project. The reason for selecting this tool for ontology building is its support for OWL–2 and it can easily integrate with JENA for creation of semantic web [13].

3.5 Ontology Reasoner: Pellet 1.5

"A reasoner is important in ontology development due to its ability to infer logic from existing entities with consistency checking and classification for minor assumption [17]".

3.6 Create and define the Classes

Protégé creates two files:
- The .prj file contains control information about the Protégé session.

- The .owl file that contains XML describing in the ontology.
- The XML file begins with a standard set of header information.

This definition begins with the definition of the namespaces that define both "University" ontology as well as RDF, OWL and other elements of the languages themselves. The tag specifies that this document is, in fact an ontology [5].

3.6.1 SuperClass and SubClass

There are several possible ways to creating class hierarchy. Subclasses are inherited from SuperClass and have all the properties of SuperClass. In the figure given below class "Thing" is the SuperClass and it has many SubClasses. One of the SubClass of "Thing" is Publication. The Publication class has further SubClasses, BookPublication and PaperPublication. Publication is the SuperClasses of BookPublication and PaperPublication.
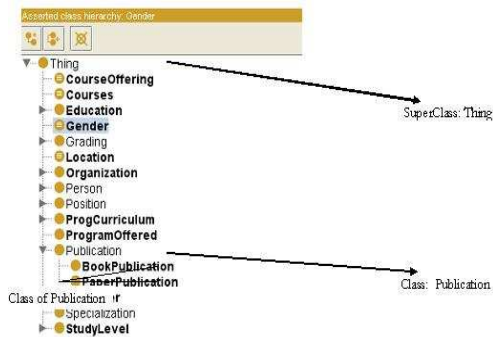

Figure 2: Class hierarchy of University Domain

3.6.2 Define the properties of the classes
Relationships are represented in Protégé using OWL Properties. There are two main types of properties in OWL, Data-type properties and Object properties.

3.6.3 Inverse Properties
"Each object property may have corresponding inverse property. If an individual 'A' is linked with individual 'B' by some property then its inverse property will link individual 'B' to individual 'A' [5]".
Example: The property hasTeach and its inverse property hasTeachBy. It describe that Instructor has hasTeach some Course and Course Offered must be TeachBy some Instructor.

3.7 OWL Object Property Characteristics
Following are the Object property characteristics, which can be defined using Protégé:

3.7.1 Functional Properties
Functional Properties states: "If a property has functional, for a particular individual, there can be at most one individual that has related to the individual through the

property. Functional properties are also known as single valued properties and features [13]".

Example: Consider the example of a functional property hasDept. If we say program Telecommunication hasDept Electronics Engineering, and we also say that Telecommunication hasDept Engineering Sciences. Then from these two relations we can infer that individual Electronics Engineering and Engineering Sciences are same. If they are not same, this can lead to inconsistency.

3.7.2 Inverse Functional Properties
Inverse Functional property states: "If a property is inverse functional then it means that the inverse property is functional. For a given individual, there can be at most one individual related to that individual through the property [13]".
Example: From the previous example, if we say hasDept is functional then the property hasProgram is inverse functional.

3.7.3 Transitive Properties
Transitive property states: "If a property P is transitive, and the property relates individual 'A' to individual 'B', and individual 'B' to individual 'C', then we can infer that individual "A" has related to individual "C" through property P [13]".
Example: If a 'StudentA' hasFriend of 'StudentB' and 'StudentB' hasFriend of 'StudentC' then 'StudentA' hasFriend of 'StudentC' also.

3.7.4 Symmetric Properties
Symmetric Property states: "If a property P is symmetric, and the property relates individual "A" to individual "B" then individual "B" is also related to individual "A" through property P [13]".
Example: If an individual student 'StudentA' is related to 'StudentB' via property hasClassFellow, then 'StudentB' is also related to 'StudentA' using that property. It is like saying that if A is a class fellow of B, then B is also class fellow of A.

3.7.5 Antisymmetric Properties
Antisymmetric property states: "A property P is said to be antisymmetric, if that property relates individual 'A' to individual 'B' then individual 'B' cannot be related to individual 'A' via that property P [13]".
Example: If an individual 'StudentA' is related to an individual 'Semester-FALL-10' via property 'hasEroll'. Then the individual 'Semester-FALL-10' cannot relate to 'StudentA' via hasEnroll property. It is like saying that a student can enroll in a semester but a semester cannot be enroll in a student.

3.7.6 Reflexive Properties
Reflexive property states: "A property is said to be reflexive when the property must relate individual 'A' to itself [13]".

Example: Here we can see that 'hasknows' is reflexive property. If an 'InstructorA' 'hasknows' to 'InstructorB', then 'InstructorA' hasknows by himself also.

3.7.7 Irreflexive Properties

"The Irreflexive property relates an individual 'A' to an individual 'B', where individual 'A' and 'B' are not the same [13]".
Example: Consider the property 'hasProgram'; an individual 'Engineering' can be relates to an individual 'Mechatronics' along the property 'hasProgram'. But 'Engineering' cannot be 'hasProgram' of itself.

3.7.8 Property Restrictions

Property restrictions are use to restrict the member of a particular class. For example, the property bookPubDate is restricting to datatype datetime as it will contain publication date.
Property restrictions fall into three main categories:

- Quantifier Restrictions
- Cardinality Restrictions
- hasValue Restrictions.

4. SEMANTIC WEB PROGRAMMING FRAMEWORK

This chapter covers the semantic web framework purpose and its operations. We integrate the university ontology in semantic web interface, which are based on the Jena Semantic Web Framework. We focus on the framework operations without the complexity of domain applications. It gives the some solutions to various programming challenges using a Semantic Web framework, such as customization, status, multithreading, and multiuser interactions.

4.1 The Semantic Web Framing

The Semantic Web is about the data, some useful semantic data. Application properly programmed to interact with the formal constructs of the Semantic Web. Framework offer consistent processing methods and concepts to simplify writing and debugging code. Consistency requires for common semantics and syntax throughout processing methods, attributes, and parameters. Semantic Web Programming languages used by the frameworks, there is no any major languages specifically designed for the Semantic Web. Other programming languages have their own views and capabilities. In semantic web requires a translation between Semantic Web data constructs and Semantic Web processing frameworks. Frameworks focus on object-oriented behaviors and it translate Semantic Web statements, classes, and such data items into programming-related classes, objects, methods, and attributes of the given programming language [14]. We adapted the JENA framework that allows substitutions and customization for optimum results in education domain.

4.2 Requirements for Semantic Web Framework:

As discussed in [14], the basic Semantic Web framework requirements are support for OWL, RDF, RDFS, and SPARQL, efficient and scalable storage of OWL, support for inference engine and easy to publish on different platforms like (LDAP repositories, SQL Database).

4.3 The Jena Semantic Web Framework

Jena is an open source semantic web framework implemented in the Java programming language, and was developed by HP labs. Programming artifacts like classes, object, function and attributes can be created from the construct of the Semantic Web using Java engine [16].

4.4 Building a Semantic web

We build semantic web with define tools and framework.
- Jena semantic web framework
- JSP for web programming language
- SPARQL for querying the data
- MySql for storing the owl
- Apache Tomcat 6.0 for Web server

4.5 Create model in database

MySQL database is used to store and retrieve ontology. For this purpose, we create an empty database with the name of universitydb in UTF-8 to develop the interface. Persistent stores are database backed Jena models. We include Jena libs and the jdbc in our classpath.

```java
public static boolean
createModel(String modelName) {
try {
IDBConnection dbcon = initDBStore();
ModelMaker maker = initModel(dbcon);

maker.createModel(modelName, true);
System.out.println("Model Created :" +
" [" + modelName + "]");

maker.close();
dbcon.close();
return true;
} catch (Exception e) {
System.out.println("Model Creation
Failed : " + e.getMessage());
}
```

This method will create a Jena model in universitydb database. This code illustrates a connection to a MySQL database using JDBC. The method declares the necessary database access information including the host address, login information, and database name. We create a database ModelMaker object through the createModel() method. The three-step process applied to obtain a ModelMaker object; create a basic model, in this case a database-backed model; and then advance the basic model to an ontology-based model. We do not need to create any relational tables and columns.

4.6 Populate owl schema into database

After creating the model, we can import an OWL file into Jena persistent store. It should be able to load data into persistent store. We can load in OWL along with RDF. We used our designed ontology to load into database with their classes, property.

4.7 Mapping Relational data to Ontology Individuals:

It is needed to convert or replicate relational data from ontology based information system. We have already design ERD of university. Then we populate data from MsSQL to mySQL and it has to be replicate into ontology-based system that works with real data.

The mapping with SQL query to RDF/OWL template. Relational data are then sent to an ontology model. The function has been implemented in Java using Jena library for ontology manipulation and MsSQL database for replication. We use relational database using JTDS connector.

4.8 Interrogating Semantic Web Data

We have populated Jena model with Semantic Web data; now we can interrogate the data. There are three basic approaches for interrogating: search, navigate, and query.
Search: A string match is performed without any semantics.
Navigate: It uses the path given by relationship to find the data.
Query: It uses a formal query language known as SPARQL.

4.8.1 Querying semantic web data

SPARQL is the W3C recommended query language use to interrogate data based on the notion of RDF. It is a simplified and effective way to query RDF. "The triple patterns in a query do not necessarily map to well-formed OWL-DL constructs. Because the way Jena stores data, we can use SPARQL to pull our data [19]".

Benefits of using SPARQL over SQL include; provision of distributed data solution, can realize data in OOP fashion, can build agents that can connect to data source, find triple relationship easier to understand than RDBMS.
SPARQL is not the direct replacement of Standard Query Language; it is an additional method of querying. "It is best in terms of merging newer and older techniques to produce a hybrid system that can handle and handled by older legacy systems and the current and future systems [18]".

4.8.2 Specification in SPARQL

The specification part of SPARQL defines the syntax that will be use by SPARQL query language for RDF. There are various data sources across which SPARQL can defines queries, whether the data stored is RDF or it can be viewed via middleware. "SPARQL contains capabilities for querying required and optional graph patterns. The results of SPARQL queries can be results sets or RDF graphs [18]".

The following terms defined in RDF Concepts and Syntax used in SPARQL:
1) Literal
2) lexical form
3) IRI
4) typed literal
5) language tag
6) plain literal
7) datatype IRI (corresponds to the Concepts and Abstract Syntax term in datatype URI)
8) blank node

4.8.3 Define Query

The example given below shows a SPARQL query to find the student who enroll in particular semester from the education data graph. The query consists of two parts: the SELECT clause identifies the variables to appear in the query results, and the WHERE clause provides the basic graph pattern to match against the data graph. The basic graph pattern in this example consists of a single triple pattern with a single variable (?s) subject in the object position and the defined Type property hasEnroll which can gives the result relating to subject and their objects . where "?s" is defined as subject stud: < http://www.semanticweb.org/ontologies/2010/1/university. owl#"> default Namespace used in ontology and Type property hasEnroll and the object "?o".

Figure 3: Query Builder for SPARQL

4.9 Reasoning Data across Semantic Web

To understand the full power of semantics, one of the important aspect is the use of reasoners across the data. Reasoners can be integrated in JENA framwork by several ways. "The Jena inference subsystem is designed to allow a range of inference engines or reasoners to be plugged into Jena. Jena inference engines are used to derive additional RDF assertions, axioms and rules associated with the reasoner [16]".

4.10 Ontology Alignment

Two different ontology can be mapped and aligned to create a single ontology using this Ontology alignment. Suppose we have two different ontology of different educational institute and we would like to combine these two ontology so that we can query the data and retrieve some semantic data. In order to align these two different names, we need to align the semantics. We create a new

---

model to hold the ontologies and then populate them.Then we require additional statements to align the two ontologies. These additions bridge the two ontologies in key areas, allowing our query to include all persons instead of just those from one or the other ontology.

A reasoner, supplies the additional inferred statements or entailments. We used Pallet reasoned to bind the schema. The Pellet reasoner implements the same interface as the Jena reasoner. This allows a program, to easily use different reasoners for different purposes.

4.11   Result and Outcome

The search based on querying relational data or keywords does not have any semantics involved in it. It performs the search on principle of reverse indexing. Whereas results obtained from the query performed on Semantic Web data using SPARQL, it is observed that the result obtained were more relevant.

4.12   Limitation

Following are the limitation or challenges face for extending semantic web

1. There is no autonomous method for creating ontology for each domain. Thus, Retrieval System based on ontology requires significant efforts for the development of ontology for each domain.

2. The need of standardized method or form of ontology developed for each domain, to avoid the need of ontology mapping.

3. How relevant the data is retrieve from the system depends upon how good ontology is developed.

5.   CONCLUSION

From the study of semantic web, we have seen that Semantic Web is the web of the future, which leaves its impact on Information system and different Knowledge areas. In this study, we design the education domain ontology using OWL-2. Initial study was based on understanding semantic web and its standards. Then the ontology was developed using the ERD method. The developed Ontology is then integrated with JENA Semantic web framework. By querying the Semantic web using SPARQL, we see that result obtained is more relevant then the same query performed on relation database.

Although the concept of Semantic Web was presented ten years back in 2000, but it still in the phase of development. Standard and Specification are available, but availability of tools and technologies to support this standard is still lacking. There is a need of substantial efforts to make it practically viable solution so that it can be easily adopted for all practical purposes.

Future Work

This study explores the way of developing ontology for education domain. The future work in extending this study includes the development of standardized vocabulary for education domain, which can be use by different educational institute. Similar work is already done by [20], which focuses on the web ontology for e-commerce. This standardized ontology will increase the visibility of different educational details like departments; degree program offered, location, admission details etc in the latest generation of search engines.

REFERENCES

[1] Finin, T.; Mayfield, J.; Joshi, A.; Cost, R.S.; Fink, C.; Information Retrieval and the Semantic. HICSS '05. Proceedings of the 38th Annual Hawaii International Conference on System Sciences, 2005.

[2] Bittencourt, I.I.; Costa, E.; Braz, L.; Pacheco, H.; Dicheva, D.;   Supporting Interoperability between Web-based Educational Systems , 39th ASEE/IEEE Frontiers in Education Conference, 2009

[3] Nicola Capuano, Sergio Miranda and Francesco Orciuoli. IWT: Ontologies and Knowledge base,HUT/TAI Research Centre Paula Silvonen, October, 2002

[4] Darina Dicheva, Sergey Sosnovsky, Tatiana Gavrilova, Peter Brusilovsky: Ontological Web Portal for Educational Ontologies .Sep 2004.

[5] Mike Dean, John Hebeler, Matthew Fisher, and Ryan Blace. Semantic Web Programming, Wiley Publishing, Inc, 2009.

[6] Yang, S. J. H., Chen, I. Y. L., & Shao, N. W. Y. Ontology Enabled Annotation and Knowledge Management for Collaborative Learning in Virtual Learning Community, Educational Technology & Society, 2004

[7] Peter F. Patel-Schneider, Patrick Hayes, and Ian Horrocks. OWL 2 Web Ontology Language Overview, W3C OWL Working Group. 27 October 2009

[8] D.B. Lenat, R.V. Guha, Building Large Knowledge-Based Systems: Representation and Inference in the Cyc; Project, Addison-Wesley, Boston, 1990.

[9] M. Uschold, M. King, Towards a Methodology for Building Ontologies, in: IJCAI95 Workshop on Basic; Ontological Issues in Knowledge Sharing, Montreal, 1995.

[10] M. Gruninger, M.S. Fox, Methodology for the design and evaluation of ontologies, in: Workshop on Basic; Ontological Issues in Knowledge Sharing, Montreal, 1995.

[11] J. Euzenat, Corporative memory through cooperative creation of knowledge bases and hyper-documents, in: Proc. 10th Knowledge Acquisition for Knowledge-Based Systems Workshop (KAW96), Banff, 1996

[12] Dershing Luo. The Information Systems Modeling with an Ontology-Based

PACIS Pacific Asia Conference on Information system. 2005

[13] Matthew Horridge. A Practical Guide To Building OWL Ontologies Using Prot´eg´e 4and CO-ODE Tools The University Of Manchester – October 16, 2007

[14] 3kbo blog, Requirements of a Semantic Web Framework, Building Semantic Web Applications October 10th, 2008

[15] M. Fernandez-Lopez, A. Gomez-Perez, A. Pazos-Sierra, J. Pazos-Sierra, Building a chemical ontology using METHONTOLOGY and the ontology design environment, IEEE Intelligent Systems & their applications 1999.

[16] Dave Reynolds, Jena: The semantic web framework, HP Labs Semantic Web Programme, October, 2009

[17] Evren Sirin, Bijan Parsia, Bernardo Cuenca Grau, Aditya Kalyapur, Yarden Katz, Pellet: A Practical OWL-DL Reasoner, Clark & Parsia, 2007

[18] Eric Prud'hommeaux, Andy Seaborne, Hewlett-Packard Labs., SPARQL Query Language W3C OWL Working Group. 15 January 2008

[19] Evren Sirin, Bijan Parsia, SPARQL-DL: SPARQL Query for OWL-DL, Workshop on OWL: Experiences and Directions – OWLED, 2007

[20] Hepp, Martin: GoodRelations: An Ontology for Describing Products and Services Offers on the Web, Proceedings of the 16th International Conference on Knowledge Engineering and Knowledge Management (EKAW2008), Acitrezza, Italy, September 29 - October 3, 2008,