An Approach to build Ontology Library (OntoLib) For Academic e-Library

Hadeel Showket AL-Obaidy[1], Nahla El Zant El Kadhi[2], Manal El Zant[3]

[1] Engineering College: Computer Engineering Department, Ahlia University, Bahrain

[2] Business College: Management Information Systems, Ahlia University, Bahrain

[3] Laboratoire d'informatique Fondamentale de Marseille (LIF), UMR 6166, France

Abstract- *The aim of this paper is to suggest an approach to build ontology library for academic portal using Semantic Web technologies: RDF, XML, ontology, Simple Protocol and RDF Query Language (SPARQL) besides ASP.NET language. Ontologies capture the semantics of information from various sources. OntoLib enables the academic portal faculty to have their research papers published in one centralized place and make it easy for them to find the research papers that belongs to their colleges. This paper defines an OntoLib as Semantic Web components that lead to an intelligent Web. It discusses ontology and its uses in semantic web for digital libraries.*

Keywords: Ontology, Knowledge Base System, RDF, OntoLib

1. Introduction

Currently, the Web is essentially syntactic. The structure of documents or resources on the Web is well defined, however, only humans can read their contents which are inaccessible to machine processing. Machines can only browse for routine treatment or linking data between different Web pages. It means that almost all Web content is intended to be read by a human user. Therefore, it cannot be handled intelligently by computer programs. Searching for information on the Web is quite imprecise, and indeed, computer has no reliable methods to deal with semantic information. Today, the objective of the semantic web is to improve the search engine to be able to retrieve the correct information needed. It is desired to have more effective methods, best accuracy, promote sharing and reusing knowledge, and also an association of semantic metadata to documents and knowledge. To solve the problem of the actual web, in this paper we suggest an approach to build ontology library for academic portal that used Semantic Web technologies such as RDF, XML, ontology, Simple Protocol and RDF Query Language (SPARQL). This paper describes the Ontology which is a major component of Semantic Web.

2. Related Works

Yuangui Lei [12] define the semantic data quality and the semantic web portal infrastructure. They specified three main components for the infrastructure, an automated metadata extraction tool that supports the extraction of high quality metadata from heterogeneous sources, an ontology driven question answering tool which makes use of the domain specific ontology and the semantic metadata to answer questions in natural language format, and semantic search engine which enhances traditional text based searching. Their infrastructure contains a source data layer, extraction layer, a semantic data layer, a semantic service layer, and a presentation layer. Viljanen et al. [10] discussed the requirements of an ontology library system to support the different phases of an ontology life cycle and related user needs for creation, publishing, maintaining and using ontologies. Viljanen et al.[11] argue that the various ontology servers on the web should be made accessible using a common API that would provide a simple but universal methods for accessing the ontology content. As a solution, they propose the LOOS API and a metadata schema for describing the services. Siricharoen [9] introduces in their paper the realistic use of the available ontologies which are provided online. The purpose of the research is to transform ontologies to Unified Modeling Language (UML) object diagram using ontology editor "Protégé". This research shows how it works efficiently with the real case study by using ontology classes in travel/tourism domain area. They specify the need to combine classes, properties, and relationships from more than two ontologies in order to generate the object model. The paper also presents a simple methodology framework which explains the process of discovering objects.

3. Semantic Web

Tim Berners-Lee, the creator of the Web, has declared that the Semantic Web is the next evolution of the Web [1, 2] which means an intelligent Web where information is stored understandable by computers in order to provide the user really seeking. Search engines help us by answering two questions: what are the pages containing a Word? And what are the most popular pages on a specific subject? The current Web is built primarily around the identifier URI, HTTP protocol and language HTML. Semantic Web is also based on URI, HTTP and the RDF language. The current Web automation capabilities are limited because the Internet has been designed to publish unstructured documents. It is difficult nowadays to access the right information needed online. For example: If you want to find a manufacturer of doors and Windows by typing the words "gates" and "windows" using the popular search engine Google, the results were not met your request and your result will include the Mr.Bill Gates pages and Microsoft Windows pages. One of the goals of the Semantic Web is to refine search on the Internet. To do this, we should add to the existing information a metadata layer that computers can develop it. The Semantic Web relies on three additional steps. First, it adds metadata to each Web resource. Then, it certifies their authenticity, and finally fixes the HTML errors. It is found that the use of XML and Resource Description Framework (RDF) should fix this problem. The XML can be used to solve this problem of research because this language has the ability to represent the semantics of data in a structured form. When searching information, it is possible to limit also the search to the documents matching particular elements. Also ambiguous words can be illustrious by the context they appear in. For example, a user searching the term “brown” in academicals documents, could be looking for papers written by Donald Brown, or papers published at Brown University, or also papers about the brown bear. So in this case, the person can specify even he wants to find <author>Brown</author>, <university>Brown</university>, or <subject>brown</subject>. W3C therefore propose to enrich existing information metadata RDF. The RDF format allows defining metadata to specify the characteristics of the information. RDF is triplets who will associate the defined metadata by group of three. One can describe a triplet as three URIs. The current Web uses links which are pairs, e.g. association "Mr. Tim BernersLee" and"W3C". Then, the Semantic Web seeks to present information by adding a third term. For this example, "founder" " <tim bernerslee=""> <fondateur> <w3c> can see the relationship between M.TBL and the W3C organization. In this case, the computer can perfectly determine what fact logically must be attached to another.

To resume, the goal of the semantic web is to express the meaning of web data in an appropriate way for automatic reasoning. This means that the descriptive data or metadata in machine readable form is to be stored on the web and used for reasoning. The challenge of the Semantic Web is to provide a language which expresses both data and rules. Then, any system of knowledge representation rules can be exported on the Semantic Web. Rules are added to the Web to give the possibility to make inferences, and answer user questions. There are two important developments of Semantic Web technologies: XML and RDF. XML allows adding an arbitrary structure to the documents without specifying the significance of structures. RDF is used by Semantic Web and it allows machines to understand documents and semantic data [8]. Figure 1 shows the Semantic Web stack diagram from the W3C.

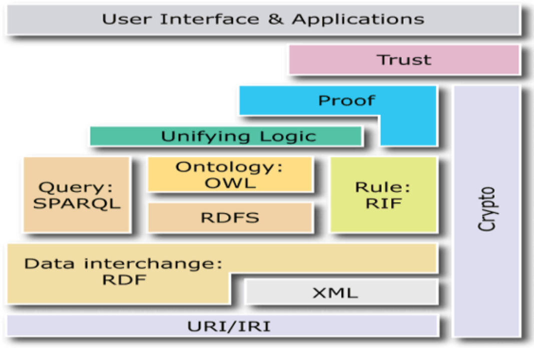

Figure 1: Semantic Web Diagram from the W3C [13].

4. Definition of Ontology

In the literature, we can find different definitions for the word ontology. Sharing knowledge and reusing them are important features of any semantic web application. The ontology is responsible for sharing the common understanding of domain between people and machine and that's important for providing the interoperability feature to the semantic web application. The ontology is defined as a set of concepts (classes) and the relationship between these concepts. Ontology, is also defined as an explicit specification of a conceptualization [6] often considered as a reusable and shareable model. Geographical ontology can be used for exploration and extraction information and also for interworking of GIS [5]. In this paper, the authors specify how the ontology is considered also as a keystone of the SMA (System Multi Agents) using a high level communication. Ontology is defined as common vocabulary for persons who need to share information in a specific domain. Ontologies are used in different domains (Geography, Biology, etc.) to share common understanding of the structure of information among people or software agents, to analyze domain knowledge, and enable reuse of domain knowledge. In our paper, we define an ontology as a description of concepts in a domain (classed, concepts) when the properties of each concept describe various features and attributes of the concept (properties, roles), and slots that describe the properties of classes and instances. According to Gruber [6], ontology is an explicit specification of conceptualization that refers to an abstract model of a domain. It represents the way we choose to express our views through words, expression of concept and elements, and relations between entities.

The Ontology or the data schema is the tool that connects people with machines and makes them communicate in a smooth way. The idea that gives the ontology its ability to learn is to have the real world conceptualization in a knowledge base and digitalize it using a readable language such as Ontology Web Language (OWL). In the ontology, every element in the real world is expressed as <owl: Thing> composed of properties and instances. Every element is drill down to classes that holds the concepts which is drill down to object properties that describes the properties of the concepts. A concept can be defined as an entity composed of three distinct elements: the terms expressing the concept into language, the meaning of the concept also called the notion or concept intension, and the objects designated by the concept. The object is composed by instances that depict the properties. The combination of them is composing the knowledge base.

5. Building Ontology

Building ontology is an iterative process that consists in different steps. The first step consists in defining the classes of the ontology, and arranging them in a taxonomic hierarchy. During this step we should define the relation between the classes and specifying the super and subclasses. The second step defines slots, describe the allowed values for them and filling in the values for slots for instances. The third step consists in creating a knowledge-base by defining individual instances, filling the slots with specific values and adding restrictions to slots. The component of ontology are, the classes which are the concepts or things that describe any object in the world (e.g. Person, table, chair, etc), the relations used to provide relations between classes (e.g. hasName, hasArticle, etc), the functions that represent the relations with one result, and the instances of the classes (e.g. An, apple, etc). The following is an example of ontology in xml showing the components of this ontology described in the previous section. Figure 2 shows the diagram of the example described above:

```
<rdf:RDF xmlns:rdf=http://www.agu.edu.bh/2010/11/5-
rdf-syntax-ns#
xmlns=http://www.agu.edu.bh/onto/schema#">

<rdf:Description rdf:about="#semantic_web">

<rdf:type rdf:resource="#article">

<imms:has_author rdf:resource="#An"/>

</rdf:Description>

<rdf:Description rdf:about="#An">

<rdf:type rdf:resource="#employee">

<imms:has_author rdf:resource="#1061"/>

</rdf:Description>
```

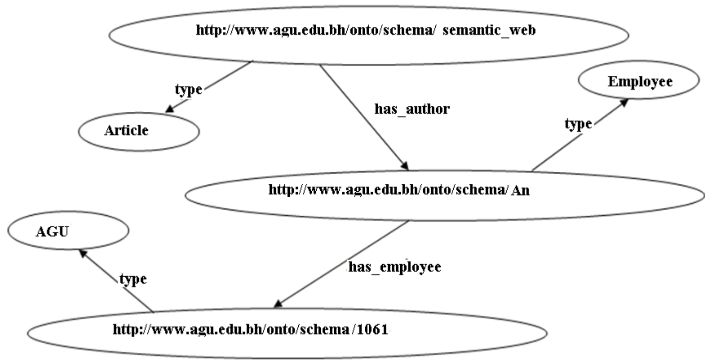

Figure (2): An ontology of Structure description

6. Resource Description framework

Resource Description Framework (RDF) is a knowledge representation language from W3C for semantic web. The RDF is an assertional logical language. It's providing the semantic web application with interoperability feature because RDF is readily for any program and facilitates data merging, no matter what schema used. Storing knowledge using this standard done by decomposing it into triples (assertions). One triple is composed of object, attribute and value. In another way it is composed of a resource (object), named property (attribute) and value for the property (value). So, any entity in the semantic web is related to another entity through a property. RDF allows structured and semi structured data to be exchanged between applications by using URI to identify each relationship between data in a triple. The triples can be expressed in three ways: tables, xml files and graphs. The easiest view is the graph view. For example: Name ('http://www.acadimic portal .edu.bh/employee/id1061", " Anne ")[7]. This example has three views table, xml and graph (see table 1).

Object	Attribute	Value
http://www.agu.edu.bh/employee/id1061	Name	Anne

```
<rdf:Description about='http://www.agu.edu.bh/employee/id1061'>
<Name>Anne</Name>
</rdf:Description>
```

Table 1: Table View of RDF example and the xml expression

7. Simple protocol and RFD Query Language

Simple Protocol and RDF Query Language (SPARQL) is a query language just like Standard Query Language (SQL) and used to perform manipulations add, update and delete the native graph stored in RDF stores. The results of the executed query using SPARQL are a set of RDF graphs, XML, JSON and HTML. The query of SPARQL should, declare Prefix using URIs, define RDF dataset and specifying the graph to be queried, identify which information should be returned as a result of the query, decide what the information to query for, and also contain the arranging query like ordering the resulted data. The following figure presents an example of SELECT query in SPARQL:

```
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
SELECT ?student?university
WHERE {?name foaf:student ?student.
?name foaf:university?university.}
```

To execute SPARQL query via HTTP, the SPARQL end point must be used for querying from RDF stores that can be accessed through Web [4].

8. The proposed System Design

This paper proposed website for academic portal library that has a semantic web technology to search in the instructors' papers. Build an OntoLib model for academic portal's library is important for a university. OntoLib enables the Academic portal faculty to have their research papers published in one centralized place and make it easy for them to find the research papers that belongs to their colleges. This research papers add to the OntoLib using a Semantic Web model and stored in a knowledge base using ontology. As mentioned above, the Semantic Web is a component of Web 3.0 that its main purpose is the intelligent addition to Web. The proposed system increases the reputation of academic portal as well as it moves the ranking of its website to upper levels. As a result, visitors of the website are expected to increase by having more knowledge available for the academic portal faculty and the visitors.

The proposed system used an OntoStudio, dotNetRDFstore, OntoMat, Ontobroker, SQL server 2005 and Visual Studio 2008 programs will be used also in this paper to achieve the proposed OntoLib and create ontologies library component. It permits academics to search for information based on semantic not syntax elements. The following are the different components for our Ontolib. The first step in designing the semantic service is to design the ontology. The resulted ontology is stored as RDF documents. Figures 3 and 4 present the use case diagram and sequence diagram of the proposed system. In this paper we present the design of the Ontology for scientific articles as it is very important part of the academic portal. The most important elements that we can search about and we should specify for any article are the paper publication (paper, thesis, etc), paper title, paper library as a Knowledge base, any related thesis, etc. Figure 5 presents a view of the ontology publication concept.

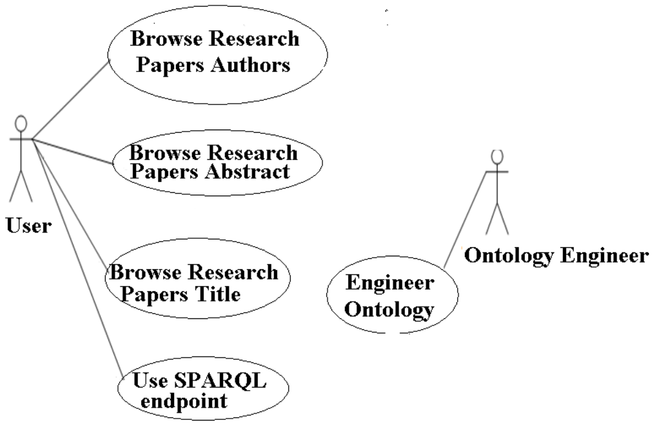

Figure (3): Use case diagram

Figure (6) visualizes the rules of the ontology. As we can see we have relations between the topic, the author and the publication. A paper is written by one or many authors. The topics and the type of the paper should be specified. Figures (7and 8) show the RDF diagram of the proposed ontology.

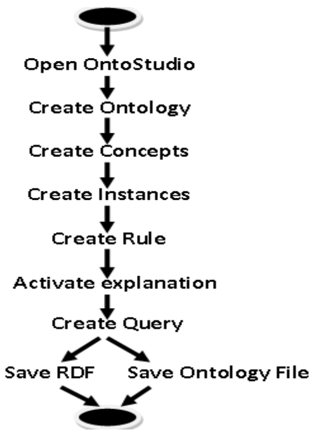

Figure (4): Activity diagram for engineering ontology.

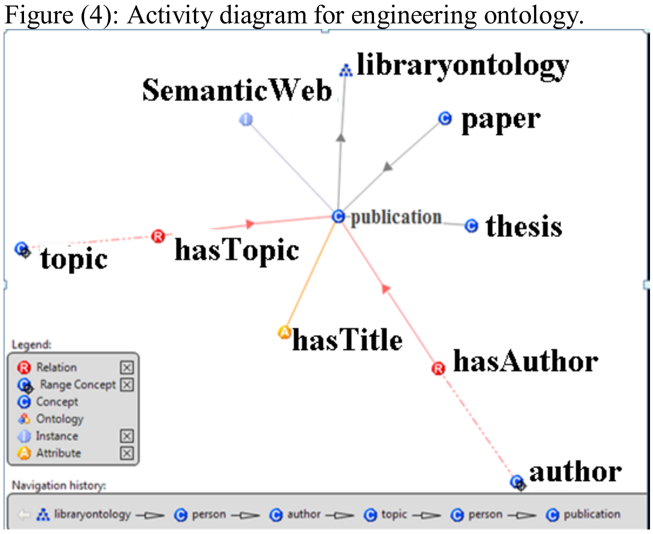

Figure (5): Diagram visualizing the ontology-publication concept.

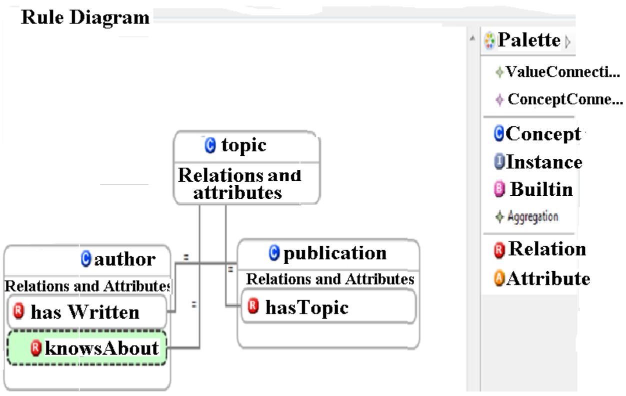

Figure (6): Rule diagram for the ontology

As mentioned before, the RDF graph is a set of triples (subject node, predicate arc, and object node). The RDF diagram of our example is shown in figure (7 and 8) and state that, for instance, the resource identified by <http:www.agu.edu.bh> is related to the resource denoted by <http:www.agu.edu.bh.LibOntology#hasTitle> via predicate <http:www.agu.edu.bh.LibOntology#instanse12475308772.

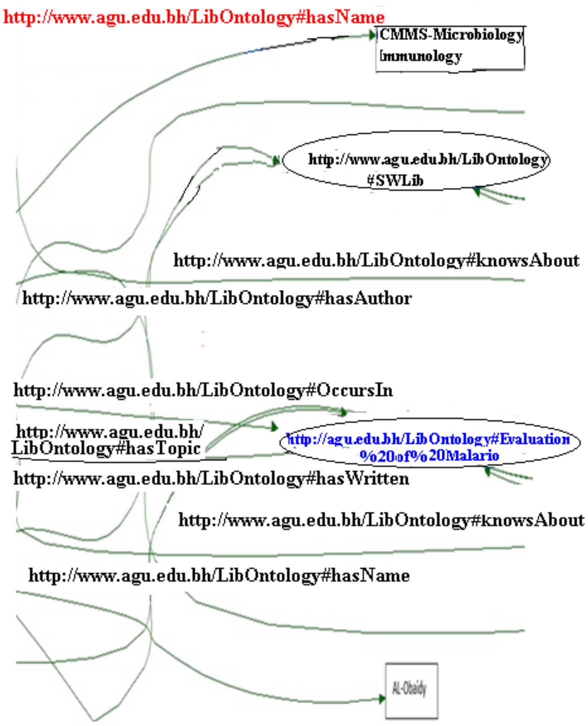

Figure (7): The RDF graph.

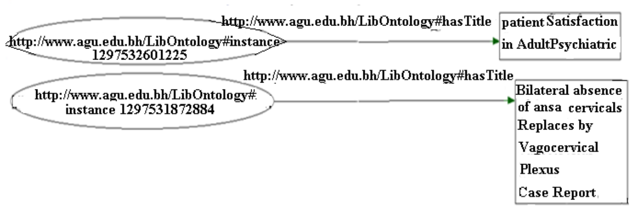

Figure (8): The RDF graph of Title concept.

9. Conclusion & Future directions

This paper introduces a proposal to develop ontology approach and show how to conduct this approach in the development of semantic academic portal. When building this ontology, it is very important to understand the knowledge base concept. This paper helps in identifying the proper ontology concepts, classes, subclasses, characterize the properties between them, shared all elements, describe the entities in those classes to describe the domain and the relationships between them and create a new knowledge from create related concepts.

This paper also presents the role of Ontologies in building Semantic Digital Libraries. Through the proposed OntoLib, the paper represents the Knowledge inside knowledge Base that can be used to build semantic infrastructure for Library 3.0. This library consists of a number of OntoLib for different domains; this can be lead to build integrated semantic libraries.

10. References

[1] Asuncion Gomez-Perez, Oscar Corcho, and Mariano Fernandez-Lopez, Ontological Engineering: With Examples from the Areas of Knowledge Management, E-commerce And the Semantic Web. First Edition (advanced Information and Knowledge Processing), 2004, ISBN: 1852335513, Springer-verlag.

[2] Berners-Lee, Tim; James Hendler and Ora Lassila, "The Semantic Web". Scientific American Magazine. http://www.sciam.com/article.cfm?id=the-semantic-web&print=true. Retrieved March 26, 2008.

[3] Berners-Lee, T., J. Hendler and O. Lassila : Semantic Web Scientific American, May 2000.

[4] Michael Science publications semantic C.Daconta, Leo J. Obrst, Kevin T.Smith,:Auide to Future of XML, Web Services and Knowledge Management, ISBN-13: 978-0471432579 2003.

[5] Nassim Dennouni, Ahmed Lehireche, Towards a spatiotemporel model for the interworking of the GIS, ICWIT 2010.

[6] T. R. Gruber. A translation approach to portable ontologies. Knowledge Acquisition, No 2, vol 5, 199-220, 1993.

[7] Nenad tojanovic, Alexander Maedche, Steffen Staab , Rudi Studer, York Sure. SEAL: a Framework for developing SEmantic PortALs, In K-CAP, pp. 155-162, 2001.What is Web 2.0: Design Patterns and Business Models for the next generation of software. O'Reilly website, 30th September 2005. O'Reilly Media Inc. Available online at: http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html [last accessed 17/01/07].

[8] Waralak Vongdoiwang Siricharoen Utilizing Ontologies Using Ontology Editor for Creating Initial Unified Modeling Language (UML) Object Model, , World Academy of Science, Engineering and Technology 64 2010.

[9] Viljanen, K., Tuominen, J., Hyv□onen, E.: Ontology libraries for production use: The Finnish ontology library service ONKI. In: Proceedings of the ESWC 2009, Heraklion, Greece, Springer.

[10] Kim Viljanen, Jouni Tuominen, Mikko Salonoja and Eero Hyvonen Noy, N.F., Shah, N.H., Whetzel, P.L., Dai, B., Dorf, M., Gri_th, N., Jonquet, C., Rubin, D.L., Storey, M.A., Chute, C.G., Musen, M.A.: Linked Open Ontology Services BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Research 37(Web Server issue).

[11] Y Lei, An Instance Mapping Ontology for the Semantic Web. In Proceedings of the Third International Conference on Knowledge Capture, Banff, Canada, 2005.

[12] H. T. Al-feel, M. Koutb, and H. Suoror, web on scope: Anew architectural model for the semantic web 1, 2008.