

# Knowledge Representation and Annotation for Semantic Web Library

Hadeel S. AL-Obaidy<sup>1</sup> and Amani Al Heela<sup>2</sup>

<sup>1</sup>Computer Engineering Department Ahlia University, Manama, Kingdom of Bahrain

<sup>2</sup>Information Technology Department Arabian Gulf University, Manama, Kingdom of Bahrain

**Abstract** - *The information contained in the World Wide Web or the web content is increasing every day. In this paper, we describe the semantic annotation process for university's library semantic web application. The step in developing the semantic web application that adds the effectiveness and reality to it is the semantic annotation for the documents published and distributed throughout the Web. The semantic annotation in this paper concerns about the research papers of the university's faculty. Semantic annotation is nothing but tagging the instances data of ontology already created with classes then map in to the related ontology classes. In this paper, two tools are going to be used for the annotation: OntoMat and OntoStudio.*

**Keywords:** Annotation, Semantic Web, Ontology, Knowledge Representation.

## 1 Introduction

The World Wide Web (WWW) is a service that needs Internet to work. It allows users to read and write information that is displayed in computers connected to the internet. What is used in the proposed system is the second and third generation of the WWW. The second generation of World Wide Web (Web 2.0) concentrates mainly on collaboration, interaction and social networking. Examples of Web 2.0 are blogs, RSS, wikis, web applications.

Tim Berners-Lee [19] has described the Semantic Web as a component of Web 3.0. The Semantic web allows for accessing information based on its meaning. A new Semantic Library model (SWLib) is considered an important need for any university. The new library website with a new design, updated information and Semantic Web model is going to change the way visitors experience the website. Currently, only few of the library websites are integrated with Semantic Web. Integrating Semantic Web with e-library is a major shift for any university's library website and allows it to be one of the leading library websites.

SWLib enables Arabian Gulf University (AGU) faculty to have their research papers published in one centralized place and makes it easy for them to find the research papers that belong to their colleges. This research papers is added to the SWLib using a Semantic Web model and through annotation process this papers are stored in an RDF store that in turn compose a knowledge base. The Semantic Web is, as mentioned above, a component of Web 3.0 which is a major intelligent addition to the Web.

This paper discusses the step in developing the Semantic Web application that adds the effectiveness to it which is the semantic annotation. This step applied for the documents published and distributed throughout the Web. Semantic annotation in this paper concerns about the research papers of the university's faculty.

## 2 Literature Review

Nicola Guarino from National Research Council and Pierdaniele Giaretta [3] from the University of Padova in their paper "Ontologies and Knowledge Bases" have clearly defined the Ontology from technological and philosophical views. They made careful analysis of Gruber's definition of ontology as a specification of a conceptualization.

Design and Implementation of Semantic Community Web Portal is the title of the paper written by Ching-Long Yeh and Chang-Gang Chen from Tatung University in Taiwan [4]. They built a semantic web portal using the RDF technology used to represent the contents of the portal. They discuss the semantic web technologies and the steps they follow to build the semantic web portal.

Another paper discusses the Extensive Markup Language (XML) and Resource Descriptive Framework (RDF) standards in depth. The paper title is "The Semantic Web - on the respective Roles of XML and RDF" and it was written by Stefan Decher, Frank van Harmelen, Jeen Broekstra, Michael Erdmann, Dieter Fensel, Ian Horrocks, Michel Klein

and Sergery Melnik [15]. These standards are used as part of this dissertation.

### 3 Web 3.0 and Semantic Web

The reporter John Markoff says in an article in The New York Times that the idea of adding meaning which is used in Web 3.0 or Semantic Web is just now emerging [6].

As Tim O'Reilly has defined Web 2.0, Nova Spivack has also defined Web 3.0 as connective intelligence that is applied through embedding intelligence in the connected data, concepts, applications and people. He rejected the view of considering Web 3.0 as Semantic Web; he includes Semantic Web is part of Web 3.0 [20].

Semantic Web has also been defined by Tim Berners-Lee, the director of World Wide Web Consortium W3C, as "a web of data that can be processed directly and indirectly by machines" and as "the extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation" [7].

The Semantic Web can be easily defined as making the machines understand the meaning of the content by applying a collection of technologies. These technologies are Resource Description Framework (RDF), RDF Schema and Web Ontology Language (OWL).

Each of these technologies will be discussed in detail in the next sections of this chapter. The building blocks of the semantic web presented by Berners Lee at the Conference XML-2000 are illustrated in the next Figure 1.

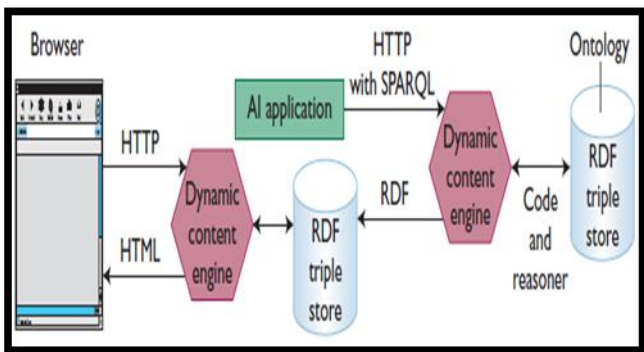


Figure 1: Semantic Web Building Blocks, Source: [8].

There are four principles which must be taken into account during developing a semantic web application:

1. All the data and entries that share the same information should be identified by Uniform Resource Identifier (URI) references.
2. The data must be provided in RDF format.

3. The URI in Hypertext Transfer Protocol (HTTP) should be linked to the RDF that belongs to it.
4. The data should be interlinked with each other.

Architecture of a sample of semantic web application (see Figure 2) has the following components:

- RDF triple store.
- Dynamic content engine.
- Artificial Intelligence (AI) application.
- Browser.

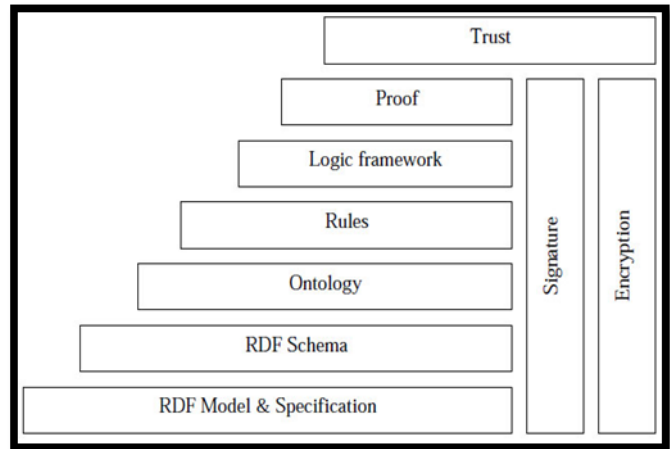


Figure 2: Architecture of a sample of Semantic Web application, Source: [17].

### 4 Annotation Process & RDF

The information contained in the World Wide Web or the web content is increasing every day. The latest survey conducted by the Internet System Consortium was on October 2010 and found that the number of hosts advertised in Domain Name Server (DNS) was 777,994,517. These hosts are the one whose responsible for serving the Web pages, one host can serve up to millions of Web pages and now imagine how many web pages with its information is available in the World Wide Web! It is a very huge number that makes it very difficult for a person to search and find the needed information from it. For that reason, Semantic Web has existed and presented to solve the problem of finding the wanted information in the World Wide Web. Its main idea is to search based on the semantics of information that use a technology to make the machines understand the information and that is obviously leading to easing and fastening the search process and overcome the problem that the World Wide Web was only provide the information to people who are the only one that understand those information.

As any process, annotation process has input and output. The input is the documents and ontology, and the output is a Resource Description Framework (RDF) document.

The first input is the documents and in this paper, the documents used as input in the annotation process are the research papers that wrote by the university's faculty.

The second input is the ontology. The ontology is the brain component of the semantic web application. It's providing the application or the machine the understanding capability. Thomas Gruber defines the ontology as “explicit specification of conceptualization” [10] while it also can be defined as the relationships that connect concepts, nodes or entities to each other.

The output is the RDF document and it is explained in details in the next paragraphs.

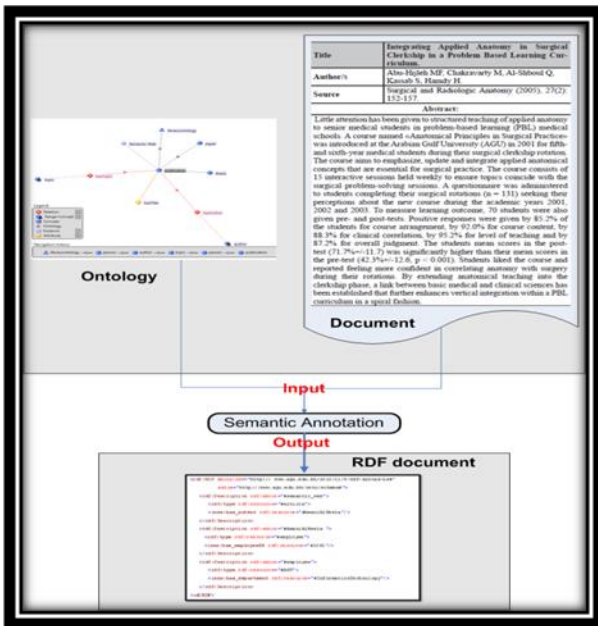


Figure 3: Diagram illustrating the annotation process

RDF is a W3C standard model used for the purpose of data interchange in the Web. 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 (3 tuples) triples. One triple is composed of object, attribute and value. In another way it composed of a resource (object), named property (attribute) and value for the property (value).

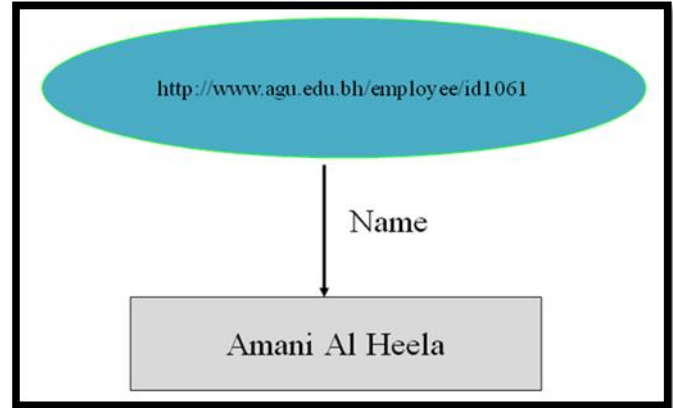
RDF allows structured and semi structured data to be exchanged between applications by using URI to identify each relationship between data in a triple.

The triples can be expressed in three ways: tables, xml files and graphs. The easiest view is the graph view. Let's take this example:

Name ('http://www.agu.edu.bh/employee/id1061', "Amani Al Heela"). This example has three views table (see Table 1), xml and graph (see Figure 4).

Table 1: Table View of RDF example

Object	Attribute	Value
<a href="http://www.agu.edu.bh/employee/id1061">http://www.agu.edu.bh/employee/id1061</a>	Name	Amani Al Heela



The xml expression of the example is:

Figure 4: Graph View of RDF example

The graph view of the example is:

```
<rdf:Description about= 'http://www.agu.edu.bh/employee/id1061\>
<Name> Amani Al Heela </Name >
</rdf:Description>
```

Simple Protocol and RDF Query Language (SPARQL) is a query language just like Standard Query Language (SQL) which is used to perform manipulations such as insert, 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 is composed o the following:

- Declaring Prefix using URIs
- Defining 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.
- The arranging query like ordering the resulted data.

This is an example of SELECT query in SPARQL:

To execute SPARQL query via HTTP, the SPARQL endpoint must be used for querying from RDF stores that can be accessed through Web.

## 5 THE PROPOSED SYSTEM DESIGN

The proposed system implemented by completing the following steps (Figure 5):

- The first step is to prepare a good design for the AGU library website that satisfies the standards.
- The web developer then converts this design to a developed website. The development language used in the proposed system is the ASP.NET language and the program used is the visual studio 2008.
- The proposed website integrated with web 2.0 applications.
- The Web 3.0 integrated to the proposed system through developing a semantic web service. The three major processes are:
  - Engineer ontology: the OntoStudio software is used for engineering the ontology of the proposed system. Ontology is the core of the proposed semantic web service that is defines the data schema for which the data will be entered.
  - Annotation: in the annotation step, OntoStudio and OntoMat are used to enrich the ontology with data. The AGU faculty papers are collected to be used in this step. This step is what this paper discusses.
  - Indexing: the RDF document that produced from the previous steps is stored in the SQL database to build RDF store.

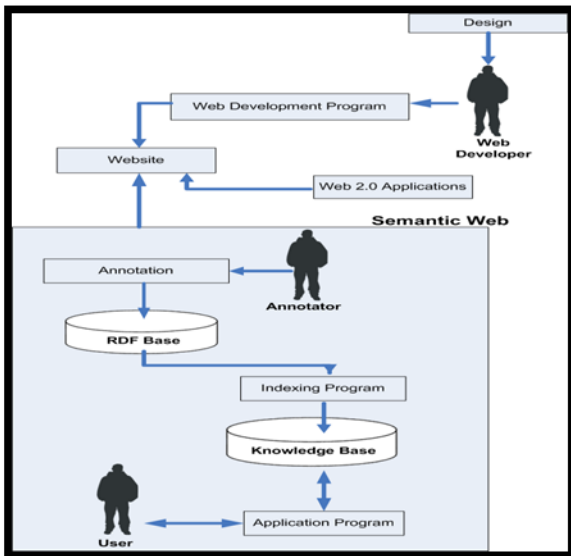


Figure 5: The conceptual architecture of the proposed website of AGU's library

The steps needed in annotation process for the proposed system are described in details in the following algorithms besides the sequence diagram in Figure 6.

### Algorithm for Annotation Process (entering paper's information) using OntoStudio

**Input:** paper information

**Output:** Knowledge Base – RDF document

Begin

Step 1: Collect the paper information: paper title, abstract and author from AGU faculty.

Step 2: Open ontology that created previously to start enriching it with papers information. The annotation is done by creating instances for the concepts.

Begin

Step 4: Open the OntoStudio software.

Loop

Step 5: Click the class that wanted to enter information to it.

Step 6: Right click the class and create new instance.

Step 7: The related attributes and relations values are entered for each instance.

End Loop

Step 8: Save the ontology with the entered information as an RDF document.

Step 9: Close OntoStudio software.

End

### Algorithm for Annotation Process (entering paper's information) using OntoMat

**Input:** paper information

**Output:** Knowledge Base – RDF document

Begin

Step 1: Open OntoMat software.

Step 2: Open ontology file.

Step 3: Open the online paper file in the specified area.

Step 4: Start annotating by creating instances for each class and drag the related information in each one.

Step 5: Save ontology file with entered data.

Step 6: Save the ontology file as RDF.

End

Step 7: Save the RDF documents into RDF store in SQL database.

Begin

Step 8: Open SQL express 2005.

Step 9: Create a database for storing the RDF documents.

Step 10: Execute dotNetRDF store manager.

Step 11: Click file then New SQL Store Manger.

Step 12: Enter the connection details: database name, username and password.

Step 13: Import the RDF document.

Step 14: Click create data store.

End

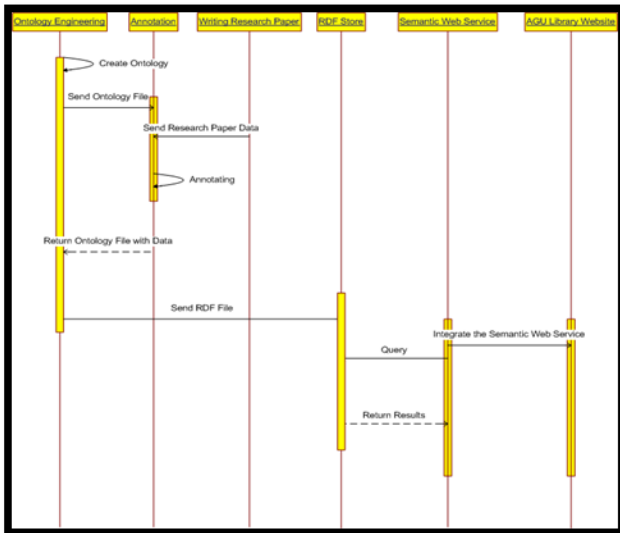


Figure 6: The sequence diagram of the proposed system

The processes used to develop the Semantic Web Service for e-library are summarized in figure 6. The annotation process this paper most concern about is shown in figure 6 after engineering ontology process; the annotating process looping to enter the research papers information and after completing it, the ontology becomes rich of data and so the knowledge base is created.

### 5.1 Annotation using OntoStudio

"The OntoStudio is an engineering environment for ontologies and for the development of semantic applications, with particular emphasis on rule-based modeling. It is the successor of OntoEdit which was distributed worldwide more than 5000 times. OntoStudio was originally developed for F-Logic but now also includes some support for OWL, RDF, and OXML. It also includes functions such as the OntoStudio Evaluator. The Evaluator is used for the implementation of rules during modeling; this procedure has been recently patented" [1].

The data of the documents (research papers) is mapped to the ontology that engineered previously in the OntoStudio. The annotation process is the process of creating new instances and entering data to it.

To feed the ontology with knowledge, the annotation step takes this role and enriches the ontology with knowledge. The following figures describe in details how the annotation process done. There are two ways of implementing the

annotation step. The first one is by using the OntoStudio software and the second is by using OntoMat. Using OntoStudio, the following figures shows the steps for annotation process.

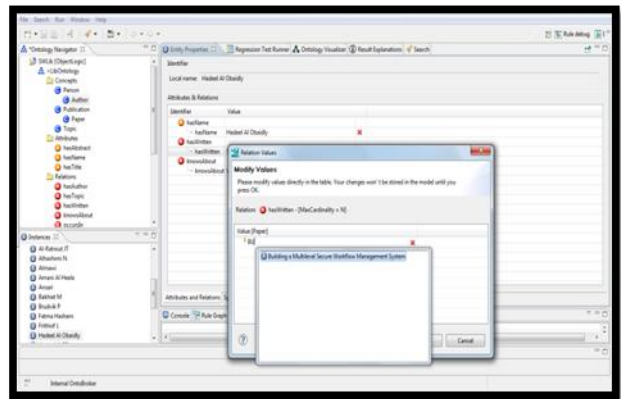


Figure 7: Annotation step – create new instance

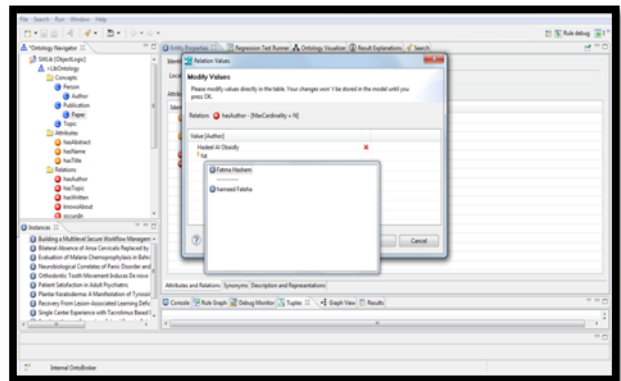


Figure 8: Annotation step – enter data for the new instance attribute

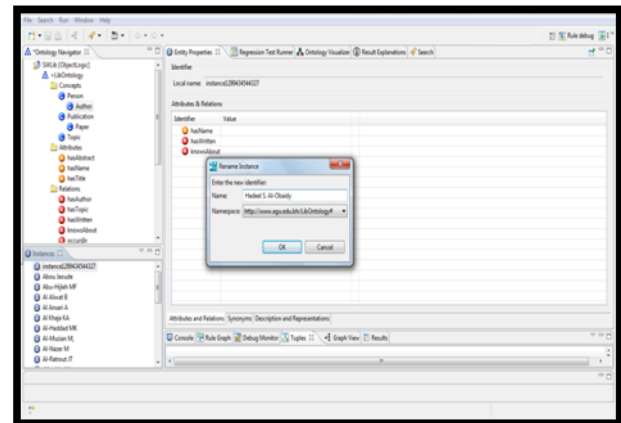


Figure 9: Annotation step – enter value for the new instance relation

### 5.2 Annotation using OntoMat

The annotation can be done by using the interactive webpage annotation tool OntoMat. It is a user-friendly and easy tool that can be used by any person. Once the OntoMat is open,

the next step is to import the ontology into OntoMat so it becomes possible to maintain the ontology and create instances, attributes and relationships. OntoMat composes of two browsers, ontology browser for viewing the ontology and instances and HTML browser that display the document that is wanted to be annotated.

The annotation process in OntoMat is just about drag and drop. Drag the part of the document and drop it to the instance of the relevant ontology's class.

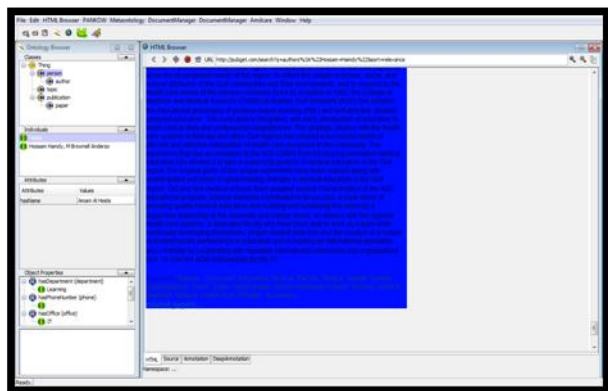


Figure 10: Annotation using OntoMat

## 6 Conclusions

The library is a very essential unit in any university; it is the unit that provides the knowledge to help the members of the university. Nowadays in the information era, the need for a website that reflects the university's library and provides access to the knowledge it holds is increasingly becoming more important. AGU is like any another global university and needs an electronic gateway to the library, which is a website. It is not an exaggeration to say that planning for developing a library website should be given the same planning and care as the library itself.

This paper supports using annotation process and RDF Semantic Web techniques for adding more value and functionality features to e-library. This can be achieved by start creating an archive of knowledge that allows the visitors to access easily. These features definitely are definitely increasing the number of visitors to the e-library and increase user satisfaction. In addition, it affects the e-library to get a higher ranking among universities which allows it to compete successfully with other the leading library websites.

This paper discussed the annotation process that is one the processes used to develop a Semantic Web Service for e-library and how important is to enrich ontology with knowledge. The input for annotation process is the documents and ontology, and the output is an RDF document that is then ready to be used by any Semantic Web Service.

### 1. FUTURE WORK

Future researches are needed to include: Developing and testing phase for the Semantic Web service for e-library. Enhancing the proposed system and adding more functionality to e-library and extending the use of Semantic Web to other services in e-library, Maximize the benefits of Semantic Web by reusing it for presenting other type of knowledge, and expand the use of library Semantic Web application to the mobile technology and develop a web application that working in WAP.

## 7 References

- [1] OntoStudio. (2011). Retrieved March 5, 2001, from <http://semanticweb.org/wiki/OntoStudio>.
- [2] ALEXANDER, B. 2006. Web 2.0: A new wave of innovation for teaching and learning. *EDUCAUSE Review*. Vol. 41, No. 2, March/April 2006, pp. 32–44. EDUCAUSE: Boulder, USA. Updated version available online at: <http://www.educause.edu/apps/er/erm06/erm0621.asp> [last accessed 14/01/2011]
- [3] Nicol Guarino, Pierdaniele Giaretta. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In *Towards Very Large Knowledge Bases*, N.J.L. Mars, Ed. Amsterdam: IOS Press; 1995.
- [4] Yeh, Ching-Long and Chen, Chang-Gang. Design and Implementation of Semantic Community Web Portal. Available at <http://www.cse.ttu.edu.tw/chingyeh/papers/DATFPortal.pdf>. [Last accessed 29/02/2011]
- [5] O'REILLY, T. 2005a. 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 17/01/2011].
- [6] J. Markoff, "Entrepreneurs See a Web Guided by Commonsense," *The New York Times*, Business, 12 Nov. 2006.
- [7] Berners-Lee, Tim; James Hendler and Ora Lassila (May 17, 2001). "The Semantic Web". *Scientific American Magazine*. <http://www.sciam.com/article.cfm?id=the-semantic-web&print=true>. [Last accessed 29/02/2011].

- [8] Berners-Lee, T., J. Hendler and O. Lassila: Semantic Web Scientific American, May 2000.
- [9] Ora Lassila & James Hendler: "Embracing 'Web 3.0'", IEEE Internet Computing 11(3):90-93, May/June 2007
- [10] A guide to Future of XML, Web Services and Knowledge Management by Michael C.Daconta, Leo J. Obrst, Kevin T.Smith, 2003
- [11] Holger Lausen, Ying Ding, Michael Stollberg, Dieter Fensel, Ruben Lara Hernandez and Sung-Kook Han, (2005), "Semantic web portals:state of the art survey" Journal of knowledge Management, Vol. 9, No. 5 (2005): 40-49.
- [12] G Küick, (2004), "Tim Berners-Lee's Semantic Web" South African Journal of Information Management, Vol.6 (1) March (2004).  
<http://www.sajim.co.za/index.php/SAJIM/article/download/297/288>. [Last accessed 06/03/2011].
- [13] Nenad Stojanovic, Alexander Maedche, Steffen Staab, Rudi Studer, York Sure. SEAL: a framework for developing SEMantic PortALs, In K-CAP, pp. 155-162, 2001
- [14] Siddharth Gupta<sup>1</sup> and Narina Thakur, (2010), "Semantic Query Optimisation with Ontology Simulation"
- [15] Stefan Decker, Frank van Harmelen, Jeen Broekstra, Michael Erdmann, Dieter Fensel, Ian Horrocks, Michel Klein, Sergey Melnik, (2000), "The Semantic Web - on the respective Roles of XML and RDF"
- [16] K Srinivas, S I Ahson, T A V Murthy, (2006), "Building a Semantic Web for Academic Networks: a conceptual architecture"
- [17] Ora Lassila and James Hendler, (2007), "Embracing Web3.0"
- [18] Leonardo Magela Cunha, (2007), "A Semantic Web Application Framework"
- [19] LÉGER, A. et al. D2.2 Successful Scenarios for Ontology-based Applications V1.0 ,2002/05/31. ,2002p. 100. Available at:<http://ontoweb.org/Members/huro/MyPublications/OntoWeb%20Deliverable%202.2/view>. [Last accessed 27/02/2011].
- [20] Definition of web 3.0. (2011). Retrieved March 1, 2011, from [http://www.webopedia.com/TERM/W/Web\\_3\\_point\\_0.html](http://www.webopedia.com/TERM/W/Web_3_point_0.html).