Applied Mathematics & Information Sciences

Annotation: An Approach for Building Semantic Web Library

Hadeel S. AL-Obaidy[1] and Amani Al Heela[2]

[1]Computer Engineering Department Ahlia University, Kingdom of Bahrain

[2] Information Technology Department Arabian Gulf University, Kingdom of Bahrain

Hadeel_showket@yahoo.com ,   roseofislam50@hotmail.com

Abstract: In this paper, we describe the semantic annotation process for developing a semantic web application to a university's library. Semantic annotation is the step that adds the effectiveness and reality to the semantic web application. It is annotating 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. Here we are going to use two tools for building annotation library, these tools are: OntoMat and OntoStudio,.

Keywords: Annotation, Semantic Web, Ontology, Knowledge Representation, Resource Definition Framework (RDF), Simple Protocol and RDF Query Language (SPARQL).

INTRODUCTION

The World Wide Web (WWW) is a service that needs Internet to work. The information distributed in the WWW is increasing every day. The Web allows users to read and write information which 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 Web (Web 2.0) concentrates mainly on collaboration, interaction and social networking. Examples of Web 2.0 are blogs, RSS, wikis, web applications. As in any field, competition is a very important factor and those who can compete, stay on and survive. In the world of universities, it is important to have a good reputation and demonstrate as much resources as possible. One useful way is to have a valuable and excellent library website is to integrated with Web 2.0 and Web 3.0[1].

Tim Berners-Lee describes the Semantic Web as a component of Web 3.0. This allows for accessing information based on its meaning. A new Semantic Library model (SWLib) is considered an important requirement for any university. The new library website containing a new design, updated information and Semantic Web model will change the way visitors experience the website. Currently, only few library websites are integrated with Semantic Web and therefore, Integrating Semantic Web with e-library will demonstrate a major shift for any university's library website by allowing it to be one of the leading library websites [2, 4]. SWLib enables faculty of the Arabian Gulf University (AGU) to have their research papers published in one centralized place. In other words, it makes it easy for them to find research

papers that belong to their colleges. These research papers are added to the SWLib using a Semantic Web model and through annotation processes these papers are stored in an RDF which composes a knowledge base. The Semantic Web is a component of Web 3.0 representing a major intelligent addition to the Web.

There are several steps for building the SWLib including: Engineer ontology, annotating, produce Resource Description Framework (RDF), create RDF store in Standard Query Language (SQL) database then develop a Semantic Web application. This paper examines the steps in developing the Semantic Web application based on annotation. These steps are applied for the documents published and distributed throughout the Web. In this paper, semantic annotation is concerned with the research papers written by the faculty of university. Semantic annotation is nothing but tagging the instances data of ontology already created with classes then map in to the related ontology classes. There are two tools are used in this paper for building Semantic annotation library: OntoMat and OntoStudio.

1. LITERATURE REVIEW

In their paper "Ontologies and Knowledge Bases", Nicola Guarino from National Research Council and Pierdaniele Giaretta [9] from the University of Padova define the Ontology from technological and philosophical views. They made careful analysis of what is Gruber's define 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 [13]. They built a semantic web portal using the RDF technology 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 Written by Stefan Decher and et al. [11] presents a work to build a semantic web. They claim that a further representation and inference layer is needed on top of the currently available layers of the WWW. To establish such a layer, they propose a general method for encoding arbitrary ontology representation languages into RDF/RDF Schema. This illustrates the extension method by applying it to a particular Ontology representation and Inference language (OIL). Holger Lausen, Ying Ding, and et al [5] conducted a study in which they integrated semantic web technologies to the web portals to increase the quality with the web portals and information consistency. They believe that semantic web technologies improve information sharing and thus it represents the next generation of web portals. Michael Erdmann, et al. [7] presents the work in ontology-based semantic annotation, which is embedded in a scenario of a knowledge portal application. Starting with seemingly good and bad manual semantic annotation, and developing an ergonomic and knowledge base-supported annotation tool. Furthermore, the annotation tool described are currently extended with mechanisms for semi-automatic information-extraction based annotation. Supporting the evolving nature of semantic content we additionally describe our idea of evolving ontologies supporting semantic annotation

2. WEB 3.0 AND SEMANTIC WEB

The reporter John Markoff says in an article published 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]. Similar to Tim O'Reilly definition of Web 2.0, Nova Spivack defines Web 3.0 as connective intelligence that is applied through embedding intelligence in the connected data, concepts, applications and people. He rejects the view of considering Web 3.0 as Semantic Web. He includes Semantic Web as part of Web 3.0 [10].

Tim Berners-Lee, the director of World Wide Web

Consortium W3C defines Semantic Web 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"[2].

However, the Semantic Web can be easily defined as making the machines understand the meaning of the content by applying collection of technologies. These technologies are Resource Description Framework (RDF), RDF Schema and Web Ontology Language (OWL)[3,4].

The building blocks of the semantic web presented by Berners Lee are illustrated in the Figure 1.
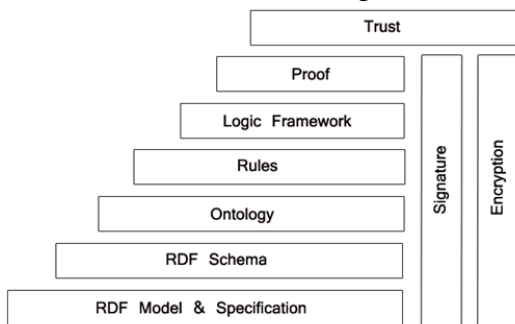

Figure 1: Semantic Web Building Blocks, Source: [2].

There are four principles must be taken into consideration in the development of 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.
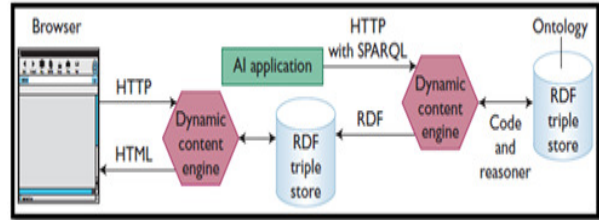4) The data should be interlinked with each other.

Figure 2: Architecture of a sample of Semantic Web application, Source: [10].

Architecture of a sample of semantic web application is illustrated in Figure 2 containing the following elements:

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

3. ANNOTATION PROCESS & RDF

The information contained in the WWW or the web content is increasing every day. A survey conducted by the Internet System Consortium on October 2010 found that the number of hosts advertised in Domain Name Server (DNS) reached to 777,994,517. These hosts are responsible for serving the Web pages, in which one host can serve up to millions of Web pages imagine how many web pages with its information is available in the World Wide Web! It is a huge number that makes it very difficult for a person to search and find the needed information from it. For this reason, Semantic Web has existed and presented to solve the problem of finding the wanted information in the World Wide Web. Its main function is to search based on the semantic of information that use a technology to make the machines understand the information. This obviously is for easing and fastening the search process and overcoming the problem that the World Wide Web was only provide the information to people who are the only one that understand this information [8].

As any process, annotation process includes input and output. The input is the documents whereas ontology

and the output is a Resource Description Framework (RDF) document.

This paper uses the input document as an input in the annotation process, in order to represent the research papers written by the university's faculty. The second input is the ontology, which is the brain component of the semantic web application. It provides the application or the machine the understanding capability. Thomas Gruber defines the ontology as "explicit specification of conceptualization" [12] while it also can be defined as the relationships that connect concepts, nodes or entities to each other. Ontology or the data schema provides the concepts with a meaning that makes the machine understand it. It is the tool that connects people with machines and makes them communicate in a smooth way. In Figure 3 the output of the RDF document is shown.
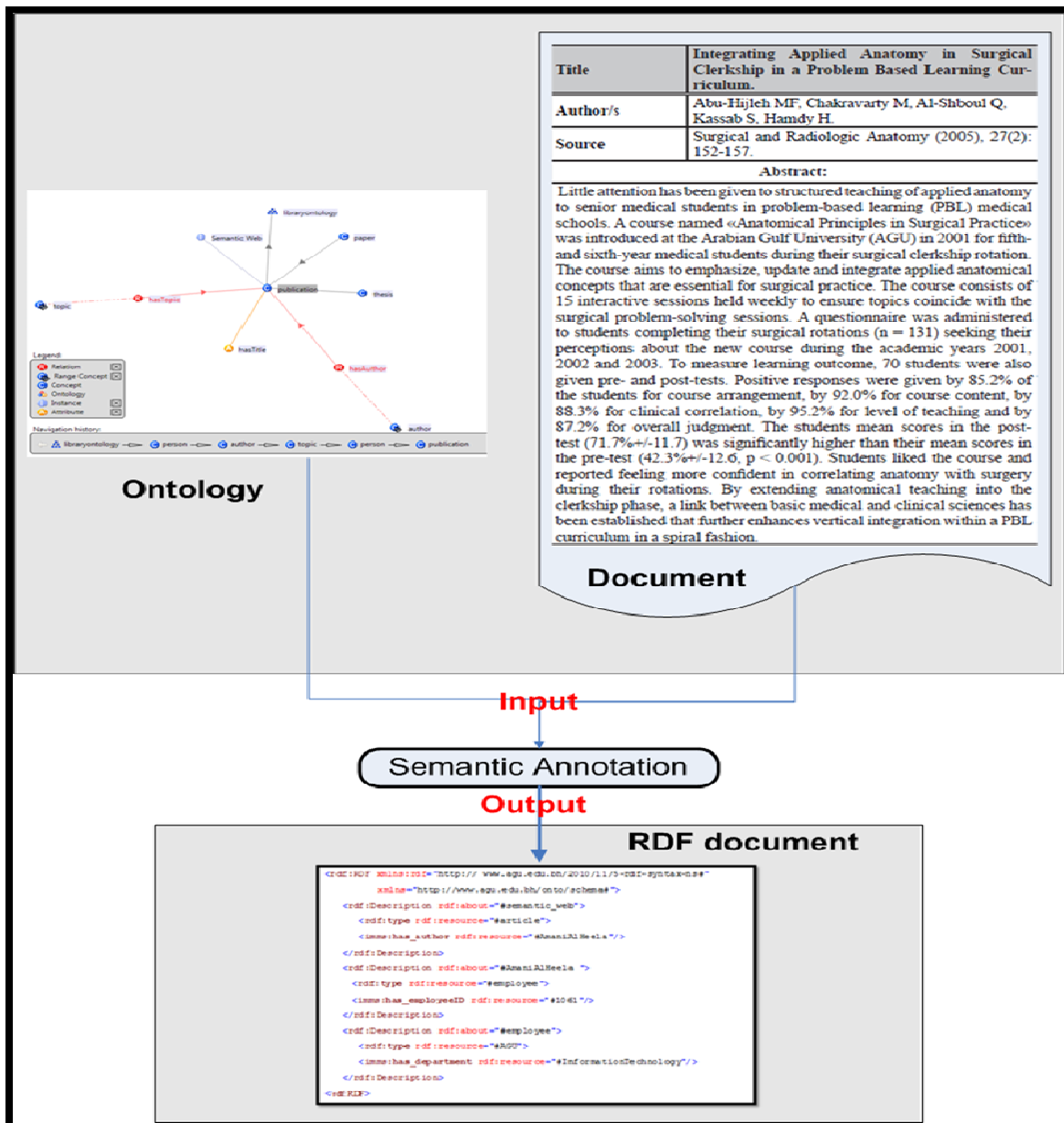

Figure 3: Diagram illustrating the annotation process

RDF is a W3C standard model used for the purpose of data interchange in the Web. It provides 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 is done by decomposing it into three 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 as in the following example: Name ('http://www.agu.edu.bh/employee/id1061",   "Amani Al Heela").   As shown in Table 1 ,the three views are explained xml and graph (see Figure 4).

Table 1: Table View of RDF example

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

The xml expression of the example is:

```
<rdf:Description about= 'http://www.agu.edu.bh/employee/id1061\>
<Name> Amani Al Heela </Name >
</rdf:Description>
```
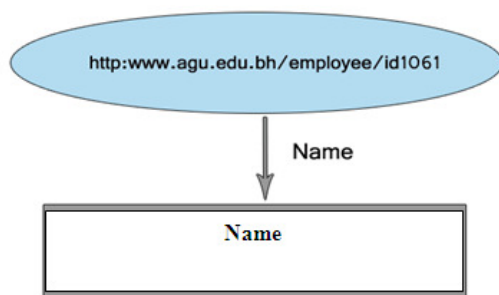
The graph view of the example is:

Figure 4: Graph View of RDF example

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 of the following elements:

- 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.

Bellow 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.

4. THE PROPOSED SYSTEM DESIGN

The current website of the library of Arabian Gulf University has been developed using Active Server Pages (ASP) language, also known as classic ASP using dynamically generated web pages. It has many cons and needs huge effort to improve it. The information it contains is not enough for a library website and not in the level of the universities' libraries websites, besides that many of its pages are under construction since 1997.

In this paper a SWLib developed taking advantage of Web 2.0 and Web 3.0 which integrated to the AGU library website. The web based application has user friendly and updated web pages that help users to easily find the correct information they need about the library. This SWLib also has an interface for professional users to execute the SPARQL queries for keeping the data in the RDF data store. Interface for annotation step is provided in this paper. It is an easy

task for the person responsible for enriching the information knowledge base. The proposed system design is the most suitable design according to the advantages of the software which are used to build the Semantic Web service for AGU e-library.

The proposed system is implemented by completing the steps as shown in Figure 5. First is to prepare a good design for the AGU library website that satisfies the required standards. Then the web developer converts this design to a developed website (in the proposed semantic library, the development language used is the ASP.NET language where as the program is the visual studio 2008) integrate web 2.0 applications, integrate Web 3.0 to the proposed system through developing a semantic web service.

The three major processes in integrating Web 3.0 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 for defining 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 and discussed in this paper, and **Indexing**: the RDF document that produced from the previous steps is stored in the SQL database to build RDF store.
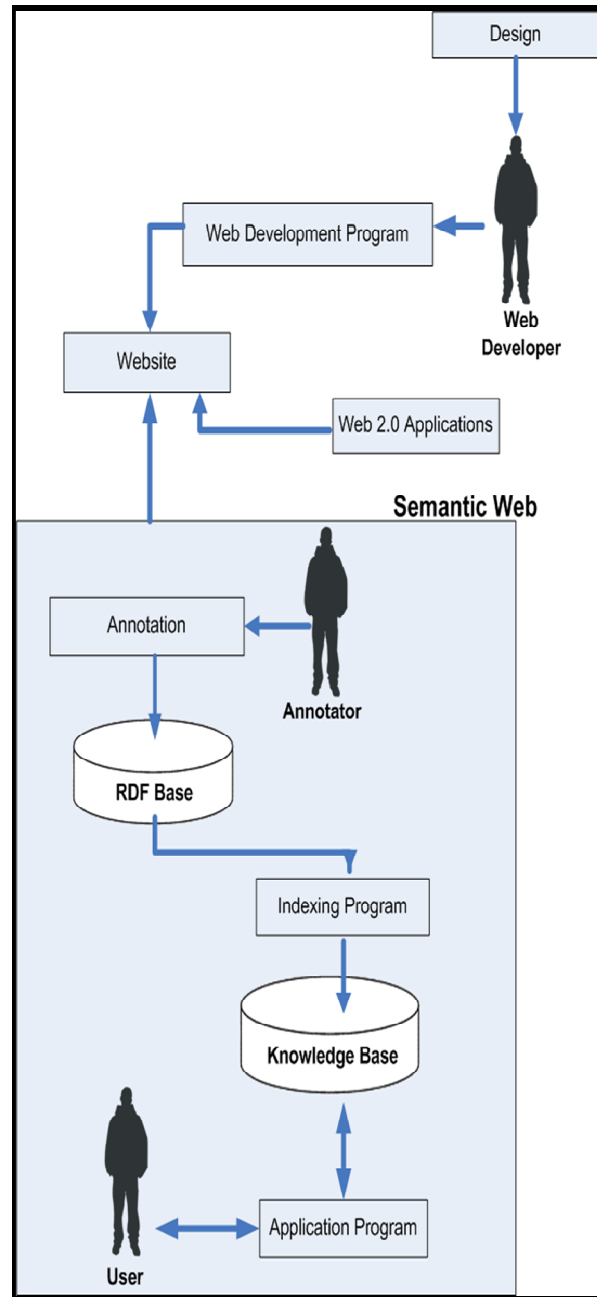

Figure 5: The conceptual architecture of the proposed website of AGU's library

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 2: Open the online paper file in the specified area.

Step 3: Start annotating by creating instances for each class and drag the related information in each one.

Step 4: Save ontology file with entered data.

Step 5: Save the ontology file as RDF.

End

Step 18: Store the RDF into RDF store in SQL database.

Begin

Step 1: Open SQL express 2005.

Step 2: Create a database for storing the RDF documents.

Step 3: Execute dotNetRDF store manager.

Step 4: Click file then New SQL Store Manger.

Step 5: Enter the connection details: database name, username and password.

Step 6: Import the RDF document.
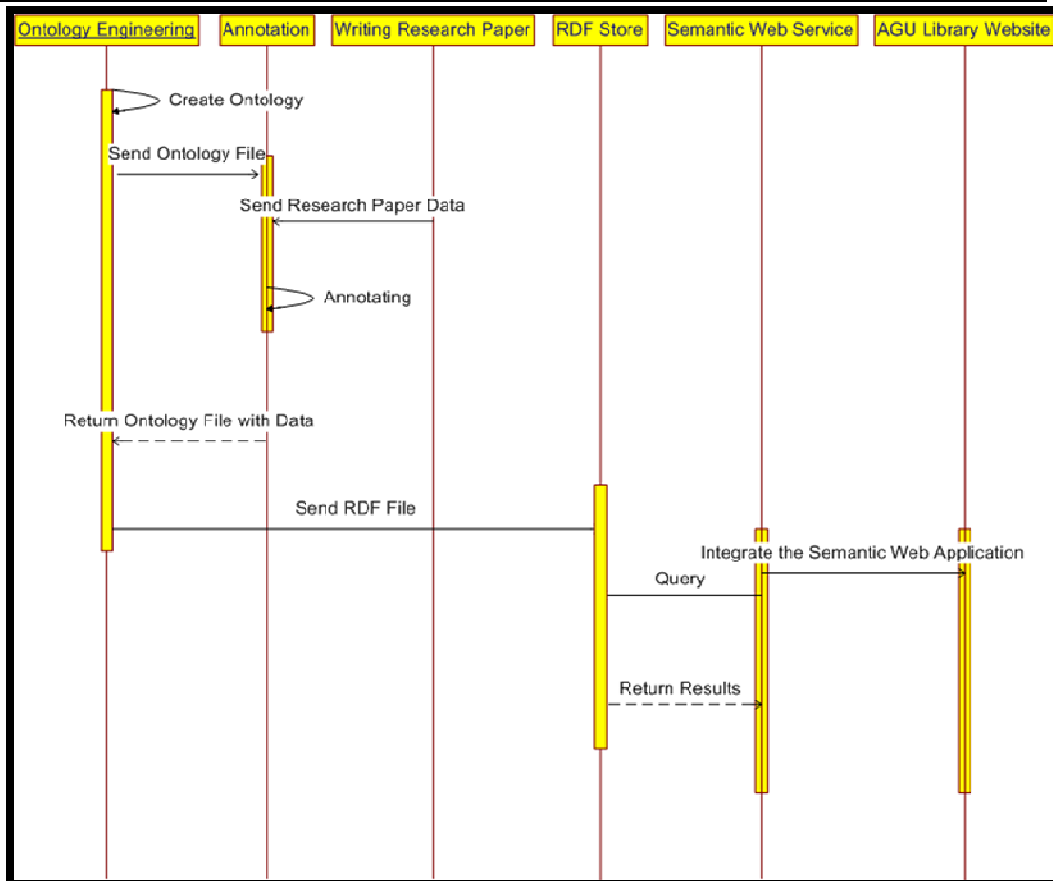
Step 7: Click create data store.

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, the subject matter of this paper 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. The successor of OntoEdit is distributed worldwide more than 5000 times. OntoStudio was originally developed for F-Logic but now includes some support for OWL, RDF, and OXML. It also includes such function as the OntoStudio Evaluator, which is used for the implementation of rules during modeling. This procedure has been recently patented.

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. Figures (7,8,9) describe in details how the annotation process are done.

There are two ways for implementing the annotation step. The first one is by using the OntoStudio software and the second is by using OntoMat.
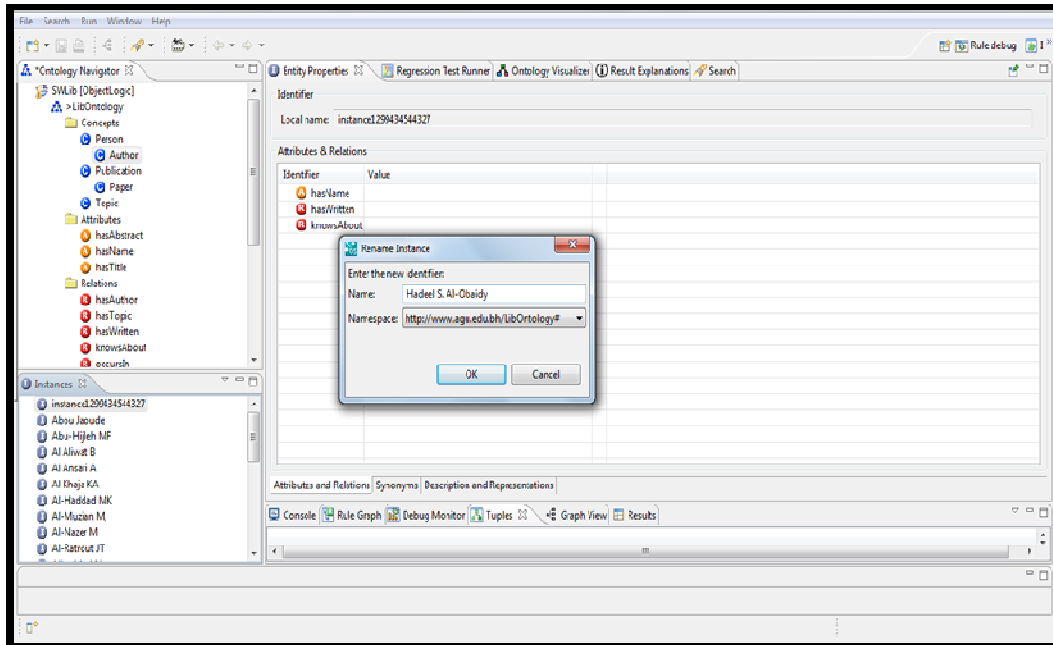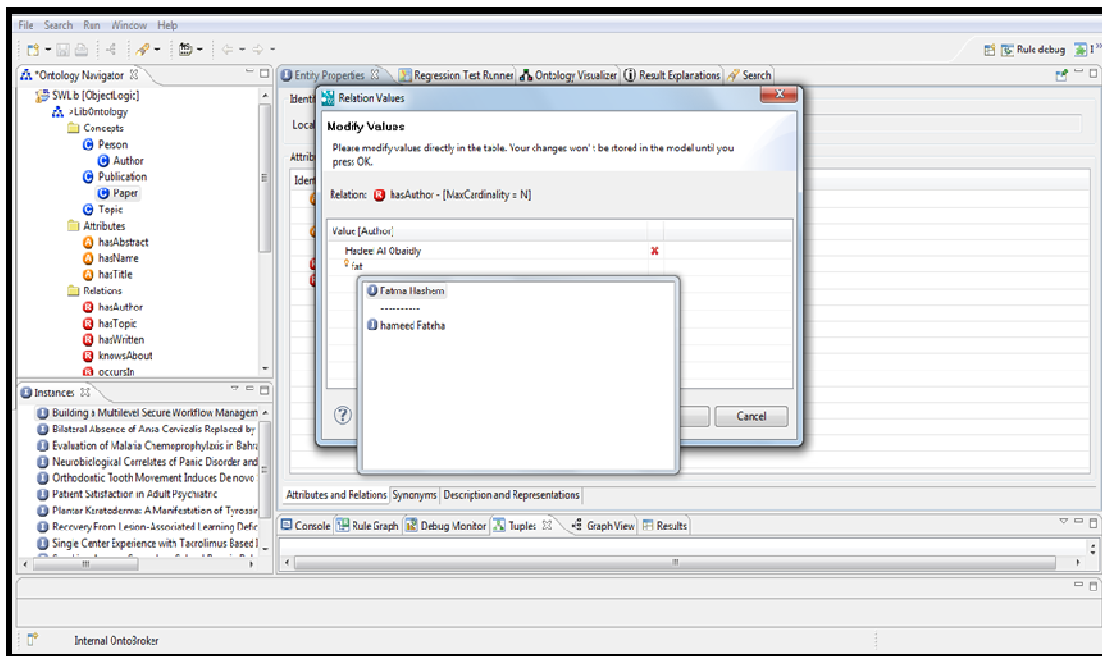

Figure 7: Annotation step – create new instance

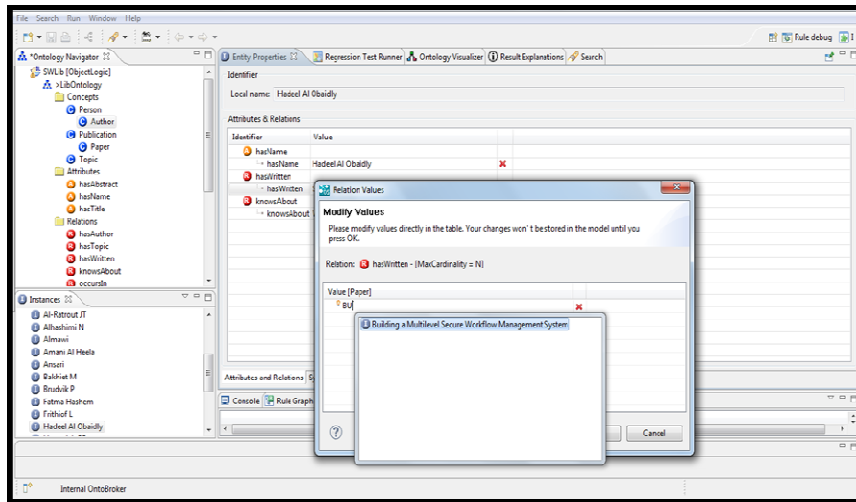Figure 8: Annotation step – enter data for the new instance attribute

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 displays 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. This process is shown in the Figure 10.
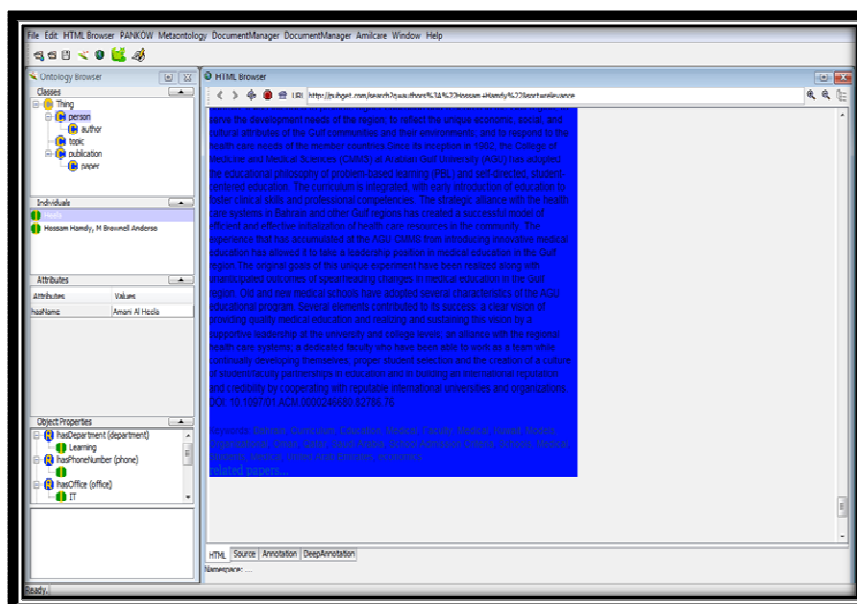

Figure 10: Annotation using OntoMat

Conclusion

Semantic Annotation is a basic technology for intelligent content. It is used for the benefit of wide range of content-oriented intelligent applications. This paper examines the annotation processes which represent one of the most important elements used in the development of a Semantic Web Service for e-library, as well as also to enrich ontology with knowledge.   Similar to any process, annotation has inputs and outputs. The inputs are the documents and ontology, whereas the outputs represent an RDF document to be used by any Semantic Web Service.

Future research is needed to shade some light on the potential use of semantic annotation technology. Still, there are several areas of research need to be undertaken to investigate the use of ontology such as developing and testing phase for the Semantic Web service for e-library, enhancing the proposed system and adding more functionality to e-library, extending the use of Semantic Web to other services in e-library, Maximizing the benefits of Semantic Web by reusing it for presenting other type of knowledge, expand the use of library Semantic Web application to the mobile technology and develop a web application that working in WAP.

REFRENCES

1. Alexander B. (2006), "Web 2.0: A new wave of innovation for teaching and learning. EDUCAUSE", Review. Vol. 41, No. 2, pp. 32–44. EDUCAUSE: Boulder, USA.

2. Berners-Lee, T., J. Hendler and O. Lassila (2001):"Semantic Web", Scientific American Magazine, 284(5):35-43. Retrieved on March 20, 2011 from : http://www.sciam.com/article.cfm?id=the-semantic-web& print=true.

3. Duineveld, A.J., Stoter, R., Weiden, M.R., Kenepa, B. and Benjamins, V.R. 2000. WonderTools? A comparative study of ontological engineering tools. International Journal of Human-Computer Studies 52:1111-1133.

4. G Kück, (2004), "Tim Berners-Lee's Semantic Web" South African Journal of Information Management"

Vol.6 (1). Retrieved on April 29, 2011 from http://www.sajim.co.za/index.php/   SAJIM/article/ download/297/288

5. Holger Lausen, Ying Ding, et al . (2005), "Semantic web portals :state of the art survey",   Journal of knowledge Management, Vol. 9, No. 5: 40-49.

6. J. Markoff, "Entrepreneurs See a Web Guided by Commonsense," The New York Times, Business, 12 Nov. 2006.

7. Michael Erdmann, et al. (2001) "From Manual to Semi-automatic Semantic Annotation: About Ontology-based Text Annotation Tools", ETAI Journal - Section on Semantic Web, No. 2.   Retrieved on April 15, 2011 from: http://www.aifb.kit.edu/web/Article474/en

8. Nenad Stojanovic, Alexander Maedche, et al ., (2001), "SEAL: a framework for developing Semantic Portals", In K-CAP, pp. 155-162.

9. Nicola Guarino, Pierdaniele Giaretta,(1995)," Ontologies and Knowledge Bases: Towards a Terminological Clarification. In Towards Very Large Knowledge Bases" , N.J.L. Mars, Ed. Amsterdam: IOS Press.

10. O'REILLY, T. (2005) , "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. Retrieved on January 25, 2011 from: http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/ 09/30/what-is-web-20.html

11. Stefan Decker, Frank van Harmelen, et al. (2000), "The Semantic Web - on the respective Roles of XML and RDF", IEEE Internet Computing, Retrieved on February 21, 2011 from: http://www.aifb.kit.edu/web/Article500/en

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

13. Yeh, Ching-Long and Chang-Gang(2011)," Design and Implementation of Semantic Community Web Portal", Retrieved on April 29, 2011 from http://www.cse.ttu.edu.tw/chingyeh/papers/DATFPortal .pdf