State of Art: Semantic Web

Najla B. Ibraheem   **Ban Sh. Mustafa**

College of Computer Sciences and Mathematics
University of Mosul, Iraq

ABSTRACT

Semantic Web is an extension to the current web. It will convert the way we use World Wide Web (WWW) by giving the machine the capability to process and infer data in web. It starts as a vision and becomes a future trend in web.

Due to the huge data that is scattered as web pages in web today, adding semantic meaning to data stored in these pages became necessary for the next age of information technology. The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users. Several tools and new technologies have been emerged to help bring this vision to reality.

In this paper, Semantic Web is defined and described with its layering architecture and supporting technologies and tools. An example is given to show how to use these tools to semantically representing data model. At last, challenges and difficulties faced building this web and made it an extension to the current web has been discussed.

Keywords: Semantic Web, Ontology, OWL, RDF

الحالة العلمية الراهنة للويب الدلالي

نجلاء بديع الدباغ   **بان شريف مصطفى**

كلية علوم الحاسوب والرياضيات، جامعة الموصل

الملخص

الشبكة المعنوية هي امتداد للشبكة الحالية ومن المؤمل باستخدام هذه الشبكة أن تتغير طريقة استخدامنا للشبكة العنكبوتية إذ تعطي للماكنة القابلية على معالجة بيانات صفحات الويب والمخزنة بنماذج بيانات معنوية وكذلك القابلية على استنتاج حقائق والحصول على بيانات جديدة من هذه النماذج.

نظرا لضخامة البيانات الموجودة حاليا في الويب والموزعة على شكل صفحات ويب, فان إضافة طرق جديدة لتمثيل هذه البيانات مع الأخذ بنظر الاعتبار مضمونها ومعناها وعلاقاتها مع البيانات الأخرى أصبحت من الضروريات لعصر تكنولوجيا المعلومات القادم. وسوف تمكن الهيكلية المعنوية للبيانات المخزنة لصفحات الويب من بناء بيئة تمكن العملاء البرمجيين من التجوال في صفحات الويب والقيام بوظائف معقدة للمستخدمين. وقد نشأت عدة أدوات وتقنيات برمجية لدعم بناء هذه الشبكة.

في هذا البحث، تم إلقاء الضوء على الشبكة المعنوية تاريخ ظهورها والنظرة المستقبلية لها. كما تم وصف معماريتها الطبقية مع شرح الأدوات والتقنيات البرمجية المطلوبة لبناء تطبيقات هذه الشبكة. وقد أعطي مثال لبيان كيفية استخدام هذه التقنيات في بناء نموذج معنوي للبيانات. وأخيرا تم مناقشة التحديات والصعوبات المواجهة لبناء الشبكة وجعلها امتداد كامل للشبكة العنكبوتية الحالية.

الكلمات المفتاحية: الويب الدلالي، الانتولوجي, OWL, RDF

1. Introduction

The current web has huge amounts of information in forms of distributed hypertext and hypermedia, coming from different resources. Web is searched and queried to access information by giving the URI address, using any search engines, or just follow links to other pages to get what is searching for. Web has a characteristic of being easy to use and explore and this gives it the wide usage and popularity. One shortcoming in this paradigm, the search answers give some irrelevant results due to huge content of web and the way search engine works. The search algorithm works by matching keywords not understanding the search request and analyzing the user needs. Thus, it works in a syntactic way, no semantics is involved [8].

To go further from this step, and can respond according to the semantic of the search request, web data should be stored in a way that let machine understand its meaning and relationships between different resources and, thus gives the ability to process it. In this way, we shall expect a more specific answer to search request and even faster. This vision is first introduced by Berners-Lee in 2001 [13]. Tim Berners-Lee originally expressed the vision of the Semantic Web as follows:

"I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A 'Semantic Web', which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The *'intelligent agents'* people have touted for ages will finally materialize".

Berners-Lee posits that if the past was document sharing, the future is data sharing.

To accomplish goals of the semantic web as versioned by Lee, new standards and technologies are developed to help machines processed and understand web data and so support richer discovery, data integration, automation, reuse of data across different applications [2].

2. Limitation of Current World Wide Web

Current World Wide Web (WWWs) are composed from a huge numbers of web pages (documents) written in Hypertext Markup Language (HTML).

HTML is a set of markup symbols or codes inserted in a file intended for display on a WWW page browser [1]. Web browser follows these markup symbols to display the page's words and images to the user. Thus, these markup symbols represent the syntax of the page representation and have nothing with the semantics of the page contents.

For example, HTML can make the document level assertions such as "this document's title is 'Computer Languages', but with no capability to assert unambiguously that, for example, Book title "Introduction to C#" have author name "Jim Blacks", Which also the author of the other book title "C# Concepts and Examples". Rather, HTML can only say how to display this information to the user. No way to describe their relations to others [2].

3. Semantic Web Solutions

Because of the limitation of HTML for defining semantic meaning and relationships between the different types of Web page contents, a future view of Semantic Web is to publish languages specifically designed for representing and manipulating web page resources (contents). Several languages are published under the

World Wide Web Consortium (W3C) Group Standardization: Resource Description Framework (RDF), Web Ontology Language (OWL), and Simple Protocol and RDF Query Language (SPARQL).

Using these technologies we can replace the content of web documents by a more descriptive data to provide a knowledge structure we have about the content and, thus add semantics to data [2]. A machine can process knowledge by using reasoner to infer new facts or answer a query about data.

By reasoning we mean deriving facts that are not expressed in knowledge base explicitly [9].

4. Ontology

Ontology is a term borrowed from philosophy. It has several definitions. One of the most known definitions is:

"Ontology is a formal specification of a shared conceptualization "[7]

Ontologies are considered one of the most concepts of the Semantic Web, it represents the Semantic Web vocabulary which can be seen as a collection of URIs with a usually informally described meaning [3].

Ontology will represent vocabularies and knowledge model for specific domain. Ontology can be built upon RDF model, RDFS model or using OWL language. Several tools are available to make easy building ontology, like protégé.

Several ontologies are available to be used or added to other ontology, you can find it in several web sites like semanticweb.org [15]. Semantic search engine is starting to appear which supports thousands of known and well written ontologies like Swoogle and SWSE (see figures 1, 2). Table 1 shows some ontologies that are currently described on semanticweb.org:

Table 1. Ontology examples and their languages

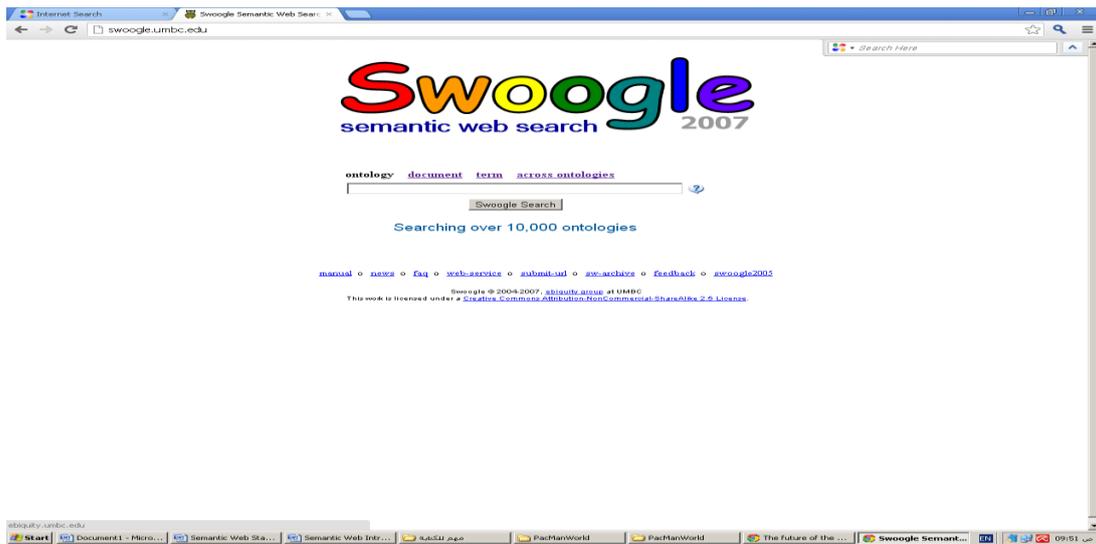| Ontology Title	Language
Dublin Core	RDF
FOAF	OWL DL
Basic Geo Vocabulary	RDF Schema
VCard RDF	RDF
Music Ontology	OWL 2

Figure 1.: SWSE Search Engine

Figure 2. Swoogle Search Engine

5. Semantic Web Stack

The Semantic Web Stack illustrates the architecture of the Semantic Web. It consists of several layers, where each layer represent, a technology or language used in the building of a semantic web application. Each layer exploits and uses the capabilities of the underlying layer. These technologies are standardized by W3C Group. All layers should be implemented to get the complete vision of semantic web [5].
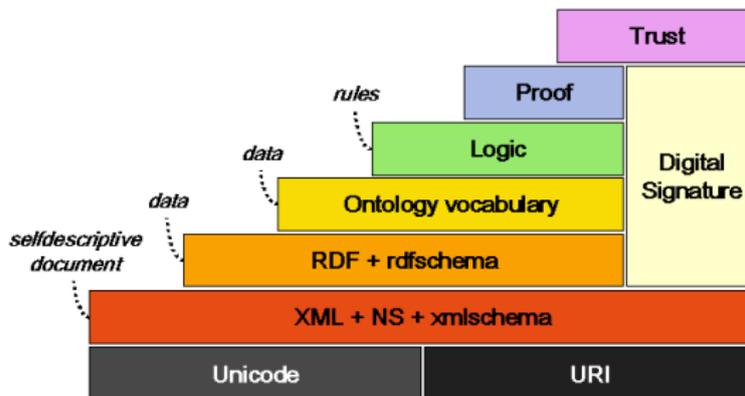

Figure 3. The Semantic Web layers

For building a semantic web applications, all resources (web data) is represented by unique identifier, which is represented by using URI (Uniform Resource Identifier), so we can build a global web on top of it. Any identifier used by web should have URI [2].

These resources represented in triples based structure, which often called "Resource Description Framework" (RDF) syntax. A triple is three URI resources (*subject-predicate-object*) in RDF. RDF enables to represent information about resources in the form of graph, this is called Giant Global Graph [16].

RDF Schema was designed to be a simple data typing model for RDF. Using RDF Schema, it is possible to create hierarchies of classes and properties by using of: rdfs:subClassOf and rdfs:subPropertyOf, and adding more advanced stuff such as

creating ranges and domains for properties .These let us give the class type of the subject (Domains) and class type or data type of object (Ranges) by using rdfs:domain and rdfs:range

5.1 OWL (Web Ontology Language)

It is a Knowledge representation (KR) language for the Semantic Web. It adds more semantic to RDFS definitions, such as, stating more cardinality constraint, restrictions of values, or adding characteristics of properties such as transitivity, functionality, and symmetric. OwL based on description logic (DL) and, thus allows using available powerful DL reasoner to infer in knowledge model [14].

For retrieving information in semantic web applications, SPARQL query language is recommended in W3C. It is used to query any RDF-based data including OWL and RDFS statements [2].

The Ontology layer supports the evolution of vocabularies, which allow the data sharing and integration between different domains. Digital Signature layer allow for detecting of documents alterations and verifies that documents come from trusted source [3].

The top layer: Logic, Proof and Trust, are currently being researched and simple application demonstrations are being constructed. With Logic layer allows for adding rules to describe more advanced relations. Trust, to add trust models to our application [2].

6. Semantic Web Projects

In this section, a very popular Semantic Web project is demonstrated: The *Friend of a Friend* (FOAF) project is creating a Web of machine-readable pages describing people, the links between them and the things they create and do. FOAF defines an open, decentralized technology for connecting social Web sites, and the people they describe. This application based on well-known FOAF ontology which is an RDF based schema to describe the relationships people have to other people and the "things" around them. Because it focused on people's online activity or identity (foaf:mbox, foaf:msnID), it is suited for describing people on Web-based Social platforms (facebook, twitter, blogspot, ...) [2][3] and [4].

7. Semantic Web Challenges [6][2]

- *Semantic web content availability:* Semantic Web content is web content annotated according to particular ontologies, which defines the meaning of the words or concepts appearing in the content. Due to the little availability of semantic web content, we need to develop new standards tools that can upgrade the current web content (HTML pages, multimedia and web services, others) to the new semantic web content.
- *Ontology availability*: Ontologies are basic components in any semantic web application because it provides the common vocabulary and semantic for specific domain. A big effort must be done to develop large numbers of used ontologies for different domains.
- *Multilingualism:* The problem of dealing with different National languages is a problem in current web and has to deal with it in Semantic Web to provide accessibility to information with different languages.
- *Standardization:* Languages and Technologies should have their acceptance from W3C as standards so it can work towards building the semantic web vision.

- *Inconsistency:* This is a contradiction that arises from combining large ontologies coming from different resources. To data sharing, we have to deal with the problem of ontology's concepts mapping.
- *Security Issues:* In semantic Web applications another security challenges arise from the misleading of information to others. Cryptography techniques can be used for consistency and identity verifier, other techniques like trust can be used over the open distributed environment.

8. RDF, OWL Data Models Example

In order to illustrate the way information is stored in semantic web, an RDF model is built to show the concepts and relationships for managing a thesis concept and its relationships. Thesis has an Author and Title.

Following is a list of RDF/XML document for the above concepts:

```
<?xml version="1.0"?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dc="http://purl.org/dc/elements/1.1/#">
  <rdf:Description rdf:about="http://author/">
    <dc:name>Jack Jimes</dc:name>
  </rdf:Description>
  <rdf:Description rdf:about="http://thesis/">
    <dc:title>Semantic Web</dc:title>
    <dc:writtenBy>Jack Jimes</dc:writtenBy>
  </rdf:Description>
</rdf:RDF>
```

In these points the above list is discussed:

- The first line of the RDF document is the XML declaration. The XML declaration is followed by the root element of RDF documents: <rdf:RDF>.
- The xmlns:rdf namespace, specifies that elements with the rdf prefix are from the namespace "http://www.w3.org/1999/02/22-rdf-syntax-ns#".
- The xmlns:dc namespace, specifies that elements with the dc prefix are from the namespace "http://purl.org/dc/elements/1.1/#".
- The <rdf:Description> element contains the description of the resource identified by the rdf:about attribute.
- The elements: <dc:name> is the properties of the resource author.
- The elements: <dc:title>, <dc:writtenBy> are properties of the resource thesis.

Below is the Triples and Graph of the resulted RDF data model:

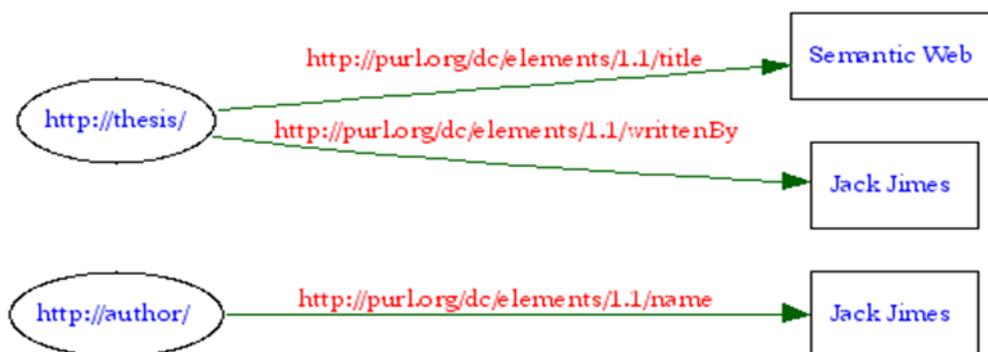| Number	Subject	Predicate	Object
1	http://author/	http://purl.org/dc/elements/1.1/name	"Jack Jimes"
2	http://thesis/	http://purl.org/dc/elements/1.1/title	"Semantic Web"
3	http://thesis/	http://purl.org/dc/elements/1.1/writtenBy	"Jack Jimes"

Figure 4. Graph of the Data Model

8.1 RDFS Data Model

In RDFS model, classes can be defined instead of rdf description. Also, we can add the subclass and subproperty logic to model. We can add to the above model the fact that the author belongs to the upper Faculty group, also we add a range and domain information for name property. Since, an RDFS class is an RDF resource, we can abbreviate using rdfs:Class instead of rdf:Description, and drop the rdf:type information. Below is a sample code to clarify these pieses of information:

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:dc="http://purl.org/dc/elements/1.1/#">
<rdfs:Class  rdf:ID="faculty" />
<rdfs:Class  rdf:ID="author">
 <rdfs:subClassOf  rdf:resource="#facuilty"/>
 </rdfs:Class>
< rdfs:Property rdf:ID:"name"
  rdfs:range   rdf:resource="#author"
  rdfs:domain  rdf:resource="string">
</rdfs:Property>
 </rdf:RDF>
```

8.2 OwL Ontology Model

More semantics can be added by using an ontology language that is based on DL for building our data model. OWL (Web Ontology Language) is one of the W3C group ontology language recommendation. In our example, we can add description class, which is a class defined by restriction constraints on its properties. A *csAuthor* class includes all authors that are belonging to Computer Science Department. This condition can be written in DL as:

 Author and (hasDept hasValue "ComputerScince").

The instance that satisfies these two conditions will belong to *csAuthor* class. Another description class is *csThesis* which is defined to have all instances that belong to Thesis class and their authors belong to *csAuthor* class. Figure 5 shows a graph for this ontology model produced by Protégé application.
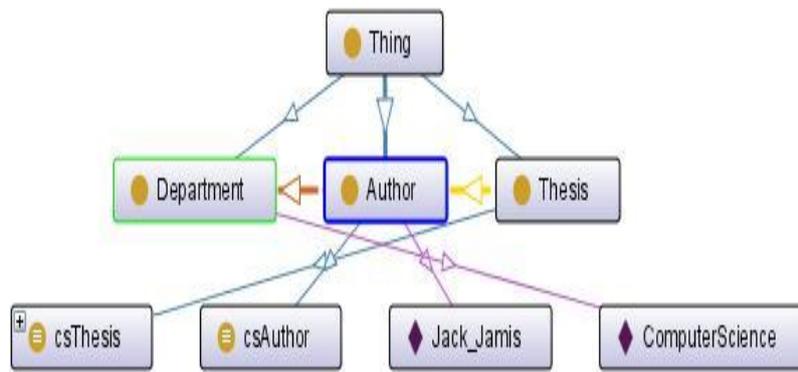
Figure 5. Ontology Example Graph Produced by Protégé Application.

Below is a listing for final ontology model exported in turtle format:

```
@prefix dc:   <http://www.owl-ontologies.com/Ontology12.owl#> .
@prefix rdfs:   <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl:   <http://www.w3.org/2002/07/owl#> .
dc:Department     owl:Class .
dc:csAuthor        owl:Class ;
   rdfs:subClassOf dc:Author ;
   owl:equivalentClass
       [ owl:Restriction ;
         owl:hasValue  dc:ComputerScience ;
         owl:onProperty  dc:hasDept
       ] .
dc:Author       owl:Class .
dc:title    owl:DatatypeProperty ;
   rdfs:domain dc:Thesis ;    rdfs:range    xsd:string .
dc:writtenBy     owl:ObjectProperty ;
   rdfs:domain dc:Thesis ;     rdfs:range dc:Author .
dc:Jack_Jamis       dc:Author .
dc:hasDept        owl:ObjectProperty ;
   rdfs:domain dc:Author ;     rdfs:range dc:Department .
dc:Thesis    owl:Class .
dc:ComputerScience       dc:Department .
dc:name    owl:DatatypeProperty ;
   rdfs:domain  dc:Author ;       rdfs:range  xsd:string .
dc:csThesis      owl:Class ;
   rdfs:subClassOf dc:Thesis ;
   owl:equivalentClass
       [ owl:Restriction ;
         owl:onProperty  dc:writtenBy ;
         owl:someValuesFrom  dc:csAuthor
       ] .
```

9. Pros and Cons for Semantic Web

Pros[18]:

- Using Semantic Web, more precise query can be answered.
- By using Semantic Web languages, we can get smarter applications with less work.
- Ontologies easy the task of sharing and linking data between applications.

Cons[18]:

- No sufficient tools still exit to the whole Semantic Web stack architecture.
- Difficult to scale
- How to make all the metadata. A big responsibility for all to produce ontologies in different used domains.
- Still suffering from the efficiency of reasoner and inference tools used.

10. Discussion

Semantic Web Vision is first introduced by Berners-Lee in 2001. From that date a big academic research efforts have been done to make this dream a reality.

The dream has been slowly to arrive, because the original vision was too close to Artificial Intelligence. Technologies and tools were inefficient. Keywords search and tagging were good enough at that time.

Today, with the huge amount of available web pages and data on WWW, a new need has been introduced in web search, business applications and services, and organization's data and information. Keywords matching search engine does not work in efficient way as before because of the growing amount of pages and data available on the web. Most of our search query will give us irrelevant answers.

Also, applications need to share their data across internet with other applications, and this rises the need to make data self describing. Meaning of data will be added to data as metadata and become part of it. Big organizations need to provide a single unified view of their data across all applications. Eventually, reducing data redundancies and bringing more meaningful and contextual data to the frontend [12]. This will let applications have intelligent and behave in a smarter way. Big organization now tends to move into adding semantics to their databases. NASA, Shell, BBC and FACEBOOK all work towards the future of semantic [11]. So, semantics will play an important role in the future of technology.

Now, several companies start to expose data to semantic web (no. of ontologies start to increase), and also the appearance of good technologies and tools for building a semantic web applications, the work towards the dream becomes closer. Now, the dream is really necessary for the future work with data and web.

REFERENCES

[1]   Margaret Rouse, Sep. 2005, "HTML (Hypertext Markup Language)", http://searchsoa.com/

[2]   http://en.wikipedia.org/wiki/Semantic_Web

[3]   http://semanticweb.org/wiki/Ontology

[4]   http://semanticweb.org/wiki/FOAF

[5]   Marja Riitta, Eric Miller, Nov. 2001, "W3C Semantic Web Activity", proceedings *Semantic Web Kick-off Seminar in Finland.*

[6]   V. Richard Benjamins, Jesús Contreras, Oscar Corcho, Asunción Gómez-pérez, 2002,"Six Challenges for the Semantic Web", In KR2002 Semantic Web Workshop.

[7]   Tom Gruber, Nov. 2006, "Where the Social Web Meets the Semantic Web". Keynote presentation at ISWC, The 5th International Semantic Web Conference.

[8]   Ian Horrocks, "Ontologies and the Semantic Web", Oxford University Computing Laboratory Oxford, UK.

[9]   Marek Obitko (advisor Vladimir Marik), 2007, "Translations between Ontologies in Multi-Agent Systems", Ph.D. dissertation, Faculty of Electrical Engineering, Czech Technical University in Prague.

[10]   Sneha chuttar, feb 15 2010, "Semantic Web pp presentation", At www.authorStream.com, [11] www.semanticweb.com

[12]   Yogesh Malik, Dec 8 2010, "The Future of the Web is Semantic T", At www.semanticweb.com

[13]   Berners-Lee, Tim, May 1, 2001, "The Semantic Web". Scientific American.

[14]   W3C, 2009, "OWL 2 Web Ontology Language Document Overview". W3C. At http://www.w3.org/TR/owl2-overview/

[15]   http://semanticweb.org.

[16]   Timbl, 2007, "Giant Global Graph", AT http://dig.csail.mit.edu.