UNIVERSITY OF CALGARY

A tutorial report for SENG 609.22

Agent Based Software Engineering

Course Instructor: Dr. Behrouz H. Far

Semantic Web: an overview

Stephane Saleh

Abstract

Semantic Web is an extension of the current Web. It gives the existing Web content a well-defined meaning: enabling machines understand and process semantic information, and cooperate with humans to realize the full potential of the Web. In this paper, we give a general description and discuss the role of each layer of the Semantic Web architecture, and compare Semantic Web and the XML based Web agents.

Introduction

Imagine you want to travel to Hawaii on the weekend. You instruct your Semantic Web agent through your handheld device to book you a flight to Hawaii. The agent retrieves information from your scheduler agent, and look up several travel services, and checks for the one that offer the best flight ticket price, and hotels rated four stars and above. The agent returns you a travel package, telling you it has asked for a cab to pick you up from the airport, and asks you if it can take care of your travel insurance.
These tasks cannot be carried out with today's World Wide Web but agents will be able to solve all kind of human problems thanks to the Semantic Web that will evolve in the future.

The World Wide Web (WWW) is a universal information space.
Most of the information on the web uses Hypertext Markup Language (HTML) that tells a computer how to display it on a browser. This information is intended for humans to read, not for machines to manipulate meaningfully.
The Semantic Web approach is an extension of the current Web. It is based on recent technologies such as the eXtensible Markup Language (XML) [1] and the Resource Description Framework (RDF) [2]. The Semantic Web is still in a very early stage of development, and is intended to give the existing Web content a well-defined meaning: enabling machines understand and process semantic information, and cooperate with humans to realize the full potential of the Web.

Semantic Web Architecture

The road map for developing the Semantic Web is proposed by Tim Berners-Lee, inventor of the WWW and the World Wide Web Consortium (W3C) [3]. The Semantic Web architecture is presented in Figure 1.

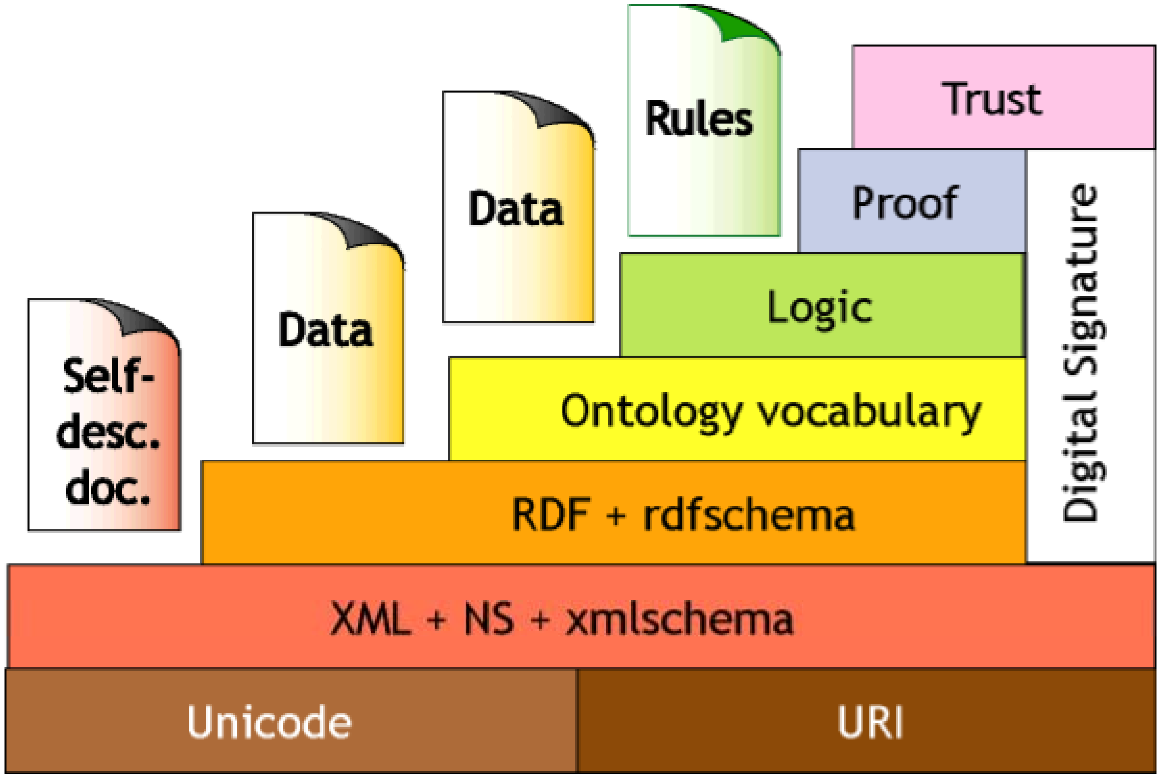

Figure 1. Semantic Web Architectural Dependencies [4]

The following paragraphs will describe briefly the different layers of the proposed Semantic Web architecture.

Uniform Resource Identifier (URI)

The base layer of Semantic Web is composed of Unicode and URI.

Any resource on the Web will be identified by a URI. The use of URIs to encode information in a document ensures that concepts are uniquely defined and can be referred to by anyone on the web.

For example, if we consider two programs *A* and *B* that have a concept named “address” in their databases, and the concept in program *A* refers to a street address and the concept in program *B* refers to a mailing box address. Program *B* that receives an “address” from program *A* will interpret it as a mailing box address and not as a street address and vice versa. Therefore, we need to differentiate the two “address” concepts. The use of a different URI for each concept solves the problem; it insures that they are distinguished from one another.

The URI is the foundation of the Web. It let systems grow on a globally decentralized basis, similar to the Internet systems, because once data is given an URI it can be referred to by anyone on the web. Moreover, it helps build intricate and complex relationships, and queries to be read and processed by machines.

eXtensible Markup Language (XML)

XML lets a user creates his/her own tags to format a document.
XML adds arbitrary structure to the document in order to self-describe data, but it does not specify the meaning of this structure.
For example, a program will be able to read the tag <address> in a document and process it but it does not understand it's semantic; what this tag is used for.
XML namespaces and schema [1] add URI to elements and attributes to prevent confusion.
In order to assign meaning to XML documents, we need to combine XML to the Resource Description Framework.

Resource Description Document Framework (RDF)

RDF is a key Semantic Web technology. RDF provides a means for publishing both human-readable and machine-processable vocabularies designed to enable the encoding, exchange, and reuse of structured data. It creates semantic meaning out of tag languages by defining the relationships between pieces of information.

RDF defines a data model of triples (subject, predicate, object). The subject defines a resource, the predicate is a property-type, and the object represents a value. This model is used to represent resources and their properties. The subject, predicate, and object are described as URIs and can be implemented in XML format.
Although, it is not the only format, RDF XML is considered to be the standard interchange format for RDF on the Semantic Web.

The following example illustrates the application and use of the RDF data model.

If we consider the two following sentences:

1. "John Doe is the author of the Semantic Web book "

2. "The author of Semantic Web book is John Doe"

These two sentences have the same meaning for humans. However, these strings are completely different for machines.
Using the triples model, RDF provides an unambiguous method of expressing semantics in a machine-readable encoding.
RDF associates properties with resources. In our example, the data model has a resource “SemanticWebBook”, a property-type of “author” and a corresponding value or object of “John Doe”.

The RDF document of the above example is as follows:

```
<RDF:RDF>
    <RDF:Description              RDF:HREF              =
“http://uriOfSemanticWebBook”>
            <DC:Title>Semantic Web</DC:Title>
            <DC:Author>John Doe</DC:Author>
            <DC:Publisher>Lafontaine</DC:Publisher>
    </RDF:Description>

</RDF:RDF>
```

RDF schemas are combined to the RDF data model to define the meaning, characteristics, and relationships of a set of resource properties.

For example, RDF schema defines a vocabulary to express the “Book” class. It specifies that a “Book” has an author, a title and a publisher property. A “Book” can be a sub-class of “Volume”. All users of a resource of type “Book” are expecting to find the mentioned properties on a book and relationship to other resources types.

Ontologies

Ontology is a formal description of terms and their interrelationships. It has a taxonomy and inference rules.
Ontology’s taxonomy takes RDF Schema a step further, by defining more in depth properties and classes.
One can assign properties to classes and allow subclasses to inherit these properties. The result is increasing the relations among entities on the Web.
The principle of inference in ontologies is about being able to derive new data from already known data. It helps computers manipulate data more effectively.
For example, an ontology can express the rule “T5D 7G4 postal code is associated to the province of Alberta” and an intelligent application using this rule can deduce that “if an address postal code is T5D 7G4, then it is located in Alberta”.

Ontologies are aimed to enhance the functioning of the Web, such as improving search engines. The Defense Advanced Research Projects Agency (DARPA) Agent Markup Language (+OIL) (DAML) [5] is an important RDF language that expresses ontologies. Applications will use DAML to relate information on a web page to the associated knowledge structures and inference rules.

Digital Signature

Digital signatures are encrypted blocks of data that will be applied to RDF documents. Digital signatures authenticate the source of information and will be used by the Trust layer.

Logic, Proof and Trust

These layers have not been developed yet.

What can be said about logic is that computers need to use a logical language to reason by inference.

Systems can use *logic* for *proof* checking. For example, a rule states, "the employee of the month has to sell ten vehicles per month". If a car salesman sells eight cars and two trucks, a system can use a mathematical rule to sum the vehicles sold by the salesman and combine it with the previous rule in order to prove that the salesman is an employee of the month.

Trust is a vital feature in Semantic Web applications. People need to trust the information they receive. Semantic Web Trust layer provides computers with digital signatures and proof mechanisms to allow them check if a source of information can be trusted or not.

Semantic Web and agents

A Web agent is a program that collects information form diverse sources on the Web, process it and exchanges the results with other programs.
Semantic Web promotes the idea that different agents, not necessarily designed to work together, can transfer data among themselves if the data come with semantics.
Web services that perform specific functions already exist on the Web.
Agents can collaborate with these Web services to execute specific functions. Agents have to locate first these services using service discovery. But, this is possible only if there is a common language that describes a service, and helps agents "understand" the functions a service is offering and how to interact with it.

Recent work on XML protocols is focused on standardizing XML based interactions between computers. This work introduces some aspects of the Semantic Web.
The use of XML technologies such as Web services Description Language (WSDL) [6], an XML-formatted language used to describe a Web service's capabilities to communicate and exchange messages, Universal Description, Discovery and Integration (UDDI) [7], a Web-based distributed directory that enables business to list themselves on the Internet and discover each other, and Simple Object Access Protocol (SOAP) [8], a platform independent protocol that provides a way for applications to communicate with each other over the Internet, allow the description and discovery of web services, and the ability to share information between them without having any prior knowledge of one another.
However, the Semantic Web is more flexible then these protocols because they impose standardization.
Semantic Web allows consumer and producer agents reach a shared understanding by exchanging ontologies. Moreover, agents can develop new reasoning capabilities by discovering other ontologies.
Semantic Web will require from some agents to use artificial-intelligence technologies to satisfy all these features.

Conclusion

The Semantic Web is “a vision: the idea of data on the Web defined and linked in a way that it can be used by machines for automation, integration and reuse. The Web can reach its full potential only if it becomes a place where data can be shared and processed by automated tools as well as by people.” [9]
Some people in the developer community argue about the usefulness of the technology, and the difficulty of using RDF.
On the other hand, Semantic Web is a logical evolution of the current Web.
It will allow the evolution of human knowledge. Concepts that are well defined will be linked into a universal Web. Software agents will analyze and process the information on the Web to help us in our daily tasks and answer questions more effectively and easily.
The Semantic Web is still at an early stage of development, and it is surely a long way from here to there.

Bibliography

[1] W3C Extensible Markup Language (XML): http://www.w3.org/XML

[2] W3C Resource Description Framework (RDF): http://www.w3.org/RDF/

[3] World Wide Web Consortium (W3C): http://www.w3.org

[4] Tim Berners-Lee, Semantic Web - XML2000: http://technetcast.ddj.com/tnc_play_stream.html?stream_id=459

[5] The DARPA Agent Markup Language with Ontology Inference Layer (DAML+OIL): http://www.daml.org/

[6] Web Services Description Language: http://www-106.ibm.com/developerworks/library/w-wsdl.html

[7] Universal Description, Discovery and Integration of Business for the Web: http://www.uddi.org/

[8] Simple Object Access Protocol: http://www.w3.org/TR/SOAP/

[9] W3C Semantic Web Activity: http://www.w3.org/2001/sw/

[10] Tim Berners-Lee, James Hendler and Ora Lassila (May 2001). *The Semantic Web.* Scientific American. http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html

[11] Edd Dumbill (2001). *The Semantic Web: A Primer*. XML.com: http://www.xml.com/lpt/a/2000/11/01/semanticweb/index.html