Web 2.0 – The Web's Edge

Instructor Teresa Pelkie

The Semantic Web – Class # 16 - Chapter 6

The Semantic Web

http://www.w3.org/2001/sw

The Quest for Information

From the early days of ARPANET to Web 2.0, the Internet has changed. Yet, if we look beyond the technological developments, we will see one common trend – we use the Internet to search for and acquire information.

The early Web gave us directories and search engines in order to find information. Web developers can code in meta data is assist in finding information. Social searches may aid in our ability to find what we are looking for.

No doubt, you have tried to find some specific piece of information, only to be frustrated with thousands of links to click on, which, in the end, did not produce the desired result.

The Semantic Web was an idea put forth by **Tim Berners-Lee** many years ago which would solve this problem by developing a Web where data can be integrated and shared.

At an MIT technology conference in February 2005, he stated:

"The Semantic Web looks at integrating data across the Web. 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. For the Web to scale, tomorrow's programs must be able to share and process data even when these programs have been designed totally independently. The Semantic Web is a vision: the idea of having data on the web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications."

At the beginning of this class, we learned that Tim Berners-**Lee** developed the HTML language and the HTTP protocol. He is credited with developing the World Wide Web, and is the chairperson of the **World Wide Web Consortium** (W3C – see http://www.w3c.org), an organization that oversees the development of many Web technologies. Tim Berners-Lee is in charge of the Semantic Web development activities at the W3C.

Human vs Computer Intelligence

Typical Web pages use text, images and page layout to present information in a way that we can understand. Computers are used to create, search and display information on the Web, but the computers themselves really can't make sense of all this information. They can't read, see relationships or make decisions like we can.

To a computer, the phrase “George Washington” is just two words. To humans, this phrase can have a variety of meanings, associations and relationships. But a computer does not know that “George Washington” is

- A name
- A first and last name
- A name of a person
- A man’s name
- What a man is
- The name of the first president of the US
- What a president is
- What the US is

Where the Semantic Web Comes In

The Semantic Web proposes to enable computers to be more able to understand meaning. Most of the content on the Web is designed for humans to read. It is not very easy for computer programs to interpret that information meaningfully. HTML does not tell us anything about the subject and nature of the content. The goal of the Semantic Web is to come up with some standards to express better the meaning of information so that computers can understand it and use it effectively. Some people refer to the Semantic Web as **Web 3.0**.

The Semantic Web is really an extension, not a replacement, of the World Wide Web. It provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. **It is a collaborative effort led by W3C** with participation from a large number of researchers and industrial partners.

While some websites are already using Semantic Web concepts, a lot of the necessary tools are still in development

The Semantic Web

The word **semantic** stands for the “**meaning of”**. The semantic of something is the meaning of something. The Semantic Web will provide the tools which will enable us to describe things in a way that computers can understand. It will do this by generating “**metadata**” – machine readable data that describes other data. The goal of the Semantic Web is to integrate “data” from diverse sources into meaningful “information” so that **automated processing** becomes possible.

The following is proposed:

- A standardized syntax for representing information will be developed
- We pages will be understood by humans and can be read and processed by machine or agents (software)
- Web searches will be accomplished by software agents and humans will receive meaningful results
- Common terminologies and knowledge structures, know as Ontologies, will be developed

The Semantic Web is about two things:

- Common formats for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents
- Language(s) for recording how the data relates to real world objects

Summary of the Semantic Web:

- A vision for the future Web as defined by Tim Berners-Lee
- Is not a separate web, but an extension of the current Web.
- Provides a way for machines to get much better at being able to process and understand the data that they merely display at present.
- Provides an emerging set of standards, markup languages, and related processing tools
- Is not about links between Web pages, but about the relationships between things or data (such as events, people, places and things)

The real power of the Semantic Web will be realized when people create many programs that collect Web content from diverse sources, process the information and exchange the results with other programs. The effectiveness of such **software agents** will increase exponentially as more machine-readable Web content and automated services (including other agents) become available

"The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

- *Tim Berners-Lee, James Hendler and Ora Lassila;* Scientific American, May 2001

Read full article at http://www.scientificamerican.com/article.cfm?id=the-semantic-web

YouTube

- **Intro to the Semantic Web** - http://www.youtube.com/watch?v=OGg8A2zfWKg
- **Tim Berners-Lee on the Semantic Web** - http://www.youtube.com/watch?v=mVFY52CH6Bc
- **Tim Berners-Lee: The next Web of open linked data** - http://www.youtube.com/watch?v=OM6XIICm_qo

Semantic Web Technologies

Two important technologies for developing the Semantic Web are already in place:

1. **XML-** Xtensible Markup Language
2. **RDF** - Resource Description Framework .

XML lets everyone create their own tags. It has become a predominant standard for labeling and tagging information. It is widely used for data exchange and integration and for defining higher level languages. Resource Description Framework RDF is used to give meaning to an object, It uses sets of triples, each triple being rather like the subject, verb and object of an elementary sentence, which are written using XML tags.

The third important technology for developing the Semantic Web is still being developed:

3. **Ontologies** - collections of information

The Resource Description Framework

– See http://www.w3.org/RDF and http://www.w3schools.com/rdf

RDF is a markup language and a W3C Specification for **describing resources** on the Web.

Putting information into RDF files makes it possible for computer programs (ie. spiders) to search, discover, find, collect, analyze and process information from the Web. RDF was designed to provide a common way to describe information so it can be read and understood by computer applications. It is not meant to display information.

RDF uses sets of triples written as XML, to display information as a graph. Each triple is like the subject, verb or predicate, and object of a sentence. In RDF, a document makes assertions that particular things (people, Web pages or whatever) have properties (such as "is a sister of," "is the author of") with certain values (another person, another Web page). This structure turns out to be a natural way to describe the vast majority of the data processed by machines. Subject and object are each identified by a Universal Resource Identifier (**URI**), just as used in a link on a Web page. (URLs, Uniform Resource Locators, are the most common type of URI.) The verbs are also identified by URIs, which enables anyone to define a new concept, a new verb, just by defining a URI for it somewhere on the Web.

By using XML, RDF information can easily be exchanged between different types of computers using different types of operating systems and application languages.

Uniform Resource Identifier (URI)
A URI is used to identify things on the Web. A URL (Uniform Resource Locator or Web address) is a type of URI.

SPARQL
The Simple Protocol and RDF Query Language (SPARQL) is a SQL-like language for querying RDF data.

Triples of the Data Model

Subject	Predicate	Object
http://www.w3schools.com	http://www.w3schools.com/rdf/title	"W3Schools.com"
http://www.w3schools.com	http://www.w3schools.com/rdf/author	"Jan Egil Refsnes"

The original RDF/XML document

```
1: <?xml version="1.0"?>
2: <rdf:RDF
3: xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
4: xmlns:si="http://www.w3schools.com/rdf/">
5: <rdf:Description rdf:about="http://www.w3schools.com">
6: <si:title>W3Schools.com</si:title>
7: <si:author>Jan Egil Refsnes</si:author>
8: </rdf:Description>
9: </rdf:RDF>
```

Graph of the data model

See an example of RDF code (above) and how it is used - http://www.w3schools.com/rdf/rdf_example.asp

---

Languages and Vocabularies

Computers do not understand what word means, and what the relationships between words mean, like we do. Therefore, the computer has to have documents to describe the words and make the connections. This is made possible by schema and ontologies.

Schema – a method for organizing information

Ontology – a vocabulary that describes the objects and how they relate to each other

RDF Schema

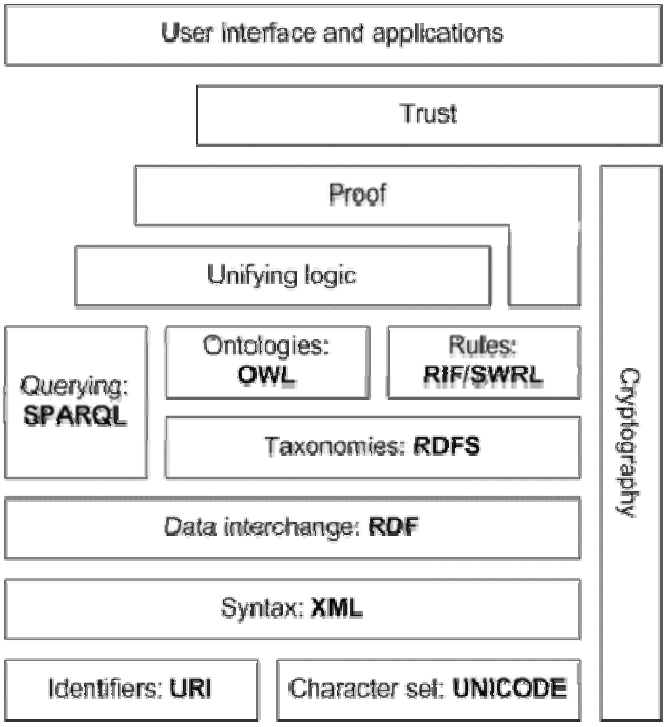
Extends the RDF language by the addition of classes and properties for better information descriptions.

Web Ontology Language - OWL - http://www.w3.org/TR/owl-features/

The OWL Web Ontology Language is designed for use by applications that need to process the content of information instead of just presenting information to humans. OWL builds on RDF and RDF Schema and adds more vocabulary for describing properties and classes. OWL has three versions: OWL Lite, OWL DL, and OWL Full.

- define meaningful relationships between different types of data
- define specific data type to generic data
- make queries and assertions between different types of data
- provide a database of information that defines relationships between different terms

The Semantic Web Stack or Architecture



It takes layers and layers of metadata, logic and security to make the Web machine-readable. Most visual representations of these layers involve a stack -- sort of a tower of blocks that represent all the layers. The stack changes and evolves as the concepts behind the Semantic Web develop

Merging with Web 2.0

The Semantic Web is based around ontologies, which are large and hard to use. Folksonomies have emerged from the Social Web. Ontologies are structured and Folksonomies are user-generated tags. They both are used to describe and give meaning to data. We may see a way to incorporate these two in the future of the Semantic Web.

Using Semantics in the XHTML Document

RDFa – see http://www.w3.org/TR/xhtml-rdfa-primer/

RDFa (or Resource Description Framework - in - attributes) is a set of extensions to XHTML which is now a W3C Recommendation. RDFa uses attributes from XHTML's meta and link elements, and generalizes them so that they are usable on all elements. This allows you to annotate XHTML markup with semantics. A simple mapping is defined so that RDF triples may be extracted.

Examples of some RDFa code found in websites:

```
<a  rel="license"  href="http://creativecommons.org/licenses/by/3.0/us/">Creative Commons License</a>
```

@property is an attribute introduced by RDFa for the specific purpose of marking up existing text in an XHTML page. In the below example, we indicate the title and author of a book.

```
<div  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <h2  property="dc:title">The trouble with Bob</h2>
  <h3  property="dc:creator">Alice</h3>
 </div>
```

 see video http://www.youtube.com/watch?v=ldl0m-5zLz4

Microformats

The official microformats.org web site defines microformats thus:

"Designed for humans first and machines second, microformats are a set of simple, open data formats built upon existing and widely adopted standards."

Microformats allow us to embed semantics into the HTML/XHTML document. They use current XHTML tags such as address, cite, and blockquote and attributes such as rel, rev, and title.

For code samples and tools to generate microformats see: http://microformats.org and http://microformats.org/code-tools/