

MARCH 2007

THE NEW FACE OF THE WEB: SEMANTIC WEB

TECHNICAL COMMUNICATION REPORT

BY CATHERINE RACETTE

ABSTRACT

The Web is an integral part of daily life. Through the use of HTML, XML, and search engines, it is possible to locate information from indexed documents (currently estimated at 10 billion pages), across the Web in a meaningful way...for people. The Web, in its current incarnation, however, is *document* focused rather than *data* focused. Integration between web page data and data in other software applications is extremely limited, and Semantic Web will change that. Technical communicators who produce content for the Web, develop for the Web or even do research through the Web will experience a new Web that connects web pages, applications, and a variety of data sources in a cohesive, *meaningful* way, not only to people, but to software applications and databases as well.

This report was compiled from internet research

through Google and the NJIT Research Navigator online tool. Resulting documents used were articles, interviews, wikis, and blogs, which provided information about the structure, goals and outlook for Semantic Web. Because more widespread implementation of Semantic Web may be years away, this will not have an immediate impact on the technical communicator, but it will change the future of technology and how we use the Web to do our work.



WHAT IS SEMANTIC WEB?

Much of the current discussion about the future of the Internet centers on how “Web 2.0” is changing the web by providing means for social networking and collaboration among web users. As a frame of reference, Semantic Web is sometimes referred to as “Web 3.0.”

The Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation (Berners-Lee, Hendler and Lassila, 2001). The goal of the Semantic Web initiative is to create a universal medium for the exchange of data where data can be shared and processed by automated tools as well as by people.

INSIDE THIS REPORT

ABSTRACT	1
WHAT IS SEMANTIC WEB?	1
HOW WILL THE SEMANTIC WEB BE USEFUL?	2
USING EXISTING CONTENT	2
SEMANTIC WEB COMPONENTS	3
CONCLUSION	3
DESCRIPTION OF TERMS	4

HOW WILL THE SEMANTIC WEB BE USEFUL?

Under the current Web structure, a user can perform a search for the web page, address, telephone number, map location, etc. for the nearest auto mechanic. S/he can use the information found to call the mechanic to schedule an appointment, or if the web site has an online scheduling tool, an appointment can be made online. However, the user must then type or copy/paste the information into his/her calendar on the computer. The calendar application on the web and the calendaring software on the computer do not understand each other. Using the Semantic Web framework, the data on the computer and the data on the web would both be recognized as “appointment” data and process the appointment and calendar entry for the user in both systems. Because the machines now *understand* that data, they can process tasks for the user.

This is just one example, but the impact of Semantic Web increases the machine’s understanding of data to perform tasks they cannot do today in the current framework.

Today’s Web is oriented toward presentation to people: this is the purpose of html tags <bold>PRINT THIS IN BOLD</bold> on Web pages. After HTML came Extensible Markup Language (XML), which separates the presentation from the meaning <LastName>Obrst</LastName>. However, to a computer this is the same as <% @&^>Obrst</% @*&^>. A machine needs something that helps interpret what the tags mean. That’s where “machine-interpretable” semantics comes in.*

-- Frank Manola, Mary Pulvermacher, and Lero Obrst (2004)

USING EXISTING CONTENT

HTML uses code to set the layout of web pages to make it understandable to humans; it enables the web browser to display the page properly. However, HTML tags don’t make it easier to access the data stored in the documents. The tags tell how a piece of information is laid out on a page, relative to other pieces of information on that page, but it doesn’t relate the data on the page to data in other systems.

XML makes this a little easier, but “for every pair of applications, in fact for each way in which they need to be linked, someone has to create an XML to XML bridge. That is, if you take XML files from two different applications, you can’t just merge them.” (Handler, Berners-Lee, Miller 2002).

The point of Semantic Web is that by using RDF (*see definitions on pp. 4 & 5*), applications will generate the data that is readable by machines that understand the relationship with other data in other systems. There’s no need to manually code information into web pages or systems; these will be built in applications that use the Semantic Web framework. No special interface will have to be programmed to allow data to be used across systems in a meaningful way.

SEMANTIC WEB COMPONENTS

The Semantic Web comprises the standards and tools of XML, XML Schema, RDF, RDF Schema and OWL. The OWL Web Ontology Language Overview describes the function and relationship of each of these components of the Semantic Web:

XML provides a surface syntax for structured documents, but imposes no semantic constraints on the meaning of these documents.

XML Schema is a language for restricting the structure of XML documents.

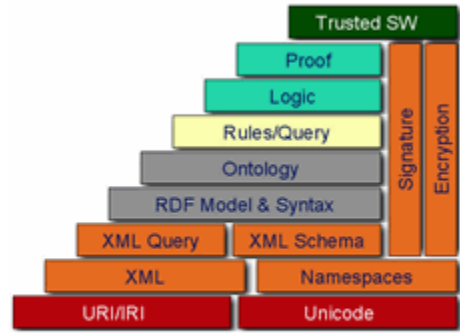
RDF is a simple data model for referring to objects ("resources") and how they are related. An RDF-based model can be represented in XML syntax.

RDF Schema is a vocabulary for describing properties and classes of RDF resources, with a semantics for generalization-hierarchies of such properties and classes.

OWL adds more vocabulary for describing properties and classes: among others, relations between classes (e.g. disjointness), cardinality (e.g. "exactly one"), equality, richer typing of properties, characteristics of properties (e.g. symmetry), and enumerated classes.

Range (2006)

See Also *Description of Terms* on page 4.



W3C Semantic Stack

CONCLUSION

It may be five to seven years before Semantic Web affects the daily workings of technical communicators and others. It will present an opportunity for professionals to integrate disparate systems. Those who are responsible for creating web content or documenting system processes will surely be impacted by Semantic Web. Ontologies will be developed in the meantime and work by the World Wide Web Consortium (W3C) will ensure that this "Web 3.0" comes to fruition. It will impact every computer user in the world and will change the way we present data and information as communicators.

*"The Semantic Web idea -- that of having data as well as documents on the web -- has been around since the start of the web. It is just more complicated to do.
-- Tim Berners-Lee (2005)*

DESCRIPTION OF TERMS

HyperText Markup Language (HTML): is the predominant markup language for the creation of web pages. It provides a means to describe the structure of text-based information in a document — by denoting certain text as headings, paragraphs, lists, and so on — and to supplement that text with interactive forms, embedded images, and other objects. HTML is written in the form of labels (known as tags), created by less-than signs (<) and greater-than signs (>). HTML can also describe, to some degree, the appearance and semantics of a document, and can include embedded scripting language code which can affect the behavior of web browsers and other HTML processors.

Extensible Markup Language (XML): is a W3C-recommended general-purpose markup language that supports a wide variety of applications. XML languages or 'dialects' may be designed by anyone and may be processed by conforming software. XML is also designed to be reasonably human-legible, and to this end, terseness was not considered essential in its structure. XML is a simplified subset of Standard Generalized Markup Language (SGML). Its primary purpose is to facilitate the sharing of data across different information systems, particularly systems connected via the Internet.

Resource Description Framework (RDF): is a language for representing information about resources in the World Wide Web. The RDF metadata model is based upon the idea of making statements about resources in the form of subject-predicate-object expressions, called triples in RDF terminology. The subject denotes the resource, and the predicate denotes traits or aspects of the resource and expresses a relationship between the subject and the object. For example, one way to represent the notion “The sky has the color blue” in RDF is as triple of specially formatted strings: a subject denoting “the sky,” a predicate denoting “has the color,” and an object denoting “blue.”

Uniform Resource Identifier (URI): is a compact string of characters used to identify or name a resource. The main purpose of this identification is to enable interaction with representations of the resource over a network, typically the World Wide Web, using specific protocols. URIs are defined in schemes defining a specific syntax and associated protocols.

Web Ontology Language (OWL): is designed for use by applications that need to process the content of information instead of just presenting information to humans. OWL is a language for defining and instantiating Web ontologies.[1] An OWL ontology may include descriptions of classes, along with their related properties and instances. OWL is designed for use by applications that need to process the content of information instead of just presenting information to humans. It facilitates greater machine interpretability of Web content than that supported by XML, RDF, and RDF Schema (RDF-S) by providing additional vocabulary along with a formal semantics. OWL is based on earlier languages OIL and DAML+OIL, and is now a W3C recommendation (i.e., a standard).

OWL was designed to provide a common way to process the semantic content of web information. It was developed to augment the facilities for expressing semantics (meaning) provided by XML, RDF, and RDF-S. Consequently, it may be considered an evolution of these web languages in terms of its ability to represent machine-interpretable semantic content on the web. Since OWL is based on XML, OWL information can be easily exchanged between different types of computers using different operating systems, and application languages. Because the language is intended to be read by computer applications, it is sometimes not considered to be human-readable, although this may be a tool issue. OWL is being used to create standards that provide a framework for asset management, enterprise integration, and data sharing on the Web.

WORKS CITED

- Berners-Lee, T., Hendler, J., & Miller, E. (2002). *Integrating Applications on the Semantic Web*. Retrieved March 17, 2007, from <http://www.w3.org/2002/07/swint>
- Berners-Lee, T., Hall, W., & Shadbot, N. (2006). *The Semantic Web Revisited*. IEEE Intelligent Systems. Retrieved March 18, 2007, from <http://www.computer.org/intelligent>
- Edmonds-Banfield, P. (2006). *Building the Semantic Web*. Orange Journal. Retrieved March 18, 2007 from <http://orange.eserver.org/issues/3-2/emonds-banfield.html>
- Manola, F., Obrst, L., & Pulvermacher, M. (2004). *The Semantic Web: A Path to Large-Scale Interoperability*. The Edge. Retrieved March 20, 2007 from http://www.mitre.org/news/the_edge/summer_04/manola.html
- Range (2006). *Semantic Web or Web 3.0 defined*. Viral One blog. Retrieved March 19, 2007 from <http://viralone.wordpress.com/2006/07/17/semantic-web-or-web-30-defined/>
- Updegrove, A. (2005). *The Semantic Web: An Interview with Tim Berners-Lee*. Retrieved March 17, 2007 from <http://www.consortiuminfo.org/bulletins/semanticweb.php>
- Wikipedia entries. Retrieved March 20, 2007 from <http://www.wikipedia.org>
- Description of Terms retrieved March 20, 2007 from <http://www.semanticwebfaq.com>, <http://www.w3c.org>, & <http://www.wikipedia.org>