THE SEMANTIC WEB

Hesam Panahi

Although use of the World Wide Web has dramatically expanded over the past decade, the underlying technologies have remained essentially the same. The fundamental problem with the web's current structure lies in its focus on information that is readable only by human beings. The shortage of machine-readable data increases the amount of human power necessary to organize and locate useful information.

The next generation of the web, the **Semantic Web**, promises to bring increased meaning to online content by incorporating concepts such as Resource Description Framework (RDF), eXtensible Markup Language (XML), and Web Ontology Language (OWL). This paper outlines how these technologies will function and their potential applications in an organizational setting.

A Possible Scenario

Acme Company's CIO, Steve, has recently implemented Enterprise Semantic Web (ESW) technology in his organization. When the company's ERP system determines that there is a low level of inventory on a key product, it launches a software agent that searches through pre-approved parts supplier databases to find the lowest priced parts. Upon determining where to purchase the parts, the software agent places an order through the supplier's system. A shipping invoice is prepared and sent by the software agent to the accounts receivable module. When the parts arrive at one of the Acme Company's warehouses, they are automatically scanned and added to inventory. The ERP receives notification that the parts has been entered into inventory, and schedules payment to the suppliers within the payment terms. This is all done without human intervention.[1]

It was recently estimated that there are more than 11.5 billion indexed pages on the World Wide Web (Gulli & Signorini 2005). This leaves us to wonder, "How do humans manage this sea of information?"

People, of course, use search engines to assist them in locating information. However, search results can be irrelevant, requiring additional searches to capture vital data. In many instances, searches end in failure and frustration. To address these problems, Google, Yahoo! and other search engine developers continue to fine tune their search algorithms to further improve the user experience. While this addresses the issue of usability, it fails to account for a major limitation -- most of the content available on the web is intended to be readable by people not "machines".

Tim Berners-Lee, inventor of the web, describes the problem:

> "The Web was designed as an information space, with the goal that it should be useful not only for human-human communication, but also that machines would be able to participate and help. One of the major obstacles to this has been the fact that most information on the Web is designed for human consumption, and even if it was derived from a database with well defined meanings (in at least some terms) for its columns, that the structure of the data is not evident to a robot browsing the web. (Berners-Lee 1998)

To improve functionality and efficiency, the next generation of the web must address this issue. Data must be given enhanced meaning (semantics) that machines can process without human intervention. Relationships among sets of data must also be recognizable. The result will be more relevant data, available in a timely manner. The term "semantic web" is intended to capture this goal.

What is the Semantic Web?

The World Wide Web is currently composed of individual pages written with HTML (HyperText Markup Language). HTML is primarily used to dictate the **presentation** and **layout** of the page; it does not provide information about the **contents** of the page. Documents are created in Unicode, a plain text format allowing for international characters. In addition, each page created has an URI (Uniform Resource Identifier), such as a web address. Users can input this URI into a user agent (i.e. Web browser) to locate the page. Pages can also have links to other URIs, effectively creating a "web" of information.

For machines to be able to interpret these pages more efficiently, Web page creators must add **metadata** to Web documents. Metadata is data about data. Creating metadata is a two-step process: metadata must first be added, and then defined. By including metadata in Web documents, web developers are structuring the document in a format that machines can understand. The primary method of creating metadata is through use of the eXtensible Markup Language (XML).

Once metadata has been added to a document using XML, the metadata is defined using the Resource Description Framework, or, RDF. RDF allows one to describe the document and the data it contains. RDF also provides the ability to share these descriptions on a global scale. Other pages can use RDF documents to retrieve information about the structural contents (metadata) of a document, thus allowing for interoperability across web applications. Although RDF is essentially XML-based, these added capabilities make it a crucial component of the Semantic Web.

As web developers use RDF to create definitions for metadata, the possibility exists for developing different domains of knowledge. A common vocabulary would also be beneficial if it were available for machines on the Semantic Web. For this reason, a language for creating ontologies on the web, the Web Ontology Language (OWL), was created. An ontology is a formal means of expressing concepts and relationships between them (Daconta, Orbst, & Smith 9). Ontologies are used in fields such as Artificial Intelligence (AI), because they allow machines (agents) to communicate with each other even when they are not in the same domain of knowledge (Gruber 1992).

> "The organization that has the best information, knows where to find it, and can utilize it the quickest wins."
>
> -Daconta, Orbst, & Smith 2003

These three levels of knowledge management, XML, RDF, and OWL, form the basis on which the Semantic Web exists.

eXtensible Markup Language (XML)

eXtensible Markup Language (XML) and HyperText Markup Language (HTML) were both derived from

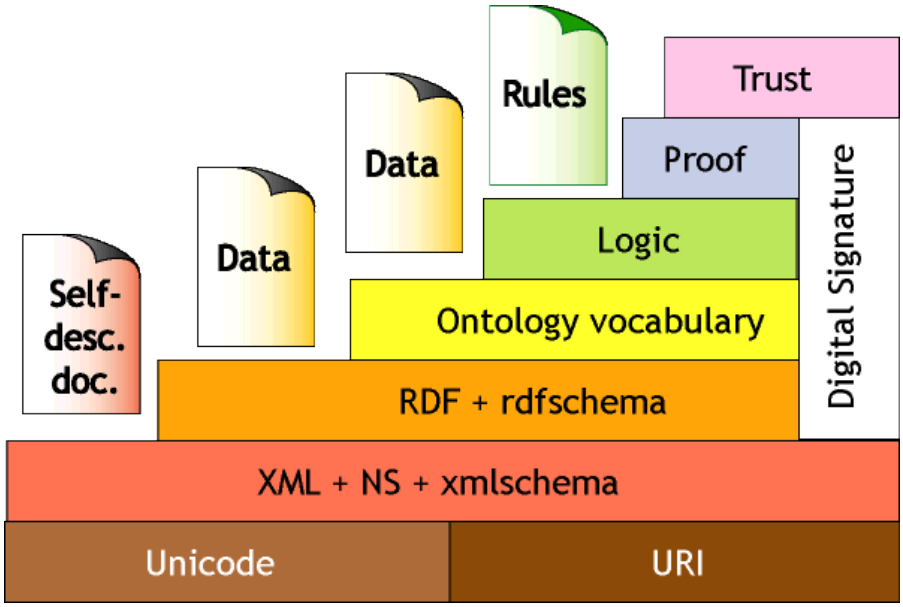

Figure 1. The layers of the Semantic Web (Berners-Lee 2000)

a more complex language, Standard Generalized Markup Language (SGML) (Antoniou & van Harmelen 23). As a result, the syntax of both is similar. However, HTML is limited to providing presentational and layout properties of a document, whereas XML can denote structural information as well.

XML itself is not a language; rather, it is the foundation on which semantic markup languages are created to be applied to a document (Daconta, Orbst, & Smith 32). To define the structure of these documents, XML Schema is used (Antoniou & van Harmelen 37). XML Schema is a definition language that dictates which elements can exist in a document's structure. If a document is said to be conforming to an XML Schema, then it will comply with the rules that have been set forth by the XML Schema (Geroimenko & Chen 7). As a result, XML Schemas provide for validation of data that is in a document. (Daconta, Orbst, & Smith 38).

In addition to adhering to an XML Schema, XML documents on the Semantic Web must also include references to at least one XML namespace. XML namespaces are used to define the origin of different names in an XML document. (Bray 1999) For example, a document may contain information about a book available on a website. Part of the document has the title of the book as well as other relevant information, such as author and publication date. Another section of the document contains information about the web page, such as page title, layout, and appearance. In this case, there may be two elements inside this segment that are considered to be a "title" (i.e. the title of the book and the title of the web page). A machine must be able to distinguish between these two items and know which one to process. By having a different namespace for each section (book and web page), the computer can be told which "title" to look for when searching for data. (Bray 1999)

> The elements in an XML document are organized in a hierarchy whereby elements are nested within other elements, creating a parent-child relationship.

To add further structure to a document, the elements in an XML document are organized in a hierarchy whereby elements are nested within other elements, creating a parent-child relationship. An example of how an XML document may be viewed conceptually can be seen in Figure 2.

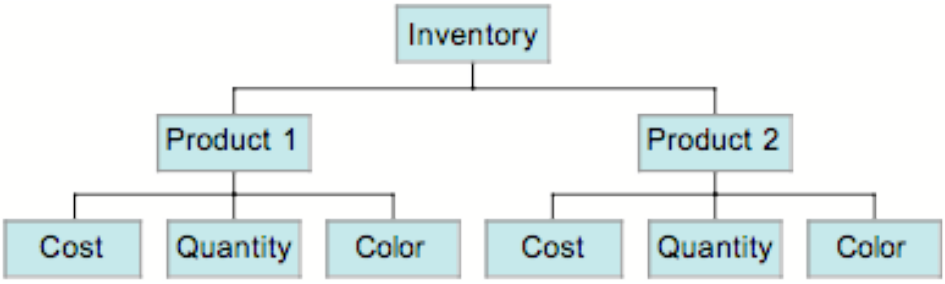

Figure 2. A visual representation of a typical XML document

```
<?xml version="1.0" encoding="UTF-8"?>
<inventory>
   <product1>
      <cost>5.00</cost>
      <quantity>200</quantity>
      <color>blue</color>
      <color>red</color>
   </product1>
...
</inventory>
```

Figure 3. Product 1's information from Figure 2 in XML format

Resource Description Framework (RDF)

Resource Description Framework (RDF) is used to convey information about one or more resources (Daconta, Orbst, & Smith, 86). For the Semantic Web, RDF's primary purpose is to describe metadata. RDF can be considered an enhanced version of XML, in that it is more scalable for global applications (Bray 2001).[2] In addition, RDF is considered more appropriate for describing the external attributes of a document (i.e. author of document and date of creation), whereas XML is used for the actual content within the document (i.e. inventory cost and quantity).

RDF syntax is composed of three parts: the resource, the property, and the statement (Antoniou & van Harmelen 63). The resource is the object that is being described, while the property is the attribute of that resource. The resource, the property, and a corresponding value for the property are used to form the RDF statement.

An example of this syntax follows: when transforming "www.ab-supply.com/parts-list/ is authored by Jim" into an RDF statement, www.ab-supply.com is the resource, author is the property, and Jim is the value. In most cases, the resource is the subject of a statement, the property is the predicate, and the value is the object. (Bray 2001)

As in using basic XML, restrictions can be added to RDF statements using RDF schema. Restrictions can be as simple as stating that the author property may only be used when describing web pages. Therefore, RDF schema provides a method for validation and data integrity. (Wrox Press 2001)

Web Ontology Language (OWL)

As the Semantic Web becomes more commonplace, different domains of knowledge will appear. Each domain will have its own description, individual concepts, and relationships among concepts (Daconta, Orbst, & Smith 182). For example, one could consider an organization a domain of knowledge. Within the organization, there are different concepts (department, employee). Each of those concepts is related to each other and these relationships must be defined (i.e. a department consists of employees; an employee manages another employee).

Although RDF and RDF Schema have some of the capabilities of an ontology, they lack the richer features that are needed for reasoning (Antoniou & van Harmelen 111). The main requirements of an ontology language are a well-defined syntax, efficient reasoning support, formal semantics, sufficient expressive power, and the convenience of expression (Antoniou & van Harmelen 110). In an attempt to standardize an ontology language for the Semantic Web, a group of professionals who oversee the development of new online standards known as the World Wide Web Consortium (W3C), have created have created the Web Ontology Language (OWL).

OWL offers many of the features that are required for rich expression of data on the Semantic Web. It can be used across systems, is an open standard, and is compatible with current web standards (W3C 2005a). The development of OWL is now completed and ready for implementation.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:page="http://www.ab-supply.com/page-syntax-ns#">

<rdf:Description
   rdf:about="http://www.ab-supply.com/parts-list/">
   <page:author>Jim</page:author>
   <page:company>AB Supply Company</page:company>
   <page:lastrevision>October 27, 2005</page:lastrevision>
</rdf:Description>
...
</rdf:RDF>
```

Figure 4. "www.ab-supply.com/parts-list is authored by Jim" in RDF format

Web Services

The primary function of web services is to allow different applications to communicate with each other (Daconta, Orbst, & Smith 57). Web services use XML-based languages to make it possible for applications to exchange data. Interoperability can occur even if the applications are on a different system platform or framework (W3C 2005b).

Web services add three concepts to the existing infrastructure of the web: SOAP, WSDL, and UDDI. SOAP (Simple Object Access Protocol) is the XML-based language that is used to communicate among different applications. WSDL (Web Service Definition Language) provides information about a service, including location, which operations it can perform, and how to send messages to the service (Daconta, Orbst, & Smith 68). UDDI (Universal Description, Discovery, and Integration) provides web services with the ability to locate and connect to each other. (Vasudevan 2001)

The introduction of web services addresses one of the primary problems in previous system implementations. As companies purchased applications for each department, the exchange of data among these systems was difficult or costly. In addition, companies in partnerships with each other encountered incompatibility issues when attempting to exchange data. By having a common language in which these systems can interact, web services eliminate communication barriers and can increase the flow of information. In fact, a 2001 study predicted that by this year, web services would reduce IT costs and increase efficiency by 30 percent (Plummer & Andrews 2001).

A 2001 study predicted that by the end of this year, web services will reduce IT costs and increase efficiency by 30 percent.

Applications of the Semantic Web

As mentioned earlier, the World Wide Web Consortium (W3C) is a group of professionals that develop and recommend standards for use on the web. Many companies are members of the W3C including computing leaders such as IBM, Hewlett Packard, and AOL. Furthermore, a number of firms whose business is not primarily based on computers and the Internet, such as Daimler-Chrysler, Nokia, and Phillips, are also involved in the process (van Harmelen 2003). The support of these companies suggests that the Semantic Web offers in for a number of industries.

Daimler-Chrysler has already started using the Semantic Web in its organization. Parts that are to be assembled into vehicles are now ordered online. By using XML based services, Daimler-Chrysler is able to ensure interoperability with vendors' systems. The result is the elimination of unnecessary paper contracts and invoices. Instead of long-term contracts with a predetermined set of suppliers, Daimler-Chrysler is able to dynamically create short-term agreements with vendors able to offer the lowest prices. Switching vendors can be done relatively quickly, thus creating a more competitive open market with greater cost savings for Daimler-Chrysler (Antoniou & van Harmelen 200).

Hewlett-Packard is also experimenting with using Semantic Web technology for printer configuration.

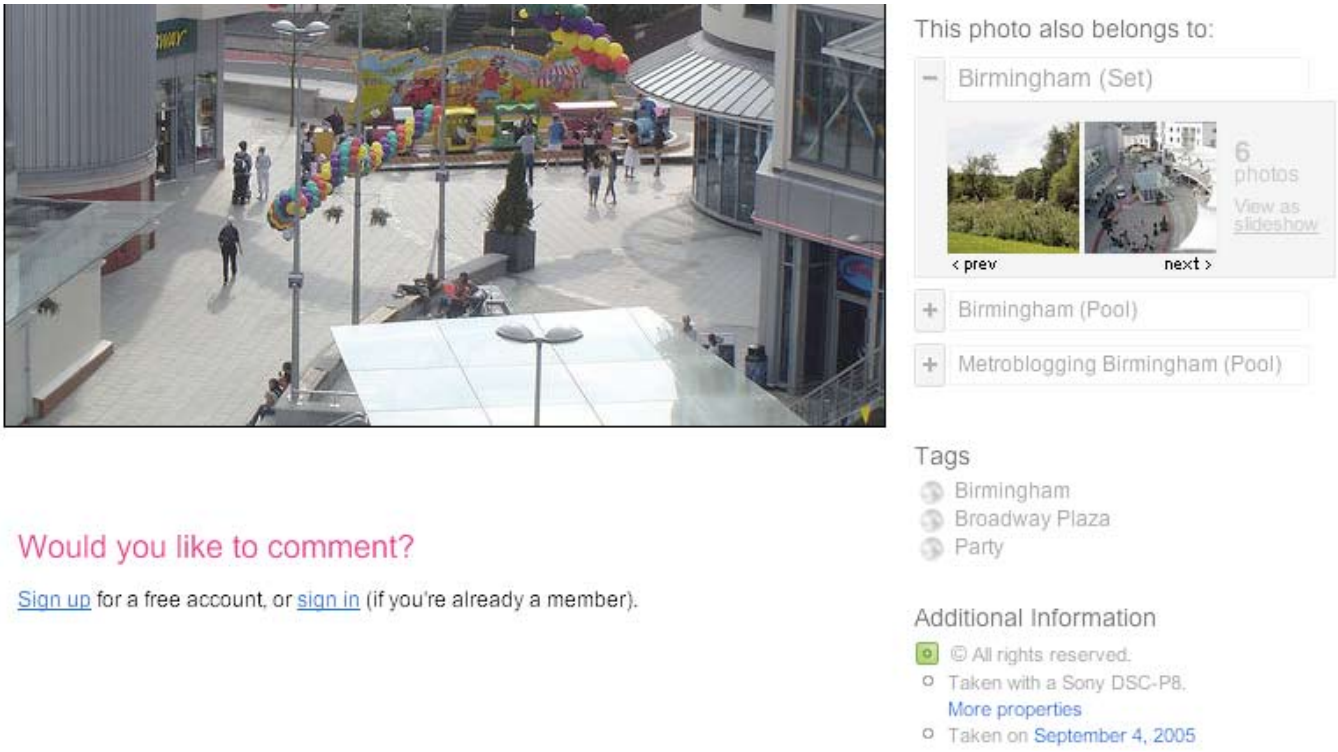

Figure 4. An image page on Flickr

Figure 5. Mappr

Each printer could potentially become a "self-describing device" (van Harmelen 2003). Walk into a building with your laptop and any available HP printers will automatically identify themselves, via a profile, to your computer. This profile will be written in a Semantic Web language to ensure interoperability. When you want to print, your device will be able to immediately determine the closest appropriate printer for the print job (van Harmelen 2003). In this case, machines are communicating with each other and performing tasks previously done manually.

A number of businesses are now offering to implement Semantic Web technologies for others. Cerebra is a company that offers "semantic technology for the enterprise" (Cerebra 2005a). Using RDF, OWL, and web services, Cerebra can assist in the deployment of decision-making systems that reason logically and make choices automatically. This operation is similar to business process management and business rule engines; however, Cerebra's solutions are more scalable and built on an open standard. Cerebra uses the term "Enterprise Semantic Web" (ESW) to describe a semantic web for use within an enterprise. Included among the many benefits Cerebra cites are faster maintenance, lower costs, and increased awareness in the decision making process. (Cerebra 2005b)

Although some applications of the Semantic Web have begun to appear in organizations, online communities have been the first to embrace this technology. One such example of semantic concepts is Flickr (http://www.flickr.com). Flickr is an online image repository where users can upload and share photos. Photos can be tagged with metadata (such as the camera used, the location where it was taken, the people or objects represented). Once tagged, users can then search for images, put them into groups (pools) with similar photos from other users, or share comments. When viewing a tagged photo, following the link of any of the tag names will result in the display of more photos with that tag from other users. By adding metadata, users are providing information that can help in organizing and locating photos.

Flickr is also a web service; developers can use the metadata provided by Flickr in their own applications. An example of this is Mappr (http://www.mappr.com). Mappr is a service that allows users to view photos by geographic location. Mappr visitors are presented with a map of the United States that is clickable; users input a tag into the search box and Mappr displays where images with the tag are located. Choosing an image brings up a list of its tags as well as Mappr's "confidence" that the image was actually taken in that location. Confidence is dependent on an image's tags; if there are multiple tags that would lead Mappr to believe an image is a certain location (for example, Ohio) then the confidence is higher (Mappr 2005). All of this is done through the Mappr user interface; interaction with Flickr web services is completed in the background.

Conclusion

Implementation of the Semantic Web is still in its early stages. Converting much of the World Wide Web to one with semantic data is a long, cumbersome process that requires support from industry, academia, and others (Clark 2003). Whether it will ever see widespread adoption remains to be seen, as various obstacles remain. The specifications for the three levels above ontology (trust, proof, and logic) in the Semantic Web have not yet been finalized (Koivunen & Miller 2001). Moreover, although the Web Ontology Language (OWL) exists, specialized ontologies will likely be needed for various industries. (Ohlms 2002).

Regardless, the potential of the Semantic Web is obvious. The Semantic Web will supplement many current IT trends, including information management, system integration, multi-device capability, e-procurement, and CRM. (Ohlms 2002). Knowledge management is an important tool for organizations, and applying metadata to information can increase the like-

lihood of the availability of relevant data when needed. Decision support systems can possess increased accuracy as a result of interoperability among systems built on open standards. Web services will make it possible for developers to draw data from other applications and present it to users with their own interface. This can be useful for executives who need to be presented with information from multiple sources simultaneously.

Many of these technologies could actually be employed within an organization today. In fact, the standards that are to be used in the Semantic Web are by and large complete; therefore, companies can begin using XML, RDF, and OWL in an intranet solution or with business partners. In doing so, organizations can bring much-needed additional meaning to the wealth of information already in their knowledge base.

Footnotes

[1] Adapted from Alper, 2005

[2] For more information about why RDF is preferred over XML for describing metadata, please visit http://www.xml.com/pub/a/2001/01/24/rdf.html?page=2#why_not_xml

References

Alper, Alan. "Disruptive Technologies: Semantic Web", 2005, available online at http://www.managingautomation.com/maonline/magazine/read/3571723?printable.

Antoniou, G., and van Harmelen, F. A Semantic Web Primer, The MIT Press, Cambridge, 2004.

Berners-Lee, T. "Semantic Web Road Map," 1998, available online at http://www.w3.org/DesignIssues/Semantic.html.

Berners-Lee, T. "Semantic Web – XML2000," 2000, available online at http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html.

Bray, T. "XML Namespaces by Example," 1999, available online at http://www.xml.com/pub/a/1999/01/namespaces.html.

Bray, T. "What is RDF?," 2001, available online at http://www.xml.com/pub/a/2001/01/24/rdf.html.

Cerebra. "Cerebra," 2005a, available online at http://www.cerebra.com.
Cerebra. "Enterprise Semantic Web," 2005b, available online at http://www.cerebra.com/downloads/brochure_ESW.pdf.

Clark, K. "Commercializing the Semantic Web," 2003, available online at http://www.xml.com/pub/a/2003/10/22/iswc.html.

Daconta, M., Orbst, L., and Smith, K. The Semantic Web, Wiley Publishing, Indianapolis, 2003.

Geroimenko, V., and Chen, C. Visualizing the Semantic Web, Springer, London, 2003.

Gruber, T. R. "Toward principles for the design of ontologies used for knowledge sharing," 1992, available online at ftp://ftp.ksl.stanford.edu/pub/KSL_Reports/KSL-93-04.ps.gz.

Gulli, A. and Signorini, A. "The Indexable Web is More than 11.5 billion pages," 2005, available online at http://www.cs.uiowa.edu/~asignori/web-size/size-indexable-web.pdf.

Koivunen, M. and Miller, E. "W3C Semantic Web Activity," 2001, available online at http://www.w3.org/2001/12/semweb-fin/w3csw.

Mappr. "About Mappr," 2005, available online at http://www.mappr.com/about/.

Ohlms, Christian. "The Future of the Semantic Web," 2002, available online at http://www.aifb.uni-karlsruhe.de/AIK/veranstaltungen/aik9/presentations/slides/020419FutureSemanticWeb.pdf.
Plummer, D. and Andrews, W. "The Hype Is Right: Web Services Will Deliver Immediate Benefits," 2001, available online at http://www.gartner.com/DisplayDocument?doc_cd=101571.

Van Harmelen, F. "The future of the search engine: 'Search and you will find'. An interview with Frank Van Harmelen," 2003, available online at http://lgxserve.ciseca.uniba.it/lei/ai/networks/03-2/vanharmelen-en.pdf.

Vasudevan, V. "A Web Services Primer," 2001, available online at http://webservices.xml.com/pub/a/ws/2001/04/04/webservices/.

W3C. "Web Ontology Language (OWL)," 2005a, available online at http://www.w3.org/2004/OWL/.

W3C. "Web Services Activity Statement," 2005b, available online at http://www.w3.org/2002/ws/Activity.

Wrox Press. "Professional XML Metadata," 2001, available online at http://www.perfectxml.com/wp/readchap.asp?CurBook=29.

Contact Information

Hesam Panahi
C.T. Bauer College of Business
University of Houston
hpanahi@uh.edu