## Chapter II

# Enabling Technologies for the Semantic Web

Kevin R. Parker, Idaho State University, USA

***Editors' Notes***

Kevin had a clear mission: to provide the Semantic Web notion for everyone. While several of the aspects of the Semantic Web will be explained further in other chapters, readers unfamiliar with the Semantic Web issues should start thinking of the importance of the Semantic Web toward more effective intelligent knowledge and learning infrastructures. We put this chapter after the corporate learning environment chapter for obvious reasons. We want to converge two pillars of critical importance: on the one hand, leading and state-of-the-art research on theoretical foundations of next generation knowledge and learning management; and on the other hand, leading edge technologies as those of Semantic Web.

This objective is diffused in every chapter of the book. We want theories and technologies to be applied in specific contexts toward the development of socio-technical systems aiming to provide a knowledge and learning driven performance.

It is again worth mentioning our involvement in the Special Interest Group on Semantic Web and Information Systems of the Association for Information Systems (*http://www.sigsemis.org*). We encourage you to visit our portal and consider becoming part of this community. An excellent point of reference for issues related to the Semantic Web is the AIS SIGSEMIS Bulletin, the official quarterly newsletter of the AIS SIGSEMIS where research papers, research center presentations, and interviews of the leaders of SW provide excellent knowledge for the field. Moreover, the International Journal on Semantic Web and Information Systems published by IDEA Group Publishing, *http://www.idea-group.com*, sponsored by AIS SIGSEMIS, provides leading edge research outcomes. It is an excellent addition to your portfolio of scientific journals.

# Abstract

*Before understanding the Semantic Web and its associated benefits, one must first be somewhat familiar with the enabling technologies upon which the Semantic Web is based. The extensible markup language (XML), uniform resource identifiers (URIs), resource definition framework (RDF), ontologies, and intelligent agents are all key to the realization of the Semantic Web. Understanding these key technologies gives readers a firm foundation before progressing to subsequent chapters. This chapter provides a broad overview of each technology, and readers new to these technologies are provided with references to more detailed explanations.*

# Introduction

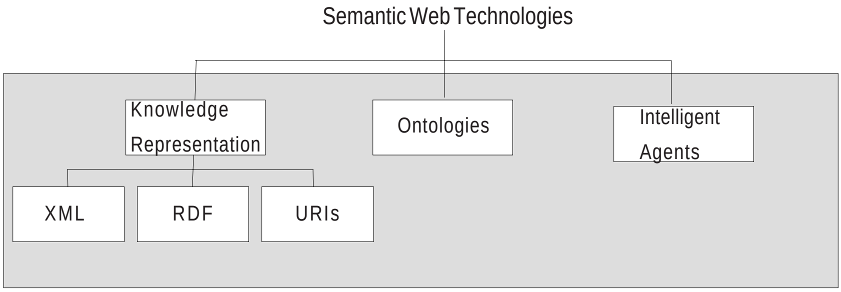
In its current form the Web makes effective searching and data exchange difficult. Today's Web pages are designed for human use, and human interpretation is required to understand the content. The Web lacks the ability to automatically link documents on the basis of semantic similarities because content is not machine-interpretable (Lassila, 2002). This means that such linking requires human intervention to ascertain the semantic context and recognize similarities between documents. *The American Heritage Dictionary* (2000) defines semantics as the "meaning or the interpretation of a word, sentence, or other language form." The critical term in this definition is *meaning*, and meaning requires understanding. Although computer software applications are unable to truly understand information, there are approaches that make it possible for applications to manipulate data "in ways that are useful and meaningful to the human user" (Berners-Lee, Hendler, & Lassila, 2001, p.40). The Semantic Web is an extension of today's Web in which documents are annotated in such a way that their semantic content is optimally accessible and comprehensible to automated software agents and other computerized tools. Thus, documents can be automatically linked on the basis of semantic similarities, eliminating the need for human reasoning to determine the meaning of Web-based data (Bonner, 2002).

There are several key technologies upon which the Semantic Web is reliant. The objective of this chapter is to introduce the reader to these concepts and to provide a brief overview of each (see Figure1).

# Background

Documents in the Semantic Web contain not only content, but also context. The meaning of the content is clearly specified so that documents can be linked to semantically similar documents to permit more effective discovery, automation, integration, and reuse across applications. "The Web will reach its full potential when it

*Figure 1*




becomes an environment where data can be shared and processed by automated tools as well as by people" (Berners-Lee & Miller, 2002, p.9).

The World Wide Web Consortium (W3C) is an association of more than 350 member organizations from all over the world that produces standards, referred to as recommendations, for the World Wide Web. It was created in October 1994 to develop common protocols to promote the evolution and interoperability of the Web. The W3C's purpose is to lead the technical evolution of the Web, and it has developed more than 80 technical specifications for the Web's infrastructure. The vision of the Semantic Web was first proposed by Tim Berners-Lee, the scientific director of the World Wide Web Consortium, in 1998.

The evolution of the Semantic Web began with early HTML documents, where a minimal set of tags specified formatting for content. Over time, designers realized that it would be helpful to integrate more meaningful tags than "head" or "bold" to express concepts like "author." The emergence of XML ensured that document syntax could be consistent and allowed applications better ways of working with groups of documents that deal with related concepts. While XML is a key building block in the evolution of the Semantic Web, the first real manifestation of the W3C's semantic work was the development of the RDF specification for encoding and sharing metadata, which describes the content, quality, condition, and other characteristics of data. RDF is based on the premise that metadata can be modeled as a set of statements that indicate a piece of information about something else (Rhyno, 2002). The creation of Semantic Web documents as well as groups of related documents, or ontologies, are the foundation of the Semantic Web (Emonds-Banfield, 2002).

## Enabling Technologies

By supplementing human-readable content with machine-comprehensible content, the Semantic Web will allow machine-processable data to span application boundaries just as human-readable documents currently do (Miller, 2003). An explicit meaning is

*Table 1. A literature review of the enabling technologies of the Semantic Web*

| Issue | References | Main Contribution |
| --- | --- | --- |
| **Semantic Web** | (Adams, 2002) | Provides a good background on enabling technologies. |
| | (Berners-Lee et al., 2001) | This is the seminal piece on the SW. |
| | (Berners-Lee & Miller, 2002) | Explains the current state of SW research and several ongoing initiatives. |
| | (Bonner, 2002) | Provides an excellent overview of the SW. |
| | (Dumbill, 2000) | A primer for the SW with a good discussion of future expectations. |
| | (Emonds-Banfield, 2002) | Fine discussion of building the SW, with a few RDF examples. |
| | (Hendler, 2003) | Excellent presentation of the future of the SW. |
| | (Kuchling, 2004) | Provides an introduction to the SW and RDF. |
| | (Lassila, 2002) | Presentation notes with an overview of the SW. |
| | (Miller, 2003) | Discusses SW research that relates to Digital Libraries. |
| | (Miller, 2002) | Overview of the SW, especially as it relates to information professionals. |
| | (Ohlms, 2002) | Presents another view on the future of the SW. |
| | (Rhyno, 2002) | Discusses the SW and libraries, but includes a good review of the enabling technologies. |
| | (Sadeh & Walker, 2003) | Explains evolution of SW. |
| | (Swartz & Hendler, 2001) | Provides an excellent overview of the SW. |
| **XML/ XHTML** | (Brooks, 2002) | Describes the use of the SW in libraries; expresses doubts about realization of the SW. |
| | (Singh et al., 2005) | Defines and explains Semantic eBusiness; includes excellent background on technologies. |
| | (W3Schools, 2004a) | Provides an introduction To XHTML. |
| | (W3Schools, 2004b) | Provides an introduction to XSLT. |
| | (Holman, 2000) | A discussion of XSLT. |
| **RDF** | (Krichel, 2002) | Provides a discussion of the SW and an introduction to RDF. |
| **Ontologies** | (Alani et al., 2003) | Provides good discussion of ontologies. |
| | (Aldea et al., 2003) | Good overview of ontologies, especially as they apply to KM systems. |
| | (Gibbins et al., 2003) | Another useful overview of ontologies. |
| | (Hendler, 2001) | Discusses ontologies, agents, and the SW. |
| | (OntoWeb, 2002) | Details and results of an EU-funded project -- an ontology-based information exchange for knowledge management and e-commerce. |
| **Intelligent Agents** | (Arai et al., 2003) | Provides material on the SW and agents. |
| | (Ermolayev et al., 2004) | Details a project on agent-enabled SW service. |
| | (Green, 2002) | Overview of the SW and the role of agents. |
| | (Kungas & Rao, 2004) | Discussion of agent interaction in the SW. |

associated with Web-based information in order to make the processing and integration of such information easier for machines to carry out automatically (Sadeh & Walker, 2003). Realization of the Semantic Web is dependent on the development of standards and technologies that allow data on the Web to be defined and linked to semantically related data (Berners-Lee & Miller, 2002).

The Semantic Web facilitates finding information by providing the enabling standards and technologies that allow communities to express data in ways in which it can more easily be integrated, merged, and effectively searched (Miller, 2002). The idea behind the Semantic Web is the creation of documents that represent information in a highly structured fashion. This representation is entirely semantic and contains no presentation format information (Emonds-Banfield, 2002). Each document will be associated with encoded metadata that provides a context for Web-based data (Bonner, 2002). Semantic Web pages are enhanced by a new set of relationships such hasLocation, worksFor, isAuthorOf, hasSubjectOf, dependsOn, and so forth, making explicit the particular contextual relationships that are implicit in the current Web (Berners-Lee & Miller, 2002). It is envisioned that semantically related documents will be more readily accessible and comprehensible to automated software agents and other computerized tools without the need for human guidance (Bonner, 2002). The Semantic Web allows such tools to follow links and facilitate the integration of data from many different sources (Berners-Lee & Miller, 2002). The overall goal of the Semantic Web is "to turn the Internet into a vast, decentralized, machine-readable database" (Bonner, 2002, p.IP02).

Semantic Web technology integrates existing technologies such as Web technology, knowledge representation technology, and Digital Libraries (Miller, 2002). According to Berners-Lee et al., (2001), the Semantic Web requires the following components:

- A knowledge representation formed by interconnected ontologies with Web application software that has access to structured collections of information and sets of inference rules that make automated reasoning possible. These applications must be linked into a single global system.
- Ontologies allow Web applications to communicate with each other by providing a common vocabulary and rules that govern how the terms in that vocabulary work together and what they mean. An ontology is a document or file that defines classes of objects and relations among them through a taxonomy and a set of inference rules. Ontologies make it possible for applications to discover meanings for the data that is encountered.
- An intelligent agent is a software program that typically gathers, sorts, and processes information found on the Web without human intervention (Adams, 2002), and exchanges the results with other programs. Agents are able to communicate on the basis of a common dialect that is established by exchanging ontologies. Even agents that were not expressly designed to work together can exchange data if that data is semantically enriched. Therefore, as machine-readable Web content becomes more common, the effectiveness of intelligent agents will continue to improve.

Realization of the Semantic Web relies primarily on five core technologies: the extensible markup language (XML), uniform resource identifiers (URIs), the resource definition framework (RDF), ontologies, and intelligent agents.

## XML

The extensible markup language (XML) and its accompanying technologies are the fundamental facilitator of the Semantic Web (Berners-Lee et al., 2001). XML provides for language customization through the definition of new tags (such as <author>) to describe the data elements used in an XML document, hence the term "extensible." Unlike

HTML, which controls how data are displayed on the Web, XML is intended to facilitate the sharing of structured text and information across the Internet. The data display remains the job of HTML. In short, XML and HTML perform complementary, rather than overlapping, functions. XML supplements presentation markup with markup that provides a context for understanding the meaning of the data, for example, <author>Berners-Lee</author>. The advantage of XML is that software programs can read the specialized tags and perform operations such as extracting bibliographic information (Adams, 2002).

The structure, content, and semantics of XML documents are defined in an associated Document Type Definition (DTD) file or in an XML Schema. XML Schemas express shared vocabularies and provide a means for defining the structure, content, and semantics of XML documents. These schemas formalize the syntax and value constraints of XML instances and facilitate the sharing of information among communities of users (Brooks, 2002). Schemas allow XML documents to be parsed, validated, and processed by application software. This provides the foundation for the capture, representation, storage, and exchange of knowledge that can be potentially accessed and shared by intelligent agents (Singh, Iyer, & Salam, 2005).

XML namespaces enable the combination, in a single XML document, of element (and sometimes attribute) names from more than one XML vocabulary. Namespaces address some of the semantic blending problems that exist in a Semantic Web (Brooks, 2002). Namespaces are useful when XML documents pull data from multiple XML sources and encounter element name collisions. For example, a relatively common XML element like <dollar> could be clarified by a reference to one namespace that provides a context as a U.S. dollar amount or to another namespace that indicates that it is a Canadian dollar amount (Brooks, 2002). XML makes it possible to provide standardized representations of data on heterogeneous systems without case-specific programming (Singh et al., 2005). However, XML namespaces are unable to solve the more serious semantic problem that stems from the rarity of precise agreement about the meaning of any common word.

A discussion of XML would be incomplete without mention of XHTML, the extensible hypertext markup language. As noted earlier, HTML was designed to display data while XML was designed to describe data. One problem inherent in HTML is that it allows developers to create poorly formed documents. An HTML document is poorly formed when tags are not properly nested, tags are not associated with end tags, tag names and attribute names are not in lowercase, attribute values are not quoted, and so forth. This is a critical shortcoming because today's market consists of different browser technologies, some of which run on computers and others that run on mobile phones and handheld devices. The latter devices do not have the resources or power to interpret the poorly formed documents that often result from the less structured HTML. The W3C defines XHTML as the latest version of HTML, with the goal of gradually replacing HTML. XHTML is almost identical to HTML 4.01, and is in fact HTML 4.01 rewritten to follow XML rules. XHTML combines all the elements of HTML 4.01 with the syntax of XML. XHTML forces designers to write "well-formed" documents that work in all browsers and that are backward compatible with older browsers and will soon play a larger role in the Semantic Web (W3Schools, 2004a).

One final W3C recommendation that merits mention is extensible stylesheet language transformations (XSLT). Holman (2000) notes that the flexibility inherent in the ability to develop specialized vocabularies makes it necessary to be able to transform information marked up in XML from one vocabulary to another. XSLT is a language for transforming an XML document into another XML document, or into another type of document that is recognized by a browser, like HTML and XHTML (W3Schools, 2004b). It provides a means of converting instances of XML that use one vocabulary into either simple text, a legacy HTML vocabulary, or XML instances that use any other vocabulary imaginable (Holman, 2000). Normally XSLT does this by transforming each XML element into an XHTML element (W3Schools, 2004b).

## URIs

Uniform resource identifiers (URIs) provide another foundation of the Semantic Web (Berners-Lee & Miller, 2002). A URI is much like a URL, but it does not have to map to a real Web address. Further, a URI can represent concepts (e.g., "author"), living entities (e.g., "Tim Berners-Lee"), and virtually anything else (Rhyno, 2002). URIs can even point to physical entities, which means that the RDF language can be used to describe devices such as cell phones and TVs, which can, in turn "advertise their functionality — what they can do and how they are controlled — much like software agents" (Berners-Lee et al., 2001, p.43). Groups can declare their specialized concepts in terms of URIs, and these concepts, in turn, can be related (broader, narrower, synonymous, and so forth). Thus, URIs provide the capability to uniquely identify not only resources, but also can indicate the relationships among resources (Berners-Lee & Miller, 2002).

## RDF

The resource description framework (RDF) leverages URIs and XML to express the meaning of Web documents in a way that specialized software can understand (Adams, 2002; Krichel, 2002). This is accomplished in part by identifying Web resources with URIs and indicating relationships among them (Brooks, 2002). RDF provides a framework within which industry vocabularies in the form of metadata can be built and exchanged by communities (Krichel, 2002). Through RDF, authors can specify the contents of pages and how those pages relate to one another and to other known bodies of data (Bonner, 2002). An RDF description can include various types of metadata such as the authors of the document, the date of its creation, the name of the sponsoring organization, intended audience, subject headings, and so forth. (Adams, 2002). "RDF Vocabularies are descriptive terms (e.g., service, book, image, title, description, rights, etc.) that are useful to communities recoding information in a way that enables effective reuse, integration, and aggregation of data" (Berners-Lee & Miller, 2002, p.9).

The premise upon which RDF is based is that metadata can be modeled as a set of statements that indicate some piece of information about something else (Rhyno, 2002). The basic unit of data in RDF is a triple, which consists of a subject (a resource identifier), a predicate (a property, characteristic, attribute, or relation), and an object (either another resource or literal data) (Bonner, 2002). The following example, adapted from Kuchling (2004), shows how four facts are represented as 3-tuples of subject, predicate (property), and object:

| ***Subject* has a *property* of an *object*** |
|---|
| Resource W has a name of “Drew” |
| ISBN 1234567890 has an author of resource X |
| Resource Y has a type of Person |
| Widget Z has the title “Mega Widget 2005” |

A resource (the subject) is linked to another resource (the object) through a third resource (the predicate) (Brooks, 2002). This RDF triple represents the fourth statement (Swartz, 2002):

_:WidgetZ <http://example.net/rdf/title> “Mega Widget 2005”.

“WidgetZ” is the subject, <http://example.net/rdf/title> (which represents has the title) is the predicate, and “Mega Widget 2005” is the object. Individual RDF statements can be combined to create an RDF document:

_: WidgetZ <http://example.net/rdf/title> “Mega Widget 2005”.

_: WidgetZ <http://example.net/rdf/description> “Gray. Rounded corners.” .

_: WidgetZ <http://example.net/rdf/price> “$9.95”.

Each element in a triple can be represented as a URI that identifies things with a unique Web address (Bonner, 2002). Elements also can be blank nodes, which identify things that do not have their own URI, and literals, which are used to represent actual values. RDF triples form webs of information about related items. “Because RDF uses URIs to encode this information in a document, the URIs ensure that concepts are not just words in a document but are tied to a unique definition that everyone can find on the Web” (Berners-Lee et al., 2001, p.40).

RDF triples are represented within an HTML or XHTML document as XML metadata. The following example from Swartz (2002) describes a catalog standard that defines both classes and properties. The URI for the catalog standard is *http://tmrc.example.org/catalog/*. It can be abbreviated in RDF as cat:. The catalog standard includes the classes Widget, Sprocket, and Frobnitz. There are also several properties that will be used in the example. The type property comes from the RDF core vocabulary, the title and description properties come from the Dublin Core Elements, and price, color, and hexcolor are defined for the catalog example. Here is an example catalog item (Swartz, 2002):

Titanium Goorplaster 27

[Frobnitz] Industrial grade.

Price: $200.47

Color: fuschia

This item can be described in RDF as:

<http:// tmrc.example.org /catalog/fg27> rdf:type cat:Frobnitz .

<http:// tmrc.example.org /catalog/fg27> dc:title "Titanium Goorplaster 27".

<http:// tmrc.example.org /catalog/fg27> dc:description "Industrial grade." .

<http:// tmrc.example.org /catalog/fg27> cat:price "$200.47" .

<http:// tmrc.example.org /catalog/fg27> cat:color _:b1 .

_:b1 dc:title "fuschia" .

_:b1 cat:hexColor "F0F" .

The "dc" that appears in several lines stands for Dublin Core and is associated with a special URI called a namespace that provides access to its content by means of an RDF Schema that, in turn, is associated with a set of metadata elements (Rhyno, 2002). The b1 notation represents blank nodes. For a more detailed explanation, see Swartz (2002).

Web authors are responsible for the creation and addition of RDF data to their Web pages (Brooks, 2002). One possible source of RDF information is databases, which store machine-processable information. Well-designed databases can handle any number of queries about the data contained within. RDF is ideally suited for publishing databases to the Web, and when they are put on the Web, everything in the database is provided with a URI which allows intelligent applications to extract data from multiple databases and fit that data together (Swartz & Hendler, 2001).

The connections between data items established by RDF help make documents more comprehensible to automated readers, but there is still a lack of context in some of the data and ambiguity about how it relates to other data (Bonner, 2002).

## Ontologies

The next element required for realization of the Semantic Web is some mechanism to formally describe the semantics of classes in the many domains of interest and the semantics of properties used in Web documents (Sadeh & Walker, 2003). Ontologies provide such a mechanism. Hendler (2001, p.30) defines an ontology as "a set of knowledge terms, including the vocabulary, the semantic interconnections, and some simple rules of inference and logic for some particular topic." Ontologies allow computers to communicate with each other by providing a common set of terms and rules that control the definitions of those terms as well as the relationships between them (Adams, 2002). For example, ontology cross references would make it possible for an application to understand that “blouse” and “dress shirt” are similar concepts (Adams, 2002). The Semantic Web requires ontologies that cover everything from factory automation to post-structural philosophy, and the Dublin Core Metadata Initiative has been working for almost a decade to build vocabularies to overcome such potential bottlenecks (Adams, 2002).

Web ontologies provide a shared and common understanding of specific domains that can be communicated between different application systems (Singh et al., 2005). They identify the relationships between objects within a given knowledge domain and usually consist of a taxonomy, definitions of relationships between objects in the given knowledge domain, and rules for drawing inferences about those objects (Bonner, 2002). “Ontologies provide richer integration and interoperability of data and permit the development of applications that search across diverse communities or merge information from them” (Sadeh & Walker, 2003, p.12). Ontologies can be used to power advanced services such as more accurate search tools, intelligent software agents, and knowledge management (Berners-Lee & Miller, 2002). The RDF working group developed RDF Schema (RDFS), an object-oriented system that provides an ontology modeling language (Singh et al., 2005). There have been several recent efforts to build on RDF and RDFS with knowledge representation languages such as ontology Web language (OWL), simple HTML ontology extensions (SHOE), DARPA agent markup language ontology language (DAML-ONT), ontology inference layer (OIL), DARPA agent markup language + ontology inference layer (DAML+OIL), and personal ontology (Personal-Ont). These ontology languages provide advanced toolkits for defining ontologies and expressing semantic data, and allow knowledge sharing among agents through the standard Web services architecture (Singh et al., 2005). They extend RDF’s simple syntax with constructs such as data types, valid data ranges, unique keys, enumerations, and other rich language elements in order to give software the linkages needed to infer connections between data that have not been precisely stated (Bonner, 2002).

According to Aldea (2003), ontologies are capable of:

1. Providing a structure to annotate the contents of a document with semantic information, which then allows the retrieval of appropriate information from those documents (Alani, 2003; Gibbins, Harris, & Shadbolt, 2003).
2. Integrating information from many different sources (the original goal of the Semantic Web) by providing a structure for its organization and facilitating the exchange of data, knowledge, and models (Lassila, 2002).
3. Ensuring consistency and correctness by formulating constraints on the content of information (OntoWeb, 2002).
4. Creating libraries of interchangeable and reusable models (OntoWeb, 2002).
5. Enabling reasoning, which allows the progression from syntactic to semantic processing and allows systems to draw inferences based on generalized rules (Lassila, 2002).

## Intelligent Agents

The final element required for the realization of the Semantic Web is an intelligent agent. Intelligent software agents are software entities that carry out operations and process information on behalf of a user or another program with some degree of independence or autonomy, directed by some awareness of the user's goals or needs. Agents are used when the software must possess human-like capabilities such as the ability to perceive and assess the environment, proactive behavior in pursuing a goal, ability to learn from their experiences, and social behavior (Ermolayev, Keberle, Plaksin, Kononenko, & Terziyan, 2004). Many different kinds of intelligent agents are designed to perform specific, specialized tasks such as searching, shopping, site management, and so forth. Many agents are cooperative, which means that they can interact and communicate with humans and/or other agents. In the context of the Semantic Web, intelligent agents typically gather, sort, and process information found on the Web without human interaction. Agents can be designed to discover content that satisfies the user's preferences and requirements (Kungas & Rao, 2004). When a user issues an information request, an intelligent agent will analyse that request and delegate it to other agents and services that it has located through the use of agent/service directories on the Web. Multiple cooperative agents work together to create an "information value chain" in which the user's search request is "packet processed" through sub assemblies of information passed between agents, each of which contributes facts to construct the answer being sought. Appropriate agents will be capable of distilling large amounts of data distributed across the Web and progressively reducing it to the desired answer (Green, 2002). "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" (Berners-Lee et al., 2001, p.42).

## Integration of Enabling Technologies

How do all the parts tie together? The Semantic Web requires that Web pages be developed (or redesigned) in XHTML, which incorporates XML. XML tags can be used to describe the contents of the document. In fact, RDF triples (subject, predicate, noun — all of which can identify the location of, content of, and relationships between resources) are expressed in an XML representation to publicize semantic connections between documents in machine-processable form. RDF schemas and ontologies describe the meaning and relationships between the various vocabularies that are used to describe Web content and allow software to convert between them to establish a common vocabulary that enables communication and understanding. Intelligent agents examine RDF schemas and ontologies and use inference to locate documents that are semantically related, parse and interpret information from those documents, and integrate data from the various sources to arrive at a solution to whatever query or problem that they are intended to address.

# Future Trends

Is the concept of a Semantic Web practical and realizable? Ohlms (2002) asserts that numerous obstacles must be overcome before the Semantic Web vision can become reality. Dumbill (2000, p.2) points out that the Semantic Web "has already been the subject of much bluster among the XML developer community and will doubtless continue to be so. Arguments rage over the usefulness of the technology, the difficulty of using RDF, and so on." Brooks (2002, p.9) observes that although the concept is attractive, "it is unclear at this time whether the degree of standardization necessary for the success of the Semantic Web is possible in the Web environment." Ohlms (2002) notes that the underlying technologies are still immature and cover only part of the Semantic Web value chain.

However, as the integral technologies evolve, the concept is coming closer to fruition. The Semantic Web vision of a machine-readable Web has possibilities for applications in most Web technologies (Dumbill, 2000). Numerous papers address the application of the Semantic Web to libraries and their resources, to knowledge management systems, and to scientific research and collaboration. There are planned applications in shared calendaring, tools for visualization, and use for querying, browsing, and visualizing semantic data. The Semantic Web promises advanced information management capabilities of discovering, filtering, and searching. There will be numerous tools for marking up images and other multimedia data to make it easier to produce Web content while authoring Web documents (Hendler, 2003). The Semantic Web will make possible "real" queries like "How many five-star hotels are there in San Francisco?" (Hendler, 2003). Ohlms (2002) lists a number of improvements that will be made possible by the Semantic Web.

1. Information management will become more precise, with more elaborate knowledge modeling, generation, navigation, and retrieval.
2. Improvement in system integration through shared metadata layers and ontologies.
3. Multi-device capability will see improvements through unambiguous definition and specification of any Web resource.
4. E-procurement will see indirect benefits through easier information management and system integration.

The future of the Semantic Web was the focus of much of the May 2004 World Wide Web Conference. Berners-Lee, in his keynote speech, predicted a second phase with fewer constraints in which many new tools and languages built on RDF will emerge. He envisions a future in which enterprises adopt the Semantic Web only to be astounded by the dramatic way in which data can be collected and formatted in order to help humans and machines interact with information. He expects to see several new applications that are integrated through RDF and OWL. He even provided examples of how diverse forms of data can be cut-and-pasted or dragged-and-dropped into a Semantic Web rule to generate events or transactions in spectacular new ways (Naraine, 2004).

There is little doubt that while the scope of the Semantic Web is ambitious, it is that very scope that properly reflects the far-reaching effect it will have on the Web (Dumbill, 2000).

# Conclusion

The enabling technologies that underlie the Semantic Web, including XML, URIs, RDF, ontologies, and intelligent agents, are rapidly maturing. These technologies promise to give meaning to the Web by incorporating well-defined semantics into Web documents. The meaning of vocabulary terms used in a particular Web document can be specified in RDF triples expressed in XML and defined by a topic-specific ontology. Agents will be able to determine the semantic linkages between Web resources by following links from Web pages to those topic-specific ontologies (Adams, 2002). Thus, using a semantically based view of web resources, intelligent agents will be able to automatically discover, interpret, and evaluate Web content (Arai, Murakami, Sugimoto, & Ishida, 2003). Further, with the advent of the Semantic Web, search engines will no longer require users to guess at proper keywords in order to locate Web resources, but will instead allow them to provide a description of the resources they are seeking. Queries will evolve beyond Boolean searches based on keywords and will instead allow natural language queries.

Information is only meaningful when associated with context, and the Semantic Web will provide that context. The Semantic Web will attribute meaning to the content of Web pages, creating an environment in which information can be readily located and integrated. The Semantic Web holds great promise that tomorrow's Web will be a Web of semantics with far greater capabilities than today's Web of text.

# References

Adams, K.C. (2002). The Semantic Web adds logic to web services. *KM World, 11*(3). Retrieved December 1, 2004, from *http://www.kmworld.com/publications/ magazine/ index.cfm?action=readarticle&Article_ID=1222&Publication_ID=67*

Alani, H., Kim, S., Millard, D., Weal, M., Hall, W., Lewis, P., & Shadbolt, N. (2003). Automatic ontology-based knowledge extraction from Web documents. *IEEE Intelligent Systems, 18*(1), 14–21.

Aldea, A., Bañares-Alcantará, R., Bocio, J., Gramajo, J., Isern, D., Kokossis, A., Jiménez, L., Moreno, A., & Riaño, D. (2003). An ontology-based knowledge management platform. In *Proceedings of the Workshop on Information Integration on the Web (IIWeb-03) at the 18th International Joint Conference on Artificial Intelligence IJCAI '03*. Acapulco, Mexico, (pp. 7-12).

*American Heritage Dictionary of the English Language* (4th ed.) (2000). Boston, MA: Houghton Mifflin Company. Retrieved January 8, 2005, from *http:// www.bartleby.com/61/83/S0248300.html*

Arai, S., Murakami, Y., Sugimoto, Y., & Ishida, T. (2003). Semantic Web service architecture using multi-agent scenario description. *Pacific Rim International Workshop on Multi-Agents (PRIMA 2003). Lecture Notes in Artificial Intelligence*, 2891, Springer-Verlag, (pp. 98-109).

Berners-Lee, T., Hendler J., & Lassila O. (2001). The Semantic Web. *Scientific American, 284*(5), 34-43.

Berners-Lee, T., & Miller, E. (2002). The Semantic Web lifts off. *Special issue of ERCIM News, 51,* 9-11.

Bonner, P. (2002). The Semantic Web. *PC Magazine, 21*(13), IP01-IP04.

Brooks, T.A. (2002). The Semantic Web, universalist ambition and some lessons from librarianship. *Information Research, 7*(4). Retrieved December 1, 2004, from *http://informationr.net/ir/7-4/paper136.html*

Dumbill, E. (2000). The Semantic Web: A primer, *O'Reilly XML.Com*. Retrieved January 4, 2005, from *http://www.xml.com/pub/a/2000/11/01/semanticweb/*

Emonds-Banfield, P. (2002). Building the Semantic Web. *Orange: An Online Journal of Technical Communication and Information Design*. Retrieved November 10, 2004, from *http://tc.eserver.org/14913.html*

Ermolayev, V., Keberle, N., Plaksin, S., Kononenko, O., & Terziyan, V.Y. (2004). Towards a framework for agent-enabled semantic Web service composition. *International Journal of Web Services Research, 1*(3), 63-87.

Gibbins, N., Harris, S., & Shadbolt, N. (2003). Agent-based Semantic Web services. In *The 12th International World Wide Web Conference (WWW2003)*, Budapest, Hungary: ACM Press.

Green, D. (2002). When the Web starts thinking for itself. *Information World Review*. Retrieved January 3, 2005, from *http://www.iwr.co.uk/News/1137710*

Hendler, J. (2001). Agents and the Semantic Web. *IEEE Intelligent Systems, 16*(2), 30-37.

Hendler, J. (2003). Semantic Web futures. Retrieved January 6, 2005, from http://www.topquadrant.com/documents/WhiteHouseHendler.ppt

Holman G.K. (2000). What is XSLT? *O'Reilly XML.Com*. Retrieved January 4, 2005, from *http://www.xml.com/pub/a/2000/08/holman/*

Krichel, T. (2002). The Semantic Web and an introduction to RDF. In M. Koenig & T.K. Srikantaiah (Eds.), *Knowledge management lessons learned: What works and what doesn't*. (pp. 191-208). Medford, NJ: Information Today Inc. (American Society for Information Science and Technology Monograph Series).

Kuchling, A.M. (2004). Introduction to the Semantic Web and RDF. Retrieved January 6, 2005, from *http://www.amk.ca/talks/2004-12-02/*

Kungas, P., & Rao, J. (2004). Symbolic agent negotiation for Semantic Web service exploitation. In *Proceedings of the Fifth International Conference on Web-Age Information Management*, Dalian, China, July 15-17, LNCS 3129, Springer-Verlag.

Lassila, O. (2002). Towards the Semantic Web. Paper presented at *Towards the Semantic Web and Web Services*, October 21-22, Helsinki, Finland.

Miller, E. (2003). Enabling the Semantic Web for scientific research and collaboration. *NSF Post Digital Library Futures Workshop*. Chatham, MA, June. Retrieved November 6, 2004, from *http://www.sis.pitt.edu/~dlwkshop/paper_miller.doc*

Miller, E. (2002). Feature interview: Eric Miller on the Semantic Web. *NewBreed Librarian, 2*(4), Retrieved November 6, 2004, from *http://www.newbreedlibrarian.org/archives/ 02.04.aug2002/interview.html*

Naraine, R. (2004). Semantic Web ready for phase two. *Earthwebnews.com*. Retrieved January 2, 2005, from *http://news.earthweb.com/dev-news/article.php/3356531*

Ohlms, C. (2002). The future of the Semantic Web: A perspective on the market adoption of Semantic Web technologies. AIK-Symposium, 9. Retrieved January 6, 2005, from *http://www.aifb.uni-karlsruhe.de/AIK/veranstaltungen/aik9/presentations/slides/020419Future SemanticWeb.pdf*

OntoWeb. (2002). IST project IST-2000-29243 OntoWeb. Retrieved December 3, 2004, from *http://www.ontoweb.org*

Rhyno, A. (2002). The Semantic Web and libraries. *InsideOLITA*. Retrieved November 12, 2004, from *http://www.hpl.hamilton.on.ca/OLITA/InsideOLITA/IO2002No6.htm*

Sadeh, T., & Walker, J. (2003). Library portals: Toward the Semantic Web. *New Library World, 104*(1184/1185), 11-19.

Swartz, A. & Hendler, J. (2001). The Semantic Web: A network of content for the digital city. In *Proceedings Second Annual Digital Cities Workshop*, Kyoto, Japan, October. Retrieved January 6, 2005, from *http://blogspace.com/rdf/SwartzHendler*

Swartz, A. (2002). RDF primer primer. Retrieved January 7, 2005, from *http://notabug.com/2002/rdfprimer/*

Singh, R., Iyer, L., & Salam, A.F. (2005). Semantic eBusiness. *International Journal on Semantic Web & Information Systems, 1*(1), 19-35.

W3Schools. (2004). Introduction to XHTML. Retrieved January 2, 2005, from *http://www.w3schools.com/xhtml/xhtml_intro.asp*

W3Schools. (2004). Introduction to XSLT. Retrieved January 4, 2005, from *http://www.w3schools.com/xsl/xsl_intro.asp*

**Internet Session:**

**Ontology Example: OWL design for a Family Tree**

*http://protege.stanford.edu/mail_archive/msg13597.html*

*Interaction:*

Examine the example for an OWL design for a family tree. Try to figure out how to create two new classes called husband and wife. Note how unwieldy the example ontology would be without proper indentation.

# Case Study

## The Race to the Semantic Web

Read the paper "August 2009: How Google Beat Amazon and eBay to the Semantic Web" (available at *http://www.ftrain.com/google_takes_all.html*) It presents a fictional account of a future article that appears a business magazine published in 2009.

*Questions:*

1. If you were directing a company such as Google, eBay, or Amazon, what would factor into your decision to embrace the Semantic Web? Analyze it from the point of view of each company.
2. Assume that you are the CIO of Amazon. List the pros and cons of converting your current set of Web pages from HTML to XHTML that incorporates RDF triples. What features might already be in place to make such a conversion easier?
3. What would it take to design a search engine to search semantically enhanced pages? Write a paper listing your main points, and then (and only then) read the paper "Information Retrieval and the Semantic Web" (available at *http://ebiquity.umbc.edu/v2.1/_file_directory_/papers/121.pdf*). Does it confirm or contradict your proposed solution? There is no right or wrong answer.

# Useful URLs

Tim Berners-Lee, Semantic Web Road Map: *http://www.w3.org/DesignIssues/Semantic.html*

Tim Berners-Lee, The Semantic Web: *http://www.ryerson.ca/~dgrimsha/courses/cps720_02/resources/Scientific%20American%20The%20 Semantic%20Web.htm*

Resource Description Framework: *http://www.w3.org/RDF/*

RDF Primer: *http://www.w3.org/TR/rdf-primer/*

Extensible Markup Language: *http://www.w3.org/XML/*

XML.Org: *http://www.xml.org/*

XHTML 1.0 The Extensible HyperText Markup Language (Second Edition): *http://www.w3.org/TR/xhtml1/*

Uniform Resource Identifier (URI) Activity Statement: *http://www.w3.org/Addressing/Activity*

World Wide Web Consortium: *http://www.w3.org/*

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

Namespaces in XML 1.1: *http://www.w3.org/TR/2004/REC-xml-names11-20040204/*

DAML+OIL (March 2001) Reference Description: *http://www.w3.org/TR/daml+oil-reference*

Simple HTML Ontology Extensions Frequently Asked Questions (SHOE FAQ): *http://www.cs.umd.edu/projects/plus/SHOE/faq.html*

W3C Semantic Web Activity: *http://www.w3.org/2001/sw/*

An introduction to ontologies: *http://www.SemanticWeb.org/knowmarkup.html*

# Further Readings

Berners-Lee, T., Hendler J. & Lassila O. (2001). The Semantic Web. *Scientific American, 284*(5), 34-43.

Berners-Lee, T., & Miller, E. (2002). The Semantic Web lifts off. *Special issue of ERCIM News, 51*, 9-11.

Devedz, V. (2004) Education and the Semantic Web. *International Journal of Artificial Intelligence in Education, 14* 39-65 Available at *http://fon.fon.bg.ac.yu/~devedzic/IJAIED2004.pdf*

Cheniti-Belcadhi, L., Henze, N., & Braham, R. (2004). An assessment framework for e-learning in the Semantic Web. In *Proceedings of the Twelfth GI- Workshop on Adaptation and User Modeling in interactive Systems (ABIS 04)*. October 2004, Berlin, Germany. Available at *http://www.kbs.uni-hannover.de/Arbeiten/Publikationen/2004/abis_assessment.pdf*

Payne, T.R. & Miller, L. (2002). Calendars, schedules and the Semantic Web. *ERCIM News*, no. 51. Available at *http://www.ercim.org/publication/Ercim_News/enw51/payne.html*

Dieng-Kuntz, R. (2002) Corporate Semantic Webs. *ERCIM News*, no. 51. Available at *http://www.ercim.org/publication/Ercim_News/enw51/dieng.html*

Hofreiter, B., Huemer, C., & Winiwarter, W. (2002) Towards syntax-independent B2B. *ERCIM News*, no. 51. Available at *http://www.ercim.org/publication/Ercim_News/enw51/hofreiter.html*

Alhulou R. & Napoli, A. (2002). Combining XML and description logics for describing and querying documents. *ERCIM News*, no. 51. Available at *http://www.ercim.org/publication/Ercim_News/enw51/napoli.html*

Singh, R., Iyer, L., & Salam, A.F. (2005). Semantic eBusiness. *International Journal on Semantic Web & Information Systems, 1*(1), 19-35.

# Possible Paper Titles/Essays

*The Effect of the Semantic Web on E-Commerce.*

*How Libraries Are Impacted by the Semantic Web*

*How the Semantic Web Can Be Used to Mark Up Multimedia Data.*

*How Search Engines Can Be Enhanced Through Semantic Data*

*Pros and Cons of Re-authoring Web Pages to Embrace the Semantic Web*