The Semantic Web … Sounds Logical!

Anthony J. Radogna Jr

Follow this and additional works at: http://scholarworks.rit.edu/theses

Recommended Citation
The Semantic Web ... Sounds Logical!

By

Anthony J. Radogna Jr.
Rochester Institute of Technology

B. Thomas Golisano College
of
Computing and Information Sciences

Master of Science in Information Technology

Thesis Approval Form

Student Name: Anthony J. Radogna

Thesis Title: Semantic Web, The Future is Upon Us

Thesis Committee

Name Signature Date

Prof. Daniel Kennedy 5/20/2004
Chair

Prof. Michael Axelrod 5/20/2004
Committee Member

Prof. Dianne Bills 5/20/04
Committee Member
Thesis Reproduction Permission Form

Rochester Institute of Technology

B. Thomas Golisano College
of
Computing and Information Sciences

Master of Science in Information Technology

Semantic Web, The Future is Upon Us

I, Anthony J. Radogna, hereby grant permission to the Wallace Library of the Rochester Institute of Technology to reproduce my thesis in whole or in part. Any reproduction must not be for commercial use or profit.

Date: 3/20/04 Signature of Author:
# The Semantic Web... Sounds Logical!

Anthony J. Radogna Jr.
Abstract

"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." -- Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web

The Semantic Web will be an enabling technology for the future because as all of life's components continue to progress and evolve, the demand on us as humans will continue to increase. Work will expect more productivity; family will demand more quality time, and even leisure activities will be technologically advanced. With these variables in mind, I believe humans will demand technologies that help to simplify this treacherous lifestyle.

As patterns already indicate, one of the driving forces of technological development is efficiency. Developers are consistently looking for ways to make life's demands less strenuous and more streamlined. The benefits of the semantic web are two-fold. Conceptually, it will enable us to be productive at home while at work, and productive at work while at home.

The Semantic Web will be a technology that truly changes our lifestyle. The Web has yet to harness its full potential. We have yet to realize that in addition to computers, other machines can actually participate in the decision-making process via the Internet. This will allow virtually all devices the opportunity to be a helpful resource for humans via the Web. It must be taken into consideration that the Semantic Web will not be separate from the World Wide Web, but an

The Semantic Web ... Sounds Logical!
Anthony J. Radogna Jr.
extension of it. It will allow information to be given a well-defined meaning, which will allow computers and people to work in cooperation. With this technology, humans will be able to establish connections to machines that are not currently connected to the World Wide Web.

“For the Semantic Web to function, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning” (Scientific American: Feature Article: The Semantic Web, 3). Using rules to make inferences, choosing a course of action, and answering questions will add functional logic to the Web. Currently the Semantic Web community is developing this new Web by using Extensible Markup Language (XML) and Resource Description Framework (RDF) and ultimately, Ontologies.
Introduction

“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.” -- Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web.

The Semantic Web will ultimately give machines the ability to use logic to automate typically manual or human tasks. More specifically, the Semantic Web will utilize existing technologies, as well as, spur the development of new and innovative standards. The combination of existing web languages with new implementation controls will allow machines to understand web technologies and communicate with each other to accomplish tasks. At this point in time the Semantic Web is just an idea, whether or not it will ultimately come to fruition has yet to be determined. By means of this document, I will endeavor to prove two things: first that the Semantic Web will become an instrument for people and business to use to increase productivity, and second that it is already on a viable path toward this goal.

Society has continued to evolve toward a more technologically based and fast-paced existence. As a result, the human-computer relationship has grown mutually dependent where people cannot accomplish certain tasks without the aid of a machine and machines cannot act independently of our commands. The
World Wide Web (WWW) has given people the ability to perform a number of routine tasks in a more efficient and convenient manner. High demands on the individual, both personally and professionally, have changed the pace of life. Efficiency is becoming the centerpiece of society, as we know it. People are now looking for efficiency to help resolve some of the stress. (Scientific American: Feature Article: The Semantic Web, 3)

Scientists, developers and inventors are competing to find solutions that provide new efficiencies for daily life. The notion of the Semantic Web is slightly different in that, in order for it to come to fruition, all aspects of research must work together. The culmination of these efforts will result in adding semantics to the current Web in order to eliminate the separation between computer and machine. Ultimately this will provide a cohesive flow of information from people to machine and conversely, machines to people. Society has not contemplated the notion of machines participating and being a useful resource to humans via the Web. (Berners-Lee, Hendler, & Lassila, 2001)

The Semantic Web is not a separate Web, but an extension, or enhancement of the current one, that will allow information to be given a "well-defined" meaning. This will enable computers to make logical deductions based on inferences that are produced from definitions applied to information throughout the web. Inferencing - within the Semantic Web - means that computers will have the capability of reading, comprehending, and deducing an action based on different
types of information. This technology will allow humans to establish connections to machines, such as appliances, that are not currently connected to the World Wide Web. This will ultimately allow computers and people to work in cooperation. (Berners-Lee, Hendler, & Lassila, 2001)

In order to accomplish the goals of the Semantic Web, developers will primarily use eXtensible Markup Language (XML), Resource Description Framework (RDF) and Ontologies. These three main components will become the backbone of this next generation Web. Together these components will add “logic” to the Web. This logic will provide the rules for inferencing, allowing machines to take a course of action to answer questions. Simply put, this logic gives machines the ability to perform the decision-making process. “For the Semantic Web to function, computers must have access to structured collections of information and sets of inference rules they can use to conduct automated reasoning”. (Scientific American: Feature Article: The Semantic Web, 3). Extensible Markup Language (XML), Resource Description Framework (RDF) and Ontologies will be the foundation technologies for building these collections of logic. (Berners-Lee, Hendler, & Lassila, 2001)
The Internet Past and Present

The Internet has gone through many phases throughout its history. It has revolutionized how people, business, and society communicate and obtain information. It has allowed individuals to collaborate and interact regardless of geographic location. The Internet was originally designed as a network of networks. The Internet came to fruition with the invention of packet switching in 1964, the introduction of the ARPANET in 1969, and the invention of TCP/IP in 1972. It wasn't until 1984 that ARPANET was renamed the "internet" and at the same time TCP/IP became its standard protocol. The most significant milestone was the creation of the World Wide Web (WWW) and HTML language in 1989.

Tim Berners-Lee, an Oxford University graduate, developed the Web as a means of communication. His basis for this communication was information sharing (http://www.w3.org/People/Berners-Lee/ShortHistory.html). The WWW allows for an ever-increasing number of individuals to access vast amounts of information (http://www.freesoft.org/CIE/Topics/57.htm). In addition to creating the Internet, Berners-Lee is also credited with forming the W3 Consortium. Since its inception, the W3C has provided a common space where ideas can be shared and materialized according to standards that are necessary.

This network of networks joins many government, university and private computers together and provides the infrastructure for the use of email, bulletin boards, file archives, hypertext documents, databases and other resources. The
growth of the WWW has increased exponentially since its creation. In 1989 there were 100,000 hosts, in 2002 there was over 200 million hosts and by 2010, analyst are projecting that 80% of the planet will be on the Internet. As of today there are over 200 million hosts, 840 million users, 31 million domain names. Most significantly, over 220 of the 246 countries in the world are active participants (Slater, 2002). The Internet has been the fastest growing communications medium invented to date. In order to illustrate how rapidly the WWW has grown, it can be compared to the radio, which took 38 years to reach 50 million people, and the television, which took 13 years to reach 50 million. The Internet accomplished this feat in only 4 years. (http://www.freesoft.org/CIE/Topics/57.htm)

The Internet has now become a necessity for both businesses and the private sector. The World Wide Web has enabled people and businesses around the world to communicate, offered them entertainment, and enabled them to utilize e-commerce. Another benefit of the WWW is efficiency. People have now adapted to a fast pace, high stress world, and are looking to technology to streamline their daily activities. People are always looking for new and innovative ways to lessen the burden of work without sacrificing quality and productivity. (Slater, 2002)

Once any new technology has been introduced and used in its original form, it is modified and improved upon in order to better serve the needs of its users. The
same can be said about the WWW. It is logical that the next step toward improving the Web would be to add semantics. It is the next logical step because society is craving a resource that can streamline hectic lifestyles. Tim Berners-Lee wanted this innovative technology to be simple enough to gain widespread acceptance. We have, no doubt, reached his initial benchmark, so the next logical step would be to continue pursuing the addition of semantics to the web. (Berners-Lee, Hendler, & Lassila, 2001)

Currently the World Wide Web is only for human readers. It contains enormous amounts of information created by businesses, people, and organizations for a wide variety of different purposes. Due to the volume of information available, it is very easy to get lost or obtain erroneous answers. The goal of the Semantic Web is to develop new standards and technologies that will allow machines to directly understand information throughout the web, and to communicate with each other for the automation of tasks. (Berners-Lee, Hendler, & Lassila, 2001)

The World Wide Web was revolutionary. Anyone can publish documents for the rest of the world to view, see and link to. It did not matter if someone in England was using a PC to view a page located in England, and that page was originally designed by a Macintosh in Rochester, NY.
The Merriam-Webster Dictionary defines semantic as, “of or relating to meaning in language” (http://www.m-w.com/). The next version of the World Wide Web will apply meanings to text in web pages. By doing so, these semantics will allow machines to infer what is described on web pages. The Semantic Web is the term given to the web in which machines will be capable of processing data. (Berners-Lee, Hendler, & Lassila, 2001)

The Semantic Web is no longer just a vision of Internet pioneers. The World Wide Web Consortium (W3C) has already acknowledged the realization of many of the components that are necessary for semantic functionality. The most important examples of this include RDF language and ontologies. RDF language was created in 1999 by the W3C as the first attempt to remedy the differences in metadata syntax and schema definition languages. (Ilanella, 1999) Ontologies came later, in 2001, with the formation of the W3C Web Ontology Working Group (http://www.w3.org/2001/sw/WebOnt/charter#L880). The Semantic Web will be a system of interconnected technologies for interoperability of machine-processable data. The Semantic Web will do for data what HTML did for textual information systems in the World Wide Web. It will provide sufficient flexibility to accommodate most databases, plus the logic rules to link them together, in a system that reaches the entire world. (http://www.straddle3.net/context/02/020619_semantic.en.html)
The Semantic Web will not be a new invention, but rather an extension of the current Web. It will take us from an age of displaying machine-readable language via web pages, to an era of machine comprehensible and processable language under the WWW. The current WWW only displays pages of information via web pages locatable by URLs, and using HTML. The problem with the current use of HTML in the World Wide Web is its degree of accessibility. HTML has limitations because it is a language that is primarily intended to link documents.

(http://www.straddle3.net/context/02/020619_semantic.en.html)

As mentioned earlier, Tim Berners-Lee is the original creator of HTML and the father of the WWW. The Semantic Web is also his vision. He foresees a web that not only performs its current function as an information database, but in addition, allows machines to communicate directly with one another in a functionally logical capacity. Berners-Lee strives to define data, and link it in a way that the information will be understandable by machines and ultimately used to create a more resourceful application of information. Not just for display purposes, but also for use across various applications and machines. In order to accomplish this, two important technologies will be used: Extensible Markup Language, known as XML, and Resource Description Framework, known as RDF. Fortunately, both of these technologies are already in place, it is just a
matter of applying them to achieve the goals of the Semantic Web. (Berners-Lee, Hendler, & Lassila, 2001)

The challenge to developing the Semantic Web will be how to provide knowledge to the machine. “For the Semantic Web to function, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning” (Berners-Lee, Hendler, & Lassila, 2001). XML allows people to add this layer of structure to information. (Berners-Lee, Hendler, & Lassila, 2001)
Universal Resource Identifiers

To this point, the Web has been a medium for posting and linking documents, rather then a medium for processing data. The Semantic Web will change this. The current allure of the Web is the power of the hypertext link, which can link concepts together by a process that simulates human thought. Information and data vary from person to person, business to business, and across societies. The semantic web will be able to comprehend these documents and data.

Proper design is a key factor in establishing this new technology. The first phase is to name every concept with a Universal Resource Identifier (URI). A URI is a short string of characters that conforms to a certain syntax. The string indicates a name or address that can be used to refer to an abstract or physical resource. This will identify "anything" (including classes, properties, or individual items).

Using URIs will allow documents and systems to reuse information.


Some examples of URIs are the URL and Mailto. A URL is an address that lets you locate a web page, for example http://www.w3.org/Addressing/. A URL lets your computer locate a specific resource (in this case, the W3C's Addressing website). In addition to URLs, there are other forms of URIs. Examples include the familiar "mailto:" URIs that are used to encode email addresses, or the less well-known "mid:" URIs which identify email messages. (Swartz & Hendler, 2001)

The Semantic Web ... Sounds Logical!
Anthony J. Radogna Jr.
The new Semantic Web will be a space where resources will be identified by URIs to be accessible to machines. It will operate under the premise of both a partial understanding and an inference from these terms. The proposed method of using URIs to represent the resources will allow systems to grow globally on a decentralized basis, which will be the same process that currently occurs with hypertext documentation on the World Wide Web. (Swartz & Hendler, 2001)

Once data has been given a URI, it then can be referenced by anyone on the web and will also allow relationships to be built, processed, or queried. URI's allow documents and systems to reuse this information. The URI's will be descriptions at the lowest level – like machine language. This unique identifier will allow people to express new concepts that they invent with little effort. It will allow concepts to be linked throughout the web. Once this structure is in place, it will open up knowledge to a meaningful analysis by machines. (Swartz & Hendler, 2001)

XML is the language that will allow people to create their own tags/URI's for Web pages or machines to process. Tim Berners Lee's plan is based on the hope that people will start publishing their data in RDF format. Both of these technologies will allow the addition of logic to the Web.
**eXtensible Markup Language**

Extensible Markup Language (XML) is a scripting technology concerned with the description and structure of data that provides a format for the combination of binary file formats with text file formats. XML is an enhancement to HTML that provides a simple way to send documents across the web. It is a subset of Standard Generalized Markup Language (SGML), which has been in use for a wide variety of commercial and industrial purposes for more than a decade. XML retains the general framework of SGML, but removes much of its complexity, while retaining information that is useful in displaying documents in a browser from HTML. As a result XML is valued for its simplicity. In effect, XML allows for multiple functions on a web browser and provides for more flexible and adaptable information identification by allowing interoperability of SGML and HTML functions. The importance of this is that it allows for user-defined document types and for the storage and transmission of these documents both on and off the web. As a result, it will permit greater interconnectivity between web-based and non-web based storage devices. (Dumbrell, 1)

Although the name implies that it is a language, XML is not a language in the conventional understanding of computer languages, but rather a standard for creating languages (Cagle et al, 15). The W3C is the body that is defining the standards for XML. The XML standard will be available to all users in the public domain. In its development of XML, W3C is aiming towards the development of interoperability capabilities for all computing devices, including hand held...
computers and wireless technology. As a result, there exists the potential for all computing devices to be interconnected and to operate without a significant degree of human involvement. (Cagle et al, 15)

In its current state, the web is intended to be read by humans, with the data from the web requiring a human interface prior to transmission to machines. The Semantic Web focuses on developing languages for expressing information in a machine-readable form (Dumbrell, 1). The advantage of the development of the Semantic Web is to increase the speed at which web users are able to find, sort and process information. In practice, machines will be able to search and sort data from the web, reducing the amount of time that is necessary for humans to search out information through the use of a browser. In order to make the Semantic Web a reality, however, it is first necessary to develop a comprehensive web language that supports machine-to-machine interaction, such as XML. The interoperable syntax of XML will provide the essential foundations for the Semantic Web. (Dumbrell, 1)
As shown in Figure 1, XML allows a person to design the document's format, and write a document within it. We begin with identifying the order that information will be displayed (i.e., animal species, animal type, food consumption, and webpage). Once this is established we can enter specific data to be displayed in this format (e.g., giraffe, herbivore, plants, and http://www.giraffes.com). XML uses formatting to enhance the meaning of documents. By doing so, it allows these documents to become more useful. They are not only comprehensible by
web browsers, but can now be used by many different programs – with each program only using the markup code that it understands. (Dumbrell, 1)

XML can be used in a number of ways to describe data. This makes it open-ended and flexible in supporting the type of widespread, ad hoc data integration envisioned for the Semantic Web.

(http://www.hpl.hp.com/semweb/swtechnology.htm#Global%20naming%20scheme)

This technology will also requires the use of a secondary technology known as RDF, which allows objects and values to be combined in the meta tags of a web page. W3C proposes to use RDF, to represent machine-readable information.

RDF is an extension of XML, which defines a general data model. (Dumbrell, 1)
Resource Description Framework

Resource Description Framework technology provides a common way of representing data across numerous applications. RDF can be used to describe any Web resource in a way that provides interoperability between machines and applications on the Web. Its main purpose is to add formalized semantics to the World Wide Web. Any, and all, items that RDF expressions describe are called resources, while at the same time, anything a URI names can also be called a resource. The difference is that a RDF not only has the ability to describe pages, parts of pages or collections of pages on the Web, it can also describe things not on the Web. However, this is only possible as long as they can be named using some URI scheme. (Bray, 2003)

"RDF is simply a data model and format that allows people to create machine-readable data. The Semantic Web will be built on top of this data." (Semantic Web, The Second Generation Web?, p.3) By using RDF, you will be making data available for use on the Semantic Web. This ties into the main premise of the Web’s global accessibility, with one advantage; the information is machine-processable. (http://www.straddle3.net/context/02/020619_semantic.en.html)

The RDF of the semantic web provides the basis of metadata interoperability across a variety of resource communities, with the metadata serving as the data that identifies the content and information contained in a web document. Currently, there are a large number of methods to create metadata, and no
single, universal standard that requires metadata to be included in all web documents. The RDF system will allow applications such as site maps, library catalogs, electronic commerce web pages, and privacy ratings to become more accessible by supporting the rapid exchange of information by means of a single metadata standard. In effect, it is a standardized method for embedding data into a web page or other type of document that will allow automated accessibility via a page. Software will evaluate whether the contents of a page meet the preset criteria guiding its search and retrieval process. (http://www.w3.org/RDF/.)

Real time searches are based on metadata. In effect, a library catalogue is a form of metadata, as it contains information about the information in the library, and allows this data to be searched to find desired information. In the context of RDF, metadata is an informational line placed on a web page for use by XML that identifies the content of the information that is on the page, much in the same manner that a library catalog identifies the content of a book. (Bray, 1)

In order for semantics to work via the Web, all information and descriptions of data will have to be housed in databases. RDF is ideally suited for publishing these databases to the Web. Each data set within the database must be given a URI in order for all people to use this information. (Swartz & Hendler, 2001)

The purpose of RDF is to establish the necessary metadata for the semantic web in a standardized format. This format will allow for an intelligent and automated search of the web for specified information based on some criteria. RDF sets rules for the use of metadata in order to facilitate interoperability, and requires
the use of metadata in the prescribed format in order to allow the data to be accessible to outside users. All resources should have metadata, which includes both URLs and individual elements of an XML document. The resource has to be named so that it is identifiable. In addition, the resource has to be assigned a property type, such as a "home page", "site map", or some other type of commonly recognized property. This is commonly referred to as the "value". This data is compressed into a metadata line, or meta tag, which enables the automated search process to query the site and determine if the site falls within the criteria specified for retrieval. (Bray, 1)

Figure 2: Basic RDF Illustration taken from: http://www.w3.org/TR/2004/REC-rdf-primer-20040210/

```xml
<?xml version="1.0"?>

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:object="http://www.w3.org/2000/10/swap/pim/object#">

  <object:Animal rdf:about="http://www.w3.org/animals/EM/lion">
    <object:species>African Lion</object:species>
    <object:website rdf:resource="www.lions.com"/>
    <object:type>carnivore</object:type>
  </object:Animal>

</rdf:RDF>
```
"Like HTML, this RDF/XML is machine processable and, using URIs, can link pieces of information across the Web. However, unlike conventional hypertext, RDF URIs can refer to any identifiable thing, including things that may not be directly retrievable on the Web (such as the person Eric Miller). The result is that in addition to describing such things as Web pages, RDF can also describe cars, businesses, people, news events, etc. In addition, RDF properties themselves have URIs, to precisely identify the relationships that exist between the linked items." (Manola & Miller, 2004)

RDF establishes the order of information in the metadata record. An RDF statement is similar to a simple sentence. The only difference is that the words are URIs. RDF is a data model based on three principles: subject (resource), property, and object (value). The subject is what the assertion is about. The property is the value that is being asserted. The object is the value of the property. The subject and property will be defined with URIs, while the object will either be defined by URI or a literal (a literal is defined as any value other than the URI, in particular a name or a subject). (Swartz & Hendler, 2001) (Bray, 1)

In practice the RDF is a three-part record identifying the resource, the property type, and the value, which supports a very rapid search and retrieval process. A Resource is anything that can have a URI; this includes all the world's Web pages, as well as individual elements of an XML document. An example of a resource is a URL.
A Property Type is a Resource that has a name and can be used as a property, for example Author or Title. In many cases, all we really care about is the name; but a Property Type needs to be a resource so that it can have its own properties. The Value can just be a string, for example "Tony Radogna" or another resource, for example "The Home-Page of http://www.textuality.com/RDF/Why.html is http://www.textuality.com." (Bray, 1)
Ontologies

XML is the preferred language of the Semantic Web and RDF is the schema that W3C has chosen. Ontologies are the third component that will make up the new Semantic Web. Ontologies are formal descriptions of concepts. They use classes to define the concept, apply properties of each concept for describing various features and attributes with slots. A slot is a defined class and a slot-constraint places restrictions on the slots and value-types place a value back on a class. Figure 3 shows a basic example of an ontology.
Figure 3: Ontology Illustration (simplified) taken from:

(http://www.hpl.hp.com/semweb/swtechnology.htm#Global%20naming%20scheme)

```
class-def animal                                      animals are a class
class-def plant                                       plants are a class

  subclass-of NOT animal                                of things that are not animals

class-def (slot) carnivore                             carnivores are a class

  subclass-of animal                                    which is a subclass of animals
  slot-constraint eats

    value-type animal                                    that eats animals

class-def (slot) herbivore                             herbivore are a class

  subclass-of animal                                    which is a subclass of animals
  slot-constraint eats

    value-type plant                                    that eats plants

class-def giaraffe                                     giaraffes are herbivore

  subclass-of herbivore                                 

class-def lion                                         lions are carnivores

  subclass-of carnivore                                 
  slot-constraint eats

    value-type herbivore                                 that eat herbivores
```
Figure 3 shows a basic example of an ontology. In this scenario the machine can infer that an animal is not a plant, thus a plant is not an animal. Further, it will assess that an animal can be a herbivore or carnivore, where the former eats plants and the latter eats animals. When a user inputs the word giraffe it will use logic to conclude that first it is an animal, and second, it is an herbivore. Ultimately it can use all of this information to deduce the following: a lion, which is read as a carnivorous animal, will in fact, eat a giraffe. Before this process the web had no way of knowing that either a giraffe or lion is an animal, much less identify the relationship between the two (i.e., that a lion will eat a giraffe).

Ontologies will provide definitions of terms and problems that exist on the current Web. XML code used within Web pages can be defined by pointing from a given page to an ontology. Ontologies will enhance the functionality of the current Web in multiple ways. For instance, they will provide accuracy in search programs by having those programs only look for pages that refer to the specific concept or idea instead of vague key words. Advanced Web page applications, or agents will have the ability to use these ontologies to relate specific information on one page to the structure and inference being used in its own application. (Berners-Lee, Hendler, & Lassila, 2001)

Ontologies will also define a common vocabulary for people, researchers, and web designers who need to share information about a particular field or domain. Numerous ontologies already exist throughout the Web. Some examples would be the taxonomies categorizing web sites, such as Yahoo and Google, while
Amazon would be an example of an ontology that categorizes products and their features. (Swartz & Hendler, 2001)

There are numerous reasons to use or develop an ontology. The idea to develop ontologies has been driven by the desire to allow for the reuse of domain knowledge. An ontology would be created to share a common definition or understanding of a specific structure of data. It will enable the reuse of specific data and allow explicit assumptions to be attached to the data. The ontology will allow this data to separate its knowledge of actual content from the operational knowledge of a site, which currently only allows it to display material. The sharing of common understanding of the structure of information among people or software agents is one of the common goals in developing ontologies. (Berners-Lee, Hendler, & Lassila, 2001)

Over the years there has been a great deal of money and time spent on research and development of ontologies. Europe has funded the development of OIL (Ontology Interface Layer), which is a language for defining ontologies. In the USA, DARPA (Defense Advanced Research Projects Agency) funds a similar project on developing an ontology language, DAML (DARPA Agent Markup Language). In late 2001 the W3C setup a working group to combine the two languages into DAML+OIL. This language is being used to define ontologies on the web. (http://www.hpl.hp.com/semweb/swtechnology.htm#Global%20naming%20scheme)
Trust and Security on the Semantic Web

Trust and security are major concerns for anyone using the WWW. The following quote from Swartz and Hendler illustrates the concern of users about the Semantic web: "Now you've probably been thinking that this whole plan is great, but rather useless if anyone can say anything. Who would trust anything from this system if anyone can say whatever they want? So you don't let me into your site? Ok, I just say I'm the King of the World and I have permission. Who's to stop me?" (Swartz & Hendler, 2001) Now that semantics will be added and more personal and significant information is exchanged to describe the data, trust and security are more crucial then ever in order to protect the interests of those using the new web.

To be effective, a framework for trust and security requires agreed upon standards for the transmission of information and security protocols. For interoperability, parties seeking to communicate must have a common standard for insuring proper identification. There are currently a number of proposals for how this security framework could operate on the semantic web. Each of them is fundamentally similar to the current automatic security protocols but designed for the specific XML language and ontological environment of the semantic web. (http://www.ninebynine.org/SWAD-E/Security-formats.html)

Access control is one of the underlying structural components of the various approaches for establishing trust and security on the semantic web. Trust and security are intended to control access to network resources by means of policies.
built into the software structure. Simple access control is keyed directly to the identity of the party seeking access and is verified by means of a software protocol. The party seeking access sends a request to the resource server. The resource server then forwards the request to the access control policy server, which returns an entry authorization or denial in a fully automated communication process between IT equipment. (http://www.ninebynine.org/SWAD-E/Security-formats.html)

In more sophisticated access control, a control policy server may request additional identity information such as a password prior to authorizing or denying access. The access control becomes more complex when it involves an online transaction. An online transaction involves the transfer of goods or money, with the primary vehicle of exchange being the Internet. In the case of an online transaction, the access server forwards information to the bank server that monitors credit card payment information. The bank server may initiate a request for additional information to verify identity that is passed through the access control server to the resource server. In practice, the level of security and trust in a specific system is based on the ontological properties of the verification protocol that the system operator constructs in the security verification software. This is significantly more sophisticated than the current system of security verification that requires only a public-private or in some cases a private-private encryption key (Kagal, Finan and Josi, 3). The proposed security system for the semantic web will continue to use the encryption key method, but will add an
additional layer of security based on specific information queries to the entities seeking access to the system. (Kagal, Finan and Josi, 3)

Because the development of the semantic web has not yet reached the point at which there is consensus of agreement on the nature of the security protocol, there are a number of proposals for security that involve varying emphasis on ontological or RDF approaches. In addition, there are proposals for systems involved in a communication to directly verify identity or to refer identity verification to a trusted third party system. (Kagal, Finan and Josi, 3)
Content Management

The objective of the semantic web is to standardize rules for metadata that support the content on a web page, in order to allow automated reading of that content. The standardization of the rules by which content is constructed will mean that the process of content management can be largely automated by a content management system that adheres to these basic rules. The content management system will follow an ontology, which serves as a model that describes the rules and relationships for the transformation of content into a web page. This will allow the content to be automatically read by software agents, which can evaluate the content for retrieval based on predetermined search parameters. (McGuiness, 1)

The metadata that is behind the visible website is designed to follow the basic rules of the semantic web. This metadata structure is composed of syntactic metadata, which describes non-contextual information about content such as language and bit rate. The metadata structure is also composed of semantic metadata, which describes domain-specific information, such as the context in which a term is used (e.g., one site may refer to "Yankee" as a member of a baseball team, and another may identify it as a slur geared toward people in the Northern United States). Within this framework established by the metadata, a company or individual is free to establish the ontology of the web site that defines the relationships between the various types of information available at that site. In effect, the ontology provides the context for the semantic metadata, with the
ontology normalized or standardized to the type of general information that is presented at the website. (Lombardi, 1)

Once both the metadata rules and the ontology are established, it will be possible to fully automate the content management process. In effect, the content manager will enter metadata about how the information is used in the organization rather than entering metadata regarding how the information will be displayed at the website, which is the current method of content management. This will effectively automate the content management process and reduce the amount of time that is necessary to manage the website. (Lombardi, 1)

In this altered context for content management, the ontological components will define the way in which software agents will operate. The agents will be able to perform applications that involve text-sensitive searching, browsing, correlation and content analysis. Evaluation will begin at the highest level of the ontological framework by means of the metadata. The agent then proceeds down the semantic metadata hierarchy searching for the specific metadata that matches the relevance level of its search criteria. (Lombardi, 1)
Goals of the Semantic Web

The basic structural goals of the semantic web are to develop an infrastructure that is universal, scaleable, decentralized, and capable of evolution. All of these goals are similar to the infrastructure goals of the current web. The operational goal of the semantic web, however, is to establish a method for search engines to seek content on the Internet, and then act on the content if it meets the parameters that have been set for action. In effect, the semantic web will have a generic and re-usable automated information retrieval protocol that requires human intervention only at the initial stage of customizing the information retrieval constraints. (Swartz & Hendler, 2001)

As previously stated, the semantic web will achieve its goals by establishing three protocols: the use of XML for syntax and structure, the development of ontological rules that define terms and their relationships, and the use of RDF that provides a model for encoding ontologically defined meaning. Once the semantic web is operational, there will no longer be a need for human intervention to evaluate the relevance of the information that has been found on the Internet. The process of assessing relevance will be based on a decision-making algorithm implemented through RDF that is preset prior to engaging in the search. (Swartz & Hendler, 2001)

The primary impact of achieving this goal for e-commerce will be a more highly automated commercial processes that require less human involvement in decision-making. In the current operational environment, e-business information
processing remains largely a manual process. The human interface is necessary to sort and complete various functions such as information evaluation, order taking and order processing. The value of this system for various e-commerce enterprises depends on the nature of the organization’s business model. For example, a firm that is in the business of providing data to customers can more swiftly and more economically access a wide variety of data that is available on the Internet in an automated fashion and sort it for relevance. (Grimes, 1)

Buyers of goods on the Internet will have the competitive advantage of being able to automatically evaluate the features and pricing of a wide variety of products that are offered for sale, with automated application functionality for both business-to-business and business-to-consumer purchasing. At the same time, vendors with highly competitive quality and pricing will have a wider market for their products due to the automated search procedures. From the internal perspective of a business engaged in e-commerce, the semantic web is likely to result in reduced operating costs due to lower employee needs and an increase in sales volume due to an automated expansion of the potential market. From the customer perspective, the semantic web is likely to result in more competitive pricing and a greater ability to comparatively shop among the providers of e-commerce goods and services. (Grimes, 1)
Current State

The semantic web is in an early state of development. There are a number of isolated pockets of individuals and companies developing websites based on semantic web metadata rules and ontologies. There is a great deal of interest in deploying semantic websites in the scientific community, particularly among physicists eager to develop grid computing that allows computing to occur on multiple interrelated computer systems (Gobel and De Roure, 5). At the same time, there are a few e-commerce sites that are deploying semantic websites in anticipation of the time in which the semantic web will become an important factor in commercial transactions. These isolated pockets, however, remain relatively uncommon. As a result, the full development of the semantic web, to a point at which the majority of websites follow this standard, is estimated to take between five and ten years. (Gobel and De Roure, 5)

A great deal of the current development of the semantic web focuses on the development of standardized XML vocabularies and ontological structures that are suited to specific businesses or industries. For example, the Distributed Management Task Force (DTMF), is a not-for-profit, vendor-neutral collaborative body that is leading the development, adoption and unification of management standards and initiatives for desktop, enterprise and Internet environments. (DTMF, 1). Another organization currently engaged in development is PapiNet, which is developing global standards for supply chain management, and is particularly focused on the paper industry. (PapiNet, 1)
Because the process of developing international standards for different industries is generally beyond the scope of any single company or individual, a consortium generally undertakes the process. The activities of these industry consortiums are mediated by W3C, which sets the overall standards for the semantic web. Over 400 member organizations are involved. Many of the metadata standards and ontologies that are under development are estimated to be ready for dissemination to various industries in a three to five year time-frame.

(http://www.w3.org/2001/12/semweb-fin/w3csw)

At the same time W3C is involved in the development of various technologies, which are essentially variations of XML for specific purposes. These technologies are in different states of development, and will provide the specific means and standards for accomplishing common tasks found in websites. A technology can categorized as a working draft, which is a proposal, the stage of a candidate recommendation, which is a technology that is well developed and has the recommendation of W3C for adoption, and the stage of recommendation, which involves a technology that has received consensus support for adoption in the semantic web. For example, Mathematical Markup Language (MATHML) is at the working draft stage, and will be used for describing mathematical notation, while XML encryption and syntax for use in encryption is at the recommendation stage. The semantic web cannot be fully functional until there is consensus on all the areas necessary for presenting information on web pages.

(http://www.w3.org/2001/12/semweb-fin/w3csw)
Conclusion

"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.” – Tim Berners-Lee

How the Semantic Web will affect people, business and the society is a question that can only be answered with time. I feel the addition of semantics to the web will allow automation, which in turn will lead to more efficiency on the job. Inevitably, increased efficiency will also lead to higher productivity and job satisfaction. With added logic, machines will reduce the monotony of routine jobs and in turn will free up time to be spent on other projects. For example, something as tedious as scheduling a meeting can become automated once machines can interface directly. With added logic the machine will be able to determine what departments and, more specifically, who from that department is a required attendee. From that point it will be able to schedule the meeting and send a notice based on everyone's availability. Normally, this activity is left entirely to the individual to perform. By automating it, that same individual can reallocate his/her time to higher-level decision making tasks. As we look further at the "big picture", increased efficiency for one individual translates into more efficient business practices for the company or industry and should also be directly correlated to a more robust economy.
The addition of semantics to the web will also result in more efficient personal and leisure activities. In the future, appliances will be able to react to schedule changes. For example, Bob Smith has a dinner prepared and placed into an oven that has refrigeration capabilities. For the purpose of this example, the oven is scheduled to go on at 5 o’clock PM. In the mean time, Bob gets a last minute meeting invitation for 5 o’clock PM, and the meeting is scheduled to last 3 hours. In normal circumstances, Bob would use a timer to set the oven and regardless of the situation, the oven would turn on at 5 o’clock. The addition of semantics to the web will enable the machine (in this case the oven) to interact with Bob’s calendar, notice the scheduled meeting, and postpone turning on the oven by 3 hours. Or, it could notice the schedule change and infer that Bob will not be home in time for dinner, in which case it will cancel the baking process entirely and keep the food refrigerated. As our lives grow increasingly more complex, the need for flexibility in the technology upon which we depend also expands. As people, one of the most important abilities we can possess is the ability to adapt quickly and efficiently to change. Unless we teach the machines that we use to adapt as well, our efforts may be confounded.

The essential property of the WWW is its universitality, its ability to link anything to anything. The Internet “changed the world” by being accessible to virtually anyone anywhere. However, to accomplish a global reach, a seed needed to be planted. The use of Hypertext to link documents and create the WWW was this seed. It was the beginning of important applications, but it did not reach its full
potential until it was linked to a single global system. Currently, the Semantic Web is a technology that is in a state comparable to this, when Hypertext was used at the advent of the current Web, in 1989. When the semantic web reaches its global potential, we will have the support of an invaluable informational resource.

Similar to the WWW, I feel the Semantic Web will have to be as decentralized as possible. In order for it to come to fruition and to become ultimately decentralized, there will have to be compromises, which will allow exponential growth. A compromise would be the reality that there will be paradoxes and unanswerable questions. However, the ultimate goal will be versatility.

On a larger note, the ability of existing companies to adopt and reap the benefits of the Semantic Web may prove to be the determining factor in not only whether a company survives, but also whether a company can be an industry leader. Ebay, Amazon and Google are examples of successful virtual marketplaces, that is, websites where people participate in online transactions. One example of the astounding impact of the Semantic Web includes these three powerhouse Internet sites. Ebay is well known as an auction house, which out sources 100% of its inventory. This process makes it an entirely manual interface, meaning users are responsible for finding and purchasing goods. Amazon, most widely recognized as a book wholesaler, has developed into a supplier of a multitude of products and holds limited inventory. While it is not an entirely manual process,
Amazon is limited by its inventory and supplier contracts, and the user specifically goes to Amazon.com to purchase goods. On the other hand, Google entered the industry strictly as a search engine, but soon realized that by using semantics within their server they were in a good position to put buyers and sellers together. In this case, web users may not initially set out to purchase goods, but in the process of researching something, Google can infer the users interest, and identify potential advertisers, suppliers and ultimately products. This idea will, no doubt, translate into industry-leading profits in just a short time.

As I have stated there are limitless uses for the Semantic Web, depending on a person’s point of view. The ability of linking your PDA, laptop, desktop, car and appliances could be the ultimate use for added semantics. Maybe it’s the fact that it will be easier than ever to find an answer to your question or a product that you might be interested in purchasing on the web. Perhaps it will be the fact that corporate decisions that used to be monotonous and hand processed can now be automated.

Although different people and organizations will find different uses, whatever an individual or organization feels is its most important application, all entities should be able to support the “grand vision” of the Semantic Web. The possibilities that it suggests are universal enough to answer a multitude of requests. Individual needs are different in scope and scale, but we can all concur that a technology as diverse as the semantic web will do no less than simplify our already too
hectic lives. There is definitely a long way to go before this vision can be realized, but there have been major strides in the right direction. The possibilities are endless, and how soon it will take off is anyone’s guess (there are estimates that range from 5 to 10 years), but there is one thing for sure, this new technology has the potential to revolutionize the world.
References

http://www.sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21

Bray, T. *RDF And Metadata.* Retrieved online April 25, 2003:
http://www.xml.com/pub/a/98/06/rdf.html


http://www.intelligententerprise.com/020328/506decision1_1.shtml.


Internet History. Retrieved online April 20, 2003:

http://www.freesoft.org/CIE/Topics/57.htm

*Introduction to Semantic Web Technologies*. Retrieved Online April 25, 2003:

http://www.hpl.hp.com/semweb/sw-technology.htm#Global%20naming%20scheme


Manola, Frank & Miller, Eric *RDF Primer*. Retrieved online May 12, 2004:

http://www.w3.org/TR/2004/REC-rdf-primer-20040210/

http://www.m-w.com/

McGuiness, D. *Ontologies Come of Age*. Retrieved online May 3, 2003:


Resource Description Framework. Retrieved online April 25, 2003 from W3C:

http://www.w3.org/RDF/.


Slater, W. F. *Internet Histroy and Growth*. (2002). Retrieved Online April 27, 2003:

http://www.isoc.org/internet/history/2002_0918_Internet_History_and_Growth.ppt

http://blogspace.com/rdf/SwartzHendler

*Uniform Resource Identifier.* Retrieved Online April 25, 2003: 

http://www.wikipedia.org/wiki/URI

*W3C Semantic Web Activity.* Retrieved Online from WWW Consortium April 25, 2003: 

http://www.w3.org/2001/12/semweb-fin/w3csw
Appendix
The Semantic Web

A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities

By Tim Berners-Lee, James Hendler and Ora Lassila

The entertainment system was belting out the Beatles' “We Can Work It Out” when the phone rang. When Pete answered, his phone turned the sound down by sending a message to all the other local devices that had a volume control. His sister, Lucy, was on the line from the doctor’s office: "Mom needs to see a specialist and then has to have a series of physical therapy sessions. Biweekly or something. I’m going to have my agent set up the appointments." Pete immediately agreed to share the chauffeuring.

At the doctor's office, Lucy instructed her Semantic Web agent through her handheld Web browser. The agent promptly retrieved information about Mom's prescribed treatment from the doctor's agent, looked up several lists of providers, and checked for the ones in-plan for Mom's insurance within a 20-mile radius of her home and with a rating of excellent or very good on trusted rating services. It then began trying to find a match between available appointment times (supplied by the agents of individual providers through their Web sites) and Pete's and Lucy's busy schedules. (The emphasized keywords indicate terms whose semantics, or meaning, were defined for the agent through the Semantic Web.)

In a few minutes the agent presented them with a plan. Pete didn't like it—University Hospital was all the way across town from Mom's place, and he'd be driving back in the middle of rush hour. He set his own agent to redo the search with stricter preferences about location and time. Lucy’s agent, having complete trust in Pete's agent in the context of the present task, automatically assisted by supplying access certificates and shortcuts to the data it had already sorted through.

Almost instantly the new plan was presented: a much closer clinic and earlier times—but there were two warning notes. First, Pete would have to reschedule a couple of his less important appointments. He checked what they were—not a problem. The other was something about the insurance company's list failing to include this provider under physical therapists: "Service type and insurance plan status securely verified by other means," the agent reassured him. "(Details?)"

Lucy registered her assent at about the same moment Pete was muttering, "Spare me the details," and it was all set. (Of course, Pete couldn't resist the details and later that night had his agent explain how it had found that provider even though it wasn't on the proper list.)

Expressing Meaning

Pete and Lucy could use their agents to carry out all these tasks thanks not to the World Wide Web of today but rather the Semantic Web that it will evolve into tomorrow. Most of the Web's content today is designed for humans to read, not for computer programs to manipulate meaningfully. Computers can adeptly parse Web pages for layout and routine processing—here a header, there a link to another page—but in general, computers have no reliable way to process the semantics: this is the home page of the Hartman and Strauss Physio Clinic, this link goes to Dr. Hartman's curriculum vitae.

The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users. Such an agent coming to the clinic's Web page will know not just that the page has keywords such as "treatment, medicine, physical, therapy" (as might be encoded today) but also that Dr. Hartman works at this clinic on Mondays, Wednesdays and Fridays and that the script takes a date range in yyyy-mm-dd format and returns appointment times. And it will "know" all this without needing artificial intelligence on the scale of 2001's Hal or Star Wars's C-3PO. Instead these semantics were encoded into the Web page when the clinic's office manager (who never took Comp Sci 101) massaged it into shape using off-the-shelf software for writing Semantic Web pages along with resources listed on the Physical Therapy Association's site.

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. The first steps in weaving the Semantic Web into the structure of the existing Web are already under way. In the near future, these developments will usher in significant new...
functionality as machines become much better able to process and "understand" the data that they merely display at present.

The essential property of the World Wide Web is its universality. The power of a hypertext link is that "anything can link to anything." Web technology, therefore, must not discriminate between the scribbled draft and the polished performance, between commercial and academic information, or among cultures, languages, media and so on. Information varies along many axes. One of these is the difference between information produced primarily for human consumption and that produced mainly for machines. At one end of the scale we have everything from the five-second TV commercial to poetry. At the other end we have databases, programs and sensor output. To date, the Web has developed most rapidly as a medium of documents for people rather than for data and information that can be processed automatically. The Semantic Web aims to make up for this.

Like the Internet, the Semantic Web will be as decentralized as possible. Such Web-like systems generate a lot of excitement at every level, from major corporation to individual user, and provide benefits that are hard or impossible to predict in advance. Decentralization requires compromises: the Web had to throw away the ideal of total consistency of all of its interconnections, ushering in the infamous message "Error 404: Not Found" but allowing unchecked exponential growth.

Knowledge Representation

For the semantic web to function, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning. Artificial-intelligence researchers have studied such systems since long before the Web was developed. Knowledge representation, as this technology is often called, is currently in a state comparable to that of hypertext before the advent of the Web: it is clearly a good idea, and some very nice demonstrations exist, but it has not yet changed the world. It contains the seeds of important applications, but to realize its full potential it must be linked into a single global system.

Traditional knowledge-representation systems typically have been centralized, requiring everyone to share exactly the same definition of common concepts such as "parent" or "vehicle." But central control is stifling, and increasing the size and scope of such a system rapidly becomes unmanageable.

Moreover, these systems usually carefully limit the questions that can be asked so that the computer can answer reliably—or answer at all. The problem is reminiscent of Gödel's theorem from mathematics: any system that is complex enough to be useful also encompasses unanswerable questions, much like sophisticated versions of the basic paradox "This sentence is false." To avoid such problems, traditional knowledge-representation systems generally each had their own narrow and idiosyncratic set of rules for making inferences about their data. For example, a genealogy system, acting on a database of family trees, might include the rule "a wife of an uncle is an aunt." Even if the data could be transferred from one system to another, the rules, existing in a completely different form, usually could not.

Semantic Web researchers, in contrast, accept that paradoxes and unanswerable questions are a price that must be paid to achieve versatility. We make the language for the rules as expressive as needed to allow the Web to reason as widely as desired. This philosophy is similar to that of the conventional Web: early in the Web's development, detractors pointed out that it could never be a well-organized library; without a central database and tree structure, one would never be sure of finding everything. They were right. But the expressive power of the system made vast amounts of information available, and search engines (which would have seemed quite impractical a decade ago) now produce remarkably complete indices of a lot of the material out there. The challenge of the Semantic Web, therefore, is to provide a language that expresses both data and rules for reasoning about the data and that allows rules from any existing knowledge-representation system to be exported onto the Web.

Adding logic to the Web—the means to use rules to make inferences, choose courses of action and answer questions—is the task before the Semantic Web community at the moment. A mixture of mathematical and engineering decisions complicate this task. The logic must be powerful enough to describe complex properties of objects but not so powerful that agents can be tricked by being asked to consider a paradox. Fortunately, a large majority of the information we want to express is along the lines of "a hex-head bolt is a type of machine bolt," which is readily written in existing languages with a little extra vocabulary.

Two important technologies for developing the Semantic Web are already in place: eXtensible Markup Language (XML) and the Resource Description Framework (RDF). XML lets everyone create their own tags—hidden labels such as or that annotate Web pages or sections of text on a page. Scripts, or programs, can make use of these tags in sophisticated ways, but the script writer has to know what the page writer uses each tag for. In short, XML allows users to add arbitrary structure to their documents but says nothing about what the structures mean.
The Semantic Web will enable machines to COMPREHEND semantic documents and data, not human speech and writings.

The triples of RDF form webs of information about related things. 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. For example, imagine that we have access to a variety of databases with information about people including their addresses. If we want to find people living in a specific zip code, we need to know which fields in each database represent names and which represent zip codes. RDF can specify that "(field 5 in database A) (is a field of type) (zip code)," using URIs rather than phrases for each term.

Ontologies

Of course, this is not the end of the story, because two databases may use different identifiers for what is in fact the same concept, such as zip code. A program that wants to compare or combine information across the two databases has to know that these two terms are being used to mean the same thing. Ideally, the program must have a way to discover such common meanings for whatever databases it encounters.

A solution to this problem is provided by the third basic component of the Semantic Web, collections of information called ontologies. In philosophy, an ontology is a theory about the nature of existence, of what types of things exist; ontology as a discipline studies such theories. Artificial intelligence and Web researchers have co-opted the term for their own jargon, and for them an ontology is a document or file that formally defines the relations among terms. The most typical kind of ontology for the Web has a taxonomy and a set of inference rules.

The taxonomy defines classes of objects and relations among them. For example, an address may be defined as a type of location, and city codes may be defined to apply only to locations, and so on. Classes, subclasses, and relations among entities are a very powerful tool for Web use. We can express a large number of relations among entities by assigning properties to classes and allowing subclasses to inherit such properties. If city codes must be of type city and cities generally have Web sites, we can discuss the Web site associated with a city code even if no database links a city code directly to a Web site.

Inference rules in ontologies supply further power. An ontology may express the rule "If a city code is associated with a state code, and an address uses that city code, then that address has the associated state code." A program could then readily deduce, for instance, that a Cornell University address, being in Ithaca, must be in New York State, which is in the U.S., and therefore should be formatted to U.S. standards. The computer doesn't truly "understand" any of this information, but it can now manipulate the terms much more effectively in ways that are useful and meaningful to the human user.

With ontology pages on the Web, solutions to terminology (and other) problems begin to emerge. The meaning of terms or XML codes used on a Web page can be defined by pointers from the page to an ontology. Of course, the same problems as before now arise if I point to an ontology that defines addresses as containing a zip code and you point to one that uses postal code. This kind of confusion can be resolved if ontologies (or other Web services) provide equivalence relations: one or both of our ontologies may contain the information that my zip code is equivalent to your postal code.
Our scheme for sending in the clowns to entertain my customers is partially solved when the two databases point to different definitions of address. The program, using distinct URIs for different concepts of address, will not confuse them and in fact will need to discover that the concepts are related at all. The program could then use a service that takes a list of postal addresses (defined in the first ontology) and converts it into a list of physical addresses (the second ontology) by recognizing and removing post office boxes and other unsuitable addresses. The structure and semantics provided by ontologies make it easier for an entrepreneur to provide such a service and can make its use completely transparent.

Ontologies can enhance the functioning of the Web in many ways. They can be used in a simple fashion to improve the accuracy of Web searches—the search program can look for only those pages that refer to a precise concept instead of all the ones using ambiguous keywords. More advanced applications will use ontologies to relate the information on a page to the associated knowledge structures and inference rules. An example of a page marked up for such use is online at http://www.cs.umd.edu/~hendler. If you send your Web browser to that page, you will see the normal Web page entitled "Dr. James A. Hendler." As a human, you can readily find the link to a short biographical note and read there that Hendler received his Ph.D. from Brown University. A computer program trying to find such information, however, would have to be very complex to guess that this information might be in a biography and to understand the English language used there.

For computers, the page is linked to an ontology page that defines information about computer science departments. For instance, professors work at universities and they generally have doctorates. Further markup on the page (not displayed by the typical Web browser) uses the ontology's concepts to specify that Hendler received his Ph.D. from the entity described at the URI http://www.brown.edu — the Web page for Brown. Computers can also find that Hendler is a member of a particular research project, has a particular e-mail address, and so on. All that information is readily processed by a computer and could be used to answer queries (such as where Dr. Hendler received his degree) that currently would require a human to sift through the content of various pages turned up by a search engine.

In addition, this markup makes it much easier to develop programs that can tackle complicated questions whose answers do not reside on a single Web page. Suppose you wish to find the Ms. Cook you met at a trade conference last year. You don’t remember her first name, but you remember that she worked for one of your clients and that her son was a student at your alma mater. An intelligent search program can sift through all the pages of people whose name is “Cook” (sidestepping all the pages relating to cooks, cooking, the Cook Islands and so forth), find the ones that mention working for a company that’s on your list of clients and follow links to Web pages of their children to track down if any are in school at the right place.

**Agents**

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 promotes this synergy: even agents that were not expressly designed to work together can transfer data among themselves when the data come with semantics.

Agents

An important facet of agents’ functioning will be the exchange of “proofs” written in the Semantic Web’s unifying language (the language that expresses logical inferences made using rules and information such as those specified by ontologies). For example, suppose Ms. Cook’s contact information has been located by an online service, and to your great surprise it places her in Johannesburg. Naturally, you want to check this, so your computer asks the service for a proof of its answer, which it promptly provides by translating its internal reasoning into the Semantic Web’s unifying language. An inference engine in your computer readily verifies that this Ms. Cook indeed matches the one you were seeking, and it can show you the relevant Web pages if you still have doubts. Although they are still far from plumbing the depths of the Semantic Web’s potential, some programs can already exchange proofs in this way, using the current preliminary versions of the unifying language.

Another vital feature will be digital signatures, which are encrypted blocks of data that computers and agents can use to verify that the attached information has been provided by a specific trusted source. You want to be quite sure that a statement sent to your accounting program that you owe money to an online retailer is not a forgery generated by the computer-savvy teenager next door. Agents should be skeptical of assertions that they read on the Semantic Web until they have checked the sources of information. (We wish more people would learn to do this on the Web as it is!)

Many automated Web-based services already exist without semantics, but other programs such as agents have no way to locate one that will perform a specific function. This process, called service discovery, can happen only when there is a common language to describe a service in a way that lets other agents “understand” both the function offered and how to take advantage of it. Services and agents can advertise their function by, for example, depositing such descriptions in directories analogous to the Yellow Pages.

Some low-level service-discovery schemes are currently available, such as Microsoft’s Universal Plug and Play, which

http://www.sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC58... 5/24/2004
focuses on connecting different types of devices, and Sun Microsystems's Jini, which aims to connect services. These initiatives, however, attack the problem at a structural or syntactic level and rely heavily on standardization of a predetermined set of functionality descriptions. Standardization can only go so far, because we can't anticipate all possible future needs.

The Semantic Web, in contrast, is more flexible. The consumer and producer agents can reach a shared understanding by exchanging ontologies, which provide the vocabulary needed for discussion. Agents can even "bootstrap" new reasoning capabilities when they discover new ontologies. Semantics also makes it easier to take advantage of a service that only partially matches a request.

A typical process will involve the creation of a "value chain" in which subassemblies of information are passed from one agent to another, each one "adding value," to construct the final product requested by the end user. Make no mistake: to create complicated value chains automatically on demand, some agents will exploit artificial-intelligence technologies in addition to the Semantic Web. But the Semantic Web will provide the foundations and the framework to make such technologies more feasible.

Putting all these features together results in the abilities exhibited by Pete's and Lucy's agents in the scenario that opened this article. Their agents would have delegated the task in piecemeal fashion to other services and agents discovered through service advertisements. For example, they could have used a trusted service to take a list of providers and determine which of them are in-plan for a specified insurance plan and course of treatment. The list of providers would have been supplied by another search service, etc. These activities formed chains in which a large amount of data distributed across the Web (and almost worthless in that form) was progressively reduced to the small amount of data of high value to Pete and Lucy—a plan of appointments to fit their schedules and other requirements.

In the next step, the Semantic Web will break out of the virtual realm and extend into our physical world. URIs can point to anything, including physical entities, which means we can use the RDF language to describe devices such as cell phones and TVs. Such devices can advertise their functionality—what they can do and how they are controlled—much like software agents. Being much more flexible than low-level schemes such as Universal Plug and Play, such a semantic approach opens up a world of exciting possibilities.

For instance, what today is called home automation requires careful configuration for appliances to work together. Semantic descriptions of device capabilities and functionality will let us achieve such automation with minimal human intervention. A trivial example occurs when Pete answers his phone and the stereo sound is turned down. Instead of having to program each specific appliance, he could program such a function once and for all to cover every local device that advertises having a volume control — the TV, the DVD player and even the media players on the laptop that he brought home from work this one evening.

The first concrete steps have already been taken in this area, with work on developing a standard for describing functional capabilities of devices (such as screen sizes) and user preferences. Built on RDF, this standard is called Composite Capability/Preference Profile (CC/PP). Initially it will let cell phones and other nonstandard Web clients describe their characteristics so that Web content can be tailored for them on the fly. Later, when we add the full versatility of languages for handling ontologies and logic, devices could automatically seek out and employ services and other devices for added information or functionality. It is not hard to imagine your Web-enabled microwave oven consulting the frozen-food manufacturer's Web site for optimal cooking parameters.

Evolution of Knowledge

The semantic web is not "merely" the tool for conducting individual tasks that we have discussed so far. In addition, if properly designed, the Semantic Web can assist the evolution of human knowledge as a whole.

Human endeavor is caught in an eternal tension between the effectiveness of small groups acting independently and the need to mesh with the wider community. A small group can innovate rapidly and efficiently, but this produces a subculture whose concepts are not understood by others. Coordinating actions across a large group, however, is painfully slow and takes an enormous amount of communication. The world works across the spectrum between these extremes, with a tendency to start small—from the personal idea—and move toward a wider understanding over time.

An essential process is the joining together of subcultures when a wider common language is needed. Often two groups independently develop very similar concepts, and describing the relation between them brings
great benefits. Like a Finnish-English dictionary, or a weights-and-measures conversion table, the relations allow communication and collaboration even when the commonality of concept has not (yet) led to a commonality of terms.

The Semantic Web, in naming every concept simply by a URI, lets anyone express new concepts that they invent with minimal effort. Its unifying logical language will enable these concepts to be progressively linked into a universal Web. This structure will open up the knowledge and workings of humankind to meaningful analysis by software agents, providing a new class of tools by which we can live, work and learn together.

Further Information:

An enhanced version of this article is on the Scientific American Web site, with additional material and links.

World Wide Web Consortium (W3C): www.w3.org/
W3C Semantic Web Activity: www.w3.org/2001/sw/
DARPA Agent Markup Language (DAML) home page: www.daml.org/

© 1996-2004 Scientific American, Inc. All rights reserved. Reproduction in whole or in part without permission is prohibited.
RDF and Metadata

By Tim Bray

This article has now been updated to incorporate changes in the RDF spec and the growth of the RDF community. You can find a newer version here: What is RDF?

The Right Way to Find Things

RDF stands for Resource Description Framework. RDF is built for the Web, but let's leave Web-land behind for a few minutes and think about how we find things in the real world.

Scenario 1: The Library

You're in a library to find books on raising donkeys as pets. In most libraries these days you'd use the computer lookup system, basically an electronic version of the old card file. This system allows you to list books by author, title, call-number, and subject. The list includes the date, author, title, and lots of other useful information, including (most important of all) where each book is.

Scenario 2: The Video Store

You're in a video store and you want a movie by John Huston. A large modern video store offers a lookup facility that's similar to the library's. Of course, the things you can search on are different (director, actors, and so on) but the results are more or less the same.

Scenario 3: The Phone Book

You're working late at a customer's office in South Denver, and it seems that a pizza is essential if work is to continue. Fortunately, every office comes equipped with a set of Yellow Pages that, properly used, can lead to quick pizza delivery.

The Common Thread

What do all these situations have in common, and what differences lie behind the scenes? First of all, each of these systems is based on metadata, or information about information. In each case, you need a piece of information (the book's location, the video's name, the pizza joint's phone number). In each case, you use metadata (information about information) to get it.

We're all used to this stuff; the usual setup is that metadata comes in named chunks (subject, director, business category) that associate lookup information ("donkeys", "John Huston", "Pizza, South Side") with the real info that you're after.
About the DMTF

Distributed Management Task Force, Inc. (DMTF), developer of the Common Information Model (CIM), is the industry organization leading the development, adoption, and interoperability of management standards and initiatives for enterprise and Internet environments.

CIM is the breakthrough standard for the exchange of management information in a platform-independent and technology-neutral way, streamlining integration and reducing costs by enabling end-to-end multi-vendor interoperability in management systems. Key technology vendors and affiliated standards groups that implement CIM enable a more integrated, cost-effective and less crisis-driven approach to management.

For more in-depth information about the DMTF and its standards, see our backgrounder (PDF).

To find out about DMTF's latest news, visit the News Room.

For more information on DMTF membership and how to join, check out our Join Us page.
W3C Semantic Web Activity

Marja-Riitta Koivunen and Eric Miller
{marja,emiller}@w3.org
W3C/MIT

1. Introduction

The World Wide Web contains huge amounts of information created by many different organizations, communities and individuals for many different reasons. Users of the Web can easily access this information by specifying URI addresses, searching, and following links to find other related resources. The simplicity of usage is a key aspect that made Web so popular; so popular in fact, it is hard to imagine life without it anymore.

This simplicity of the current web has a price. It is very easy to get lost, or discover irrelevant or unrelated information with all that is available. For instance, if we search for something as simple as research papers written by a person named "Eric Miller" we will find all kinds of other information starting from Web diaries or phonebooks that mention "Eric" and/or "Miller" somewhere. Similar problems arise if you search for resources about "Marja" as "Marja" could equally well refer to a first name of a person, or to a berry in Finnish.

The goal of the Semantic Web is to develop enabling standards and technologies designed to help machines understand more information on the Web so that they can support richer discovery, data integration, navigation, and automation of tasks. With Semantic Web we not only receive more exact results when searching for information, but also know when we can integrate information from different sources, know what information to compare, and can provide all kinds of automated services in different domains from future home and digital libraries to electronic business and health services [Berners-Lee 2001].

With the Semantic Web we can associate semantically rich, descriptive information with any resource. For instance, by adding metadata about document creation, we can search for documents that have metadata specifying Eric Miller as a "writer". With a bit more metadata we can also search only documents under the category of "research papers". In Semantic Web we not only provide URIs for documents as we have done in the past, but to people, concepts and relationships. In the above case for example, by giving unique identifiers to the person, the role "writer" and the concept of "research paper" we make very clear who the person is, and the corresponding relation between this person and a particular document. Furthermore, by making clear which person we are talking about we can differentiate the plethora of "Eric Miller's". We can also combine descriptive information from different sites and learn more about this person in differing contexts; in his roles as an author, as a manager, as a developer, etc.
The Semantic Web: A Primer
By Edd Dumbill

Introduction

The Semantic Web lies at the heart of Tim Berners-Lee's vision for the future of the World Wide Web. Along with others at the W3C Berners-Lee is working on the infrastructure for this next stage of the Web's life. But the question "What is the Semantic Web?" is being asked with increasing frequency. While mainstream media is content with a high level view, XML developers want to know more, and they want to discover the substance behind the vision.

Accusations of fuzziness about the Semantic Web (SW) have been levelled at Berners-Lee, and it is certainly true that he has yet to deliver the long-awaited "Semantic Web Whitepaper." However, there are interesting suggestions in his Semantic Web Roadmap text which give details about the direction he wants to go. Furthermore, SW activity at the W3C and MIT/LCS has been increasing in intensity, and community involvement with RDF, a key SW technology, has increased markedly over recent months.

In his Roadmap document, Berners-Lee contrasts the Semantic Web with the existing, merely human-readable Web: "the Semantic Web approach instead develops languages for expressing information in a machine processable form." This is perhaps the best way of summing up the Semantic Web -- technologies for enabling machines to make more sense of the Web, with the result of making the Web more useful for humans.

Given that goal, it's unsurprising that the scope of the Semantic Web vision is somewhat broad and ill-defined. There are many ways to solve the problem and many technologies that can be employed. Some XML developers have a "well-formed" prejudice against, as they cheerily call it, the "Pedantic Web" because of the strong links with RDF (not everyone's favorite technology) and the definite view taken on URIs. But to perceive the SW only in this light would be a mistake. Technical peeves aside, the value of the Semantic Web is to solve real problems in communication. First and foremost this means radically improving our ability to find, sort, and classify information: an activity that takes up a large part of our time.

The development of the Semantic Web is well underway. This development is occurring in at least two areas: from the infrastructural, all-embracing, position as espoused by the W3C/MIT and other academically-focused organizations, and also in a more directed application-specific fashion by those using web technologies for electronic business.

Abstract

This memo examines some current standards for data exchange formats and protocols related to security, authorization policy and trust management. The goal of this memo is to explore the interaction between various security- and trust-related mechanisms, and to identify enough of the purpose and content of these existing standards to indicate how they might be integrated in a Semantic Web trust and policy management framework.

SWAD Europe

This memo has been prepared for the SWAD-Europe project, as a strand of CLRC/RAL participation.

© 2002 CCLRC

Table of Contents

1. Introduction
   1.1 Structure of this memo
2. Interactions between autonomous communicating parties
   2.1 A trust framework for simple data transfer
   2.2 Example of an extended transaction
3. Resource Description Framework (RDF)
4. Secure MIME (S/MIME)
   4.1 Content and concepts
      4.1.1 Signed data
      4.1.2 Encrypted (enveloped) data
      4.2 Syntax
      4.3 Relationship to trust management
      4.4 Related specifications
5. Open PGP (OpenPGP)
   5.1 Content and concepts
      5.1.1 Signed and encrypted data
      5.1.2 Cleartext signature framework
      5.1.3 Web of trust
      5.2 Syntax
      5.3 Relationship to trust management
      5.4 Related specifications
6. XML digital signatures (XMLDSIG)
   6.1 Content and concepts
   6.2 Syntax
   6.3 Relationship to trust management
   6.4 Related specifications
   6.5 Note

The Grid: An Application of the Semantic Web

Carole Goble
Department of Computer Science
University of Manchester
Oxford Road
Manchester, M13 9PL, UK
carole@cs.man.ac.uk

David De Roure
Dept of Electronics & Computer Science
University of Southampton
Southampton,
SO17 1BJ, UK
dder@ecs.soton.ac.uk

Abstract

The Grid is an emerging platform to support on-demand “virtual organisations” for coordinated resource sharing and problem solving on a global scale. The application thrust is large-scale scientific endeavour, and the scale and complexity of scientific data presents challenges for databases. The Grid is beginning to exploit technologies developed for Web Services and to realise its potential it also stands to benefit from Semantic Web technologies; conversely, the Grid and its scientific users provide application pull which will benefit the Semantic Web.

What is the Grid?

The Grid is “flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources—what we refer to as virtual organizations.” Foster et al, 2001[1].

Large-scale science and engineering are undertaken through the interaction of people, heterogeneous computing resources, information systems, and instruments, all of which are geographically and organizationally dispersed. “The Grid” is an emerging platform to support coordinated resource sharing and problem solving on a global scale for data-intensive and computer-intensive applications [1]. The name arose from an analogy with an electricity power grid: computing and data resources would be delivered over the Internet seamlessly, transparently and dynamically as and when needed, just like electricity. Thus, the overall motivation for Grids is to facilitate the routine interactions of resources in order to support large-scale science and engineering.

The Grid was originally focused on sharing computational power and resources for advanced science and engineering. The ‘metacomputing’ projects of the early 1990s set out to build virtual supercomputers using networked computer systems. The target applications were, and primarily continue to be, large-scale science. For example, trans-national experiments, such as the particle physicist’s quest to find the Higgs boson by building a Large Hadron Collider. This device generates petabytes of data in a few seconds and the complex analyses can take months of computational processing [2].

Increasing the computational power by combining large numbers of geographically diverse systems raises the issues of scalability and heterogeneity. Scalability brings a number of challenges: the inevitability of failure of components, the need for automation, the need to exploit the locality of resources due to network latency, and the increasing number of organisational boundaries, emphasising authentication and trust issues. Larger scale applications may also result from the composition of other applications, which increases the complexity of systems. Heterogeneity is addressed by middleware, such as the Globus Toolkit [3], to provide uniformity through a standard set of interfaces to the underlying resources.

Early Grid middleware exploits a range of protocols such as LDAP for directory services and file store queries [4], GridFTP for large-scale reliable data transfer and SSL for security. Higher level functionality, such as tolerant scalable data replication [5], exploit these. Some attention has been paid to data intensive rather than compute intensive Grid use; for example, the Storage Request Broker provides applications with uniform access to distributed file storage [6]. However, research and development activities relating to the Grid have generally focused on applications where data was stored in files, and there is little support for transactions, relational database access or distributed query processing [7].

The Grid community is now actively developing fundamental mechanisms for the interaction of any kind of resource including documents, databases, instruments, archives and people. Support for data interaction is focused on consistent access to databases from Grid applications and coordinated access to databases from Grid applications [8].
March 28, 2002

http://www.intelligententerprise.com/020328/506decision1_1.jhtml

The Semantic Web

By Seth Grimes

Tim Berners-Lee invented the World Wide Web in 1989 by creating a language for presenting and linking content, an information-interchange protocol, and basic client/server software. By 1994, the year he founded the World Wide Web Consortium (W3C) at the Massachusetts Institute of Technology, his attention had turned to embedding meaning in his creation.

This interest was long before the birth of well-loved sites (depending on your interests) such as SlashDot, The Onion, and Gap.com Applying business intelligence (BI) and knowledge management (KM) jargon, these kinds of sites are "stovepipe" systems, designed and used primarily for a single purpose, whether furthering technology development, commentary, or mass-market retail. Regardless of their purposes, these isolated systems constitute "islands of information." Much of the work in the BI and KM worlds is to build bridges among such islands. The thought is that a whole is greater than the sum of its parts, that acquiring new, varied data sources will expand the scope of analyses and strengthen their reliability. Fully integrating, rather than just occasionally drawing from those sources and finding relationships in the integrated whole, can bring additional rewards.

Building Bridges

In the BI world, developers build bridges using harmonized metadata and coordinated access methods. KM adds value by classifying and cataloging unstructured information to create analytic metadata. But, BI bridges are often one-directional and purpose-built, designed to support ad hoc queries and periodic reports, aggregation and consolidation via scorecards and portals, and electronic data interchange (EDI) extractions from specific sources to specific users. Generalized interchange, where information can flow in more than one direction and the user and source haven't negotiated protocols in advance, requires standardized, published interfaces with common metadata structures and definitions.

Interchange standards such as the extensible markup language (XML) and efforts such as the Dublin Core metadata project - which seeks to promote "the widespread adoption of interoperable metadata standards and developing specialized metadata vocabularies for describing resources that enable more intelligent information discovery systems" - are significant advances on the generalization road.

Searching for Meaning

An Idiot's Guide to the Resource Description Framework

Renato Iannella
DSTC Pty Ltd
The University of Queensland, 4072, AUSTRALIA
Phone: +61 7 3365 4310 Fax: +61 7 3365 4311
Email: renato@dstc.edu.au

First Published: 1998-09-03
Updated: 1999-01-25
Status: Conforms with RDF Syntax Specification 1999-01-05

Introduction

The Resource Description Framework (RDF) - developed by the World-Wide Web Consortium (W3C) - provides the foundation for metadata interoperability across different resource description communities. One of the major obstacles facing the resource description community is the multiplicity of incompatible standards for metadata syntax and schema definition languages. This has led to the lack of, and low deployment of, cross-discipline applications and services for the resource description communities. RDF provides a solution to these problems via a Syntax specification (W3C, 1999a) and Schema specification (W3C, 1998a).

RDF is based on Web technologies and, as a result, is lightweight and highly deployable. RDF provides interoperability between applications that exchange metadata and is targeted for many application areas including; resource description, site-maps, content rating, electronic commerce, collaborative services, and privacy preferences. RDF is the result of members of these communities reaching consensus on their syntactical needs and deployment efforts.

The objective of RDF is to support the interoperability of metadata. RDF allows descriptions of Web resources - any object with a Uniform Resource Identifier (URI) as its address - to be made available in machine understandable form. This enables the semantics of objects to be expressible and exploitable. Once highly deployed, this will enable services to develop processing rules for automated decision-making about Web resources.

RDF is based on a concrete formal model utilising directed graphs that elude to the semantics of resource description. The basic concept is that a Resource is described through a collection of Properties called an RDF Description. Each of these Properties has a Property Type and Value. Any resource can be described with RDF as long as the resource is identifiable with a URI as shown in Figure 1.

Internet History

1969 - Birth of a Network

The Internet as we know it today, in the mid-1990s, traces it origins back to a Defense Department project in 1969. The subject of the project was wartime digital communications. At that time the telephone system was about the only theater-scale communications system in use. A major problem had been identified in its design - its dependence on switching stations that could be targeted during an attack. Would it be possible to design a network that could quickly reroute digital traffic around failed nodes? A possible solution had been identified in theory. That was to build a "web" of datagram network, called an "catenet", and use dynamic routing protocols to constantly adjust the flow of traffic through the catenet. The Defense Advanced Research Projects Agency (DARPA) launched the DARPA Internet Program.

1970s - Infancy

DARPA Internet, largely the plaything of academic and military researchers, spent more than a decade in relative obscurity. As Vietnam, Watergate, the Oil Crisis, and the Iranian Hostage Crisis rolled over the nation, several Internet research teams proceeded through a gradual evolution of protocols. In 1975, DARPA declared the project a success and handed its management over to the Defense Communications Agency. Several of today's key protocols (including IP and TCP) were stable by 1980, and adopted throughout ARPANET by 1983.

Mid 1980s - The Research Net

Let's outline key features, circa-1983, of what was then called ARPANET. A small computer was a PDP-11/45, and a PDP-11/45 does not fit on your desk. Some sites had a hundred computers attached to the Internet. Most had a dozen or so, probably with something like a VAX doing most of the work - mail, news, EGP routing. Users did their work using DEC VT-100 terminals. FORTRAN was the word of the day. Few companies had Internet access, relying instead on SNA and IBM mainframes. Rather, the Internet community was dominated by universities and military research sites. It's most popular service was the rapid email it made possible with distant colleagues. In August 1983, there were 562 registered ARPANET hosts (RFC 1296).

UNIX deserves at least an honorable mention, since almost all the initial Internet protocols were developed first for UNIX, largely due to the availability of kernel source (for a price) and the relative ease of implementation (relative to things like VMS or MVS). The University of California at Berkeley (UCB) deserves special mention, because their Computer Science Research Group (CSRG) developed the BSD variants of AT&T's UNIX operating system. BSD UNIX and its derivatives would become the most common Internet programming platform.
Introduction to Semantic Web Technologies

This is intended to give someone new to the Semantic Web a basic overview of the technologies involved, and a guide to where to go to find out more.

The basis for the augmented functionality of the Semantic Web is

- A global naming scheme (URIs);
- A standard syntax for describing data (rdf);
- A standard means of describing the properties of that data (rdf-schema);
- A standard means of describing relationships between data items (ontologies);
- The means to support trust and security.

Global naming scheme

If any Semantic Web application is to be able to access and use data from any other such application, every data object and every data schema/model must have a unique and universal means of identification. These identifiers are called URIs (Universal Resource Identifiers).

top

Standard Syntax - RDF

The computer industry has agreed, by and large, to use XML (Extensible Markup Language) to represent not only human readable documents, but data in general. The XML standards give a syntactic structure for describing data. Unfortunately, XML can be used in many different ways to describe the same data. This makes it too open and arbitrary to support the type of widespread and ad hoc data integration envisaged for the Semantic Web. The semantic web vision proposes to represent machine processable information using RDF (Resource Description Framework), which extends XML. RDF defines a general common data model that adheres to web principles. The W3C are strong supporters of this approach.

RDF provides a consistent, standardised way of describing and querying internet resources, from text pages and graphics to audio files and video clips. It gives syntactic interoperability, and provides the base layer for building a Semantic Web. RDF defines a directed graph of relationships. These are represented by object-attribute-value triples i.e. an object O has an attribute A with value V, often written as A(O,V). For instance, telnet(janet_bruten, 3128700) represents the fact that the person object Janet Bruten has the telnet number 312-8700.
Introduction to Semantic Web Technologies

Figure 1: A simple directed graph

Further information

top

Describing properties - RDF Schema

RDF itself is a composable and extensible standard for building data models. To support the definition of a specific vocabulary for a data model, which can itself be published, another layer is required. RDF schema allows a designer to define and publish the vocabulary used by an RDF data model, i.e define the data objects and their attributes. For instance, it might define that people have a phone attribute. RDFS also uses class and subclass, so that hp_employee could be defined as a sub-class of person.

Both RDF and RDF-Schema are based on XML and XML-Schema. The existence of standards for describing data (RDF) and data attributes (RDF Schema) enables the development of a set of readily available tools to read and exploit data from multiple sources. The degree to which different applications can share and exploit data is sometimes termed syntactic interoperability. The more standardised and widespread these data manipulation tools are the higher the degree of syntactic interoperability, and the easier and more attractive it becomes to use the Semantic Web approach as opposed to a point solution.

Further Information

top

Describing relationships between data items

If data is to be truly "understandable" by multiple applications, and therefore become information, semantic interoperability is required. Syntactic interoperability is all about parsing data correctly. Semantic interoperability requires mapping between terms, which in turn requires content analysis. This requires formal and explicit specifications of domain models, which define the terms used and their relationships. Such formal domain models are sometimes called ontologies. Ontologies define data models in terms of classes, subclasses, and properties. For instance, we might define a herbivore to be a subclass of animals that eats plants. Figure 2 shows a very simple example ontology for animals.
Smarter Content Publishing

In: Articles

By Victor Lombardi

Published on August 6, 2002

Do you remember the days before WYSIWYG word processors when you had to markup the text, much like we markup web documents using HTML? I don't. I started word processing using WYSIWYG applications on the Macintosh and only heard about markup in 1993 when I first saw a website. After all those years of refining the word processor to become a more efficient tool, I had to wonder why we reverted to manually creating markup for the web.

One reason is that the web is not print. The need to create hypertext within pages requires more control over documents. Another is that web pages can also be applications, so access to the "guts" of the page is needed to insert programming code. And ideally, we would like to separate presentation from content, enabling us to format the content in different ways for different purposes. How can we achieve all this with the efficiency and usability we've come to expect of other publishing tools?

Content Management Systems

While WYSIWYG web page editors are now available, more of our needs can be met using an entire content management system (CMS). CMS offers us:

- WYSIWYG editing or preview
- The ability to create content components that can be used repeatedly
- A centralized place to edit content
- Separation of presentation and content by using presentation templates
- A user interface customized for the task of content management and publishing

While CMS was once the domain of large companies with big budgets and talented programmers, low cost or free CMS programs such as Movable Type are now powering thousands of simple, personal sites. There is clearly a trend of easier publishing and less costly software becoming available on a widespread basis. The next step in this trend is to make CMS more efficient.

There are several steps in the traditional content management process that require manual work which could be made more efficient. One area is telling the system how to assemble content components into pages. Publishing content onto a page usually requires several steps:

- Input the content itself, along with the corresponding metadata
- Specify how that content should be displayed in relation to similar content, such as its placement in a list and that list's order
- Specify how that group of content should be displayed on a page

If we can provide pre-determined rules for the system to assemble the content components, we could simply
The Resource Description Framework (RDF) is a language for representing information about resources in the World Wide Web. This Primer is designed to provide the reader with the basic knowledge required to effectively use RDF. It introduces the basic concepts of RDF and describes its XML syntax. It describes how to define RDF vocabularies using the RDF Vocabulary Description Language, and gives an overview of some deployed RDF applications. It also describes the content and purpose of other RDF specification documents.
Ontologies Come of Age

Deborah L. McGuinness

Associate Director and Senior Research Scientist
Knowledge Systems Laboratory
Stanford University
Stanford, CA 94305
dlm@ksl.stanford.edu

Abstract
Ontologies have moved beyond the domains of library science, philosophy, and knowledge representation. They are now the concerns of marketing departments, CEOs, and mainstream business. Research analyst companies such as Forrester Research report on the critical roles of ontologies in support of browsing and search for e-commerce and in support of interoperability for facilitation of knowledge management and configuration. One now sees ontologies used as central controlled vocabularies that are integrated into catalogues, databases, web publications, knowledge management applications, etc. Large ontologies are essential components in many online applications including search (such as Yahoo and Lycos), e-commerce (such as Amazon and eBay), configuration (such as Dell and PC-Order), etc. One also sees ontologies that have long life spans, sometimes in multiple projects (such as UMLS, SIC codes, etc.). Such diverse usage generates many implications for ontology environments.

In this paper, we will discuss ontologies and requirements in their current instantiations on the web today. We will describe some desirable properties of ontologies. We will also discuss how both simple and complex ontologies are being and may be used to support varied applications. We will conclude with a discussion of emerging trends in ontologies and their environments and briefly mention our evolving ontology evolution environment.

Introduction: The web’s growing needs
We may be poised for the next major evolution of online environments. In the early days of the web, HTML pages were generated by hand. The pages contained information about how to present information on a page. Early adopters took to the web quickly since it provided a convenient method for information sharing. Arguably, the generation of tools for machine generation and management of web pages allowed the web to really take off. Tool platforms allowed non-technical people to generate and publish web pages quickly and easily. The resulting pages typically included content and display information and targeted human readers (rather than targeting programs or automatic readers).

The web continues to grow at an astounding rate with web pages ubiquitously integrated into many aspects of business and personal life. However, web pages still preserve much of their character of being aimed at human consumption. Thus, applications such as search still require humans to review results pages in order to find the right answer to their queries. While search engine advances such as Google [Google 2000] improve the

Mission and Goals

papiNet is the global initiative to develop, maintain and promote the implemental standard electronic transaction standards to facilitate the flow of information among the parties engaged in the buying, selling, and distribution of forest, paper and wood products.

The set of standards is referred to as the papiNet standard. The standard includes common terminology and standard business documents that support the entire supply chain.

The goals are to improve the reach and richness of communication throughout the supply chain, increase efficiencies in transactions and marketplace activities, and support interoperability among trading partners. The papiNet standards are freely available.

papiNet facilitates open communications and commerce within the forest and paper products industry and across industry boundaries. Critical mass is achieved by key players up and down the global supply chain.

© 2004 AF&PA, IDEAAlliance, and papiNet G.I.E.

webmaster
sematic web, the second-generation web?

[abstract]: Still in a very early stage of development, semantic web vision is designed as the second-generation web, building the foundation of a new information space. The W3C presents the semantic web as an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation. The semantic web is one of emerging and converging technology trends that are shaping the Web's future. But without solving intellectual property issue, there is a danger that the next Internet revolution won't happen. It could mean that the Semantic Web doesn't really happen.

[background]: [ source + references + grafik ]

[keywords]: aggregation, daml, metadata, oil, p2p, rdf, semantic web, syndication, uri, w3c, web service, writeable web, xml

[date]: June 19, 2002

>a new conceptual information space

"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. Leaving aside
the artificial intelligence problem of training machines to behave like people, the Semantic Web approach instead develops languages for expressing information in a machine processable form... The Semantic Web is a web of data, in some ways like a global database."

> from 'Semantic Web roadmap' by Tim Berners-Lee, September, 1998

From a web of a human readable documents to a web of processable data too. "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 using it in various applications... 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."

> from W3C Semantic Web Activity Statement

"The Semantic Web is an extension of the current Web that will allow to find, share, and combine information more easily... Ultimate goal - the design of enabling technologies to support machine facilitated global knowledge exchange." >from 'The Semantic Web' by Eric Miller, W3C Semantic Web Activity Lead. WWW2002, May 2002

semantic web vision:

> the semantic web by eric miller
may 2002
www2002 w3c track

> berners-lee: prepare for next-gen web now
by c. aaley. boston.internet, april 19, 2002
it's the same challenge we faced with the web: bringing the vision into reality

> web founder seeks simplicity
by a. weber. software development magazine. june 2001
globally unique identifiers key to semantics and services on the web

> semantic web to be berners-lee's second success?
by eric van der vlist. xmlhack, may 4, 2001
a report from www10. on semantic web skeptics

> not done yet, says berners-lee
by edd dumbill. o'reilly network, may 2, 2001
about berners-lee opening keynote at www10

> semantic web
by t. berners-lee, j. hendler, o. lassila. scientific american, may 2001
new form of web content meaningful to computers will
In response to a request, a one page looking back on the development of the Web from my point of view.

The World Wide Web: A very short personal history

There have always been things which people are good at, and things computers have been good at, and little overlap between the two. I was brought up to understand this distinction in the 50s and 60s and intuition and understanding were human characteristics, and that computers worked mechanically in tables and hierarchies.

One of the things computers have not done for an organization is to be able to store random associations between disparate things, although this is something the brain has always done relatively well. In 1980 I played with programs to store information with random links, and in 1989, while working at the European Particle Physics Laboratory, I proposed that a global hypertext space be created in which any network-accessible information could be refered to by a single "Universal Document Identifier". Given the go-ahead to experiment by my boss, Mike Sendall, I wrote in 1990 a program called "Worldwideweb", a point and click hypertext editor which ran on the "NeXT" machine. This, together with the first Web server, I released to the High Energy Physics community at first, and to the hypertext and NeXT communities in the summer of 1991. Also available was a "line mode" browser by student Nicola Pellow, which could be run on almost any computer. The specifications of UIDs (now URIs), HyperText Markup Language (HTML) and HyperText Transfer Protocol (HTTP) published on the first server in order to promote wide adoption and discussion.

The dream behind the Web is of a common information space in which we communicate by sharing information. Its universality is essential: the fact that a hypertext link can point to anything, be it personal, local or global, be it draft or highly polished. There was a second part of the dream, too, dependent on the Web being so generally used that it became a realistic mirror (or in fact the primary embodiment) of the ways in which we work and play and socialize. That was that once the state of our interactions was on line, we could then use computers to help us analyse it, make sense of what we are doing, where we individually fit in, and how we can better work together.

The first three years were a phase of persuasion, aided by my colleague and first convert Robert Cailliau, to get the Web adopted. We needed Web clients for other platforms (as the NeXT was not ubiquitous) and browsers Erwise, Viola, Cello and Mosaic eventually came on the scene. We needed seed servers to provide incentive and examples, and all over the world inspired people put up all kinds of things.

Between the summers of 1991 and 1994, the load on the first Web server ("info.cern.ch") rose steadily by a factor of 10 every year. In 1992 academia, and in 1993 industry, was taking notice. I was under pressure to define the future evolution. After much discussion I decided to form the World Wide Web Consortium in September 1994, with a base at MIT the USA, INRIA in France, and now also at Keio University in Japan. The Consortium is a neutral open forum where companies and organizations to whom the future of the Web is important come to discuss and to agree on new common computer protocols. It has been a center for issue raising, design, and decision by consensus, and also a fascinating vantage point from which to view that evolution.

With the dramatic flood of rich material of all kinds onto the Web in the 1990s, the first part of the dream is largely realized, although still very few people in practice have access to intuitive hypertext creation tools. The second part has yet to happen, but there are signs and plans which make us confident. The great need for information about information, to help us categorize, sort, pay for, own information is driving the design of languages for the web designed for processing by machines, rather than people. The web of human-readable document is being merged with a web of machine-understandable data. The potential of the mixture of humans and machines working together and communicating through the web could be immense.
Tim Berners-Lee: A short history of web development

Back to main Bio

http://www.w3.org/People/Berners-Lee/ShortHistory.html
Good

We were standing behind the plate-glass window looking out at the blank horizon. Water lay in sheets on the ground, flooding several Aboriginal humpies to the depth of a foot or more. The owners had heaped their gear on to the roofs. The water was awash with refuse.

A short way off to the west was the old administrator's house, of two storeys, which had since been given over to the community. The roof was still on and there were floors and fireplaces. But the walls, the window sashes and the staircase had all been burnt for firewood.

We looked through this X-ray house into the yellow sunset. On both upper and lower floors sat a ring of dark figures, warming themselves over a smoky fire.

'They don't give a fuck for walls,' said Red, 'but they do like a roof for the rain.'

Bruce Chatwin, *The Songlines*

Meta

SGML represents three interrelated concurrent views of a document:

a) An element structure containing instances of one or more documents types, reflecting, for example, the unformatted document, the document formatted for electronic display, the document formatted as camera-read copy, etc..

b) An entity structure, reflecting the physical organization of the document into separately accessible pieces. Depending on the actual storage system, these pieces could be individual files, locations in memory, parts of a single file, etc.

c) A sequence of characters, formulated according to the rules of SGML to represent both the element structure and entity structure. A character string is the simplest and most universal data structure, hence the most easily interchanged.

Charles F. Goldfarb, *The SGML Handbook*

*Good* and *Meta* change four times per month. Copyright © 1994-2004 Textuality Services, Inc.