Special Section

An Overview of W3C Semantic Web Activity

by Eric Miller and Ralph Swick

Both authors of this article are with the World Wide Web Consortium (W3C). Eric Miller is Semantic Web Activity Lead and can be reached by e-mail at em@w3.org; Ralph Swick is Technology and Society Domain Technical Lead, e-mail: swick@w3.org.

The Semantic Web is an extension of the current Web in which the meaning of information is clearly and explicitly linked from the information itself, better enabling computers and people to work in cooperation. The World Wide Web Consortium (W3C) Semantic Web Activity, in collaboration with a large number of researchers and industrial partners, is tasked with defining enabling standards and technologies to allow data on the Web to be defined and linked in such a way that it can be used for more effective discovery, automation, integration and reuse across various applications. The Web can reach its full potential if it becomes a place where data can be shared and processed by automated tools as well as by people.

The Semantic Web fosters and encourages greater data reuse by making it available for purposes not planned or conceived by the data provider. Suppose you want, for example, to locate news articles published in the previous month about companies headquartered in cities with populations under 500,000 or to compare the stock price of a company with the weather at its home base or to search online product catalogs for an equivalent replacement part for something. The information may be there in the Web, but currently only in a form that requires intensive human processing.

The Semantic Web will allow two things. First, it will allow this information to surface in the form of data, so that a program doesn't have to strip the formatting, pictures and ads off a Web page and guess at how the remaining page markup denotes the relevant bits of information. Second, it will allow people to write (or generate) files that explain – to a machine – the relationship between different sets of data. For example, one will be able to make a "semantic link" between a database with a "zip-code" column and a form with a "zip" field to tell the machines that they do actually mean the same thing. This will allow machines to follow links and facilitate the integration of data from many different sources. When the relationships among data are fully accessible to our machines, our machines will be able to help us browse those relationships and interpret the data as well as assess the appropriateness of the data for our intended purposes.

This notion of being able to "semantically link" various resources, such as documents, images, people or concepts, is an important one. With semantic links we can move from the current Web of simple relationships like "links-to" to a more expressive, semantically rich Web – a Web where we can incrementally add meaning and express a whole new set of relationships (hasLocation, worksFor, isAuthorOf, hasSubjectOf, dependsOn, etc.). These relationships can make explicit the particular contextual relationships that are either implicit or expressed in the current Web only in prose that is impossible for machines to interpret. This enhancement in turn opens doors for a whole new set of effective information integration, management and automated services.

The Semantic Web is a place where strongly controlled (or centralized) metadata vocabulary registries can flourish alongside special-purpose, small community or even "private" vocabularies. The Semantic Web technology supports free comingling of vocabularies as well as the ad-hoc definition of new relationships to construct data descriptions. In addition, instructions for processing data in specific ways can be expressed in the Semantic Web using the same technologies used to describe the data. So discovery mechanisms that work for data will also work for procedures to operate on the data. Trust mechanisms to permit an application to evaluate whether specific data or procedures are suitable for use in a given context are simply more data and relationships in the Semantic Web architecture; that is, they are an integral part of the Semantic Web vision.

The development of the Semantic Web is well underway in at least two very important areas: (1)

from the infrastructural and architectural position defined by W3C and (2) in a more directed application-specific fashion by those leveraging Semantic Web technologies in various demonstrations, applications and products. This article provides a brief introduction to both of these developmental areas with a specific focus on those in which the W3C is directly involved.

More information on the Semantic Web, including additional projects, products, efforts and future directions, is available on the Semantic Web home page (www.w3.org/2001/sw/).

Enabling Standards

Uniform Resource Identifiers (URIs) (www.w3.org/Addressing/) are a fundamental component of the current Web and are in turn a foundation for the Semantic Web. URIs provide the ability for uniquely identifying resources of all types – not just Web documents – as well as relationships among resources. An additional fundamental contribution toward the Semantic Web has been the development of the Extensible Markup Language (XML) (www.w3.org/XML/). XML provides an interoperable syntactic foundation upon which the languages to represent relationships and meaning are built. The Resource Description Framework (RDF) (www.w3.org/RDF/) family of languages leverages XML, URIs and the Web to provide a powerful means of expressing and representing these relationships and meaning.

The W3C Semantic Web Activity (www.w3.org/2001/sw/) plays a leadership role in both the design of specifications and the open, collaborative development of technologies focused on representing relationships and meaning and the automation, integration and reuse of data. The base level RDF 1.0 standard was defined in 1999. RDF 1.0 and RDF Schema (RDF Vocabularies) are currently being refined based on implementation experience, and more expressive higher layers are being addressed.

The base level standards for supporting the Semantic Web are currently being refined by the RDF Core (www.w3.org/2001/sw/RDFCore/) Working Group. This group is chartered to revise and formalize the original RDF Model and Syntax Recommendation (www.w3.org/TR/1999/REC-rdf-syntax-19990222/), which provides a simple, yet powerful, assertional framework for representing information in the Web. Additionally, this group is tasked to layer upon this general descriptive framework a simple means for defining RDF Vocabularies (www.w3.org/TR/rdf-schema/). RDF Vocabularies are descriptive terms such as *service, book, image, title, description* or *rights* that are useful to communities interested in recoding information in a way that enables effective reuse, integration and aggregation of data. Additional deliverables include a precise semantic theory (www.w3.org/TR/rdf-mt/) associated with these standards useful for supporting future work, as well as a primer (www.w3.org/TR/rdf-primer/) designed to provide the reader the basic fundamentals required to effectively use RDF in their particular applications.

The Web Ontology (www.w3.org/2001/sw/WebOnt/) Working Group standards efforts are designed to build upon the RDF core work a language, OWL (www.w3.org/TR/owl-ref/), for defining structured, Web-based ontologies. Ontologies can be used by automated tools to power advanced services such as more accurate Web search, intelligent software agents and knowledge management. Web portals, corporate website management, intelligent agents and ubiquitous computing are just some of the identified scenarios (www.w3.org/TR/webont-req/) that helped shape the requirements for this work.

Semantic Web Advanced Development (SWAD)

Code modules such as libwww (www.w3.org/Library/) accelerated the early deployment of the Web, and to a similar end the W3C is devoting resources to the creation and distribution of components to assist in the deployment of the Semantic Web.

These W3C Semantic Web Advanced Development initiatives are designed to work in collaboration with a large number of researchers and industrial partners to stimulate various complementary areas of development that will help facilitate further deployment and future standards work associated with the Semantic Web.

SWAD DAML. SWAD DAML is a project within the Defense Advanced Research Project Agency (DARPA) Agent Markup Language (DAML) (www.daml.org/) Program. The SWAD DAML (www.w3.org/2000/01/sw/daml) project combines research and development to define the architectural layering of the languages of the Semantic Web infrastructure. SWAD DAML builds critical components of that infrastructure and demonstrates how those components can be used by practical, user-oriented applications. It both seeks to define a logic language framework on top of RDF and the OWL vocabulary and to build basic tools for working with RDF, OWL and this logic framework.

To demonstrate some practical applications of these tools to manipulate structured information, SWAD DAML is deploying them to maintain the ongoing activities of the W3C, including access control, collaboration, document workflow tracking and meeting management. Another component of SWAD DAML is focused on the informal and often heuris-

tic processes involved in document management in a personalized information environment. Integrated into SWAD DAML will be tools to enable authors to control terms under which personal or sensitive information is used by others, a critical feature to encourage sharing of semantic content.

SWAD-Europe. SWAD-Europe (www.w3.org/2001/sw/Europe/) aims to highlight practical examples of where real value can be added to the Web through Semantic Web technologies. The focus of this Advanced Development initiative is on providing practical demonstrations of how (1) the Semantic Web can address problems in areas such as sitemaps, news channel syndication, thesauri, classification, topic maps, calendaring, scheduling, collaboration, annotations, quality ratings, shared bookmarks, Dublin Core (http://dublincore.org/) for simple resource discovery, Web service description and discovery, trust and rights management and (2) effectively and efficiently integrate them.

The focus of the SWAD-Europe deliverables are to exploit the enabling standards that have already been developed and not to depend upon future technologies identified with the Semantic Web architecture. Thus, the SWAD-Europe work is demonstrating the potential of what can be built on existing Semantic Web standards.

SWAD-Europe will additionally engage in exploratory implementation and pre-consensus design in such areas as querying and the integration of multiple Semantic Web technologies. This effort will provide input and experiences to future standards work.

SWAD Simile. Under the SWAD initiatives, W3C is also working with Hewlett-Packard (www.hp.com/), Massachusetts Institute of Technology (MIT) Libraries (http://libraries.mit.edu/), and MIT's Laboratory for Computer Science (MIT LCS) (www.lcs.mit.edu/) on Project Simile (http://web.mit.edu/simile/www/). Simile seeks to enhance interoperability among digital assets, schemas, metadata and services across distributed individual, community and institutional stores and across value chains to provide useful end-user services by drawing upon the assets, schemas and metadata held in such stores. Simile will leverage and extend DSpace (http://dspace.org/), also developed by MIT and HP, enhancing DSpace's support for arbitrary schemas and metadata, primarily through the application of RDF and Semantic Web techniques. The project also aims to implement a digital asset dissemination architecture based upon Web standards, enabling services to operate upon relevant assets, schemas and metadata within distributed stores.

The Simile effort will be grounded by focusing on well-defined, real-world cases in the libraries domain. Since parallel work is underway to deploy DSpace at a number of leading research libraries, we hope that such an approach will lead to a powerful deployment channel through which the utility and readiness of Semantic Web tools and techniques can be demonstrated compellingly in a visible and global community.

SWAD Oxygen. The Oxygen Project (http://oxygen.lcs.mit.edu/), a joint effort of the MIT LCS and the MIT Artificial Intelligence Laboratory (MIT AI), is designed to make pervasive, human-centered computing a reality through a combination of specific user and system technologies. Oxygen's user technologies directly address human interaction needs: automation (http://oxygen.lcs.mit.edu/Automation.html), individualized knowledge access (http://oxygen.lcs.mit.edu/KnowledgeAccess.html) and collaboration (http://oxygen.lcs.mit.edu/Collaboration.html) technologies help us perform what we want to do in the ways we like to do them. In Oxygen, these technologies enable the formation of spontaneous collaborative regions that provide support for recording, archiving and linking fragments of meeting records to issues, summaries, keywords and annotations.

A goal of the Semantic Web is to foster similar collaborative environments – human-to-human and human-to-machine – and the W3C is working with project Oxygen to help realize this goal. The ability for "anyone to say anything about anything" is an important characteristic of the current Web and is a fundamental principal of the Semantic Web. Knowing who is making these assertions is increasingly important in trusting these descriptions and enabling a "Web of Trust." The Annotea (www.w3.org/2001/Annotea/) advanced development project provides the basis for associating descriptive information, comments, notes, reviews, explanations or other types of external remarks with any resource. Together with XML digital signatures, the Annotea project will provide a test-bed for "Web-of-Trust" Semantic Web applications.

Education and Outreach

To fulfill its leadership role and facilitate the effectiveness and efficiency of the W3C Semantic Web Activity, a strong focus on education and outreach is important. The RDF Interest Group (www.w3.org/RDF/Interest/) continues to be an extremely effective forum in which developers and users coordinate public implementation, share deployment experiences of RDF and help each other promote the Semantic Web.

Arising out of RDF Interest Group discussions are several

The Semantic Web provides an infrastructure that enables not just Web pages, but databases, services, programs, sensors, personal devices and even household appliances to both consume and produce data on the Web.

public issue-specific mailing lists, including RDF-based calendar and group scheduling systems (http://lists.w3.org/Archives/Public/www-rdf-calendar/), logic-based languages (http://lists.w3.org/Archives/Public/www-rdf-logic/), queries and rules for RDF data (http://lists.w3.org/Archives/Public/www-rdf-rules/) and distributed annotation and collaboration (http://lists.w3.org/Archives/Public/www-annotation/) systems. Each of these discussion groups focuses on complementary areas of interest associated with the Semantic Web activity.

Future education and outreach plans include the formation of a Semantic Web education and outreach group designed to develop strategies and materials to increase awareness among the Web community of the need for and benefits of the Semantic Web and to educate the Web community regarding best practice solutions and enabling technologies associated with the Semantic Web.

Conclusion

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. It is based on the idea of having data on the Web defined and linked such that it can be used for more effective discovery, automation, integration and reuse across various applications.

The Semantic Web provides an infrastructure that enables not just Web pages, but databases, services, programs, sensors, personal devices and even household appliances to both consume and produce data on the Web. Software agents can use this information to search, filter and prepare information in new and exciting ways to assist Web users. New languages make significantly more of the information on the Web machine-readable to power this vision and will enable the development of a new generation of technologies and toolkits.

The seeds of the Semantic Web have been present within the Web from the time of Tim Berners-Lee's original Web proposal. For the Web to reach its full potential, it must grow and incorporate this Semantic Web vision, providing a universally accessible platform that allows data to be shared and processed by automated tools as well as by people. The W3C Semantic Web Activity is a multi-faceted program of basic research, collaborative technology development and open consensus-based standards setting to bring the Semantic Web to a reality and open the door to a whole new set of effective information integration, management and automation services.

For Further Reading

Resource Description Framework (RDF)
W3C Working Draft (work in progress) 11 November 2002, RDF Primer; www.w3.org/TR/rdf-primer/

Semantic Web Home Page
W3C, Semantic Web; www.w3.org/2001/sw/

URIs
W3C, Naming and Addressing: URIs, URLs, ...; www.w3.org/Addressing/

Web Ontology Language (OWL)
W3C Working Draft (work in progress) 12 November 2002, Web Ontology Language (OWL) Reference Version 1.0; www.w3.org/TR/owl-ref/