The Semantic Web: An Overview

By: Ms. Surabhi Jolly
Asst. Librarian
ICFAI University
Dehradun.

Ms. J. Sredevi
Asst. Librarian
Bundelkhand University
Jhansi.

Abstract:

The World Wide Web is a universal information space. As a medium for human exchange, it is becoming mature, but we are just beginning to build a space where automated agents can contribute –just beginning to build the Semantic Web. The RDF Schema design [RDF Schema] and XML Schema design [XML Schema] began independently but now a common model is needed where they fit together as interlocking pieces of the semantic web technology.

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(2001). 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. For the web to reach its full potential it must evolve into this semantic web, providing a universally accessible platform that allows data to be shared and processed by automated tools as well as by people.

In the semantic web data itself becomes part of the web and is able to be processed independently of application, platform or domain. This is in contrast to the world wide eb as we know it today, which contains virtually boundless information in the form of documents. We can use computers to search for these documents but they still have to be read and interpreted by humans before any usual information can be extrapolated. Computers can present you with information but can't understand what information is well enough to display the data that is most relevant in a given circumstance.

The semantic web on the other hand, is about having data as well as documents on the web so that machines can process, transform, assemble and even act on the data in useful ways. The World Wide Web is a universal information space. As a medium for human exchange, it is becoming mature, but we are just beginning to build a space where automated agents can contribute.

The Vision of the Semantic Web

Still in its definition stage, the term Semantic Web is perhaps new to many people, even to those within IT circle. But the problems it aims to address are ones we are been struggling to solve for decades – issues such as information overload, stovepipe systems and poor content aggregation (Daconta , Orbst, and Smith ,2003) . The fundamental roots of these problems are the lack of semantic integration among data sets and the lack of semantic definitions in individual systems, the lack of semantic integration among data sets and the lack of semantic interoperability across disparate system. The semantic web extends beyond the capabilities of the current web and existing information technologies, enabling more effective collaborations and smarter decision –making .It is an aggregation of intelligent websites and data stores accessible by an array of semantic technologies , conceptual frameworks , and well –understood contracts of interaction to allow machines to do more of the work to respond to service requests- whether that be taking on rote search processes , providing better information relevance and confidence , or performing intelligent reasoning and brokering.

What the Semantic Web Is and what Not

Semantic Web is not a new and distinct set of websites

The semantic web is an extension of the current World Wide Web, not a separate set of new and distinct websites. It builds on the current World Wide Web constructs and topology, but adds further capabilities by defining machine – processable data and relationship standards along with richer semantic associations. Existing sites may use these constructs to describe information within web pages in ways more readily accessible by outside processes such as search engines, spider search technology, and parsing scripts. The protocols underlying the semantic web are meant to be transparent to existing technologies that support the current World Wide Web.

The Semantic web is not being constructed with just human accessibility in mind

The current web relies mainly on text markup and data link protocols for structuring and interconnecting information at a very coarse level. These protocols are used primarily to describe and link documents in the forms presentable for human consumption (but that has useful hooks for first order machine searching and aggregation). Semantic Web protocols define and connect information at a much more refined level. Meanings are expressed in formats understood and processed more easily by machines in ways that can bridge structural and semantic differences within data stores. This abstraction and increased accessibility means that current web capabilities can be augmented and extended – and new, powerful ones introduced.

The semantic web is not some magical piece of artificial intelligence

The concept of machine –understandable documents does not imply some form of magical artificial intelligence that allows machines to comprehend human beings. It only indicates a machine's ability to solve a well defined problem by performing well defined operations on existing well –defined data .Current search engines perform capabilities that would have been magical 20 years ago, but that we recognize now as being the result of IP protocols, HTML, the concept of websites, web pages, links, graphical browsers, innovative search and ranking algorithms, and, of course, a large number of incredibly fast servers and fast disk storage arrays. Semantic web capabilities will likewise be the result of a logical series of interconnected progressions in Information technology and knowledge representation formed around a common base of standards and approaches.

Potential Benefits of Semantic Web

Most people are capable of using the web to, say, find the Swedish word for "club", renew a library book, or find the cheapest "car" and buy it. But if you ask a computer to do the same thing, it wouldn't know where to start. That is because web pages are designed to be read by humans, not machines. The semantic web is a project aimed to make web pages understandable by computers, so that they can search websites and perform actions in a standardized way.

The potential benefits are that computers can harness the enormous network of information and services on the web. Your PC could, for example, automatically find the nearest dentist to where you live and book an appointment for you that fits in with your diary.

Semantic interoperability, represents a more limited or constrained subset of the vision of the semantic web. Significant returns , however , can still be gained by using semantic –based tools to arbitrate and mediate the structures , meanings and contexts within relatively confined and well understood domains for specific goals related to information sharing and information interoperability. In other words semantic interoperability addresses a more discrete problem set with more clearly defined endpoints.(Pollock and Hodgson,2004) Semantic technologies can also provide a loosely connected overlay on top of existing Web services and XML frameworks , which in turn can offer greater adaptive capabilities than those currently available. They can also make immediate inroads in helping with service discovery and reconciliation, as well as negotiations of requests and responses across different vocabularies.

Semantic computing is an emerging discipline being formed and shaped as this is written .As such, there are many definitions and interpretations, and even a few low intensity philosophical wars being waged among thought – leaders and practitioners. That said , the release of RDF and OWL as W3C Recommendations earlier in the year has created a greater commonality in expression.

Semantic technologies differ from database schemes, data dictionaries, and controlled vocabularies in an important way. They have been designed with the connectivity in mind allowing different conceptual domains to work together as network.

Components of the Semantic Web

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

- XML provides a surface syntax for structured documents, but impose no semantic constraints on the meaning of these documents.
- XML Schema is a language for restricting the structure of XML documents.
- RDF is a simple data model for referring to objects ("resources") and how they are related. An RDF – based model can be represented in XML Syntax.
- RDF Schema is a vocabulary for describing properties and classes of RDF resources, with a semantics for generalization –hierarchies of such properties and classes.
- OWL adds more vocabulary for describing properties and classes: among others, relations between classes, equality, richer typing of properties and characteristics of properties (e.g. Symmetry), and enumerated classes.

Metadata Publishing and Management Tools

The process of creating a metadata about a document or data item can occur when that item is creating or authored , when it is imported into a content management system or a website , or when it is used or viewed by the users . It can also be added by some other explicit or implicit action at any point over the course of existence of that data item. In other words metadata creation is just not one time occurrence. Metadata can continue to accumulate and can be modified at any time by conceivably any number of people.

At content creation, authors typically connect information such as the subject, creator, location, language, and copyright status with a particular document. This information makes the document much more searchable. RSS consists essentially of this type of information, providing news reading applications with significantly expanded capabilities for searching and filtering information. The creation of metadata is only one step in the process. Metadata management tools are needed in order to maintain metadata vocabularies, perform metadata-driven queries, and provide visualization tools for monitoring changes in areas of interest. An example of a website that uses Meta data as a key aspect of creating a collaborative and shared system of data is Flickr, a site for people to easily upload and share digital photos.

Modeling Tools (Ontology creation and modification)

Modeling tools are used to create and modify ontologies. Knowledge modelers used them to create and edit class structures and model domains. The tools often have an interface that is similar to a file system directory structure or bookmark folder interface. They also tend to offer the ability to import, transform, and re-purpose, in whole or in part, existing ontological structures that are often in the form of database schemas, product catalogues, and yellow pages listings. Other prominent feature includes advanced mechanisms for organizing, matching, and associating similar terms and concepts.

Also, because it is a common practice for modelers to create smaller interconnected ontologies instead of a single large monolithic model- primarily for better reusability and ease of use –support for splitting , merging and connecting models can be an important capability in the ontology editor. Some editors even support collaborative work methods and rich visualization and graphical interaction modes.

Ontologies

Arriving at the right ontology is often a critical element of successful implementation of semantics-based projects. Even more so than database design, ontology creation is a highly specialized field. Not only are there not as yet a sizeable number of skilled practioners, it can take considerable time to arrive at an ontology that successfully captures a conceptual domain. As a result, it is important to look at existing bodies of work that can be used (and reused) in lieu of having to create something from scratch. Likely sources of existing ontologies can typically be located in close association with ontology modeling tools. Use of proprietary ontologies may be contingent upon licensing of the modeling tools, a practice which is not unreasonable considering the efforts expended to develop the ontologies. Other ontologies, however, may be open and free for use of commercial and non commercial purposes, much in the vein of Linux, JBoss, Musicbrainz, and other open sources and data repositories.

Current ontology development efforts vary in scope and size. Some ontologies have been developed specifically in answer to localized implementations such as reconciling charts of accounts or health care records , areas where the emphasis is primarily on information interoperability.- arbitrating between syntaxes ,structures and semantics –and less on logic programming . Other ontology development efforts take a more top down approach under the assumption that a shared view of a wide knowledge domain is critical to wide spread proliferation of adaptive computing and intelligent reasoning capabilities. There is significant advocacy in these latter circles on the establishment of an enterprise wide common upper ontology under the belief that it will provide the foundation for any number of domain ontologies. New domain ontologies could be extension of , and fully compliant with , this upper ontology. Existing ontologies and legacy data models could be mapped to this upper ontology, which theoretically would constitute a significant number of the steps towards achieving greater semantic interoperability across domains.

Mapping Tools (Ontology Population)

Once an ontology model is created, it needs to be populated with data. This process is usually accomplished by linking various data sources to the concepts in an ontology using a mapping tool. Once “maps” have been created, a query in one data source could be transformed by its map to the ontology and then from ontology to the data sources using their maps. Th corresponding data could then be returned in the same manner without any of the data stores knowing or carrying about the others. In other words, each data source may have a unique “map” to an overarching ontology that acts as a pivot table among the various sources and targets. Providing this abstraction layer requires some effort on the part of creating the ontology and then creating the data maps, but once this has been done each data source can interoperate with other data sources strictly within run-time processes. Bringing new data sources onboard will, in most cases, have little or no effect on existing data sources.

This process drastically reduces the amount of data value mapping and semantic conflict resolution that typically takes place using current enterprise application approaches – approaches that that up to now typically require n-squared mappings or alternatively, exporting to hard coded, inflexible and explicit standards. The modeling and mapping makes the process far less political and far more flexible and adaptable. Network interface and unicorn are two vendors with tools of this type. Tools that aggregate normalize and map unstructured data forms including Word, RTF, Text files, and HTML. Semagix is a leading vendor for unstructured data.

Mediation Engines

Mediation engines are automated tools that can dynamically transform data among different syntaxes, structures and semantics using models instead of hard –wired transformation code. They are critical components of any interoperability architecture. Using data maps, ontologies, and other forms of conceptual model, mediation engines are run –time processes that provide an abstraction layer between heterogeneous data sets , allowing organizations to essentially agree to disagree about how data and information should be represented. Mediation engines typically work with highly structured data .Unstructured and semi –structured data must first be bound to a schema prior to creating the mediation maps.

Interface Engines

Interface engines are software tools that derive new facts or association from existing information. It is often said that an interface engine emulates the human capability to arrive at a conclusion by reasoning. In reality, inferencing is not some mythical artificial intelligence capability but, rather, a quite common approach in data processing. One can think of a complex data mining exercise as a form of inferencing. By creating a model of the information and relationships, we enable the reasoner to draw logical conclusions based on the model. A common example of an inference is to use models of people and their connections to other people to gain new knowledge. Exploration of these network graphs can enable inferences about relationships that may not have been explicitly defined.

With the addition of rule and logic languages, however, greater leaps in conceptual understandings, learning and adaptation can take place, although implementations with these types of capabilities are, as yet, few and far between. Both free and commercial versions of inference engines are available. For example, Jena, an open source Java framework, for writing Semantic web applications developed by HP labs, has a reasoner subsystem. Jena reasoner includes a generic rule based inference engine together with configured rule sets for RDFS and for the OWL –Lite subset of OWL Full.

Other Components

Ordinary web pages are a good source of instance information; many tools for populating ontologies are based on annotation of web pages. W3C Annotea project offers free annotation tools. Commercial vendors include Ontoprise and Lockheed – Martin. Several software vendors, including Semagix, Siderian Software and Entopia offer products that use ontologies to categorize information and to provide improved search and navigation.

Semantic Web Services

A web service is a software system designed to support interoperable machine to machine interaction over a network. A web service has an interface described in a machine processable format using Web Services Description Language (WSDL). The combination of WSDL, UDDI, and SOAP form a triad of technologies that will shift the entire market towards service oriented architectures (SOA). Together these technologies provide directory, component look up, and exchange protocol services on top of HTTP or SMTP network protocol. Microsoft, IBM, and most other large software vendors have embraced the concepts and languages that underline the WEB services model and an increasing number of books and industry articles point to the benefits of adopting service oriented architecture. Web services, however, are not without shortcomings. Security issues have long been a concern but several solutions that address these issues have been introduced over the last several years. Perhaps the most significant improvement opportunities for Web services that remain are in the areas of

(a) Flexible look up and discovery
(b) Information management and schema transformation.

Fundamentally, Web service technologies handle messages in a loosely coupled manner but they do not currently bridge differences in description terminologies nor they inherently enable the recipient to understand a message that has been sent. With Web services, these part of exchange rely on custom coded solutions and/or widespread community agreement upon some kind of document exchange standard.

Semantic Interoperability

Formally put, the use of semantic technologies makes it possible to describe the logical nature and context of information being exchanged, while allowing for maximum independence among communicating parties. The results are greater transparency and more dynamic communication among information domains irrespective of business logic, processes and work flows (Pollock and Hodgson, 2004)

The technical vision is one where flexible information models, not inflexible programs or code , are used to drive dynamic , self healing ,and emergent infrastructures for the sharing of mission critical data in massively scaleable environments. Recent advances in taxonomy and thesaurus technology, context modeling approaches, inferencing technology, and ontology driven interoperability can be applied in a cohesive framework that dramatically changes the way information is managed in disperse, decentralized communities of knowledge (Pollock and Hodgson, 2004)

One of the key advantages of using semantic interoperability approaches is that they do not necessarily require the replacement of existing integration technologies, databases, or software applications. A semantic framework made up of various semantics based components and application program interfaces (APIs) can be deployed with web services or traditional middle ware APIs to leverage existing infrastructure investments, and yet still to provide massive benefits by virtually centralizing the query, transformation and business rules metadata that flows through the network infrastructure's pipes. As such the software will fit into the customer's existing IT ecosystem with low overhead for installation, minimal coding, and maximum reusability.

Intelligent Search

Related in some regards to semantic interoperability is the area of intelligent search. As mentioned above, semantic interoperability techniques can allow queries native to one system to be federated to other non –native systems. This eliminates the need to convert system to a universal query language and enables system to continue maintaining the information they have in there current formats. By overlaying a virtual on top of the data sources, queries can be defined in a universal manner, thereby enabling access to all mapped assets. Federated searching can also be made smarter by making searches more semantically precise. In other words, searches can be broadened to include concepts, or narrowed to include only specific key words. The depth and granularity – of such searches enables the specification of the search that the individual desires.

Another aspect of intelligent search is the ability to make searches more relevant to the person searching by making use of identity and relationship information. Relationship among people and information about them can be key links to greater relevance and confidence. Despite investments in knowledge management systems, many people still rely their personal network of friends, neighbour, co-workers, and others to locate experts of find trusted information. Personal relationships are also useful in sales situations and in many organizational interactions. Social networking schemas and software are making broad use of this.

Conclusion

Semantic Web technologies will enhance the utility of web services when they are widely deployed. On the web, the many service providers will need to be able to advertise their services to an extremely wide and varied audience of service users. Providing brokering capabilities, the ability of service users to be automatically matched with service users, is difficult, and revolves around the same sort of vocabulary mapping issues that database exhibit. In the current implementations, the services describe inputs, outputs, ports, and other aspects of calling each other.

Moreover, the information that links one service description to another doesn't need to rely on the luck of having a common term in natural language to merge on. An outside source, whether it be a different user or developer, a separate thesaurus, or even a random fact found on someone else's web page, can express mapping information. Thus one could come along and say my author calls a "lorry" is equivalent to what you call a "truck", and from then on merge the graphs, a connection between lorry and truck can be found.

In short, the semantic web already provides the kinds of languages and tools needed to attack the problem. We're not that far from the time when you can click on the web page for the meeting, and your computer, knowing that is indeed a form of appointment, will pick up all the right information, and understand it enough to send it to all the right applications. Further, it will evoke those applications directly (using web services) needing little or no human intervention. The business market for this integration of data and programs is huge, and we believe the companies who choose to start exploiting Semantic Web technologies will be the first to reap the rewards.

References:

1. Michael C. Daconta , Leo J . Obrst, Kevin T.Smith(2003-05-30). *The semantic web: A guide to the future of XML, web services, and Knowledge management,* John Wiley & sons.
2. Dieter Fensel , Wolfgang Wahlster, Henry Lieberman , James handler. *Spinning the semantic web: Bringing the world wide web to its full potential*
3. John Davies , Dieter Fensel , Frank Van Harmelen : Towards semantic Web : Ontology –Driven Knowledge Management , John Miley and Sons .
4. Grigoris Antoniou , Frank van Harmelen :A Semantic Web Primer . The MIT Press
5. Jefeery T. Pollock , Ralph Hodgson. Adaptive Information : Improving business through Semantic Interoperability, Grid Computing ,and Enterprise Integration , John Wiley.
6. Steffen Staab, Rudi Studer . Handbook on Ontologies , Heidelberg: Springer Verlag
7. Vladimir Geroimenko, Chaomei Chen . Visualising the Semantic Web , Springer Verlag.