THE SEMANTIC WEB: A SUCCESSOR OF THE WORLD WIDE WEB?

MV Slabbert (MACI Consultants)

The World Wide Web is based primarily on documents written in HTML. This is a language that is useful for describing and visual presentation, but has limited ability to classify the blocks of text on a page. The Semantic Web is namely a project that intends to create a universal medium for information exchange by giving meaning, in a manner understandable by machines, to the content of documents on the Web. Thus, the Semantic Web extends the ability of the World Wide Web through the use of standards, markup languages and related processing tools, aiming at making information in Web-based documents more accessible. This article explores the Semantic Web as a natural successor of the World Wide Web.

Key phrases: Semantic Web, World Wide Web Consortium

INTRODUCTION

The Semantic Web (SW) is in many senses a more complex structure than the World Wide Web (WWW) itself, due to various reasons. It is currently being constructed and/or synthesised from a wide variety of sources and purposely being structured to achieve a range of sophisticated processes to approximate the functions of a human agent in optimising choices from a given set of possibilities. Numerous difficulties exist in achieving a workable structure that would justify the effort and expenditure required to develop it to the stage where it would become self-sustaining, and (ideally) commercially profitable.

The Semantic Web is primarily a vision of software engineer Tim Berners-Lee, following the development and highly successful outcome of the World Wide Web (WWW) from his early days at CERN. Having designed and implemented the various software 'bridges' (linkages) that were required to launch the connections between the independent PCs, databases, and various other electronic devices into 'live' interactive networks, Berners-Lee eventually evolved the concept of a more sophisticated and powerful mechanism using the WWW as the underlying structure.

Berners-Lee (2002) conceives of the Semantic Web as being more user-friendly and 'intelligent' compared to the WWW, which is a mechanistic operation dealing almost exclusively with documents. He has been working extensively for at least a decade on the development of the Semantic Web from the software side to give machines the power to make intelligent assessment of data stored in cyberspace, and not merely retrieve data. The "road map" is a document outlining a plan for achieving a set of connected applications for data on the Web in such a way as to form a consistent logical web of data, termed the Semantic Web (Berners-Lee 1998).

The functioning of the basic existing Web (WWW) is very much a tracing of linkages between data stored in endless different forms, and the intention of the user in discovering or tracing specific information. Culture, language, and personal experience amongst other aspects all contribute to the user's interpretation of the

nature of the relationship between the symbols used (e.g. words) and the ideas attributed to them. Consequently, the 'intelligent assessment' part of analysing the information found this way is fraught with ambiguities, requiring the attention of a human evaluator.

Ultimately this extraordinary variety of possible interpretations, while giving richness to a language in its numerous permutations, is contributing substantially to the problems in structuring the Semantic Web.

DEFINITION OF THE SEMANTIC WEB

Developing a definition for the Semantic Web is clearly a similarly problematical task, since it is so dependent on the user's viewpoint.

The choice of the term *semantic* in describing this more advanced form of the Web is no doubt related to the aspect of linguistics which deals with the meanings given to words (or concepts). It embraces the changes that can occur with these meanings given a contextual framework to provide numerous additional connotations; and the residual influence over time that changes such meanings for the community communicating with the language.

The Semantic Web is fundamentally a web of encoded data, in some ways like a global database. The Semantic Web development adds to the Web formats for representing data and its semantics - the meaning in terms of what rules can be applied and how it can be transformed into other data. This will lead to much greater clarity in complex communications, for example when an invoice is sent with some accompanying simple mathematics which describes its role in commerce transaction. It will lead to much greater re-use of the data, and much easier analysis of what is going on.

Berners-Lee (1998, 2001, 2002), in various papers, has outlined and discussed the evolution of the WWW and then the Semantic Web as a natural extension of this – but having considerably more structure, form and logical complexity. In contrast to the Web which operates as a vast storehouse of documents for users to trawl, the Semantic Web is being purposely developed as a medium for data and information that can be processed automatically where information has such well-defined meanings that interpretation by machine becomes possible '… better enabling computers and people to work in cooperation' (Berners-Lee *et al.* 2001).

Since the Semantic Web is still very much in the process of development, a precise and concise definition is difficult to formulate, but consists rather of a range of characteristics, what it may/may not accomplish (as far as current vision extends), the many difficulties and obstacles that exist, and a variety of goals and possibilities. The core issue is that in general, computers have no reliable way to process the semantics of all the data in WWW cyberspace. The Semantic Web is creating a paradigm shift in dealing with the many different structured databases and their

interaction. Such 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" (Berners-Lee *et al.* 2001). Hence 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" (Berners-Lee *et al.* 2001)

Giving some explicit meanings to the concept of the Semantic Web, Marshall & Shipman (2003) have analysed the work of Berners-Lee (and others), providing an extensive analysis of the Semantic Web from several perspectives.

Depending on the needs of the user, the Semantic Web is portrayed as:
- a universal library, to be readily accessed and used by humans in a variety of information use contexts
- the backdrop for the work of computational agents completing sophisticated activities on behalf of their human counterparts
- a method for federating particular knowledge bases and databases to perform anticipated tasks for humans and their agents (Marshall & Shipman 2003).

The underlying issue, however, still remains – will the Semantic Web be sufficiently attractive to users on a large enough scale to expand and grow in the same way as the WWW ?

FEASIBILITY OF THE SEMANTIC WEB

The short answer to this feasibility query is 'yes' if only because various groups of highly educated people have already spent many man-years of work, and plenty of effort in realising the potential of this vision of Berners-Lee, and developing the specific mechanisms of the Semantic Web model for well over a decade already.

Since there are specific mathematical tools (algorithms) available to define data, theoretically such a system can indeed operate. Further, there is already a form of the Semantic Web functioning as work is increasingly being formatted in the appropriate ways. In addition, many languages, publications and tools have already been developed and disseminated.

Therefore, had it not been feasible, the whole World Wide Web Consortium (W3C), the major developmental consortium overseeing the Semantic Web creation, would have collapsed long ago. Clearly there are various problems that still need to be solved, according to the available literature, but they are not the kind that would prevent the Semantic Web from functioning at all, but rather affect its scope, depth and power.

From the perspective of artificial intelligence (AI), the mathematics of Boolean algebra has reached an insurmountable barrier, according to leading researcher Jaques Hale of the Butler Group, UK (Flood 2001). Hale foresees a positive outcome

from this alternative route using the 'grid computing' resources of the Internet and the clearness of Berners-Lee's vision in creating an intelligent structure out of the Web. According to Berners-Lee, the development of the Semantic Web has already crossed the barrier from the realm of the technologists to more generalised areas with wider application.

However, other viewpoints are not as sanguine - having far less certainty of a positive future since Semantic Web technologies are still so under-developed. According to Palmer (2004) 'there seems to be little consensus about the likely direction and characteristics of the early Semantic Web'. Interestingly enough, according to Flood (2001) both Microsoft with a product called ".NET" and Sun Microsystems with "ONE framework" have been attempting to pre-empt the Semantic Web from a commercial perspective, albeit on a rather smaller scale. However, the aspects of 'ontologies' (collections of information), where an ontology is a document or file that formally defines the relations among terms (traditionally an ontology is a theory about the nature of existence, of what types of things exist) and 'trust' seem to be the main areas likely to create the most obstacles in large-scale Semantic Web functioning.

OPERATION OF THE SEMANTIC WEB

A considerable part of constructing the Semantic Web model arises from the functioning of computers, and the constantly evolving nature of this technology, both hardware and software. Berners-Lee *et al.* (2001) refer to the emergence of this Semantic Web as bringing "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". So the general approach has been to re-define raw data wherever it is stored, so that it becomes associated with a meaning in some way that can be readily translated into other contexts. 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 form that a machine is capable of processing.

The Semantic Web approach means putting in hidden code indicating how each bit of data is placed and related to other data from the syntax perspective while also indicating the sense of each word being used. In general, the Semantic Web requires a whole new inbuilt infrastructure to enable the model of 'machine intelligence' to become a reality – a formidable task despite the apparent ease of operation that is so often depicted in science fiction scenarios.

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. Berners-Lee expects that some (software) agents will exploit available artificial-intelligence technologies to create more complicated value chains automatically on demand, adding to the yield from the networks of the Semantic Web. But the Semantic Web will provide the

overall foundations and the framework to make such technologies more feasible and rewarding.

As at 2001, the task before the Semantic Web community was finding a way to add logic to the Web - this means to use rules to make inferences, choose courses of action and answer questions, approximating artificial intelligence. Berners-Lee *et al.* (2001) indicated that a mixture of mathematical and engineering decisions complicates 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. Two important technologies for developing the Semantic Web are already in place: eXtensible Markup Language (XML) and the Resource Description Framework (RDF). The goal of this approach in constructing the Semantic Web is to enable machines to comprehend documents and data semantically, and not as human speech and writings which is entirely another mechanism (Dumbill 2003).

A basic element incorporated in the model is the use of a web identifier mechanism (Universal Resource Identifier, known as the acronym URI) in the syntax. The Semantic Web, in naming every concept simply by a URI, lets anyone express new concepts that they invent with minimal effort. Subject and object are each identified by a URI, just as used in a link on a Web page. (URLs, the acronym for Uniform Resource Locators and referring to a web address, are the most common type of URI.) The verbs are also identified by URIs, which enables anyone to define a new concept, a new verb, just by defining a URI for it somewhere on the Web. Its unifying logical language will enable these concepts to be progressively linked into a universal Web which 'will open up the knowledge and workings of humankind to meaningful analysis by software agents, providing a new class of tools' (Berners-Lee *et al.* 2001).

In developing a solution to this problem, a third basic component of the Semantic Web was introduced, namely ontologies, with a somewhat extended meaning for Semantic Web applications. The most typical kind of ontology for the Web has taxonomy and a set of inference rules.

Berners-Lee (2002) perceives that ultimately, when people create many programs that collect Web content from diverse sources, process the information and exchange the results with other programs, the real power of the Semantic Web will be realised through the synergy of numerous software agents interacting automatically with each other: approximating a kind of initiative ability by software agents. 'Even agents that were not expressly designed to work together can transfer data among themselves when the data come with semantics', according to Berners-Lee's vision.

From the website of the W3C states that "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 principle of the Semantic Web".

THE SEMANTIC WEB AND THE TYPICAL TRADITIONAL INFORMATION RETRIEVAL PROBLEMS

Since the infrastructure of the Semantic Web is required to be considerably more sophisticated as well as being a radically different approach in some ways from the traditional Web, the answer whether the Semantic Web would be able to solve the traditional information retrieval problems, is a qualified 'maybe'. There are a range of similarities; ideally the most effective and productive aspects of the Web are being transferred into the design of the Semantic Web. At the same time, a whole selection of new parameters is coming into place to accomplish their necessary purposes – but which could well create their own drawbacks in practice.

Another issue is to introduce a form of verification into the Semantic Web system, so that sustainable levels of credibility can allow mechanisms of 'reasoning' to happen in the process of determining the correct/appropriate/optimal outcome of a search or instruction. Inferences drawn from a large variety of sources require a way to check their varying degrees of truth, probability, personal opinion, or factual correctness. This is being done through establishing 'trust relationships' (Dumbill 2003).

Like the Internet, the Semantic Web needs to be as decentralised as possible where computers must have free access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning. Semantic Web researchers, in contrast to artificial-intelligence researchers, accept that paradoxes and unanswerable questions are a price that must be paid to achieve versatility and flexibility.

CLOSURE

A considerable body of work has established a form of Semantic Web, with many hopes, visions and intentions woven into it. Given the tremendous success of the WWW which evolved without any control or purposeful guidance (and possibly because of that freedom), there are various pros and cons surrounding the future of its possible successor, the Semantic Web, which makes the future path difficult to predict.

According to the literature, the year 2004 is a pivotal stage when the consortium and groups involved in establishing the groundwork, are supposed to have completed their work.

Should there be any possible commercial gains to be made, then the future of the Semantic Web is probably assured. The extent of this will no doubt depend on just how easy it is to operate in practice. Again, it might take some time before there is a sufficiently large enough body of re-formatted data, for this to develop.

BIBLIOGRAPHY

BERNERS-LEE T. 1998. Semantic Web road map. [Internet: http://www.w3.org/ DesignIssues/Semantic.html.]

BERNERS-LEE T, HENDLER J & LASSILA O. 2001. The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. [Internet: http://www.scientificamerican.com/article.cfm?articleID= 00048144-10D2-1C70-84A9809EC588EF21&catID=2.]

BERNERS-LEE T. 2002. Commemorative lecture: the World Wide Web - past, present and future. [Internet: http://www.w3.org/2002/04/Japan/Lecture.html.]

DUMBILL E. 2003. The Semantic Web: a primer. [Internet: www.xml.com/pub/a/semanticweb/.]

FLOOD G. 2001. From Web to grid. *Information World Review,* 175:28. [Also available at Internet: http://orion.learned.co.uk/iwr/archive/175feature_001. asp.]

MARSHALL CC & SHIPMAN FM. 2003. Which Semantic Web? [Internet: http://www.csdl.tamu.edu/ ~marshall/ht03-sw-4.pdf.]

PALMER SB. 2004. The Semantic Web: an introduction. [Internet: http:// infomesh.net/swintro/.]

WORLD WIDE WEB CONSORTIUM. 2004. Various documents. [Internet: http://www.w3.org]