Part 1--The Semantic Web

The World Wide Web's simplicity was a key factor in its rapid adoption. But as it grows ever larger and more complex, that simplicity has begun to hinder our ability to make intelligent use of the vast store of data on the Web. In response to that challenge, the World Wide Web Consortium (W3C) has spearheaded an effort to create an extension of the Web that brings meaning and order to web data. It's called the Semantic Web, and at its core is the Resource Description Framework (RDF), an application of XML. However, whilst the creation of the Semantic Web is dependant on RDF, it is conceivable that one day the tail will wag the dog, and the Semantic Web become just one application of RDF. In current literature, papers on RDF are replete with descriptions of the Semantic Web, and papers on the Semantic Web include discussions on RDF as if their functions were conditionally interdependant. Also there are voices that question the desirability of the Semantic Web, some claim that RDF is a language giving a universal application advantage, and others that it is best used selectively.

What Is The Semantic Web?

The Semantic Web, conceived as a worldwide network of information linked in such a manner as to be easily processable by machines operating as a globally linked database, was thought up by Tim Berners-Lee, inventor of the WWW, URIs, HTTP, and HTML. Currently a dedicated team of people at the World Wide Web consortium (W3C) are working to improve, extend and standardize the system, and many languages, publications, tools and so on have already been developed. However, Semantic Web technologies are still very much in their infancies, and although the future of the project in general appears to be bright, there seems to be little consensus about the likely direction and characteristics of the early Semantic Web.

What's the rationale for such a system?

Data generally hidden away in HTML files is often useful in some contexts, but not in others. The majority of data on the Web, is in this form at the moment, but difficult to use on a large scale, because there is no global system for publishing data in such a way that it can be easily processed by everyone. Information about local sports events, weather, travel, and television, is presented by numerous sites, but all in HTML. The problem is that in some contexts, it is difficult to use this data in a manner needed by particular applications. It seems at first glance that developing a consistent universal data format would not be too great a task. However, it has proved in the past that it is almost impossible to get two companies to agree on a specific definition of "data". There has never been complete agreement as to data exchange formats. The arrival of XML offers hope of a wide acceptance of its syntax or syntactic rules which no one faction can claim.

The Semantic Web is generally built on syntaxes which use URIs to represent data, usually in triples based structures: i.e. many triples of URI data that can be held in databases, or interchanged on the world Wide Web using a set of particular syntaxes developed especially for the task. These syntaxes are called "Resource Description Framework" syntaxes. (See Page 4)

URI--Uniform Resource Identifier

A URI is simply a Web identifier: like the strings starting with "http:" or "ftp:" found on the World Wide Web. Anyone can create a URI, and the ownership of them is clearly delegated, so they form an ideal base technology with which to build a global Web. The World Wide Web is such a thing: anything that has a URI is considered to be "on the Web".

The syntax of URIs is carefully governed by the IETF, who published RFC 2396 as the general URI specification. The W3C maintains a list of URI schemes

Creating the Semantic Web

How, however, do we create a web of data that machines can process? The first step is a paradigm shift in the way we think about data. Historically, data has been locked away in proprietary applications. Data was seen as secondary to processing the data. This incorrect attitude gave rise to the expression "garbage in, garbage out," or GIGO. GIGO basically reveals the flaw in the original argument by establishing the dependency between processing and data. In other words, useful software is wholly dependent on good data. Computing professionals began to realise that data was important, and it must be verified and protected. Programming languages began to acquire object-oriented facilities that internally made data first-class citizens. However, this "data as king" approach was kept internal to applications so that vendors could keep data proprietary to their applications for competitive reasons. With the Web, Extensible Markup Language (XML), and now the emerging Semantic Web, the shift of power is moving from applications to data. This also gives us the key to understanding the Semantic Web. The path to machine-processable data is to make the data smarter.

The Semantic Web is not specifically for the World Wide
The Role of XML

How does XML fit into the Web? XML is the syntactic foundation layer of the Semantic Web. All other technologies providing features for the Semantic Web will be built on top of XML. Requiring other Semantic Web technologies (like the Resource Description Framework) to be layered on top of XML guarantees a base level of interoperability. The technologies that XML is built upon are Unicode characters and Uniform Resource Identifiers (URIs). The Unicode characters allow XML to be authored using international characters. URIs are used as unique identifiers for concepts in the Semantic Web. Is XML enough? The answer is no, because XML only provides syntactic interoperability. In other words, sharing an XML document adds meaning to the content; but, only when both parties know and understand the element names, in particular in cases when different words have equivalent meanings.

Are The Tools Available?

While implementing the Semantic Web on the Internet is still a vision, the building blocks for the Semantic Web are being deployed in small domains and prototypes. Gradually the pieces are falling into place to make the promise a reality. Over the past five years, we have seen a paradigm shift away from proprietary "stovepiped systems" and toward open standards. The W3C, the Internet Engineering Task Force (IETF), and Organization for the Advancement of Structured Information Standards (OASIS) have had widespread support from corporations and academic institutions alike for interoperability. The support of XML has spawned support for XML-based technologies, such as SOAP-based Web services that provide interoperable interfaces into applications over the Internet. RDF provides a way to associate information. Using XML as a serialization syntax, RDF is the foundation of other ontology-based languages of the Semantic Web. XML Topic Maps (XTM) provide another mechanism for presenting taxonomies of information to classify data. Web services provide a mechanism for software programs to communicate with each other. Ontology languages (OWL, DAML+OIL) are ready for prime time, and many organizations are using these to add semantics to their corporate knowledge bases. This list could go on and on. Currently, there is an explosion of technologies that will help to fulfill the vision of the Semantic Web. Helping the Semantic Web’s promise is industry’s current focus on Web services. Organizations are beginning to discover the positive ROI of Web services on interoperability for Enterprise Application Integration (EAI). The next big trend in Web services will be semantic-enabled Web services, where we can use information from Web services from different organizations to perform correlation, aggregation, and orchestration.

Is Everyone Agreed?

Not everyone, those with not such long memories of the claims for Self Organising Systems, Artificial Intelligence etc of the sixties stand alongside the refugees from the dot-com era in regarding the claims for a super web to be proven. Open Source systems including Linux are now only now looking convincing despite the support of some illustrious companies. When new developments are simple to install and operate they are rapidly adopted, as was Windows. Establishing the Semantic Web in a a relatively short time means re-education and re-writing on a global scale with the consequent costs. At the moment there is a school of though that claims that the Semantic Web will enable users to extract from the “Global Information Database” the most relevant information to satisfy their needs.

In many cases where the objectives are not narrowly defined, the answers may simply lead to more questions. Fundamentally the famous ‘Gain/Bandwidth Rule says that as the field of knowledge widens the quality falls. Again as the store of information increases so does the effort to maintain its fidelity. Finally, information has a nasty habit of becoming irrelevant as time moves on. This is nowhere more obvious than on the current Web where it would not be out of place to accompany each paper with a 'Sell by Date' setting a time scale on its relevance.

Is improved technology the answer?

The answer of the protagonists is yes. They say that computing power has brought us thus far, and will continue to enable us to progress.

They say: "When you connect cell phones to PDA’s to personal computers, you have more brute-force computing power by several orders of magnitude than ever before in history. More power makes more layers possible. For example the virtual machines of Java and C# were conceived more than twenty years ago; however, they were not practical until the computing power of the 1990s was available."

Sadly, history has shown that progress is made by asking the right questions and correctly interpreting the answers.
This is why the emergence of the Resource Description Framework as an intellectual aid may prove to be more important than millions of more gigabytes, since it offers the chance of formulating the right questions.