Semantic Web

Rui G. Pereira
Universidade de Beira Interior, Portugal

Mário M. Freire
Universidade de Beira Interior, Portugal

THE WEB

The World Wide Web (WWW, Web, or W3) is known as the largest accessible repository of human knowledge. It contains around 3 billion documents, which may be accessed by more than 500 million worldwide users. In only 13 years since its appearance in 1991, the Web suffered such a huge growth that it is safe to say there is no phenomenon in history that can compare to it. It reached such importance that it became an indispensable partner in the lives of people (Daconta, Obrst & Smith, 2003).

Researching information on the current Web is supported through the use of robust and practical applications known as search engines and directories. But the fast and unorganized growth of the Web is making it difficult to locate, access, present, and maintain online trustful content for an increasing number of users. Difficulties in the search of Web contents are associated with the use of non-structured, sometimes heterogeneous information, and with the ambiguity of Web content. Thus, one of the limitations of the current Web is the lack of structure of its documents and the information contained in them. Besides, information overload and poor aggregation of contents make the current Web inadequate for automatic transfers of information (Berners-Lee, Hendler & Lassila, 2001; Lu, Dong & Fotouhi, 2002; Moura, 2001).

Another limitation is the fact that the current Web uses only a human-oriented type of communication. Information on the Web is conceived to have human beings as its consumers, not to be understandable by machines or software agents. In these circumstances, only human beings can understand and manipulate online information. Because of this, the current Web is not exploring all of its potentials (Berners-Lee, Hendler & Lassila, 2001).

Berners-Lee et al. (2001) appear as founders of the next generation Web, to which is given the name Semantic Web. The idea of the Semantic Web is to bring the Web to its full potential. It will have a positive impact on all levels, from individual users to large companies.

THE SEMANTIC WEB

The vision of the Semantic Web, as proposed by Berners-Lee et al. (2001), is the evolution of the current Web to one where data and services are understandable and usable by humans as well as computers. It is important to say that the Semantic Web is not artificial intelligence (Berners-Lee, 1998). The objective of the Semantic Web is not to make computers understand the human language. Processing and relating of contents do not mean an intelligent processing in the same concept that is used by the artificial intelligence researchers. The challenge of researchers of the Semantic Web is to define a universal language for the expression of data and a set of inference rules that computational agents can use to process it.

The idea of the Semantic Web is to define and link the data in a way that it can be used by machines, not just for display purposes, but for automation, integration, and reuse across heterogeneous applications. For that purpose, data needs to be application-independent, classifiable, editable, and part of a large information ecosystem (known as ontology). This type of data, embodied with semantic information, is known as smart data. With smart data, machines can interpret, manipulate, and make inferences about it. The Semantic Web can be defined as a machine-processable Web composed of smart data; the power is moving from applications to data (Berners-Lee,

Table 1. Main advantages of the Semantic Web

Main Advantages of the Semantic Web
• It is a way to describe and expand the current Web, adding a concept layer to it. • Allows machine-readable, interpretable and editable web content. • Offers a way to enable semantic annotations that could be easily organized and found. • Enhances search mechanisms with the use of "Ontologies" - relationships and axioms between concepts, allowing the standardization of web annotations, service descriptions, and web data to be meaningfully related. • Enables software agents to carry sophisticated tasks automatically – through the use of smart data. • Allows better communication between platform-independent software agents. • Enables the use of levels of trust for information.

Hendler & Lassila, 2001; Daconta, Obrst & Smith, 2003; Zhao & Sandahl, 2003). The main advantages of the Semantic Web can be grouped into several points, as shown in Table 1.

As the current Web allows the sharing of documents between previously incompatible computers, the Semantic Web intends to go beyond, allowing stovepipe systems, hardwired computers, and other devices to share document embedded contents (Berners-Lee, Hendler & Lassila, 2001; Daconta, Obrst & Smith, 2003).

Although very promising, the implementation of the Semantic Web is of enormous complexity. The first challenge consists of establishing standards that define an intelligent and universal form of content of Web pages in order to support better interpretation by the machines. The second stage consists of developing programs that obtain and share data from several sources. Past these two stages, it will become necessary to develop software agents that can generate new information based on the available one (Moura, 2001).

The Semantic Web is on its first steps. Therefore, its ideology and current structure must be seen as the base for future development of the Web and not as a final prototype. Researchers should avoid predictions that are too optimistic or incorrect, as happened with the first researchers of artificial intelligence in the 1950s and 1960s. It is important that Semantic Web researchers keep their feet on the ground.

Other point of attention is the resistance that human beings have to changes, new ideas, and, in particular, to having a complete understanding of the Semantic Web. The Semantic Web needs the network effect to become a reality. It is very important to alert people that a single Semantic Web page does not have any power by itself. In fact, the development of integrated tools for the Semantic Web that could be easily used and tested by every Web user is still not a reality. Research activities for these new tools need to take into account that the way people are currently writing Web pages must be changed as little as possible.

Another point is the implementation of the Semantic Web, which can create more problems than it is trying to solve. To be successful, the Semantic Web needs to be simple and powerful for all Web users. According to Clark (2002), "the Semantic Web is most often criticized by its detractors for having, in the end, very little to do with reality; or, put less pointedly, for being easier to dream about than to implement" (p. 1). Champion (2002) suggested three general threads in the critiques of the Semantic Web initiative: first, W3C members have a higher priority of Web services interoperability than Semantic Web affords; second, to date, progress of Semantic Web efforts has not shown any really useful example for Web users; third, the Semantic Web vision, as other visions in recent history, is unlikely to live up to its promise. Even though the Semantic Web may yet seem a remote dream, researchers believe that the Semantic Web will be adopted by users in the future—real conclusions on the viability of the Semantic Web are still in years to come.

THE STRUCTURE OF THE SEMANTIC WEB

The structure of the Semantic Web presented by W3C is based on a Semantic Web stack, as shown in Figure 1. As we move up in the stack, technologies represent