Chapter I

The Syntactic and the Semantic Web

Jorge Cardoso
University of Madeira, Portugal

ABSTRACT

This chapter gives an overview of the evolution of the Web. Initially, Web pages were intended only for human consumption and were usually displayed on a Web browser. New Internet business models, such as B2B and B2C, required organizations to search for solutions to enable a deep interoperability and integration between their systems and applications. One emergent solution was to define the information on the Web using semantics and ontologies in a way that it could be used by computers not only for display purposes, but also for interoperability and integration. The research community developed standards to semantically describe Web information such as the resource description framework and the Web Ontology Language. Ontologies can assist in communication between human beings, achieve interoperability among software systems, and improve the design and the quality of software systems. These evolving Semantic Web technologies are already being used to build semantic Web based systems such as semantic Web services, semantic integration of tourism information sources, and semantic digital libraries to the development of bioinformatics ontologies.

MOTIVATION FOR THE SEMANTIC WEB

The World Wide Web (WWW) was developed in 1989 at the European Laboratory for Particle Physics (CERN) in Geneva, Switzerland. It was Tim Berners-Lee who developed the first prototype of the World Wide Web intended to serve as an information system for physicists.

By the end of 1990, Tim Berners-Lee had written the first browser to retrieve and view hypertext documents and wrote the first Web server—the software, which stores Web pages on a computer for others to access. The system was originally developed to allow information sharing within internationally dispersed working groups. The original WWW consisted of documents (i.e., Web pages) and links between documents.

Browsers and Web server users grew. They became more and more attractive as an information sharing infrastructure. The Web became even more interesting as the amount of available...
information of every sort increased. A Web page can be accessed by a URL (uniform resource locator) through the hypertext transfer protocol (HTTP) using a Web browser (e.g., Internet Explorer, Netscape, Mozilla, Safari).

Currently, the World Wide Web is primarily composed of documents written in HTML (Hyper Text Markup Language), a language that is useful for visual presentation. HTML is a set of “markup” symbols contained in a Web page intended for display on a Web browser. Most of the information on the Web is designed only for human consumption. Humans can read Web pages and understand them, but their inherent meaning is not shown in a way that allows their interpretation by computers.

The information on the Web can be defined in a way that can be used by computers not only for display purposes, but also for interoperability and integration between systems and applications. One way to enable machine-to-machine exchange and automated processing is to provide the information in such a way that computers can understand it. This is precisely the objective of the semantic Web—to make possible the processing of Web information by computers.

The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. (Berners-Lee, Hendler, et al., 2001)

The next generation of the Web will combine existing Web technologies with knowledge representation formalisms (Grau, 2004).

The Semantic Web was made through incremental changes, by bringing machine-readable descriptions to the data and documents already on the Web. Figure 1 illustrates the various developed technologies that made the concept of the Semantic Web possible. As already stated, the Web was originally a vast set of static Web pages linked together. Many organizations still use static HTML files to deliver their information on the Web. However, to answer to the inherent dynamic nature of businesses, organizations are using dynamic publishing methods which offer great advantages over Web sites constructed from static HTML pages. Instead of a Web site comprising a collection of manually constructed HTML pages, server-side applications and database access techniques are used to dynamically

<table>
<thead>
<tr>
<th>Encoding</th>
<th>Static</th>
<th>Dynamic</th>
<th>Syntax</th>
<th>Semantic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creation</td>
<td>Manually</td>
<td>Generated by server-side applications</td>
<td>Generated by applications based on schema</td>
<td>Generated by applications based on models</td>
</tr>
<tr>
<td>Users</td>
<td>Humans</td>
<td>Humans</td>
<td>Humans and applications</td>
<td>Humans and applications</td>
</tr>
<tr>
<td>Paradigm</td>
<td>Browse</td>
<td>Create/Query/Update</td>
<td>Integrate</td>
<td>Interoperate</td>
</tr>
<tr>
<td>Applications</td>
<td>Browsers</td>
<td>Browsers</td>
<td>Process, Integration, EAI, BPM, Workflows</td>
<td>Intelligent agents, Semantic engines</td>
</tr>
</tbody>
</table>

Figure 1. Evolution of the Web