Category: Concepts, Trends, and Challenges

Emergent Semantic Web

Mahesh S. Raisinghani
Texas Woman's University, USA

Tapas R. Sahoo
University of Dallas, USA

INTRODUCTION

In less than a decade, the World Wide Web has become popular because of the depth of information it provides and the simplicity of its usage by simple clicks through related and interlinked pages. However, the amount of information and the numerous formats in which it is presented are simply overwhelming, and it is not uncommon to get overloaded with irrelevant or unrelated information. For example, a simple search task of finding books written by an author named David Flower would fetch hundreds of pages that merely contain the words David and/or Flower.

The Web contains information on millions of Web pages interwoven by the use of hyperlinks and presented in rich HTML (hypertext markup language) formats, such as images, graphics, audio, and video. This rich presentation capability makes the Web highly readable for humans, but adds no meaning to the information when read by computers.

The Semantic Web, which is considered to be the next evolution of the current Web, would qualify information with well-defined meaning. This added meaning to data, called metadata, would enable computers and people to work in cooperation (Hendler, Berners-Lee, & Miller, 2002). In addition to having hyperlinked pages containing media objects, the Semantic Web will also contain resources pointing to real-world objects such as people, places, organizations, and events. These objects will be linked based on their real-world relationships.

Another goal of the Semantic Web is to develop enabling standards and technologies designed to help machines understand more information on the Web so that they can support richer discovery, data integration, navigation, and automation of tasks (Berners-Lee, Hendler, & Lassila, 2001). The current Web has the potential of becoming the largest database system, but it suffers from its foundation as a presentation media. This article addresses issues involved in effectively storing and managing data on the Web and focuses on various research activities in this direction.

The Semantic Web is a vision that will extend the current Web to give well-defined meaning to information, enabling computers and people to work in better cooperation. A collaborative effort between the World Wide Web Consortium (W3C) and a large number of researchers and industrial partners is defining standards and technologies required for building the Semantic Web. This effort will enable data to be understood by machines and will be used for effective discovery, automation, integration, and reuse across applications.

BACKGROUND

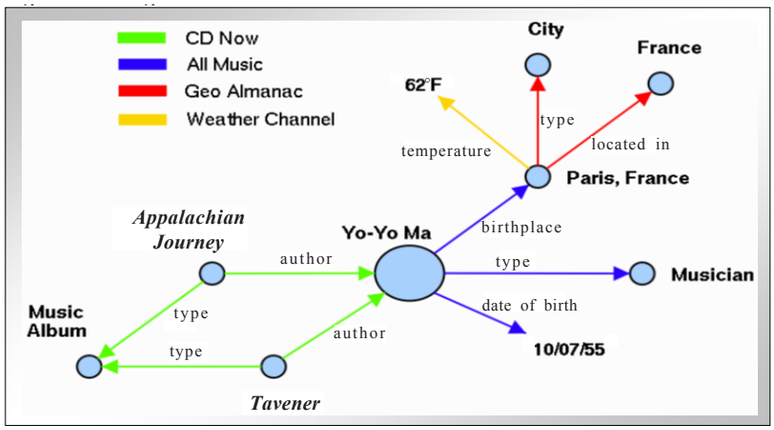
The Semantic Web is not just a web of documents; it is a web of relations between resources representing real-world objects, such as people, places, and events. It includes documents describing explicit relationships between objects and containing semantic information intended for automated processing by the machines.

Figure 1 shows a small chunk of the Semantic Web corresponding to the cellist Yo-Yo Ma (Guha, McCool, & Miller, 2003). It contains objects such as the city of Paris, the musician Yo-Yo Ma, the music album *Appalachian Journey*, and so forth. It is clear from the diagram that many different sources such as CD Now, All Music, Geo Almanac, and the Weather Channel have published different types of information about Yo-Yo Ma.

The Semantic Web extends the cumulative knowledge about any resource in a distributed fashion. This example illustrates the basic idea behind building the Semantic Web.

To transform the novel idea behind the Semantic Web into a reality, the designers of the Semantic Web are following a bottom-up approach to deal with the complexities involved in such a gigantic structure. They are building simple components for specific purposes that can be glued together in a layered structure. Figure 2 displays this layered architecture of the Semantic Web along with its various components. Some of the major components are described in the following section.

Figure 1. A segment of the Semantic Web (Source: Semantic Search)



SEMANTIC WEB ARCHITECTURE

In this section, some of important architectural components of the Semantic Web are analyzed and a proposal is derived.

Infrastructural Components

Uniform Resource Identifiers

Each item on the Web is considered a resource, and uniform resource identifiers or URIs are used to uniquely identify them (Swartz, 2002). URIs can be assigned to real-world objects like persons, places, books, and so forth. The most common form of URI is the universal resource locator (URL), which represents the address of a unique Web page on the Internet. However, the primary function of a URI is to identify a resource in lieu of providing an address of a specific file on the Web.

Resource Description Framework

In order to automate the understanding of data by machines, metadata have to be added to describe the data contained on the Web. The resource description framework (RDF) is the standard followed by W3C to process metadata on the Semantic Web (Brickley & Guha, 2000). RDF is a framework to create statements about resources in a machine-readable format and is based on the idea of identifying things using URIs and describing resources in terms of simple properties and property values. This enables RDF to represent simple statements about resources as directed, labeled graphs of nodes and arcs representing the resources and their properties and values (Manola & Miller, 2003).

Ontologies

Ontologies are ways to describe the meaning and relationships between terms. RDF is used to create these descriptions that help computers know how to use different terms.

The ontology for a domain enumerates and gives semantic descriptions of concepts in the domain of discourse, defining domain-relevant attributes of concepts and various relationships among them. For example, an ontology that describes wines will include concepts like vintages, wine regions, wineries, and grape varieties. It will also include relations such as by whom a wine is

Figure 2: Architecture of the Semantic Web (Source: Semantic Web-XML 2000, http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html)