ABSTRACT

Semantic technologies are a new wave of computing, using explicit representation of meaning to enable data interoperability and more powerful and flexible information services and transactions. At the core of semantic technologies are ontologies, which capture meaning explicitly and may be used to manipulate and reason over information via its semantics. Unlike traditional data schemas or models, ontologies are capable of representing far more complex relations, may be linked directly to the data they describe, and have a formal logical semantics, facilitating automated deductive reasoning. This chapter introduces the vision of semantic technologies, and provides an overview of the approach and the techniques developed to date. It provides both an executive summary and an orienting framework for reading more technical material.

INTRODUCING THE VISION

I have a dream for the Web [in which computers] become capable of analysing all the data on the Web—the content, links, and transactions between people and computers. A “Semantic Web,” which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The “intelligent agents” people have touted for ages will finally materialize. (Berners-Lee & Fischetti, 1999, p. 169)

Technology visionaries like Sir Tim-Berners Lee, the inventor of the World Wide Web, have long dreamed of such a seamless information technology platform (Berners-Lee & Fischetti, 1999) to support distributed business and government and personal interactions, as well as other information-based activities like research, learning, and entertainment. The benefits
Overview of Semantic Technologies

of sharing and using knowledge seamlessly, globally, and on demand hold great promise for the future of economics, government, health, the environment, and all areas of human life. Semantic technologies, which are designed to process information at the level of its meaning, hold the key for delivering this vision.

The amount of worldwide digital data generated annually is now measured in exabytes ($10^{18}$ bytes), (Lyman, & Varian, 2003) providing access to unprecedented amounts of information. While methods and technologies to store data and retrieve it reliably and securely over distributed environments are well-developed and generally highly effective, the ready availability of vast amounts of data is, in itself, not enough. Each data store is designed within its own organization or business unit for a specific purpose, and the resulting vocabularies, data formats, data structures, data value relationships, and application processing vary considerably from one system to another. Faced with information overload and a spectrum of incompatibility, most organizations are experiencing a constant struggle to find, assemble, and reconcile even a portion of the potentially relevant and useful data, even within the enterprise itself, and the potential benefits of leveraging the knowledge implicit in this data are largely untapped.

Semantic Technologies are a new wave of computing (Niemann, Morris, Riofrio, & Carnes, 2005) that enable one system to make use of the information resident in another system, without making fundamental changes to the systems themselves or to the way the organization operates. In the same way that a universal power adaptor enables an Australian appliance to be plugged into a PowerPoint in Europe, the U.S., or Asia without the need to change the local power grid, semantic technologies enable semantic interoperability for IT systems with different data structures, formats, and vocabularies, without changing the core systems themselves. By providing more effective ways to connect systems, applications, and data, greater capabilities like intelligent search, automated reasoning, intelligent agents, and adaptive computing become possible, and the potential to leverage existing information for far greater benefits becomes realizable.

HARNESSING SEMANTICS

Typically, each IT system reflects the unique missions, work flows, and vocabularies of its own organization. Differences in syntax, structure, and the concepts used for representation prevent the interoperability of information across systems and organizations. Whilst middleware and data exchange standards like XML (Bray, Paoli, Sperberg-McQueen, Maler, & Yergeau, 2006) address some of the problems, they provide only a partial solution. The main obstacle in achieving efficient and seamless system integration is the lack of effective methods for capturing, resolving, and using meaning, a field referred to as "semantics."

To date, information processing has been primarily at the syntactic or symbol-processing level, whilst the semantic level—the level of the meaning of the information—has been relatively inaccessible to machine processes. The knowledge of exactly what the data means resides in the mind of the database architect, system designer, or business analyst, or, if made explicit, in a document or diagram produced by these people. Such documentation is not in an executable form and without a direct function in the live system it quickly becomes out of date. On the other side of the coin, the understanding of the needs and wants of the information consumer resides in their mind, and traditionally there has been no way for them to represent this directly or to match their needs with the system.

Semantic technologies provide the capability to handle information on the basis of its meaning, or semantics. The core idea of semantic technologies is to use logical languages to make the structure and meaning of data explicit, and to attach this information directly to the data, so that at run-time, automated procedures can determine whether and how to align information across systems. By enabling this "semantic interoperability" across systems, a linked virtual data structure is created, where the relevant data can be searched, queried, and reasoned over across multiple native data stores based on its common meaning.