

Spinning the Semantic Web

*Are you looking for a method for managing Web data in a useful and orderly format?
Your search is over. This handbook is both a good introduction for the metadata
neophyte and a useful reference for the semantically hip.*

Reviewed by Carl Bedingfield

Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential, edited by Dieter Fensel, James Hendler, Henry Lieberman, and Wolfgang Wahlster, The MIT Press, 2003

The Internet, and its graphical overlay, the World Wide Web, started off as a simple message and document exchange medium, a practical tool intended to simplify the work of academics and researchers. However, because of the early recognition of its commercial potential, application developers, service providers and marketers worked hard to expand the power of the Internet by adding security, flexibility, sound, images, color, formatting, layout and even money management. Their efforts have arguably turned the Net from a cozy village into a dazzling cosmopolitan resource, combining the cultural impacts of billboard, shopping mall, pawn shop and peep show, to name just a few.

The utility of this standardized medium of information exchange has led to its wild popularity, and this popularity to problems of a scale not foreseen by its starry-eyed developers. One of the biggest problems is that information on the Net is typically stored and tagged in a wide variety of inconsistent ways. The World Wide Web has thus become a victim of its own scale. Even a well-constructed search request usually turns up thousands or millions of results, the bulk of which are obviously irrelevant. Obvious to the human reader, that is.

"Spinning the Semantic Web" proposes a broader vision for the Internet, and a framework for addressing many of the problems that hamper the Net as it reaches toward its ultimate potential. The claim of the semantic Web is that the Net will be more useful when information is catalogued and stored in a way that allows successful automated (and manual) methods to search, retrieve and utilize information.

As an example of the difficulty of locating and using information on the Web, imagine that you are asked to design a general-purpose, automated, price-finding software agent. Your program might start by parsing page after web page of HTML code to identify

likely sites that offer products for sale, perhaps by looking for key words (price, quantity, sale). Having found such a page, the agent might then look for a currency character, perhaps further restricting the search to entries having a decimal marker preceded by at least one numeric character, and followed by two. Your agent would need to expand the search to identify table columns containing any of these indicators, correctly matching product descriptions with prices. Words such as "each," "quantity," "per kilogram," and "per meter" must also be considered. To further complicate the task, a growing number of pages use dynamic and database-driven pages, which present product and pricing information only in response to a customer's specific selection.

This crude example only begins to demonstrate the challenges of coping with the current structure of the Web. Information is not stored in any standardized format, nor tagged with useful identifying information. The Semantic Web offers a means of managing Web data in an orderly and useful format.

The task of defining and implementing a Semantic Web offers both opportunity and challenge. In the mid-'90s, the World Wide Web Consortium (W3C) began to explore the use of metadata (data which describes data) to enrich HTML Web content. The W3C activities on metadata morphed into the Semantic Web Activity, chartered in February 2002. The W3C defines the Semantic Web as "the representation of data on the World Wide Web. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming."

Much of the significant work of these "researchers and industrial partners" is collected in "Spinning the Semantic Web." Following an inspiring introduction by Net patriarch Tim Berners-Lee, the book presents three sections (Languages and Ontologies, Knowledge Support, and Dynamic Aspect) in fifteen chapters. Each chapter is an academic-quality paper describing the work of one of the various groups working on the Semantic Web. The chapters range from introductory material, theoretical explanations filled with acronyms and detailed definitions, to descriptions of actual projects. These projects focus on the application of these semantic structures and processes to real-world problems. The technologies described include intelligent agents, markup languages, knowledge bases, and natural language processing, to name just a few.

The book is edited well. The chapters are organized in a logical sequence, and the thoughts flow smoothly. Despite the fact that the book is assembled from multiple sources, it presents a complete and well-formed picture of the discipline. The discussions

are of adequate depth to accurately represent the work being described, without losing the reader in unnecessary detail. The broadness of the chapter topics, and the multiple, slightly different, views of the subject matter, serve to give depth to the topic, adding realism.

In short, the book is, as the flyleaf claims, the first handbook for the semantic Web. It manages to provide both a good introduction for the metadata neophyte, and a useful reference for the Semantically hip. The task of improving the usefulness of the Web is an on-going, multi-track effort. "Spinning the Semantic Web" effectively describes one of the leading industry directions toward that end.

Carl Bedingfield is a principal member of the technical staff of a large telecommunications company. His work includes the design and implementation of new services.

Source: Ubiquity, Volume 4, Issue 41, Dec. 10 - 16, 2003