

Spinning the Semantic Web,

**edited by Dieter Fensel, James Hendler, Henry Lieberman and
Wolfgang Wahlster**

MIT Press, 478 p., 2003. ISBN 0-262-06232-1

Review by

Guy Lapalme

RALI-DIRO

Université de Montréal

This book deals with an interesting and important problem, namely considering the World Wide Web as a *web of meaning* instead of a *web of links*. It looks at ways of creating richer modes of representation and tagging of information than what is currently possible with HTML, which is mainly limited to presentation aspects. A richer representation mode could enable new applications for both humans and machines because by using semantics for inferences and thus going around the limitations of the different modes of presenting the same information.

Spinning the Semantic Web was published by MIT Press, in 2003 as a compendium of 15 articles based on talks presented at a Seminar in Dagstuhl, Germany in March 2000. It comprises 15 articles from 37 authors and 4 editors, most of them well established researchers in the domain. The articles are divided into three parts: language and ontologies, knowledge support and dynamic aspect. The foreword of the book by Tim Berners-Lee (the *inventor* of the web) is based on a talk given at the World Wide Web Consortium conference in December 1997. Given that the web is changing so rapidly, one would have imagined that this book could be out of date because it reviews works from more than three years ago. But overall, points made in the book are still relevant which can be either seen positively (these researchers have a good insight) or negatively (real evolution is always slow even in these Internet times). As the ideas put forward in this article are not yet fulfilled after all these years, it shows to what extent the Semantic Web is still far ahead because the semantic gap is quite hard to leap over.

The editors have unified the presentation of the chapters, which are professionally typeset using a uniform format and numbering across the book. Unfortunately, some parts of XML code (e.g. figures 4.2 or 4.7) are badly aligned: initial tabulations seem to have been ignored or some long lines have been wrapped to the start of the next line; this is unfortunate because these figures become unintelligible given the high level of nesting needed for even simple XML examples. Although there is a common index for all the articles, no other links are done within the different chapters even though there are many references to works described in the same book; for example, figure 6.1 is the same as the one given in the foreword but no cross-references are done between them. Chapters have their own list of references and endnotes and can be read in any order. The editors have thus lost an excellent opportunity of doing what they are preaching, because these chapters would have benefited from a rich network of semantic links between them. This would have made a nice test bed for the ideas put forward in the book. It could be done in a subsequent electronic version of the book.

The first chapter, written by the editors, is a very good introduction to the notion of semantic web, defined as follows: languages for machine understandable meta-information with tools using these languages for supporting finding, access, presentation and maintenance of information. Realizations of these new types of information are illustrated in the domain of electronic commerce (B2C - Business to Consumers) and electronic business (B2B – Business to Business). This introduction is a good way of making sense of the most of the many acronyms that are used in the rest of the book but unfortunately it does not make any explicit references to other chapters although they motivate briefly the somewhat arbitrary division between chapters into the following three parts:

1. *Languages and ontologies* which describe formalism for annotating objects
2. *Knowledge support* tools for using the languages and ontologies in applications
3. *Dynamic aspect* which goes further than the current web technologies for the interaction with the user

The quality and interest of the other chapters are quite uneven. There are enough interesting chapters to make this book a worthy buy but unfortunately some of them are not up to the level of interest and readability one would expect in this series of book. As chapters are independent, we review them separately (giving the corresponding part number in parentheses) and list them according to *our order of decreasing interest*.

Chapter 6 (part 1) is by far the best introduction to ontologies that we have read: it presents a useful classification and compares them to other kinds of knowledge classifications: list of terms, glossary, is-a hierarchy, frames and logical constraints. It then presents the use of ontologies, their acquisition, their implications and their needs. This chapter is a must read by anybody working with ontologies and looking for a good overview of the domain.

Chapter 4 (part 1) is a very interesting discussion about the relative merit of ontologies, XML schemas and RDF (Resource Description Framework) schemas which share a certain number of properties but which differ mainly in their relative level of abstraction. It also shows interesting translation mechanisms between these different languages.

Chapter 12 (part 3) presents *Semantic Gadgets*, which are small devices that communicate with their environment and make appropriate suggestions or negotiations with other gadgets. It describes a challenging thought experiment about some possible use of these gadgets: going to a museum, finding a parking place, recommending an order of the visit, suggesting buying a gift, etc... This description is very well done and is an excellent motivation for the need of more complex ontologies for even these simple devices even though a full implementation is clearly far off.

Chapter 8 (part 2) is devoted to the Task-Centered Knowledge support and makes a very good point for considering the Web as a *place for doing* instead of only a *place for finding*. There is a very interesting and thorough example in the domain of aerospace customer service; unfortunately the bad layout renders its almost unreadable, but once we get over this the chapter is well worth reading.

Chapter 10 (part 2) tackles the problem of complex relationships in the semantic web through the presentation of *InfoQuilt* in which one can describe knowledge about domains (ontologies),

resources about them, interdomain relationships, functions and simulations. A challenging problem is described: finding possible relations between earthquakes and nuclear tests. Usual queries are replaced by so-called Iscapes (Information Scapes) that are given in natural language but that should be formalized in XML by a process unfortunately not described in the chapter.

Chapter 2 (part 1) describes SHOE (Simple HTML OntologiE) is a very good introduction to the domain of ontologies and a clear description of SHOE and a comparison with other mark-up languages, knowledge representation languages and datalog. It makes a clear distinction between categories, relations and inference rules. SHOE has been applied within a system with more than 40 000 facts about Computer Science departments in the USA. It has also been applied in a web site about food safety risks.

Chapter 13 (part 3) gives a compelling comparison between the static and dynamic semantics of the web. A solution given is through GLUE, an actor based programming language embedded within HTML. Although this language is relatively clean and simple, it does not seem to have sufficient novelty or advantages to replace Javascript, which is now a *de facto* standard for creating client-side dynamic HTML pages.

Chapter 14 (part 3) motivates the needs for an external semantic annotation of Web content in order to adapt it to different output devices. This problem is well known and better described in other chapters of the books. The chapter shows an example of adaptation of a single content to a small device, a process usually solved nowadays with XSLT transformations of Cascading Style Sheets.

Chapter 9 (part 2) deals with mobility of knowledge and its interoperability through translation. It discusses the difficulties of interoperability between knowledge bases. The problems are well identified at a high level of abstraction. They suggest solving these problems with yet another knowledge base tool, which is described by a screen dump and a simplistic example without any reference to any of the problems previously identified.

Chapter 7 (part 2) describes SESAME, an architecture for the efficient storage of large quantities of RDF data. As the description is at a very high level, it is not useful to appreciate the power of the system.

Chapter 11 (part 2) describes SEAL (SEmantic portAL) through some methodological aspects of building ontologies and shows the implementation of a web site based on these patterns and so called "intelligent means of accessing information". The description is monotonous, long-winded and unexciting. This important subject could have been presented in a more dynamic way and would have made a much more interesting reading.

Chapter 15 (part 3) describes Task Achieving Agents within a planning framework, which is well explained but has only tenuous links with the web except for its display. All situations are described at the conceptual level except for a cell phone display problem which arrives out of nowhere at the end of the article.

Chapter 3 (part 1) relates the tortuous history of the development of DAML (DARPA Agent Markup Language) through contracts with various teams. Some trivial parts are well described

while others are left unspecified. The examples given are too simple to make clear the power of the formalism.

Chapter 5 (part 1) is about UPML (Unified Problem solving Method development Language), an architecture for describing heuristic and dynamic reasoning. Unfortunately the chapter is hard to follow because it combines many different projects without any explicit relationships between them.

Even with the uneven quality of its chapters, this book is a good introduction to the domain of the Semantic Web for anybody wanting to get the most out of the Internet of the future.

Terminologists will appreciate the chapters of the first part dealing with ontologies whose importance is starting to be acknowledged and that show the power and richness of the words.

Although the book does not cover all aspects of the Semantic Web, it deals with the most important parts and it provides a much more solid introduction to the domain than trying to sort out the many documents at semanticweb.org or than typing "semantic web" into Google...