The Semantic Web: Year One*

Jérôme Euzenat[1] and Amedeo Napoli[2]

[1]INRIA Rhône-Alpes
[2]LORIA

October 2003

Book review of Dieter Fensel, James Hendler, Henry Lieberman, Wolfgang Wahlster (eds.), "Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential", The MIT Press, Cambridge (MA US), 2003, 0-262-06232-1, 479pp. US$40

The Semantic Web began at the end of the last century as an object of investigation and experimentation. It has grown into a subject of huge interest to several research communities and to Web users at large. Many people consider it the next generation of the Web, a Web that machines themselves can handle to help users.

The few available books on the Semantic Web cover its programming or research aspects. Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential, however, is a kind of handbook on the Semantic Web's first steps and the principles behind them. The book stemmed from a Daghstul seminar in 2000, at the beginning of the Semantic Web development, and features contributions from some of the main actors in that development. It includes general chapters on the subject and several chapters relating early experiments and system designs (indeed, these chapters were written just before the OWL, or Web Ontology Language, era). In this way, this well-crafted book can be a starting point and reference book for understanding the forthcoming Semantic Web design issues.

From a general point of view, Spinning the Semantic Web contains no formalisation (it is not very technical, except that some chapters display computer code). Rather, the book describes systems, languages, and their designs.

In the book's foreword, Tim Berners-Lee offers a techno-philosophical motivation of the Semantic Web. In the introduction, editors Dieter Fensel, James Hendler, Henry Lieberman, and Wolfgang Wahlster present an extensive overview of applications the Semantic Web is meant to enable and the required technologies, and then describe the book itself. These two sections are complementary: while Berners-Lee focuses on information and linking, Fensel and his colleagues center on applications and resources. They present the Semantic Web infrastructure as a set of languages, ontologies, and intelligent information access, and applications developed on top of these. This roughly parallels the rest of the book structure.

*This is the author version of the text published in *IEEE Intelligent systems* 18(6):76–78, November-December 2003 [DOI: 10.1109/MIS.2003.1249173]

The core of the book comprises three parts: Languages and Ontologies, Knowledge Support, and Dynamic Aspect. We found that the chapters tend to fall into two categories: ones raising problems and ones offering solutions. This distinction is not absolute; presenting a solution requires presenting the problem it solves, and presenting a problem is often followed by some tracks to solve it. However, it is useful for the purposes of this review.

Languages and ontologies

To be understood by machines, the Semantic Web pages (or resources) should be expressed in a language designed for this purpose. Such a language is a crossbreed of knowledge representation languages from artificial intelligence and markup languages from document engineering. This part of the book covers various languages designed from the Semantic Web perspective. It includes materials about languages that are not widely available, providing a good foundation for understanding them and their design rationale. Of course, all of the chapters that present languages provide solutions for expressing knowledge on the Web.

Chapter 2 describes SHOE (Simple HTML Ontology Extensions) and the developments surrounding it. SHOE is based on HTML and is a precursor of the current Semantic Web infrastructure. This chapter explains what is new in using a knowledge representation language in a distributed environment: interoperability, evolution, and scalability.

Chapter 3 covers DAML-ONT (Darpa Agent Markup Language-Ontology). You can think of DAML-ONT as a frame language on top of RDFS (Resource Descriptive Framework Schema).

Chapter 4 introduces OIL (Ontology Inference Layer), RDFS, and XML (Extensible Markup Language) Schema. This chapter is an interesting overview of these languages and their merits. It also details specific points, such as the way to embed OIL (and thus OWL) within RDFS.

Chapter 5 discusses Web services with UPML (Unified Problem-Solving Method Development Language) and IBrow, an intelligent broker for retrieving a set of knowledge components from the Web. With IBrow, the reasoning components themselves must be available on the Web, and the broker selects the set of suitable components for the ongoing task. So, you can think of UPML as a precursor of Web services description languages. The chapter briefly describes how to use the Protégé editor in this framework.

Chapter 6 provides different definitions of ontology and the kind of applications ontologies can support. Because this chapter is rather introductory, exploring the jungle of multiple definitions, you can read it before the other chapters in this part.

These chapters would have been clearer and more understandable if they had shared a common example, but this is difficult to achieve when the chapters have different authors. This part of the book does show well the need for each successive language to have increased expressive power (especially the chapter on OIL, RDFS, and XML Schema). The emphasis on languages predating DAML+OIL and OWL should help readers understand the genesis of more recent languages.

Knowledge support

The second part covers tools and architectures that will take advantage of the Semantic Web resources. Two chapters in this part provide solutions. Chapter 7 takes a look at Sesame, an architecture for efficient storage and expressive querying of large quantities of metadata. The chapter includes a simple presentation of RDF and RDFS, as well as a query language for RDFS. In particular, it explains how answers to queries differ at the syntactic, structural, and semantic levels. For semantic-level querying, the authors suggest using RQL (RDF Query Language, an extension of Ob-

ject Query Language for the Semantic Web) or a similar query language. The chapter also discusses a concrete experiment with the PostGres database management system.

Chapter 11 promotes ontologies as a sound basis for communications of all types on the Web. This position is based on experience with Ontobroker and relies on Ontobroker's application to a laboratory Semantic Web site (a Web site generated from data systematically related to an ontology). The chapter also discusses modelling and designing ontologies (ontology engineering) and comparing ontologies. The authors describe the SEAL tool suite, built around Ontobroker, and discuss topics important for Semantic Web sites, such as ranking knowledge and personalising the user experience. The chapter tells how to include such a site within the Semantic Web through knowledge export in RDF.

The other chapters deal primarily with problems. Chapter 8 is oriented toward industry. It advocates interpreting resources in regard to tasks. It also provides a systematic analysis of querying a set of resources in terms of the frequency of queries and ability to anticipate them. The chapter describes a help desk application that exploits the Web and knowledge-based resources.

However, it does not go all the way in providing a task-oriented view of the application. In particular, the chapter gives no key for generalising the approach to the Web.

Chapter 9 presents the risks of interoperability through ontologies because of granularity mismatches, inconsistencies, and modelling mismatches between ontologies. The author suggests using knowledge bases hyperlinked to their sources and elaboration substrate. These "resilient hyperknowledge bases" should provide the required mobility to ontologies. This perspective would be more integrated with the Semantic Web if the author had defined it in terms of the Web, which remains the ideal context for linking to background sources.

Chapter 10 deals with relationships as a fundamental element for supporting semantics in knowledge systems. In queries that involve several domains at once, interoperability problems exist between ontologies. One proposed solution is to model relationships between ontologies. This chapter includes several interesting examples that you could use for designing systems for the Semantic Web that take into account complex relations.

Because most chapters in this part of the book focus on problems rather than solutions, the "support" aspect is lacking. Nonetheless, the discussion of Semantic Web design problems and general principles for addressing them provide much food for thought.

Dynamic aspect

In this part, the book shifts its focus from the Semantic Web evolution to its dynamic use. Again, two chapters present problems. In Chapter 12, the problem is opportunistic communication between heterogeneous and unprepared knowledge sources. This chapter covers ubiquitous computing: small, handheld, wireless devices for computing the needs of everyday life. The ultimate goal is for automated systems to discover and use services without human intervention. However, the story presented in this chapter, a visit to a museum, illustrates a recurring question: When the Semantic Web starts handling everyday decisions, where should it stop? This old question does not receive any answer here.

Chapter 13 discusses Web page dynamic aspect during user interaction. The author advocates including this procedural element in the Semantic Web through a dedicated prototype-based language embedded in XML pages.

The last two chapters offer solutions. Chapter 14 describes a content adaptation engine that adapts requested documents in accordance with the capabilities of the client sending the request (transcoding). This chapter concerns itself strictly

with syntactic transformation ("semantic" denotes the transformation context) but shows the way to extend such methods to semantic descriptions.

Chapter 15 proposes the Inova language for representing issues, tasks, and plans and the OP3 (Open Planning Process Panel) framework for representing the planning process. This chapter provides a methodology for integrating planners and users within the Semantic Web. Regrettably, Chapter 15 is not tied more to the UPML presentation in Chapter 5.

This part of the book highlights the various ways the Semantic Web could handle dynamicity. Dynamicity plays an important role when Semantic Web resources must be used in context, but current development does not widely address it.

Spinning the Semantic Web is a good, substantial book on the Semantic Web origin and design. However, we would have expected better coordination between chapters. Cross-references would have been useful for both understanding the material and comparing approaches (the index partly serves this purpose).

The book also contains several conflicting ontology definitions and RDF presentations. A more integrated construction (even explaining disagreements) would have contributed to readability. Finally, from a historical standpoint, we would have liked an indepth presentation of the philosophy behind Ontobroker and the DAML program.

Nonetheless, we warmly thank the editors for gathering in a single volume these enlightening contributions documenting both the historical and technical aspects of the Semantic Web early development—most of the contributions would not have been available otherwise. We highly recommend it to everyone wanting to work in the field. This comprehensive reference could help many newcomers grasp the spirit behind the Semantic Web.

As we mentioned before, this book is a handbook on the Semantic Web in the making. The answers it provides will be realised only "by construction" in the forthcoming languages and tools for the Semantic Web. We encourage a follow-up book that integrates all the presented approaches in their presentation and, above all, in their implementation.

Jérôme Euzenat is a research scientist at INRIA Rhône-Alpes. Contact him at jerome.euzenat@inrialpes.fr.
Amedeo Napoli is a senior research scientist at LORIA Nancy. Contact him at amedeo.napoli@loria.fr.