Preface

Objectives of the Book

This book is all about the Semantic Web.

From its basics, the Semantic Web can be viewed as a collection of standards and technologies that allow machines to understand the meaning (semantics) of information on the Web. It represents a new vision about how the Web should be constructed so that its information can be processed automatically by machines on a large scale.

This exciting vision opens the possibility of numerous new applications on the Web. Since 2001, there have been many encouraging results in both academic world and real application world. A whole suite of standards, technologies, and related tools have been specified and developed around the concept of the Semantic Web.

However, such an extensive and rapid progress of the Semantic Web has presented a steep learning curve for those who are new to the Semantic Web. Understanding its related concepts, learning the core standards and key technical components, and finally reaching the point where one can put all these into real development work require a considerable amount of effort.

To facilitate this learning process, a comprehensive and easy-to-follow text is a must. This book, A Developer’s Guide to the Semantic Web, serves this exact purpose. It provides an in-depth coverage on both the What-Is and How-To aspects of the Semantic Web. From this book, you will not only obtain a solid understanding about the Semantic Web but also learn how to combine all the pieces together to build new applications on the Semantic Web. More specifically,

- it offers a complete coverage of all the core standards and technical components of the Semantic Web. This coverage includes RDF, RDFS, OWL (both OWL 1 and OWL 2), and SPARQL (including features offered by SPARQL 1.1). Other related technologies are also covered, such as Turtle, microformats, RDFa, GRDDL, and SKOS;
- it provides an in-depth description of multiple well-known applications and projects in the area of the Semantic Web, such as FOAF, semantic Wiki, SearchMonkey by Yahoo!, Rich Snippets by Google, Open Linked Data Project, and DBpedia Project;
• it explains the key concepts, core standards, and technical components in the context of examples. In addition, the readers will be taken in a step-by-step fashion through the development of each example. Hopefully for the first time ever, such teaching method will ease the learning curve for those who have found the Semantic Web a daunting topic;
• it includes several complete programming projects, which bridge the gap between What-Is and How-To. These example applications are real coding projects and are developed from the scratch. In addition, the code generated by these projects can be easily reused in the readers’ future development work.

Intended Readers

The book is written with the following readers in mind:

• software engineers and developers who are interested in learning the Semantic Web technology in general;
• Web application developers who have the desire and/or needs to study the Semantic Web and build Semantic Web applications;
• researchers working in research institutes who are interested in the Semantic Web research and development;
• undergraduate and graduate students from computer science departments, whose focus of work is in the area of the Semantic Web;
• practitioners in related engineering fields. For example, data mining engineers whose work involves organizing and processing a large amount of data by machines.

The prerequisites needed to understand this book include the following:

• working knowledge of Java programming language and
• basic understanding of the Web, including its main technical components such as URL, HTML, and XML.

Structure of the Book

This book is organized as follows:

• Chapters 1–6 cover the basic concept, the core standards, and technical components of the Semantic Web. The goal of these chapters is to show you the What-Is aspect about the Semantic Web.

Chapter 1 introduces the concept of the Semantic Web by using a simple example. With this example, the difference between the traditional Web and the Semantic
Web is clearly revealed. Further discussion in this chapter helps you to establish a solid understanding about the concept of the Semantic Web.

Chapter 2 covers RDF in great detail to give you a sound technical foundation to further understand the Semantic Web. If you are not familiar with RDF, you should not skip this chapter, since everything else is built upon RDF. In addition, Turtle format is presented in this chapter, which will be used to understand the material presented in Chap. 6.

Chapter 3 goes on with other RDF-related technologies, including Microformats, RDFa, and GRDDL. If you prefer to get a full picture about the Semantic Web as quickly as possible, you can skip this chapter. However, the material presented in this chapter will be necessary in order to understand Chap. 8.

Chapter 4 presents RDF schema and also introduces the concept of ontology. You should not skip this chapter since Chap. 5 is built upon this chapter. SKOS is also presented in this chapter; you can skip it if you are not working with any existing vocabularies in knowledge management field.

Chapter 5 discusses OWL in great detail and covers both OWL 1 and OWL 2. This is one of the key chapters in this book and should not be skipped. Unless RDF schema can satisfy the needs of your application, you should spend enough time to understand OWL, which will give you the most updated information about latest ontology development language.

Chapter 6 covers SPARQL. This is another chapter that you should carefully read. Working on the Semantic Web without using SPARQL is like working with database systems without knowing SQL. Note that SPARQL 1.1 is covered in this chapter as well. At the time of this writing, SPARQL 1.1 has not become a standard yet, so when you are reading this book, note the possible updates.

• Chapters 7–11 provide an in-depth discussion of some well-known Semantic Web applications/projects in the real application world. These chapters serve as a transition from knowing What-Is to understanding How-To in the world of the Semantic Web.

Chapter 7 presents FOAF (Friend of A Friend) project. The FOAF ontology is arguably the most widely used ontology at this point. The goal of this chapter is to introduce you to a real-world example in the social networking area. Since the modeling of this domain does not require any specific domain knowledge, it is easy to follow and you can therefore focus on appreciating the power of the Semantic Web. This chapter should not be skipped, not only because of the popularity of the FOAF ontology but also because this ontology has been used frequently in the later chapters as well.

Chapter 8 presents Google’s Rich Snippets and Yahoo!’s SearchMonkey; both are using RDFa and microformats as the main tools when adding semantic markups. These are important examples, not only because they are currently the major Semantic Web applications developed by leading players in the field but also because they show us the benefits of having the added semantics on the Web.
Chapter 9 discusses the topic of Semantic Wiki, together with a real-world example. This chapter represents the type of Semantic Web applications built by using manual semantic markup. After reading this chapter, you should not only see the power of the added semantics but also start to understand those situations where manual semantic markup can be a successful solution.

Chapter 10 presents DBpedia in great detail. DBpedia is a well-known project in the Semantic Web community, and a large number of real-world Semantic Web applications take advantage of the DBpedia datasets directly or indirectly. Also, DBpedia gives an example of automatic semantic markup. Together with Chap. 9, where manual semantic markup is used, you have a chance to see both methods at work.

Chapter 11 discusses the Linked Open Data project (LOD) as a real-world implementation example of the Web of Data concept. For the past several years, LOD has attracted tremendous attention from both the academic world and the real application world. In fact, DBpedia, as a huge dataset, stays in the center of the LOD cloud. Therefore, LOD together with DBpedia becomes a must for anyone who wants to do development work on the Semantic Web. More specifically, this chapter covers both the production and the consumption aspects of Linked Data; it also provides application examples that are built upon LOD cloud. In addition, this chapter explains how to access LOD programmatically, which should be very useful to your daily development work.

- Chapters 12–15 are the section of How-To. After building a solid foundation for development work on the Semantic Web, this section presents three different running applications that are created from scratch. The methods, algorithms, and concrete classes presented in these chapters will be of immediate use to your future development work.

Chapter 12 helps to build a foundation for your future development work on the Semantic Web. More specifically, it covers four major tool categories you should know, namely development frameworks, reasoners, ontology engineering tools, and other tools such as search engines for the Semantic Web. This chapter also discusses some related development methodology for the Semantic Web, such as the ontology-driven software development methodology. Furthermore, since ontology development is the key of this methodology, this chapter also presents an ontology development guide that you can use.

Chapter 13 covers a popular development framework named Jena to prepare you for your future development work on the Semantic Web. More specifically, this chapter starts from how to set up Jena development environment and then presents a Hello World example to get you started. In what follows, this chapter covers the basic operation every Semantic Web application needs, such as creating RDF models, handling persistence, querying RDF dataset, and inferencing with ontology models. After reading this chapter, you will be well prepared for real development work.
Developing applications for the Semantic Web requires a set of complex skills, and this skill set lands itself on some basic techniques. In Chap. 13, you have learned some basics. Chapter 14 continues along the same path by building an agent that implements the Follow-Your-Nose algorithm on the Semantic Web. After all, most Semantic Web applications will have to be based on the Web, so moving or crawling from one dataset to another on the Web with some specific goals in mind is a routine task. Follow-Your-Nose method is one such basic technique. Besides implementing this algorithm, Chap. 14 also introduces some useful operations, such as how to remotely access SPARQL endpoints.

Chapter 15 presents two additional Semantic Web applications from scratch. The first application helps you to create an e-mail list that you can use to enhance the security of your e-mail system. The second one is a ShopBot that runs on the Semantic Web, and you can use it to find products that satisfy your own specific needs. These two projects are discussed in great detail, showing how applications on the Semantic Web are built. This includes RDF documents handling, ontology handling, inferencing based on ontologies, and SPARQL query handling, just to name a few.

Where to Get the Code

The source code for all the examples, application projects in this book can be downloaded from the author’s personal Web site, www.liyangyu.com

Acknowledgment

My deepest gratitude goes to Dr. Weirong Ding, a remarkable person in my life, for supporting me in all the ways that one can ever wish to be supported. It is not nearly as possible to list all the supports she gave me, but her unreserved confidence in my knowledge and talents has always been a great encouragement for me to finish this book. Being the first reader of this book, she has always been extremely patient with many of my ideas and thoughts, and interestingly enough, her patience has made her a medical doctor who is also an expert of the Semantic Web. And to make the readers of this book become experts of the Semantic Web, I would like to share something she always says to me: “never give yourself excuses and always give 200% of yourself to reach what you love.”

I would like to thank Dr. Jian Jiang, a good friend of mine, for introducing me to the field of the Semantic Web and for many interesting and insightful discussions along the road of this book.

My gratitude is also due to Mr. Ralf Gerstner, senior editor at Springer. As a successful IT veteran himself, he has given me many valuable suggestions about the content and final organization of this book. The communication with him is always quite enjoyable: it is not only prompt and efficient, but also very insightful
and helpful. It is simply my very good luck to have a chance to work with an editor like Ralf.

Finally, I would like to express my love and gratitude to my beloved parents for their understanding and endless love. They always give me the freedom I need, and they accept my decisions even when they cannot fully understand them. In addition, I wanted to thank them for being able to successfully teach me how to think and speak clearly and logically when I was at a very young age, so I can have one more dream fulfilled today.

Atlanta, GA, USA

Liyang Yu

September 2010