Semantic Web and Web 2.0

In this final chapter, we provide the reader with a peek at the two main directions the Web has been heading in recent years: the Semantic Web and Web 2.0.

- **The Semantic Web**: So far, the Web has mainly been used as a communication medium of content and services for human consumers. Since the first emergence of the idea of the Semantic Web, there has been a trend toward accompanying content with machine-interpretable metadata to allow computers to “understand” the meaning of that content and, thus, to facilitate the access, retrieval, and exchange of whatever data that is on the Web. Huge funding and effort have been invested in this research area, and we will illustrate the principles, technologies, and main applications of the Semantic Web.

- **Web 2.0**: Since its first usage of the Web as a mere one-way communication medium, where there was a clear distinction between content providers and content consumers, things have radically changed. Currently, we see a clear trend of users acting not only as content consumers, but also as content providers. Social networks, wikis, blogs, and folksonomies, equipped with newer and more powerful technologies like AJAX, Flash, or RSS, represent what is commonly called Web 2.0. In addition, while the potential and possibilities of Web 2.0 are far from exhausted, the term Web 3.0 has already been coined. We will attempt to clarify the concepts behind these terms, and give the reader a view of what they are all about.

Semantic Web and Web 2.0 are currently hot research topics, both in academia and industry. Given their relative novelty, their results are not yet as stable as those of the topics discussed so far in this book, and their application is less widespread. Nevertheless, in this final chapter we will provide the reader with some insight into these topics, and point out how existing Web engineering methods have been adapted to cope with the additional requirements that came along.
9.1 The Semantic Web

A book on Web engineering would not be complete without a discussion on the Semantic Web. The Semantic Web is an extension to the current Web, and started out as a vision of one of its creators, Tim Berners-Lee. He realized that one of the major shortcomings of the traditional Web is the fact that its information is mostly human-interpretable, not machine-interpretable. To fully grasp this problem, consider the example of figure 9.1, showing two very simple Web pages, each representing a “mouse.” You, the reader, will immediately notice that on the left Web page an animal is shown, and on the right page a computer accessory is shown. You do so by interpreting the image and using this knowledge to attach the correct interpretation to the word “mouse” in the sentence “This is a mouse.” In other words, the meaning or semantics of the word “mouse” is provided by you, the human reader, using contextual information (in this case, the accompanying image). For a machine, however, it is impossible to grasp what is on each page. In fact, a machine only detects two pages, each containing an image and a string. It does not know what is represented in the image; it cannot interpret the sentence, nor can it correctly distinguish and interpret the two occurrences of the word “mouse.” For a computer, the information contained in the page is readable, but not interpretable.

![Fig. 9.1. Two different meanings for “mouse”](image)

This problem is illustrated by the fact that all major search engines are mainly restricted to syntax-based search queries, as they are not able to grasp the semantics of the information on the Web. Also, for automated agents crawling the Web searching for relevant information, the lack of semantics is problematic. Imagine a shopping agent that searches for the cheapest 17-inch screen available. It will run into all kinds of problems related to the lack of