The Semantic Web: A Brain for Humankind

Dieter Fensel, *Vrije Universiteit Amsterdam*
Mark A. Musen, *Stanford University*

O riginally, the computer was intended as a device for computation. Then, in the 1980s, the PC developed into a system for games, text processing, and PowerPoint presentations. Eventually, the computer became a portal to cyberspace—an entry point to a worldwide network of information exchange and business transactions. Consequently,

technology that supports access to unstructured, heterogeneous, and distributed information and knowledge sources is about to become as essential as programming languages were in the 60s and 70s.

The Internet—especially World Wide Web technology—was what introduced this change. The Web is an impressive success story, in terms of both its available information and the growth rate of human users. It now penetrates most areas of our lives, and its success is based on its simplicity. The restrictiveness of HTTP and (early) HTML gave software developers, information providers, and users easy access to new media, helping this media reach a critical mass.

Unfortunately, this simplicity could hamper further Web development. What we're seeing is just the first version of the Web. The next version will be even bigger and more powerful—but we're still figuring out how to obtain this upgrade.

Growing complexity

Figure 1 illustrates the growth rate of current Web technology. It started as an in-house solution for a small group of users. Soon, it established itself as a worldwide communication medium for more than

10 million people. In a few years, it will interweave one billion people and penetrate not just computers but also other devices, including cars, refrigerators, coffee machines, and even clothes.

However, the current state of Web technology generates serious obstacles to its further growth. The technology's simplicity has already caused bottlenecks that hinder searching, extracting, maintaining, and generating information. Computers are only used as devices that post and render information—they don't have access to the actual content. Thus, they can only offer limited support in accessing and processing this information. So, the main burden not only of accessing and processing information but also of extracting and interpreting it is on the human user.

The Semantic Web

Tim Berners-Lee first envisioned a Semantic Web that provides automated information access based on machine-processable semantics of data and heuristics that use these metadata. The explicit representation of the semantics of data, accompanied with domain theories (that is, *ontologies*), will enable a Web that provides a qualitatively new level of service. It will weave
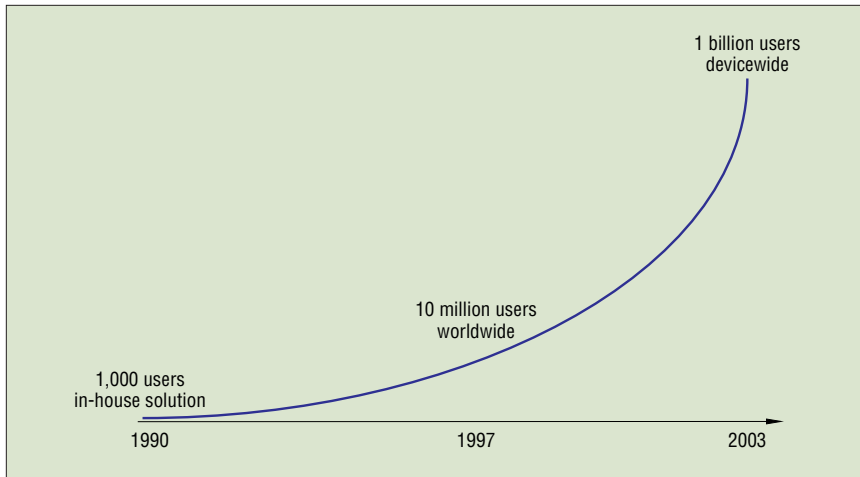
Figure 1. The growth rate of current Web technology.

Labels in figure:
- 1 billion users devicewide
- 10 million users worldwide
- 1,000 users in-house solution
- 1990   1997   2003

A key technology for the Semantic Web is ontologies. In "Creating Semantic Web Contents with Protégé-2000," Natalya F. Noy, Michael Sintek, Stefan Decker, Monica Crubézy, Ray W. Fergerson, and Mark A. Musen provide excellent tool support for manually building ontologies based on Protégé-2000. However, even with an excellent tool environment, manually building ontologies is labor intensive and costly. Alexander Maedche and Steffen Staab, in "Ontology Learning for the Semantic Web," try to mechanize The ontology building process with machine learning techniques. ■

together an incredibly large network of human knowledge and will complement it with machine processability. Various automated services will help the user to achieve goals by accessing and providing information in a machine-understandable form. This process might ultimately create an extremely knowledgeable system with various specialized reasoning services—systems that can support us in nearly all aspects of our life and that will become as necessary to us as access to electric power.

This gives us a completely new perspective of the knowledge acquisition and engineering and the knowledge representation communities. Some 20 years ago, AI researchers coined the slogan "knowledge is power." Quickly, two communities arose:

- knowledge acquisition and engineering, which deals with the bottleneck of acquiring and modeling knowledge (the human-oriented problem), and
- knowledge representation, which deals with the bottleneck of representing knowledge and reasoning about it (the computer-oriented problem).

However, the results of both communities never really hit the nail on the head. Knowledge acquisition is too costly, and the knowledge representation systems that were created were mainly isolated, brittle, and small solutions for minor problems.

With the Web and the Semantic Web, this situation has changed drastically. We have millions of knowledge "acquisitioners" working nearly for free, providing up to a billion Web pages of information and knowledge. Transforming the Web into a "knowledge Web" suddenly put knowledge acquisition and knowledge representation at the center of an extremely interesting and powerful topic: Given the amount of available online infor-

mation we already have achieved, this Knowledge (or Semantic) Web will be extremely useful and powerful. Imagine a Web that contains large bodies of the overall human knowledge and trillions of specialized reasoning services using these bodies of knowledge. Compared to the potential of the Semantic Web, the original AI vision seems small and old-fashioned, like an idea of the 19th century. Instead of trying to rebuild some aspects of a human brain, we are going to build a brain of and for humankind.

In this issue

The work and projects described in this special issue provide initial steps into such a direction. We start with Michel Klein's tutorial, which introduces the current language standards of the Semantic Web: XML, XMLS, RDF, and RDFS.

James Hendler—who has already helped us all by successfully initiating and running a large DARPA-funded initiative on the Semantic Web—reveals his vision of the Semantic Web. On the basis of a standard ontology language, he sees software agents populating the Semantic Web, providing intelligent services to their human users. In "OIL: An Ontology Infrastructure for the Semantic Web," Dieter Fensel, Ian Horrocks, Frank van Harmelen, Deborah L. McGuinness, and Peter F. Patel-Schneider propose such an ontology standard language. OIL and DAML+OIL are the basis of a semantic working group of the W3C that should soon develop a standardization approach. Sheila A. McIlraith, Tran Cao Son, and Honglei Zeng, in "Semantic Web Services," and Jeff Heflin and James Hendler, in "A Portrait of the Semantic Web in Action," describe intelligent services on top of such services, based on query and reasoning support for the Semantic Web.

T h e   A u t h o r s

Dieter Fensel is an associate professor at the Division of Mathematics and Computer Science, Vrije Universiteit, Amsterdam, and a new department editor for Trends & Controversies. After studying mathematics, sociology, and computer science in Berlin, he joined the Institute AIFB at the University of Karlsruhe. His major subject was knowledge engineering, and his PhD thesis was on formal specification language for knowledge-based systems. Currently, his focus is on using ontologies to mediate access to heterogeneous knowledge sources and to apply them in knowledge management and e-commerce. Contact him at the Division of Mathematics and Computer Science, Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, Netherlands; dieter@cs.vu.nl; www.cs.vu.nl/~dieter.
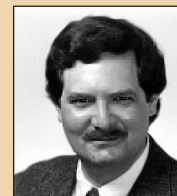
Mark A. Musen is an associate professor of medicine (medical informatics) and computer science at Stanford University and is head of the Stanford Medical Informatics laboratory. He conducts research related to knowledge acquisition for intelligent systems, knowledge-system architecture, and medical-decision support. He has directed the Protégé project since its inception in 1986, emphasizing the use of explicit ontologies and reusable problem-solving methods to build robust knowledge-based systems. He has an MD from Brown University and a PhD from Stanford. Contact him at Stanford Medical Informatics, 251 Campus Dr., Stanford Univ., Stanford, CA 94305; musen@smi.stanford.edu; www.smi.stanford.edu/people/musen.