

JUST A MATTER OF SEMANTICS ...

While others might still occupy themselves with Web 2.0, visionary developers are already working on the next Internet generation. **WEB 3.0** is becoming intelligent—a kind of Google with brains.



By Manfred Buchner

WHEN FRANÇOIS BRY saw a computer mouse for the first time in 1974, the graduate fresh out of high school was deeply impressed: “Many people regarded it as screwball idea. However, I regarded this small thing as a visionary development.” Nowadays, Bry—now a professor of IT at the University of Munich—is dealing with a subject that in a few years time could become as commonplace as the mouse is today: the Semantic Web.

While the Internet community worships the Golden Calf called Web 2.0, and discovering the joys of its own media production at Wikipedia, in blogs and at video-swap

portals such as YouTube, visionaries are already seeing a completely new Internet emerging: Web 3.0.

THE SEMANTIC WEB—which could be freely translated as “intelligent Web”—is an idea dreamed up by Tim Berners-Lee, the inventor of the World Wide Web. Even before the great Internet gold rush, the computer scientist, who was knighted by the Queen of England, had anticipated the dangers of the approaching data flood and the threat of information chaos. His rescue plan: He wants to enable computers to not only read data but also to understand it. As

assistants with brains, they are supposed to utilize the Internet’s mines of information substantially better than before, thus fulfilling Berners-Lee’s vision. According to that dream, “the Web is a place where the emotional capabilities of people coexist with the logical thinking of a computer in an ideal, powerful combination.”

The road to that place is a long one, however, as proven by entries into search engines that produce endless lists of alleged hits. Entering the words “When does the next plane from Berlin to Munich depart?” into Google (without quotation marks) will result in a list of 95,600 pages—but no answer. According to an empirical study carried out by the University of Düsseldorf, more than two-thirds of the results shown in Google are worthless, because they lack relevance. Furthermore, many information sources remain unused, because Google and its counterparts mostly do not file documents in databases or data in tables at all. These are a part of the “Invisible Web,” the extent of which is significantly larger than the visible portion of the global knowledge labyrinth.

THIS OFTEN FRUITLESS searching costs time and money. Bill Gates complained about the resulting losses in productivity in a speech to chief executives from around the world at the CEO Summit 2006 in Redmond, Washington. According to estimates by his company, employees spend around a third of their working time searching for information, which costs on average \$18,000 a year per employee. With the Semantic Web, relief is in sight. In his capacity as Director of the World Wide Web Consortium (W3C), Tim Berners-Lee has defined technical standards and procedures for the required extension of the Web. This involves two specific goals: First to facilitate the exchange of data and to make all databases mechanically retrievable, and second, to organize the correlation of data to people and to the context. “Google with brains” is how François Bry, professor at the Institute for Informatics at

the University of Munich, refers to the capabilities of the new Internet, which surfers can already become acquainted with sporadically today. At the online mail-order company Amazon, for example. Each and every page view of visitors is recorded there. This enables the generation of personal recommendations, such as book titles. Other examples are Google Maps and Windows Live Local. Whoever types in an address here is automatically shown a city map with street details—even in 3-D, upon request. This is made possible by linking text and graphics files.

YET THAT IS JUST the beginning. By combining familiar methods such as business intelligence, expert systems or data mining with new semantic processes, the digital behavior of users—including their environment—can be systematically investigated in the future. For this purpose, all data inputs and outputs, used documents and accessed Web sites and e-mails will be evaluated. The PC automatically generates personalized, time- and situation-specific

EU PROJECT

The EU research project **REVERSE (REasoning on the WEb with Rules and SEmantics)** is designed to bring intelligence to the Internet. Some examples from the list of topics:

- Rule Markup Languages: harmonized markup languages with tools to categorize terms
- Composition and Typing: methods and rules for software interoperability on the Internet
- Reasoning-Aware Querying: logical conclusions for Web inquiries
- Decision aids for event, time and geographical data
- Personalized systems for user-adaptive information and teaching systems

information and promotion suggestions from the user profiles produced. This is provided above all by the Resource Description Framework (RDF), the key technology of the Semantic Web. This programming language functions in a similar way to the syntax model “subject–predicate–object” in grammar lessons at school. With the assistance of so-called metadata, information categories are formed, through which computers can identify, for example, the author, the draft date, the title or other components of a document. Matching real terms to metadata produces machine-readable data encyclopedias—the computer’s memory.

MUCH OF THIS is still at the research level, however. Just one look at the language world of the Semantic Web, which is teeming with exotic terms such as “annotation,” “ontology,” “reasoning” and “taxonomies,” proves it. The subject of the EU project Reverse (see box), for instance, is developing of semantic methods, based on standards set by the W3C committees. Around 100 scientists from 27 European research organizations are involved in the project, which comes at a cost of €5.5 million and is coordinated by Professor François Bry together with colleagues from Sweden’s Linköping University. Gartner experts estimate that it will take between five and 10 years before the Intelligent Web is available in any major form. But patience will pay off, for the benefits are enormous. According to a current study by the “Fraunhofer-Institut für Rechnerarchitektur und Softwaretechnik” (Fraunhofer Institute for Computer Architecture and Software Technology), companies can anticipate efficiency gains of up to 50 percent in business processes. According to Web visionary Tim Berners-Lee, there is a simple reason why progress is not made more rapidly: the conservative attitude of users: “It is difficult to convince them of the new technology. After all, people also found it difficult to comprehend the World Wide Web before it existed.” ||