



INFORMATIONWEEK.COM

The Next Web

The Semantic Web, the next step in the Web's evolution, promises even more dramatic changes

By David M. Ewalt, InformationWeek
Oct 14, 2002 (12:00 AM)

URL: <http://www.informationweek.com/story/IWK20021010S0016>

Burst bubble and a few thousand failed business plans aside, it's clear that the World Wide Web really has changed the way we do business. But there's something missing from the equation. We've got more data than ever, with everything from sales records to engineering specs available online, but finding the needle we need in the haystack of the Web isn't getting easier.

The trick may be to build a new Web, one based on the meanings and syntax of our language but invisible to the humans who speak it. This world, dubbed the Semantic Web by the researchers and academics who are planning it, is one meant for machines, not people. The goal is to allow computers not just to process and move our words and data but to understand them.

"The last seven years, the Web has been very focused on giving value to human eyeballs," says Prabhakar Raghavan, chief technology officer at knowledge-management software company Verity Inc. "Over the next seven years, the interesting eyeballs will belong to computers."

To be accurate, the Semantic Web is an extension of today's Web rather than an entirely new one. Championed by the father of the World Wide Web, Tim Berners-Lee, it will probably take the form of specialized tags inserted inside HTML documents that don't just identify the page but also help computers understand what it's about. "It's annotating any form of content, any type of data, with additional semantic relationships that are machine processible, so a machine can infer a little more about what's contained in that content," says Alexander Linden, director of emerging trends and technology at market-analysis firm Gartner.

Are you having flashbacks? Weren't XML and other Web services supposed to let computers talk directly to one another? That's true, but the Semantic Web is being built to push that concept much further. XML by itself lets machines identify information within very narrowly defined boundaries -- for example, go to Web sites and download something explicitly labeled as "<price>." But a Semantic Web service knows that "price" can be the same thing as "cost," that it can be measured in "dollars," and take the form "\$X,XXX.XX." Therefore, it can get information from a site where the XML tag says "price" as well as one where it says "cost," or one with no identification at all, just some data that looks like "\$5,000.00."

That's because at the heart of the Semantic Web are dictionaries that draw direct relationships between terms. The Semantic Web knows magazines are also called publications, people work for a company, and so on. Any program running a semantic search would see the tags in a document

and access a dictionary to define them and figure out relationships before proceeding -- like when you type www.informationweek.com in your browser, and the program accesses a dictionary, or name server, to find out what computer actually hosts that site.

The ability to find information more easily won't just make life easier but should provide distinct returns for businesses. "If search gets better, people are going to be able to find you easier, and your employees are going to be more productive," says Alden Hart, CTO of the Adrenaline Group, a technology consulting firm.

Consider the hypothetical case of an automobile manufacturer that needs to find the perfect part for a new car it's developing. The carmaker could instruct a semantic search tool to find nuts that are lightweight, very resistant to heat, of a certain size, cost less than a penny, and can be delivered at the same time each week. By accessing the relatable semantic tags in product catalogs from a variety of suppliers, a program could compare, contrast, and evaluate the options, presenting the carmaker with a list of nuts that best meet its criteria. That wouldn't be possible without semantic tags, says Eric Miller, Semantic Web activity lead for the World Wide Web Consortium. "Not everyone says a 'cog' is a 'cog,'" he says.



If the Semantic Web lives up to its promise, it will mark a major change in the way businesses think, says Enigmatec chief technology officer Johnson-Watt.

The Semantic Web might also make monitoring companies' financial performance and dealings more automated. By placing semantic tags in Securities and Exchange Commission filings and other public documents, regulators or investors could create programs to automatically alert them to red flags such as insider stock selling, says Duncan Johnson-Watt, CTO of business-intelligence software maker Enigmatec Corp. Law-enforcement agencies could use similar technology to track funds while looking for terrorist or other illegal activity. "It's very hard to spot patterns that might be of interest," Johnson-Watt says.

The Semantic Web is the brainchild of Berners-Lee, who has never been satisfied with the way the Web works, so he's pushing for a smarter version. It's largely an academic effort today, led by the World Wide Web Consortium, a nonprofit standards organization founded in 1994. The group is supported mostly by MIT, the French research institute INRIA, the Defense Advanced Research Projects Agency, the European Union, and Japan's Keio University. Its 446 members include executives

from companies one would expect, such as Hewlett-Packard, IBM, and Microsoft, but their roles are largely to serve on advisory boards. And there's widely varying opinions on just how quickly some form of the Semantic Web will offer commercial potential.

A few vendors are already working on what can be considered early-generation semantic technologies. Verity's knowledge-management software, capable of automatically analyzing and tagging documents, can identify patterns in customer E-mails and service records, helping catch potential problems before they get worse, CTO Raghavan says. "The system is constantly extracting names, like 'Ford' and 'Firestone,' and then a mining application can look through that and note an unusual concurrence," he says. "We're taking a bag of words, extracting structure, and building on it." By adding that structure to unstructured data, businesses will make it easier for employees to find critical data, increasing efficiency and allowing them to make more informed decisions, he says.



Verity's knowledge-management software tags E-mail patterns and catches potential problems, Raghavan

CAVE

IBM is making significant investments in research, even creating the

Institute of Search and Text Analysis in California. It will focus on issues critical to the development of the Semantic Web, says Nelson Mattos, director of information integration at IBM Research. "This really is about the next generation of Web technology," he says, "and IBM is going to play a key role in defining what it's all going to look like."

Others say the Semantic Web isn't worth spending money on until the standard develops further. Researchers at Microsoft consider the effort promising, but they aren't doing any research projects that address the Semantic Web. "We'll keep an eye open, and if it begins to pan out and have some successes, we'll get involved," says Jack Breese, a director at Microsoft Research. The same wait-and-see attitude is espoused at the star of Web searching, Google Inc., which has no plans to link its efforts to semantics.

Others contend that businesses already can get part of what the Semantic Web promises -- processing unstructured data -- through other means. Autonomy Corp. makes software that uses pattern-matching algorithms to scan text, identifying key ideas based on the placement or frequency of words that are associated with certain concepts. "The Semantic Web is an interesting evolutionary step, but it's dealing with a problem that we've been addressing for years now -- the increasing amount of unstructured data within businesses," says Ron Kolb, director of technology strategy.

Faculty and students at the R.H. Smith School of Business at the University of Maryland are in the final stages of testing an integrated portal using Autonomy's software. Students and faculty identify a number of concepts and ideas related to their course of study. Next, automated agents search data sources on the Web, crawling through legal news, filings, and so on, reading each document and quickly summarizing its content. Should a document seem relevant to someone's research, the software will make it available to that person through the custom portal. Though the school is just finishing testing, the portal has already proven beneficial. "We've seen real improvements in the quality of research," CIO Sandor Boyson says. "It's increased our ability to provide support to our faculty and be at the cutting edge in terms of research." Though he's experimenting with emerging portal technology, Boyson doesn't expect much from the Semantic Web soon. "To talk about this sort of gigantic interoperability is premature," he says.

Clearly, companies could use a way to more efficiently structure data and probe information anywhere it's stored. But tagging unstructured data is only the first step. The real power comes afterward, if machines can look at that data and understand its meaning and context in the world. That will require agents able to search data in different sources using such an understanding, so that ZIP codes tagged with semantic markers are identifiable to agents regardless of the format of the underlying data. Then data stores around a company can be treated as one homogeneous whole. "You can think of [semantics] as a sort of thin veil over your existing data to treat it more like a relational model," the World Wide Web Consortium's Miller says.

Of course, there's a reason that rather smart companies such as Microsoft are playing wait-and-see on the Semantic Web. It's far from ready for wide use, and significant obstacles need to be hurdled before anyone can see the technology's benefits. Like any computing environment, in order for the Semantic Web to succeed, a critical mass of users and vendors need to agree on certain standards and protocols for how it all will work. Most technology managers are familiar with XML, a tagging language that defines data in Web pages and documents. It will be the first step toward making the Semantic Web work, acting as the underlying standard for writing tags. But don't think that because XML is involved, the Semantic Web is just a fancy name for Web services. It's more of a way for businesses to continue leveraging and benefiting from XML, says IBM's Mattos. Web services will remain a piece of the larger Semantic Web but not the whole, he says (see story, below). Beyond XML, developers must agree on a common vocabulary or framework

to define different semantic concepts.

There's a new general-purpose language, called Resource Description Framework, being hashed out by a number of organizations, including the World Wide Web Consortium. That's one reason why commercial use of the Semantic Web is years away. "It's such a hard thing to get three people in a room to agree on a topic, so to try to get a whole profession to agree is a tough nut to crack," says James Lester, chief scientist at LiveWire Logic Inc., which uses linguistic software for call-center automation. "There's going to be competing standards until one wins, and there's going to be some teeth knocked out in the process."

But if the true Semantic Web is years off, why should technology managers pay attention? Mattos says it's a matter of being ready for its emergence as a critical piece of businesses' infrastructure. Enigmatec's Johnson-Watt says IT will be expected to identify opportunities it creates. IT departments "are lightning rods for the business. And if the Semantic Web is delivered, it will mark a major change in the way businesses think," he says.

The World Wide Web Consortium's Miller says there's a key role for business-technology executives to play as the standards develop, particularly in trying early elements to see if the consortium is on the right track in developing useful business tools. "Take them for a test drive, give feedback, help explain where you ran into problems," Miller says. "It sounds a little altruistic, but we believe all ships are going to rise on this, and the sooner they rise, the more financially beneficial it will be for all." The Semantic Web would also change the approach businesses take to data in that its value would be defined less by the program it's held in. "Try to realize that your data is far more important than the application that accesses it," he says.

The Semantic Web is one of those business-changing ideas that leaders need to start getting their minds around today, LiveWire Logic's Lester says. "It's hard to imagine, in much the same way that 10 years ago it would have been hard for us to imagine the impact that the Web has had on commerce," he says. "But this is going to have a similar impact."

Photo of Johnson-Watt by Jonathan Olley/Corbis Saba.

Photo of Raghavan by Olivier Laude.



Copyright ©2001 [CMP Media LLC](#)