



Get more return from your  
customer communication.

**EMC<sup>2</sup>**  
where information lives™

Read the 15-Minute Guide. >>

## Semantics in Practice

By Nancy Davis Kho - November 2010 Issue, Posted Nov 12, 2010

SHARE | Print Version | Page 1 of 3 next >>

The definition of the word "semantics" sounds straightforward enough: According to Merriam-Webster, it is the study of meanings. However, if you ask a technologist, an editor, or an advertising rep, you may come up with widely differing interpretations. They might say it's a practical application of artificial intelligence; a way to streamline back-office operations and make websites stickier; or a means of contextualizing advertisements for higher efficacy.

None of them is wrong.

Ironically, it is the very breadth, depth, and range of possibility inherent in semantic technology that can prevent content companies from experimenting with it, though it may be one of the most useful commercial innovations of the past decade. The murkiness of the word itself, not to mention the standards, acronyms, and jargon that can dominate the discussion of semantics, only adds to the confusion. David Siegel, author of *Pull: The Power of the Semantic Web to Transform Your Business* (Portfolio, 2009), says, "Semantics was a bad word from the get-go. A better word would have been unambiguous."

As both understanding of the possibilities inherent in the semantic web and tools to harness it have matured, semantic technology has finally gained a foothold in practical business applications, in areas from search to back-office processing to advertising. Siegel believes that enterprise adoption has reached a critical growth point. "I'd say we're solidly 1% of the way there, in terms of adoption for the enterprise," says Siegel. "And that's very good."

**Sponsor text:**

Read EMC's 15-minute guide to Enterprise Content Management today. Click Here to Download the White Paper.

Particularly for the content industry, semantic technology offers a compelling story. At the June 2010 SemTech conference in San Francisco, Bob DuCharme of TopQuadrant, a provider of semantic web technologies, pointed out that publishing and semantics share much in common. "The publishing industry has lots of real data and metadata already, and they have experience in developing vocabularies," DuCharme noted.

Tom Tague, vice president of platform strategies at ThomsonReuters who leads the OpenCalais initiative, a web service that automatically creates rich semantic metadata, agrees that within the content world, semantics has found its beachhead. "The future is here," Tague says, citing a favorite aphorism. "It's just not evenly distributed yet."

### A Brief History

Semantics first made a big splash with a 2001 article in *Scientific American* by Tim Berners-Lee, James Hendler, and Ora Lassila called "The Semantic Web." The article laid out a future on the web in which common data formats and definitions would enable people (and machines) to share information at a much more granular level, moving through disparate databases to find interrelated information based on a shared "aboutness." Instead of the document in which it resided, the lowest common denominator on the semantic web envisioned by the authors was the information itself.

To provide a simple example, it's the difference between typing "hotel Berlin" into a search engine from 10 years ago, and then again in 2010. Back in 2000, the searcher might have received long lists of links to various hotels (and one movie) called "Hotel Berlin," and each link would have to be opened and read individually. These days, a search engine utilizing semantic technologies, such as Bing, brings back a page with a map showing locations of various hotels, along with links to Berlin attractions and Berlin tours, and lists of hotels presorted into price categories. By understanding the meaning and intent behind the search term, semantics makes it easier to find the correct information quickly.

To achieve that vision, producers of content needed to add metadata to information resources. It was critical that data standards and ontologies—that is, formalized vocabularies of terms—were agreed upon so that anyone putting data onto the web could agree that "revenue" in one database was the same as "revenue" in another. There are a number of techniques for doing this, and the applicability of each depends on the complexity and flexibility required by a specific application.

For instance, the resource description framework (RDF) publishing standard is an eXtensible Markup Language (XML)-based standard for describing resources that exist on the web, intranets, and extranets, including metadata such as title, author, modification date of a webpage, copyright, and licensing information about a web document.

Underlying RDF are "triples" that match a subject with a predicate and an object, each of which describe a particular aspect of the subject. One example of a triple might be "chicken" "is" "animal." Another is "chicken" "is" "bird," and still another is "chicken" "has" "feathers." The number of triples associated with a particular subject is nearly infinite, which provides the flexibility that is touted as a major advantage of semantics. As new categories emerge that may relate to "chicken," such as "bird flu," a new triple can be easily added.

Other data organization techniques, such as Web Ontology Language (OWL), Simple Knowledge Organization System (SKOS), and Rule Interchange Format (RIF), can be used alone or in combination to describe information parameters. Purpose-driven standards have also emerged to solve specific vertical information needs on the semantic web, such as eXtensible Business Reporting Language (XBRL), designed to facilitate the exchange of business and financial information on the web. Siegel says, "XBRL is a winning example of how to make things interoperable and shareable. It's not overly flexible, but it's well specified."

SHARE | Print Version | Page 1 of 3 next >>

Innovate

Subscribe to our complimentary e-newsletters at:  
[www.dialog.com/signup](http://www.dialog.com/signup)

Authoritative answers enriched by ProQuest

[www.dialog.com](http://www.dialog.com) Dialog

Search EContent:

 GO

Infocentral  
POWERED BY PROQUEST

Enterprise Search  
Information Today, Inc. center

EContent  
Solutions  
DIRECTORY

EMC<sup>2</sup>  
where information lives™

Capture,  
connect,  
and go

EMC<sup>2</sup> | captiva

Transform your  
business →

- Research Centers
- Content Commerce
- Content Creation & Digital Publishing
- Content Delivery
- Content Distribution
- Content Integration
- Content Management
- Content Security
- Digital Asset Management
- Intranets and Portals
- KM & Collaboration
- Mobile & Wireless Content
- News/Finance/Business
- Online Community
- Premium Content Services
- Rich Media
- Sci-Tech/Medical
- Search Technology
- Taxonomy
- Web Services

- Columns
- After Thought
- Agile Minds
- Behind the Firewall
- Edit This
- Eureka
- Flexing Your Content
- Follow the Money
- Guest Columns
- I Column Like I CM
- Info Insider
- Info Pro
- Media Redux
- Screen Play
- Technology Watch

- Departments
- Case Studies
- eReader
- Faces of EContent
- Peopleware

- In Focus
- EContent 100
- EContent 100 Videos
- Past Issues

- Services
- About EContent
- Advertising
- Subscribe to EContent Magazine
- EContent Xtra Newsletters
- Solutions Directory

RSS Feeds

**KMWorld**  
Nov. 16-18  
Renaissance  
Washington, DC Hotel

### SPONSORED LINKS

- Read the 15 minute guide to Enterprise Content Management
- Read the 15-Minute Guide to Best Practices in Correspondence Management
- ITIResearch.com - A collection of market research and reports for executive management and business & IT professionals
- Need instructional help with Online Video? Check out our new resource - OnlineVideo.net
- Try the Wall Street Journal Professional free for two weeks.