

[Tech Notes are short articles discussing library-related technology.]

WiLS [Tech Note:] The Semantic Web

The age of the “semantic web” is arriving, in dribs and drabs. First envisioned by Tim Berners-Lee some ten years ago, there are now a number of web sites that implement semantic content to some degree. It has also gotten a big boost by the announcement that Yahoo! Search will soon be able to search for semantic web elements within web pages [1]. Can Google be far behind?

Currently, web pages convey their content to people (and search engine web crawlers) in a non-uniform and scattered way. The semantic web, as its name suggests, brings content with *meaning* to the worldwide web. Conventional web searches can only return matches on the terms the user types in. Semantic web searching takes this further by capturing the relationships between concepts of search terms. For example, a web page might contain a telephone directory of people working in a library. It could be in alphabetical order and include title (Youth Services Librarian) or area of specialization (East Asian Literature). This web content and its meaning are easily absorbed by people viewing it, given that it is a well-constructed directory. People are a lot smarter than web crawlers. In the case of a public library site, the fact that John is a Youth Services Librarian could be encoded in a machine readable format. Another encoding, either on the library’s staff web page or, more likely, somewhere else entirely, might be a semantic web entry that conveys the fact that sometimes a Youth Services Librarian is called a Children’s Librarian. When someone searches with a sufficiently powerful, semantically-enabled search engine and types in the library’s name and “children’s librarian,” the library’s information for John would be displayed.

Semantic web content can be embedded in any web page, and includes a set of formal statements in a structured format describing relationships and attributes of the objects and concepts. This content is almost always invisible to people browsing the page; it is meant to be ingested by web crawlers or intelligent browsing software. It conveys meaning to this software in a more recognizable way than the text of a typical web page.

The format most often used for semantic web content is called RDF, for resource description framework. This is the same format used to represent Dublin Core, a set of metadata elements used to describe digital and other information objects. RDF is particularly good at representing metadata, data that describes digital or other objects. But it can also carry information about the *relationships* between objects and concepts. To use an example from a particularly good *Scientific American* article about the semantic web [2], it is possible to encode the following statements in RDF:

Flipper is a Dolphin
Dolphin is a subclass of Mammal

A piece of software called an inference engine can then make the logical conclusion, also representable in RDF:

Flipper is a mammal.

Admittedly this is an elementary example, but it tantalizingly hints that much more complicated relationships can be inferred given sufficient data.

In addition to RDF, semantic intelligence can also be obtained from various “microformats,” simple sets of information conveyed according to established, open data standards. The Yahoo! Search effort includes support of a number of microformats. That’s important, because there are many sites that encode data using microformats but not RDF.

What about practical applications of the semantic web? Even my simple and simplistic examples of the library staff search and Flipper the mammal should suggest the potential power of the semantic web. And so far it has mainly been potential. As I mentioned, Yahoo! Search will be incorporating semantic web elements into its search engine, enabling much richer search possibilities. But until now the semantic web has faced the classic chicken-and-egg problem: why embed semantic content in web pages when there are no tools to harvest the data? Why build tools to harvest data when there is no data available? Although Yahoo! may be breaking through this problem, there are few other easily-found implementation examples.

There are two “case studies” in the *Scientific American* article I mentioned earlier[3]. One discusses drug discovery using semantic web principles, while the other is about capturing and analyzing emergency room public health data. They are quite interesting but are unavailable to the public.

The semantic web will enrich searching by both the public and information professionals. And it’s coming soon to your Internet-connected computer. Really.

Takeaways From This Article

- The potential of the semantic web will soon translate into reality.
- Yahoo! Search’s adoption of semantic web searching will provide a richer return of information to its users.
- RDF (Resource Description Framework) is a format for representing semantic information.
- Microformats can also represent semantics.
- Watch for other search engines to weigh in with their own version of semantic search.

[1] “The Yahoo! Search Open Ecosystem,” Posting to Yahoo! Search Blog, March 13, 2008, viewable at <http://www.ysearchblog.com/archives/000527.html> .

[2] “The Semantic Web in Action,” *Scientific American*, December 2007, by Lee Feigenbaum *et al.*, pp. 93-97.

[3] *Ibid.*

A couple of other online articles may be of interest:

“Yahoo embraces the Semantic Web?,” Posting to Semantic Web Blog, March 13, 2008, viewable at <http://blogs.zdnet.com/semantic-web/?p=114> .

“Looking for a dominant Semantic Web search engine,” Posting to Semantic Web Blog, March 17, 2008, viewable at <http://blogs.zdnet.com/semantic-web/?p=112> .

The Semantic Web Blog, viewable at <http://blogs.zdnet.com/semantic-web/>, is a good source for Semantic Web news.

A *Scientific American* article from May, 2001 detailing a rosy feature for the then-nascent semantic web is available online at <http://www.sciam.com/article.cfm?id=the-semantic-web&catID=2> (by Tim Berners-Lee, James Hendler and Ora Lassila)

—Tom Zillner