

Department Editor: Niek Wijngaards

The Semantic Web: What, Why, How, and When

Frank van Harmelen , *Vrije Universiteit Amsterdam*



Notwithstanding the World Wide Web's enormous success and impact, it currently has a significant limitation: it's accessible only for humans. Machines ship the information across the network and paint it on our screens, but they don't help us at all to select the information, interpret it, compare different information sources, draw conclusions from the information, and act on it.

It's easy to see why this is so: computers understand only the Web pages' structure and layout and have no access to their intended meaning. The Semantic Web aims to enrich the existing Web with a layer of machine-interpretable metadata so that a computer program can draw conclusions (that is, derive new information) predictably (that is, all programs that receive this information draw the same conclusions from it).

WHY DO WE NEED THE SEMANTIC WEB?

The Semantic Web has two main motivators. The first is *data integration*, which is a significant bottleneck in many IT applications. Current solutions to this problem are mostly ad hoc: each time, a specific mapping is made between the data models (schemas) of the data sources involved. If the data sources' semantics were described in a machine-interpretable way, the mappings could be constructed at least semiautomatically.

The second motivator is *more intelligent support for end users*. If computer programs can infer consequences of information on the Web, they can give better support in finding information, selecting information sources, personalizing information, combining information from different sources, and so on.

HOW WILL WE ACHIEVE THE SEMANTIC WEB?

To achieve these goals, we must meet the challenges of providing

- ◆ A syntax for representing metadata
- ◆ Vocabularies for expressing the metadata
- ◆ Metadata for lots of Web pages

We're well under way to meeting the first challenge. The W3C (World Wide Web Consortium) has defined such open standards for metadata syntax as RDF (Resource Description Framework) and OWL (Web Ontology Language), and support for these standards from both industry and academia is rapidly increasing.

Also, professional groups increasingly are building metadata vocabularies (or *ontologies*). Large ontologies exist for medical terminology, genomics, geographic information systems, and law, just to mention a few. These terminologies are typically hand built, but systems are rapidly getting better at learning them semiautomatically from large volumes of text. An important open problem is that of automatically finding translations between different terminologies that have been designed for the same domain (the "ontology mapping problem").

For obvious reasons, we'll have to rely on computers to obtain large amounts of metadata on the Web. We can't expect general-purpose, widely applicable solutions. After all, that would beg the question of machine-understandable semantics that motivates the Semantic Web in the first place. Instead, we'll have to apply many special-purpose techniques: automatic concept extraction from natural language in limited domains, exploiting the schemas from the data pages that are used to generate many Web pages, mining meaningful terms in URLs, and so on.

WHEN WILL THE SEMANTIC WEB BE A REALITY?

The only thing that's certain about the future is its unpredictability, and this is doubly true of anything Web related. Nevertheless, some patterns are emerging:

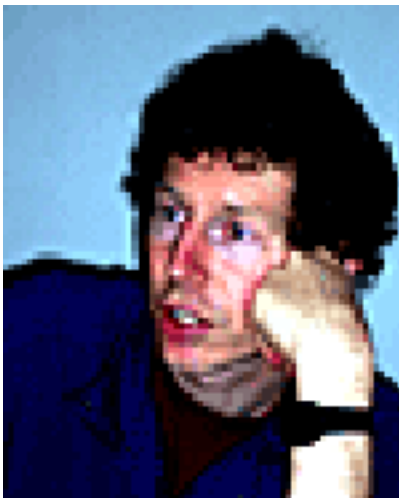
- ◆ The use of RDF (and to a lesser extent OWL) in data integration will likely be the first application area. Large companies such as Boeing and Verizon are investing in this technology.
- ◆ Rather than a World Wide Semantic Web, we'll first see "islands of semantic interoperability" in large organizations, professional communities, market segments, and so on.
- ◆ Most metadata will be semantically rather lightweight. We won't have a full-blown heavy formalization of large bodies of knowledge in the style of knowledge representation. Rather, we'll have lightweight hierarchies of terms, in the style of thesauri.
- ◆ Web Services interoperability, business-to-business e-commerce, and improved precision for search engines will likely be important application drivers for the Semantic Web. Business-to-consumer e-commerce, personalization, and autonomous information agents will be much further in the future.

WHERE CAN I FIND OUT MORE?

For RDF and OWL overviews, visit the W3C Web site, <http://w3.org>.

Somewhat outdated, but still very good, is Sean Palmer's "The Semantic Web: An Introduction," at <http://infomesh.net/2001/swintro>.

Grigoris Antoniou and I have written *A Semantic Web Primer*, which we consider to be the first real textbook on the Semantic Web. It contains numerous additional pointers. MIT Press will publish it this summer. For more information, see www.cs.vu.nl/~frankh.



Frank van Harmelen is a professor of knowledge representation and reasoning in the Artificial Intelligence Department at the Vrije Universiteit Amsterdam. Contact him at frank.van.harmelen@cs.vu.nl.

Related Links

Purchase or Digital Library members login:

- "E-Science: The Grid and the Semantic Web," *IEEE Intelligent Systems* (Jan./Feb. 2004) <http://csdl.computer.org/comp/mags/ex/2004/01/x1065abs.htm>
- "The Zen of the Web," *IEEE Internet Computing* (Sept./Oct. 2003) <http://csdl.computer.org/comp/mags/ic/2003/05/w5030.pdf> **FREE**