**Editor: James Hendler**
University of Maryland
hendler@cs.umd.edu

# Research Challenges and Perspectives of the Semantic Web

**Jérôme Euzenat,** *INRIA Rhône-Alpes*

Accessing documents and services on today's Web requires human intelligence. The interface to these documents and services is the Web page, written in natural language, which humans must understand and act upon.

The Semantic Web will augment the current Web with formalized knowledge and data that computers can process. In the future, some services will mix human-readable and structured data so that both humans and computers can use them. Others will support formalized knowledge that only machines will use. This will enable

- Computers that assist human users in tasks. Computers will be able to "understand" data in ways they cannot today.
- A more open market for information processing and computer services. This will result in new applications and services created from combinations of existing services.

To identify directions and goals for the Semantic Web, and its challenges and opportunities, 20 European and US researchers met at a National Science Foundation-European Union Strategic Workshop from 3 to 5 October 2001, in Sophia-Antipolis, France. The researchers came from the areas of knowledge representation and engineering, databases, the World Wide Web, and human–machine communication. Their ultimate goal was to determine the major research needed and major breakthroughs expected in the future.

The workshop's first two days consisted of participant presentations covering four topics: languages, resources and infrastructure, clients and human interface, and Semantic Web application areas. After each session, the participants identified topics for further discussion. On the third day, they split into four working groups to discuss future research perspectives and agendas. Each group covered a different topic: languages and inferences, infrastructure, human-related issues, or ontologies. For extensive information about the workshop, including the full report, visit www.ercim.org/EU-NSF/semweb.html. The following sections summarize the workshop's context and recommendations.

## Formalizing knowledge

We can view the Semantic Web as an infrastructure for supplying the Web with formalized knowledge in addition to its informal content. No consensus exists on how far the formalization should go; the possibilities range from metadata schemes (such as the Dublin Core metadata markers) to full-fledged logical representation languages. This infrastructure currently exists only for particular applications (for example, SHOE [Simple HTML Ontology Extensions] and Ontobroker) and is limited to a small part of the Web.

One of the challenges of Semantic Web development is designing a framework that can encompass all these different understandings of the Web. We can only attain the Semantic Web's full benefits when computers are able to relate information from various sources.

## Applications

Like the Web, the Semantic Web is not an application; it is an infrastructure on which many different applications (such as e-commerce) will develop. Characterizing the Semantic Web's killer application will be as hazardous as predicting the Web's killer application was 10 years ago. The use does not precede the technology, but it explodes when the technology is available.

Semantic Web applications seem to fall into two types:

- *Applications for organizations*. Examples include ontology-based marketplace development for business-to-business e-commerce and the bioinformatic knowledge grid, which seamlessly interconnects biological data and knowledge bases with available computing resources.
- *Applications for the masses*. One example is an intelligent personal assistant that gathers information, filters the relevant information, and composes it into a coherent picture for the user (for instance, a travel assistant).

More generally, corporate or personal knowledge management can exploit the Semantic Web, which will provide value to any semantically annotated resource by facilitating its retrieval when appropriate.

Three application areas in particular have attracted attention: e-commerce, knowledge management, and bioinformatics. Some of these areas could provide "seeding applica-


1094-7167/02/$17.00 © 2002 IEEE

tions"—that is, applications that are either testbeds for, or early adopters of, Semantic Web techniques. (In this regard, the bioinformatics community could be for the Semantic Web what the physics community has been for the Web.) The workshop report thus devotes a section to these scenarios.

## Who are involved?

The Semantic Web's development will involve many areas of computer science, including

- *Artificial intelligence:* reasoning mechanisms, knowledge representation languages, approximate and rough computing, and learning and resource discovery
- *The Web:* profiling, identification, and XML-based languages and technologies
- *Databases:* storage, fault tolerance, security, transactions, and query languages
- *Agents:* distributed computing, communication languages, and interaction and cooperation protocols
- *Theoretical computer science and (computational) logic:* languages, theorem proving, and semantics
- *Systems:* reliability, mobility, and (Web) security
- *Computational linguistics and pattern recognition:* knowledge acquisition from primary sources, lexical resource use for ontology development, pragmatics, and question-and-answering
- *Document engineering and digital libraries:* transformation, markup, and indexing
- *Human–computer interfaces:* computer-supported collaborative work, work factor evaluation, and communication studies
- *Social and human sciences:* ontology validation experiments and social informatics

Moreover, the Semantic Web must not be separated from other issues: personalization (and thus privacy concerns), mobility (reliability concerns), and publication (security concerns). These issues and topics are quite traditional, but the Semantic Web drives them to extremes; using semantically grounded languages makes computer actions more powerful and security and privacy threats more acute.

Nevertheless, for the Semantic Web to happen will require more than just technology. Economics and social matters are also important. A trade-off between these factors must be found that will lead to a value-adding, appealing, and easy-to-use Semantic Web. This is why we should encourage many applications on top of a solid infrastructure so that the fittest will raise the Semantic Web's utility.

## Requirements

The key requirement for the Semantic Web is interoperability. If machines are to exploit Web resources, they must be able to access and use them. Consequently, the resources must be open and understandable, not hidden in a proprietary system that publishes information only in a human- or tool-oriented format. They must be invokable and published in an open, structured, and rich format that lets the machines make the best of them.

Identifying the Semantic Web with a particular technology (search engines, knowledge representation, natural language processing, and so on) or language (XML, RDF, DAML+OIL, and so on) is certainly not accurate.

Realizing the Semantic Web will require several layers of development (see Figure 1). The infrastructure will allow for identifying, locating, and transforming resources robustly and safely. Languages are necessary for expressing the Semantic Web's content; these languages' semantics are sanctioned by inference engines. Resources such as ontologies, transformations, metadata, and databases must feed these two base layers. Applications that run on devices will exploit these resources. The layers correspond roughly to the working groups that were formed during the workshop.

## Recommendations

Concerning Semantic Web research and development, the workshop participants' recommendations fell into four broad categories.

First, *identification and localization* is important for Semantic Web reasoning, annotating, and computing. It amounts to agreeing on how to identify resources, compare or equate two identifiers, and localize Web resources for easier access. This involves research in languages, infrastructures, and ontologies. The semantics of representation languages must take identity into account, and the assumptions related to object identity must be explicit in ontologies and manageable by the infrastructure. The infrastructure must support the localization of and access to identified resources.

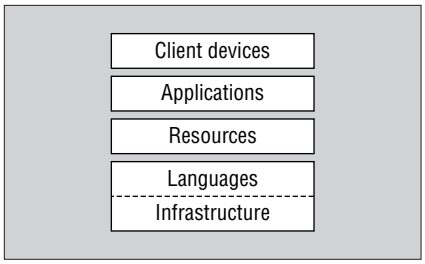


**Figure 1. A high-level layered view of the Semantic Web.**

Second, we must *assess relationships and reduce differences among semantic models* across languages and modeling styles. Heterogeneity is intrinsic to the Semantic Web; no language will be suitable for all purposes, no model will be applicable to all cases, and no ontology will cover the infinite potential applications. Because a semantic description of information and knowledge is available, we can address heterogeneity by

- Developing layered and modular representation languages for the Semantic Web
- Studying the impact of modeling assumptions on interoperability
- Providing a transformation infrastructure
- Articulating and composing services and transformations
- Supporting reuse and evolution through metrics for comparing models and distributed version support

Third, *tolerant and safe reasoning* is necessary, and we must characterize the accuracy of its results. A variety of reasoning methods will be necessary for different applications (from fetching to theorem proving), and the quality of their required results will vary (from "anything will do" to "100-percent certified"). This will involve

- Developing tolerant computing techniques to cope with messy metadata and the Web's open character
- Providing an infrastructure for implementing safe computing with proven properties
- Developing new computational models for trust, proofs, and rewards on the Web

Finally, *facilitating Semantic Web adoption* is critical. It will depend first on resource availability and then on the Semantic Web's ease of use. This can be supported by

- Foundational ontology development
- Well-crafted ontology libraries
- Text mining
- Ontology and metadata learning
- Growth model and acceptance factor studies

- Incidental-knowledge capture
- Consensus-building tools
- Unobtrusive support for collaboration

Concerning research funding, the participants made these recommendations:

- Support worldwide collaboration among researchers, because it allows consensus on the worldwide level that the Web requires (and not just the continental level). This will involve funding non-project-focused work, which is necessary for producing reports, surveys, and studies.
- Encourage both open-source development of high-quality components and nonprofit shelter organizations for software (such as the Apache Software Foundation).
- Support efforts to build Semantic Web seeding applications. Obviously, we first need a set of existing applications before we can improve them.
- Provide educational support (for example, teaching materials and starter kits).

Expectations are high for the Semantic Web, because information overload currently reduces the Web's usability. Also, the lack of interoperability among Web services is an obstacle to realizing the Web's promise. The Semantic Web's development and use could benefit both providers and users and result in increased economic, social, and cultural activity, as happened with the Web in the 1990s.

The Semantic Web will happen because we need it. It is necessary for improving the information infrastructure. It is necessary for commercial companies that want to interoperate better, either in business-to-business e-commerce or in worldwide enterprise. It is necessary for citizens who want better service from suppliers, service organizations, and government and more effective privacy protection.

The Semantic Web could happen now, in 10 years, or in a century. This depends mainly on what we expect from it. It is happening now, if you think of the many initiatives for marking up resources (syndication, open catalogs, and annotations). It will happen in several years, if you think of the possibility of agents realizing one of the ambitious scenarios presented in the workshop report. But, in the end, the Semantic Web will have fully succeeded when no one talks anymore of "the Semantic Web" but simply calls it "the Web." ■

## Acknowledgments

The European Union's Future Emergent Technology program and the National Science Foundation's Computer and Information Science and Engineering directorate held the workshop, which was organized by the European Research Consortium in Informatics and Mathematics.

**Jérôme Euzenat** is a research scientist at INRIA Rhône-Alpes. He leads the Exmo team investigating the exchange of formal knowledge mediated by computers. He is particularly interested in the relationships between representations including abstraction, granularity, versioning, and transforming representations. He holds a PhD and habilitation in computer science, both from Grenoble 1 University. Contact him at INRIA Rhône-Alpes, 655 avenue de l'Europe, Montbonnot Saint-Martin, 38334 Saint-Ismier, France; jerome.euzenat@inrialpes.fr.