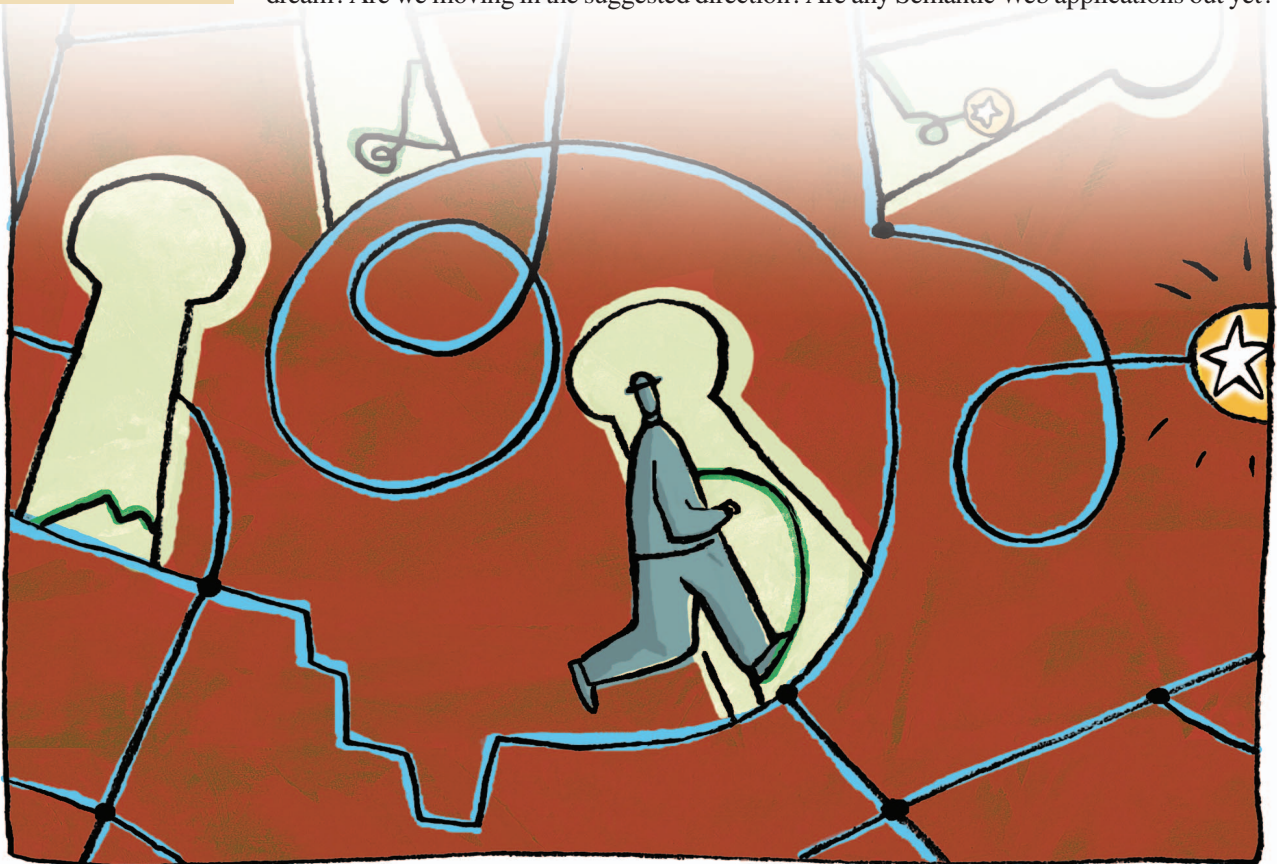


Semantic Web Challenge 2003

Michel Klein, *Vrije Universiteit Amsterdam*

Ubbo Visser, *University of Bremen*

For several years now, the Semantic Web's promise has been exciting computer science researchers. With the publication of futuristic scenarios in *Scientific American*, we all know what kind of dreams could eventually become true. However, what's the status nowadays? What kind of things can be realized with today's techniques? How far are we from realizing the dream? Are we moving in the suggested direction? Are any Semantic Web applications out yet?



Those questions were the direct reason for starting a Semantic Web Challenge. "Show us what kind of application you can create with today's Semantic Web techniques" was the challenge we proposed to both academic researchers and people from industry. Many research results were already available when we started the challenge. They involve language design, storage systems, and ontology modeling. However, there weren't yet attractive, integrated examples of what the Semantic Web could provide.

The challenge's aim is twofold. First, it should help researchers show our society what kind of applications the Semantic Web can provide. Second, the competition should be fun for researchers and stimulate their creativity. The first edition of the challenge was in the fall of 2003. The plan is to continue for at least five years. Each year will focus on a different specific goal reflecting Semantic Web developments at the time.

What is a Semantic Web application?

For such a contest, we first needed to define what a Semantic Web application actually is. After discussion with several experts, we formulated a set of minimal requirements. First, the application must use information sources that

- Are geographically distributed.
- Have diverse ownerships—that is, there's no control of evolution.
- Are heterogeneous (syntactically, structurally, and semantically).
- Contain real-world data—that is, the sources must be more than toy examples.

Second, the application must assume an open world; that is, it assumes that the information is never complete. Third, the application must use some formal description of the data's meaning.

Besides these minimal criteria, we determined several desirable qualities. The application should use data sources for other purposes or in another way than originally intended. It also should use the contents of multimedia documents. Users should be able to access the application in multiple languages or with devices other than a PC. The application should exploit both static and dynamic knowledge—for example, a combination of static ontologies and dynamic workflows. Finally, the application should be scalable (in terms of the amount of data used and

of distributed components working together).

The 2003 challenge

The specific challenge for 2003 was to apply Semantic Web techniques to build an online application that deduces, combines, and integrates information to help users perform tasks. A specific requirement was to incorporate at least two heterogeneous XML data or information sources that aren't controlled by the application's authors and that allow different viewpoints.

Procedure and evaluation

All submitted applications had to be described by a short paper and be accessible online. We evaluated them with the help of the Semantic Web Challenge Advisory Board, whose background varied from machine learning to knowledge representation. At least three reviewers evaluated each application according to the minimal criteria and desirable qualities. In addition, evaluators could give extra points for any notable feature that the criteria or qualities didn't cover (for example, easy access or a nice user interface).

Submissions

The 10 applications we received are diverse both geographically and organizationally. They come from different types of organizations (companies, research organizations, and universities) in six different countries (Japan, Taiwan, the US, the Netherlands, the UK, and Germany).

The submissions' application domains are also diverse: environmental data, scientific publications in medicine, earthquake data, Web logs, bioinformatics, news, and computer science research. Most applications aim to coherently present information from different sources to users. Some applications provide additional help for this in the form of a query interface.

The submissions use a vast amount of data that are surprisingly diverse. One example is the application from National Taiwan University's Department of Computer Science and Information Engineering. Their application integrates eight bioinformatics Web information sources to provide an integrated tool for single-nucleotide polymorphism analysis. These Web-based information sources are geographically distributed (in five countries, three continents) and have diverse ownership. They're heterogeneous in all aspects and contain real-world data or provide real-world services.

Another example is the application from the Jet Propulsion Laboratory at the California Institute of Technology. It integrates earth science knowledge from various sources, including available gazetteers, earthquake data from the US Geological Survey, US Central Intelligence Agency databases on countries, and geographic polygons.

The submissions also show different data acquisition strategies. Several applications use ready-made RSS (RDF Site Summary) feeds from Web logs and news sites, some crawl the Web and extract data from HTML pages, and others interpret existing databases. Also, some applications use email message archives and existing online RDF descriptions. Although the submissions use many different sources, they can incorporate only a minority of these sources automatically. Most sources require manually formulated translations.

The winners

The competition was close; ranking the applications was difficult. However, the jury agreed on these winners:

First place went to CS AKTive Space, from the University of Southampton. This application provides a way to explore the UK computer science research domain across multiple dimensions for multiple stakeholders, from funding agencies to individual researchers.

Second place went to SECO (*semantic collaboration*), from the University of Southern California's Information Sciences Institute. This application enables collaboration in online communities. It collects RDF data from the Web, stores it in an index, and makes it accessible via a Web interface.

Third place went to AnnoTerra, from Science Systems & Applications. This application presents enhanced earth science news feeds by making focused semantic searches on NASA knowledge catalogs using concepts and relationships from the earth science domain.

The other applications appearing in this issue deal with semantic portals (SEAL), provide access to multiple information sources in the area of life sciences (DOPE), integrate satellite images to locate buildings (BuildingFinder), and integrate heterogeneous information sources with respect to distinct thematic topics, namely urban and environmental planning, tourism, labor, and education (GeoShare).

Most of the applications can be accessed and used online via <http://challenge.semanticweb.org>.

Conclusions

The challenge's results lead us to several conclusions. First, and probably the most important, there *are* Semantic Web applications that can be used already or that are about to become a product.

Second, the submitted applications all use heterogeneous sources and standard languages (XML, RDF, and OWL), and meet the minimal requirements. In general, they use relatively straightforward ontologies. Most of these ontologies define concepts and subclass relations between them, but some also specify other relations between the concepts and further characteristics of these relations. For most of these ontologies, RDF Schema provides sufficient representation; OWL's additional expressivity isn't required. About half of the ontologies were specifically created for the applications; the other half are reused. Few of the ontologies contain more than 100 concepts. Most of the ontologies function as a schema for the data, while others guide the user in selecting or finding information. The important outcome is that all the applications use simple ontologies.

What does this mean? Are these applications only shallow examples of what we can expect in the future? Another explanation is

quite simple: most ontologies that will be accessible over the Internet will be shallow; they'll be mostly taxonomies. This could be sufficient for most applications, and—to be a bit more provocative—this could be the future of ontologies for the Semantic Web. The next Semantic Web Challenges will monitor the development of this issue, among others.

Third, we see that several applications already exist that use the Semantic Web infrastructure to help users interpret and collect information. Some of these require manual translation of data, but some applications work directly with data available in the RDF format. Up to now, the infrastructure's reasoning capabilities haven't seen much use because most ontologies are relatively simple. The future will reveal whether this will change.

How far are we from realizing the dream? Judge for yourself! This issue of *IEEE Intelligent Systems* presents seven state-of-the-art Semantic Web applications, their goals, the underlying techniques, and their pros and cons. Do you think you can do it better? There's a new challenge in 2004: see <http://challenge.semanticweb.org/>! ■

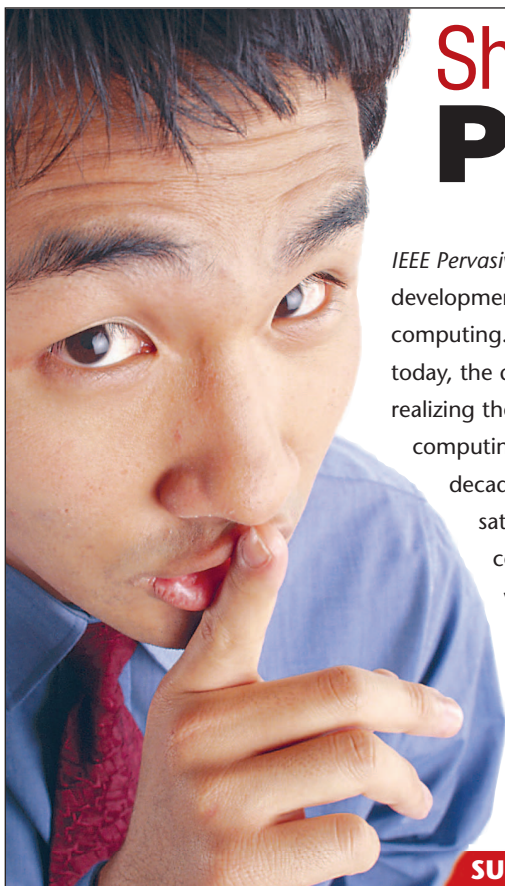
The Authors



Michel Klein is a post-doctoral researcher in the Knowledge Representation and Reasoning group at Vrije Universiteit. His research interests include the use of ontologies for information integration, representational issues of ontologies, and support for the dynamic aspects of knowledge representation on the Web. Contact him at Vrije Univ., Computer Science Div., De Boelelaan 1081a, 1081 HV, Amsterdam, NL; michel.klein@cs.vu.nl; www.cs.vu.nl/~mcklein.



Ubbo Visser is an assistant professor at the University of Bremen's Center for Computing Technologies. His research includes knowledge representation and reasoning components for technological, spatial, and temporal issues with regard to the Semantic Web. Contact him at the TZI, Universitaetsallee 21-23, 28359 Bremen, Germany; visser@tzi.de; www.tzi.de/~visserx.



Shhh. Pervasive Computing, Pass it on...

IEEE Pervasive Computing delivers the latest developments in pervasive, mobile, and ubiquitous computing. With content that's accessible and useful today, the quarterly publication acts as a catalyst for realizing the vision of pervasive (or ubiquitous) computing Mark Weiser described more than a decade ago—the creation of environments saturated with computing and wireless communication yet gracefully integrated with human users.

Editor in Chief: M. Satyanarayanan
Carnegie Mellon Univ. and Intel Research Pittsburgh

Associate EICs: Roy Want, Intel Research; Tim Kindberg, HP Labs; Gregory Abowd, Georgia Tech; Nigel Davies, Lancaster University and Arizona University



UPCOMING ISSUES:

- ✓ Ubiquitous System Software
- ✓ First Response
- ✓ Energy Harvesting and Conservation



SUBSCRIBE NOW! www.computer.org/pervasive/subscribe.htm