The Zen of the Web

Jeff Heflin

Michael N. Huhns
University of South Carolina - Columbia, huhns@sc.edu

Follow this and additional works at: https://scholarcommons.sc.edu/csce_facpub

Part of the Computer Engineering Commons

Publication Info
http://ieeexplore.ieee.org/servlet/opac?punumber=4236
© 2003 by the Institute of Electrical and Electronics Engineers (IEEE)
The Zen of the Web

In contrast to the original Web’s content, which was designed for human use and comprehension, the Semantic Web’s content is for computer use and understanding. To date, however, most efforts have focused on the understanding rather than use. This special issue of IC focuses on the use of the Web by computer systems and agents. By supporting the notion of “getting work done,” the Semantic Web will become more useful, valuable, and pragmatic.

Many organizations are attempting to make the Web computer-friendly via Web services, but current incarnations of these technologies are subject to several limitations:

- A Web service knows only about itself — not about its users, clients, or customers.
- Web services are not designed to use and reconcile ontologies among each other or with their clients.
- Web services are passive until invoked; they can’t provide alerts or updates when new information becomes available.
- Web services do not cooperate with each other or self-organize, although they can be composed by external systems.

We invited researchers and developers to submit articles that address some of these issues and describe future aspects of Web technologies. Collectively, the articles show how to harmonize Web services’ behaviors and reconcile and exploit Web sources’ semantics.

Ontologies and the Semantic Web

The goal driving the Semantic Web is to automate Web-document processing. To that end, researchers are developing languages and software that add explicit semantics to XML’s content-structuring aspects. A Semantic Web language lets users create ontologies that specify standard terms and machine-readable definitions. Information resources (such as Web pages and databases) then commit to one or more ontologies, thus specifying which sets of definitions are applicable to a specific resource. For example, an ontology about animals might explicitly state that the class Dog is a subclass of Mammal and that the classes Mammal and Fish are disjoint. Logical reasoning systems can use these statements to deduce additional information that was not explicitly stated about the terms in the resource.

For the past 10 years, knowledge-representation researchers have studied the use
of ontologies for sharing and reusing knowledge.² Although there is some disagreement regarding what constitutes an ontology, most include a taxonomy of terms (“a Car is a Vehicle,” for example) and a language for expressing the terms and their relationships. A good definition, provided by Guarino, is that an ontology is “a logical theory that accounts for the intended meaning of a formal vocabulary.”³ Most ontology languages provide mechanisms for extending existing ontologies, which gives users the option of customizing and including domain-specific information.

The Semantic Web is based on the idea of numerous ontologies providing definitions that information resources can commit to. When two sources commit to the same ontology, the same meaning is intended for any term from that ontology. In this decentralized vision, any source can commit to any ontology or create a new one. Thus, the Semantic Web is essentially a distributed approach to creating standard vocabularies.

Several Semantic Web languages exist — from early developments such as Simple HTML Ontology Extensions (SHOE)⁴ and Ontobroker⁵ to more recent entries like the DARPA Agent Markup Language+Ontology Interchange Language (DAML+OIL)⁶ and OWL, the Web Ontology Language⁷ — and they all have different features. SHOE is based on the datalog data model (commonly used for deductive databases) and has mechanisms for supporting ontologies that evolve over time. Ontobroker is based on frame logic and has the tightest integration with existing HTML. With DAML+OIL, an international committee of researchers worked to standardize the best features from preceding Semantic Web languages. It is essentially an expressive description logic with a resource description framework (RDF) syntax. DAML+OIL’s success prompted the World Wide Web Consortium (W3C) to form the Web Ontology working group (www.w3.org/2001/sw/WebOnt/), which is chartered to produce OWL. Designed to clarify and simplify DAML+OIL, this language is now a candidate recommendation and could become an official W3C specification as early as the end of 2003.

The challenges include
- getting information into the appropriate format;
- scaling Semantic Web technology to handle “Web size” data;
- creating, maintaining, and integrating ontologies;
- using the Semantic Web to describe and compose Web services;
- handling inconsistent data; and
- determining what to trust.

A frequent criticism of the Semantic Web is that nobody would be willing to enter data in the necessary structured format. To a certain extent, this is a “chicken and egg” problem: If there were significant content that adhered to Semantic Web principles, more systems and agents would use the Semantic Web for search tasks; if it were used in more searches, more content providers would be willing to provide information in the specified format. Nonetheless, we must simplify the process of providing content for the Semantic Web to succeed.

One solution for reaching this goal lies in the use of wrappers. Much of the Web’s content is currently produced from databases, and manually creating wrappers that could export such content in a semantic language is relatively simple. Researchers have also used various machine-learning techniques to generate wrappers for semi-structured Web pages (that is, large portions of the pages have a regular format). Clearly, the Semantic Web can benefit from this work.

Another concern is whether the tools developed for the Semantic Web can truly handle “Web scale” data concerning billions of Web pages. In particular, knowledge bases are often derived from AI systems that do not typically support this level of scalability. We are making some progress in developing systems and benchmarks (see the “Further Reading” sidebar, next page), but clearly we have much work to do in this area.

The most important question is where the ontologies will come from. Ontology design is a skill that is not widely found in the workforce. Current tools, such as Protégé,⁸ provide only limited help, and they have not been widely used outside of prototyping projects and research groups. Fortunately, we can view ontology design as an extension of logical database design, which means that training data modelers could be a promising approach. To increase sharing and minimize duplicate efforts, we will have to create large ontology libraries. The DAML Web site (www.daml.org) pro-
Recent years have seen a flurry of activity focused on addressing many of the challenges described in this introduction. These selected readings will introduce you to current progress.

### Generating Semantic Data

### Semantic Web Query Systems

### Scalable Semantic Web Systems

### Ontology Integration

### Semantic Web Services
also hold diametrically opposed views on many topics. If semantic search engines will be gathering and combining information for us, we must be able to determine how much we can trust their answers. Given many possible answers, the search engines should ideally rank them by level of confidence. However, a significant problem is that trust is subjective: one person might consider another’s trusted source to be totally biased. Thus, users must be able to adapt any method for calculating trust to their preferences.

The Articles
The three articles in this issue have the characteristics we wanted: they focus on ways to bridge the gap between the meanings (semantics) of Web sources and the behavior of Web services, on integrating and reconciling different Web sources’ semantics, and on integrating and reconciling different Web services’ behaviors.

In “Autonomous Semantic Web Services,” Paolucci and Sycara describe an agent-based view of Web services that promises not only behavioral autonomy, but also semantic harmony. They describe DAML-S and present a prototype system in which several Web services interoperate appropriately because of their adherence to it.

In “Synthesizing an Integrated Ontology,” Ben-eventano and colleagues describe a framework for extracting and integrating information from Web sources that have different semantics and syntax and range from semistructured to fully structured. The framework produces a global view, represented by an incrementally constructed ontology, which enables applications to reconcile and integrate the different sources’ semantics.

Ko and Neches end the theme section with “Web Services for Large-Scale Tasks.” They examine the problem presented by independently developed Web services, which constitute bits of functionality that are difficult for systems and users to compose into larger, more complicated behavioral components. The challenges are similar to those faced in reusing code. To ameliorate the problem, the authors developed Eurasia, a framework that lets end users compose services and test the combined behavior. The resulting distributed Web-based information systems are easier to develop and maintain than conventional systems.

Lanterns for the Journey
This set of articles doesn’t necessarily illuminate everything that is going on with the Semantic Web, but it does illustrate the type of work that is leading the way. We look forward to continued work in this area and to the day when the Web serves us more actively and with more enlightenment.

Acknowledgments
We thank all those who submitted their work, as well as the hard-working reviewers who gave their precious time to ensure the high quality of IEEE Internet Computing articles.

References

Jeff Heflin is an assistant professor of computer science at Lehigh University. A pioneer of Semantic Web research, he wrote the first PhD dissertation on the subject and, with James Hendler and Sean Luke, created SHOE. Heflin is a member of the committee that designed DAML+OIL and an invited expert to the W3C Web Ontology working group. His recent research interests include using ontologies to integrate heterogeneous systems, building scalable Semantic Web systems, and developing theories of distributed ontology systems. Contact him at heflin@cse.lehigh.edu.

Michael N. Huhns is a professor of computer science and engineering and director of the Center for Information Technology at the University of South Carolina. His research interests are in the areas of multiagent systems, agent-based Web services, and enterprise integration. He is a member of the editorial boards of IEEE Internet Computing and five other journals. Huhns is a founding member of the International Foundation for Multiagent Systems. Contact him at huhns@sc.edu.