



The Zen of the Web

Jeff Heflin
Lehigh University

Michael N. Huhns
University of South Carolina

In contrast to the original Web's content, which was designed for human use and comprehension, the Semantic Web's¹ content is for computer use and understanding. To date, however, most such 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 comprises 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.

Challenges

Although a standardized Web ontology language will be a major step forward, several challenges remain to be addressed before the Semantic Web can become a “pragmatic Web” – an online environment that not only helps computer systems find information, but also helps ordinary people accomplish tasks and get practical work done.

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, the process of providing content must clearly be simplified 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. Some progress is being made in developing systems and benchmarks (see the “Further Reading” sidebar), but there is clearly 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, ontology design can be viewed 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-

For Further Reading

Recent years have seen a flurry of activity focused on addressing many of the challenges described in this intro. These selected readings will introduce you to current progress.

Generating Semantic Data

- S. Dill et al., "SemTag and Seeker: Bootstrapping the Semantic Web via Automated Semantic Annotation," *Proc. 12th Int'l World Wide Web Conf. (WWW 03)*, W3C, 2003, pp. 178–186.
- S. Handschuh, S. Staab, and R. Volz, "On Deep Annotation," *Proc. 12th Int'l World Wide Web Conf. (WWW 03)*, W3C, 2003, pp. 431–438.
- J. Wang and F. Lochovsky, "Data Extraction and Label Assignment for Web Databases," *Proc. 12th Int'l World Wide Web Conf. (WWW'03)*, W3C, 2003, pp. 187–196.

Semantic Web Query Systems

- J. Broekstra, A. Kampman, and F. van Harmelen, "Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema," *Proc. 1st Int'l Semantic Web Conf. (ISWC'02)*, LNCS 2342, Springer, 2002, pp. 54–68.
- M. Sintek and S. Decker, "TRIPLE — A Query, Inference and Transformation

Language for the Semantic Web," *Proc. 1st Int'l Semantic Web Conf. (ISWC'02)*, LNCS 2342, Springer, 2002, pp. 364–378.

Scalable Semantic Web Systems

- V. Christophides et al., "On Labeling Schemes for the Semantic Web," *Proc. 12th Int'l World Wide Web Conf. (WWW 03)*, W3C, 2003, pp. 544–555.
- Y. Guo, J. Heflin, and Z. Pan, "Benchmarking DAML+OIL Repositories," *Proc. 2nd Int'l Semantic Web Conf. (ISWC 03)*, (to appear).
- A. Halevy et al., "Piazza: Data Management Infrastructure for Semantic Web Applications," *Proc. 12th Int'l World Wide Web Conf. (WWW 03)*, W3C, 2003, pp. 556–567.

Ontology Integration

- A. Doan et al., "Learning to Map between Ontologies on the Semantic Web," *Proc. 11th Int'l World Wide Web Conf. (WWW 02)*, W3C, 2002, pp. 662–673.
- J. Heflin and J. Hendler, "Dynamic Ontologies on the Web," *Proc. 17th Nat'l Conf. Artificial Intelligence (AAAI 00)*, AAAI/MIT Press, 2000, pp. 443–449.
- D. McGuinness et al., "An Environment for Merging and Testing Large Ontologies," *Proc. 7th Int'l Conf. Principles of*

Knowledge Representation and Reasoning (KR 00), Morgan Kaufmann, 2000, pp. 483–493.

- N. Noy and M. Musen, "PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment," *Proc. 17th Nat'l Conf. Artificial Intelligence (AAAI 00)*, AAAI/MIT Press, 2000, pp. 450–455.
- L.M. Stephens and M.N. Huhns, "Consensus Ontologies: Reconciling the Semantics of Web Pages and Agents," *IEEE Internet Computing*, vol. 5, no. 5, 2001, pp. 92–95.

Semantic Web Services

- S. Narayanan and S. McIlraith, "Simulation, Verification and Automated Composition of Web Services," *Proc. 11th Int'l World Wide Web Conf. (WWW 02)*, W3C, 2002, pp. 77–88.
- M. Paolucci et al., "Semantic Matching of Web Services Capabilities," *Proc. 1st Int'l Semantic Web Conf. (ISWC 02)*, LNCS 2342, Springer, 2002, pp. 333–347.
- D. Trastour, C. Bartolini, and C. Preist, "Semantic Web Support for the Business-to-Business E-Commerce Lifecycle," *Proc. 11th Int'l World Wide Web Conf. (WWW 02)*, W3C, 2002, pp. 89–98.

vides an index of more than 200 existing DAML ontologies, but libraries with much more sophisticated search capabilities will eventually be required. When ontologies are used in production, an important consideration is how to manage dependencies when they must be modified.

Although shared ontologies enable interoperability, developers will inevitably use different ones to describe the same domain in many cases. We must therefore be able to translate, align, and merge them. Ideally, we should be able to publish interontology mappings in ontology format, so that others can reuse the information. Additionally, Semantic Web ontologies will have to evolve to meet their users' needs. Effective ways to manage such changes in highly distributed and decentralized environments are essential to success.

The DAML for Services (DAML-S) initiative is attempting to define how to describe a Web ser-

vice using DAML+OIL. This work focuses on three types of knowledge: a profile of the service, a process model that describes how it works, and a description of how to invoke it. Using this information, researchers are looking at creating match-making services that can find a service that is capable of performing a task. They are also looking at composing sets of Web services to accomplish tasks that no single service could perform.

A key problem is that the Semantic Web, as the product of many individuals who will often disagree, will be inconsistent as a whole. Research must focus either on ways to identify consistent subsets, or on reasoning methods that are not trivialized by inconsistency (as with first-order logic).

On a related note, we need ways to determine what to trust. This is already a significant problem on the Web today where people publish misleading or blatantly false information. Different groups

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 et al. 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 "Composing 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

1. T. Berners-Lee, J. Hendler, and O. Lasilla, "The Semantic Web," *Scientific Am.*, vol. 284, no. 5, May 2001, pp. 34–43.
2. N. Noy and C. Hafner, "The State of the Art in Ontology Design," *AI Magazine*, vol. 18, no. 3, 1997, pp. 53–74.
3. N. Guarino, "Formal Ontology and Information Systems," *Proc. Formal Ontology and Information Systems*, IOS Press, 1998, pp. 3–15.
4. J. Heflin, J. Hendler, and S. Luke, "SHOE: A Blueprint for the Semantic Web," *Spinning the Semantic Web*, D. Fensel et al., eds., MIT Press, 2003. pp. 29–63.
5. S. Decker et al., "Ontobroker: Ontology Based Access to Distributed and Semistructured Information," *Proc. 8th Working Conference on Database Semantics: Semantic Issues in Multimedia Systems*, Kluwer Academic, 1999, pp. 351–369.
6. J. Hendler and D. McGuinness, "The DARPA Agent Markup Language," *IEEE Intelligent Systems*, vol. 15, no. 6, 2000, pp. 72–73.
7. M. Smith, C. Welty, and D. McGuinness, "OWL Web Ontology Language Guide," W3C candidate recommendation, Aug. 2003; www.w3.org/TR/owl-guide/.
8. N.F. Noy et al., "Creating Semantic Web Contents with Protege-2000," *IEEE Intelligent Systems*, vol. 16, no. 2, 2001, pp. 60–71.

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.