Embracing "Web 3.0"

Ora Lassila • *Nokia Research Center*
James Hendler • *Rensselaer Polytechnic Institute*

In an article published in *The New York Times* this past November, reporter John Markoff stated that "commercial interest in Web 3.0 — or the 'Semantic Web,' for the idea of adding meaning — is only now emerging."[1] This characterization caused great confusion with respect to the relationships between the Semantic Web and the Web itself, as well as between the Semantic Web and some aspects of the so-called Web 2.0. Some wanted to reject the term "Web 3.0" as too business-oriented; others felt that the vision in the article was only part of the larger Semantic Web vision, and still others felt that, whatever it was called, the Semantic Web's arrival in the Business section of *The New York Times* reflected an important coming of age.

With the Resource Description Framework (RDF) and Web Ontology Language (OWL) — the languages that power the Semantic Web — becoming standards and new technologies reaching maturity for embedding semantics in existing Web pages and querying RDF knowledge stores, something exciting is clearly happening in this area.

Semantic Web Background

With more than 10 years' work on the Semantic Web's foundations and more than five years since the phrase became popular, it's an opportune moment to look at the field's current state and future opportunities. From a humble beginning as a methodology for machine-interpretable metadata and through a "world-embracing" vision of a new era of software (often — erroneously, in our opinion — attributed as science fiction), the Semantic Web has matured into a set of standards that support "open" data and a view of information processing that emphasizes information rather than processing.

From one viewpoint, the Semantic Web is the symbiosis of Web technologies and knowledge representation (KR), which is a subfield of artificial intelligence (AI) concerned with constructing and maintaining (potentially complex) models of the world that enable reasoning about themselves and their associated information. As such, we can understand the Semantic Web through the lessons learned from the Web's development and adoption, as well as (perhaps somewhat painfully) from the deployment of AI technologies.

On the Web, we've seen the emergence of some completely new business models that do indeed work, despite initially seeming infeasible. These include the models introduced or perfected by Netscape (creating a community by giving stuff away), Amazon and eBay (marketplaces), and Yahoo! and Google (advertising-supported sites). Sharing data (or *content*, as it's often called when discussing the Web) has unexpected and serendipitous outcomes — once you make something available, you have no idea how some people will use it. The long-tail phenomenon — for example, aggregate sales of low-selling items, such as specialized books, surpassing the total number of best-sellers sold — defies traditional thinking about business models, but it's important to the new Web-based economy. Web sites don't really exist in isolation — linking is what makes search engines work and gives the "blogosphere" its power.

From the euphoria surrounding AI in the 1980s through the hangover of the "AI winter" in the 1990s, we've learned what doesn't work: you can't sell a stand-alone "AI application." These technologies make sense only when embedded within other systems. Tools are hard to sell and often fail to make good business sense (they certainly don't make sense according to venture capitalists). Finally, thinking of AI itself, we observe that reasoning engines are a means to an end, rather than the end itself; how you use them is more important than the mere fact that you use them.

1089-7801/07/$25.00 © 2007 IEEE

More recently, many people have become excited about Web 2.0. Although we abhor both the term and the use of version numbers, we see that the movement is rife with interesting phenomena. Web 2.0 is mostly a social revolution in the use of Web technologies, a paradigm shift from the Web as a publishing medium to a medium of interaction and participation. From the Semantic Web viewpoint, however, the most interesting technical aspects are

- *Folksonomies* (or "tagging") provide an organic, community-driven means of creating structure and classification vocabularies; they often succeed where traditional mechanisms for defining ontologies have failed or at least proven cumbersome.
- *Microformats* — the use of HTML markup to decode structured data (with the underlying thinking that human-readable representation now comes free) — are a step toward "semantic data." Although not in Semantic Web formats, microformatted data is easy to transform into something like RDF or OWL for Semantic Web agents to process. W3C is working on new approaches, such as Gleaning Resource Descriptions from Dialects of Languages (GRDDL) and RDFa, to standardize the linking of structured data with instructions on how to transform or embed data into existing Web resources.

Since the 2004 completion of the RDF and OWL standards, we've seen a lot of experimentation (and confusion) regarding the right representation language to use for any particular application. Not surprisingly, subsets and extensions of these languages have started to appear — most notably, versions of RDF(S) that borrow a small number of features from OWL (though remaining simpler than "OWL Lite"). Other developers have invested considerable effort in revisiting and extending the functionality in the OWL standard, which is now emerging as OWL 1.1.

Given that much of the current work was presented at academic conferences, that new journals have sprung up relating to semantic technologies, and that much of the language design happened in academic labs and corporate research centers, some have understandably assumed that the Semantic Web is primarily a research vision that's not yet ready for prime time. However, we're starting to see considerable development within the applications space and, as the "Web 3.0" article revealed, this work is emerging in an important and exciting way.

Web 2.0 is mostly a social revolution in the use of Web technologies, a paradigm shift from the Web as a publishing medium.

Web 3.0

Although Semantic Web proponents have long seen evidence of growing interest, the technology's success has become far more evident in the past few months. This is largely because of the maturing of the RDF languages and the technologies that support them.

Oracle's July 2005 release of RDF support in its Spatial 10.2g database product provided the legitimacy that some felt the language lacked. As people experimented with RDF databases, they found significant advantages over traditional structured databases in many cases, especially with respect to embedding data on the Web. As Microsoft put it in its December 2006 *Connected Services Framework 3.0 Developer Guide*:[2]

> There are two main benefits offered by a profile store that has been created by using RDF. The first is that RDF enables you to store data in a flexible schema so you can store additional types of information that you might have been unaware of when you originally designed the schema. The second is that it helps you to create Web-like relationships between data, which is not easily done in a typical relational database.

As RDF acceptance has grown, the need has become clear for a standard query language to be for RDF what SQL is for relational data. The SPARQL Protocol and RDF Query Language (SPARQL),[3] now under standardization at the W3C, is designed to be that language. As Nova Spivack, CEO of Web startup Radar Networks (www.radarnetworks.com), put it in a February 2007 blog, "There is a huge amount of interest in SPARQL at the moment, and there are already a growing number of SPARQL endpoints popping up around the Web. These new SPARQL endpoints are to data what Web sites were to documents."

Numerous players of various sizes are now focusing in different areas of the Semantic Web space. UK-based Garlik (www.garlik.com), for example, uses Semantic Web technologies for the "control of personal data in the digital world." Specifically, the company is working to let users discover what's known about them on the Web to see what the aggregation of this information (exposed via an RDF store) reveals. Dave Beckett, an engineer at Yahoo announced in November 2006 that the Yahoo Food site (http://food.yahoo.com) is being powered by OWL and RDF, as well as several other technologies. Teranode (www.teranode.com), among others, is exploring the use of Semantic Web

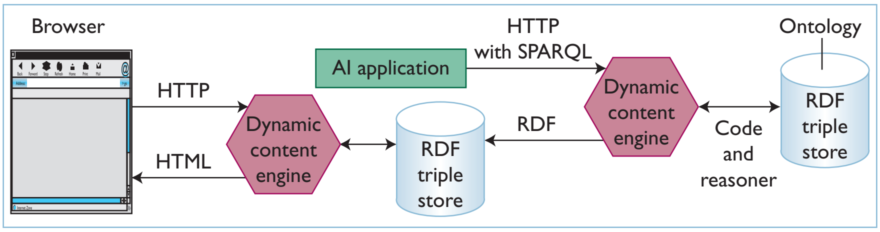

Figure 1. Sample "fractal" architecture of Semantic Web applications. Dynamic content engines, backed by RDF triple stores, act as both producers and consumers of "semantic" data. Data exchange can be facilitated using, for example, SPARQL.

technologies for scientific data integration, particularly in the biology sector. Joost (www.joost.com), the new Internet TV platform that made big news in February in announcing a partnership with Viacom, uses RDF extensively. In fact, Joost announced recently that it will provide its open source RDF back-end technology to the Apache Foundation (www.apache.org), making it much more widely available for use by Web developers.

It's interesting to note how little of this effort focuses on what was once thought to be the Semantic Web's major business sector: the integration of enterprise data assets via ontologies. It isn't that such work isn't going on — Oracle, IBM, and several startups are all providing important capabilities in that area — but embedding RDF and OWL on the Web, via the all-important URI mechanism, is a key part of the emerging excitement over Semantic Web technologies. Whereas the research community is widely exploiting the AI technologies that motivate, in particular, the OWL DL sublanguage, the languages' more "Webby" features — sometimes referred to as the "dark side" of the Semantic Web[4] — are powering the Web 3.0 technology space.

Beyond Web 3.0

How do we see the future of the Semantic Web and, most importantly, the application of Semantic Web technologies for "mainstream" IT problems and systems? With Web 3.0, these technologies are finding fertile ground in multi-tiered Web applications in which the middle tier can be implemented using an RDF *triple store*: a component that allows, among other things, the integration of heterogeneous data sources and repositories. SPARQL's emergence as the standard query language for RDF lets many data stores expose themselves as SPARQL endpoints, thus enabling flexible data exchange among systems. It is leading the way toward Web applications that exhibit a kind of "fractal" structure, with patterns in which one component uses another as a data source (via SPARQL, for example) and acts as a data source to yet another component (see Figure 1). Such architectures open up new possibilities for the original vision of Web services and loosely coupled distributed systems.

Essentially, we can view Semantic Web technology as a novel approach to interoperability: application developers can defer to the runtime accessible semantics of a dialogue between two information systems even after the systems have been deployed. By using reasoning mechanisms to access *implied* information within conversations of explicit statements, and by enabling systems to dynamically add capabilities by acquiring new ontologies and data to reason over, the Semantic Web lets us build future-proof systems that have a chance of "doing the right thing" even in unexpected situations. This approach is particularly amenable to scenarios and situations in which interoperability is critical — for example, the ubiquitous computing vision of environments with pervasive embedded computation. To connect, say, your handheld device to a dynamically changing set of dozens, if not hundreds, of other systems that are often beyond your ownership or control (and potentially hostile), requires fundamentally new approaches to ensuring interoperability. No longer can we expect a priori standardization of every pair-wise interaction between all possible systems we anticipate interacting with; indeed, we can't even anticipate all future scenarios.

Operating in such an open-ended world requires mechanisms for limiting the decision-making scope. For example, when seeking a particular kind of new service to use, you'll want to limit the set of candidates to something that's contextually relevant (such as those that are offered in your current location). Similarly, traditional access-control mechanisms might not scale to situations in which we have an open-ended set of systems and users: we need new decision-making mechanisms to enforce more flexible policies. From the representational viewpoint, Semantic Web technologies offer the possibility of implementing these kinds of technological frameworks and platforms. We claim that context-awareness and policy-awareness are complementary rather than separate mechanisms — think of policies (and their enforcement) as a particular kind of context.

In the longer term, given that Semantic Web technologies are maturing as a means of describing things, we could use their representational power to describe things in the real world. One view is that the physical objects will become Web-accessible in that we will be able to represent them via metadata. Just like applying semantic technologies to problems of interoperability in ubiquitous computing environments, describing physical things will expand our scope beyond the current Web. This is not unlike when some argue that Web services merely exploit mechanisms and technologies developed for the Web, but really have nothing to do with it. Semantic Web efforts provide an approach to constructing flexible, intelligent information systems; some are Web-based applications, but we're certainly not limited to those. At the same time, the application scope for Web technologies is expanding elsewhere. For example, W3C has started an initiative dubbed the Ubiquitous Web, acknowledging the benefits of expanding the Web's reach beyond our desktop and laptop computers to other types of devices and situations. The synergies between ubiquity and semantics are an exciting area in which we expect to see significant future work.

About six years ago, we outlined a vision for the Semantic Web,[5] including a view in which data described in a machine-interpretable way, coupled with a means for defining vocabularies and ontologies, would lead to a revolution in new Web applications. In one of the article's asides, we reflected that we couldn't really predict what the Semantic Web's "killer application" would be. Rather, we claimed, "the abilities of the Semantic Web are too general to be thought about in terms of solving one key problem or creating one essential gizmo. It will have uses we haven't dreamed of." From enterprise data integration to the coming generation of Web TV, the current variety of Semantic-Web-powered applications make it clear that this was an understatement.

Although many aspects of the Semantic Web are yet to be explored, and much research remains to be done, this technology is clearly transitioning into a serious player in the modern Web universe. We might not like the term "Web 3.0," but we enthusiastically embrace the technologies it is bringing to the field.

References

1. J. Markoff, "Entrepeneurs See a Web Guided by Commonsense," *The New York Times*, Business, 12 Nov. 2006.
2. *Connected Services Framework 3.0 Developers Guide*, Microsoft, 2006; http://msdn2.microsoft.com/en-us/library/aa303446.aspx.
3. E. Prud'hommeaux and A. Seaborne, "SPARQL Query Language for RDF," W3C working draft, 26 Mar. 2007; www.w3.org/TR/rdf-sparql-query/.
4. J. Hendler, "The Dark Side of the Semantic Web," *IEEE Intelligent Systems*, vol. 22, no. 1, 2007, pp. 2–4.
5. T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific Am.*, May 2001.

Ora Lassila is a research fellow at the Nokia Research Center in Cambridge, Massachusetts, and is completing a PhD in computer science from the Helsinki University of Technology. Previously, he's been a software engineer, an entrepreneur, a venture capitalist, and has done research in knowledge representation and its application to production scheduling and logistics planning. Lassila's work in the 1990s on machine-interpretable metadata for Web resources contributed to the Semantic Web's launch. He is an elected member of the W3C Advisory Board. Contact him at www.lassila.org.

James Hendler is the Tetherless World Chair at Rensselaer Polytechnic Institute, where he directs the Future of Information project and has appointments in the computer science and cognitive science departments. He also serves as the associate director of the Web Science Research Initiative, headquartered at MIT, and is a member of the W3C's Semantic Web Coordination Group. Hendler is editor in chief of *IEEE Intelligent Systems* and serves on the board of reviewing editors for *Science*. Contact him at www.cs.rpi.edu/~hendler.