Web 3.0 Emerging

Jim Hendler, *Rensselaer Polytechnic Institute*

While Web 3.0 technologies are difficult to define precisely, the outline of emerging applications has become clear over the past year.

I was recently at a meeting where a colleague, significantly younger than I, referred to Facebook as "so 2008!" It was a reminder of just how fast things are changing on the World Wide Web's leading edge.

An open environment with a large number of eager early adopters, the Web is an amazing platform for rapid prototyping, application integration, and innovation. Just as the media and businesses are coming to grips with Web 2.0 phenomena like social networking, Wikipedia and its many offspring, the growth of the blogosphere, and even microblogging, a new generation of technologies is emerging.

There has been some confusion over what to call this rising wave of innovation, which is in part a realization of the Semantic Web vision expressed in 2001 (T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific American*, vol. 284, no. 5, pp. 35-43).

Because these technologies are largely based on mashups that occur at the data, rather than application, level, and often involve the read-write nature of Web 2.0 applications, there has been a tendency to give this new evolutionary stage of the Web its own name: Web 3.0. We can thus essentially view Web 3.0 as Semantic Web technologies integrated into, or powering, large-scale Web applications.

A GOOD YEAR FOR WEB 3.0

Last year was a rewarding one for those of us involved in the Web 3.0 world.

Evidence that something exciting was happening could be seen at the May 2008 Semantic Technology Conference, which drew more than 1,000 attendees from 35 countries; the LinkedData Planet Conference in June; the 7th International Semantic Web Conference in October; and the first Web 3.0 Conference and Expo, also in October. In addition, numerous companies demonstrated new products at various semantic technologies conferences held on six continents.

Despite the overall economic downturn, Web 3.0 also had success business-wise in 2008. The year got off to a good start with the news in January that Metaweb Technologies had received more than $42 million in second-round funding for continued development of its Freebase "social database" (http://venturebeat.com/2008/01/14/shared-database-metaweb-gets-42m-boost).

Semantic search technology got a boost in June when Microsoft acquired Powerset, a company that uses semantic technologies to enhance search engine capabilities (http://news.cnet.com/8301-13953_3-9982015-80.html).

And in October, Radar Networks' Twine, a semantically enhanced social-networking Web application, moved out of beta with more than 50,000 users (http://blogs.zdnet.com/semantic-web/?p=220), and it is growing rapidly.

New technologies are transitioning from universities to start-ups, and several start-ups have come out of stealth into beta. In addition, several large companies such as Yahoo are integrating semantic technologies on the Web (www.w3.org//2001/sw/sweo/public/UseCases/yahoo).

WEB 3.0 ENABLERS

While the specific nature of Web 3.0 technologies are difficult to define precisely, the outline of emerging applications has become clear over the past year. Key enablers are a maturing infrastructure for integrating Web data resources and the increased use of and support for the languages developed in the World Wide Web Consortium

0018-9162/09/$25.00 © 2009 IEEE Published by the IEEE Computer Society

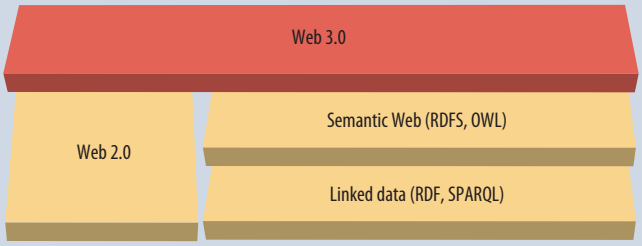

Figure 1. Web 3.0 extends current Web 2.0 applications using Semantic Web technologies and graph-based, open data.

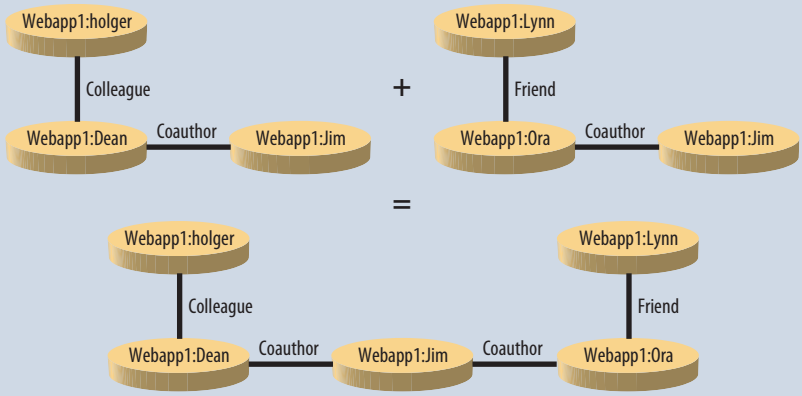

Figure 2. RDF provides a model for merging graphs based on URIs.

(W3C) Semantic Web Activity (www.w3.org/2001/sw).

As Figure 1 shows, the application of these technologies, integrated with the Web frameworks that power the better-known Web 2.0 applications, is generally becoming the accepted definition of the Web 3.0 generation.

The base of Web 3.0 applications resides in the Resource Description Framework (RDF) for providing a means to link data from multiple websites or databases. With the SPARQL query language, a SQL-like standard for querying RDF data, applications can use native graph-based RDF stores and extract RDF data from traditional databases.

Once the data is in RDF form, the use of uniform resource identifiers (URIs) for merging and mapping data from different resources facilitates development of multisite mashups. (For more on the use of RDF in Web-application development, see J. Hendler and O. Lassila, "Embracing Web 3.0," *IEEE Internet Computing*, May/June 2007, pp. 90-93.)

RDF Schema (RDFS) and the Web Ontology Language (OWL) provide the ability to infer relationships between data in different applications or in different parts of the same application. These Semantic Web languages allow for the assertion of relationships between data elements, which developers can use, via custom code or an emerging toolset, to enhance the URI-based direct merging of data into a single RDF store.

In RDF, if we can recognize two data elements with the same URI, then we can join them in a merged graph.

For example, if we know from one dataset that www.example.org/webapp1#Jim (hereafter abbreviated as Webapp1:Jim) has a relation with some other elements and from a different dataset that the same URI has a relation with different data elements, we can unify the two chunks of data into a single graph, as Figure 2 shows.

This graph-merging capability makes it possible to link different datasets together using direct assertions. This might not seem that significant, but on the Web a little information can go a long way in producing interesting new mashups and combined applications.

As such, if our application has a mechanism by which we can assert Webapp1:Jim owl:sameAs www.wikipedia.org/en/James_Hendler. We could then unify information extracted from Wikipedia in RDF with the information from Webapp1. (In fact, DBPedia, http://dbpedia.org, represents the information boxes from Wikipedia in RDF, which allows just such kinds of merging.)

RDF Schema and OWL also provide mechanisms for making numerous inferences about the classes in which various data elements fall and for inferring whether two URIs represent the same or different elements.

Thus, we could assert foaf:email rdf:type owl:inverseFunctionalProperty, which states that any two users with the same foaf:email property should be assumed to be the same user. (FOAF, for friend of a friend, is a commonly used vocabulary for describing properties of people; www.foaf-project.org.)

So if Webapp1:Jim has a particular e-mail address and we determine that Webapp2:JimH has that same e-mail address, we could infer that the information in Facebook describes the same person as in Webapp1, which describes the same person as my Wikipedia page.

INTEGRATING DATA

While many Web 3.0 technologies might seem to be familiar to those in the AI knowledge representation field, the key difference is the Web

naming scheme provided by URIs coupled with the simple and scalable inferencing in Web 3.0 applications (which typically only use a small subset of the OWL language). This combination makes it possible to create large graphs that can underlie large-scale Web applications.

Further, more companies are providing tools for manipulating RDF data, which is helping to accelerate the development of this emerging market.

The term "linked data" is often used to describe the evolving RDF development space, and "Semantic Web" is increasingly being used to describe coupling linked data with RDFS and OWL. These capabilities can be used in numerous different environments, and many current Semantic Web applications are being deployed within industries to do enterprise data integration and related functions.

The term "Web 3.0," in turn, now commonly describes the use of one or both of these capabilities underlying a large-scale Web application, typically including Web 2.0 technologies or approaches.

It is worth noting that several early Web 3.0 applications do not use RDF and OWL directly. However, these applications are increasingly creating SPARQL APIs or RDF exports of their data, as the ability to integrate data using these standards is seen as an opportunity for cross-marketing and more open applications.

Beyond technology, there is another reason Web 3.0 is starting to get "the buzz." As the original Web grew, a key problem was finding the right pages, and Google became a world power by filling that niche.

Web 2.0 has several dominant applications—including Flickr, Wikipedia, Facebook, MySpace, and YouTube—that found ways to deliver the functionality of the "read-write Web" to large user groups, and to great effect. Plenty of money is left to be made in Web 2.0, but these marquee applications dominate the market.

With Web 3.0, on the other hand, the explosion of data on the Web has emerged as a new problem space, and the game-changing applications of this next generation of technology have yet to be developed. To put it another way, in a few years we may be hearing "<your Web 3.0 application goes here> is so 2012," and boy, won't that be great! C

Jim Hendler *is the Tetherless World Chair of Cognitive and Computer Science and the Assistant Dean for Information Technology at Rensselaer Polytechnic Institute. Contact him at hendler@cs.rpi.edu.*

Editor: Simon S.Y. Shim, Dept. of Computer Engineering, San Jose State Univ., San Jose, CA; simon.shim@sjsu.edu