Web 3.0 emerges...

Jim Hendler
Tetherless World Professor of Computer and Cognitive Science
Assistant Dean of Information Technology and Web Science
Rensselaer Polytechnic Institute
http://www.cs.rpi.edu/~hendler
@jahendler (twitter)
The Next Generation Web

Web 3.0 Conference & Expo is focused on bringing together the key proponents and components delivering the promise of next generation web applications, technologies and business utilization.

Web 3.0 showcases in case-study format, the explosive, game-changing promise and disruptive opportunities as they develop in order to help the entire community realize the promise of Web 3.0.

Semantic Web takes that hyper-data and enriches it with meaning through semantic web standards, technologies and strategies, and ultimately leads to the next generation Web.

Linked Data brings traditional hyperlinks and today's "web of documents" into the era of an interconnected, standards-based "web of linked data".
Tetherless World Constellation

On the Web -- links are critical!

Web page

Any Web Resource

HTML

On the Semantic WEB -- links are critical!

RDF is like the web!

2003 slide
Sem Web models start from RDF...

2003 slide

DOC1
<mind:Person rdf:id="Hendler">  
  <mind:title jobs:Professor>  
  <jobs:placeOfWork http://www.cs.umd.edu>  
</mind:Person>
• ~2006: Web app developers discover the Semantic Web

2008 examples include sites from "regular" Web players such as Dow Jones, Reuters and Yahoo!
Facebook's Graph API: The Future Of Semantic Web?

"There are two important themes behind everything we're delivering today," says Bret Taylor, head of Facebook's platform products in the facebook developer blog, about the recent announcements at the f8 conference in San Francisco. Facebook introduced Open Graph protocol, and the Graph API as the next evolution in the Facebook platform.

"First, the Web is moving to a model based on the connections between people and all the things they care about. Second, this connections-based Web is well on its way to being built and providing value to both users and developers — the underlying graph of connections just needs to be mapped in a way that makes it easy to use and interoperable."

Facebook introduced three new components of Facebook Platform two of which the Open Graph protocol, and the Graph API. The API provides access to Facebook objects like people, photos, events etc. and the connections between them like friends, tags, shared content etc. via a uniform and consistent URI to access the representation. Every object can be accessed using the the URL https://graph.facebook.com/ID, where ID stands for the unique ID for the object in the social graph. Every connection (CONNECTION_TYPE) that the facebook object supports can be examined using the URL https://graph.facebook.com/ID/CONNECTION_TYPE.
Google, Twitter and Facebook build the semantic web

02 August 2010 by Jim Giles
Magazine issue 2771. Subscribe and save

A TRULY meaningful way of interacting with the web may finally be here, and it is called the semantic web. The idea was proposed over a decade ago by Tim Berners-Lee, among others. Now a triumvirate of internet heavyweights - Google, Twitter and Facebook - are making it real.

The defining characteristic of the semantic web is that information should be stored in a machine-readable format. Crucially, that would allow computers to handle information in ways we would find more useful, because they would be processing the concepts within documents rather than just the documents themselves.

Imagine bookmarking a story about Barack Obama: your computer will store the URL, but it has no way of knowing whether the content relates to politics or, say, cookery. If, however, each web page were to be tagged with information about its content, we can ask the web questions and expect sensible answers.

...
Google Snaps Up Metaweb in Semantic Web Play

Acquisition of firm offering a free, open source database of real-world information figures to inform smarter search as Google builds out semantic technology.

July 12, 2010
By Kenneth Doctor. More stories by this author.

Google’s shopping spree doesn’t show any signs of letting up, as the Web behemoth continues to open its wallet to acquire firms with niche technologies that can augment its developing business lines.

The latest target is Metaweb, a firm specializing in semantic Web technology that Google has acquired to improve its search engine. Specifically, Google thinks that Metaweb’s database of real-world items can help it produce ordered and accurate results in response to specific, complex search queries. DevX has the details.

In the latest in a string of acquisitions, Google has unveiled plans to purchase Metaweb, a firm that maintains a vast open source database of information about objects in the real world.

Through the acquisition, Google (NASDAQ: GOOG) aims to provide a smarter search engine that will be able to retrieve more specific answers to complex queries.

Read the full story at DevX:
Google Snaps Semantic Web Firm Metaweb

TAGS: Google, semantic Web, acquisition, search engine, Metaweb
Semantic Enterprise: What Are The Gorillas Doing? (Oracle, IBM, HP, Cisco, Microsoft and SAP)

By Bernard Lunn on Aug 03, 2010 07:09 AM

In *Crossing The Chasm* terminology, 'gorillas' are the dominant vendors. Simple message for start-ups - don't mess with them!

In this post, we want to understand what the gorillas are doing to apply semantic web technology to the enterprise. The gorillas in this market are: Oracle, IBM, HP, Cisco, Microsoft and SAP.

**Oracle: Embrace & Extend**

Oracle is active in the semantic web. This matters to them. They cannot afford another database management system based on RDF to replace Oracle and MySQL. Oracle look at RDF as just another thing to store - like objects. They saw off the threat from object management systems and they aim to see off any threat from RDF triples.
Two very different sorts of use cases

- cf. US National Center for Biotechnology Information, "Oncology Metathesaurus"
  - 50,000+ classes, ~8 people supporting full time, monthly updates, mandated for use by NIH-funded cancer researchers
    - OWL DL rigorously followed
    - Provably consistent
- cf. Friend of a Friend (Foaf)
  - 30+ classes, Dan Brickley and Libby Miller made it, maintained by consensus in a small community of developers
    - Violates DL rules (undecidable)
    - Used inconsistently
Widely varying use

NCBI Oncology Ontology
- “High use” in medical community (≈1200 users)
- Very "trusted" information (provenance from NCBI)
- Primarily terminological (relationships between cancer-related concepts), not data-oriented

FOAF
- >60M Foaf people (not necessarily distinct individuals) in hundreds of applications touched by a large community (≈1,000,000,000 users)
- Used by a number of large providers
  - If you use LiveJournal, you have a FOAF file
    - Also flickr, ecademy, tribe, joost, ...
    - And you can export Foaf from Facebook and many other social networking sites
- Becoming de facto standard for open social networking
The argument for NCBI seems compelling

- When "folksonomy" isn't enough...

Which one do you want your doctor to use?
But the cost is too high

Formal modeling finds its use cases in verticals and enterprises
- Where the vocabulary can be controlled
- Where finding things in the data is important

Example
- Drug discovery from data
  - Model the molecule (site, chemical properties, etc) as faithfully and expressively as possible
  - Use "Realization" to categorize data assets against the ontology
    - Bad or missed answers are money down the drain

But the modeling is very expensive and the return on investment must be very high!
- Which is part of why the "expert systems revolution" wasn't one
- Became part of the technology tool kit, a useful niche in the programming pantheon, but didn't change the world

Analogy: the pre-web hypertext world
• OWL is based on RDF, a language designed for the (Semantic) Web
  – Built with Web architecture in mind
    • Exploits Web infrastructure, respects W3C TAG recommendations
      – Internationalization, accessibility, extensibility
  – Fits the **Web culture**
    • Open and extensible, supports communities of interest
      – *If you don't like my ontology, extend it, change it, or build your own*
    • Fits the Web application development paradigm
      – Scales like "databases"
  – With some new ways of linking to formal models
    • **Heavy use of a small amount of RDFS and a tiny bit of OWL**
    • Generally used "like it sounds" not like the formal model
      – Example "owl:sameAs" debate

“linked data” often used to describe this low semantics Semantic Web

Analogy: the World Wide Web
"Linked Data" approach finds its use cases in Web Applications (at Web scales)

- **A lot of data, a little semantics**
- **Finding anything in the mess can be a win!**

**Example**

- Declare simple inferable relationships and apply, at scale, to large, heterogeneous data collections
  - *eg.* Use InverseFunctional triangulation to find the entities that can be inferred to be the same
    - **These are "heuristics" not every answer must be right (qua Google)**
    - But remember *time = money!*
<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 1, 2009</td>
<td>“Openness will strengthen our democracy and promote efficiency and effectiveness in Government.” --- President Obama</td>
</tr>
<tr>
<td>May 21, 2009</td>
<td>data.gov online</td>
</tr>
<tr>
<td>December 8, 2009</td>
<td>“Open Government Directive” released</td>
</tr>
<tr>
<td>May 21, 2010</td>
<td>data.gov relaunch 6.4 billion triples</td>
</tr>
<tr>
<td>June 30, 2009</td>
<td>Putting Government Data online</td>
</tr>
<tr>
<td>January 19, 2010</td>
<td>data.gov.uk online</td>
</tr>
</tbody>
</table>
• Built around linked data with top-down push from “Number 10”
• Originated with 3rd parties doing Sem Web bottom up

• May 21 2010, embraced by data.gov
Data.gov community

Tetherless World Constellation
Pump through to Google Viz for demos
A Typical Mashup: CASTNET

1. Convert raw dataset into linkable RDF
2. Query multiple RDF dataset via SPARQL endpoint
3. Drill down for details
4. Surf to EPA applications

Data Mashup
Visualization Mashup
Web Application Mashup

Adding some Web magic

Tetherless World Constellation

Web Analytics

Social Data Networks

External Links
NTIA internet study vs. libraries

Tetherless World Constellation

Map of the United States showing data on poverty, broadband, and internet access.
State Internet Statistics versus Twitter Geocodes

Random Sampling of Streaming Tweets, 2/12/2010
Tweets including the word "haiti", 1/22/2010-1/29/2010
Visualization can help identify data errors

Tetherless World Constellation

National Wild Fire Statistics and Budget

Correlates fires, acres burned, and agency budgets
Visualization can help identify data errors

Were there really no fires in 1985?
Presents a challenge – different ontologies
Presents a challenge – different ontologies

Tetherless World Constellation
Combining data from different sites

Tetherless World Constellation

Comparing US (USAID) and UK (DFID) Global Foreign Aid

- Aid to Indonesia (in USD)
  - USAID
  - DFID

- USAID Categories (in USD)
  - Economic/Security Assistance
  - Development Assistance
  - Child Survival and Health
  - Other USAID Assistance

- DFID Categories (in USD)
  - Education
  - Health
  - Social Services
  - Government and Civil Society
  - Economic
  - Research
  - Humanitarian Assistance
  - Non Sector Allocable
Presents a challenge – different ontologies
Tetherless World Constellation

An RPI touch – supersize it
This on the Web takes us into “Web 3.0”

Web 3.0 extends current Web applications using Semantic Web technologies and graph-based, open data.
Web 3.0: The Dawn of Semantic Search

James Hendler, Rensselaer Polytechnic Institute

Emerging Web 3.0 applications use semantic technologies to augment the underlying Web system's functionalities.

IEEE Computer, Jan 2010; IEEE Computing Now, Feb 2010 (free)
Tetherless World Constellation

Web 3.0 Applications

Lots More
Significant and growing commercial interest...

- Web: Google, Amazon, Travelocity...
- Web 2.0: Facebook, Wikipedia, YouTube, Twitter...
- Web 3.0: the big ones are still out there
• The Semantic Web is going just fine thank you
  – People asking “how,” not why
• So far the commercial driver has been “weak semantics”
  – In the enterprise
• Web 3.0 adds semantics as a value add to regular Web functionality
  – Semantic search
  – Semantic match
  – Semantic etc
• The big one is still out there