Current Status and Future Promise of the Semantic Web

James Hendler
University of Maryland
College Park, MD

Ora Lassila
Nokia Research Center
Cambridge, MA

November 2006

Before We Get Started…

• Two views of the Semantic Web:
  • implementing SEMANTIC applications using web technologies
  • using semantic technologies to support new WEB applications

• Exploring the relationship (and tension) between the two over time

• This talk is
  • a retrospective, a status report
  • an “interpretation”
  • some thoughts on the future…
1990’s: “Pre-history”

- **Rebirth of Artificial Intelligence (end of “AI Winter”)**
  - “big” AI applications
    - Deep Blue, Mars Rover, Deep Space 1, ...
    - embedded vs. stand-alone
  - Web AI
    - IR, statistical NLP, machine learning
    - lots of data!
- **Emergence of the Web**
  - new ways of doing things
    - new business models (even new social models)
    - new technology
  - “dot-com” boom
- **Early forays into “meta-content”**

---

**Exploration of KR applications on the Web**

**Applicability Across Domains**

**“The Web”**
2000-2001: What Did We Believe?

- Jim: Semantic Web and the advent of pervasive computing (March 2000)

- Jim: Roadmap from the “old” Web to the Semantic Web (October 2001)

- Ora: Semantic Web and the advent of pervasive computing (June 1999)

- Ora: Roadmap from the “old” Web to the Semantic Web (October 2001)
2000-2001: “Early Years”

• “Dot-Com” optimism still prevails: 
  easy to explore new directions
  
  • Government meddles with semantics
    • DARPA’s DAML program; EU follows
    • DAML+OIL
  
  • Web community discovers metadata
    • W3C Metadata Activity
    • RDF

“3-pronged” attack:
- DARPA
- EU IST
- W3C

Model Complexity

Applicability Across Domains

Semantic Web research & new standards (DARPA, EU, …)
Original Outline (July 2000)

Scientific American Articles mentions

1. Semantic Web Vision (TBL)
2. What are the analysts? (in sequence)
   - Screen Scouring (Ont and TBL)
   - Data on Web (Ont and TBL)
   - Zip code link between Dmoz (Ont)

   “Then, a miracle occurs”

III. What can you do with it? (not necessarily in sequence)
   - Self-describing documents (Ont)
   - Logic to emit... (TBL)
   - Services and Advertising (Ont)
   - Digital Signatures, Authentication, and Trust (TBL)

Funded Research
WG activity
Recommendation

2001

• Research, experimentation, early demonstrations
• Reminiscent of the early days of the Web
2003

- Early government adoption
- Emerging corporate interest

2005

- Commercial tools
- Lots of open source software
- Scalability
2006: You Are Here!

Then a Miracle Occurs...

"I think you should be more explicit here in step two."

from What's so Funny about Science? by Sidkey Harris (1977)
Significant Corporate Activity

- Semantic (Web) technology companies starting & growing
  - Siderean, SandPiper, SiberLogic, Ontology Works, Intellidimension, Intellisophic, TopQuadrant, Data Grid, ...

- Bigger players buying in
  - Adobe, Cisco, HP, IBM, Nokia, Oracle, Sun, Vodaphone...
  - announcements/use in 2005-2006
  - integrator and contractor uptake: Northrop Grumman buys Tucana, Lockheed-Martin uses SiberLogic in FCS, WebMethods buys Cerebra, ...
  - tools being announced: AllegroGraph, TopBraid, ...

- Government projects in and across agencies
  - US, EU, Japan, Korea, China, ...

- Life sciences/pharma an increasingly important market
  - Health Care and Life Sciences Interest Group at W3C

- Many open source tools available
  - Kowari, RDFLib, Jena, Sesame, Protégé, SWOOP, Onto(XXX), Wilbur, …

50+ Semantic Web press releases within one month!
Growing Government Activity (US&EU)

• Agencies moving beyond the “talk” phase
  • primarily prototyping, but first acquisitions starting
• Example:
  • NASA is developing an enterprise data strategy around using existing data via Semantic Web integration

Lots of activities across NASA

• Science, Engineering, and Mission all have SWT production or development efforts in place
• Now focus in on re-using the data systems we already have in place
• Agency wide integration planning is underway for building a federation of models into an integrated information service across all disciplines

There's a Lot Out There!

2,120,000 hits on "RDF filetype:rdf"
13,600 hits on "ontology filetype:owl"

(Paid ads)
More OWL Use

- The OWL namespace has been declared by 113,000 SWDs (8%) and actually used by 108,000 (7%)
- The RDFS namespace enjoys more use, being declared by 677,000 (47%) and used by 538,000 (37%) SWDs
- owl:Class is the most used term from the OWL namespace with about 1,800,000 instantiations in 68,000 SWDs
- significant use of two OWL equality assertions: owl:sameAs (280,000 assertions in 17,000 SWDs) and owl:equivalentClass (70,000 assertions in 4,300 SWDs) – their common use may be an indication of increased ontology alignment

Semantic WEB

- Data harvesting & visualization
- A little Semantics goes a long way
- Web-based social networks
Enterprise Information Integration

- Deployment of semantic technologies is easier in a “controlled” environment
  - such as a corporate intranet
- Key benefits from Semantic Web Technology:
  - reuse of installed clients and servers
    - careful design of SW languages for Web compatibility
  - leave data in place, integrate through an RDF store
    - analogous to 3-tiered Web application
  - heterogeneity supported by ontologies

"Corporate Semantic Web", Gartner "hot pick" for 2006
2006: The Gap Is Closing

Semantic Web applications of varying complexity and applicability

Model Complexity

Applicability Across Domains

SEMANTIC Web Lessons

• What we learned from AI...
  • embedded AI succeeded, stand-alone did not
  • tools are hard to sell
  • reasoners are a means, not an end
  • knowledge engineering bottleneck

• ...applied in the Web context
  • futureproofing
    • URIs are important
  • good standards evolve
    • languages (RDFS, OWL, RIF, …)
    • content!
Semantic WEB Lessons

- Web needed high value sites
  - personal (homepages, pets)
  - public (hobbyists, govt)
- As these linked up, new functionality emerged
  - Yahoo, Alta Vista, ...
- New business models followed...
  - "give it away" (Netscape)
  - marketplace (Amazon)
  - advertising (Yahoo, Google)
- What do we need?
  - Open Source Tools
  - Open Source Datasets
  - Open Source Harvesters

The “Layer Cake” is Evolving…

2001
2006
New Languages Underway

- **RIF: Rules Interchange Format**
  - representing rules on the Web
  - linking rule-based systems together
- **SPARQL: Query language for (distributed) triple stores**
  - the “SQL of the Semantic Web”
- **GRDDL/RDFa: Integration of HTML and Semantic Web**
  - “embedding” RDF-based annotation on traditional Web pages
- **OWL: New features, specialized subsets**
  - RDF++/OWL Mini – simplification, identity, scaling to large datasets
  - OWL 1.1 – additional expressivity for OWL constructs
- **And more…**
  - multimedia annotation, Web-page metadata annotation, Health Care and Life Sciences (LSID), privacy, etc.

Linking Is Power!

[Diagram showing Applicability Across Domains and Model Complexity]

- Model Complexity
- Applicability Across Domains
Semantic Web vs. “Web 2.0”

- Data with formal semantics
  - RDF, OWL
  - SPARQL, RIF
- Spontaneous information integration (finally!)
- Semantic Web services, agents
- Strong emphasis on open standards

- New social phenomena: blogs, wikis, tagging, folksonomies
- New user interfaces
  - AJAX (or: “Rich User Experience”)
- “New” kinds of data
  - microformats, RSS
  - “mash-ups”
  - Web services
- Plays “fast & loose” with standards

Semantic Web & “Web 2.0”

- What is their relationship?
- Will they stay separate? Does that even make sense?
Semantic Web & “Web 2.0”

• NO! Considerable synergies exist

Exploiting “Web 2.0”

• Vast amounts of “semi-engineered” knowledge
  • Flickr: tens of millions of keyword-tagged photos
  • microformatted Web documents
  • Wikipedia: thousands of carefully documented subjects (in a hierarchy, with disambiguation, …)

• Generate “persistent” URIs
  • "Tank" [hypertext reference]
  • "Tank" [hypertext reference] (small town in Pakistan)

• Remember: Anything with a URI can be linked to the Semantic Web!
Linking of “Web 2.0” & Semantic Web

- Using informal Knowledge Engineering (KE) to bootstrap "formal" KE
- Extending formal KE from tags/wiki

Looking Further Out

Applicability Across Domains

Model Complexity
Where Are the Agents?

• “Brave New Applications”
  • operate autonomously in “unanticipated” situations
  • exhibit robustness in the face of
    • changing, inconsistent and unexpected data
    • variations in reliability, trust
  • capable of serendipitous behavior, opportunism
• Move from the “tool use” of personal computing to systems that work on our behalf
• (Semantic) Web services as “plumbing” for agents
  • emerging as we speak…

Pervasive Computing & Semantic Web

• Pervasive Computing is an interoperability nightmare!
  • instead of sometimes connecting a handful of devices, dynamically connect/disconnect/reconnect possibly hundreds of devices
• Today, high cost of ensuring interoperation
  • any interaction has to be specifically designed/engineered
  • heavy emphasis on application-specific standardization
  • spontaneous interoperability is next to impossible
• The vision is largely contingent on getting unanticipated “encounters” of devices to work
  • how do you behave in a situation not covered by a standard?
  • not “future-proof”
• Semantic Web is a good match
  • (It is an “interoperability technology”)
Other Emerging Trends

- **Semantic Web Services**
  - crucial for linking “programs” into the mix
  - “plumbing” for agents...
- **Scaling Semantic Web stores to database sizes**
- **Information extraction and semantics ("Web 3.0")**
  - can we “retrofit” semantics on the existing Web?
- **Semantic Web information creation**
  - can we make it so we don't have to retrofit in the future?
    - tools that help embed the semantics as a document is created
    - better dynamic integration of structured data into the Semantic Web
  - “Semantic Desktop”

---

Summary

- **Most things we predicted have happened**
  - (or are happening at the moment…)
- **Some things happened faster than we anticipated**
  - triple store scaling
  - reasoner performance actually matters
  - ontologies are there (but very little linking)
- **Some things are yet to materialize (but we are hopeful)**
  - public information sources (as RDF, OWL, …)
  - digital convergence, pervasive computing just emerging
  - little progress on agents

*Now go out there and make some money off this…!*
Any Questions?

The Semantic Web

“A Little Web Goes A Long Way”

“A Little Semantics Goes A Long Way”