Semantic Web

Getting from here to there… … or not.

Obstacles 1

- Usefulness
  - What can syllogisms do for us? Or not do?
  - http://www.shirky.com/writings/semantic_syllogism.html
- Ability to define logic in a way machines can “understand”
- Flexibility vs. Efficiency
  - Is there such thing as too heterogeneous?
- Standards, Compatibility and Politics
  - Can’t we all just get along?
  - (remember the browser wars?)

Obstacles 2

- Legacy of HTML Web
  - If search engines begin to favor semantically encoded pages, will legacy pages be lost?
  - Who’s going to rewrite the billions of pages?
  - Will the SW bring economic efficiencies worth the cost of the transition?
- Populating the SW
  - Metadata is expensive! There needs to be significant incentives for authors to produce metadata. Critical mass needed.

Flexibility vs Efficiency

- RDF is arguably more flexible but currently less efficient than object oriented and XML databases, typically requiring ten times as much memory. This is because
  - in RDF everything is represented as a graph even if there is a more efficient representation
  - every node, and sometimes every arc, is indexed
  - all arcs and most nodes are URIs, and few if any systems use namespace compression

Compatibility

- RDF/XML is a layer built on top of XML but which is not entirely compatible with XML
- Therefore, standard XML tools cannot be used
  - cannot use schema validation languages, so lots of RDF/XML documents contain mistakes
  - it cannot be manipulated and repurposed using XSLT
- Have to use a specialist RDF parser like ARP which is 5 – 20 times slower than XML parsers
- This is largely the result of a rift between the RDF community and the XML community…

According to Tim

1. The syntax of RDF/XML is sufficiently scrambled and arcane that it is neither human-writeable nor human-readable.
2. The RDF/XML syntax makes heavy use of qnames that is neither intuitive to humans nor conforms particularly well to Web Architecture, which requires that everything significant be identified by URI.
3. People who care about metadata have no trouble thinking in terms of resource/property/value triples.
4. Alternatives like N3 that make the RDF triples evident in syntax suffer in comparison to the RDF/XML syntax because they lack XML’s widely deployed base of software, i18n facilities, and APIs.
5. The notion that RDF can be mixed into XML transparently enough to be unobtrusive has failed resoundingly in the marketplace.

Tim Bray, Co-Inventor of XML http://tbray.org/ongoing/
Adoption

- In order for the Semantic Web to become populated (and therefore useful) at least one of 2 things must happen:
  - There must be significant incentives for companies and individual authors to enter metadata into the SW
  - Or, entering metadata into the SW must become cheap and easy

Potential Incentives 1

- Competition
  - Sellers would be compelled to use SW if buyers started using SW agents to locate products
  - Chicken and egg problem – SW isn’t much use to buyers until sellers adopt it
- Law
  - If companies were required by law.
    - For instance, all public filings to the SEC would be made through SW. Then financial institutions as well as government could automate their data collection process which is currently labor intensive and expensive.

Potential Incentives 2

- Closed-loop Systems
  - The intelligence community is looking at adopting RDF and the SW model for sharing heterogeneous data between agencies (www.org/ongoing)
    - DARPA is currently funding much of the W3C’s SW work in hopes of building the DARPA Agent Markup Language (DAML) (http://src.net/np/news/article.php/220971)
- Profit
  - If a search engine or other outfit were willing to pay web authors to populate their sites with proprietary metadata. The metadata, which would be trusted and exclusive to the search engine, would give the search engine the ability to provide agent-like services and higher quality searches.

Making Metadata Cheap

- Software
  1. Metalog “Towards the Semantic Web”
    - Pseudo natural language interface makes metadata entry easy (?)
    - Powerful reasoning engine for query processing
  2. IsaViz
    - IsaViz is a visual environment for browsing and authoring RDF models represented as graphs.
    - No query processing
  3. Grassroots Semantic Web Tools

Metalog v2.0b

Metalog v2.0b
Grassroots Semantic Web Tools

- Ontologies
  - Don’t require user to define ontology first, let them encode and infer ontology
- Ontology Equivalencies
  - Simple interface that allows users to align equivalent concepts from different ontologies
  - Preserve and share these maps
- Naïve Query Processing
  - Software uses alignment maps, query rewriting rules to make naïve queries work across different schema

Time to Vote…