The Policy-Aware Web

Jim Hendler

http://www.cs.umd.edu/~hendler

The Semantic Web

Lots of papers showing up on the "semantic web" – But what is it?
And where is it going?

(Berners-Lee, 03)
On the Web -- links are critical!

Web page

HTML

<a href="http://…">

Any Web Resource

URI

On the Semantic WEB -- links are critical!

RDF

URI

URI

labeling the links
“Use the Links”

Burkitt’s Lymphoma

Web

Semantic Web

-Oncogene(MYC):
  Found_In_Organism(Human).
  Gene_Has_Function(Transcriptional_Regulation).
  Gene_Has_Function(Gene_Transcription).
  In_Chromosomal_Location(8q24).
  Gene_Associated_With_Disease(Burkitts_Lymphoma)

8q24
PVT1

“MODEL”

PVT
Rearrangement of a DNA sequence homologous to a cell-virus junction fragment in several Moloney murine leukemia virus-induced rat thymomas

PubMed

"Use the Links"
The SEMANTICS is in the links (e.g. to ontologies)!
A very old idea in new clothes

- Scientists communicate by use of models
  - c.f. Physical
  - c.f. Mathematical
  - c.f. Organizational

Models expose semantics
Sem Web Languages

- Graph
- Labeled graph
- Graph + limited logic
- Rules/Logic

- Data Dictionary (Vocab)
- Ontology
- Ontology
- Ontology

- RDF
- RDF Schema
- OWL
- ??

All of these languages add semantic modeling primitives to XML - so you can “do this in XML” per se, but it is reinventing the wheel.
Web Modeling Languages

• Resource Description Framework (RDF)
  – Few, but important, constraints
  – A basic, extensible assertional language

• RDF Schema (RDFS)
  – Weak structuring of sets of terms (taxonomy-esque)
  – Class and property hierarchies
  – Domain and Range constraints

• The Web Ontology Language, OWL
  – Stronger structuring of sets of terms (ontologies)
  – Everything in RDFS plus
    • Complex Class constructors (unionOf, intersectionOf)
    • Additional property features (inverse, transitive)
    • Class local property type and cardinality constraints
    • And more
Example Sem Web Roadmap

**Short-term**
- Ontology Focus
  - Use, creation, tools, editing, collaboration, sharing, learning/modifying
- Data and Multimedia applications
- Support collaboration/sharing

**Mid-term**
- “Rule” Focus
  - Development/personalization
  - Access control
- Help users explore/analyze relationships in data
- Privacy and trust issues

**Long-term**
- "Agent" Focus
  - Learning/modifying, model-based interaction, higher level of automation (less human in the loop)
- “Google for information”
Now

• Companies getting into the act
  – IBM SNObase ontology management system
  – Oracle discusses (some) forthcoming support at W3C Life Science Workshop
  – Licensed versions of a scalable triple store currently for sale
    • First results for 1B triples
    • Many of the features of an RDBMS in TKS
      – Concurrency control
      – Security model
      – Soap and linking tools
    • OWL support coming
  – HP Labs open-source Jena API

• Many open source tools becoming available for experimentation/academic use
  – Kowari, RDFLib, 3Store…
A lot is out there already
Collaborative Editing

A screenshot of the SWOOP v2.2 beta3 interface, showing an unsatisfiable concept and its complement. The concept involves a Koala as a human and related annotations and properties. The interface includes options for concise format, properties tree, and a section for annotations and intersection of concepts.
Multi-ontology support
Tools for markup...
Of different media and types
Of different media and types

The scene from the James Bond movie where the guy throws his hat at a statue

The story that ran on NHK television from 0847-0903 on 2001-09-11 (GMT + 9)
Semantic Web Application Portals

Disclaimer: This site is for demo purposes only! Consider all data as fake.

- Location
  - City
  - Country
- Person
  - George W. Bush (1)
  - Saddam Hussein (2)
  - William Jefferson Clinton (2)
  - Yitzhak Rabin (1)
- Agent
  - Terrorist
  - Adam Frey (1)
  - Aziz Eknati (1)
  - Ferhat Urga (1)
  - Mahamad Elhajami (1)
  - Sirrini Fettah (1)
  - Khalid Bin Muhammad Al-Jubair (1)
  - Mohamed Jaa (1)
  - Muhammad Said Al Hasan (1)
- Terrorist Action
  - Abd al-Hadi Iyad, Farouk Al-Tunisi, Abdennour Ben Habib Jedid (1)
  - Abdal, Abdullah Alkafi (2)
  - Fadi, Fajij Musharaf (4)
  - Ferhat Yenici, FY (1)
  - Fadil Alkhairi (1)
  - Ibrahim Khat (1)
  - Ryas Kuncak (7)
  - Mustafa Cabuk (1)
  - Mikaelsson (3)
  - Possibly wife of Must Cabuk (1)
  - Ramzi Youssef (2)
  - Razza, Razik Iqbal (3)
  - Hazim, Hazem Mustafa (6)
  - masked (1)
  - unknown (1)
  - unknown (10)
  - unnamed (11)
Info views...

Abu Muhammad, Ayman al-Zawahiri

Depictions:

<table>
<thead>
<tr>
<th>Abu Muhammad, Ayman al-Zawahiri</th>
<th>Physical Feature</th>
<th>Male, Olive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Member Of</td>
<td>Al Jihad, Egyptian Islamic Jihad</td>
<td></td>
</tr>
<tr>
<td>Citizenship</td>
<td>Egypt</td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>Afghanistan</td>
<td></td>
</tr>
<tr>
<td>Knows</td>
<td>Osama Bin Laden</td>
<td></td>
</tr>
<tr>
<td>Al Jihad, Egyptian Islamic Jihad</td>
<td>Member</td>
<td>Abu Muhammad, Ayman al-Zawahiri</td>
</tr>
<tr>
<td>Osama Bin Laden</td>
<td>Knows</td>
<td>Abu Muhammad, Ayman al-Zawahiri</td>
</tr>
</tbody>
</table>

Disclaimer: This site is for demo purposes only! Consider all data as fake.
Not tied to specific domains

Bio-Central

Disclaimer: This is a demonstration website! Consider all information as contrived and unfounded.

We would like to extend our thanks to the MINDSWAP group at University of Maryland, for the use of their software, valuable ideas and indispensable help in putting this site together.

leukemia

Depictions:

<table>
<thead>
<tr>
<th>Leukemia</th>
<th>Type</th>
<th>Feeding Event, Behavioural Event</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Information content</td>
<td>badCell</td>
</tr>
<tr>
<td></td>
<td>Has Hypothetical Explanation</td>
<td>cancer</td>
</tr>
</tbody>
</table>
Not tied to specific domains

http://www.mindswap.org
(Semantic Web - 24/7)
With Info Contexts
Extending the Portal Technology
Shared Calendar + photos + talks + ...
Semantic Web Portal

Project goals
- Connecting unstructured inputs to rich metadata
- Bringing in information from "open" sources – provenance
  - tracking who entered the information
  - tracking where the information came from
- Using (multiple) ontologies in OWL
  - To help with annotation
  - To organize information display
  - To manage import/export of unstructured and semi-structured data
- Image and subimage annotation and display
- Demonstrating new information portal management tools enabled by the Semantic Web

Technologies demoed
- The Web Ontology Language OWL
- RDF Query for page generation
  - Equivalence and inverse inferencing
  - Limited classification (OWL DL)
  - RDF/XML syntax for data exchange
  - RDF/N3 syntax for human readability
- Sem Web Integration w/standard Web tools
  - HTTP protocol based (standard server software)
  - Universal Resource Identifiers for web-based access
  - XSLT and Perl for portal presentation (HTML-like view)
Sem Web status

<table>
<thead>
<tr>
<th>2003/1</th>
<th>Research (etc., etc!)</th>
<th>Common Formats</th>
<th>Web Standards</th>
<th>Wide deployment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>PAW cwm</td>
<td>KIF</td>
<td>logic framework</td>
<td>Trusted systems</td>
</tr>
<tr>
<td>Rules</td>
<td>PCA</td>
<td>n3</td>
<td>crypto</td>
<td>Inter-engine interop</td>
</tr>
<tr>
<td></td>
<td>etc</td>
<td>RuleML</td>
<td>rules...</td>
<td>Declarative ebusiness</td>
</tr>
<tr>
<td></td>
<td>Ontology</td>
<td>DAML</td>
<td>OWL</td>
<td>Rule indexes</td>
</tr>
<tr>
<td></td>
<td>cyc</td>
<td>OIL</td>
<td></td>
<td>Web of meaning</td>
</tr>
<tr>
<td>Data</td>
<td>KR</td>
<td>MCF</td>
<td>RDF-S</td>
<td>Cross-App interop</td>
</tr>
<tr>
<td></td>
<td>DL</td>
<td>ASN.1</td>
<td>RDF</td>
<td>Interop within App</td>
</tr>
<tr>
<td>Markup</td>
<td>GML troff</td>
<td>SGML</td>
<td>XML</td>
<td>(Berners-Lee, 03)</td>
</tr>
</tbody>
</table>
Extending the Portal Technology
Shared Calendar + photos + talks + …

Who can see What??
Motivation: Access and Privacy Control

1) If $X$ is AC rep of $Y$, $X$ can delegate W3C member access rights in $Y$.
2) Kari is AC rep of Elisa.

1) If $X$ is employee of Elisa, $X$ has W3C member access rights.
2) Tiina is employee of Elisa.

Tiina: I have W3C member access rights

Proof: Alan 1, Alan 2, Kari 1, Kari 2

2004 NSF National Priorities ITR to UMCP and MIT (Hendler, Berners-Lee, Weitzner- PIs)
So what was hard about OWL?

Q. How is OWL different from earlier ontology languages?

A. OWL is a *Web* Ontology language. Where earlier languages have been used to develop tools and ontologies for specific user communities (particularly in the sciences and in company-specific e-commerce applications), they were not defined to be compatible with the architecture of the World Wide Web in general, and the Semantic Web in particular.

OWL rectifies this by providing a language which uses the linking provided by RDF to add the following capabilities to ontologies:

- Ability to be distributed across many systems
- Scalable to Web needs
- Compatible with Web standards for accessibility and internationalization.
- Open and extensible

- Challenge in OWL was to make the ontologies "live on the Web"
  - Descriptions (Class, properties, etc.) must be 1st Class citizens of the WWW
    - i.e. must use URIs
  - OWL without RDF would just be yet another KR language
Open, Distributed Rules Challenges

• Common Notation
  – "Small matter of standardization"
    • N3, SWRL, RuleML

• Identity vs. privacy
  – How do you identify yourself w/o violating the very privacy concerns we hope to address?
    • Current identity schemes are centralized and universal
    • Can we do a distributed ID model (maybe email based)?

• Inconsistency
  – In logic "P ^ -P => Q"
    • On Web it better not!
      (Supported(Bush) ^ --Supported(Bush)) => you owe me $1000
    • Can we use a paraconsistent logic solution?
Annotated Paraconsistent Logic
(in 25 words or less)

• Traditional Logic
  \[ P \land \lnot P \Rightarrow Q \] (P and \lnot P are inconsistent)

• Annotated Logic
  – \[ P;X \land \lnot P;Y \] are not inconsistent
  – \[ P;X \land \lnot P;X \Rightarrow Q;X \] but not \[ Q;Y \]
  – \[ P;X \land \lnot (P;X) \] is inconsistent and must be avoided (but this is easily checked if inference of RHS is restricted)
On the Web

Annotations represent document contexts

\[ X;Y \text{ and } -(X;Y) \text{ cannot co-occur} \]
(\text{unless Web is broken})

(modulo temporal change, but that's another talk)
Another Cool thing…

• What is a rule of logic?
  – In traditional philosophy it relates to "Truth"
    • What is truth on the Web?
      – Ex: How many cows are in Texas?
    – On the Web, we could use an idea of agreed upon rules, grounded at URI
      • Social definition of truth via shared contexts
        – Ex: Because Mom said so…
Truth on Web Pages [based on Heflin et al, 1998]

- Inference rules could be used to determine the credibility of claims
  - I might believe the claims made by a reliable Newspaper
    - Trustable(x) :- x; reliableNewspaper.
  - And I could establish the Washington Post as reliable...
    - i.e. I assert:
    - or if I infer it
      - ReliableNewspaper(X) :-
        X owl:class ReliableNewspaper; http://MediaWatchList.
      - (?) reliableNewspaper(X) :-
        X owl:class ReliableNewspaper; src ^ trusted(src).

- The rules are "grounded" in a testable way
  - If I can HTTP-get the fact, then it is asserted
Rule Sets could be shared

- You can ground your sources
  - \( X \maps X; \text{src} \land \text{src} \text{owl:} \text{class} \text{TrustedSource}; \text{http://}\ldots/\text{myMomSet.rdf} \)

- Or infer trusted sources based on other rule sets
  - \( X \maps X; \text{src} \land \text{src} \text{owl:} \text{class} \text{TrustedSource}; \text{http://ex.com/RushLimbaughSet.rdf} \)
  - \( X \maps X; \text{src} \land \text{src} \text{owl:} \text{class} \text{TrustedSource}; \text{http://ex.com/UnabomberRules.rdf} \)
  - \( \text{http://www.rushLimbaugh.com/truths.rdf} \)
Policy Aware WEB

(A) User requests a resource.

(B) 401 error provides access rules.

(C) Proof is generated and pointer is sent in new HTTP-Get request.

(D) Proof is checked, and confirmed, and the transaction succeeds.
You owe me $30

Oh yeah? Prove it

{Purchased(user1, book1, AOL); www.confirm com#t1221122}
{Priceof(book1, $30), AOL- historyDB#29293910}
{Purchase(a,b,c) & Priceof(b,d) -> Owes(a,c,d); www. ont.com/prodont}

The check is in the email!
Conclusions

• Information lives in specific contexts
  – The Semantic Web helps us place information into these (multiple) contexts.

• Control of information requires control of contexts
  – Explication of policies
    • Linked in a Web-like way
  – Integrated directly into the Web
    • With extensions for rules and proofs
  – Is really hard
    • Issues of identity, inconsistency, grounding, change over time
  – But holds great potential
    • Personal Control of your information spaces

• "Policy-Aware" Web project (joint between UMCP and MIT)
  – Goal: make this real!

Other projects

• Today I didn't talk about:

  – Semantic Grid (Scientific computing)
    • OGSI/OWL hybrid
  – Distributed ontology creation
    • Global federated ontology
    • SWOOP/Annotea
    • Distributed DL’s: e-connection framework
  – Semantic Markup for Science
    • NASA Missions documents: http://semspace.mindswap.org
    • Eco-informatics (NSF ITR)
    • Scientific publishing

  – Semantic Web Services
    • Distributed Workflow creation
    • AI planning and SWS
    • OWL-S API
  – Semantic Web tool kit
    • Ontology editor/browser
    • OWL DL Reasoning
    • OWL back-end support
    • OWL blogging tools
    • Multimedia markup (beyond images)

  – Database Interoperability using semantics
    • OWL-DB

See http://www.mindswap.org/papers for a list
(OWL/RDF generated, of course)
MIND SWAP

• Maryland Information and Network Dynamics Laboratory, Semantic Web and Agents Project
  
Jim Hendler  Aditya Kalyanpur  Daniel Krech  Jordan Katz
Bijan Parsia  Taowei Wang  Ron Alford  Daniel Hewlett
Bernardo Cuenca-Grau  Vladimir Kolovski  Kendall Clark  Meem Mahmoud
Jen Golbeck  Chris Halaschek  Michael Grove  Chris Testa
Evren Sirin  Naiwen Lin  Amy Alford

• Corporate Research Partners:
  – Fujitsu Laboratory of America, College Park
  – Lockheed Martin Advanced Technology Laboratories
  – NTT Corp
  – SAIC Corp., Kevric Corp, Top Quadrant

• Govt Funding:
  – NSF, NGA, US Army Research Laboratory, DARPA, DoD, NIST

http://www.mindswap.org
(OWL-powered Semantic Web page)