Robert Engelmore Memorial Lecture

Knowledge is Power: A view from the Semantic Web

Jim Hendler

[w/thanks to the incredible MINDSWAP research group
visit us on the (Semantic) Web at
http://www.mindswap.org ]

Robert Engelmore Memorial Lecture
Knowledge is Power

• But in knowledge resides the power. Because of the importance of knowledge in expert systems and because the current knowledge acquisition method is slow and tedious, much of the future of expert systems depends on breaking the knowledge acquisition bottleneck and in codifying and representing a large knowledge infrastructure.

— R. Engelmore, E. Feigenbaum, 1993
"… documents on the web describe real objects and imaginary concepts, and give particular relationships between them... The title document to a house describes a house and also the ownership relation with a person. ... This means that machines, as well as people operating on the web of information, can do real things.

Tim Berners-Lee, 1994
Knowledge representation … is currently in a state comparable to that of hypertext before the advent of the web: it is clearly a good idea, and some very nice demonstrations exist, but it has not yet changed the world. It contains the seeds of important applications, but to unleash its full power it must be linked into a single global system.

As we publish more info- how do we control access ...
Imagine if

• Every PDF document on the Web contained metadata in a machine readable form as to who created it, when, with what tool, and some of the key document features
Imagine if

• We could
  – Take a photo from our phone
  – Upload it to the Web
  – Search for a vocabulary relating to the content of the photo
  – Easily use that vocabulary to annotate the image
    • Relating it to existing concepts in that space
  – Send it to a Web portal
  – And have it show up indexed under appropriate terms
Imagine if

- There was a distributed knowledge base, accessible through almost every computer on Earth, with information about millions of people, places, things, transactions, processes, services, ... Indexed against thousands of ontologies in a standard KR language with dozens of open-source (and commercial) tools for parsing, serializing, browsing, editing, storing, searching and inferencing these ...
Now stop imagining
The Web

IAAI-05 Robert S. Engelmore Memorial Lecture
Knowledge as Power: A View from the Semantic Web
James Hendler, University of Maryland

The emerging semantic web focuses on bringing KR-like capabilities to Web applications in a Web-friendly way. The ability to put knowledge on the Web, share it, and reuse it through standard Web mechanisms provides new and interesting challenges to artificial intelligence. In this talk, Hendler explores the similarities and differences between the semantic web and traditional AI knowledge representation systems, and sees if he can validate the analogy "the semantic web is to KR as the Web is to hypertext."

Jim Hendler is a professor and director of the Joint Institute for Knowledge Discovery at the University of Maryland. One of the inventors of the semantic web, Hendler was the recipient of a 1995 Fulbright Foundation Fellowship, is a former member of the US Air Force Science Advisory Board, and is a Fellow of the American Association for Artificial Intelligence. He is also the former chief scientist of the Information Systems Office at the US Defense Advanced Research Projects Agency (DARPA), was awarded a US Air Force Exceptional Civilian Service Medal in 2002, and is a member of
The emerging semantic web focuses on providing structure to Web applications in a Web-friendly format. Knowledge on the Web, share it, and find it. This new technology mechanisms provides new and interesting challenges to artificial intelligence. In this talk, Hendler explores the similarities and differences between the semantic web and traditional AI knowledge representation systems, and sees if he can validate the analogy “the semantic web is to KR as the Web is to hypertext.”

Jim Hendler is a professor and director of the Joint Institute for Knowledge Discovery at the University of Maryland. One of the inventors of the semantic web, Hendler was the recipient of a 1995 Fulbright Foundation Fellowship, is a former member of the US Air Force Science Advisory Board, and is a Fellow of the American Association for Artificial Intelligence. He is also the former chief scientist of the Information Systems Office at the US Defense Advanced Research Projects Agency (DARPA), was awarded a US Air Force Exceptional Civilian Service Medal in 2002, and is a member of...
The emerging semantic web focuses on adding semantics to Web applications in a Web-first environment. It allows for the sharing of knowledge on the Web, enabling new and interesting challenges to artificial intelligence. In this talk, Jim Hendler explores the similarities and differences between the semantic web and traditional AI knowledge representation systems, and seeks to validate the analogy between the semantic web and as the Web is to hypertext.

Jim Hendler is a professor and director of the Joint Institute for Knowledge Discovery at the University of Maryland. One of the inventors of the semantic web, Hendler was the recipient of a 1995 Fulbright Foundation Fellowship, is a former member of the US Air Force Science Advisory Board, and is a Fellow of the American Association for Artificial Intelligence. He is also the former chief scientist of the Information Systems Office at the US Defense Advanced Research Projects Agency (DARPA), was awarded a US Air Force Exceptional Civilian Service Medal in 2002, and is a member of several editorial boards.

Adding Semantics -- NOT!
The emerging semantic web focuses on developing tools and techniques to create and share knowledge on the Web, share it, and use it. This allows for new and interesting challenges to artificial intelligence. In this talk, Hendler explores the similarities and differences between the semantic web and traditional AI knowledge representation systems, and sees if he can validate the analogy “the semantic web is to KR as the Web is to hypertext.”

Jim Hendler is a professor and director of the Joint Institute for Knowledge Discovery at the University of Maryland. One of the inventors of the semantic web, Hendler was the recipient of a 1995 Fulbright Foundation Fellowship, is a former member of the US Air Force Science Advisory Board, and is a Fellow of the American Association for Artificial Intelligence. He is also the former chief scientist of the Information Systems Office at the US Defense Advanced Research Projects Agency (DARPA), was awarded a U.S. Air Force Exceptional Civilian Service Medal in 2002, and is a member of...
Semantics is about "real objects and imaginary concepts and the particular relations between them!" (cf. Berners-Lee 94)

And the field that knows how to do this is AI!
(We call it "knowledge representation"!)

James Hendler is a professor of Computer Science at the University of Maryland, College Park. He is the founder of the Knowledge Discovery and Reasoning Group at the University of Maryland and is a member of the Information Systems Laboratory of the University of Maryland. He is also a co-founder of the Semantic Web Company, a company that develops software for the Semantic Web. He is a member of the National Academy of Engineering and the National Academy of Sciences. He is also a fellow of the American Association for the Advancement of Science.
Sem Web Language Stack

• The most used slide about the Semantic Web
  – Berners-Lee, 01: Semantic Web Layercake (this version ca. 03)
AI languages for the Web

<table>
<thead>
<tr>
<th>Semantic Networks (BUT…)</th>
<th>Frame Language (BUT…)</th>
<th>KR Logic-Lite (BUT…)</th>
<th>Next up: OPS5-ish (BUT…)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDF</td>
<td>RDF Schema</td>
<td>OWL</td>
<td>Sem Web Rules Language</td>
</tr>
</tbody>
</table>

All of these languages add semantic modeling primitives to XML.
## AI STANDARDS for the Web

<table>
<thead>
<tr>
<th>Semantic Networks (BUT…)</th>
<th>Frame Language (BUT…)</th>
<th>KR Logic-Lite (BUT…)</th>
<th>Next up: OPS5-ish (BUT…)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDF</td>
<td>RDF Schema</td>
<td>OWL</td>
<td>Sem Web Rules Language</td>
</tr>
</tbody>
</table>

RDF, RDFS and OWL are W3C Recommendations!!!
BUT...

- **Using World Wide Web technology!**
  - **Standard Languages** *(W3C: Feb, 2004)*
    - If it doesn't import/export RDF/OWL, don't buy it!
  - **Web-based approach**
    - No new servers and boxes - use your current Web infrastructure!!!
    - Documents for exchange -- no need for exotic new security protocols, etc.
  - **Eventual network effect**
    - When it is time to open up beyond intranets, we know the technology can scale!

- **Don't underestimate the importance of these!**

Robert Engelmore Memorial Lecture

July 13, 2005
This is not your father's KR…

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

(OWL FAQ - W3C)
Open and extensible

• Example: I have created the world's most comprehensive ontology on feline leukemia
  – Over 45,000 cancer-related classes
  – Over 47,000 common sense terms
And it fits on this slide
• Sem Web languages allow linking of
  – multimedia
  – databases
  – services
    • web
    • Grid
  – meta-data repositories
• Or any other Web resource!
• Including other Semantic Web resources
  – partial mappings just fine
  – this creates a web of models (semantics) much like
    the current web is a web of texts

• Network effect as mappings provide links to
  linked resources

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).

NCI Cancer Ontology (OWL)

BioMedCentral Metadata (XML)

EPA Vocabulary (RDFS)

EPA data set (XML)
And Web services come to the party too

**input xsd:complex="oncogene"**

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).

**output xsd:complex="RiskType"**

```xml
<owl:Class rdf:about="http://annotation.semanticweb.org/iswc/iswc.daml#RiskIndicator">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="http://annotation.semanticweb.org/iswc/iswc.daml#name"/>
      <owl:allValuesFrom rdf:resource="http://www.w3.org/2000/10/XMLSchema#string"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```
Get a B&N price (In Euros)
Of a particular book
In its German edition
And it's real…

- Companies getting into the act
  - Oracle to support RDF in database 10.2
  - IBM SNObase ontology management system
  - Adobe embeds RDF in all content
  - HP, Cisco, Nokia, Sun … announcements/use in '05
  - Start ups in the space (Cerebra, Siderean, SandPiper, …)
    - And failure already (Tucana)
      - Whose IP was bought by large contractor (Northrup Grumman)

- Many open source tools available
  - Open source (Kowari) scalable triple store
    - 100,000,000+ triples
    - Supports RDFS, OWL support coming
  - RDFLib, 3Store…
  - Jena, Sesame …
  - Protégé, SWOOP, Onto(***)...
<table>
<thead>
<tr>
<th>Organization</th>
<th>Usage</th>
<th>Generic Capability</th>
</tr>
</thead>
<tbody>
<tr>
<td>US Intel Agency</td>
<td>Unstructured intel document analysis</td>
<td>Scalable cross-document co-reference &amp; link analysis</td>
</tr>
<tr>
<td>US Dept. of Defense</td>
<td>Federated search and analysis</td>
<td>Ontology-enhanced search &amp; discovery with cross-document co-reference &amp; link analysis</td>
</tr>
<tr>
<td>Legal Research Ctr.</td>
<td>Unstructured legal document search and analysis</td>
<td>Ontology-enhanced search with cross-document co-reference &amp; link analysis</td>
</tr>
<tr>
<td>Viewpoint Systems</td>
<td>Quality test &amp; measurement analysis</td>
<td>Advanced manufacturing data analysis with rapidly changing data schemas</td>
</tr>
<tr>
<td>General Motors</td>
<td>Centralized meta-data repository; EAI software platform</td>
<td>Change impact analysis &amp; configuration management</td>
</tr>
<tr>
<td>Enterprise Software Company</td>
<td>RDF/OWL representation of products and services</td>
<td>Dynamic Web site navigation (SwoRDFish)</td>
</tr>
</tbody>
</table>

(From David Wood) July 13, 2005
Semantic Web Commercial Roadmap

Short-term:
- EII
- RDF triple stores
- OWL as "semantic technology" standard
- Database and Image Markup
- "RDF/OWL in the business enterprise"

"intranet" uses

Coming:
- Personal Data Integration
- Home enterprise
- Social Network-based apps
- Small business portals
- "Rules and policy"

"OEM" market

And even:
- Small business EAI
- Every doctor, dentist, laundromat, etc. has a "supply chain"
- Do business your way, map to their ASPs
- "Like visicalc did to reports"

Disruptive technology

Robert Engelmore Memorial Lecture

July 13, 2005
Sem Web "Technology" Roadmap

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

**Mid-term**
- "Rule" Focus
  - Development/personalization
- Access control
- Help users explore/analyze relationships in data

**Long-term**
- Automation Focus
  - learning/modifying, model-based interaction
- "information factory"
Research Example: Access and Privacy Control

URI variable

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

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

$Tiiina$: 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)
Robert Engelmore Memorial Lecture
July 13, 2005
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 are here…

Documents, linked to
Images, annotated with
Ontologies, linked to
Other ontologies, describing
Databases, exported as
RDF graphs, inputted to
Services, which designate
Documents, linked to
(ad infinitum)
The other shoe falls

- I gave a talk at KR called "The Semantic Web: KR's worst nightmare"
  - KBs in AI are assumed to be consistent
  - KR&R stresses expressivity over simplicity

- On the Web, you cannot assume these!
So

• Semantic Web Languages (RDF/RDFS/OWL)
  – Are not-very-expressive-KR-language standards
    • Not KIF, or even KL-ONE
  – Create non-persistent KBS
    • Servers come and go
    • Ontologies change over time
  – And can't be kept consistent
    • Disagreement, error, dishonesty…
In short

• Semantic Web Languages (RDF/RDFS/OWL)
  – Are not-very-expressive-KR-language standards
    • Like HTML is to SGML
  – Create non-persistent KBs
    • Like the 404 error (w/o which there is no Web)
  – And can't be kept consistent
    • Like the blog-space and Web 2.0
• Which is why they might be just what the doctor ordered…
Knowledge representation … is currently in a state comparable to that of hypertext before the advent of the web … it is clearly a good idea, and some very nice demonstrations exist, but it has not yet changed the world. It contains the seeds of important applications, but to unleash its full power it must be linked into a single global system.