Introduction to the Semantic Web

Jim Hendler
Rensselaer

http://www.cs.rpi.edu/~hendler
The Fellowship of the (Semantic) Web
Ontological Conundrum

• The progress of the Semantic Web has been hampered by significant confusion as to what an ontology, and especially a Web ontology is.
  – Two separate visions (or perhaps two end points on what are a continuum) have caused significant confusion
• And the confusion blurs an important message
  – Both uses have proven valuable in the real world!!
• Our goal in this Tutorial is to reduce this confusion
Ontology: the "Expressive" view

- **Ontology** as Barad-Dur (Sauron's tower):
  - Extremely powerful!
  - Patrolled by Orcs
    - Let one little hobbit in, and the whole thing could come crashing down
Ontology:

- Ontology as Barad-Dur (Sauron's tower):
  - Extremely powerful!
  - Patrolled by Orcs
  - Let one little hobbit in, and the whole thing could come crashing down

Decidable Logic basis
Inconsistency is the bane of this view

1537 classes, 1 modeling error = failure!
ROI: Reasoning over (Enterprise) data

- This "big O" Ontology finds 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
ontology: the RDFS view

• ontology and the tower of Babel
  – We will build a tower to reach the sky
  – We only need a little ontological agreement
    • Who cares if we all speak different languages?
ontology: the RDFS view

• ontology and the tower of Babel
  - We will build a tower
  - We only need a little ontological agreement
  - Who cares if we all speak different languages?

Genesis 11:7 Let us go down, and there confound their language, that they may not understand one another's speech. So the Lord scattered them abroad from thence upon the face of all the earth: and they left off to build the city.
Boundaries are the bane of this view

Tabulator and Linked Open Data
ROI: Web 3.0

• The "small o" ontology finds 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!*
O asks: how can you ignore soundness?

- Twine recommends some people I may want to connect to
  - What is correctness in this case?
    - If I find some folks I like this way, I use twine more. Surprises can be fun.
    - But if it does a "bad" job, I may go elsewhere
o asks O: Why do you need expressiveness?

- Often "folksonomy" isn't enough!

Which one do you want your doctor to use?
A big problem for O

- Ontology mapping
Is not a big problem for o

Slogan: A little semantics goes a long way
A big problem for o

• What do we do with all this stuff?

* The primary goal is to show how they add value to the very large triple store. This can involve anything from helping people figure out what is in the store via browsing, visualization, etc; could include inferencing that adds information not directly queryable in the original dataset; could involve showing how ontological information could be tied to part(s) or the whole of the dataset; etc.
  * The tool or application has to make use of at least a significant portion of the data provided by the organizers.
  * The tool or application is allowed to use other data that can be linked to the target dataset, but there is still an expectation that the primary focus will be on the data provided.
  * The tool or application does not have to be specifically an end-user application, as defined for the Open Track Challenge, but usability is a concern. The key goal is to demonstrate an interaction with the large data-set driven by a user or an application. However, given the scale of this challenge, solutions that can be justified as leading to such applications, or as crucial to the success of future applications, will be considered.

(ISWC 2008 - Open Web, Billion Triple Challenge -

A big problem for o

• What do we do with all this stuff?

* The primary goal is to show how they add value to the very large triple store. This can involve anything from helping people figure out what is in the store via browsing, visualization, etc; could include inferencing that adds information not directly queriable in the original dataset; could involve showing how ontological information could be tied to part(s) or the whole of the dataset; etc.

* The tool or application has to make use of at least a significant portion of the data provided by the organizers.

* The tool or application is allowed to use other data that can be linked to the target dataset, but there is still an expectation that the primary focus will be on the data provided.

* The tool or application does not have to be specifically an end-user application, as defined for the Open Track Challenge, but usability is a concern. The key goal is to demonstrate an interaction with the large data-set driven by a user or an application. However, given the scale of this challenge, solutions that can be justified as leading to such applications, or as crucial to the success of future applications, will be considered.

(ISWC 2008 - Open Web, Billion Triple Challenge -

Is well understood in O

Slogan: Knowledge is power
We use the same word…

**Sem Web Modeling**

<table>
<thead>
<tr>
<th>Graph</th>
<th>Labeled graph</th>
<th>Graph + limited logic</th>
<th>Logic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ontology</td>
<td>Data Dictionary</td>
<td>Ontology</td>
<td>Ontology</td>
</tr>
<tr>
<td>RDF</td>
<td>RDF Schema</td>
<td>OWL</td>
<td>KIF?</td>
</tr>
</tbody>
</table>
But $O \neq o$
Why does this matter

- Different issues of concern
  - Confuses messaging

- Effort is spent in different parts of the space
  - i.e. scaling vs. modeling
    - Leads to confusion in costs, esp. for interested parties
    - Starting out: You must know which O/o you're going after

- Different "first-concern" tools for the different models
  - Big O: ontology creation and modeling
  - Small o: triple store and SPARQL

- …
Tensions

- There are also some serious tensions between these models
  - Base in RDF (links) vs. XML (validation)
  - Soundness and Completeness
    - Big O: Mandatory
    - Small o: Impossible
  - Consistency impossible to maintain in large scale distributed efforts
    - Error, Disagreement, Fraud
- Business Model
  - Enterprise v. Web Scale
Not Irreconcilable Differences

Cf. Cleveland Clinic "Semantic DB" effort

OR ≠ XOR
Today you'll hear about

- Ontologies
  - OWL
  - Ontology engineering
  - Ontology Design

- Using Semantics - principles
  - Semantic Interoperability
  - Semantic Web Services

- Using Semantics - applications
  - Semantic Search
  - Linked Data
  - Semantic Web Applications
And now…

QuickTime™ and a decompressor are needed to see this picture.

On with the show!