The Fellowship of the (Semantic) Web

Prof. J. Hendler

RPI

HENDLER @ CS.RPI.EDU
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!!

• My goal in this talk is to try to reduce this confusion
  – Which could be a real first for a college Professor!
Outline

• Intro (now done)
• some imprecise analogies for motivation
• The kinda technical stuff
• Some shameless self-promotion
• Boffo Conclusion
Ontology: the OWL DL 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
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?

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!*
ROI: Web 3.0

- ~2006: Web app developers discover the Semantic Web
O asks o: 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 browsing, visualization, etc. 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;
* 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 may 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 key goal is to demonstrate an interaction with the largest of this challenge, solutions that can be justified as leading to such applications, or as crucial to the success of future applications (ISWC 2008 - Open Web, Billion Triple Challenge -)

Is well understood in O

Slogan: Knowledge is power
We use the same word…

<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

• ...

Rensselaer
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
Which is why RDFS/OWL matter

• From the Original W3C OWL Faq*
  – Q. How is OWL different from earlier ontology languages?
  – A. OWL is a Web Ontology language.
    • Ability to be distributed across many systems
    • Scalable to Web needs
    • Compatible with Web standards for accessibility and internationalization.
    • Open and extensible

• Interoperability is lost if these two towers grow too far apart
  – It is important that RDFS and OWL remain WEB languages
    • Which is much harder in "O" than "o"
    • (W3C members: make sure your AC rep is watching the OWL 2 space with this in mind)
Where can I learn more?

Or Ask the Experts

- Submit your questions on line for Tim, Nova, Wendy and Nigel
  - Vote on the ones you want to hear
- Watch the Webcast
- Or show up in person
  - We're buying the drinks!

http://tw.rpi.edu/launch/
Boffo Conclusion

• There's no ontology ontology
  – The term is used in many ways in our community
• This causes great confusion
  – And hurts our technology adoption
• The Web is a big place
  – Room for more than one vision
  – But they must play nice to get the network effect
• Which leads to…
... the Semantic **WEB** Vision

Interoperability trumps homogeneity ever time!