Round Trip Transformation, (starting from OBO)

Prof. Daniel P. Miranker, presenting for
Syed Hamid Tirmizi
Department of Computer Science
The University of Texas at Austin
hamid@cs.utexas.edu
OBO & OWL: The Common Standard Mapping

The result of a standards process
  – but without a parent body and associated formal process

Contributors:

• Syed Hamid Tirmizi (Univ. of Texas at Austin)
• Stuart Aitken (Univ. of Edinburgh)
• Dilvan Moreira (Univ. of São Paulo)
• Chris Mungall (Lawrence Berkeley National Lab.)
• Juan Sequeda (Univ. of Texas at Austin)
• Nigam H. Shah (Stanford University)
• Daniel P. Miranker (Univ. of Texas at Austin)
Goal: Interoperability Between Two Worlds

OBO: Open Biomedical Ontology Language

- Began as the infrastructure for the Gene Ontology (GO)
  - Likely the world’s best known and most used ontology
  - Operational Semantics
- ~100 biomedical ontologies of various kinds

OBO Foundry
http://www.obofoundry.org/
US National Institutes of Health, (NIH), sponsors

• OBO Biomedical ontologies:
  – Model Organisms (Yeast, Drosophila, Mouse, Human…)
    • Anatomy
    • Development
    • Phenotypes
  – Health
    • Human disease nomenclature
    • Including vocabularies that are part of U.S. gov’t health care system.
  – Standards
    • Relations Ontology, a form of upper ontology

• Biomedical ontologies often viewed as an ends.
OWL: Web Ontology Language for the Semantic Web

- Semantic Web, a means
  - Inference
    - Formal semantics
    - GUIDs
- Expansive support
  - Piecemeal
    - Academic research groups, various funding sources
    - Companies, various sizes, various funding sources
  - Shepherded by the W3C
- Some important biomedical ontologies are in OWL
  - The NIH, National Cancer Institute (NCI) Thesaurus
  - BioPAX (Biological Pathways Exchange)
The Problem: OBO or the Semantic Web!

• In the Miranker lab: The Morphster project

• Morphster: Image driven ontology editing
  – Productivity tool for systematic biologists
  – Embodies knowledge capture, data integration, workflow
    ● Single taxon descriptions
    ● Morphological phylogenetic study

• Morphster has to do both:
  – Biodiversity data, (GBIF), OWL adherents
  – Anatomy and phenotypes, OBO
An Informally Gathered Standards Process

- Each contributor has their own motivating story.
- Each contributor initially developed a mapping system.
The Process

• At the start of 2007,
  – An invitation was broadcast to all interested parties.
  – OBO constructs were enumerated.
  – A shared Google spreadsheet was created,
    • One tab for each contributors mapping
    • One tab for the consensus.

• A wiki page was created for discussion on the mapping

• Artifacts of this process can be found at
Thesis:
The organization of the Semantic Web hierarchy (layer cake) transcends the Semantic Web
Thus, the Semantic Web hierarchy itself can be leveraged to study other systems.
Two layer cakes prove useful for identifying:
- What is the same
- What is different

about the two systems
Miranker lab has done this for SQL as well

Basis for Ultrawrap: SPARQL endpoint for relational databases
A DR2 like system, except
- completely automatic
- leverages SQL optimizer
## Mapping Examples

*What is the same.*

<table>
<thead>
<tr>
<th>OBO</th>
<th>OWL</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Typedef]</td>
<td>&lt;owl:TransitiveProperty</td>
</tr>
<tr>
<td>id: part_of</td>
<td>rdf:about=&quot;...#part_of&quot;&gt;</td>
</tr>
<tr>
<td>name: part of</td>
<td><a href="">rdfs:label</a>part of&lt;/rdfs:label&gt;</td>
</tr>
<tr>
<td>is_transitive: true</td>
<td>&lt;/owl:TransitiveProperty&gt;</td>
</tr>
</tbody>
</table>

**Example A** Simple transformations: name, transitivity

<table>
<thead>
<tr>
<th>Term</th>
<th>OWL</th>
</tr>
</thead>
<tbody>
<tr>
<td>id: ZFA:0000434</td>
<td>&lt;owl:Class rdf:about=&quot;...#ZFA_0000434&quot;&gt;</td>
</tr>
<tr>
<td>name: skeletal system</td>
<td><a href="">rdfs:label</a>skeletal system&lt;/rdfs:label&gt;</td>
</tr>
<tr>
<td>is_a: ZFA:0001439</td>
<td>&lt;rdfs:subClassOf</td>
</tr>
<tr>
<td></td>
<td>rdf:resource=&quot;...#ZFA_0001439&quot;/&gt;</td>
</tr>
<tr>
<td></td>
<td>&lt;/owl:Class&gt;</td>
</tr>
</tbody>
</table>

**Example B** Transformation of ‘is-a’
Mapping Examples

Example C  Transformation of a relationship

<table>
<thead>
<tr>
<th>Term</th>
<th>Transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>id: ZFA:0001439</td>
<td>&lt;owl:Class rdf:about=&quot;...#ZFA_0001439&quot;&gt;</td>
</tr>
<tr>
<td>name: anatomical system</td>
<td><a href="">rdfs:label</a>anatomical system&lt;/rdfs:label&gt;</td>
</tr>
<tr>
<td>relationship: part_of ZFA:0001094</td>
<td><a href="">owl:Restriction</a></td>
</tr>
<tr>
<td></td>
<td>&lt;owl:onProperty rdf:resource = &quot;...#part_of&quot; / &gt;</td>
</tr>
<tr>
<td></td>
<td>&lt;owl:someValuesFrom rdf:resource = &quot;...#ZFA_0001094&quot; /&gt;</td>
</tr>
<tr>
<td></td>
<td>&lt;/owl:Restriction&gt;&lt;/rdfs:subClassOf&gt;</td>
</tr>
<tr>
<td></td>
<td>&lt;/owl:Class&gt;</td>
</tr>
</tbody>
</table>

Example D  Transformation of obsolete term

<table>
<thead>
<tr>
<th>Term</th>
<th>Transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>id: ZFA:0000437</td>
<td>&lt;owl:Class rdf:about=&quot;&amp;oboInOwl;ObsoleteClass&quot;/&gt;</td>
</tr>
<tr>
<td>name: stomach</td>
<td>&lt;owl:Class rdf:about=&quot;...#ZFA_0000437&quot;&gt;</td>
</tr>
<tr>
<td>isObsolete: true</td>
<td><a href="">rdfs:label</a>stomach&lt;/rdfs:label&gt;</td>
</tr>
<tr>
<td></td>
<td>&lt;rdfs:subClassOf rdf:resource=&quot;&amp;oboInOwl;ObsoleteClass&quot;/&gt;</td>
</tr>
<tr>
<td></td>
<td>&lt;/owl:Class&gt;</td>
</tr>
</tbody>
</table>
Differences

Required in OWL, missing from OBO

1. Globally unique identifiers
   - OBO has a local identification scheme for its concepts
   - OWL classes and properties need global IDs
   - Special consideration required to complete roundtrip

Missing from OWL, part of OBO

2. Synonyms
   - Various kinds (possibly) emerging from biomedical domain
   - Lack of semantics and documentation creates problems

3. Subsets
   - OWL does not have an exact match to this OBO construct
Mapping OBO IDs to URIs - I

• Any string can be an OBO identifier
• Preferred ID syntax: <IDSPACE>:<LOCALID>
• ‘idspace’ tag can be added to an OBO ontology header to make GUID possible
  – E.g. “idspace: GO http://www.go.org/owl#”
  – Read: GO is an ID space that refers to the given URI

  – Example: **GO:0000001** maps to
    http://www.go.org/owl#GO_0000001
Mapping OBO IDs to URIs - II

• ID space is not defined in the ontology header
  – Each ontology also has a default base URI
  – `<default_base_URI>`: http://www.bioontology.org/…#

• ID is of the form: `<IDSPACE>:<LOCALID>`
  – Example: SO:0000001 maps to
    `<default_base_URI>SO_0000001`

• ID is of the form: `<LOCALID>`
  – Example: ABC001 maps to
    `<default_base_URI>UNDEFINED_ABC001`
Implementation

• Java implementation is a part of official Gene Ontology source
  – http://sourceforge.net/projects/geneontology/
  – Also in tools like OBO-Edit and Morphster

• Web service available for online conversion
  http://www.cs.utexas.edu/~hamid/oboowl.html
  http://www.youtube.com/watch?v=GYnFMq0W_8g

• Already converted OBO Foundry ontologies
  http://www.berkeleybop.org/ontologies/
Towards Formal Semantics for OBO

• OWL is formally defined, while OBO has operational semantics

• A semantics document for OBO can be mechanically derived using the mapping and the semantics for corresponding OWL elements
• OWL has a larger construct set than OBO

• To make roundtrips possible, any editing in OWL must be limited to a defined set of constructs
  – i.e. the constructs used in the mappings

• We call this subset \textit{OWL-Bio}
Interconnecting OBO and the Semantic Web

• Roundtrip transformations on arbitrary ontology
• Both OBO and OWL are moving targets.

• OWL-Bio, raise awareness
  
  “A problem clearly stated is a problem half solved”
  
  Dorothea Brande
Thank you
Any questions?