Exploring and Using the Semantic Web

Mathieu d’Aquín
KMi, The Open University – m.daquin@open.ac.uk
Exploring the Semantic Web

What??

Ontologies

Vocabularies

Linked Data

RDF documents
Example: Exploring DBPedia

- “DBpedia is a community effort to extract structured information from Wikipedia and to make this information available on the Web. DBpedia allows you to ask sophisticated queries against Wikipedia, and to link other data sets on the Web to Wikipedia data.” -- [http://wiki.dbpedia.org/About](http://wiki.dbpedia.org/About)
- Provides a huge, multi-domain knowledge base, accessible through the Web, using Semantic Web standards, mainly SPARQL ([http://dbpedia.org/sparql](http://dbpedia.org/sparql))
- Example: “what are the names of German musicians who were born in Berlin?”
  
  ```sparql
  PREFIX dbo: http://dbpedia.org/ontology/
  SELECT ?name
  WHERE {?
  ?person foaf:name ?name . }
  ```
Example: Exploring DBPedia

- “DBpedia is a community effort to extract structured information from Wikipedia and to make this information available on the Web. DBpedia allows you to ask sophisticated queries against Wikipedia, and to link other data sets on the Web to Wikipedia data.” -- http://wiki.dbpedia.org/About
- Provides a huge, multi-domain knowledge base, accessible through the Web, using Semantic Web standards, mainly SPARQL (http://dbpedia.org/sparql)
- Example: “what are the names of German musicians who were born in Berlin?”
  
  PREFIX dbo: http://dbpedia.org/ontology/
  SELECT ?name
  ?person foaf:name ?name . }
  → Alexander Marcus, Klaus Voormann
Linked Open Data ( LOD) cloud:
http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData
Other Tools to Explore the Linked Data Cloud…

• Instance of OpenLink Virtuoso on amazon EC2 loaded with data from the LOD cloud
• And also semantic web browsers such as Tabulator, RKBExplorer, etc.

• But… the LOD cloud is not everything…
• Many ontologies, vocabularies and datasets of various sizes and complexities are being created all the time
Finding Semantic Information on the Web
With Semantic Web Search Engines

Watson:  http://watson.kmi.open.ac.uk

Sindice:  http://sindice.com

Swoogle:  http://swoogle.umbc.edu/

Falcons:  http://iws.seu.edu.cn/services/falcons/objectsearch/

SWSE:  http://swse.deri.org/
Example: Watson (architecture)
Example: Watson (interface)

http://watson.kmi.open.ac.uk
So, what can you find on the Semantic Web?

- **Size:**
  - LOD Cloud: 4.5 Billion Triples (estimated)
  - Sindice: claim 54.65 million documents (June 2009)
- **An analysis of a sample of the Watson collection (not up to date!):**

<table>
<thead>
<tr>
<th>Language</th>
<th>Complexity</th>
<th>Size of documents</th>
<th>Topic coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>DL</td>
<td>OWL</td>
<td>Total</td>
<td>DL OWL Full</td>
</tr>
<tr>
<td>$	ext{AC}(D)$</td>
<td>21375 (84%)</td>
<td>3644 (59%)</td>
<td>3365 (78%)</td>
</tr>
<tr>
<td>$	ext{AC}$</td>
<td>2455 (10%)</td>
<td>1406 (23%)</td>
<td>281 (6.5%)</td>
</tr>
<tr>
<td>$	ext{ACH}(D)$</td>
<td>293 (1%)</td>
<td>105 (1.5%)</td>
<td>68 (1.5%)</td>
</tr>
<tr>
<td>$	ext{ALC}$</td>
<td>105 (&lt;1%)</td>
<td>94 (1.5%)</td>
<td>44 (1%)</td>
</tr>
<tr>
<td>$	ext{ACH}$</td>
<td>101 (&lt;1%)</td>
<td>54 (&lt;1%)</td>
<td>28 (&lt;1%)</td>
</tr>
<tr>
<td>$	ext{ALC}$</td>
<td>102 (&lt;1%)</td>
<td>43 (&lt;1%)</td>
<td>27 (&lt;1%)</td>
</tr>
</tbody>
</table>
Using It

How??

Using It

Through SPARQL

Linking to Other Datasets

Reusing in Ontologies

Through API

Building Applications!!!
What to use?

• SPARQL endpoints, and such…
  – DBPedia’s SPARQL endpoint
  – Virtuoso LOD instance
  – All other LOD nodes
  – The Talis Platform
  – Etc.

• Semantic Web search engines services and APIs
  – **Watson**: complete, open SOAP and REST API, with Java client
  – **Sindice**: REST API for entity lookup, etc.
  – **Swoogle**: simple search REST API
  – **Falcons**: simple search REST API
  – **SWSE**: SPARQL endpoint!

• Other nice (online!) tools:
  – **OpenCalais Semantic Proxy**: for entity recognition and annotation in web pages
  – **Textwise SemanticHacker**: for document classification
  – **Triplify, D2R, etc.**: for transforming databases into semantic web resources
  – …
Example: Watson API

- Provide Semantic Web application developers with the ability to efficiently:
  - Locate (find) Semantic Web documents online using advanced search functions
  - Explore the documents, automatically extracted metadata and content
  - Query the documents, to exploit online knowledge in an homogeneous way
- In a set of lightweight APIs, and without having to download the data or use any other dedicated infrastructure.
Some applications we developed

Ontology Reuse: The Watson Plugin

Semantic Relation Discovery: Scarlet

Question Answering: PowerAqua

Folksonomy Enrichment

Semantic Browsing: PowerMagpie

And also: Word sense disambiguation, Query Expansion, Synonym Discovery…
A demonstrator: Wahoo!

Find on the Semantic Web terms to specialize or generalize a query to a search engine (Yahoo).
The technical stuff…

- Find ontologies containing the keyword “researcher”
  http://watson.kmi.open.ac.uk/API/semanticcontent/keywords?q=researcher

- … exactly “researcher” in the label or id of a class
  http://watson.kmi.open.ac.uk/API/semanticcontent/keywords?q=researcher&scope=LN+Label&ent=Class&match=Exact

- Find entities corresponding to “researcher” in ontology

- Find subclasses and superclasses of an entity

- The rest is interface stuff and call to Yahoo!
Other Example: Scarlet
### Relation Discovery: Scarlet

**How does cholesterol relate to OrganicChemical?**

**Strategy:**
- Use one ontology
- Use more ontologies

**Other parameters:**
- Find first relation
- Find all relations
- Use inheritance depth 1

The following relations were found:

- cholesterol subClass OrganicChemical

**Because:**
- Cholesterol - subClassOf -> Lipid
  - In: http://www.co-ode.org/ontologies/galen
- Lipid - subClassOf -> OrganicChemical
  - In: http://onto.cs.yale.edu:8080/umls/UMLSinDAML/NET/SRSTR.daml

**Examples:**
- River vs. waterway
- Cocaine vs. narcotic
- Water vs. Solid
- Branch vs. Tree
- Coal vs. Industry
- Fish vs. Lobster
- Cholesterol vs. OrganicChemical
- Apple vs. Meat
Other Example: The Watson Plugin

- Choose an entity to search
- Get entities from online ontologies
- Integrate statements into the edited ontology
More?

Exploring, using… and contributing to the Semantic Web!

- Cupboard: an ontology hosting system
  http://cupboard.open.ac.uk
- Using semantic web resources to improve the quality on the Semantic Web. Manual rating of ontologies, trust network between ontologies and ontology providers, improving links between ontologies and datasets through reuse.

A lot left to do!

- Understanding the notions of relevance of Semantic Web information, of agreement between ontologies…
- Detecting and managing the relations between ontologies. Making the network of resources explicit…
Thanks!

Hands-on session this afternoon
Using Semantic Web search engines
Using Semantic Web APIs
Building ontologies by reuse
Publishing an ontology with Cupboard

Now:
questions? Comments?
Other topics of interest?

Contact: m.daquin@open.ac.uk