RKBExplorer.com:

Anatomy of a Semantic Web Application

Hugh Glaser & Ian Millard
4th December 2008, Seoul
Context

• CSAKTiveSpace
  – AKT Project
  – 2003 Semantic Web Challenge winner

• ReSIST - EU Network of Excellence in Resilient Systems
  – Knowledge-enabled infrastructure

• KISTI work on Co-reference
  – 2008
User Interaction

- Semantic MediaWiki
- Custom form interfaces
- Google Maps
- Raw Knowledge Browser

- RKBExplorer
- Why do you think that? information
Focusing on a Person
Knowledge Sources

- Partners
- Publications
- Funding Agencies
- Project Wiki
- Courseware
- Resilient-Explicit Computing

- Wide range, don’t just look where you expect to find
Some Underlying Sources

<table>
<thead>
<tr>
<th>acm.rkbexplorer.com</th>
<th>italy.rkbexplorer.com</th>
</tr>
</thead>
<tbody>
<tr>
<td>budapest.rkbexplorer.com</td>
<td>kaunas.rkbexplorer.com</td>
</tr>
<tr>
<td>citeseer.rkbexplorer.com</td>
<td>kisti.rkbexplorer.com</td>
</tr>
<tr>
<td>cordis.rkbexplorer.com</td>
<td>laas.rkbexplorer.com</td>
</tr>
<tr>
<td>courseware.rkbexplorer.com</td>
<td>lisbon.rkbexplorer.com</td>
</tr>
<tr>
<td>darmstadt.rkbexplorer.com</td>
<td>newcastle.rkbexplorer.com</td>
</tr>
<tr>
<td>dblp.rkbexplorer.com</td>
<td>nsf.rkbexplorer.com</td>
</tr>
<tr>
<td>deepblue.rkbexplorer.com</td>
<td>pisa.rkbexplorer.com</td>
</tr>
<tr>
<td>deploy.rkbexplorer.com</td>
<td>rae2001.rkbexplorer.com</td>
</tr>
<tr>
<td>epsrc.rkbexplorer.com</td>
<td>resex.rkbexplorer.com</td>
</tr>
<tr>
<td>eurecom.rkbexplorer.com</td>
<td>roma.rkbexplorer.com</td>
</tr>
<tr>
<td>ft.rkbexplorer.com</td>
<td>southampton.rkbexplorer.com</td>
</tr>
<tr>
<td>ibm.rkbexplorer.com</td>
<td>ulm.rkbexplorer.com</td>
</tr>
<tr>
<td>ieee.rkbexplorer.com</td>
<td>unlocode.rkbexplorer.com</td>
</tr>
<tr>
<td>irit.rkbexplorer.com</td>
<td>wiki.rkbexplorer.com</td>
</tr>
</tbody>
</table>

Range from a few 100 to more than 10,000,000 “facts”
For example

- Statistics for repository kisti.rkbexplorer.com
  - Last data assertion 2008-09-18 17:16:41
  - Number of triples 12815162
  - Number of symbols 3239105
  - Size of RDF dataset 671M
Semantic Web/Linked Data issues

- The system supports state of the art facilities:
  - Browsing
  - Resolvable URIs
  - SPARQL endpoint
  - CRS (Co-reference knowledge)
  - RDF publishing
  - Semantic Sitemap
  - On a separate domain
Communication

- Ontologies
  - General Scientific Endeavour
  - Domain-specific
  - Support (geospatial, etc)
- Open Local Knowledge – HTTP
  - Resolvable URIs
  - SPARQL
- Uses Remote Knowledge
  - Resolves URIs with caching
Components 1

• Semantic Web infrastructure throughout
• Triplestore for each source
  – Putting the Web in Semantic Web
  – Stores RDF – (Subject, Predicate, Object)
  – We use 3store
• Linked Data
  – 303 and content negotiation architecture with caching
Components 2

- Co-Reference Subsystem
  - CRS – more later
- Community of Practice Analysis
  - Why do you think that?
- Ontology Mapping
  - Dealing with other Ontologies
- NLP for text classification
- Caching everywhere
Components 3

- Application Middleware
  - URI Equivalence Closure
  - RDF Graph Closure
- Semantic Sitemap
  - Facilitate Search Engines
Co-Reference

• Co-Reference is a Big Problem
  – Identifying multiple URIs for one resource
  – Rejecting incorrectly conflated resources
  – Publishing
  – Using

• Coldstart
  – A serious problem
  – Nothing is linked to anything
Complete Co-Reference Information

This service computes the equivalence class within the known URIs for a specified URI by consulting all relevant CRS knowledge bases.

Equivalent URIs...

1. (Canon) http://acm.rkbexplorer.com/id/person-407157
4. http://dblp.rkbexplorer.com/id/people-1ec5a600299222dd6374695e5f21405-906423eb148125a96e5c573dc5a15e43c

The following diagram shows the interconnectivity between the CRS knowledge bases which maintain the context-dependent representation of coference for each of the RKBExplorer domains.

Seungwoo Lee

Showing information queried from all repositories ...
CRS – Consistent Reference Service

- A service to manage and publish co-referent information
- Identify co-referent pairs using a set of tools
- Assert into the CRS
- Query the CRS
  - URI$_i$ -> { URI$_1$, ..., URI$_i$, ..., URI$_n$ }
- Recommend a Canon
CRS continued

- CRS Policies are defined by context
  - Often one per Triplestore
  - Can be many per Triplestore for different purposes
  - May not be associated with a particular Triplestore
- Maintenance
  - Provenance
  - Rollback
- Can be used to infer owl:sameAs
Dealing With Non-SPARQL KBs

• The RKBExplorer application uses SPARQL to query the KBs
  – But needs to access data from KBs that only offer resolvable URIs
• So resolve such a URI
• Cache the RDF with associated resolved RDF locally
• Query the local cache
Dealing With Different Ontologies

• The RKBExplorer application uses a particular ontology
  – Some KBs will use different ontologies
  – Eg kisti.rkbexplorer.com

• One solution
  – Represent the ontology relationship in RDF (as far as possible)
  – Resolve the URI through the mapping service to get RDF in the required ontology
Concluding Remarks

• Major Data Fusion using Semantic Web Technologies
• Many things can be cast in a Semantic Web framework
• Linked Data works pretty well
• RDF works pretty well
• A little Ontology goes a long way
• Co-Reference is the biggest problem
  – But is tractable