Searching on the Open Semantic Web using a URI Identity Management Approach
Afra Jaffri [a.o.jaffri@ecs.soton.ac.uk]
Dependable Systems and Software Engineering group
Electronics and Computer Science
University of Southampton

Aim of Research
In order for Semantic Web search engines and other applications to work with the increasing amount of RDF data that is being made available on the Web, there needs to be a URI management system that will track URI usage and coreference between URI's. The system will:

- Detect and group together URI’s referring to the same resource
- Integrate with Linked Data
- Provide a query mechanism so groups of URI’s can be quickly discovered by Semantic Web agents
- Track the Provenance of URI’s
- Provide search functionality through the use of enhanced keywords mapped to URI’s

Why do we need URI Identity Management?
URI’s for ‘Hugh Glaser’:
- http://acm.rkbexplorer.com/rdf/resource/P112732
- http://www.ics.soton.ac.uk/info/12person-00021

URI’s for ‘Spain’:
- http://dbpedia.org/resource/Spain
- http://www4.wiwiss.fu-berlin.de/factbook/resource/Span
- http://www.geonames.org/25130769
- http://www.4.wiwiss.fu-berlin.de/eurostat/resource/countries/Esp%20s

What is the problem with owl:sameAs?

Solution

Consistent Reference Services

URI’S referring to the same resource are grouped together in ‘Bundles’
A bundle has properties:

- Coref:hasCanonicalReference – One URI in a bundle can be made to be the canonical representation i.e. The preferred URI
- Coref:hasEquivalentReference – The URI’s in a bundle are grouped together using this predicate
- Coref:updatedOn – The date of the last update to the bundle

Knowledge Manager
Matches keywords to URI’s using rdfs:label and similar properties

Evaluation Method
- Also precision and recall metrics on URI’s of known resources in order to determine coverage of URI coreference
- Perform usability testing on input and output interfaces
- Compare results between linkage with owl:sameAs and linkage with CRS’s
- Normalize a theory of URI linking

This work is supported under the ReSIST Network of Excellence (NoE) which is sponsored by the information Society Technology (IST) priority of the EU Sixth Framework programme (FP6) under contract number IST-4-026764-NOE