

Managing Co-reference on the Semantic Web

The Co-reference Resolution Service

Hugh Glaser, Afraz Jaffri, Ian Millard
20 April 2009

Background

- ‘Resilience for Survivability in IST’ Project
 - Dependable Systems, Fault Tolerance, Security
 - We have developed an “Explorer” application to facilitate exploration and discovery of related people, publications, domain concepts, projects and organisations.
- Multiple data sources
 - Separate triple-store for each source
 - SPARQL endpoint
 - Linked Data compliant URI resolution

Problem

- Significant overlap of people, publications, etc
- Need some way of mapping between co-reference identifiers
- Algorithms for detecting co-references are often based on heuristics, not absolutely certain
 - sameAs too strong a semantics? not appropriate?
 - different contexts?
- Require easy management, maintenance, updates...
 - loose notion of provenance

Co-reference Resolution Service

- Manage co-reference knowledge as first class entity
 - represent “bundles” of equivalent URIs
 - nominate a canonical identifier for each bundle (canon)
- When new co-references detected, merge bundles together
 - recording history, enabling rollback
- Deprecation of old URIs
- Re-publishing of co-reference / equivalence data
- Simple, lightweight implementation and interfaces

Example: Languages of the World

- Besides ReSIST, collaborating with a number of others
- Bernard Vatant, LingVoj
 - Linked Data repository of languages
- David Baxter, OpenCyc
 - Linked Data resources describing ontology of everyday common concepts
 - Versioning system, URIs contain date component
 - Used CRS to maintain external co-reference links
 - Deprecation for older versions of concepts

CRS “global closure” service

- <http://www.rkbexplorer.com/sameAs/?uri=.....>
- Wrapper around “follow your nose” expansion
 - ask appropriate domain CRS if known
 - or ask all CRSes if no known starting point
- Return equivalence data in a number of formats
- Co-references for 17 million URIs, in 5 million bundles

