

# An Infrastructure for Managing URI Synonymity on the Semantic Web

Afraz Jaffri, Hugh Glaser and Ian Millard

Dependable Systems and Software Engineering Group  
School of Electronics and Computer Science  
University of Southampton  
{a.o.jaffri,hg,icm}@ecs.soton.ac.uk

**Abstract.** The Semantic Web vision involves the production and use of large amounts of RDF data. There have been recent initiatives amongst the Semantic Web community, in particular the Linking Open Data activity and our own ReSIST project, to publish large amounts of RDF that are both interlinked and dereferenceable. The proliferation of such data gives rise to millions of URIs for non-information resources such as people, places and abstract things. Our Consistent Reference Services provide a standard way of managing multiple URIs and finding URI equivalences. The CRS has been designed for use with all linked data Providers and is currently running on a live linked data site.

## 1 Background

The issue of identity has become a central area of Semantic Web research. Whilst existing in theory for a number of years, practical solutions are now required to solve the URI Identity Crisis [1]. The Linking Open Data initiative has led to an explosion in the number of URIs used to identify different entities, which has also provided new impetus into finding a solution for managing URI coreference. The increase in the number of information sources being exposed as RDF has also led to an increase in the number of URIs used to identify different entities. It is often the case that data in different repositories will hold information regarding identical entities. The multiplicity of URIs leads to the problem of *coreference*, where different URIs are used to describe the same entity.

We subscribe to the belief that the meaning of a URI may change according to the context in which it is used [2]. For example the URIs that refer to Spain given above could refer to -Spain the political entity or -Spain the geographic location or -Spain the football team. Some people would be happy to use each URI interchangeably because they do not care about the precise definition, whereas others will want a URI that specifically matches their intended meaning. There is a requirement to have some form of a system that deals with URIs about the same resource that are not exactly identical. The semantics of *owl:sameAs* are too strong and other alternatives like *rdfs:seeAlso* do not fit the intended purpose. Such a requirement is vital if data is to

be cleanly linked together in a consistent fashion. The CRS system that we have developed fulfills just such a requirement.

## 2 Consistent Reference Service

The Consistent Reference Service (CRS) has been created in order to manage coreference between the millions of URIs that are accumulating on the Semantic Web. The CRS introduces the concept of a *bundle* to group together resources that have been deemed to refer to the same concept within a given context. Different bundles may be used to group together URIs of the same resource in different contexts.

Each CRS can use different algorithms to identify equivalent resources. For example, the algorithms to detect equivalence amongst authors are different from the algorithms used to detect equivalence between countries.

The CRS can be treated as any other knowledge base, in that it contains knowledge about a particular URI. Our infrastructure implements the current best practice on how to serve linked data and uses cool URIs. As an example we use the URI <http://southampton.rkbexplorer.com/id/person-21> to represent the non-information resource, ~~“Hugh Glaser”~~

At present in the Linked Data world, no procedure exists for finding a complete set of synonyms for a given URI. We use URIs of example locations of CRSes with a possible set of URI synonyms in each. Our infrastructure enables the easy retrieval of synonyms of a given URI that relies on nothing more than URI dereferencing and ~~“following your nose”~~. The complete CRS system and URI Synonym Finder will be presented in the poster.

## 3 Summary

The issue of coreference is becoming all the more relevant with linked data that is now available online. This has caused a real debate on how to manage coreference in the Semantic Web [3]. Our proposed solution is meant to highlight the issues involved in building a URI management system and how such a system can be used to handle billions of triples of RDF data.

## References

1. Halpin, H. Identity, Reference and Meaning on the Web, Proceedings of the Workshop on Identity, Meaning and the Web (IMW06) at WWW2006, Edinburgh, Scotland.
2. Booth, D. URIs and the Myth of Resource Identity, Proceedings of the Workshop on Identity, Meaning and the Web (IMW06) at WWW2006, Edinburgh, Scotland.
3. W3C Mailing List Discussion Thread, Terminology Question Concerning Web Architecture and Linked Data, <http://lists.w3.org/Archives/Public/semantic-web/2007Jul/0049.html>