Exploring RDF Usage and Interlinking in the Linked Open Data Cloud using ExpLOD

Shahan Khatchadourian

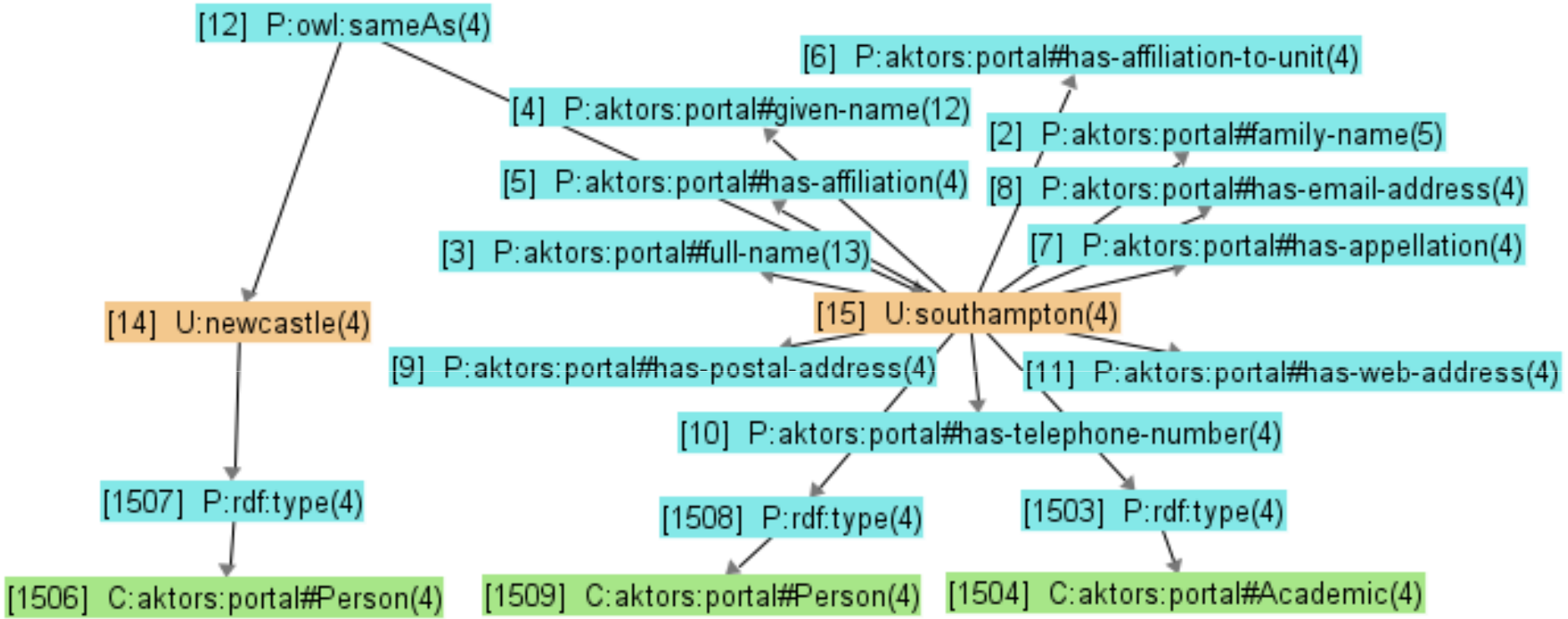
Mariano P. Consens

University of Toronto

Motivation

- Challenge to LOD consumption: understanding the data that is out there
  - How can we describe the instances/entities and the metadata (that is actually *used*)?
    - What classes and predicates describe an instance?
  - How are cloud datasets interlinked?
    - How are equivalent instances described across interlinked datasets?
- ExpLOD: tool to create and explore summaries of the structure of linked data

Sample ExpLOD Summary



Shows: Classes (Person, Academic) and predicates from Southampton with the classes (Person) of the equivalent resources from Newcastle

Also: resource counts, can generate SPARQL

ExpLOD's RDF Summaries

To generate a summary (using bisimulation contraction)

- Do not have to know the schema/ontology beforehand
- Neighbourhoods of interest can be specified using SPARQL (or path regular expressions)
- Granularity can be controlled with OLAP-style hierarchies for labelling (grouping) summary nodes

After summaries (describing the usage) are created

- Summaries can be explored visually and interactively
- Summaries can be exported (eg in RDF) – and can be used to enrich datasets descriptions (VoID, sitemaps)
- Generate SPARQL queries - guaranteed to return non-empty answers

RDF Usage Demonstration

- Which classes are instantiated (jointly)
- Which predicates are instantiated (jointly)
- How classes and predicates are used to describe equivalent instances across interlinked datasets

Summary

- ExpLOD enables powerful exploration that facilitates linked data consumption

  More information in ESWC 2010 paper “ExpLOD: Summary-based Exploration of Interlinking and RDF Usage in the Linked Open Data Cloud”

What’ s next

- Support for query-based labelled bisimulation within SPARQL (a generalization of grouping)
- Implementing on top of map-reduce paradigm for large-scale exploration

Thanks

Paper Reference:

Shahan Khatchadourian, Mariano P. Consens,
ExpLOD: Summary-based Exploration of Interlinking and RDF Usage in the Linked Open Data Cloud,
ESWC, 16 pages, LNCS 6089, 2010

shahan@cs.toronto.edu

consens@cs.toronto.edu