DBpedia linkage analysis leveraging on entity semantics

Author: David Fuchs
Supervisor: Vojtěch Svátek
Department of Information and Knowledge Engineering at University of Economics, Prague

Motivation
- Linked open data are an abundant source of information, but with varying levels of quality. In this work we thus:
  - quantitatively analyse the connectivity of linked open datasets with Dbpedia and
  - study in depth the semantics of artwork entities on Wikidata, analyse their consistency and the consistency of their interlinking with Dbpedia.

Experiments
- analysis of interlinking towards Dbpedia
  - accessibility – 43 % datasets published using Semantic Web access mechanisms (SPARQL, RDF dump, dereferenceable URIs)
  - uniqueness – 0.8 % duplicates, 0.1 % for lexico-linguistic datasets, 1.3 % for encyclopaedic datasets
  - consistency of interlinking – less than 0.1 % overall
  - currency – 50 % of datasets were updated at most 5 years before the data for this work was collected

- analysis of consistency of artwork entities within Wikidata
  - mismatch between assignment to class and attached property
  - properties indicating different classes
  - 3 000 cases of inconsistency found (1.5 % of examined entities), estimated 40 000 inconsistencies (22 % of examined entities)

- analysis of consistency of artwork entities across Wikidata and Dbpedia
  - annotations of DBpedia entities
  - 16 % of examined DBpedia entities inconsistently interlinked with Wikidata

Conclusion
- identified data quality issues
- provided to the DBpedia community a list of entities inconsistently interlinked with Wikidata for data cleaning

Breakdown by the year of last modification of datasets