Towards the Natural Ontology Of Wikipedia

Andrea Giovanni Nuzzolese [1 2], Aldo Gangemi [2 3], Valentina Presutti [2], and Paolo Ciancarini [1 2]

[1]Department of Computer Science and Engineering, University of Bologna, Italy
[2]STLab-ISTC, National Research Council, Italy
[3]LIPN, Université Paris 13, Sorbone Cité, UMR CNRS, France

Motivation

DBpedia [1] has ~4M entities
- 33% of entities are untyped
  - ↳ Limited extensional coverage

YAGO [2] and DBPO [3] are the main reference ontologies for DBpedia
- YAGO is based on Wikipedia categories
- DBpedia is based on Infobox templates
  - ↳ Their domain coverage is constrained and potentially limited

YAGO and DBPO types have different level of granularities
- ↳ It is difficult to reuse DBpedia knowledge with good precision

Goal

To identify **ORA***, i.e., the "natural" ontology of Wikipedia by means of Tìpalo [6], which is built on top of FRED [7]
- By processing natural language definitions of DBpedia entities
  - ↳ ORA reflects the richness of terms used and agreed by the crowds for defining entities in Wikipedia

To obtain an alternative or a complement resource for DBPO and Yago
- ↳ More accurate usage of DBpedia in Semantic Web applications, e.g., mash-up tools, recommendation systems, exploratory search tools, etc.

* ORA is the italian translation of the english NOW

Result

Typed **3,023,890** entities with associated taxonomies of types
- ~79% of DBpedia v3.8 entities are covered
  - ↳ Missing results are due to the lack of matching Tìpalo heuristics

The resulting ontology includes
- **585,474** distinct classes organized in a taxonomy
- **396,375** refs:subClassOf axioms among classes
  - ↳ Polysemy among classes is avoided by means of word-sense disambiguation

25,480 classes are aligned to **20,662** sunsets through owl:equivalentClass axioms
- These classes are also aligned to WordNet supersenses and to a subset of Dolce+DnS Ultra Lite classes through owl:subClassOf axioms
  - ↳ There exists **4,818** synonym classes in the ontology

Example

```
...

dbpedia:The_Marriage_of_Heaven_and_Hell
    a ora:Book_102870092

dbpedia:Book_of_Revelation
    a ora:CanonicalBook_106394865 .
    ora:CanonicalBook_106394865
    rdfs:subClassOf ora:Book_106394865 ;
    rdfs:label "Canonical Book"@en-US .

ora:Book_102870092
    owl:equivalentClass wn:synset-book-noun-2 ;
    rdfs:label "Book"@en-US .

ora:Book_106394865
    owl:equivalentClass wn:synset-book-noun-10 ;
    rdfs:subClassOf wn:supersense-noun_communication ,
    d0:InformationEntity ;
    rdfs:label "Book"@en-US .

...
```

ORA is available for download at http://stlab.istc.cnr.it/stlab/ORA

1. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S. (2009). Dbpedia - a Crystallization Point for the Web of Data. J. Web Sem., 7(3):154-165, 2009.
2. Suchanek, F. M., Kasneci, G., Weikum, G. (2007). Yago: A Core of Semantic Knowledge. In Proceedings of WWW, New York, NY, USA. ACM Press.
3. DBpedia Ontology, *http://dbpedia.org/ontology*
4. Gangemi, A., Navigli, R., Velardi, P. (2003). The ontowordnet project: extension and axiomatization of conceptual relations in wordnet. In WordNet, Meersman, pages 3-7. Springer.
5. Dolce Ultra Lite ontology, *http://www.ontologydesignpatterns.org/ont/dul/DUL.owl*
6. Gangemi, A., Nuzzolese, A.G., Presutti, V. , Draicchio, F., Musetti, A., Ciancarini, P. (2012). Automatic Typing of DBpedia Entities. In Proceedings of ISWC, 1:65-81.
7. Presutti, V., Draicchio, F., Gangemi, A. (2012). Knowledge extraction based on discourse representation theory and linguistic frames. In Proceedings of EKAW: 114-129.