**Improving Multilingual DBpedia: Interlingual Property Mapping**

Mart Busger op Vollenbroek, Olivier Louwaars, Xianchao Zeng

Faculty of Arts

University of Groningen

Abstract

In this project, the target is to improve the multilingual DBpeida. Our approach is to develop a multilingual mapping tool that automatically map properties in different languages into the same ontology. Through the research period, we took the Dutch DBpedia (nl.dbpedia.org) version as an object of study. We found two main problems to be developed in a Dutch DBpedia page. In the original situation, it might either not able to automatically update value under a same property in different language or cause error when searching in a SPARQL query. We attempt to use the semantic web knowledge to build up a mapping method between multilingual properties. The result shows a high feasibility of our attempt in both mapping the multilingual properties and improving the SPARQL search quality.

1 Introduction

(About DBpeida: Introduction of DBpeida )

“ DBpedia is a crowd-sourced community effort to extract structured information from Wikipedia and make this information available on the Web. DBpedia allows users to ask sophisticated queries against Wikipedia, and to link the different data sets on the Web to Wikipedia data. This work may make it easier for the huge amount of information in Wikipedia to be used in some new interesting ways. Furthermore, it might inspire new mechanisms for navigating, linking, and improving the encyclopedia itself.

The English version of the DBpedia knowledge base describes 4.58 million things, out of which 4.22 million are classified in a consistent ontology, including 1,445,000 persons, 735,000 places (including 478,000 populated places), 411,000 creative works (including 123,000 music albums, 87,000 films and 19,000 video games), 241,000 organizations (including 58,000 companies and 49,000 educational institutions), 251,000 species and 6,000 diseases.

In addition, we provide localized versions of DBpedia in 125 languages. All these versions together describe 38.3 million things, out of which 23.8 million are localized descriptions of things that also exist in the English version of DBpedia. The full DBpedia data set features 38 million labels and abstracts in 125 different languages, 25.2 million links to images and 29.8 million links to external web pages; 80.9 million links to Wikipedia categories, and 41.2 million links to YAGO categories. DBpedia is connected with other Linked Datasets by around 50 million RDF links. Altogether the DBpedia 2014 release consists of 3 billion pieces of information (RDF triples) out of which 580 million were extracted from the English edition of Wikipedia, 2.46 billion were extracted from other language editions. Detailed statistics about the DBpedia datasets in 24 popular languages are provided at Dataset Statistics. ”

2. DBpedia Ontology

(DBpeida Ontology: introduction of ontology, why property should be linked in ontology)

The DBpedia Ontology is a shallow, cross-domain ontology, which has been manually created based on the most commonly used infoboxes within Wikipedia. The ontology currently covers 685 classes which form a subsumption hierarchy and are described by 2,795 different properties.

Since the DBpedia 3.7 release, the ontology is a directed-acyclic graph, not a tree. Classes may have multiple superclasses, which was important for the mappings to schema.org. A taxonomy can still be constructed by ignoring all superclasses except the one that is specified first in the list and is considered the most important.

The DBpedia Ontology currently contains about 4,233,000 instances. The table below lists the number of instances for several classes within the ontology:

| **Instances per class** | |
| --- | --- |
| **Class** | **Instances** |
| Resource (overall) | 4,233,000 |
| Place | 735,000 |
| Person | 1,450,000 |
| Work | 411,000 |
| Species | 251,000 |
| Organisation | 241,000 |

3. Property Mapping

DBpedia properties also have properties of their own and while those subproperties are already mapped in the English DBpedia ontology. However, in other languages they are poorly mapped or sometimes not mapped at all. The problem with this is that the property itself exists in multiple languages, but does not ‘know’ that of itself. Hence, changing the value in one language does not change it in the rest. Therefore, we want to extend and map these sub properties in DBpedia for other languages than English.

We firstly find out 2 main problems in a Dutch DBpedia page. The first problem is the properties of a domain are not annotated/linked in an otology, while it is linked in English version. The second problem is that different spelling of a same property exist in Dutch version, somehow making the data inconsistent. In such situation, it is not applicable to use a certain language to search using a SPARQL query.

Research questions:

1. How can ontologies and their methods in different languages be tied together (mapped)?
2. How is DBpedia being mapped, and how can we use these methods to contribute?
3. How to conduct and analyze the evaluation for mapping the multilingual properties for a DBpedia ontology?

4.Method

5.Experiment

5.1 Data Set

5.2 Results

5.3 Evaluation

6. Related Work

7. Conclusion

Reference: