DOI: 10.20396/revpibic2720192119

Consistency Handling in Dbpedia Evolution.

Túlio B. S. Martins*, Julio Cesar dos Reis.

Abstract
DBpedia is a huge resource available in the Web of Data. It is relevant to update this dataset based on new information appearing in the Wikipedia. However, this operation can provide inconsistencies in the dataset. Although existing literature has defined tools to update DBpedia, there is a lack of studies related to understand and detect inconsistencies in different languages of the database in its evolution. In our research we define and detect specific types of inconsistencies and offer a solution that can be applied to the updates of DBpedia Live, a tool that mantains DBpedia updated constantly. Our research shows that the proposed solution can turn the DBpedia more reliable over time.

Key words:
DBpedia, Inconsistency, Evolution.

Introduction

With the continuous growth of Web of data there is an increase in demand on how to handle big chunks of relational data. With that purpose, Linked Data was born and one of it's biggest examples is DBpedia: a database of interlinked resources based on Wikipedia. For the DBpedia to update itself, it relies on the Wikipedia's continuous updates as well, and to remain up-to-date with it's information constantly the DBpedia Live tool was created. However, many inconsistencies arise when extracting the information from the Wikipedia and converting it to Linked Data and as a result the DBpedia database becomes inconsistent and unreliable for use. That inconsistency is amplified in language chapters of the database other than English. The work presented here shows a method of classification of the inconsistencies, their identification and then a proposed solution for correction of multiple inconsistencies that affect DBpedia in non-english languages.

Results and Discussion

DBpedia information is stored in RDF (Resource Description Framework) data in the formats of triples. Each triple *(s, p, o)* contains a subject *s* a predicate *p* (a property) and an object *o.* In our data analyzed, the triples each refer to an object that is also a resource. We classified our inconsistencies in two types: range inconsistency and domain inconsistency. Each inconsistency refers to the resource that referenced in the triple that is not of the correct type of the property utilized.

For the detection of these inconsistencies we analyzed the extraction method through the DBpedia Live Extraction Manager, the tool inside DBpedia Live that handles extracting information from Wikipedia and converting it into triples. Using our detection algorithm on 40000 triples in the Spanish DBpedi we found the source of most of the inconsistencies in non-english databases: 1) too many resources were not properly defined and 2) some properties were not well translated to other languages.

Based on these problems we developed a solution algorithm: utilizing queries and similar triples present in the English DBpedia we looked to confirm a resource's type and alter inserted the correct definition. Then, we inserted the solution algorithm into the extraction workflow generating a different dataset. We then ran our inconsistency detection algorithm on the new dataset and compared the results.

Chart 1. Inconsistency Detection on 40000 triples on the Spanish DBpedia.

Original Extraction		
Type of Inconsistency	Quantity of triples	Percentage
Range Inconsistency	13434	33,58%
Domain Inconsistency	5355	13,38%
Extraction with Algorithm		
Type of Inconsistency	Quantity of triples	Percentage
Range Inconsistency	6941	17,35%
Domain Inconsistency	3046	7,62%

As we can see from the results the proposed solution algorithm could reduce the total of inconsistencies in over 40%. Considering the Wikipedia is altered by users and it is natural that some consistencies will remain it still

Conclusions

With the data obtained we concluded that the detection of inconsistencies in non-english DBpedia chapters can be done to better filter the triples that are inserted constantly. We also offered a solution on DBpedia Live that helps turn such database far more consistent than it previously was. Such consistency is important to make the database reliable for general use and this work shows progress to making it reliable and aids the correction of future inconsistencies

Acknowledgement

This work is financially supported by the São Paulo Research Foundation (FAPESP).