

Datasets for resolving range violations in DBpedia

Introduction

A range violation error is a problem when an object of a triple does not have a type required by the range of the triple's predicate. Based on our observation, 18.7% of DBpedia triples whose predicate is a mapping-based object property with a defined range are suffering from this kind of error. These errors could be probably due to

- The predicate is incorrect and should be replaced
- The range of the predicate is incorrect and should be refined
- The object is incorrect and should be replaced
- The object misses the correct (required) type

In this work, we focus only on erroneous triples that are due to incorrect objects. For example,

Subject	Predicate	Object	Required Type
dbr:Sedo	dbo:locationCountry	dbr:Cologne	dbo:Country
dbr:Safi_Airways	dbo:targetAirport	dbr:New_Delhi	dbo:Airport
dbr:Stewart_Steven	dbo:employer	dbr:Daily_Mail	dbo:Organisation

One way to fix these errors is replacing the incorrect objects with the correct objects which have the required types. For instance,

- dbr:Cologne → dbr:Germany
- dbr:New_Delhi → dbr:Indira_Gandhi_International_Airport
- dbr:Daily_Mail → dbr:Daily_Mail_and_General_Trust

Datasets

The eight datasets we publish on this page were used to evaluate our work [1]. Each of them is a csv file containing one hundred erroneous triples $\langle s, p, o \rangle$ of a particular object property p (in DBpedia) together with their correct objects ans which we manually find based on the information from corresponding Wikipedia pages. The eight properties and their ranges are

Property p	Range (r_p)
dbo:locationCountry	dbo:Country
dbo:language	dbo:Language
dbo:targetAirport	dbo:Airport
dbo:routeEnd	dbo:RouteOfTransportation
dbo:formerTeam	dbo:SportsTeam
dbo:college	dbo:EducationalInstitution
dbo:employer	dbo:Organisation
dbo:birthPlace	dbo:Place

File format:

s,p,o,ans
http://dbpedia.org/resource/Terrytoons,http://dbpedia.org/ontology/locationCountry,http://dbpedia.org/resource/271_North_Avenue,http://dbpedia.org/resource/United_States
http://dbpedia.org/resource/Novotel,http://dbpedia.org/ontology/locationCountry,"http://dbpedia.org/resource/Évry,_Essonne",http://dbpedia.org/resource/France
http://dbpedia.org/resource/Quest_Arabiya,http://dbpedia.org/ontology/locationCountry,
http://dbpedia.org/resource/Abu_Dhabi,http://dbpedia.org/resource/United_Arab_Emirates
...

Download datasets from <http://ri-www.nii.ac.jp/FixRVE/>

Contact

Piyawat Lertvittayakumjorn pl1515@imperial.ac.uk
Ryutaro Ichise ichise@nii.ac.jp

Reference

1. Piyawat Lertvittayakumjorn, Natthawut Kertkeidkachorn, Ryutaro Ichise. Resolving Range Violations in DBpedia. In Proceedings of 7th Joint International Semantic Technology Conference (JIST). 2017