esDBpedia++

"An enhanced and extended esDBpedia"

Ontology Engineering Group

DBpedia español

#esDBVienna2018 @marianorico

Thanks to TIN2016-78011-C4-4-R (AEI/FEDER, UE) and RTC-2016-4952-7

Detailed info

https://goo.gl/vbS8GB


esDBpedia
In a big, big galaxy…

esDBpedia has *been* without *modifications* *for the last years, but...*

esDBpedia has *been* without *modifications* *for the last years, but...*

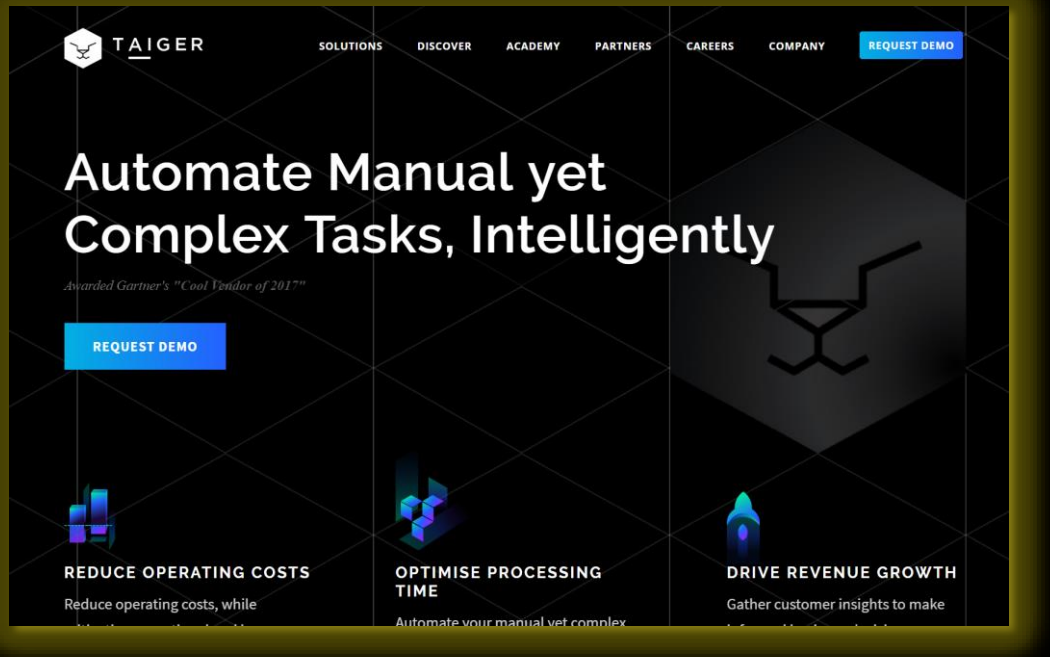
esDBpedia has *been* without *modifications* *for the last years, but...*

esDBpedia has been without modifications for the last years, but...

We've got reinforcements!

esDBpedia++

Company funded 2y project + Public Money



GOBIERNO DE ESPANA
MINISTERIO DE ECONOMIA Y COMPETITIVIDAD

Ontology Engineering Group

DBpedia español

TAIGER's objective

Enhance its Text Analytics in-the-cloud service

- Semantic annotation
- Semantic disambiguation

Our research

Enhance & extend esDBpedia

- Enhance the quality
- Extend to proprietary datasets

Enhancing quality

New techniques

ML + linked data

NLP + linked data

Extending the dataset

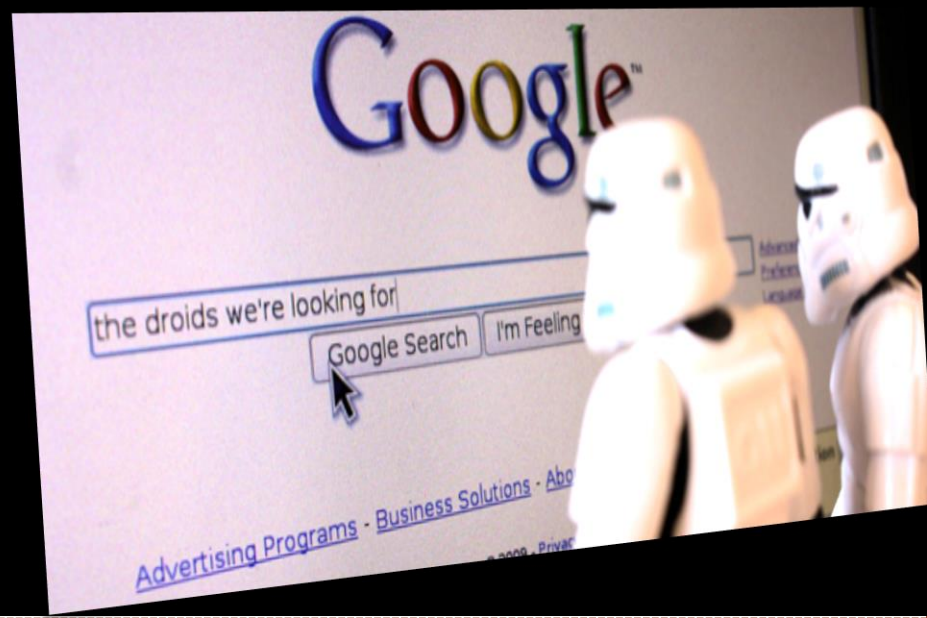
Alignment (ontologies & data)

Dataset 1

Dataset 2

ML + linked data

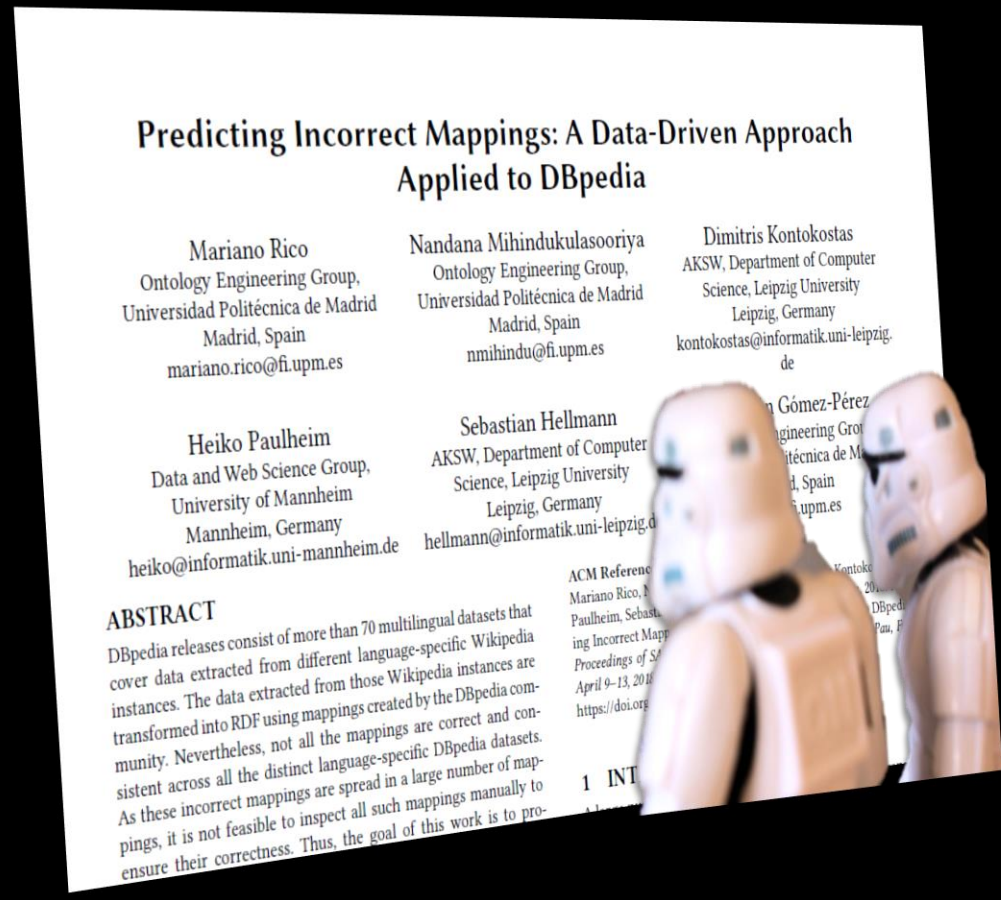
Supervised learning



ML + linked data

Supervised learning

Wrong mapping detection



Predicting Incorrect Mappings: A Data-Driven Approach Applied to DBpedia

Mariano Rico
Ontology Engineering Group, Universidad Politécnica de Madrid
Madrid, Spain
mariano.rico@fi.upm.es

Nandana Mihindukulasooriya
Ontology Engineering Group, Universidad Politécnica de Madrid
Madrid, Spain
nmihindu@fi.upm.es

Dimitris Kontokostas
AKSW, Department of Computer Science, Leipzig University
Leipzig, Germany
kontokostas@informatik.uni-leipzig.de

Heiko Paulheim
Data and Web Science Group, University of Mannheim
Mannheim, Germany
heiko@informatik.uni-mannheim.de

Sebastian Hellmann
AKSW, Department of Computer Science, Leipzig University
Leipzig, Germany
hellmann@informatik.uni-leipzig.d

ABSTRACT

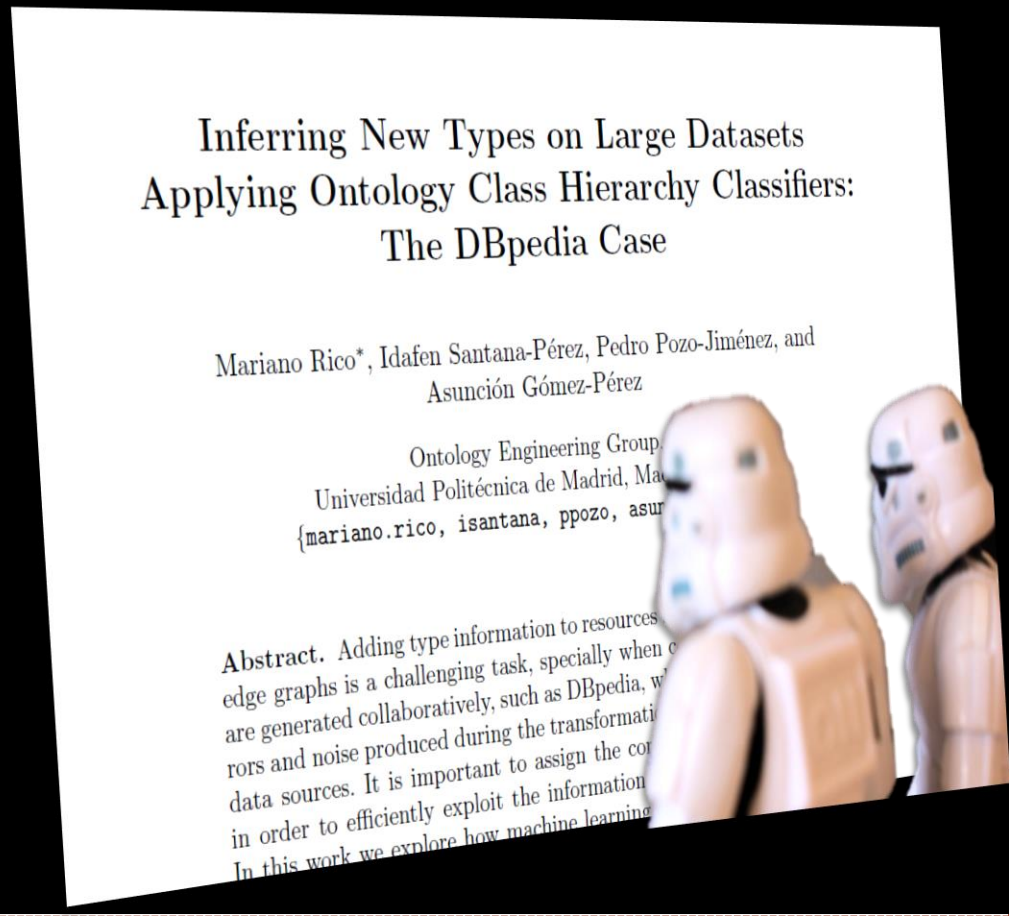
DBpedia releases consist of more than 70 multilingual datasets that cover data extracted from different language-specific Wikipedia instances. The data extracted from those Wikipedia instances are transformed into RDF using mappings created by the DBpedia community. Nevertheless, not all the mappings are correct and consistent across all the distinct language-specific DBpedia datasets. As these incorrect mappings are spread in a large number of mappings, it is not feasible to inspect all such mappings manually to ensure their correctness. Thus, the goal of this work is to pro-

ML + linked data

Supervised learning

Wrong mapping detection

Inferring resource types


Inferring New Types on Large Datasets
Applying Ontology Class Hierarchy Classifiers:
The DBpedia Case

Mariano Rico*, Idafen Santana-Pérez, Pedro Pozo-Jiménez, and
Asunción Gómez-Pérez

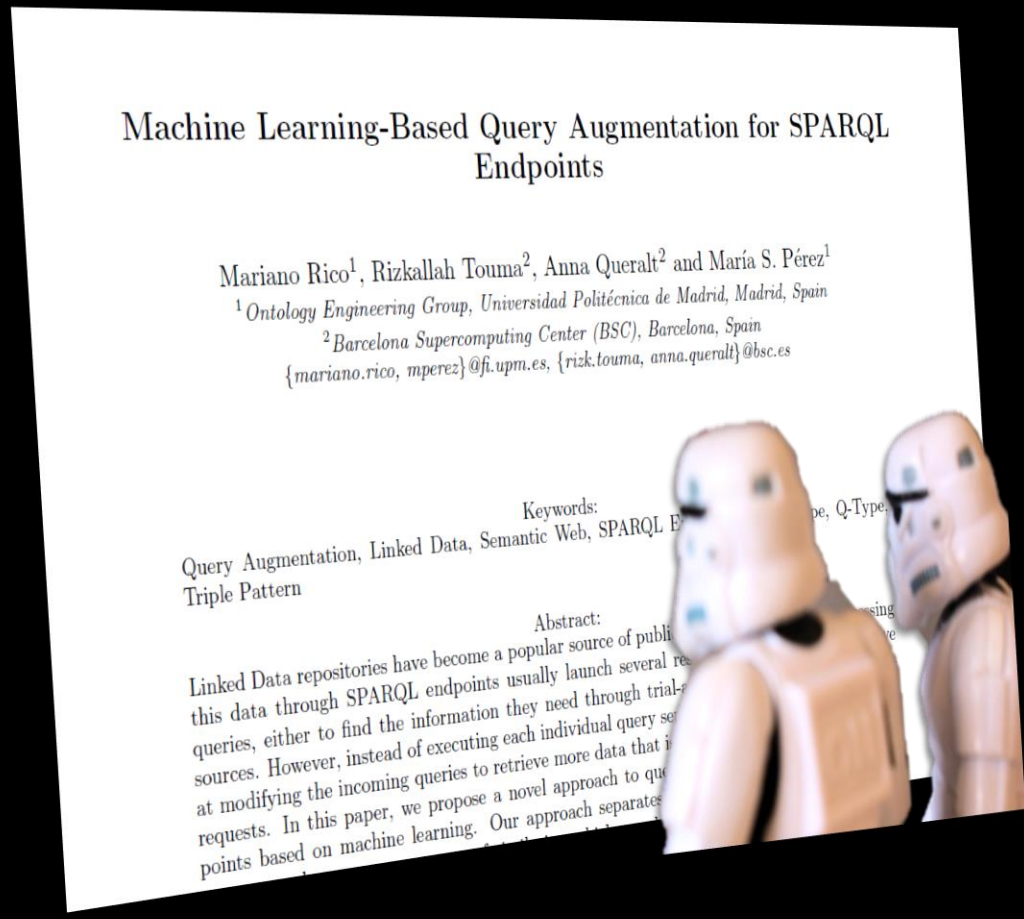
Ontology Engineering Group
Universidad Politécnica de Madrid, Ma
{mariano.rico, isantana, ppozo, asu

Abstract. Adding type information to resources
edge graphs is a challenging task, specially when
are generated collaboratively, such as DBpedia,
rors and noise produced during the transformati
data sources. It is important to assign the co
in order to efficiently exploit the information
In this work we explore how machine learning

ML + linked data

Supervised learning

- Wrong mapping detection
- Inferring resource types
- Predicting next query



Machine Learning-Based Query Augmentation for SPARQL Endpoints

Mariano Rico[1], Rizkallah Touma[2], Anna Queralt[2] and María S. Pérez[1]
[1] Ontology Engineering Group, Universidad Politécnica de Madrid, Madrid, Spain
[2] Barcelona Supercomputing Center (BSC), Barcelona, Spain
{mariano.rico, mperez}@fi.upm.es, {rizk.touma, anna.queralt}@bsc.es

Keywords:
Query Augmentation, Linked Data, Semantic Web, SPARQL E
Triple Pattern

Abstract:
Linked Data repositories have become a popular source of publi
this data through SPARQL endpoints usually launch several re
queries, either to find the information they need through trial-
sources. However, instead of executing each individual query se
at modifying the incoming queries to retrieve more data that i
requests. In this paper, we propose a novel approach to qu
points based on machine learning. Our approach separates

ML + linked data

Wrong mapping detection

Google Summer of Code 2018 Projects

Search for a project, organization, or student

VIEW BY ORGANIZATION

STUDENT

OpenML

Search

HELP

SIGN IN

Explore

- Data
- Task
- Flow
- Run
- Study
- Task type
- Measure
- People

Predicting wrong DBpedia mappings

Created 05-05-2017 by Mariano Rico Visibility: public

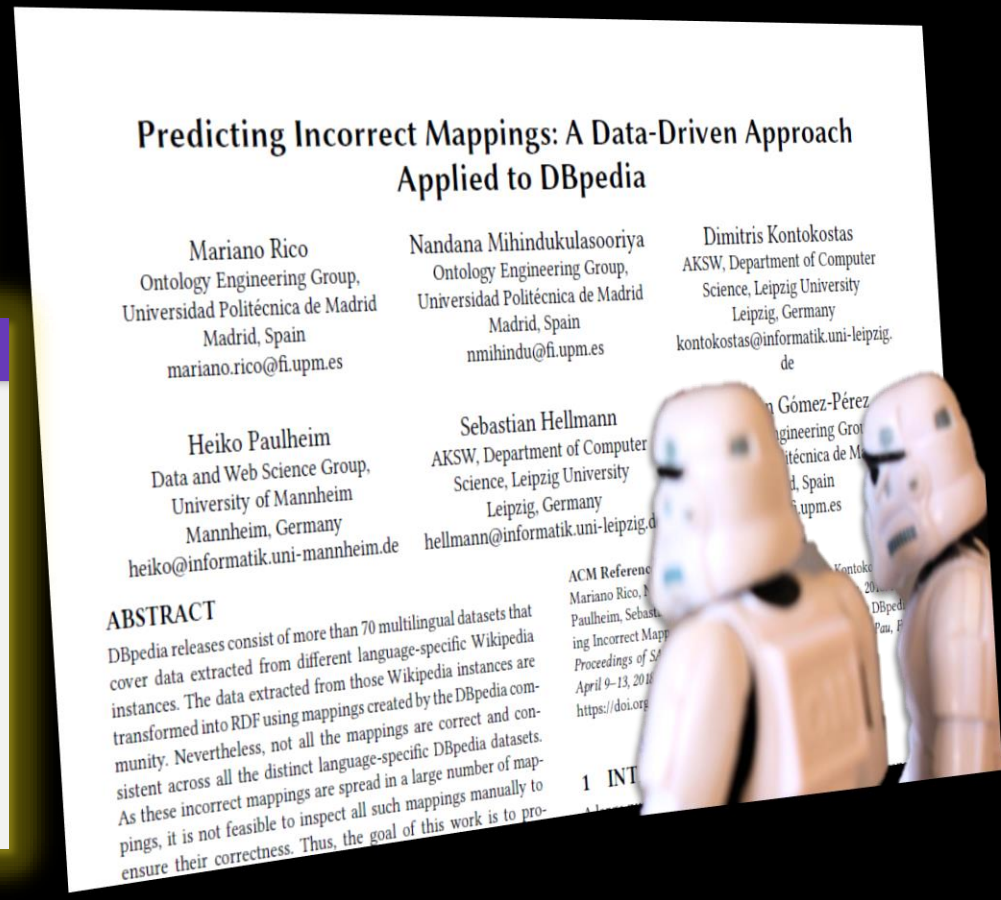
DESCRIPTION | 10 DATA SETS | 4 TASKS | 0 FLOWS | 3 RUNS

Edit

This was an study started by Nandana and Mariano in 2016. We started with unsupervised methods, but we could not find good clusters. En 2017 we started with annotated data and here we are.

Summary

> DBpedia releases consist of more than 70 multilingual datasets that cover data extracted from different language-specific Wikipedia instances. The data extracted from those Wikipedia instances are transformed into RDF using mappings created by the DBpedia community. Nevertheless, not all the mappings are correct and consistent across all the distinct language-specific DBpedia datasets. As these incorrect mappings are spread in a large number of mappings, it is not feasible to inspect all such mappings manually to



Predicting Incorrect Mappings: A Data-Driven Approach Applied to DBpedia

Mariano Rico
Ontology Engineering Group, Universidad Politécnica de Madrid
Madrid, Spain
mariano.rico@fi.upm.es

Nandana Mihindukulasooriya
Ontology Engineering Group, Universidad Politécnica de Madrid
Madrid, Spain
nmihindu@fi.upm.es

Dimitris Kontokostas
AKSW, Department of Computer Science, Leipzig University
Leipzig, Germany
kontokostas@informatik.uni-leipzig.de

Heiko Paulheim
Data and Web Science Group, University of Mannheim
Mannheim, Germany
heiko@informatik.uni-mannheim.de

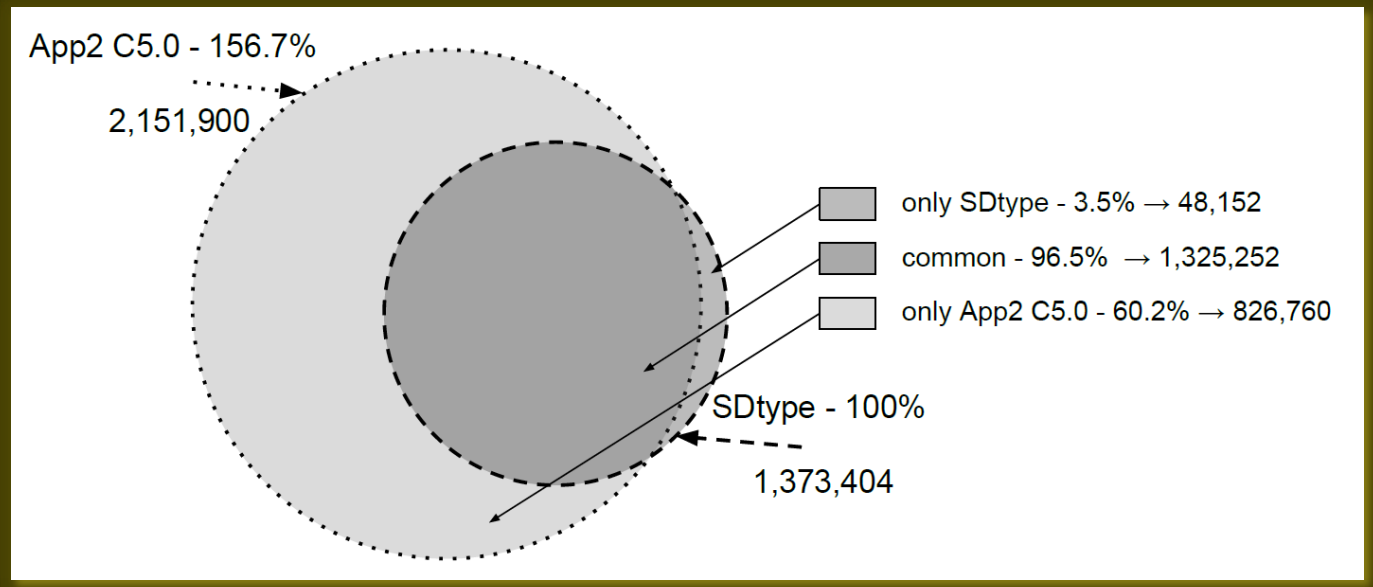
Sebastian Hellmann
AKSW, Department of Computer Science, Leipzig University
Leipzig, Germany
hellmann@informatik.uni-leipzig.d

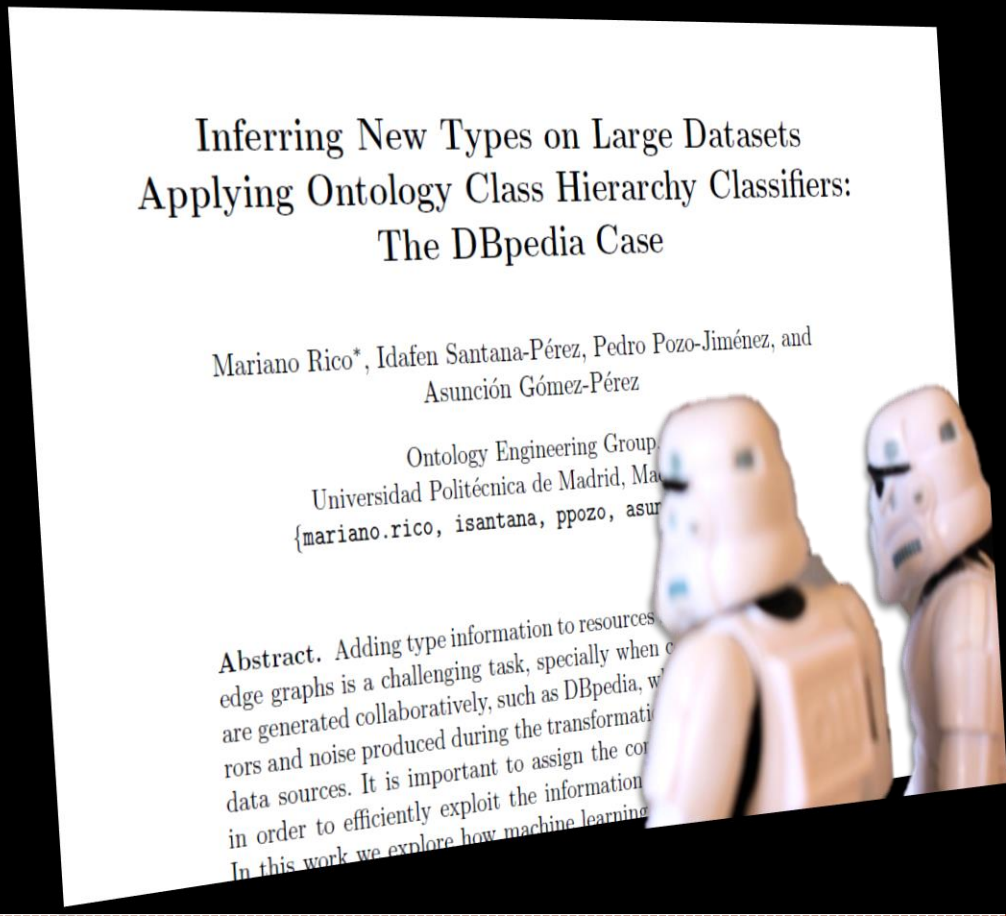
ABSTRACT

DBpedia releases consist of more than 70 multilingual datasets that cover data extracted from different language-specific Wikipedia instances. The data extracted from those Wikipedia instances are transformed into RDF using mappings created by the DBpedia community. Nevertheless, not all the mappings are correct and consistent across all the distinct language-specific DBpedia datasets. As these incorrect mappings are spread in a large number of mappings, it is not feasible to inspect all such mappings manually to ensure their correctness. Thus, the goal of this work is to pro-

ML + linked data

Inferring resource types





Inferring New Types on Large Datasets Applying Ontology Class Hierarchy Classifiers: The DBpedia Case

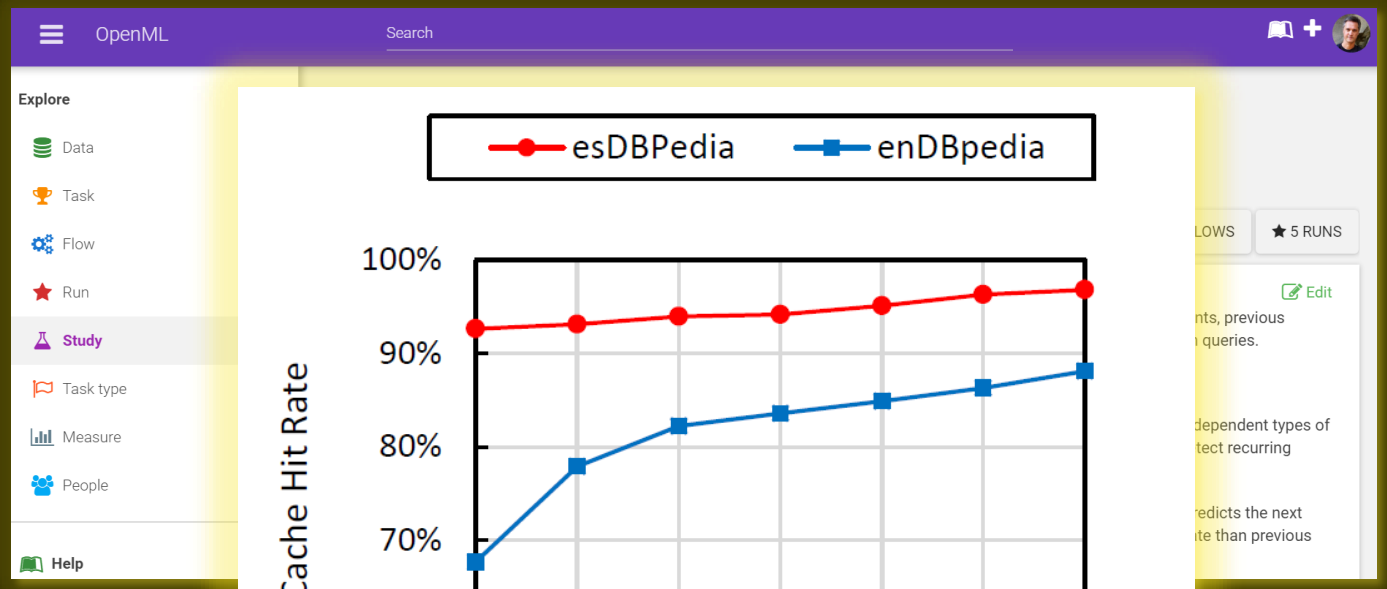
Mariano Rico*, Idafen Santana-Pérez, Pedro Pozo-Jiménez, and Asunción Gómez-Pérez

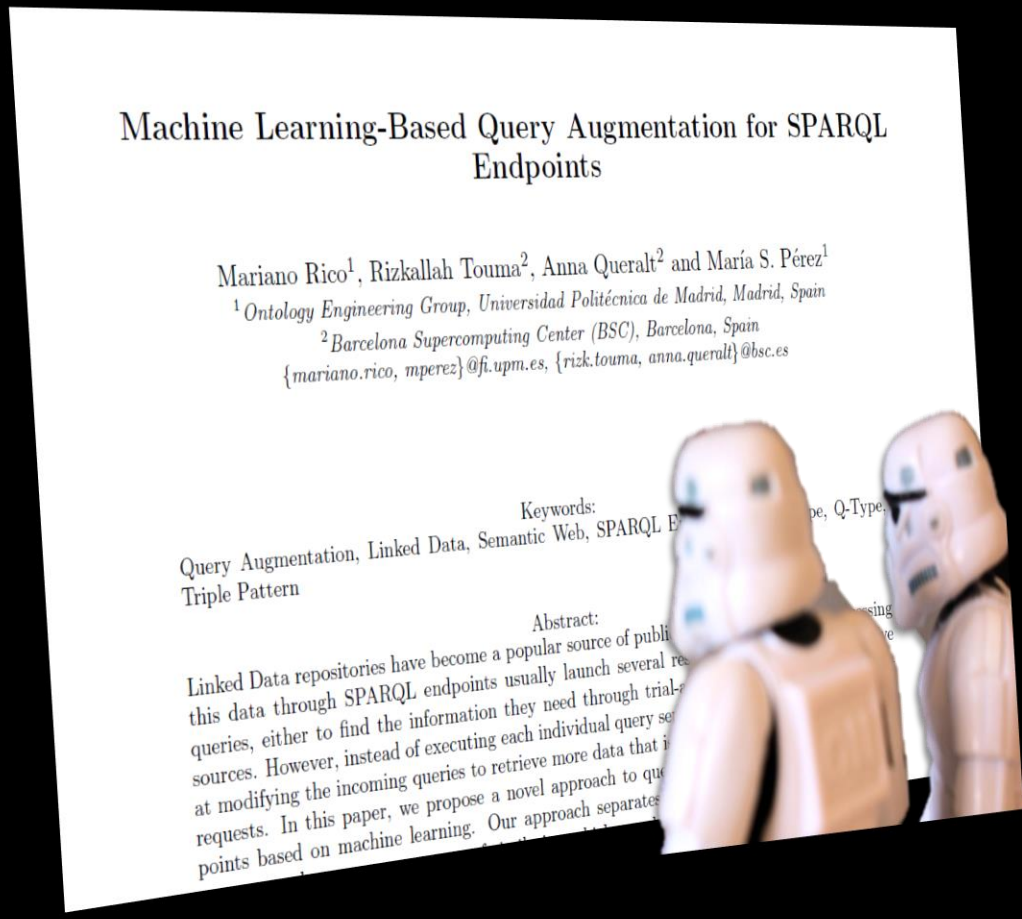
Ontology Engineering Group
Universidad Politécnica de Madrid, Ma
{mariano.rico, isantana, ppozo, asur

Abstract. Adding type information to resources
edge graphs is a challenging task, specially when
are generated collaboratively, such as DBpedia,
rors and noise produced during the transformati
data sources. It is important to assign the co
in order to efficiently exploit the information
In this work we explore how machine learning

ML + linked data

Predicting next query





Machine Learning-Based Query Augmentation for SPARQL Endpoints

Mariano Rico[1], Rizkallah Touma[2], Anna Queralt[2] and María S. Pérez[1]

[1] Ontology Engineering Group, Universidad Politécnica de Madrid, Madrid, Spain

[2] Barcelona Supercomputing Center (BSC), Barcelona, Spain

{mariano.rico, mperez}@fi.upm.es, {rizk.touma, anna.queralt}@bsc.es

Keywords:

Query Augmentation, Linked Data, Semantic Web, SPARQL E

Triple Pattern

Abstract:

Linked Data repositories have become a popular source of publi

this data through SPARQL endpoints usually launch several re

queries, either to find the information they need through trial-

sources. However, instead of executing each individual query se

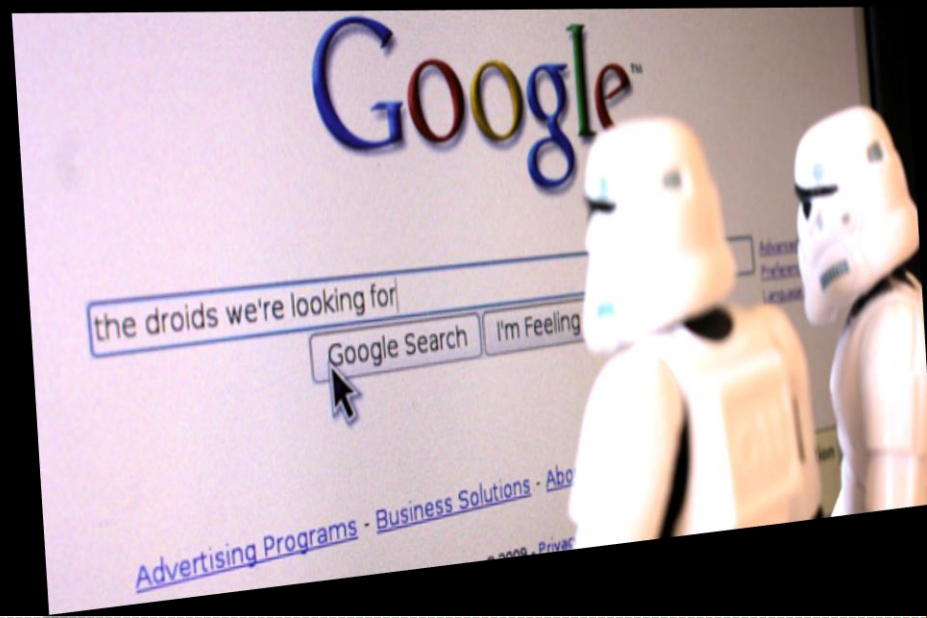
at modifying the incoming queries to retrieve more data that i

requests. In this paper, we propose a novel approach to qu

points based on machine learning. Our approach separates

NLP + linked data

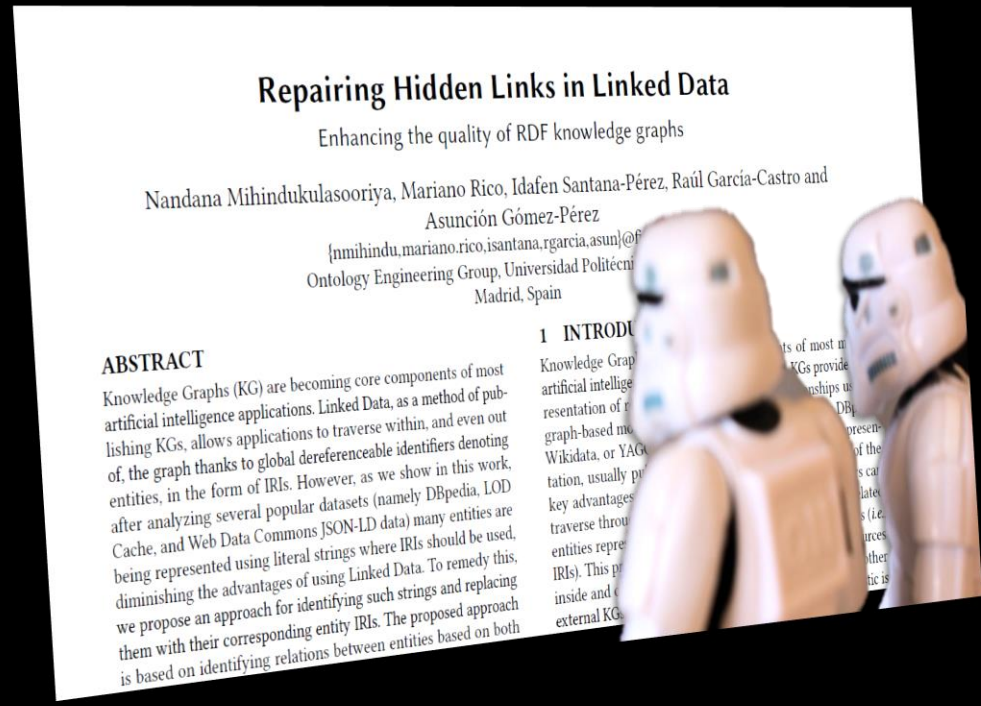
NER & similarity



NLP + linked data

NER & similarity

“Resurrection” of URIs



Repairing Hidden Links in Linked Data

Enhancing the quality of RDF knowledge graphs

Nandana Mihindukulasooriya, Mariano Rico, Idafen Santana-Pérez, Raúl García-Castro and Asunción Gómez-Pérez

{nmihindu,mariano.rico,isantana,rgarcia,asun}@f

Ontology Engineering Group, Universidad Politécn

Madrid, Spain

ABSTRACT

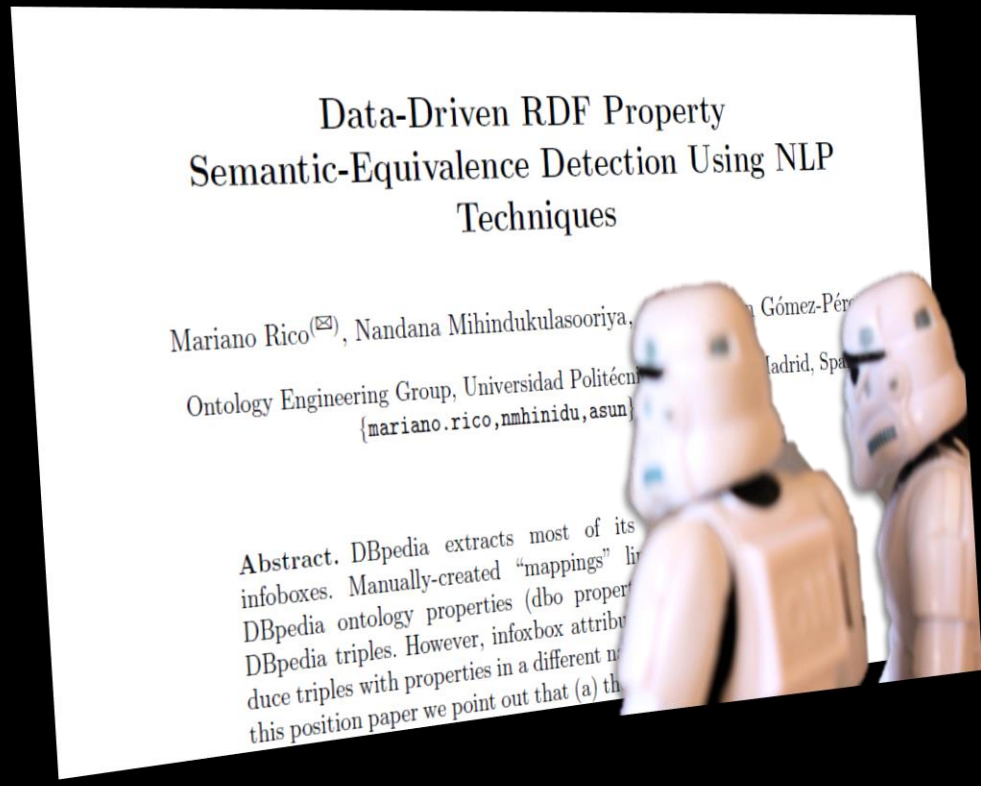
Knowledge Graphs (KG) are becoming core components of most artificial intelligence applications. Linked Data, as a method of publishing KGs, allows applications to traverse within, and even out of, the graph thanks to global dereferenceable identifiers denoting entities, in the form of IRIs. However, as we show in this work, after analyzing several popular datasets (namely DBpedia, LOD Cache, and Web Data Commons JSON-LD data) many entities are being represented using literal strings where IRIs should be used, diminishing the advantages of using Linked Data. To remedy this, we propose an approach for identifying such strings and replacing them with their corresponding entity IRIs. The proposed approach is based on identifying relations between entities based on both

NLP + linked data

NER & similarity

"Resurrection" of URIs

dbp properties to enhance SPARQL queries



NLP + ML + linked data

Not published yet

DBpedia resource type from a textual description

NLP + linked data

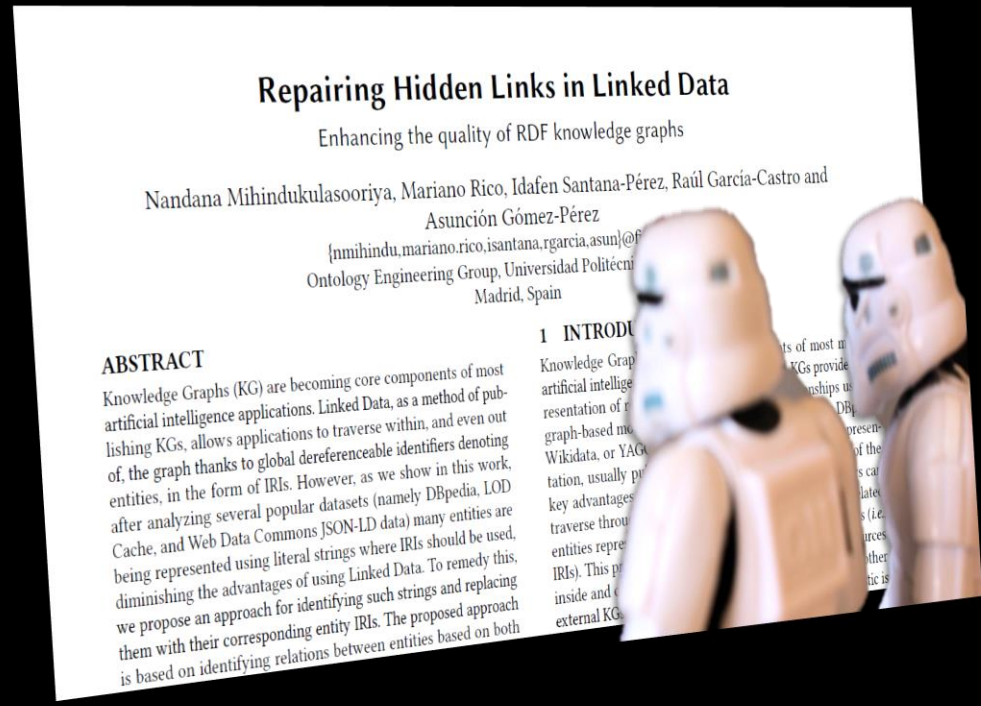
“Resurrection” of URIs

Table 7: String disambiguation performance

Class	# of strings	Disambiguated		Prec.	Recall
		#	Correct		
dbo:Athlete	100	51	50	98.04%	50%
dbo:SportsTeam	100	44	44	100%	44%
dbo:SportsEvent	100	58	55	94.83%	56%
Total	300	153	149	97.38%	49.67%

Table 8: Improvements in connectivity

Graph	edges	com-ponents	isolated	largest component	
				size	% total
Original	828,310	119,623	112,331	168,128	54.28
Repaired	1,035,912	99,137	93,507	192,805	64.16
Δ	+207,602	-20,486	-18,824	+24,677	+9.88
	+25.06%	-17.12%	-16.76%	+14.68%	



Repairing Hidden Links in Linked Data

Enhancing the quality of RDF knowledge graphs

Nandana Mihindukulasooriya, Mariano Rico, Idafen Santana-Pérez, Raúl García-Castro and Asunción Gómez-Pérez

{nmihindu,mariano.rico,isantana,rgarcia,asun}@f

Ontology Engineering Group, Universidad Politécn

Madrid, Spain

ABSTRACT

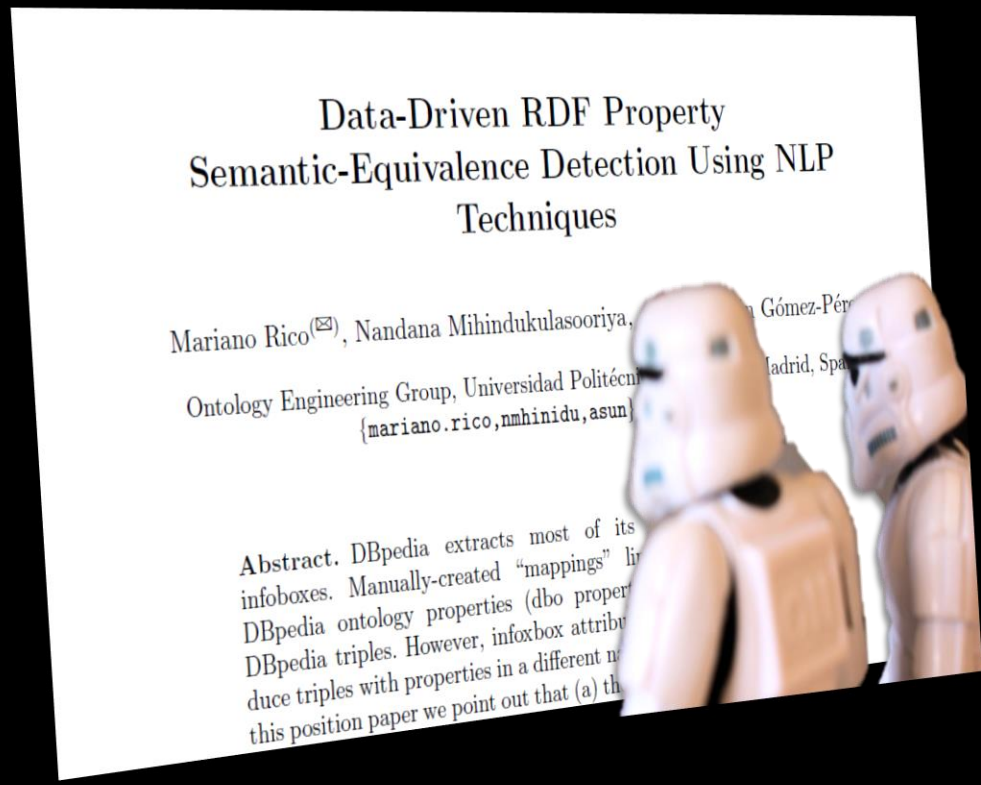
Knowledge Graphs (KG) are becoming core components of most artificial intelligence applications. Linked Data, as a method of publishing KGs, allows applications to traverse within, and even out of, the graph thanks to global dereferenceable identifiers denoting entities, in the form of IRIs. However, as we show in this work, after analyzing several popular datasets (namely DBpedia, LOD Cache, and Web Data Commons JSON-LD data) many entities are being represented using literal strings where IRIs should be used, diminishing the advantages of using Linked Data. To remedy this, we propose an approach for identifying such strings and replacing them with their corresponding entity IRIs. The proposed approach is based on identifying relations between entities based on both

NLP + linked data

dbp properties to enhance SPARQL queries

Table 2. Example of dbp → dbo property mappings. Δ_1 is the enhancement for the example query in Listing 1.1.

DBpedia dbo prop		dbp prop				Δ_1
		Syntactic		Semantic		
English	birthPlace	birthPlace birthPlac birthPlace birthPlace oplaceOfBirth birthPalce birthPlae birthPlcace birthPlace nbirthPlace birthdplace birthPalce birthPace birthPlaxe birthPlace brithPlace birghPlace biRthPlace birth placebirth placeOfBirth placOfBirth birthPlaceOf	birthplace birthdplace PlaceOfBirth birthPlace. birthPlacE birthplace	placeofbirth cityofbirth cityOfBirth birthLocation laceOfBirth		cityofbirthPlace 350%
Spanish	birthPlace	lugarDeNacimiento lugarNacimiento lugardenacimiento lugarNaciento	lugarNacimiento lugarnacimiento lugarNacimento	ciudaddenacimiento ciudadDenacimiento paisdenacimiento birthPlace	paisNacimiento birthplace placeOfBirth	221%
German	birthPlace	geburtsort placeOfBirth placeofbirth	birthplace	birthPlace geburtsland	countryofbirth	134%



Thanks a LOD!

esDBpedia++

"An enhanced and extended esDBpedia"

Ontology Engineering Group

DBpedia español

#esDBVienna2018 @marianorico

Thanks to TIN2016-78011-C4-4-R (AEI/FEDER, UE) and RTC-2016-4952-7

NLP + ML + linked data

Not published yet

Resource type from textual description

NLP + ML + linked data

Not published yet

Resource type from textual description