LODsyndesisIE: Entity Extraction from Text and Enrichment using Hundreds of Linked Datasets

Michalis Mountantonakis and Yannis Tzitzikas

{mountant|tzitzik}@ics.forth.gr

Institute of Computer Science, Foundation for Research and Technology (FORTH-ICS), GREECE, and
Computer Science Department, University of Crete, GREECE

1. The problem

It is not trivial to **extract information** for the **entities** of a text from **multiple RDF datasets simultaneously.**

Existing Entity Extraction tools **associate** each **recognized entity** to a **single dataset**.

2. The proposed research prototype LODsyndesisIE

LODsyndesisIE is a research prototype that offers **Entity Extraction from text** and **Entity Enrichment** for the **extracted entities,** by using **hundreds of RDF datasets simultaneously**.

✓It uses any **combination** of three popular **entity recognition tools** for **identifying** the **entities of a text**.

✓It exploits **LODsyndesis,** which supports **cross-dataset identity reasoning,** for **extracting more information** for each identified entity from **400 RDF datasets**.

DBpedia Spotlight, CoreNLP, WAT, LODsyndesis

3.The process of LODsyndesisIE

1. Input Text from user

> The Lord of the Rings Film Series is a series of three epic fantasy adventure films and feature an ensemble cast including Elijah Wood, Ian McKellen, Liv Tyler, Viggo Mortensen, Sean Astin, Cate Blanchett, John Rhys-Davies, Christopher Lee, Billy Boyd, and Orlando Bloom.

2. Entity Recognition and Text Annotation ↓ (DBpedia Spotlight, CoreNLP, WAT)

> **The Lord Of The Rings Film Series** is a series of three epic **Fantasy** adventure films and feature an **Ensemble Cast** including **Elijah Wood**, **Ian Mckellen**, **Liv Tyler**, **Viggo Mortensen**, **Sean Astin**, **Cate Blanchett**, **John Rhys-davies**, **Christopher Lee**, **Billy Boyd**, and **Orlando Bloom**.

3. Information Extraction by using LODsyndesis

3a. Find Related Facts

Key Entity	Relationship	Entity 2	Provenance
The Lord Of The Rings Film Series	http://schema.org/actor	Elijah Wood	http://www.wikidata.org/ http://yago-knowledge.org

3b. Export annotated text in RDFa format

```
<span vocab="http://schema.org/" typeof="Person" about="http://dbpedia.org/resource/Peter_Jackson">
<a property="url" href="https://demos.isl.ics.forth.gr/lodsyndesis/RunQuery?URI=dbpedia.org$resource
<span property="name">Peter Jackson</span></a>
<a property="sameAs" href="http://sw.opencyc.org/concept/Mx4rvsvTu5wpEbGdrcN5Y29ycA"></a>
<a property="sameAs" href="http://id.loc.gov/authorities/names/no96017219"></a>
<a property="sameAs" href="http://www.idref.fr/035340061/id"></a>
<a property="sameAs" href="http://data.linkedmdb.org/resource/actor/25090"></a>
<a property="sameAs" href="http://yago-knowledge.org/resource/Peter_Jackson"></a>
<a property="sameAs" href="http://data.linkedmdb.org/resource/producer/9893"></a>
<a property="sameAs" href="http://data.bnf.fr/ark:/12148/cb13182829t#foaf:Person"></a>
```

3c. For each recognized entity, browse or export in several formats … (CSV, RDF, HTML, {REST API})

Datasets
DBpedia
Freebase
YAGO
BBC
VIAF
IDREF

14 Datasets Found

…Its provenance

…All its triples

Elijah Wood

Predicate	Value	Provenance
dbp:occupation	Actor	DBpedia
foaf:familyName	Wood	VIAF
frb:birthDate	1981-01-28	Freebase

837 Triples Found

…All its URIs

Equivalent URIs
http://dbpedia.org/resource/Elijah_Wood
http://viaf.org/viaf/34660423
http://www.wikidata.org/entity/Q483771

16 URIs Found

…Its K most related datasets

Dataset1	Dataset 2	Triples
Freebase	DBpedia	571
Freebase	YAGO	557
Freebase	Wikidata	553

Top Pairs of Datasets with the most Triples

4. Applications of LODsyndesisIE

A. Comparison of Entity Recognition Tools

"Which entity extraction tool is more effective for my text?"

DBpedia Spotlight, WAT, CoreNLP

B. Data Integration and Enrichment

"I want to enrich the data for each entity of my text from many datasets."

C. Dataset Discovery and Selection

"Give me the 4 most relevant datasets for the entities of my text."

WIKIDATA, DBpedia, Freebase, yago select knowledge

D. Data Veracity

"Is this true?"

Elijah Wood was an actor in **The Lord of the Rings**.

✓ Confirmed by WIKIDATA, DBpedia

5. TRY LODsyndesisIE

TRY *LODsyndesisIE* and its REST API

https://demos.isl.ics.forth.gr/LODsyndesisIE

Follow us on twitter*: https://twitter.com/lodsyndesis*

FORTH INSTITUTE OF COMPUTER SCIENCE — **FORTH-ICS-ISL**

UOC-CSD

PRESENTED BY

eswc20