Benchmarking Entity Linking for Question Answering over Knowledge Graphs

Guillermo Echegoyen Blanco
Universidad Nacional de Educación a Distancia
gblanco **at** lsi.uned.es

Overview

Introduction

Entity Linking

Def: Link parts of a Natural Language passage to their corresponding node in a Knowledge Graph. Usually comprises:

- Recognize the entity mention in the text.
- Disambiguate the mention.

Motivation

- Asses the impact of Entity Linking on a Question Answering task over a KG.
- Actual collections for QA are easy for EL.

Outcomes

Outcomes

Our main contributions are:

- QA Datasets characterization
- Semi-automatic method to generate EL dataset.
- Release large benchmark dataset and baseline for EL in QA.

Experiments

Experiments

Datasets

- QALD {1-4} Unger et al. 2014) $\leq$ 200 QA pairs each
- LC-QuAD (Trivedi et al. 2017) 5K QA pairs

Example

Q: List all episodes of the first season of the HBO television series **The Sopranos**

$$\Downarrow$$

EL: *http ://dbpedia.org/resource/The_Sopranos*

Difficulty

- Given the Question, how easy is the EL?

EL Casuistry Total	**QALD-1** **15**	**QALD-2** **84**	**QALD-3** **89**	**QALD-4** **202**
Identical to DBP uri	73.33	85.71	84.27	79.21
Missing tokens		4.76	5.62	9.9
Additional tokens	20.0	1.19	1.12	0.5
Lexical variation	6.67	5.95	5.62	8.42
Other		2.38	3.37	1.98

Conclusions

Questions?

References

References

Priyansh Trivedi et al. “Lc-quad: A corpus for complex question answering over knowledge graphs”. In: *International Semantic Web Conference*. Springer. 2017, pp. 210–218.

Christina Unger et al. “Question Answering over Linked Data (QALD-4)”. In: (2014). URL: https://hal.inria.fr/hal-01086472/.