Entity Linking in the ParlaMint Corpus

An exploration of linking actors in parliamentary debates

Ruben van Heusden, Jaap Kamps, Maarten Marx

University of Amsterdam, IRLab

Table of contents

1. Introduction

2. Entity Linking

3. Experiments

4. Conclusion & Next Steps

5. Questions

Intro

ParlaMint

- Collection of parliamentary proceedings from 17 European countries
- Unified in the TEI XML format
- Corpora are (automatically) linguistically annotated

Named Entities

Why are we interested in Named Entities?

- Important *anchors* in text
- Interactions between Named Entities (across languages)
- Linking entities across languages

Some ParlaMint Examples of Entities

St George, Mr Speaker, Chester, Tim Draycott, Charlotte Leslie, Jess, Halton

Entity Linking

Entity Linking

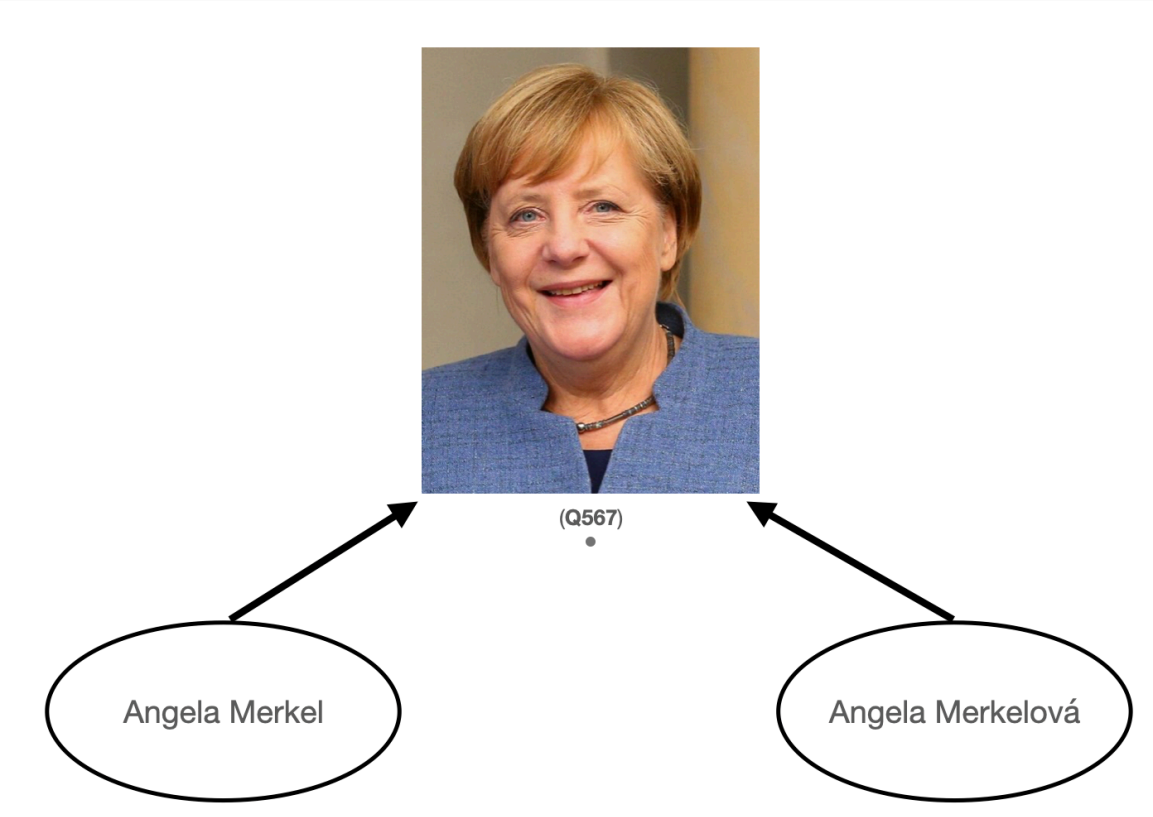
- Linking mentions of entities (persons, organization, locations etc) to a knowledge base

Entity Linking Example

Angela Merkel is the former chancellor of Germany. **Merkel** grew up in East-Germany.

Q-Items

- Cross-lingual identifiers from WikiData



Experiments

Naive Approach

1. Pick a (multilingual) Entity Linker
2. Run this entity linker over all languages
3. Use these results

Naive Experiment

1. Pick 20 ’international’ entities
2. Find mentions that include this entity as a string
3. Run a system on these entities (WikiData)

Naive Experiment

Country	Accuracy Score
PL	0.33
CZ	0.37
HR	0.29
IS	0.67
LV	0.16
BG	0.77
NL	0.91

Table 1: Accuracy of the WikiData system on a set of 20 entities,taken directly from ParlaMint

Why? Inflections, ambiguous entities

Ideal Case

Ideal Case Experiment

- Three systems
  1. YAGO [2]
  2. DPBedia [1]
  3. WikiData
- Selection of 100 local politicians from 10 countries for baseline tests

Usage of *Q-items* for multilingual Entity Linking

Entity Linking Baseline

Country	DBPedia	WikiData	YAGO
NL	0.97	**0.98**	0.56
DE	0.58	**0.94**	0.60
FR	0.95	**0.97**	0.95
CZ	0.31*	**0.95**	0.87
HU	0.75	**0.90**	0.73
EN	0.74	**0.87**	0.78
IT	0.18*	0.95	**0.97**
IS	0.67*	**1.00**	0.85
DK	0.69	**0.96**	0.79
TR	0.52	**0.97**	0.71
Mean	0.74	**0.94**	0.73

Table 2: Accuracy of DBPedia, WikiData and YAGO on 100 local politicians from 10 countries.

Baselines (2)

- Differences between language performance for DBPedia and YAGO
- Best performance for WikiData

Baselines (3)

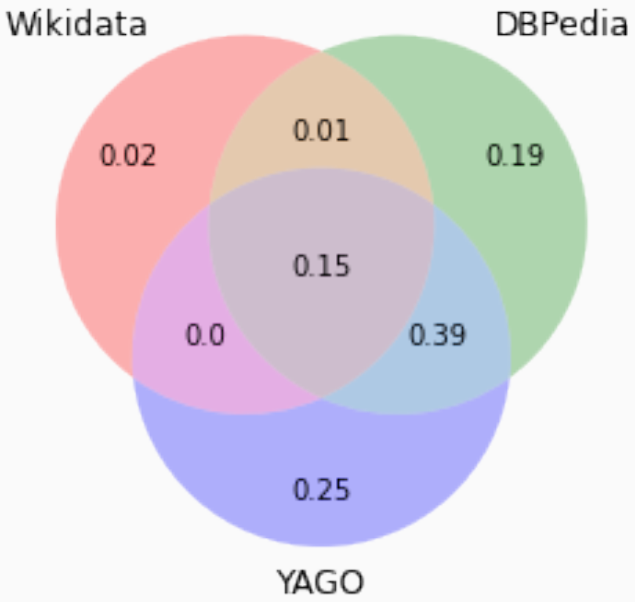


Figure 1: Error Analysis of the three systems used in the baseline experiment

Conclusion & Next Steps

Conclusion & Next Steps

- Differing coverage in systems for different languages
- Naive Approach does not work very well
- Developing / using algorithms to solve problems

Questions

References i

P. N. Mendes, M. Jakob, A. García-Silva, and C. Bizer.
Dbpedia spotlight: shedding light on the web of documents.
In *Proceedings of the 7th international conference on semantic systems*, pages 1–8, 2011.

F. M. Suchanek, G. Kasneci, and G. Weikum.
Yago: a core of semantic knowledge.
In *Proceedings of the 16th international conference on World Wide Web*, pages 697–706, 2007.