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
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
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
1. Pick 20 ‘international’ entities
2. Find mentions that include this entity as a string
3. Run a system on these entities (WikiData)
### Table 1: Accuracy of the WikiData system on a set of 20 entities, taken directly from ParlaMint

<table>
<thead>
<tr>
<th>Country</th>
<th>Accuracy Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>PL</td>
<td>0.33</td>
</tr>
<tr>
<td>CZ</td>
<td>0.37</td>
</tr>
<tr>
<td>HR</td>
<td>0.29</td>
</tr>
<tr>
<td>IS</td>
<td>0.67</td>
</tr>
<tr>
<td>LV</td>
<td>0.16</td>
</tr>
<tr>
<td>BG</td>
<td>0.77</td>
</tr>
<tr>
<td>NL</td>
<td>0.91</td>
</tr>
</tbody>
</table>

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

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

**Table 2:** Accuracy of DBPedia, WikiData and YAGO on 100 local politicians from 10 countries.
• Differences between language performance for DBPedia and YAGO
• Best performance for WikiData
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
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. 