Query Interpretations from Entity-Linked Segmentations

Motivation
- Search queries can be ambiguous
- Identifying entities may resolve ambiguities
- Derive entity-based interpretations
- Interpretations help to diversify search results

Entities in Queries
Explicit entities:
- `nile` → `Nile` (river)
- `source of the nile` → `Source of the Nile`

Implicit entities:
- `source of the nile` → `Blue Nile`
- `source of the nile` → `White Nile`

Related entities:
- Egypt, Sudan, The Settlers of Catan, ...

Query Interpretation Approach

1. Entity recognition
   - Identify explicitly mentioned entities
   - Recall-oriented implementation
     - `nile` → `Nile` (river)
     - `source of the nile` → `Source of the Nile`

2. Query segmentation
   - Split query into semantically coherent segments
     - Approach by Hagen et al. (2011)
     - `source | of the | nile`
     - `source of the nile`

3. Segment-entity linking
   - Link segments to semantically coherent entities
     - `source | of the | Nile (river)`
     - `Source of the Nile`

4. Interpretation scoring
   - Compute weighted sum of
     - Commonness scores
     - Relatedness scores
     - Context scores

Query Interpretation Evaluation

Comparison of our query interpretation method to various entity linking approaches equipped with a greedy interpretation finding step (Hasibi et al., 2014).

<table>
<thead>
<tr>
<th></th>
<th>Complete Matches</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recall</td>
<td>Precision</td>
</tr>
<tr>
<td>Our approach</td>
<td>0.295</td>
<td>0.336</td>
</tr>
<tr>
<td>Dexter</td>
<td>0.230</td>
<td>0.312</td>
</tr>
<tr>
<td>Nordlys EL</td>
<td>0.189</td>
<td>0.278</td>
</tr>
<tr>
<td>Radboud EL</td>
<td>0.144</td>
<td>0.199</td>
</tr>
<tr>
<td>Smaph</td>
<td>0.176</td>
<td>0.243</td>
</tr>
<tr>
<td>Dandelion</td>
<td>0.166</td>
<td>0.226</td>
</tr>
<tr>
<td>TagMe</td>
<td>0.165</td>
<td>0.216</td>
</tr>
<tr>
<td>Babelfy</td>
<td>0.112</td>
<td>0.160</td>
</tr>
<tr>
<td>TextRazor</td>
<td>0.098</td>
<td>0.131</td>
</tr>
<tr>
<td>FEL</td>
<td>0.133</td>
<td>0.173</td>
</tr>
<tr>
<td>Ambiverse</td>
<td>0.007</td>
<td>0.011</td>
</tr>
</tbody>
</table>

Conclusion
Contributions
- Entity-based query interpretation
- New annotated dataset: 2,800 queries
- Recall-oriented explicit entity recognizer
- Segmentation-based query interpretation method
- Comparison to entity linking and interpretation methods

Future Work
- Implicit entity recognition
- Query interpretation for conversational systems

Data: webis.de/data/webis-qinc-22.html
Code: github.com/webis-de/WSDM-22