Integrating Query Interpretation Components into the Information Retrieval Experiment Platform

WOWS 2024

Marcel Gohsen
Benno Stein

Bauhaus-Universität Weimar
Webis
Query Interpretation

source of the nile
Query Interpretation

source of the nile

Nile / Sources

Blue Nile River

Atbarah

White Nile
In particular, the Nile is the primary water source of Egypt, Sudan and South Sudan. Additionally, the Nile is an important economic river, supporting...
Query Interpretation

source of the nile

Nile/Sources

- Blue Nile River
- Atbarah
- White Nile

Nile

In particular, the Nile is the primary water source of Egypt, Sudan and South Sudan. Additionally, the Nile is an important economic river, supporting...

White Nile · Blue Nile · Nile Delta · River source

Nile River | Delta, Map, Basin, Length, Facts, Definition ...

Its reputed source is a spring, considered holy by the Ethiopian Orthodox Church, from which a small stream, the Abay, flows down to Lake Tana (T'ana), a fairly...

Physiography · White Nile River · Nile River basin · Climate and hydrology
source of the nile

OR

source of the nile
Query Interpretation

⟨source | of the | nile⟩

OR

⟨source of the nile⟩
Query Interpretation

(source | of the | kb:Nile) OR (kb:Source_of_the_Nile)
Query Interpretation
Kasturia et al. 2022

Entity Linking

nile → kb:Nile

Query Segmentation
< source | of the | nile >
< source of the nile >

Segment-Entity Linking

< source | of the | kb:Nile >
< kb:Source_of_the_Nile >

Interpretation Scoring
< source | of the | kb:Nile > → 1.0
< kb:Source_of_the_Nile > → 0.8
...
Entity index:

- nile → [kb:Nile,...]
- ...
Entity Linking
Kasturia et al. 2022

Entity index:
- nile → [kb:Nile,...]
- nile river → [kb:Nile,...]
- ...
Entity Linking
Kasturia et al. 2022

Entity index:
- nile → [kb:Nile, kb:Nile_(band),...]
- nile river → [kb:Nile,...]
- ...

Nile

From Wikipedia, the free encyclopedia

For other uses, see Nile (disambiguation)
Entity Linking
Kasturia et al. 2022

Entity index:
- nile → [kb:Nile, kb:Nile_(band),...]
- nile river → [kb:Nile,...]
- ...

Entity linking process
1. Segment query into all possible n-grams
2. Request entity index for each segment
3. Compute commonness scores for each segment-entity pair
4. Remove all entities with commonness of zero
Query Segmentation
Hagen et al. 2012, Hagen et al. 2013

Rank all possible segmentations by the sum of pre-computed segment scores.

The segment score of a segment $s$:

- $s$ **is not** a Wikipedia title
  - Occurrence frequency according to Google n-grams

- $s$ **is** a Wikipedia title
  - $1 +$ Occurrence frequency of most frequent 2-gram in $s$
Query Segmentation
Hagen et al. 2012, Hagen et al. 2013

Rank all possible segmentations by the sum of pre-computed segment scores.

The segment score of a segment $s$:

- $s$ **is not** a Wikipedia title
  - Occurrence frequency according to Google n-grams

- $s$ **is** a Wikipedia title
  - $1 + \text{Occurrence frequency of most frequent 2-gram in } s$

Cut off segmentations with a ratio of less than 0.66:

<table>
<thead>
<tr>
<th>Rank</th>
<th>Segmentation</th>
<th>Score</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>⟨source</td>
<td>of the</td>
<td>nile⟩</td>
</tr>
<tr>
<td>2</td>
<td>⟨source of the nile⟩</td>
<td>333.4 million</td>
<td>0.67</td>
</tr>
<tr>
<td>3</td>
<td>⟨source of the</td>
<td>nile⟩</td>
<td>35.6 million</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Query Interpretation

Combine segmentations with linked entities to build interpretation candidates.

Rank interpretation candidates by a weighted sum of

- Entity commonness
- Relatedness between entities
- Context score between an entity and unlinked segments

Relatedness and context scores are computed with Wikipedia-based joint word-entity embeddings.¹

¹Wikipedia2Vec wikipedia2vec.github.io/wikipedia2vec/
Query Interpretation

Combine segmentations with linked entities to build interpretation candidates.

Rank interpretation candidates by a weighted sum of

- Entity commonness
- Relatedness between entities
- Context score between an entity and unlinked segments

Relatedness and context scores are computed with Wikipedia-based joint word-entity embeddings.¹

¹Wikipedia2Vec [wikipedia2vec.github.io/wikipedia2vec/](http://wikipedia2vec.github.io/wikipedia2vec/)

The query entity linking and interpretation approaches are available as dockerized software and have been integrated into TIREx.
Query Interpretation Analytics

As components of TIREx,…

- 89,289 entities in 2,544 queries from 31 datasets and
- 2,304 interpretations of 1,225 queries from 18 datasets

…have been identified, which can be used and accessed through TIREx.
Query Interpretation Analytics

As components of TIREx, . . .

- 89,289 entities in 2,544 queries from 31 datasets and
- 2,304 interpretations of 1,225 queries from 18 datasets

. . . have been identified, which can be used and accessed through TIREx.

As part of our preliminary analysis, we have found that . . .

- most queries contain entities (98%)
- the number of entities in a query correlates with its length ($\rho = 0.63$)
- most queries are ambiguous (1.8 interpretations per query)
- the number of interpretations does not correlate with the number of relevant documents ($\rho \approx 0$)
## Query Interpretation Examples

Example queries from TREC Web Track 2009 and 2012 datasets.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Interpretation</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>\langle kb:Barack_Obama</td>
<td>family tree\rangle</td>
</tr>
<tr>
<td>2</td>
<td>\langle kb:Barack_Obama</td>
<td>kb:Family_Tree\rangle</td>
</tr>
<tr>
<td>3</td>
<td>\langle obama</td>
<td>kb:Family_Tree\rangle</td>
</tr>
</tbody>
</table>

*For query “obama family tree”*

<table>
<thead>
<tr>
<th>Rank</th>
<th>Interpretation</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>\langle kb:Pork_tenderloin\rangle</td>
<td>0.92</td>
</tr>
</tbody>
</table>

*For query “pork tenderloin”*

<table>
<thead>
<tr>
<th>Rank</th>
<th>Interpretation</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>\langle kb:The_Last_Supper_(Leonardo)</td>
<td>painting\rangle</td>
</tr>
<tr>
<td>2</td>
<td>\langle last supper</td>
<td>kb:Painting\rangle</td>
</tr>
<tr>
<td>3</td>
<td>\langle kb:The_Last_Supper_(Leonardo)</td>
<td>kb:Painting\rangle</td>
</tr>
</tbody>
</table>

*For query “last supper painting”*
Summary
We contributed...
- query entity linking and interpretation software as TIREx components
- 89,289 entity candidates and 2,304 interpretations for 31 and 18 datasets
- preliminary analytics of entities and interpretations
Summary

We contributed...

- query entity linking and interpretation software as TIREx components
- 89,289 entity candidates and 2,304 interpretations for 31 and 18 datasets
- preliminary analytics of entities and interpretations

Thank you! Find more information in the paper and in the following repositories.

www.github.com/webis-de/query-entity-linking
www.github.com/webis-de/query-interpretation
## Evaluation

### Query Entity Linking

<table>
<thead>
<tr>
<th>Entity Linking Tools</th>
<th>MicR</th>
<th>MicR*</th>
<th>MacR</th>
<th>MacR*</th>
<th>MicP</th>
<th>MacP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our approach</td>
<td>0.838</td>
<td>0.859</td>
<td>0.668</td>
<td>0.670</td>
<td>0.035</td>
<td>0.126</td>
</tr>
<tr>
<td>Nordlys ER</td>
<td>0.735</td>
<td>0.776</td>
<td>0.543</td>
<td>0.548</td>
<td>0.002</td>
<td>0.009</td>
</tr>
<tr>
<td>TagMe</td>
<td>0.333</td>
<td>0.410</td>
<td>0.385</td>
<td>0.401</td>
<td>0.328</td>
<td>0.399</td>
</tr>
<tr>
<td>Babelfy</td>
<td>0.320</td>
<td>0.398</td>
<td>0.383</td>
<td>0.398</td>
<td>0.293</td>
<td>0.289</td>
</tr>
<tr>
<td>Smaph</td>
<td>0.314</td>
<td>0.390</td>
<td>0.399</td>
<td>0.413</td>
<td>0.431</td>
<td>0.463</td>
</tr>
<tr>
<td>Dandelion</td>
<td>0.302</td>
<td>0.373</td>
<td>0.414</td>
<td>0.428</td>
<td>0.431</td>
<td>0.500</td>
</tr>
<tr>
<td>Nordlys EL</td>
<td>0.293</td>
<td>0.359</td>
<td>0.579</td>
<td>0.593</td>
<td>0.780</td>
<td>0.731</td>
</tr>
<tr>
<td>Dexter</td>
<td>0.267</td>
<td>0.332</td>
<td>0.359</td>
<td>0.372</td>
<td>0.481</td>
<td>0.462</td>
</tr>
<tr>
<td>FEL</td>
<td>0.250</td>
<td>0.309</td>
<td>0.313</td>
<td>0.324</td>
<td>0.273</td>
<td>0.333</td>
</tr>
<tr>
<td>TextRazor</td>
<td>0.216</td>
<td>0.265</td>
<td>0.372</td>
<td>0.380</td>
<td>0.511</td>
<td>0.445</td>
</tr>
<tr>
<td>Radboud EL</td>
<td>0.213</td>
<td>0.263</td>
<td>0.498</td>
<td>0.507</td>
<td>0.789</td>
<td>0.627</td>
</tr>
<tr>
<td>Falcon</td>
<td>0.204</td>
<td>0.251</td>
<td>0.226</td>
<td>0.234</td>
<td>0.397</td>
<td>0.368</td>
</tr>
<tr>
<td>Ambiverse</td>
<td>0.011</td>
<td>0.013</td>
<td>0.259</td>
<td>0.259</td>
<td>0.750</td>
<td>0.263</td>
</tr>
</tbody>
</table>

### Entity Recognition Tools

<table>
<thead>
<tr>
<th>Entity Recognition Tools</th>
<th>MicR</th>
<th>MicR*</th>
<th>MacR</th>
<th>MacR*</th>
<th>MicP</th>
<th>MacP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWS Comprehend</td>
<td>0.229</td>
<td>–</td>
<td>0.476</td>
<td>–</td>
<td>0.604</td>
<td>0.616</td>
</tr>
<tr>
<td>MITIE</td>
<td>0.114</td>
<td>–</td>
<td>0.358</td>
<td>–</td>
<td>0.797</td>
<td>0.463</td>
</tr>
<tr>
<td>Flair NER</td>
<td>0.129</td>
<td>–</td>
<td>0.374</td>
<td>–</td>
<td>0.787</td>
<td>0.487</td>
</tr>
<tr>
<td>LingPipe NER</td>
<td>0.109</td>
<td>–</td>
<td>0.321</td>
<td>–</td>
<td>0.497</td>
<td>0.410</td>
</tr>
<tr>
<td>DeepPavlov</td>
<td>0.048</td>
<td>–</td>
<td>0.269</td>
<td>–</td>
<td>0.478</td>
<td>0.305</td>
</tr>
<tr>
<td>Stanford NER</td>
<td>0.011</td>
<td>–</td>
<td>0.257</td>
<td>–</td>
<td>0.563</td>
<td>0.261</td>
</tr>
<tr>
<td>OpenNLP</td>
<td>0.000</td>
<td>–</td>
<td>0.246</td>
<td>–</td>
<td>0.000</td>
<td>0.246</td>
</tr>
<tr>
<td>No-Entity Baseline</td>
<td>0.000</td>
<td>0.000</td>
<td>0.246</td>
<td>0.246</td>
<td>0.000</td>
<td>0.246</td>
</tr>
</tbody>
</table>
**Evaluation**

**Experiment: Query Interpretation**

Comparison to pairs of *greedy interpretation finding* [Hasibi et al., 2014] and various entity linking approaches.

**Data:**
- Test split of our new query interpretation dataset (544 queries)
- Interpretations that are at least “moderately likely”
## Evaluation

### Query Interpretation Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Recall</th>
<th>Precision</th>
<th>F_1</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our approach</td>
<td>0.295</td>
<td>0.336</td>
<td>0.283</td>
<td>47</td>
</tr>
<tr>
<td>Dexter</td>
<td>0.230</td>
<td>0.312</td>
<td>0.246</td>
<td>282</td>
</tr>
<tr>
<td>Nordlys EL</td>
<td>0.189</td>
<td>0.278</td>
<td>0.207</td>
<td>1,533</td>
</tr>
<tr>
<td>Radboud EL</td>
<td>0.144</td>
<td>0.199</td>
<td>0.155</td>
<td>200</td>
</tr>
<tr>
<td>Smaph</td>
<td>0.176</td>
<td>0.243</td>
<td>0.190</td>
<td>116,425</td>
</tr>
<tr>
<td>Dandelion</td>
<td>0.166</td>
<td>0.226</td>
<td>0.177</td>
<td>74</td>
</tr>
<tr>
<td>TagMe</td>
<td>0.165</td>
<td>0.216</td>
<td>0.175</td>
<td>99</td>
</tr>
<tr>
<td>Babelfy</td>
<td>0.112</td>
<td>0.160</td>
<td>0.124</td>
<td>49</td>
</tr>
<tr>
<td>TextRazor</td>
<td>0.098</td>
<td>0.131</td>
<td>0.105</td>
<td>367</td>
</tr>
<tr>
<td>FEL</td>
<td>0.133</td>
<td>0.173</td>
<td>0.141</td>
<td>22</td>
</tr>
<tr>
<td>Ambiverse</td>
<td>0.007</td>
<td>0.011</td>
<td>0.009</td>
<td>53</td>
</tr>
</tbody>
</table>