Query Interpretations from Entity-Linked Segmentations

WSDM 2022

Vaibhav Kasturia\textsuperscript{1}

Marcel Gohsen\textsuperscript{2}

Matthias Hagen\textsuperscript{3}

\textsuperscript{1,3} Martin-Luther-Universität Halle-Wittenberg

\textsuperscript{2} Bauhaus-Universität Weimar

\textsuperscript{2,3} Webis
What is the intent?

source of the nile
What is the intent?
Is it ...
What is the intent?
Is it ... or ...?

source | of the | Nile

Source_of_the_Nile
What is the intent?
What does Google think?

Google search for 'source of the nile'

Nile/Sources

- Blue Nile River
- Atbarah
- White Nile
What is the intent?
What does Google think?

⇒ Where is the board game?
source of the nile

Explicit entities:
Computational Problems
Entity Linking

source of the Nile

Explicit entities:

Nile
Computational Problems

Entity Linking

source of the Nile

Explicit entities:

Nile

Source_of_the_Nile
Computational Problems
Entity Linking

source of the Nile

Explicit entities:

Nile

Source_of_the_Nile

Implicit entities:
Computational Problems
Entity Linking

Explicit entities:

![Nile](image1)

source of the Nile

![Source_of_the_Nile](image2)

Implicit entities:

White_Nile

![White_Nile](image3)
Computational Problems
Entity Linking

Explicit entities:

source of the Nile

Implicit entities:

White Nile
Blue Nile
Computational Problems
Entity Linking

Explicit entities:

source of the Nile

Implicit entities:

White Nile  Source of the Nile  Blue Nile
Computational Problems
Query Interpretation

Entity Linking

Query
Segmentation

Entities
Segmentation Skeletons

Segment-Entity Linking

Entity-based Interpretations

Query
Entities
Segmentation
Segment-Entity Linking
Entity-based Interpretations & Confidence Scores

source of the nile

nile → Nile

source of the nile → ...

source | of the | Nile

Source_of_the_Nile

= 1.0

source | of the | Nile

Source_of_the_Nile

= 0.8

source | of the | Nile

Source_of_the_Nile

...
Computational Problems

Query Interpretation

Query

Entity Linking

Query Segmentation

Entities

Segmentation Skeletons

Segment-Entity Linking

Entity-based Interpretations

Interpretation Scoring

Entity-based Interpretations & Confidence Scores

source of the nile

nile → Nile

source | of the | Nile

Source_of_the_Nile

source | of the | Nile = 1.0

Source_of_the_Nile = 0.8

...
Computational Problems

Query Interpretation

Entity Linking

Query Segmentation

Entities

Segmentation Skeletons

Segment-Entity Linking

Entity-based Interpretations

Interpretation Scoring

Entity-based Interpretations & Confidence Scores

source of the nile

nile → Nile

source | of the | nile

source of the nile → ...

source of the nile → ...

source | of the | Nile

Source_of_the_Nile

... 

source | of the | Nile

Source_of_the_Nile

= 1.0

= 0.8

... 

source | of the | Nile

Source_of_the_Nile

... 

source | of the | Nile

Source_of_the_Nile

= 1.0

= 0.8

...
Computational Problems
Query Interpretation

Query → Entity Linking → Entities → Segmentation → Segmentation Skeletons → Segment-Entity Linking → Entity-based Interpretations → Interpretation Scoring → Entity-based Interpretations & Confidence Scores

source of the nile

nile → Nile

source | of the | Nile

Source_of_the_Nile

... = 1.0

... = 0.8
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></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
- Combination annotated dataset: 2,800 queries
- Recall-oriented explicit entity recognizer
- Segmentation-based query interpretation method
- Comparison to existing entity linking and interpretation methods
Conclusion

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

Future work:
- Implicit entity recognition
- Query interpretation for conversational systems
Conclusion

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

Future work:
- Implicit entity recognition
- Query interpretation for conversational systems

Thank you! :-}