Vaibhav Kasturia[1] **Marcel Gohsen**[2] Matthias Hagen[1]

[1]Martin-Luther-Universität Halle-Wittenberg, [2]Bauhaus-Universiät Weimar

[1]first.last@informatik.uni-halle.de [2]first.last@uni-weimar.de

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:

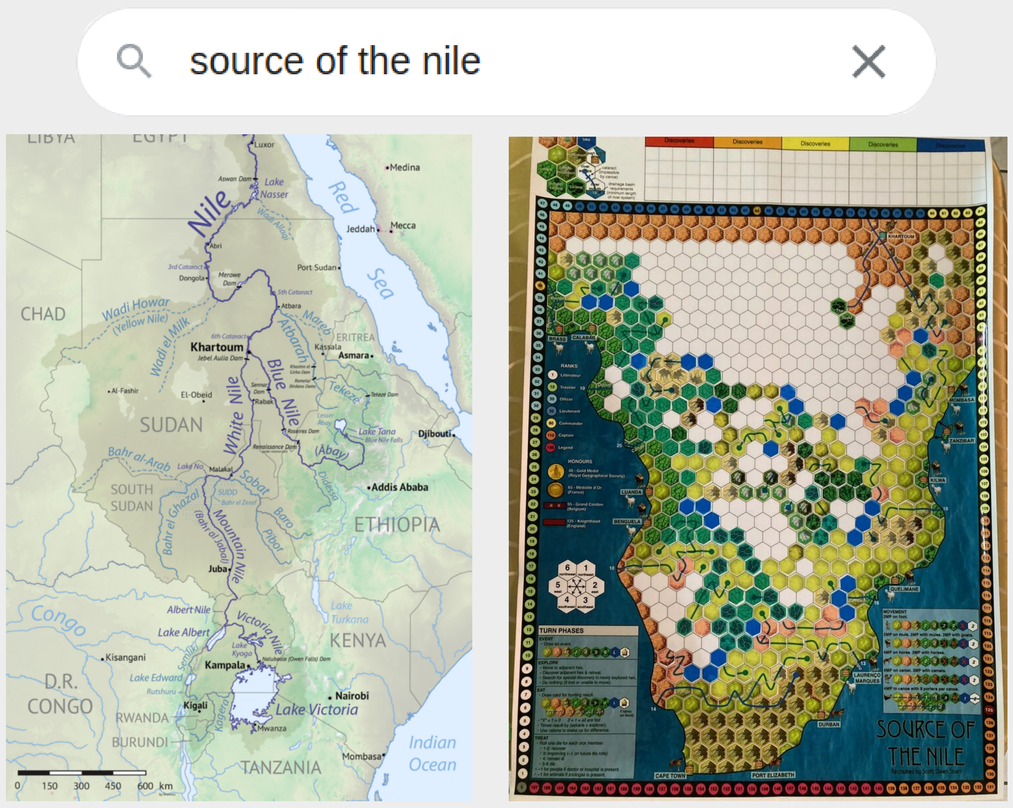
`source of the nile` → `Blue_Nile`

`source of the nile` → `White_Nile`

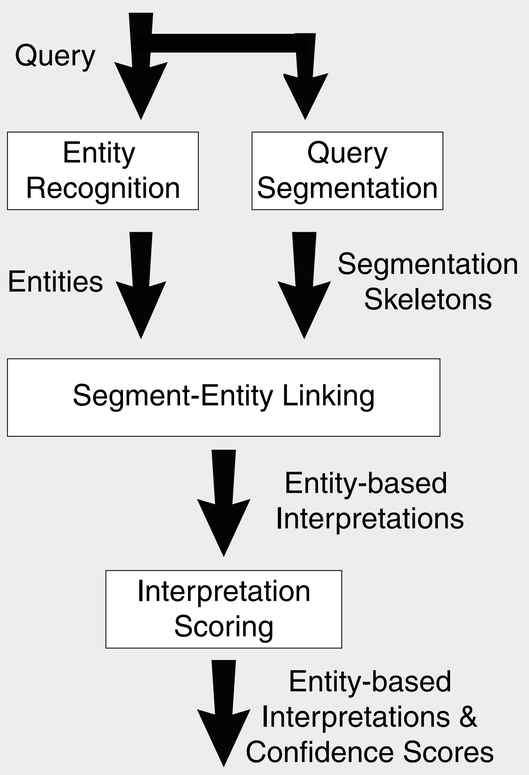
Related entities:

`Egypt`, `Sudan`, `The_Settlers_of_Catan`, ...

Ambiguous Query Example



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

New Dataset: Webis-QInC-22

Combined queries from
- Yahoo Search Query Log To Entities
- ERD 2014 (Carmel et al., 2014)
- DBpedia-Entity v2 (Hasibi et al., 2017)

Manually annotated 2,800 queries with
- Explicit, implicit and related entities
- Entity-based query interpretations
- Diffculty and relevance assessments

Query Length	Count	Mentions	Explicit Entities	Interpretations
1	206	0.86	2.47	2.79
2	610	1.08	2.16	2.60
3	775	1.17	2.07	2.40
4	540	1.34	2.00	2.13
5	290	1.51	1.65	1.60
6	154	1.56	1.74	1.81
7	96	1.76	1.56	1.47
8–14	129	1.83	1.38	1.30
1–14	2,800	1.27	1.99	2.22

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).

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

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