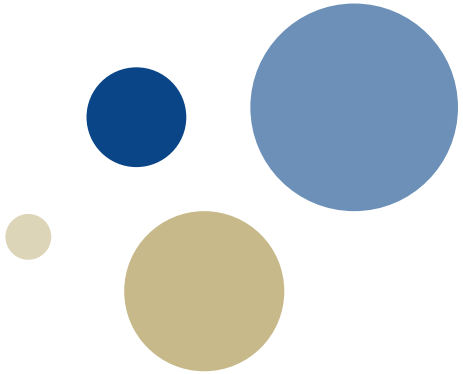
NTNU – Trondheim
Norwegian University of
Science and Technology



Entity Linking in Queries: Tasks and Evaluation

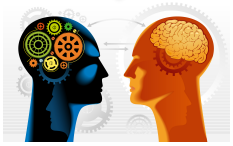
Faegheh Hasibi, Krisztian Balog, and Svein Erik Bratsberg

ICTIR conference, September 2015

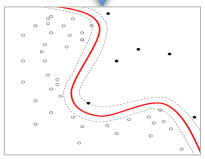
Entity linking

Definition from Wikipedia:

In natural language processing, **entity linking**, **named entity disambiguation** (NED), **named entity recognition and disambiguation** (NERD) or **named entity normalization** (**NEN**)[1] is the task of determining the identity of entities mentioned in text. It is distinct from named entity recognition (NER) in that it identifies not the occurrence of names (and a limited classification of those), but their reference.



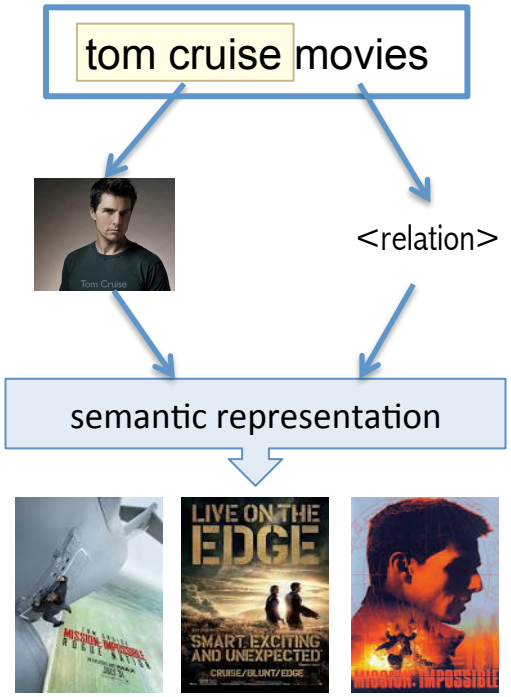
https://en.wikipedia.org/wiki/Natural_language_processing



https://en.wikipedia.org/wiki/Statistical_classification

Why entity linking in queries?

- ~70% of queries contain entities
- To exploit semantic representation of queries



Improves:

- Ad-hoc document retrieval
- Entity retrieval
- Query understanding
- Understanding users' task (Tasks track, TREC)

J. Pound, P. Mika, and H. Zaragoza. Ad-hoc object retrieval in the web of data. *In Proc. of WWW '10.*

It is different …

Different from conventional entity linking:

- Limited or even no context
- A mention may be linked to **more than one** entity

{France, FIFA world cup}
or
{France national football team, FIFA world cup}

In this talk

How entity linking should be performed for queries?

- **Task:**
  "Semantic Mapping" or "Interpretation Finding"?
- **Evaluation metrics**
- **Test collections**
- **Methods**

In this talk

How entity linking should be performed for queries?

- **Task:**
  "Semantic Mapping" or "Interpretation Finding"?
- **Evaluation metrics**
- **Test collections**
- **Methods**

Entity linking

- Output is **set of entities**
- Each mention is linked to a single entity
- Mentions do not overlap
- Entities are explicitly mentioned

obama mother	the music man	new york pizza manhattan
{Barack Obama}	{The Music Man}	{New York City, Manhattan}

Semantic mapping

- Output is **ranked list of entities**
- Mentions can overlap and be linked to multiple entities
- Entities may not be explicitly mentioned
- Entities do not need to form semantically compatible sets
- False positive are not penalized

obama mother	the music man	new york pizza manhattan
Ann Dunham Barack Obama	The Music Man The Music Man (1962 film) The Music Man (2003 film) …	New York City New York-style pizza Manhattan Manhattan pizza ...

Interpretation finding

- Output is **set(s) of semantically related entity sets**
- Each entity set is an interpretation of the query
- Mention do not overlap within a set

obama mother	the music man	new york pizza manhattan

{{Barack Obama}}	{
	{The Music Man}
	{The Music Man (1962 film)},
	{The Music Man (2003 film)}
	}
		{
		{New York City, Manhattan},
		{New York-style pizza, Manhattan}
		}

D. Carmel, M.-W. Chang, E. Gabrilovich, B.J. P. Hsu, and K. Wang. ERD: Entity recognition and disambiguation challenge, *2014.*

Tasks summary

	Entity Linking	Semantic Mapping	Interpretation Finding
Entities explicitly mentioned?	Yes	No	Yes
Mentions can overlap?	No	Yes	No*
Results format	Set	Ranked list	Sets of sets
Evaluation criteria	Mentioned entities found	Relevant entities found	Interpretations found
Evaluation metrics	Set-based	Ranked-based	Set-based

* Not within the same interpretation

Entity linking requirements are relaxed in semantic mapping.

In this talk

How entity linking should be performed for queries?

- **Task:**
  "Semantic Mapping" or "Interpretation Finding"?
- **Evaluation metrics**
- **Test collections**
- **Methods**

Evaluation

- Macro-averaged metrics (precision, recall, F-measure)
- Matching condition:
  - Interpretation sets should exactly match the ground truth

$$\mathrm{P} = \frac{|I \cap \hat{I}|}{|I|} \qquad \mathrm{R} = \frac{|I \cap \hat{I}|}{|\hat{I}|}$$

System query interpretation

$$I = \{E_1, \ldots, E_n\}$$

Ground truth query interpretation

$$\hat{I} = \{\hat{E}_1, \ldots, \hat{E}_m\}$$

What if $I = \emptyset$ or $\hat{I} = \emptyset$?

D. Carmel, M.-W. Chang, E. Gabrilovich, B.J. P. Hsu, and K. Wang. Entity recognition and disambiguation challenge, *2014.*

Evaluation (revisited)

Solution:

$$\mathrm{P}_{\mathrm{int}} = \begin{cases} |I \cap \hat{I}|/|I|, & I \neq \emptyset \\ 1, & I = \emptyset, \hat{I} = \emptyset \\ 0, & I = \emptyset, \hat{I} \neq \emptyset \end{cases}$$

$$\mathrm{R}_{\mathrm{int}} = \begin{cases} |I \cap \hat{I}|/|\hat{I}|, & \hat{I} \neq \emptyset \\ 1, & \hat{I} = \emptyset, I = \emptyset \\ 0, & \hat{I} = \emptyset, I \neq \emptyset \end{cases}$$

System output **matches** ground truth.

System output **does not match** ground truth.

This evaluation is methodologically correct, but **strict**.

Lean evaluation

- Partial matches are not rewarded in $\mathrm{P_{int}}$, $\mathrm{R_{int}}$
- **E.g.** {{New York City, Manhattan}} ≠ {{New York City}, {Manhattan}}

Solution: Combine them with **entity-based** metrics.

$$\mathrm{P} = \frac{\mathrm{P_{int}} + \mathrm{P_{ent}}}{2} \qquad \mathrm{R} = \frac{\mathrm{R_{int}} + \mathrm{R_{ent}}}{2}$$

$$\mathrm{P_{ent}} = \begin{cases} |E \cap \hat{E}|/|E|, & E \neq \emptyset \\ 1, & E = \emptyset, \hat{E} = \emptyset \\ 0, & E = \emptyset, \hat{E} \neq \emptyset \end{cases} \qquad \mathrm{R_{ent}} = \begin{cases} |E \cap \hat{E}|/|\hat{E}|, & \hat{E} \neq \emptyset \\ 1, & \hat{E} = \emptyset, E = \emptyset \\ 0, & \hat{E} = \emptyset, E \neq \emptyset \end{cases}$$

In this talk

How entity linking should be performed for queries?

- **Task:**
  "Semantic Mapping" or "Interpretation Finding"?
- **Evaluation metrics**
- **Test collections**
- **Methods**

Test collections - ERD

The ERD challenge introduced two test collections:

- ERD-dev (91 queries)[1]
- ERD-test (500 queries)
  - Unavailable for traditional offline evaluation

Annotation rules:

- The longest mention is used for entities
- Only proper noun entities are annotated (e.g., companies, locations)
- Overlapping mentions are not allowed within a single interpretation

 ERD-dev is not suitable for training purposes (small)

[1] http://web-ngram.research.microsoft.com/erd2014/Datasets.aspx

Test collections - YSQLE

Yahoo Search Query Log to Entities (YSQLE)

- 2398 queries, manually annotated with Wikipedia entities
- Designed for training and testing entity linking systems for queries

Issues:

- Not possible to automatically form interpretation sets
  - E.g. Query “france world cup 1998”
- Linked entities are not necessarily mentioned explicitly
  - E.g. Query “charlie sheen lohan” is annotated with Anger Management (TV series)
- Annotations are not always complete
  - E.g. Query “louisville courier journal” is not annotated with Louisville, Kentucky

YSQLE is meant for the **semantic mapping** task

Yahoo! Webscope L24 dataset - Yahoo! search query log to entities, v1.0. http://webscope.sandbox.yahoo.com/

Test collections - Y-ERD

Y-ERD is manually re-annotated based on:

- YSQLE annotations
- ERD rules

Additional rules:

- Site search queries are not linked
  - E.g. Query “facebook obama slur” is only linked to Barack Obama
- Clear policy about misspelled mentions
  - Two versions of Y-ERD is made available

Y-ERD is made publicly available
http://bit.ly/ictir2015-elq

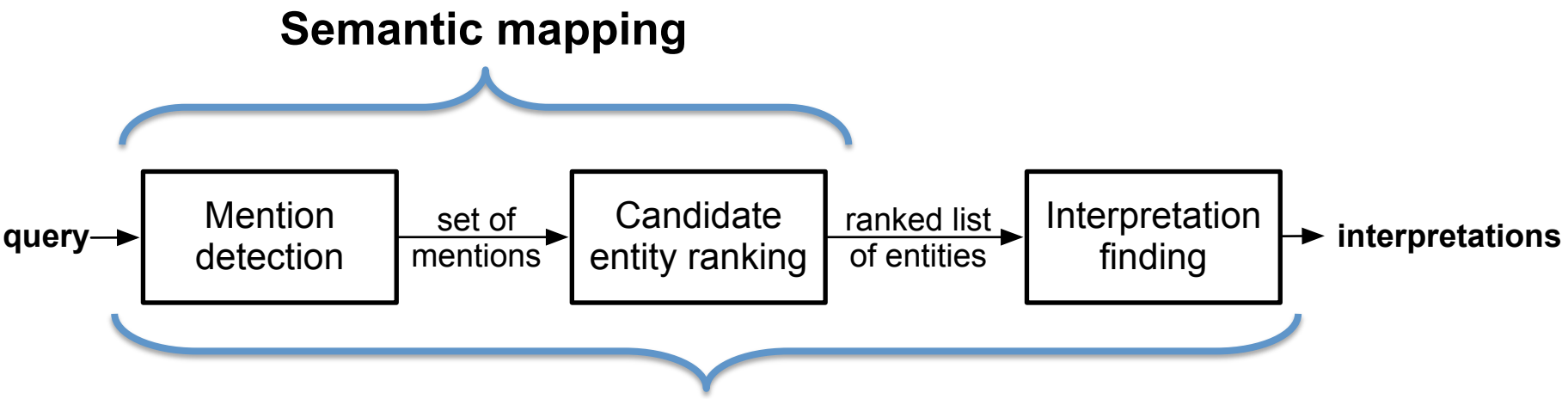
In this talk

How entity linking should be performed for queries?

- **Task:**
  "Semantic Mapping" or "Interpretation Finding"?
- **Evaluation metrics**
- **Test collections**
- **Methods**

Methods

Pipeline architecture for two tasks:



The goal of entity linking in queries

Mention detection

Entity name variants are gathered from:

- **KB:** A manually curated knowledge base (DBpedia)
- **WEB:** Freebase Annotations of the ClueWeb Corpora (FACC)

	YSQLE	Y-ERD	ERD
KB	0.7489	0.7976	0.8556
Web	0.9127	0.9716	0.9956
KB+Web	**0.9163**	**0.9724**	**1.0000**

Recall in mention detection step

E. Gabrilovich, et. al. FACC1: Freebase annotation of ClueWeb corpora, 2013.

Methods

Pipeline architecture for two tasks:

Candidate entity ranking

Ranking using language models:

$$P(e|q) = \frac{P(q|e)P(e)}{P(q)} \propto P(e)P(q|\theta_e)$$

Scores should be comparable across queries

– $P(q)$ should be considered

Mixture of Language Models (MLM)

$$\sum_{f \in F} \mu_f P(t|\theta_{e_f})$$

$$P(e|q) = P(e) \frac{\prod_{t \in q} P(t|\theta_e)^{P(t|q)}}{\prod_{t \in q} P(t|C)^{P(t|q)}}$$

P(q) - Query length normalization

- P. Ogilvie and J. Callan. Combining document representations for known-item search. In *Proc. of SIGIR '03*, 2003
- W. Kraaij and M. Spitters. Language models for topic tracking. In Language Modeling for Information Retrieval, 2003.

Candidate entity ranking

Combining MLM and Commonness:

$$P(e|q) \propto \arg\max_{m \in q} \underbrace{P(e|m)}\underbrace{P(q|e)}$$

Commonness
Probability of entity ***e*** being the link target of mention ***m***

$$\frac{n(m,e)}{\sum_{e'} n(m,e')}$$

Query length normalized MLM score

Candidate entity ranking

Semantic mapping results on YSQLE:

	MAP	S@1	MRR
CMNS	0.6334	0.5751	0.6442
MLM	0.4582	0.3601	0.4638
MLMc	0.6228	0.5413	0.6312
MLMcg	**0.7078**	**0.6403**	**0.7151**
TAGME	0.6230	0.6016	0.6385

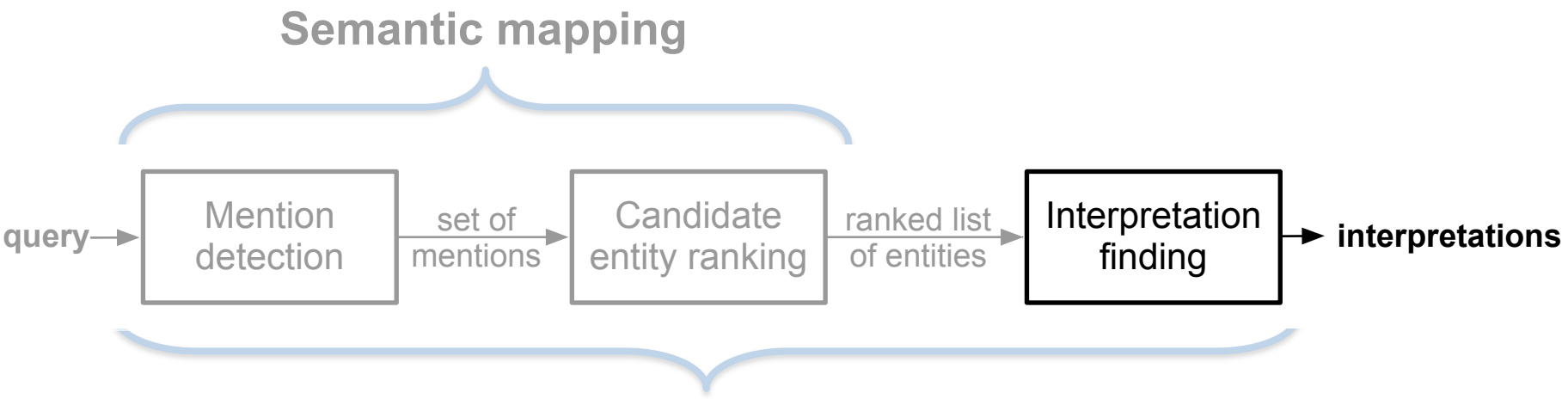
TAGME is an entity linking system.

- Should not be evaluated using rank-based metrics
- Should not be compared with semantic mapping results

P. Ferragina and U. Scaiella. TAGME: On-the-fly annotation of short text fragments. In *Proc. of CIKM 2010*.

Methods

Pipeline architecture for two tasks:



Interpretation finding

Greedy Interpretation Finding (GIF):

Example query: “jacksonville fl riverside”

Mention	Entity	Score
“jacksonville fl”	Jacksonville Florida	0.9
“jacksonville”	Jacksonville Florida	0.8
“riverside”	Riverside Park (Jacksonville)	0.6
“jacksonville fl”	Naval Station Jacksonville	0.2
“riverside	Riverside (band)	0.1

Step 2:
Pruning containment mentions

Step 1:
Pruning based on a score threshold (0.3)

Step 3:
Forming interpretation sets

{ {Jacksonville Florida, Riverside Park (Jacksonville)} }

Interpretation finding

	Method	Strict eval.			Lean eval.		
		P	R	F	P	R	F
Y-ERD	TopRanked	0.4554	0.4542	0.4545	0.4771	0.465	0.4689
	TAGME	0.6647	0.6642	0.6643	0.6821	0.6853	0.6815
	GIF-CMNS	0.6927	0.6938	0.6929	0.7093	0.7072	0.7062
	GIF-MLM	0.5259	0.5254	0.5255	0.5363	0.5387	0.5361
	GIF-MLMc	0.6351	0.6354	0.6348	0.6422	0.642	0.6409
	GIF-MLMcg	**0.7191**	**0.7213**	**0.7195**	**0.7305**	**0.7308**	**0.7288**
ERD-dev	TopRanked	0.3846	0.3645	0.3700	0.4231	0.3837	0.3956
	TAGME	0.7143	0.7015	0.7051	**0.7418**	**0.7372**	**0.7333**
	GIF-CMNS	0.5824	0.5824	0.5824	0.6071	0.5962	0.5998
	GIF-MLM	0.5824	0.5608	0.5659	0.5934	0.5718	0.5760
	GIF-MLMc	0.7253	0.7037	0.7088	0.7445	0.7174	0.7234
	GIF-MLMcg	**0.7143**	**0.7125**	**0.7114**	0.7335	0.7262	0.7260

Take home messages

- Entity linking in queries is different from documents
- Different flavors, different evaluation criteria:
  - Interpretation finding **(yes)**
  - Semantic mapping **(no)**
- Ultimate goal should be interpretation finding
- SM and EL should not be compared to each other
- Resources are available at http://bit.ly/ictir2015-elq

SIGIR
Special Interest Group
on Information Retrieval

Thanks!