Entity Linking

KB

Wikipedia

DBpedia http://wiki.dbpedia.org/

YAGO https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/

Freebase [over]

KnowItAll http://www.gabormelli.com/RKB/KnowItAll_System

ReadTheWeb http://rtw.ml.cmu.edu/rtw/

Probase https://www.microsoft.com/en-us/research/project/probase/

Wikidata https://www.wikidata.org/wiki/Wikidata:Main_Page

Pipeline

- Identifying Target Mentions (NER tool or n-gram)
  - Mentions in the input text that should be Wikified
- Identifying Candidate Titles (Search, String matching)
  - Candidate Wikipedia titles that could correspond to each mention
- Candidate Title Ranking (Learning to Rank, Graph)
  - Rank the candidate titles for a given mention
- NIL Detection and Clustering
  - Identify mentions that do not correspond to a Wikipedia title
  - Entity Linking: cluster NIL mentions that represent the same entity.

Wikifiers

TagMe: https://tagme.d4science.org/tagme/

Illinois Wikifier: https://cogcomp.cs.illinois.edu/page/demo_view/Wikifier

Ed-visee Wikifier: http://lagos.lti.cs.cmu.edu:8080/disambiguate.jsp

Microsoft: https://www.microsoft.com/cognitive-services/en-us/entity-linking-intelligence-service

DBpedia Spotlight: http://dbpedia-spotlight.github.io/demo/

Wikifiers

TagMe: https://tagme.d4science.org/tagme/

Unable to reproduce: Hasibi, Faegheh, Krisztian Balog, and Svein Erik Bratsberg. "On the Reproducibility of the TAGME Entity Linking System." *European Conference on Information Retrieval*. Springer International Publishing, 2016.

Rationale is simple and quick: Ferragina, Paolo, and Ugo Scaiella. "Tagme: on-the-fly annotation of short text fragments (by wikipedia entities)." *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 2010.

Wikifiers

Illinois Wikifier: https://cogcomp.cs.illinois.edu/page/demo_view/Wikifier

- **Clickable output**
- **Only supports English**
- **Heavily relies on Popularity (sometimes)**

Wikifiers

Ed-visee Wikifier: http://lagos.lti.cs.cmu.edu:8080/disambiguate.jsp

- **Support English, Chinese, and Spanish**
- **Change between NER/n-gram**
- **Change threshold and other settings**

Wikifiers

Microsoft: https://www.microsoft.com/cognitive-services/en-us/entity-linking-intelligence-service

- Not free
- Blackbox
- Good performance

Sample Text

BERLIN -- Usain Bolt crossed the finish line, saw his **record-setting** time on the clock and spread his arms as if he were soaring like a bird.

Bolt ran his latest unforgettable race at Olympic Stadium in Berlin, the **history-filled** home of the 1936 Olympics where Jesse Owens became the world's biggest track star. Bolt lives in Owens' stratosphere now, having set the 100 world record three times and also owning the 200-meter record thanks to the 19.30 he ran in Beijing to break Michael Johnson's **12-year-old** mark.

Now he has added the world championship, last won by Gay in 2007, to his Olympic title.

A Stanford professor estimated he could've gone 9.55 if he'd run full out through the line in Beijing. Bolt almost made that guy look like a genius.

Experiments

(demo)

Problems

- Pre-train model vs Emerging entities

http://stats.grok.se/en/201506/donald%20trump

- Running time

- Errors are compounded