WIKIPEDIA

Entity linking

In natural language processing, **entity linking**, **named entity linking** (NEL),[1] **named entity disambiguation** (NED), **named entity recognition and disambiguation** (NERD) or **named entity normalization** (**NEN**)[2] is the task of determining the identity of entities mentioned in text. For example, given the sentence "Paris is the capital of France", the idea is to determine that "Paris" refers to the city of Paris and not to Paris Hilton or any other entity that could be referred as "Paris". NED is different from named entity recognition (NER) in that NER identifies the occurrence or mention of a named entity in text but it does not identify which specific entity it is.

Entity linking requires a knowledge base containing the entities to which entity mentions can be linked. A popular choice for entity linking on open domain text are knowledge-bases based on Wikipedia,[2][3] in which each page is regarded as a named entity. NED using Wikipedia entities has been also called **wikification** (see Wikify! an early entity linking system[4]). A knowledge base may also be induced automatically from training text[5] or manually built.[6]

Named entity mentions can be highly ambiguous; any entity linking method must address this inherent ambiguity. Various approaches to tackle this problem have been tried to date. In the seminal approach of Milne and Witten, supervised learning is employed using the anchor texts of Wikipedia entities as training data.[7] Other approaches also collected training data based on unambiguous synonyms.[8]. Kulkarni *et al.* exploited the common property that topically coherent documents refer to entities belonging to strongly related types.[9]

Entity linking has been used to improve the performance of information retrieval systems[2] and to improve search performance on digital libraries.[10][11] NED is also a key input for Semantic Search.[12] For example, entity linking has been successfully used to validate the output of Named-entity recognition methods by using an underlying assumption that every text has a specific scope and context, and therefore entities mentioned in that text should have a semantic relationship between them.[13] Semantic similarity measures can then be used to filter out errors by identifying entities out of context, even when multiple interpretations are possible for the same entity.

Entity Linking evaluation campaigns (https://www.nist.gov/tac/2013/KBP/EntityLinking/index.html) are organized by the U.S. National Institute of Standards and Technology (NIST) in the context of the Knowledge Base Population task (https://www.nist.gov/tac/2013/KBP/) of the Text Analysis Conference.

Contents

Wikification services

- Wikify![4]
- Wikifier.org [14], from the Jožef Stefan Institute, Slovenia
- Illinois Wikifier [15]

- Wikipedia miner Wikifier [16]
- Exalabs Wikifier [17]

General purpose systems

An overview of selected systems is provided by the GERBIL General Entity Annotation Benchmark Framework [18] [19].

- AIDA (http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/aida/) [20] [21]
- Babelfy (http://babelfy.org/) [22] [23]
- DBpedia Spotlight (http://demo.dbpedia-spotlight.org/)
- Dexter (http://dexter.isti.cnr.it/)
- Entityclassifier.eu (http://entityclassifier.eu/thd/about/) [24]
- FRED (http://wit.istc.cnr.it/stlab-tools/fred)
- FREME e-Entity (http://www.freme-project.eu/the-project/what-is-freme/)
- FOX+ AGDISTIS (http://aksw.org/Projects/FOX.html) [25] [26]
- KEA [27]
- NERD-ML (http://nerd.eurecom.fr/documentation) [28]
- TagMe (https://services.d4science.org/web/tagme/documentation) [29]
- WAT [30]
- xLisa (http://km.aifb.kit.edu/sites/xlisa/) [31]

See also

- Controlled vocabulary
- Explicit semantic analysis
- Geoparsing
- Information extraction
- Linked data
- Named entity
- Named-entity recognition
- Record linkage
- Word sense disambiguation
- Author Name Disambiguation

References

1. Hachey, Ben; Radford, Will; Nothman, Joel; Honnibal, Matthew; Curran, James R. (2013-01-01). "Artificial Intelligence, Wikipedia and Semi-Structured ResourcesEvaluating Entity Linking with Wikipedia" (http://www.sciencedirect.com/science/article/pii/S0004370212000446). *Artificial Intelligence*. **194**: 130–150. doi:10.1016/j.artint.2012.04.005 (https://doi.org/10.1016/j.artint.2012.04.005).
2. M. A. Khalid, V. Jijkoun and M. de Rijke (2008). The impact of named entity normalization on information retrieval for question answering (http://ilps.science.uva.nl/biblio/impact-named-entity-normalization-information-retrieval-question-answering). Proc. ECIR.
3. Xianpei Han, Le Sun and Jun Zhao (2011). Collective entity linking in web text: a graph-based method (http://www.nlpr.ia.ac.cn/2011papers/gjhy/gh133.pdf). Proc. SIGIR.
4. Rada Mihalcea and Andras Csomai (2007)Wikify! Linking Documents to Encyclopedic Knowledge (https://digital.library.unt.edu/ark:/67531/metadc31001/m2/1/high_res_d/Mihalcea-2007-Wikify-Linking_Documents_to_Encyclopedic.pdf). Proc. CIKM.

5. Aaron M. Cohen (2005). Unsupervised gene/protein named entity normalization using automatically extracted dictionaries. Proc. ACL-ISMB Workshop on Linking Biological Literature, Ontologies and Databases: Mining Biological Semantics, pp. 17–24.
6. Wikidata
7. David Milne and Ian H. Witten (2008). Learning to link with Wikipedia. Proc. CIKM.
8. Zhang, Wei; Jian Su; Chew Lim Tan (2010). "Entity Linking Leveraging Automatically Generated Annotation". *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*.
9. Kulkarni, Sayali; Singh, Amit; Ramakrishnan, Ganesh; Chakrabarti, Soumen (2009). *Collective annotation of Wikipedia entities in web text*. Proc. 15th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD). doi:10.1145/1557019.1557073 (https://doi.org/10.1145/1557019.1557073). ISBN 9781605584959.
10. Hui Han, Hongyuan Zha, C. Lee Giles, "Name disambiguation in author citations using a K-way spectral clustering method," ACM/IEEE Joint Conference on Digital Libraries 2005 (JCDL 2005): 334-343, 2005
11. [1] (http://www.dnb.de/EN/Wir/Projekte/Abgeschlossen/nerd.html)
12. STICS (https://stics.mpi-inf.mpg.de)
13. Grego, Tiago; Couto, Francisco (2013). "Identification of chemical entities in patent documents". *PLoS ONE*. doi:10.1145/1557019.1557073 (https://doi.org/10.1145/1557019.1557073).
14. "Wikifier" (http://wikifier.org). *wikifier.org*.
15. "Cognitive Computation Group" (http://cogcomp.cs.illinois.edu/page/software_view/Wikifier). *cogcomp.cs.illinois.edu*.
16. "Wikipedia Miner Wikifier" (http://www.nzdl.org/wikification/). *www.nzdl.org*.
17. "Exalabs - Wikifier" (http://labs.exalead.com/project/wikifier). *labs.exalead.com*.
18. "GERBIL — Agile Knowledge Engineering and Semantic Web (AKSW)" (http://aksw.org/Projects/GERBIL.html). *aksw.org*.
19. "AKSW/gerbil: Available APIs and approaches" (https://github.com/AKSW/gerbil/wiki/Available-APIs-and-approaches). *GitHub*.
20. Hoffart, J., Yosef, M. A., Bordino, I., Fürstenau, H., Pinkal, M., Spaniol, M., Taneva, B., Thater, S., and Weikum, G. (2011). Robust disambiguation of named entities in text (http://www.anthology.aclweb.org/D/D11/D11-1072.pdf). In EMNLP
21. "AIDA: Accurate Online Disambiguation of Named Entities in Text and Tables" (http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/aida/). *www.mpi-inf.mpg.de*. Max-Planck-Institut für Informatik.
22. Moro, A., Raganato, A., and Navigli, R. (2014).Entity Linking meets Word Sense Dis- ambiguation: a Unified Approach (http://wwwusers.di.uniroma1.it/~navigli/pubs/TACL_2014_Babelfy.pdf)
23. "Babelfy™ Multilingual Word Sense Disambiguation and Entity Linking together!" (http://babelfy.org). *babelfy.org*.
24. "THD Demo" (http://entityclassifier.eu). *entityclassifier.eu*.
25. Usbeck, R., Ngomo, A. N., Röder, M., Gerber, D., Coelho, S. A., Auer, S., and Both, A. (2014). AGDISTIS - graph-based disambiguation of named entities using linked data. In ISWC (https://www.researchgate.net/profile/Ricardo_Usbeck/publication/264412537_AGDISTIS_-_Graph-Based_Disambiguation_of_Named_Entities_using_Linked_Data/links/553df6de0cf2c415bb0f85fd.pdf)
26. "AGDISTIS — Agile Knowledge Engineering and Semantic Web (AKSW)" (http://aksw.org/Projects/AGDISTIS.html). *aksw.org*.
27. "KEA - Named Entity Recognition, Disambiguation, Linking" (http://s16a.org/kea). *s16a.org*. Semantic Multimedia, University of Potsdam.
28. "NERD: Named Entity Recognition and Disambiguation" (http://nerd.eurecom.fr/). *nerd.eurecom.fr*.
29. "TAGME: on-the-fly annotation of short text fragments!" (https://tagme.d4science.org/tagme/). *tagme.d4science.org*. Retrieved 2016-09-25.
30. https://pdfs.semanticscholar.org/edbd/1c220c69e6c655b7b6c956dc227e3a0f021c.pdf
31. http://www.vldb.org/pvldb/vol7/p1693-zhang.pdf