Location-Aware Named Entity Disambiguation

Maithrreye Srinivasan
University of Alberta
Edmonton, Alberta, Canada
maithrre@ualberta.ca

Davood Rafiei
University of Alberta
Edmonton, Alberta, Canada
drafiei@ualberta.ca

ABSTRACT

Named Entity Disambiguation (NED) and linking has been traditionally evaluated on natural language content that is both well-written and contextually rich. However, many NED approaches display poor performance on text sources that are short and noisy. In this paper, we study the problem of entity disambiguation for short text and propose a location-aware NED framework that resolves ambiguities in text with little other contextual cues. We show that the spatial dimension is crucial in disambiguating named entities and that the location inference is less utilized in many NED systems. Our proposed framework integrates (in an unsupervised manner) spatial signals that are readily available for many sources that emit short text (e.g., micro-blogs, search queries, and news streams). Our evaluation on news headlines and tweets reveals that a simple spatial embedding improves the accuracy of competitive baseline NED approaches from the literature by 8% for the news headlines and by 4% on tweets.

CCS CONCEPTS

• **Information systems → Information retrieval**;

KEYWORDS

Entity Disambiguation, Spatial Information, Location Inference

ACM Reference Format:
Maithrreye Srinivasan and Davood Rafiei. 2021. Location-Aware Named Entity Disambiguation. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21), November 1–5, 2021, Virtual Event, QLD, Australia.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3459637.3482135

1 INTRODUCTION

Names mentioned in documents and articles are often ambiguous, for example, referring to more than one candidate entity in a knowledge base. Resolving these ambiguities, referred to as Named Entity Disambiguation (NED), is an essential component of many automatic language understanding tools including semantic search [4, 46], knowledge base and knowledge graph population [27, 37], question answering [44] and chatbots [43]. However, resolving name ambiguities has always been a challenge with name variations and aliases, abbreviations, spelling errors, Polysemy (e.g., *John Smith may refer to American Actor or New Zealand Cricketer*), Metonymy (e.g., *The White House can refer to American Administration*), etc.

Many NED systems resolve the ambiguities of mentions using various local and global features. *Local features* often include the string similarity between a mention and a candidate entity, entity type (e.g. person, organization), prior probability, etc., and they fail to distinguish between mentions with the same surface text. For example, one cannot detect if the mentions of *Malone*, a last name, refers to *Karl Malone* or *Jeff Malone*, both professional basketball players. NED systems overcome these problems using *global features* based on the assumption that mentions from the same document are expected to be semantically related and coherent around the topic of the document [18, 20, 38]. However, sparse text lacks contextual information for these global features and the prior probability does not always give a correct mapping.

As an example, consider the following two short texts; *T1* is a headline from the NYT corpus [42] and *T2* is a text from a tweet collection [11].

```
(T1) Smith back in action after recovery
(T2) He's a Spurs lad, and we can't blame him for this season...
```

Passing these short texts to different NED systems results in different mappings of the named entity mentions, probably due to different prior probabilities of each system. For example, in text *T1*, "Smith" is mapped to *Tommy Smith*, a New Zealand Footballer born in 1990, by one system [26], *Agent Smith*, a fictional character from the movie The Matrix, by another system [15], *Adam Smith*, a Scottish Economist, by a third system [34], and *Will Smith*, an American actor, by a fourth system [9]. Similarly, in text *T2*, "Spurs" is incorrectly mapped to *San Antonio Spurs* by multiple systems [9, 15, 26, 34]. However, knowing that headline T1 is originated or published in Sydney, Australia, one may say with some confidence that the mention of "Smith" refers to *Steve Smith*, an Australian cricketer. Similarly, if we take into consideration the fact that tweet T2 is posted by someone from London, England, we may link "Spurs" to *Tottenham Hot Spur F.C.* and not *San Antonio Spur*.

State of the art NED systems (e.g. [18, 26, 32]) mainly target long text, and many of these systems do not perform well in disambiguating entities in short text due to the lack of context. Our goal in this paper is to explore additional cues that are readily available for short text in the form of metadata. In particular, tweets and news headlines possess rich metadata information, notably temporal and spatial data, often recorded and emitted by capturing devices (e.g., mobile phones and GPS-enabled cameras). On the other hand, the spatial signal is a crucial property not just of physical entities such as countries mountains, and rivers but also of persons, organization, artifacts, and events. such as sports leagues, battles, etc. For example, *Manchester United Football Club* is based in *Greater Manchester, England* and *Mona Lisa* is in *Louvre, Paris*. These spatial

CIKM '21, November 1–5, 2021, Virtual Event, QLD, Australia

ACM ISBN 978-1-4503-8446-9/21/11...$15.00
https://doi.org/10.1145/3459637.3482135

signals for candidate entities may be collected from relations such as *happendIn, isLocatedIn, bornIn, diedIn* in a knowledge base or a knowledge graph. Not all entities are popular around the globe. For example, we can see that *San Antonio Spur* is popular in the US and *Tottenham Hot Spur F.C.* is popular in the UK, and that knowing the location of an ambiguous mention can help the NED process. Hence, we study the problem of how spatial information can be modeled and used for entity disambiguation. Our approach is to compute location signatures for entities from different sources where entities are mentioned (e.g. tweets) or discussed (e.g. Wikipedia pages) and use the spatial dimension for linking the mentions to candidate entities.

Our contribution can be summarized as follows:

- We propose a framework for integrating spatial signals in disambiguating named entities in short text. To the best of our knowledge, our work is the first studying the problem in the context of short text.
- As most of the entities have a relationship with multiple locations beyond their primary home location, we develop an algorithm to construct location signatures for entities and context mentions that reflect the importance of different locations associated with the entities.
- We evaluate our model on two data sets - headlines from the New York Times archives and Tweet with rich geographical information. Our work significantly improves, in terms of accuracy and the F1 measure, several baselines from the literature, including some considered the state of the art.

The rest of the paper is organized as follows: the related work is reviewed next, followed by our framework in Section 3. Our spatial signatures are discussed in Section 4 and our evaluation is presented in Section 5. Section 6 concludes the paper.

2 RELATED WORKS

Named entity disambiguation has a long history with some early work on record linkage where the task is to find out if two records in a database represent the same entity [13]. With the advent of Wikipedia, NED systems leverage this resource to detect ambiguous mentions and to link them to entities in Wikipedia. For example, mentions may be linked to candidate entities in Wikipedia based on the overlap between the context of each mention and that of a candidate Wikipedia page [5, 8] and/or using the Wikipedia hyperlink structure to measure the semantic relatedness between candidate entities [32]. In general, NED can be considered as a ranking model where each mention has a set of candidate entities and the model has to choose for each mention a unique entity with the highest confidence score. This confidence score may be estimated based on both local and global features using probabilistic graph models [20, 24, 26, 38, 47], learning to rank techniques [51], and neural networks [9, 21, 25]. The local features may include lexical features including string similarity [51],bag-of-words [2, 45],keyphrases [26], n-grams [19], word embeddings [9], and statistical features such as the prior probability [10, 52] and entity type [35]. The global features find coherence between the mentions of entities in a document[7, 30, 32, 40].

We are not aware of more than a few works that use spatio-temporal features for named entity disambiguation. In particular, Fang and Chang [11] use spatio-temporal signals in a weakly supervised fashion for linking entities mentioned in tweets. They use a binning method to divide both time and space into discrete equal-sized bins. Agarwal et al. [1] use the publication year of documents to build a temporal vector for each entity and show that using the temporal feature helps the disambiguation process.

There are also a handful of NED works that are applicable to short text. TagMe [15] provides on-the-fly annotation for short text, where it compares mentions to anchor texts from Wikipedia and treats pages under the same anchor text as their possible senses. The authors compute a scoring function that aims at a collective agreement among the mappings. Babelfy [34] combines word sense disambiguation and entity linking by running a random walk algorithm. KEA [49], a NED model for tweets, uses features such as surface text similarity, DBpedia types, and a co-occurrence analysis of mentions within Wikipedia articles for disambiguation. WAT [16], an extension of TagMe, uses a vote-based approach for local and a graph-based for global disambiguation. These works on short text are not using the spatial dimension hence they are orthogonal to our work.

3 LOCATION AWARE NED FRAMEWORK

Given an input text with a set of mentions $\{m_1, m_2, \ldots, m_n\}$ and a candidate set of entities from a knowledge base, we want to map each mention to either a unique entity in our candidate set or $\emptyset$ for out-of-KB entities.

Objective function Our framework combines prior probability, context-similarity and spatial similarity into an objective function, and we want to find an assignment of candidates to mentions that maximizes this objective function. Let *cad(m_i)* denote the set of candidates of mention m_i. Our objective function can be written as

$$\underset{e_{j,i} \in cad(m_i)}{\text{argmax}} \Bigg(\alpha . \sum_{i=1}^{n} prior(m_i, e_{j,i}) + \beta . \sum_{i=1}^{n} cxtSim(m_i, e_{j,i}) + \gamma . \sum_{i=1}^{n} locSim(m_i, e_{j,i}) \Bigg) \tag{1}$$

where $\alpha+\beta+\gamma = 1$, and $prior()$, $cxtSim()$ and $locSim()$ respectively denote the prior probability, the context similarity and the spatial similarity of a mention and a candidate entity.

Tagging mentions Examples of input text are a tweet, a news headline and a search query. The input text can be tagged and the mentions of named entities can be detected using standard tools (e.g. the Stanford NER tagger[17]). We assume the input has some location cues in the form of either mentions of locations or some geo-coordinates, for our approach to be applicable.

Finding candidates Public knowledge bases (e.g. Yago[41] and DBpedia[3]) provide a good collection of candidates, where each entity has a short name and a set of paraphrases, constructed from Wikipedia disambiguation pages, redirects, and links. For news headlines, the names are expected to be accurate and one may select as candidates all entities where either the name or a paraphrase matches the mention fully. A full match will not work for tweets and search queries, which generally have misspellings, abbreviations, and unreliable capitalization. For such input, candidates may be selected using a k-gram matching approach [29].

Prior popularity The prior probability of an entity is a context-independent feature and denotes the probability that a given mention m refers to a candidate entity e. It is often estimated from Wikipedia as the fraction of anchor texts that contain m and link to the entity page of e. For example, "Obama" refers to "Barack Obama" in 60.5% of the occurrences and to "Obama, Fukui (location)" in 2.4% of the cases.

Context similarity The context similarity of a mention and a candidate entity is studied in the literature. We use the keyphrase-based similarity of Hoffart et al. [26], which measures the mutual information between the keyphrases of an entity and the words that appear within the context window of a mention. The keyphrases of an entity are extracted from the Wikipedia article of the entity, the anchor texts of links to the article, category names, etc. The context window of a mention is expected to have some of these keyphrases. For instance, the keyphrases of "the Beatles" include *rock band, best-selling band, music of Liverpool* and other phrases, and these phrases are expected to appear in the context window of the mentions of the entity.

Spatial similarity With the spatial features expressed as vectors, the similarity between the feature vector of an entity and that of a mention may be measured using Cosine or the inner product. We discuss the spatial features in more detail in the next section.

4 SPATIAL SIGNALS AS FEATURES

There are some challenges in using spatial information as features for named entity disambiguation. First, locations references can be ambiguous. For example, the term *Paris* by itself refers to 34 different places around the world (e.g., Paris, France; Paris, Ontario; Paris, Arkansas). Linking location references to a geographic database (e.g. Geonames [1]) is studied on its own in the literature (e.g. [28]) and is outside the scope of this paper. A simple strategy is to resolve location ambiguities based on population. A major challenge in our case is that the locations of entities and mentions may not match but can be related under spatial relationships such as containment, neighbourhood, etc. This is addressed next.

Spatial signature of entities The spatial signature of entities can be constructed from their Wikipedia pages since relevant locations are often mentioned there. Our experiment with the NER in spaCy [2] shows that the tool is very effective in tagging locations, reaching a precision of 0.96 or higher on both New York Times corpus and Wikipedia pages. Let the initial location signature of an entity e, denoted by l_e^c, include all locations mentioned in the entity page of e and their frequencies. One may do a spatial smoothing to account for locations that are relevant but not listed in the Wikipedia page of an entity. This is useful in news-style corpus and micro-posts where neighborhood events are reported. For example, although the headline "`Teen violent assault at a Catholic high school south of Edmonton`" is reported in *Edmonton*, the actual event takes place in *Leduc*, a neighborhood of Edmonton. An exponential smoothing is desirable to assign higher weights to nearby locations.

[1]www.geonames.org
[2]https://spacy.io/

Let l_e^s include all locations in l_e^c with their counts and all other locations that are relevant but not in l_e^c with counts set to zero. The relevant locations of l may include its siblings and ancestors in a location hierarchy. For smoothing the counts, one may select a few different geographical ordering of locations (e.g., east to west, west to east, etc) and propagate the weights in the ordering direction. Suppose $l_e^s = \{(l_1, c_1), (l_2, c_2), \ldots, (l_n, c_n)\}$ is one such ordering. Then updating c_i with $\delta c_i + (1-\delta)c_{i-1}$ for $\delta \in [0,1]$ and $i = 2, \ldots, n$ will give a smoothed signature. In our experiments, we do this smoothing for siblings under two orderings (east to west and west to east) and δ set to 0.6 based on cross-validation.

A mismatch between locations can also happen if they are reported at different levels of dispersion. For example, the headline "`Floyd was killed in police custody`" will be reported under the location *United States* in global news, whereas it is reported under *Minneapolis* in the local news. To handle inference between different levels of location dispersion, we may transfer weights from the lower level of dispersion to upper levels while constructing the signature. This can be done in our smoothing by updating the weight of a parent l_i based on the sum of the weight of its children in the signature vector, i.e.

$$c_i = \delta c_i + (1-\delta) \sum_{l_j \in children(l_i)} c_j$$

where $\delta \in [0,1]$, set to 0.5 based on cross-validation.

Spatial signature of mentions A location signature can be constructed for mentions based on the location cues in the surrounding text as well as in the metadata description. Under the inheritance hypothesis [39], named entities inherit the location of an article (e.g. the location where a headline is published or a tweet is posted). The location signature can also be smoothed following the same algorithm discussed for entity signatures.

5 EXPERIMENTAL EVALUATION

This section presents an experimental evaluation of our algorithm. We review our dataset and preprocessing, describe our experimental setup and present an evaluation of our algorithm, in comparison with different baselines from the literature, on two datasets: a NYT-headlines dataset and a geotagged Tweets dataset.

5.1 Datasets

The standard dataset for NED evaluation (e.g., MSNBC news-wire articles [8], CoNLL-YAGO [26], TAC KBP [27], and ACE 2004 NED [40], AQUAINT [32]) are high-quality news articles with most entities mentioned at least once by their full names. To gauge the effectiveness of our spatial signatures, we evaluated our named entity disambiguation on short text, with both formal and informal structure. We used a subset of the New York Times archive (originally containing 1.8 million articles published between 1987 to 2007) [42]), extracting only headlines that have no more than two mentions per headline. This resulted in 2340 headlines. To validate our model in informal text, we used tweets that had location context or were geotagged with latitude and longitude coordinates. The Locke collection [31], Habib collection [23] and Micro post-collection [6] are all re-annotated by Habib et al. [22] with a total of 5535 named entities in tweets linked to Wikipedia. Less than 2% of those tweets had any location context in the tweet for us to use in our evaluation. Hence, we used a subset of a tweet dataset from Farazi and Rafiei [12] that originally contained 53 Million geo-tagged tweets, mostly from the US and Canada. We randomly took only 3490 tweets that had at least one tagged named entity mention for pre-processing and added 314 tweets from a dataset by Fang and Chang [11]. We cleaned the tweets by removing hashtag symbols (#), retweets (RT), @ symbols, and did a text segmentation using the ekphrasis[1] library. We used the Stanford NER with a precision around 0.67 on our tweets dataset, after evaluating other NER Tools including spaCy, NLTK[2], and TwitterNER [33].

We annotated 1762 entities from the NYT subset and 1015 entities from the Tweet subset, with corresponding entities in the Wikipedia 2020 dump. We removed NER errors, out of KB entities and entities with full names in input text. Priors and Keyphrases in AIDA were collected from Wikipedia 2014 and our spatial signatures were constructed from Wikipedia 2020. To avoid any discrepancy between different Wikipedia versions as a source for our features, we annotated only those entities that existed in both years. Our dataset and spatial signatures are publicly available[1].

5.2 Evaluation result

We used standard accuracy, precision, recall, and the F1 measure as evaluation metrics aggregated across mentions (micro-averaged). We evaluated our model against the feature set of AIDA, as we added the spatial dimension to the feature set and retrained the model to get new weights for the features. The weights for prior, context and spatial feature were $\alpha = 0.21$, $\beta = 0.33$, $\gamma = 0.46$ for the NYT headlines and $\alpha = 0.5$, $\beta = 0.2$, $\gamma = 0.3$ for the tweet subset based on cross validation. We chose the α, β and γ value that gave top accuracy on different trials. This feature weighting highlights the importance of each feature class for the dataset. The tweet subset had more prominent entities hence larger prior weights, whereas the NYT subset had more varied locations around the globe and a larger weight to spatial features. From Table 1, we see a significant rise in the accuracy when combining the prior and the context feature with a spatial dimension for both datasets.

Table 2 compares the micro F1 of our model with different NED systems available via GERBIL [48] and to our re-implemented diaNED [1]. Many of the tweets in our tweets dataset did not have a temporal metadata, hence we only evaluated diaNED on the NYT subset using the year of publication as temporal context. Though we could not directly compare our model with the systems from GERBIL as they differ in the candidate generation and entity selection process, cynically, the table shows that our model enhances the NED quality by adding the spatial signatures. Table 2 shows that incorporating spatial signature as a local feature gives a performance comparable to that of models with global features.

To further study the importance of spatial signals, we probed the results of our model in the NYT-subset to see how location-awareness helps improve disambiguation quality based on different entity types. From Table 3, we can see that the location feature helps entity types person and organization with the spatial signals bounding these entities, but only a very moderate improvement is observed for the location type.

Feature set	NYT subset	Tweet subset
prior	45.7	33.2
+spatial	59.1	36.4
prior+context	65.4	40.4
+spatial	**73.8**	**44.9**

Table 1: Micro-accuracy of our model with and without spatial feature

NED system	NYT subset	Tweet subset
xLisa [50]	54.1	35.8
AGDISTIS [47]	58.4	33.7
WAT [16]	52.9	33.2
TagMe 2 [15]	60.2	43.7
Babelfy [34]	63.1	41.2
KEA [49]	57.9	40.8
AIDA [26]	65.7	42.6
WNED [20]	72.1	41.4
reimpl. diaNED-1[1]	68.9	-
our model	**74.5**	**45.6**

Table 2: Micro-F1 of various NED systems on NYT-subset and Tweet-subset

Entity type	w location	w/o location
PER	40.6	27.9
ORG	36.9	18.3
LOC	15.2	14.7

Table 3: Micro-accuracy of our model on NYT subset with and without spatial awareness.

Here are two examples from our tweet subset that shows how the location signal helps linking to correct entities in Wikipedia.

```
(T1) Bro was my favorite player after Kobe..(USA)
(T2) Drake My Go To For Everyyy Mood (Texas, USA)
```

In *T1*, for the mention *Kobe*, the cue *USA* helps to differentiate between the candidate *Kobe, a city in Japan*, and the gold candidate *Kobe Bryant*. Similarly, In *T2*, the location cue *Texas, USA* helps to differentiate between Ervin Drake, American song writer, and Drake (Rapper); the latter has a stronger tie to Texas and is the correct entity in this case.

6 CONCLUSIONS

This paper proposes a NED approach that explicitly considers location signals to aid in the disambiguation process. Our evaluation results show that location-awareness improves the NED quality when the entities in the text hold some regional attachments. As a possible future direction, our work may be extended for a more robust resolution of entities with global dispersion (e.g., Justin Bieber, who tours around the globe). Also our spatial features may be studied in the context of neural models for short text [14, 36].

ACKNOWLEDGMENTS

This research was partially supported by the Natural Sciences and Engineering Research Council and by a grant from Huawei.

[1] https://github.com/cbaziotis/ekphrasis
[2] www.nltk.org
[1] https://github.com/maithrreyes/LocationAwareNED

REFERENCES

- [1] Prabal Agarwal, Jannik Strötgen, Luciano Del Corro, Johannes Hoffart, and Gerhard Weikum. 2018. DiaNED: Time-aware named entity disambiguation for diachronic corpora. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 686–693.
- [2] Amit Bagga and Breck Baldwin. 1998. Entity-Based Cross-Document Core f erencing Using the Vector Space Model. In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*. 79–85. https://www.aclweb.org/anthology/P98-1012
- [3] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. 2009. Dbpedia-a crystallization point for the web of data. *Journal of web semantics* 7, 3 (2009), 154–165.
- [4] Roi Blanco, Giuseppe Ottaviano, and Edgar Meij. 2015. Fast and space-efficient entity linking for queries. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. 179–188.
- [5] Razvan Bunescu and Marius Pasca. 2006. Using encyclopedic knowledge for named entity disambiguation. (2006).
- [6] Amparo E Cano, Giuseppe Rizzo, Andrea Varga, Matthew Rowe, Milan Stankovic, and Aba-Sah Dadzie. 2014. Making sense of microposts:(# microposts2014) named entity extraction & linking challenge. In *Ceur workshop proceedings*, Vol. 1141. 54–60.
- [7] Xiao Cheng and Dan Roth. 2013. Relational Inference for Wikification. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. 1787–1796. http://aclweb.org/anthology/D/D13/D13-1184.pdf
- [8] Silviu Cucerzan. 2007. Large-scale named entity disambiguation based on Wikipedia data. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*. 708–716.
- [9] Yotam Eshel, Noam Cohen, Kira Radinsky, Shaul Markovitch, Ikuya Yamada, and Omer Levy. 2017. Named Entity Disambiguation for Noisy Text. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*. 58–68.
- [10] Anthony Fader, Stephen Soderland, Oren Etzioni, and Turing Center. 2009. Scaling wikipedia-based named entity disambiguation to arbitrary web text. In *Proceedings of the IJCAI Workshop on User-contributed Knowledge and Artificial Intelligence: An Evolving Synergy, Pasadena, CA, USA*. 21–26.
- [11] Yuan Fang and Ming-Wei Chang. 2014. Entity linking on microblogs with spatial and temporal signals. *Transactions of the Association for Computational Linguistics* 2 (2014), 259–272.
- [12] Sara Farazi and Davood Rafiei. 2019. Top-K frequent term queries on streaming data. In *2019 IEEE 35th International Conference on Data Engineering (ICDE)*. IEEE, 1582–1585.
- [13] Ivan P Fellegi and Alan B Sunter. 1969. A theory for record linkage. *J. Amer. Statist. Assoc.* 64, 328 (1969), 1183–1210.
- [14] Zhifan Feng, Qi Wang, Wenbin Jiang, Yajuan Lyu, and Yong Zhu. 2020. Knowledge-Enhanced Named Entity Disambiguation for Short Text. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*. 735–744.
- [15] Paolo Ferragina and Ugo Scaiella. 2010. Tagme: on-the-fly annotation of short text fragments (by wikipedia entities). In *Proceedings of the 19th ACM international conference on Information and knowledge management*. 1625–1628.
- [16] P Ferragina and U Scaiella. 2012. Fast and Accurate Annotation of Short Texts with Wikipedia Pages. *IEEE Annals of the History of Computing* 29, 01 (2012), 70–75.
- [17] Jenny Rose Finkel, Trond Grenager, and Christopher D Manning. 2005. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*. 363–370.
- [18] Amir Globerson, Nevena Lazic, Soumen Chakrabarti, Amarnag Subramanya, Michael Ringaard, and Fernando Pereira. 2016. Collective entity resolution with multi-focal attention. (2016). https://doi.org/10.18653/v1/P16-1059
- [19] Stephen Guo, Ming-Wei Chang, and Emre Kiciman. 2013. To link or not to link? a study on end-to-end tweet entity linking. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 1020–1030.
- [20] Zhaochen Guo and Denilson Barbosa. 2014. Robust entity linking via random walks. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. 499–508.
- [21] Nitish Gupta, Sameer Singh, and Dan Roth. 2017. Entity linking via joint encoding of types, descriptions, and context. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2681–2690.
- [22] Mena Habib and Maurice van Keulen. 2015. Need4tweet: a twitterbot for tweets named entity extraction and disambiguation. In *Proceedings of ACL-IJCNLP 2015 System Demonstrations*. 31–36.
- [23] Mena B Habib and Maurice Van Keulen. 2012. Unsupervised Improvement of Named Entity Extraction in Short Informal Context Using Disambiguation Clues.. In *SWAIE*. 1–10.
- [24] Xianpei Han, Le Sun, and Jun Zhao. 2011. Collective entity linking in web text: a graph-based method. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. 765–774. https://doi.org/10.1145/2009916.2010019
- [25] Zhengyan He, Shujie Liu, Mu Li, Ming Zhou, Longkai Zhang, and Houfeng Wang. 2013. Learning entity representation for entity disambiguation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 30–34.
- [26] Johannes Hoffart, Mohamed Amir Yosef, Ilaria Bordino, Hagen Fürstenau, Manfred Pinkal, Marc Spaniol, Bilyana Taneva, Stefan Thater, and Gerhard Weikum. 2011. Robust disambiguation of named entities in text. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. 782–792. https://doi.org/anthology/D11-1072
- [27] Heng Ji, Ralph Grishman, Hoa Trang Dang, Kira Griffitt, and Joe Ellis. 2010. Overview of the TAC 2010 knowledge base population track. In *Third text analysis conference (TAC 2010)*, Vol. 3. 3–3.
- [28] Ehsan Kamalloo and Davood Rafiei. 2018. A coherent unsupervised model for toponym resolution. In *Proceedings of the 2018 World Wide Web Conference*. 1287–1296.
- [29] Grzegorz Kondrak. 2005. N-gram similarity and distance. In *International symposium on string processing and information retrieval*. Springer, 115–126.
- [30] Sayali Kulkarni, Amit Singh, Ganesh Ramakrishnan, and Soumen Chakrabarti. 2009. Collective annotation of wikipedia entities in web text. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. 457–466.
- [31] Brian Locke and James Martin. 2009. Named entity recognition: Adapting to microblogging. *Senior Thesis, University of Colorado* (2009).
- [32] David Milne and Ian H Witten. 2008. Learning to link with wikipedia. In *Proceedings of the 17th ACM conference on Information and knowledge management*. 509–518.
- [33] Shubhanshu Mishra and Jana Diesner. 2016. Semi-supervised Named Entity Recognition in noisy-text. In *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*. 203–212.
- [34] Andrea Moro, Alessandro Raganato, and Roberto Navigli. 2014. Entity linking meets word sense disambiguation: a unified approach. *Transactions of the Association for Computational Linguistics* 2 (2014), 231–244.
- [35] Dávid Márk Nemeskey, Gábor András Recski, Attlia Zséder, and Andras Kornai. 2010. Budapestacad at tac 2010. (2010).
- [36] Feng Nie, Shuyan Zhou, Jing Liu, Jinpeng Wang, Chin-Yew Lin, and Rong Pan. 2018. Aggregated semantic matching for short text entity linking. In *Proceedings of the 22nd Conference on Computational Natural Language Learning*. 476–485.
- [37] Heiko Paulheim. 2017. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web* 8, 3 (2017), 489–508.
- [38] Maria Pershina, Yifan He, and Ralph Grishman. 2015. Personalized page rank for named entity disambiguation. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 238–243. https://doi.org/10.3115/v1/N15-1026
- [39] Jiangwei Yu Rafiei and Davood Rafiei. 2016. Geotagging named entities in news and online documents. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. 1321–1330.
- [40] Lev Ratinov, Dan Roth, Doug Downey, and Mike Anderson. 2011. Local and global algorithms for disambiguation to wikipedia. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies*. 1375–1384. ttps://www.aclweb.org/anthology/P11-1138.
- [41] Thomas Rebele, Fabian Suchanek, Johannes Hoffart, Joanna Biega, Erdal Kuzey, and Gerhard Weikum. 2016. YAGO: A multilingual knowledge base from wikipedia, wordnet, and geonames. In *International semantic web conference*. Springer, 177–185.
- [42] Evan Sandhaus. 2008. *The New York Times Annotated Corpus LDC2008T19*. Technical Report. Web Download. Philadelphia: Linguistic Data Consortium,.
- [43] Mingyue Shang, Tong Wang, Mihail Eric, Jiangning Chen, Jiyang Wang, Matthew Welch, Tiantong Deng, Akshay Grewal, Han Wang, Yue Liu, et al. 2021. Entity Resolution in Open-domain Conversations. *NAACL-HLT 2021* (2021), 26.
- [44] Daniil Sorokin and Iryna Gurevych. 2018. Mixing context granularities for improved entity linking on question answering data across entity categories. *arXiv preprint arXiv:1804.08460* (2018).
- [45] Tadej Štajner and Dunja Mladenić. 2009. Entity resolution in texts using statistical learning and ontologies. In *Asian Semantic Web Conference*. Springer, 91–104.
- [46] Chuanqi Tan, Furu Wei, Pengjie Ren, Weifeng Lv, and Ming Zhou. 2017. Entity Linking for Queries by Searching Wikipedia Sentences. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 68–77.
- [47] Ricardo Usbeck, Axel-Cyrille Ngonga Ngomo, Michael Röder, Daniel Gerber, Sandro Athaide Coelho, Sören Auer, and Andreas Both. 2014. AGDISTIS-graph-based disambiguation of named entities using linked data. In *International semantic web conference*. Springer, 457–471. https://doi.org/10.3233/978-1-61499-419-0-1113

[48] Ricardo Usbeck, Michael Röder, Axel-Cyrille Ngonga Ngomo, Ciro Baron, Andreas Both, Martin Brümmer, Diego Ceccarelli, Marco Cornolti, Didier Cherix, Bernd Eickmann, et al. 2015. GERBIL: general entity annotator benchmarking framework. In *Proceedings of the 24th international conference on World Wide Web*. 1133–1143.

[49] Jörg Waitelonis and Harald Sack. 2016. Named Entity Linking in# Tweets with KEA.. In *# Microposts*. 61–63.

[50] Lei Zhang and Achim Rettinger. 2014. X-lisa: Cross-lingual semantic annotation. *Proceedings of the VLDB Endowment* 7, 13 (2014), 1693–1696.

[51] Zhicheng Zheng, Fangtao Li, Minlie Huang, and Xiaoyan Zhu. 2010. Learning to link entities with knowledge base. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. 483–491. http://www.aclweb.org/anthology/N10-1072

[52] Stefan Zwicklbauer, Christin Seifert, and Michael Granitzer. 2016. Robust and collective entity disambiguation through semantic embeddings. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 425–434.