Improving Language-Dependent Named Entity Detection

Gerald Petz, Werner Wetzlinger, Dietmar Nedbal

Improving Language-Dependent Named Entity Detection

Gerald Petz[0000-0002-8012-5369], Werner Wetzlinger[0000-0003-2368-7127], and Dietmar Nedbal[0000-0002-7596-0917]

University of Applied Sciences Upper Austria, Steyr, Austria
{gerald.petz,werner.wetzlinger,dietmar.nedbal}@fh-steyr.at

Abstract. Named Entity Recognition (NER) and Named Entity Linking (NEL) are two research areas that have shown big advancements in recent years. The majority of this research is based on the English language. Hence, some of these improvements are language-dependent and do not necessarily lead to better results when applied to other languages. Therefore, this paper discusses TOMO, an approach to language-aware named entity detection and evaluates it for the German language. This also required the development of a German gold standard dataset, which was based on the English dataset used by the OKE 2016 challenge. An evaluation of the named entity detection task using the web-based platform GERBIL was undertaken and results show that our approach produced higher F1 values than the other annotators did. This indicates that language-dependent features do improve the overall quality of the spotter.

Keywords: Entity Recognition, Entity Detection, Language-dependent, Dataset Development, Gold Standard, NER

1 Introduction

The recognition of named entities is an important starting point for many tasks in the area of natural language processing. Named Entity Recognition (NER) refers to methods that identify names of entities such as people, locations, organizations and products [1, 2]. It is typically broken down into the two subtasks entity detection (or "spotting") and entity classification. In many application scenarios, however, it is not only of interest which types of entities are contained in a text, but also how the entities can be semantically linked to a knowledge base. The task of correctly disambiguating and linking the recognized named entities in a text into a knowledge base with an external definition and description is referred to as Named Entity Linking (NEL) [3]. The overall goal is to make sense of data in the context of an application domain [4].

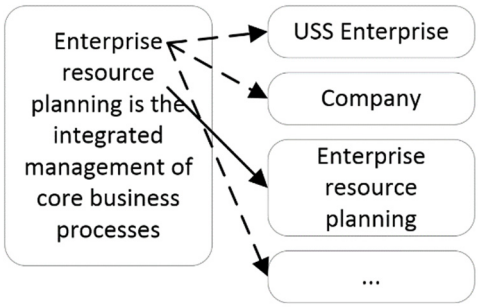

Fig. 1: Illustration for the entity linking task

The whole pipeline (including the aforementioned tasks of NER and NEL) is strongly dependent on the knowledge base used to train the named entity extraction algorithm [5]. Most approaches for linking entities leverage on the use of Wikipedia (wikipedia.org), Dbpedia (dbpedia.org), Freebase (freebase.com) or YAGO (yago-knowledge.org) as the knowledge base. Although widely used and the largest online encyclopedia with millions of articles, Wikipedia may not be sufficient for more specific domains and contexts. For example, in the German Wikipedia, only some large Austrian organizations are represented, names of persons are rare, etc. Moreover, the English Wikipedia does not hold this specific information either.

Among others, Piccinno and Ferragina [6] recognized that recent research tends to focus its attention on the NEL step of the pipeline by trying to improve and implement new disambiguation algorithms. However, ignoring the issues raised by entity recognition leads to the introduction of many false positives, which provoke a significant loss in the overall performance of the system. It would therefore be better to first try to improve the quality of the NER spotter.

Another problem area relates to differences in the language itself. It has been acknowledged that linguistically motivated and thus language aware spotting methods are more accurate than language independent methods [7]. The German language has a lot of differences for example in the use of upper and/or lowercase, compound nouns or hyphens to concatenate nouns. However, improvements in a certain language usually come at the expense of ease of adaptation to new languages. In addition, the established NER/NEL challenges and tasks of the scientific community like the OKE challenges [8], the NEEL challenge series [9], or the ERD challenges [10] are in the English language and therefore language-dependent improvements are often not in the focus of the research.

Moreover, the results from different tools need to be comparable against certain quality measures (cf. Section 4.1) based on the same dataset. Frameworks addressing the continuous evaluation of annotation tools such as GERBIL [11, 12] can be used for comparison, but evaluation datasets provided by GERBIL as “gold standards” are only available for the English language as well.

The objective of this paper therefore is to (i) develop an approach for language-aware spotting and (ii) to evaluate the proposed spotting approach for the German language.

After an analysis of the state of the art in spotting methods in general (Section 2), the paper focuses on possibilities to optimize the spotter for a certain language in Section 3. In Section 4, evaluation measures are discussed, followed by an analysis of available datasets for evaluation. Additionally, we show how a German dataset was developed and used for evaluation purposes. Section 5 presents results of the experiments and final conclusions are drawn in Section 6.

2 State of the Art in Entity Detection (Spotting)

As mentioned above the entity detection ("spotting") is an important task in the area of NEL; a couple of authors emphasize the importance of a correct entity spotting in order to avoid errors in later stages of the entity linking task. [6, 13]

Several approaches to the spotting task can be identified in the literature:

- NER tagger. Some tools and approaches rely on existing implementations of NER taggers such as Standford NER tagger or OpenNLP Named Entity Recognition in order to spot surface forms of entities [14–18]. The Standford NER tagger is an implementation of linear chain Conditional Random Field (CRF) sequence models, the OpenNLP NER is based on a Maximum Entropy model.
- POS tags and rules. A couple of authors use part of speech (POS) taggers and/or several rules in order to identify named entities [19–34]. The rules range from simple rules such as "capitalized letter" (if a word contains a capitalized letter the word will be treated as a spot), stop word lists, "At Least One Noun Selector"-rule to complex, combined rules.
- Dictionary based techniques. The majority of approaches leverage techniques based on dictionaries [6, 19, 31, 35–45]. The structure of Wikipedia provides useful features for generating dictionaries:
  - Entity pages: Each page in Wikipedia contains a title (e.g. "Barack Obama") that is very likely the most common name for an entity.
  - Redirect pages: Wikipedia contains redirect pages for each alternative name of an entity page. E.g. "Obama" is a redirect page to "Barack Obama".
  - Disambiguation pages: Disambiguation pages in Wikipedia are used to resolve conflicts with ambiguous article titles. E.g. "Enterprise" may refer to a company, to aircrafts, to Star Trek, and many more. Disambiguation pages are very useful for extracting aliases and abbreviations.
  - Bold phrases: Bold phrases in the first paragraph of a Wikipedia entry can contain useful information such as abbreviations, aliases or nicknames. E.g. the bold phrase in the page "Barack Obama" contains the full name ("Barack Hussein Obama II").
  - Hyperlinks in Wikipedia pages: Pages in Wikipedia usually contain hyperlinks to other pages; the anchor texts of these hyperlinks may provide synonyms and other name variations.
- Methods based on search engines. Some authors try to use web search engines such as Google to identify candidate entities [46–49].

- Computational techniques. A couple of authors leverage heuristic based methods or machine learning methods. Some approaches expand the surface forms by searching the textual context based on heuristic pattern matching [46, 47, 50]. Other authors use N-Grams [31, 33, 44, 51, 52], others experiment with CRF [25, 26], Topic Modeling [53, 54], Naïve Bayes and Hidden Markov Models [55]. Last but not least one can find approaches based on Finite-state machines [14, 15, 30, 34].

The majority of the papers use dictionary approaches. Nevertheless, the above mentioned approaches are usually combined, e.g. [6] leverages OpenNLP NER, a dictionary approach based on Wikipedia with utilization of several features such as anchor texts, redirect pages, etc.

The authors usually provide measures (recall, precision, F1) of the effectiveness of their approaches; unfortunately, these measures cannot be directly compared because usually different datasets are used and the approaches are optimized towards these datasets.

3 TOMO Approach to Optimize Spotter for the German Language

This section details our approach to optimizing a spotter ("TOMO") for the German language with a focus on the spotting phase within the entity linking pipeline . The base system used is Dexter [36, 37], an open-source framework (available at https://github.com/dexter/dexter) that implements a dictionary spotter using Wikipedia content. Fig. 2 shows the approach comprising the construction and the annotation process using a dictionary.

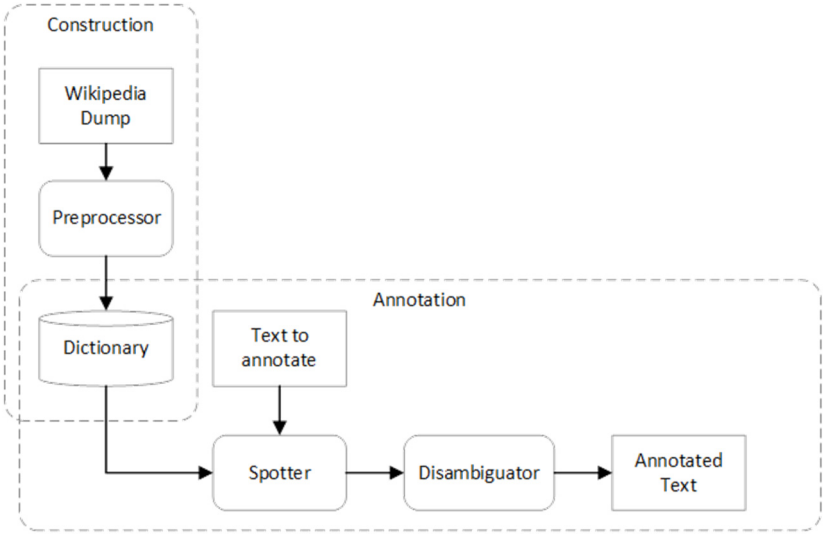

Fig. 2: Basic TOMO Architecture

The annotation process involves a spotter and a disambiguator with an annotated text as output. The spotter detects a list of candidate mentions in the input text, and retrieves for each mention a list of candidate entities [36]. When spotting a text, individual fragments or words from the text ("shingles") are compared with the dictionary of up to six words ("n-grams"). Before being able to use the dictionary for NER, the dictionary needs to be filled with known entities first. Therefore, each Wikipedia article is processed using the title of the article as well as all internal links (anchors within Wikipedia) as spots for the dictionary. In addition, the measures of mention frequency (mf), link frequency (lf) and document frequency (df) are calculated and stored as well. Both in the construction of the dictionary and in the annotation of a text based on this dictionary the text fragments (shingles and known entities) go through a cleaning pipeline with a series of replacements. The cleaning pipeline in pseudocode is as follows:

```
foreach (article in knowledgebase)
  listOfSpots = preprocess(getTitle(article))
  listOfSpots = preprocess(getAnchors(article))
  calculateMeasures(listOfSpots)

preprocess(textfragment)
  clean(textfragment)
  filter(textfragment)
  map(textfragment)
```

A "cleaner" performs a simple transformation of a text fragment (e.g. transform a text to lowercase, remove symbols, remove quotes, unescape Javascript, clean parenthesis, etc.). A "filter" allows the removal of a given text fragment if it does not respect a filter constraint (e.g. delete text fragments that are below the threshold for commonness, have less than three characters, consist only of numbers or symbols, etc.). A "mapper" returns several different versions of the spot (e.g. a "quotes mapper" generates from [dave "baby" cortez] the spots [dave "baby" cortez], [baby], [dave cortez]) [56].

Moreover, for simplification purposes, many tools use lowercase filters. Full text search indices such as Lucene also imply such a lowercase behavior per default, which in many tasks (e.g. search engine querying, microblogging analysis) makes sense. In our setting, lowercase simplification is responsible for introducing several spotting errors (e.g. the sentence "the performance is worse": the word "worse" translates to "schlechter" in German and the spotter identifies this word as a candidate entry for the Wikipedia page "Carl Schlechter"). In German language, only nouns and proper names are written with capitalized initial letters.

Language-aware preprocessing pipeline

In a setting where typing errors are relatively rare (e.g. in press releases, formal documents) the application of a case sensitive setting is therefore a reliable and straightforward approach to increase the precision of the spotter for the German language [57].

Another important aspect of a language-aware approach is the correct usage of the code page. For the English language, the US-ASCII code page is the preferred setting,

as it uses less space than other code pages. In the German language, many named entities contain non US-ASCII characters, like umlaute or the German eszett. Using US-ASCII filters, these characters are replaced by their English representation (umlaut a gets replaced by an "a", etc.). This sometimes changes the whole meaning of the word, as the English replacements are also used in German language and this gets even worse in combination with lowercase filtering. E.g. the sentence "we made this", with its German translation "Wir haben das gemacht": the word "made" translates to "gemacht" in German and this is disambiguated to "Gemächt" (the male genitalia). The UTF-8 code page can be used as a solution to this problem.

Additionally, some minor issues may occur due to differences in the language of the Wikipedia syntax itself. For instance, it is possible to link images within Wikipedia with the common English terms "File:" or "Image:", but the German Wikipedia additionally allows the deprecated terms "Datei:" or "Bild:" as well. Such filters therefore also need to be aware of differences in the German language in order to improve spotting.

4 Evaluation Measures and Datasets

In this section, we discuss the evaluation of spotting named entities in the German language. This includes which measures, tools and datasets to use.

4.1 Measures and Benchmarking

To ensure comparability across different NER and NEL system evaluations the most common measures are *precision*, *recall*, *F1* and *accuracy*.

Precision. Precision considers all spots that are generated by the system and determines how correct they are compared to a gold standard. Consequently, the precision of a spotting system is calculated as the fraction of correctly spotted entity mentions compared to all spotted mentions generated by a particular system.

$$precision = \frac{correctly\ spotted\ mentions}{mentions\ spotted\ by\ the\ system} \quad (1)$$

Recall. Recall is a measure that describes how many of the spots of a gold standard are correctly identified by a system. It is the fraction of correctly spotted entity mentions by a particular systems compared to the all entity mentions that should be spotted according to a selected gold standard.

$$recall = \frac{correctly\ spotted\ mentions}{manually\ annotated\ mentions} \quad (2)$$

F1. To generate a single measure for a system from recall and precision, the measure F1 was developed. It is defined as the harmonic mean of precision and recall as shown in equation 3.

$$F1 = \frac{2*precision*recall}{precision+recall} \quad (3)$$

GERBIL. As also minor differences between these measures exist, we use the web-based benchmarking system GERBIL (gerbil.aksw.org) to evaluate these measures for our system and compare them with others. GERBIL is an entity annotation system that provides a web-based platform for the comparison of annotators [11]. Currently it incorporates 13 annotators and 32 datasets for evaluating the performance of systems. The evaluation is done using uniform measuring approaches and well established measures like the aforementioned recall, precision and F1 [12]. Consequently, GERBIL can be used for benchmarking different annotators. External tools can be added to the GERBIL platform by providing an URL to a REST interface of the tool. Besides the integrated datasets, GERBIL allows for the use of user-specified datasets. As GERBIL is based on Natural Language Programming Interface Format (NIF), user-specified datasets also have to be uploaded using this format. Additionally, GERBIL provides Java classes for implementing APIs for datasets and annotators to NIF. Due to these features, GERBIL is also used by challenges (e.g. OKE challenge) as platform for evaluating the performance of contestants.

4.2 Dataset

To test an entity linking system a gold standard dataset must be provided. This dataset has to include all sentences to analyze, spots to be linked and links to a knowledge base for correct disambiguation. Systems are then ranked by comparing the above-mentioned evaluation measures (recall, precision and F1) they score in relation to this dataset. There are already a number of English corpora to test entity recognition and entity linking systems. Some of them emerged from challenges that compare the results of multiple algorithms and systems to assess the performance of different approaches. For example the datasets of the OKE challenge as part of the European Semantic Web Conferences 2016 (2016.eswc-conferences.org), the NEEL challenge of the Microposts Workshop at the World Wide Web Conference 2016 (microposts2016.seas.upenn.edu), or the ERD challenge at the SIGIR 2014 Workshop ERD'14 (sigir.org/sigir2014) are publicly available.

Requirements and Review of Existing Datasets. To test the performance of our approach to spot entities in the German language we had to select or develop a dataset. For this task, we defined the following requirements:

- The dataset has to be available in German to test the performance of the spotter for German texts.

- The dataset should be testable via the GERBIL web service. Thus, it should be already available in GERBIL or encoded in NIF format.
- The dataset should be widely used, specifically by new systems, to be able to compare our results with leading systems and approaches.
- The dataset should be independent of a certain domain (e.g. only articles about economics).
- The content of the dataset should be comprised of natural language in encyclopedia entries or news. Specific content like tweets or queries were not of interest, since these datasets often have just very few spots with average entries per documents lower than 2.0.
- The dataset should include co-references to evaluate the performance improvements of future enhancements of our system.

We examined existing datasets and their suitability for our requirements. Table 1 shows the results of this literature review which showed that nearly all available datasets are for the English language.

Table 1. Comparison of gold standard datasets

Datasets	Type	Lang	Co-Refs	Docs	Linked Entities	Avg. Entity/ Doc
ACE2004	news	en		57	257	4.44
AIDA-Yago2/CoNLL	news	en		231	4485	19.97
AQUAINT	news	en		50	727	14.54
Dbpedia Spotlight	news	en		58	330	5.69
Derczynski - Broad Twitter Corpus	tweets	en		9551	12117	1.27
ERD2014	queries	en		91	59	0.65
GERDAQ	queries	en		992	1706	1.72
IITB	webpages	en		103	11249	109.22
KORE50	news	en		50	144	2.86
MSNBC	news	en		20	658	32.50
Microposts2014 test dataset	tweets	en		1165	1458	1.25
Microposts2015 test dataset	tweets	en		2027	2382	1.18
Microposts2016 test dataset	tweets	en		3164	738	2.46
N3-RSS-500	news	en		500	1000	1.00
N3-Reuters-128	news	en		128	880	4.85
OKE 2015 Task 1 evaluation dataset	encyclopedia	en	x	101	664	6.57
OKE 2015 Task 2 evaluation dataset	encyclopedia	en	x	99	99	1.00
OKE 2016 Task 1 evaluation dataset	encyclopedia	en	x	55	340	6.18
OKE 2016 Task 2 evaluation dataset	encyclopedia	en		50	50	1.00
N3-News-100	news	de		100	1547	15.47
Meij	tweets	en		502	812	1.87
LinkedTV	news	de		150	1346	8.97
GerNED	news	de		2468	1664	0.67
Ritter	tweets	en		2394	1672	0.70

The three German datasets found were not appropriate for our requirements because they were too domain specific (News-100, LinkedTV), possess only the "classic" named entities (persons, locations, etc.), had no co-references defined, and/or are not publicly accessible (GerNED). Since none of these datasets fitted our requirements, we decided to develop a new dataset to evaluate the spotter against German texts.

Development of a German gold standard dataset. We chose to develop a new German dataset based on the evaluation dataset of the OKE challenge 2016 ("OKE 2016 Task 1 evaluation dataset") for several reasons. Since the content of this dataset originated from Wikipedia articles, it covers a wide range of topics. Therefore, it also consists of natural language and not of tweets or search queries. Furthermore, documents are long enough to contain multiple spots (6.18 average entities per document) and they include co-references as well. Additionally, the English version of the dataset is coded in NIF format and is already integrated in GERBIL. Finally, with 55 documents and 340 entities, we considered this dataset to be of an appropriate length.

To develop the new dataset based on the dataset, we conducted a multi-step approach that consisted of the following tasks:

1. Identify all documents and included spots in the NIF file of the OKE 2016 Task 1 evaluation dataset.
2. Translate all documents in this dataset using Google Translate (translate.google.com)
3. Adjust the initial Google translation by improving German grammar, word order, etc. by native speakers.
4. Identify all English spots of dataset in the German translation.
5. Identify the corresponding entities in the German knowledge base (de.wikipedia.org).
6. Link the spots to the identified knowledge base entities using links in a HTML file.
7. Transform the HTML file to NIF using a converter.

This process was not straightforward and a number of problems occurred that were mainly based on ambiguities in steps 5 and 6:

- Because the English Wikipedia is more than twice as large as the German Wikipedia, some spots had no representation in the German knowledge base. This was mainly the case with persons (e.g. Andrew McCollum, James Alexander Hendler) and organizations (e.g. Kirkcaldy High School, American Association for Artificial Intelligence).
- Literal translation by Google Translate led to surface forms that were wrong or unusual (e.g. "artificial intelligence researcher" was translated to "künstlicher Intelligenzforscher").
- Translation by Google Translate led to a sentence structure and grammar that was sometimes unusual or incorrect for German sentences.
- In some cases, it was not clear which term was the correct German translation of the English term in the specific context of sentences (e.g. "independent contractor" was

translated to “unabhängiger Auftragnehmer” by Google Translate, but “freier Mitarbeiter” was considered to be the correct translation for the context of the sentence).
- In a few cases, it was not clear to which entity in the German knowledge base an entity should be linked (e.g. the English term “treasury” can be translated based on traditional British or American interpretations of the word as “Finanzministerium” or “Schatzamt”, but is now also used in its English form as a department of corporations.)

In order to cope with these uncertainties, three researchers (German native speakers) independently identified the corresponding German surface form for the spot based on the translated text. For every spot that led to different surface forms or links the different solutions from the three authors were discussed and a majority decision was made by voting.

As a result, some spots were not available in the German knowledge base and therefore the resulting dataset has fewer spots than the original. Since not all systems currently support co-references, we developed two versions of the dataset. One with co-references and one without co-references (15 documents incorporated a total of 24 co-references). The resulting corpus can be downloaded here: *https://github.com/HCSolutionsGesmbH/OKE-Challenge-German*.

5 Experiments and Results

Based on the discussion described in Section 4 we built different test cases to evaluate the changes between a language independent (n/a) and an explicit German language (de) setting. In addition, we considered case sensitivity as a test case for our experiments and built a model based on a case sensitive (cs) and case insensitive (cis) setting. These model characteristics led to the four different test cases shown in Fig. 3.

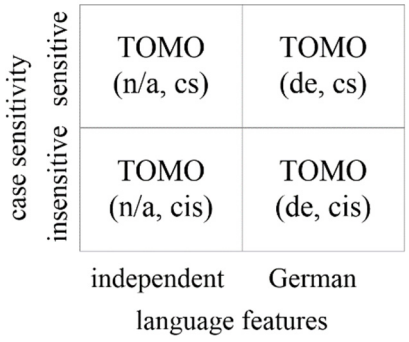

Fig. 3: Test cases

Using the German case sensitive model, we experimented with the commonness threshold. We cut out 0%, 5%, 10%, 15% and 20% of spots with the lowest commonness and evaluated the resulting F1 scores. Results showed that using all spots (i.e. not cutting

out any) led to the highest F1 score. Thus, this setting resulted in the highest recall without lowering the precision too much.

Consequently, we tested all four test cases using this setting in GERBIL with our developed dataset that is based on the recent German Wikipedia dump from 2017/05/01. Table 2 shows the resulting scores for recall, precision and F1.

Table 2. Evaluation of spotting results

Annotator	Recall	Precision	F1
TOMO (n/a, cis)	0.8447	0.4575	0.5684
TOMO (n/a, cs)	0.8599	0.4707	0.5823
TOMO (de, cis)	0.8447	0.4575	0.5684
TOMO (de, cs)	0.8569	0.4811	**0.5866**
AIDA	0.3109	0.6516	0.3983
Babelfy	0.4336	0.3599	0.3689
DBpedia Spotlight	0.4139	0.5077	0.4197
Dexter	0.3537	0.6176	0.4236
Entityclassifier.eu NER	0.767	0.4888	0.5689
FOX	0.4523	**0.6777**	0.4996
FRED	**0.9519**	0.1878	0.303
FREME NER	0.2353	0.3485	0.2643
Kea	0.541	0.5953	0.5399
WAT	0.495	0.6442	0.5163

The annotators TagMe 2, xLisa-NER and xLisa-NGRAM of GERBIL did not produce any results (the GERBIL experiment reported: “The annotator caused too many single errors.”) and could not be evaluated. FRED produced the highest recall. As FRED aims at producing formal structure graphs from natural language text and is based on a dictionary comprising different knowledge bases including WordNet aims at a high recall. On the other side it automatically translates the input text beforehand which may lead to a decrease in precision [45]. FOX combines the results of several state-of-the-art NER tools by using a decision-tree-based algorithm which performed best on the precision measure [58, 59]. In addition, the tool is capable of automatically detecting German language text input. Results also show that the TOMO approach using the German language setting in combination with the case sensitive model achieved the highest F1 score among all tested annotators.

6 Conclusions and Future Work

The paper discusses an approach for language-aware spotting and evaluates the proposed spotting approach for the German language. The results indicate that language-dependent features do improve the overall quality of the spotter. This is necessary, because errors introduced in the spotting phase have an effect on the disambiguation step and can hardly be corrected. A limitation of this work is that the performance metrics of TOMO vs. other systems are only partially comparable, because the annotators were

either developed only for the English language or do not take into account any language specifics. However, we were able to show that language-dependent features improve spotting quality. With the availability of a dataset in German and English language, it is possible to directly compare the performances of the systems for different languages.

When the authors of this paper developed the German corpus a lot of discussions about which surface forms should be linked to the knowledge base arose. For example this text (taken from OKE 2016 Task 1 evaluation dataset) contains several links (shown as underlined words): "Ray Kurzweil grew up in the New York City borough of Queens. He was born to secular Jewish parents who had emigrated from Austria just before the onset of World War II." It is not quite clear, why "parents" are linked to an entity, but some text fragments that are probably more in need of explanation such as "Jewish" or "World War II" are not spots. Wikipedia contains a separate page that provides guidelines for linking. These guidelines suggest for example, not to link everyday words, but to link to other articles that will help the reader to understand the context more fully [60]. However, every gold standard obviously represents a certain way of thinking. Furthermore, performing well on a certain gold standard just means the system replicates a certain way of thinking very well.

Further research work should include a discussion and development of guidelines or rules which terms should be annotated in a gold standard dataset in order to align the different evaluation datasets. Furthermore, a population of a cross-language and cross-domain gold standard in order to evaluate annotation systems for different purposes would be of value for the community.

Acknowledgements

This research was supported by HC Solutions GesmbH, Linz, Austria. We have to express out appreciation to Florian Wurzer, Reinhard Schwab and Manfred Kain for discussing these topics with us.

The TOMO Named Entity Linking is part of TOMO ® (http://www.tomo-base.at), a big data platform for aggregating content, analyzing and visualizing content.

References

1. Petasis, G., Spiliotopoulos, D., Tsirakis, N., Tsantilas, P.: Large-scale Sentiment Analysis for Reputation Management. In: Gindl, S., Remus, R., Wiegand, M. (eds.) 2nd Workshop on Practice and Theory of Opinion Mining and Sentiment Analysis (2013)
2. Derczynski, L., Maynard, D., Rizzo, G., van Erp, M., Gorrell, G., Troncy, R., Petrak, J., Bontcheva, K.: Analysis of Named Entity Recognition and Linking for Tweets. Preprint submitted to Elsevier (2014)
3. Rizzo, G., van Erp, M., Troncy, R.: Benchmarking the extraction and disambiguation of named entities on the semantic web. In: 9th International Conference on Language Resources and Evaluation (LREC'14), pp. 4593–4600 (2014)

4. Holzinger, A.: Introduction to Machine Learning and Knowledge Extraction (MAKE). Machine Learning and Knowledge Extraction 1, 1–20 (2017)
5. Rizzo, G., Troncy, R., Hellmann, S., Brümmer, M.: NERD meets NIF: Lifting NLP extraction results to the linked data cloud. In: LDOW, 5th Workshop on Linked Data on the Web, April 16, 2012, Lyon, France. Lyon, FRANCE (2012)
6. Piccinno, F., Ferragina, P.: From TagME to WAT: A New Entity Annotator. In: Proceedings of the First International Workshop on Entity Recognition & Disambiguation, pp. 55–62. ACM, New York, NY, USA (2014)
7. Daiber, J., Jakob, M., Hokamp, C., Mendes, P.N.: Improving Efficiency and Accuracy in Multilingual Entity Extraction. In: Proceedings of the 9th International Conference on Semantic Systems, pp. 121–124. ACM, New York, NY, USA (2013)
8. Nuzzolese, A.G., Gentile, A.L., Presutti, V., Gangemi, A., Garigliotti, D., Navigli, R.: Open Knowledge Extraction Challenge. In: Gandon, F., Cabrio, E., Stankovic, M., Zimmermann, A. (eds.) Semantic Web Evaluation Challenges: Second SemWebEval Challenge at ESWC 2015, Portorož, Slovenia, May 31 - June 4, 2015, Revised Selected Papers, pp. 3–15. Springer International Publishing (2015)
9. Rizzo, G., Pereira, B., Varga, A., van Erp, M., Cano Basave, A.E.: Lessons Learnt from the Named Entity rEcognition and Linking (NEEL) Challenge Series. Semantic Web Journal (in press) (2017)
10. Carmel, D., Chang, M.-W., Gabrilovich, E., Hsu, B.-J., Wang, K.: ERD'14: Entity Recognition and Disambiguation Challenge. SIGIR Forum 48, 63–77 (2014)
11. Usbeck, R., Röder, M., Ngonga Ngomo, A.-C.: GERBIL – General Entity Annotator Benchmarking Framework (2015)
12. Röder, M., Usbeck, R., Ngonga Ngomo, A.-C.: GERBIL's New Stunts: Semantic Annotation Benchmarking Improved (2016)
13. Hachey, B., Radford, W., Nothman, J., Honnibal, M., Curran, J.R.: Evaluating Entity Linking with Wikipedia. Artificial Intelligence 194, 130–150 (2013)
14. Mendes, P.N., Jakob, M., Garcia-Silva, A., Bizer, C.: DBpedia Spotlight: Shedding Light on the Web of Documents. In: Proceedings of the 7th International Conference on Semantic Systems, pp. 1–8. ACM, New York, NY, USA (2011)
15. Mendes, P.N., Jakob, M., Bizer, C.: DBpedia: A Multilingual Cross-domain Knowledge Base. In: Proceedings of the International Conference on Language Resources and Evaluation (LREC), pp. 1813–1817 (2012)
16. Rizzo, G., Troncy, R.: NERD: evaluating named entity recognition tools in the web of data. In: ISWC 2011, Workshop on Web Scale Knowledge Extraction (WEKEX'11), October 23-27, 2011, Bonn, Germany. Bonn, Germany (2011)
17. Hoffart, J., Yosef, M.A., Bordino, I., Fürstenau, H., Pinkal, M., Spaniol, M., Taneva, B., Thater, S., Weikum, G.: Robust Disambiguation of Named Entities in Text. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 782–792. Association for Computational Linguistics, Stroudsburg, PA, USA (2011)
18. Charton, E., Gagnon, M., Ozell, B.: Automatic Semantic Web Annotation of Named Entities. In: Butz, C., Lingras, P. (eds.) Advances in Artificial Intelligence, 6657, pp. 74–85. Springer Berlin Heidelberg (2011)
19. Eckhardt, A., Hreško, J., Procházka, J., Smrž, O.: Entity Recognition Based on the Co-occurrence Graph and Entity Probability (2014)

20. Zhao, S., Li, C., Ma, S., Ma, T., Ma, D.: Combining POS Tagging, Lucene Search and Similarity Metrics for Entity Linking. In: Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B. et al. (eds.) Web Information Systems Engineering – WISE 2013, 8180, pp. 503–509. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
21. Zhang, L., Dong, Y., Rettinger, A.: Towards Entity Correctness, Completeness and Emergence for Entity Recognition (2015)
22. Moro, A., Raganato, A., Navigli, R.: Entity Linking meets Word Sense Disambiguation: a Unified Approach. Transactions of the Association for Computational Linguistics 2, 231–244 (2014)
23. Leaman, R., Gonzalez, G.: BANNER: an executable survey of advances in biomedical named entity recognition. In: Pacific Symposium on Biocomputing, 13, pp. 652–663 (2008)
24. Cucerzan, S.: 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), pp. 708–716. Association for Computational Linguistics, Prague, Czech Republic (2007)
25. Dojchinovski, M., Kliegr, T.: Entityclassifier.eu: Real-Time Classification of Entities in Text with Wikipedia. In: Blockeel, H., Kersting, K., Nijssen, S., Železný, F. (eds.) Machine Learning and Knowledge Discovery in Databases, 8190, pp. 654–658. Springer Berlin Heidelberg (2013)
26. Kliegr, T.: Linked hypernyms: Enriching DBpedia with Targeted Hypernym Discovery. Web Semantics: Science, Services and Agents on the World Wide Web, - (2014)
27. Tonelli, S., Giuliano, C., Tymoshenko, K.: Wikipedia-based WSD for multilingual frame annotation. Artificial Intelligence 194, 203–221 (2013)
28. Goudas, T., Louizos, C., Petasis, G., Karkaletsis, V.: Argument Extraction from News, Blogs, and Social Media on AI, SETN 2014, Ioannina, Greece, May 15-17, 2014. Proceedings. In: Likas, A., Kalles, D., Blekas, K. (eds.) Artificial Intelligence: Methods and Applications - 8th Hellenic Conference on AI, SETN 2014, Ioannina, Greece, May 15-17, 2014. Proceedings, pp. 287–299. Springer (2014)
29. Ritter, A., Clark, S., Mausam, Etzioni, O.: Named Entity Recognition in Tweets: An Experimental Study. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 1524–1534. Association for Computational Linguistics, Stroudsburg, PA, USA (2011)
30. Olieman, A., Azarbonyad, H., Dehghani, M., Kamps, J., Marx, M.: Entity linking by focusing DBpedia candidate entities. In: Carmel, D., Chang, M.-W., Gabrilovich, E., Hsu, B.-J.(., Wang, K. (eds.) the first international workshop, pp. 13–24 (2014)
31. Chiu, Y.-P., Shih, Y.-S., Lee, Y.-Y., Shao, C.-C., Cai, M.-L., Wei, S.-L., Chen, H.-H.: NTUNLP approaches to recognizing and disambiguating entities in long and short text at the ERD challenge 2014. In: Carmel, D., Chang, M.-W., Gabrilovich, E., Hsu, B.-J.(., Wang, K. (eds.) the first international workshop, pp. 3–12
32. Barrena, A., Agirre, E., Soroa, A.: UBC entity recognition and disambiguation at ERD 2014. In: Carmel, D., Chang, M.-W., Gabrilovich, E., Hsu, B.-J.(., Wang, K. (eds.) the first international workshop, pp. 79–82 (2014)

33. Noraset, T., Bhagavatula, C., Downey, D.: WebSAIL wikifier at ERD 2014. In: Carmel, D., Chang, M.-W., Gabrilovich, E., Hsu, B.-J.(., Wang, K. (eds.) The first international workshop, pp. 119–124 (2014)
34. Lipczak, M., Koushkestani, A., Milios, E.: Tulip: Lightweight Entity Recognition and Disambiguation Using Wikipedia-Based Topic Centroids. In: Carmel, D., Chang, M.-W., Gabrilovich, E., Hsu, B.-J.(., Wang, K. (eds.) the first international workshop, pp. 31–36 (2014)
35. Petasis, G., Spiliotopoulos, D., Tsirakis, N., Tsantilas, P.: Sentiment Analysis for Reputation Management: Mining the Greek Web. In: Likas, A., Blekas, K., Kalles, D. (eds.) Artificial Intelligence: Methods and Applications, 8445, pp. 327–340. Springer International Publishing (2014)
36. Ceccarelli, D., Lucchese, C., Orlando, S., Perego, R., Trani, S.: Dexter: An Open Source Framework for Entity Linking. In: Proceedings of the Sixth International Workshop on Exploiting Semantic Annotations in Information Retrieval, pp. 17–20. ACM, New York, NY, USA (2013)
37. Ceccarelli, D., Lucchese, C., Orlando, S., Perego, R., Trani, S.: Dexter 2.0 - an Open Source Tool for Semantically Enriching Data. In: Horridge, M., Rospocher, M., van Ossenbruggen, J. (eds.) Proceedings of the ISWC 2014 Posters & Demonstrations Track, pp. 417–420 (2014)
38. Ferragina, P., Scaiella, U.: 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, pp. 1625–1628. ACM, New York, NY, USA (2010)
39. Mihalcea, R., Csomai, A.: Wikify!: Linking Documents to Encyclopedic Knowledge. In: Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management, pp. 233–241. ACM, New York, NY, USA (2007)
40. Ratinov, L., Roth, D., Downey, D., Anderson, M.: Local and Global Algorithms for Disambiguation to Wikipedia. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, pp. 1375–1384. Association for Computational Linguistics, Stroudsburg, PA, USA (2011)
41. Agirre, E., Soroa, A.: Personalizing PageRank for word sense disambiguation. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, pp. 33–41. Association for Computational Linguistics, Athens, Greece (2009)
42. Agirre, E., de Lacalle, O.L., Soroa, A.: Random Walks for Knowledge-Based Word Sense Disambiguation. Computational Linguistics 40, 57–84 (2014)
43. Milne, D., Witten, I.H.: An open-source toolkit for mining Wikipedia. Artificial Intelligence, Wikipedia and Semi-Structured Resources 194, 222–239 (2013)
44. Kemmerer, S., Großmann, B., Müller, C., Adolphs, P., Ehrig, H.: The neofonie NERD system at the ERD challenge 2014. In: Carmel, D., Chang, M.-W., Gabrilovich, E., Hsu, B.-J.(., Wang, K. (eds.) the first international workshop, pp. 83–88 (2014)
45. Gangemi, A., Presutti, V., Reforgiato Recupero, D., Nuzzolese, A.G., Draicchio, F., Mongiovì, M., Alani, H.: Semantic Web machine reading with FRED. SW, 1–21 (2016)
46. Lehmann, J., Monahan, S., Nezda, L., Jung, A., Shi, Y.: LCC Approaches to Knowledge Base Population at TAC 2010. In: TAC 2010 Proceedings Papers (2010)
47. Han, X., Zhao, J.: NLPR_KBP in TAC 2009 KBP Track: A Two-Stage Method to Entity Linking. In: TAC 2009 Workshop (2009)

48. Dredze, M., McNamee, P., Rao, D., Gerber, A., Finin, T.: Entity Disambiguation for Knowledge Base Population. In: Proceedings of the 23rd International Conference on Computational Linguistics. Coling 2010, pp. 277–285 (2010)
49. Monahan, S., Lehmann, J., Nyberg, T., Plymale, J., Jung, A.: Cross-Lingual Cross-Document Coreference with Entity Linking. In: Proceedings of the Text Analysis Conference. (2011)
50. Jain, A., Cucerzan, S., Azzam, S.: Acronym-Expansion Recognition and Ranking on the Web. In: 2007 IEEE International Conference on Information Reuse and Integration, pp. 209–214. IEEE (2007)
51. Hakimov, S., Oto, S.A., Dogdu, E.: Named Entity Recognition and Disambiguation Using Linked Data and Graph-based Centrality Scoring. In: Proceedings of the 4th International Workshop on Semantic Web Information Management, p. 4. ACM, New York, NY, USA (2012)
52. Milne, D., Witten, I.H.: Learning to Link with Wikipedia. In: Proceedings of the 17th ACM Conference on Information and Knowledge Management, pp. 509–518. ACM, New York, NY, USA (2008)
53. Han, X., Le Sun: A Generative Entity-mention Model for Linking Entities with Knowledge Base. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, pp. 945–954. Association for Computational Linguistics, Stroudsburg, PA, USA (2011)
54. Han, X., Le Sun: An Entity-topic Model for Entity Linking. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 105–115. Association for Computational Linguistics, Stroudsburg, PA, USA (2012)
55. Carpenter, B.: Phrasal Queries with LingPipe and Lucene: Ad Hoc Genomics Text Retrieval. In: Ellen M. Voorhees, Lori P. Buckland (eds.) Proceedings of the Thirteenth Text REtrieval Conference, TREC 2004. National Institute of Standards and Technology (NIST) (2004)
56. Ceccarelli, D., Lucchese, C., Orlando, S., Perego, R. and Trani, S.: SpotManager, https://github.com/dexter/dexter/blob/eeced3782f958f070f2448413f413e10e9df2281/dexter-core/src/main/java/it/cnr/isti/hpc/dexter/spot/clean/SpotManager.java
57. Neumann, G., Backofen, R., Baur, J., Becker, M., Braun, C.: An information extraction core system for real world German text processing. In: Grishman, R. (ed.) the fifth conference, pp. 209–216
58. Speck, R., Ngonga Ngomo, A.-C.: Named Entity Recognition using FOX. In: International Semantic Web Conference 2014 (ISWC2014), Demos & Posters (2014)
59. Speck, R., Ngonga Ngomo, A.-C.: Ensemble Learning for Named Entity Recognition. In: The Semantic Web - ISWC 2014, 8796, pp. 519–534. Springer International Publishing (2014)
60. Wikipedia:Manual of Style/Linking, https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Linking