Annotating Domain-Specific Texts with Babelfy: A Case Study

Michael Färber
University of Freiburg
Freiburg, Germany
michael.faerber@cs.uni-freiburg.de

Kristian Noullet
University of Freiburg
Freiburg, Germany
kristian.noullet@gmail.com

Boulos El Asmar
BMW Group
Munich, Germany
boulos.el-asmar@bmw.de

ABSTRACT

Freely available large knowledge graphs, such as DBpedia, Wikidata, and YAGO, generally provide a very solid representation of general knowledge, making them a good basis for text annotation. However, when it comes to annotating domain-specific text documents, these knowledge graphs need to be used with care. Moreover, publications describing real-world use cases of entity linking based on such knowledge graphs are surprisingly rare. In this paper, we describe the use case of annotating customer feedback texts written in German based on Babelfy as the text annotation service. We perform a manual evaluation of the annotations and show that Babelfy annotates around 85% of all annotations correctly. This makes Babelfy as a text annotation method and BabelNet as its knowledge graph a valid baseline for developing a custom knowledge graph and entity linking method, respectively.

CCS CONCEPTS

• **Applied computing** → **Annotation**; • **Software and its engineering** → **Use cases**; • **Information systems** → *Document representation*;

KEYWORDS

Babelfy, knowledge graph, text annotation, entity linking, customer feedback, domain specificity

ACM Reference Format:
Michael Färber, Kristian Noullet, and Boulos El Asmar. 2018. Annotating Domain-Specific Texts with Babelfy: A Case Study. In *Proceedings of 1st International Workshop on EntitY REtrieval (EYRE'18).* Lingotto, Turin, Italy, 4 pages.

1 MOTIVATION

Recently, freely available and large knowledge graphs, such as DBpedia, Wikidata, and YAGO, have been shown to cover a very wide range of entities and facts about those entities [3]. However, they are mainly restricted to modeling general knowledge but not domain-specific knowledge. This is understandable, because creating knowledge graphs is a nontrivial task with much effort involved [10], and for specific domains, the amount of structured or semi-structured data that can be used as starting point for a knowledge graph's creation is typically very limited.

As a consequence, instead of creating a completely new custom knowledge graph, an alternative is to use one of the existing large knowledge graphs and to accept the shortcomings concerning domain specificity. In particular, when it comes to using this knowledge graph for linking phrases in given texts to entities in this knowledge graph (i.e., *entity linking*, see Figure 1), for example for the purpose of enabling semantic search on these texts, this typically means to gain a lower performance (i.e., lower precision and recall values) than when texts of general knowledge are used as input. The crucial question is whether this performance of entity linking approaches based on general knowledge graphs is still acceptable for a given industrial use case with domain-specific, non-English texts.

In this paper, we present a case study in which we pursue this question. More specifically, we are given customer feedback texts written in German language from the automotive industry. These texts are positive and negative feedbacks of clients concerning services provided by a car manufacturer and should be annotated by knowledge graph entities to enable a semantic search (thereby resolving ambiguities in the language) and enhance data integration.

In our use case, we selected BabelNet[1] [9] as the knowledge graph and Babelfy[2][3] [7] as the text annotation service. Our decision was based on the following main aspects: (1) BabelNet is obtained by combining data from Wikipedia, Wikidata, OmegaWiki, Wordnet, Open Multilingual WordNet, and Wiktionary. Thus, it provides a wide range of entities and general concepts and does not only cover named entities but also actions and other items. Moreover, due to the combination of multiple knowledge graphs and other knowledge repositories, the chances are high that this knowledge graph (BabelNet) also contains domain-specific entities and concepts, which is especially useful for our use case. (2) BabelNet contains labels and descriptions for entities not only in English but numerous other languages. Indeed, BabelNet is considered to be the largest available multilingual knowledge graph [7] to date. This makes BabelNet especially appealing to be used for non-English texts. In our use case, we have German texts at our disposal. Using the entity linking service Babelfy, which is based on BabelNet, allows us to annotate the texts with BabelNet entities and to have a language-independent knowledge representation of the texts.

In summary, we target the following research question in this paper: Given domain-specific, German texts that should be used for semantic search systems, is it reasonable to apply Babelfy as an annotation service and to use BabelNet as its underlying knowledge graph?

The rest of this paper is structured as follows: In Section 2, we present related works on annotating domain-specific, non-English texts in general and using Babelfy as the text annotation service. In Section 3, we describe our process of annotating our texts with Babelfy. Our manual evaluation of the annotations is described in

EYRE'18, October 2018, Lingotto, Turin, Italy

[1] https://babelnet.org/
[2] http://babelfy.org/
[3] Babelfy is released under the Babelfy Non-Commercial License, see https://babelnet.org/license.

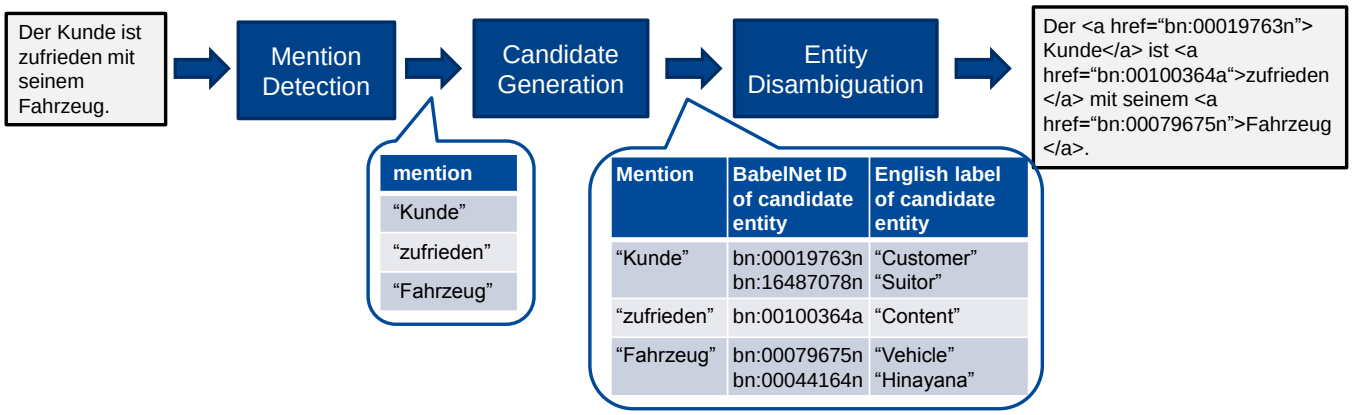

Figure 1: General steps of entity linking exemplified by the input sentence "Der Kunde ist zufrieden mit seinem Fahrzeug" (in English: "The customer is satisfied with his vehicle"). BabelNet URIs are abbreviated for better readability.

Section 4, followed by a discussion on noteworthy findings concerning the annotation of domain-specific, non-English texts (based on Babelfy and BabelNet as the knowledge graph). We conclude in Section 6 with a summary and outlook on future work.

2 RELATED WORK

Although the use case of annotating domain-specific texts and texts in languages other than English is presumably very common, not many publications exist that describe case studies about the performance and lessons learned of annotating domain-specific, non-English texts. We assume that this fact can be traced back to the following principal reasons:

(1) Case studies typically deal with real-world data (e.g., texts written by companies' clients) that cannot be published due to data privacy and copyright reasons. Even publishing metadata, aggregations, or statistical key figures of such texts is often not possible. Published case studies concerning the annotation of texts often deal with rather non-domain-dependent texts, such as news articles [5].

(2) In many cases, researchers and developers might conclude that no general findings concerning the processing of texts with Semantic Web technologies can be made and that the performance of text annotation tools depends on each individual use case. However, as entity linking applied in the field of digital humanities [1, 4] shows, general conclusions can still be made and lessons can be learned in terms of when to use which approach or how to adapt approaches and resources. This includes, for instance, how to deal with common annotation errors, such as non-verbatim mentions (e.g., misspellings, alternative writings), wrong capitalization, unseen or novel words, and irregular word context.

As outlined in the Introduction, we decided to use Babelfy as the text annotation service and BabelNet as the underlying knowledge graph due to its wide coverage of entities (originating from various knowledge graphs) and due to its support of non-English texts as input. In the past, Babelfy has been selected many times as the text annotation service for annotating English and non-English texts. In particular, when it comes to annotating non-English texts, Babelfy has been used considerably often [6, 8, 11–13] and has been found to perform among the best approaches for non-English texts [11, 12]. More specifically, Rosales-Mendez et al. [12] evaluated entity linking systems for non-English texts. The motivation was that such systems are typically developed primarily for English texts only, although in real-world settings (and especially industrial contexts), non-English texts often need to be processed. Among other systems, Babelfy showed the best performance for Spanish texts in this paper. In [11], the authors compared entity linking systems using English, German, and Spanish texts as input. Through experiments, they confirmed that Babelfy is a good choice when it comes to annotating German texts. While Babelfy generally displayed the best results in German and Spanish, it often had the lowest precision when it came to English.

While several publications on evaluating Babelfy with multiple languages exist, publications concerning evaluations of Babelfy with domain-specific texts are scarce, to the best of our knowledge. Among others, Moro et al. [8] evaluated the quality of Babelfy annotations on the large English corpus MASC 3.0, which covers various genres of written and spoken text. The authors came to the conclusion that roughly 70% of the named entities and of the word sense annotations were correct. As we will see, in our experiments, we achieved considerably higher precision values (about 85%). Note, however, that we used non-English texts as input, confirming the hypothesis of [11] that Babelfy performs particularly well for non-English texts.

3 AUTOMATIC TEXT ANNOTATION WITH BABELFY

In total, we had several thousand customer feedback texts as input for the automatic text annotation with Babelfy. In order to annotate these texts, we used Babelfy's Web API,[4] sending each document separately. Figure 1 shows the general steps of entity linking, using BabelNet as an underlying knowledge graph and an input sentence being relatively close to a real-world one.

[4] http://babelfy.org/guide

URIs	BabelNet	DBpedia
Total number	1,432	606
Distinct number	615	205

Table 1: Recognized annotations.

	being correct	being relevant
Number of surface forms	1,214	1,372

Table 2: Number of surface forms manually evaluated as "being correct" and "being relevant."

4 MANUAL EVALUATION OF ANNOTATIONS

Given our initial data set (customer feedback texts in German), which we annotated with the help of Babelfy, we sampled *30 text documents* to manually evaluate Babelfy's performance on our texts. The number of characters per document ranged from four up to 2,625. On average, each document had a length of 894.1 characters.

The 30 documents (4,374 words in total) contained in total *1,432 annotations* (615 unique annotations, i.e., with different BabelNet URIs; see Table 1). Out of these 1,432 annotations, almost half of them (606 annotations, 205 unique annotations) could also be linked to DBpedia.[5]

We evaluated the annotations by means of the evaluation metric *precision*. This describes the ratio to which found annotations are also correct (i.e., linked to the correct entity in the knowledge graph BabelNet). The judgments were performed by an expert familiar with the domain of the texts and with knowledge graphs. Of the 1,432 annotations, 1,214 were manually judged as having a correctly linked URI. This corresponds to a *precision* value of **0.847**.

In addition, we were interested in an overall estimation of which of the phrases in our text documents were "relevant" in the context of an information retrieval system (which is, for instance, using a filtered bag-of-words representation of the text documents) and what the overlap of these relevant phrases to the annotated phrases is. In other words, phrases were relevant if they were judged as salient or representative of the overall document. Note that these judgments on relevance can be considered as subjective to some degree and only give a rough estimation.[6,7] In total, 1,372 phrases were judged manually as relevant to the context (independently on whether or not these phrases were automatically annotated through Babelfy; see Table 2). This corresponds to 31.4% (1372/4374) of all phrases in the texts. About 74.85 % (1,027 / 1,372) of these relevant phrases were also annotated by Babelfy. 892 phrases (of 1447 tokens in the input) were judged as both "correct" (linking to the correct entity) and "relevant."

[5]Note that each annotation provides a corresponding BabelNet URI and that only for a subset of them are DBpedia URIs also provided.

[6]The problem with human judgments is that they are rather idiosyncratic and variable in general, posing an issue. The success of an IR system depends on how good it is at satisfying the information needs of the people. Letting these people identify relevant phrases is therefore valid at least to some degree.

[7]Note also that this "relevance" differs from recall.

5 DISCUSSION

Generally, it can be stated that both aspects of the used entity linking system and aspects of the used knowledge graph influence the usefulness and applicability of semantic annotation. In the following, we describe a few aspects of the used knowledge graph BabelNet before considering aspects of the used entity linking system Babelfy in the context of our use case. Note that knowledge graphs can be evaluated in more depth according to various data quality dimensions [3]. Also, entity linking systems can be evaluated according to typical error types [2].

Adaptability and up-to-dateness of BabelNet. BabelNet is a static knowledge graph and updated only from time to time (roughly once a year). However, especially for company internal use cases, adaptations of the used knowledge graph might be needed (e.g., including entities and facts that are particularly relevant to the company's businesses and processes). Using BabelNet as the knowledge graph off the shelf and Babelfy as the entity linking method, such requirements are deliberately omitted. In our case, we can assume that the topics written in the customer feedbacks do not change significantly over time. However, in order to also cover domain-specific entities, which users might query in semantic search systems, adaptations of BabelNet as a knowledge graph might be unavoidable. Luckily, due to the option of also using BabelNet locally (for research purposes), it is possible to use BabelNet's multilingual indices very efficiently and adapt them as needed, while at the same time not compromising data privacy.

Domain coverage of Babelfy. Aside from noun phrases, Babelfy tries to accurately detect verbs, adverbs, and adjectives, among others. These, unfortunately, led in our case to disproportionately many linking errors compared to nouns. We noticed for our texts that verbs in particular were incorrectly linked to entities from unrelated domains, such as from the medical and juridical domains (e.g., "angezeigt" was recognized as *inform police of a crime* rather than *displayed on a screen*).

Mention detection of Babelfy. Our text partially contained domain-specific phrases, in particular from the automotive industry sector. Thus, phrases such as "Boardcomputer" (in English: "board computer") were not detected by Babelfy in our texts. However, according to our investigations, based on the list of phrases that were judged manually as "relevant" and as being "not annotated at all," we can derive that most annotated entities are represented within BabelNet (although not always correctly linked in the texts). Furthermore, because we processed real-world data from users, mention detection suffered from grammar errors and incorrect word usage.

Disambiguation errors of Babelfy. Unfortunately, even initially seemingly non-ambiguous phrases in the text can be (and indeed have been) mistakenly linked to rarely occurring entities. For instance, the phrase "kleines Fahrzeug" (in English: "small vehicle") is linked to *Hinayana*, which is a special Sanskrit term for "the earliest system of Buddhist doctrine." However, the phrase should actually be linked to *Vehicle*.

Especially when it comes to real-world, noisy textual data, entity disambiguation suffers from missing or limited contexts. As a consequence, for instance, co-references (e.g., personal pronouns like "es" [in English: "it"]) cannot be resolved and are linked to wrong

entities. In real-world use cases, Babelfy therefore needs to be used with care.

Summary. To the best of our knowledge, there exist no publications stating necessary performance thresholds (e.g., in terms of precision) for entity linking in real-world applications (especially, in industrial settings). Consequently, we cannot give an absolute answer to our research question concerning the applicability of Babelfy and BabelNet in domain-specific contexts. Generally, Babelfy seems to be promising for general and slightly domain-dependent domains. However, in the industrial sector, minimum requirements concerning the precision of text annotations are presumably set very high. Thus, we believe that Babelfy's performance in our use case (obtaining a precision of 0.835) does not meet these requirements. As a possible solution, the process of semantic annotation could be performed in a semi-automated fashion (i.e., annotations could be approved by experts). However, this would require considerable human resources and efforts. If there is a possibility to create a custom knowledge graph [10] and an associated custom entity linking system, we recommend this solution. Institutions of various kinds (international companies, small and medium enterprises, and universities) have already deployed knowledge graphs for their purposes [10] or are on the verge of doing so. The use cases are thereby located in diverse domains, such as health care, the media industry, and cultural heritage. In this context, Babelfy could be used as a baseline for evaluating custom entity linking systems.

6 CONCLUSION

In this paper, we presented a case study in which we automatically annotated customer feedback texts written in German with knowledge graph entities. We thereby used Babelfy as the text annotation method and BabelNet as its underlying knowledge graph. We evaluated the performance of Babelfy given the annotations for our texts and obtained a precision value of 0.835. This shows that BabelNet might be useful as a starting point for semantically annotating texts written in German and being domain-dependent to some degree. In real-world settings, however, very high precision values are presumably needed to prevent adverse effects on a company's business and/or customers. Approving the automatically created annotations manually might be a solution. Especially when it comes to very domain-specific texts, the creation of a custom knowledge graph and a corresponding entity linking method might be favorable.

For the future, we plan on evaluating the entity linking process in more detail. Specifically, we will evaluate the single steps of mention detection and entity disambiguation, and we will also evaluate an overall recall rate. Because a high precision value is crucial for our industrial use case, we plan creating a custom knowledge graph with a corresponding entity linking component. For this purpose, Babelfy as a text annotation service and BabelNet as a knowledge graph will serve as valid baselines. This custom knowledge graph can then be combined with freely available knowledge graphs. Concerning entity linking, we plan to develop approaches for reconciling closed-domain custom text annotation methods with general domain annotation methods such as Babelfy.

REFERENCES

[1] Sebastian Arnold, Felix A. Gers, Torsten Kilias, and Alexander Löser. 2016. Robust Named Entity Recognition in Idiosyncratic Domains. *CoRR* abs/1608.06757 (2016). arXiv:1608.06757

[2] Adrian Brasoveanu, Giuseppe Rizzo, Philipp Kuntschik, Albert Weichselbraun, and Lyndon J. B. Nixon. 2018. Framing Named Entity Linking Error Types. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC'18)*.

[3] Michael Färber, Frederic Bartscherer, Carsten Menne, and Achim Rettinger. 2018. Linked data quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. *Semantic Web* 9, 1 (2018), 77–129.

[4] Francesca Frontini, Carmen Brando, and Jean-Gabriel Ganascia. 2015. Semantic Web based Named Entity Linking for Digital Humanities and Heritage Texts. In *Proceedings of the First International Workshop Semantic Web for Scientific Heritage (SW4SH'15)*. 77–88.

[5] Georgi Kobilarov, Tom Scott, Yves Raimond, Silver Oliver, Chris Sizemore, Michael Smethurst, Christian Bizer, and Robert Lee. 2009. Media Meets Semantic Web – How the BBC Uses DBpedia and Linked Data to Make Connections. In *Proceedings of the 6th European Semantic Web Conference (ESWC'09)*. 723–737.

[6] Simon Mille, Miguel Ballesteros, Alicia Burga, Gerard Casamayor, and Leo Wanner. 2016. Multilingual Natural Language Generation within Abstractive Summarization. In *Proceedings of the 1st International Workshop on Multimodal Media Data Analytics (MMDA'16)*. 33–38.

[7] Andrea Moro, Francesco Cecconi, and Roberto Navigli. 2014. Multilingual Word Sense Disambiguation and Entity Linking for Everybody. In *Proceedings of the ISWC 2014 Posters & Demonstrations Track a track within the 13th International Semantic Web Conference (ISWC'14)*. 25–28.

[8] Andrea Moro, Roberto Navigli, Francesco Maria Tucci, and Rebecca J. Passonneau. 2014. Annotating the MASC Corpus with BabelNet. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. 4214–4219.

[9] Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network. *Artificial Intelligence* 193 (2012), 217–250.

[10] Jeff Z Pan, Guido Vetere, Jose Manuel Gomez-Perez, and Honghan Wu. 2017. *Exploiting Linked Data and Knowledge Graphs in Large Organisations*. Springer.

[11] Henry Rosales-Méndez, Aidan Hogan, and Barbara Poblete. 2018. VoxEL: A Benchmark Dataset for Multilingual Entity Linking. In *Proceedings of the 17th International Semantic Web Conference (ISWC'18)*.

[12] Henry Rosales-Méndez, Barbara Poblete, and Aidan Hogan. 2017. Multilingual Entity Linking: Comparing English and Spanish. In *Proceedings of the 5th International Workshop on Linked Data for Information Extraction (LD4IE'17)*. 62–73.

[13] Henry Rosales-Méndez, Barbara Poblete, and Aidan Hogan. 2018. What should Entity Linking link?. In *Proceedings of the 12th Alberto Mendelzon International Workshop on Foundations of Data Management (AMW'18)*.