

KORE 50^{DYWC}: An Evaluation Data Set for Entity Linking Based on DBpedia, YAGO, Wikidata, and Crunchbase

Kristian Noullet, Rico Mix, Michael Färber

Karlsruhe Institute of Technology (KIT)

kristian.noullet@kit.edu, ucvde@student.kit.edu, michael.farber@kit.edu

Abstract

A major domain of research in natural language processing is named entity recognition and disambiguation (NERD). One of the main ways of attempting to achieve this goal is through use of Semantic Web technologies and its structured data formats. Due to the nature of structured data, information can be extracted more easily, therewith allowing for the creation of knowledge graphs. In order to properly evaluate a NERD system, gold standard data sets are required. A plethora of different evaluation data sets exists, mostly relying on either Wikipedia or DBpedia. Therefore, we have extended a widely-used gold standard data set, KORE 50, to not only accommodate NERD tasks for DBpedia, but also for YAGO, Wikidata and Crunchbase. As such, our data set, KORE 50^{DYWC}, allows for a broader spectrum of evaluation. Among others, the knowledge graph agnosticity of NERD systems may be evaluated which, to the best of our knowledge, was not possible until now for this number of knowledge graphs.

Keywords: Data Sets, Entity Linking, Knowledge Graph, Text Annotation, NLP Interchange Format

1. Introduction

Automatic extraction of semantic information from texts is an open problem in natural language processing (NLP). It requires appropriate techniques such as named entity recognition and disambiguation (NERD), which is the process of extracting named entities from texts and linking them to a knowledge graph (KG). In the past, several methods have been developed for this task, such as DBpedia Spotlight (Mendes et al., 2011), Babelfy (Moro et al., 2014a; Moro et al., 2014b), MAG (Moussallem et al., 2017), X-LiSA (Zhang and Rettinger, 2014), TagMe (Ferragina and Scaiella, 2010) and AGDISTIS (Usbeck et al., 2014). These methods face a lot of challenges, such as ambiguity among short pieces of text – corresponding to a potentially large number of entities (known as surface forms) – and overall hard-to-disambiguate mentions like forenames or typographical errors in texts. Thus, it is important to provide a holistic evaluation of these methods in order to identify potential strengths and weaknesses which help to improve them.

In the past, there have been a number of approaches aiming for “proper” entity linking system evaluation, such as CoNLL (Sang and Meulder, 2003), ACE (Doddington et al., 2004) or NEEL (Rizzo et al., 2015). Besides the development of entity linking systems, such approaches focus on the creation of general benchmark data sets which test and improve annotation methods and therefore entity linking. For evaluating NERD, typically gold standards are used which require manual annotation making their creation a time-consuming and complex task. However, current state-of-the-art evaluation data sets are subject to limitations:

1. The majority of evaluation data sets for NERD rely on the KGs of DBpedia or Wikipedia. In other words, for most of them, entity linking systems cannot be realistically evaluated on other KGs at present.
2. Data sets often contain data from specific domains such as sports-, celebrities- or music-related topics.

Therefore, annotation methods can only be evaluated on a small part of the KG concerning a specific domain.

3. Some data sets are not available in the RDF-based NLP Interchange Format (NIF) which allows interoperability between NLP tools, resources and annotations. This leads to an inconvenient use for developers.

There are data sets which take into account the points of domain-specificities and making use of overall accepted formats, but few allow evaluation based on KGs other than DBpedia and Wikipedia. For example, the KORE 50 (Hofart et al., 2012) and DBpedia Spotlight (Mendes et al., 2011) data sets are publicly available, converted into NIF and contain news topics from different domains. However, they both rely on DBpedia – “the central point of the Linked Open Data movement,” according to (Röder et al., 2014) – and thus are not capable of delivering results on other KGs. In order to allow further exploration of entity linking systems’ inherent strengths and weaknesses, it is of interest to evaluate their abilities on different KGs. Our goal is to provide an evaluation data set that considers each of our listed limitations and can be used with ease by other developers as well, therefore allowing for a broader spectrum of application. With KORE 50, we chose an existing popular data set to ensure a solid foundation to build upon, entity linking system comparability across varying domains, as well as a heightened ease of switching between evaluation KGs, therefore potentially increasing the likelihood of use in future systems.

For this paper, we release three sub-data sets – one for each KG which entities within the documents link to, namely YAGO, Wikidata, and Crunchbase.

We decided upon the mentioned KGs because they cover cross-domain topics (meaning: general knowledge), especially YAGO and Wikidata. Since the KORE 50 data set covers a broad range of topics (sports, celebrities, music, business etc.), they are suitable candidates for being linked to. Besides DBpedia, YAGO and Wikidata are widely used

Corpus	Topic	Format	# Documents	# Mentions	# Entities	Avg. Entity/Doc.
KORE 50	mixed	NIF/RDF	50	144	130	2.86

Table 1: Main features of the KORE 50 data set.

in the domain of NLP. Crunchbase, on the other hand, represents a different perspective altogether, making up an interesting and potentially highly valuable addition to our data set. It is largely limited to the domains of technology and business and is of considerable size (e.g., around half a million people and organizations each, as well as 200k locations (Färber, 2019), etc.). Its addition to the midst of our more famous KGs allows for more detailed evaluation relating to domain-dependency as well as up-to-dateness, among other things. Our data set is publicly available online at <http://people.aifb.kit.edu/mfa/datasets/kore50-lrec2020.zip>.

Furthermore, annotation of texts based on the above-mentioned KGs shows whether KORE 50 adequately evaluates entity linking systems on varying KGs or just covers specific domains. Therefore, KORE 50^{DYWC} is not limited to evaluating NERD systems, but can also be used to analyse KGs in terms of advantages and disadvantages with regard to varying tasks – potentially limited to (or enhanced by) domain-specific knowledge. To guarantee for qualitative and accurate results, we had multiple annotators working independently on the annotations (for further details, see Section 5.).

The rest of the paper is structured as follows. We outline related work (Section 2.), prior to presenting our annotation process and the findings obtained (Section 3.). Furthermore, we present the data set format (Section 4.). After describing our data set analysis, as well as our findings in this regard, we come to an end with our conclusion and potential future work (Section 6.).

2. Related Work

There exist a plethora of data sets for entity linking result evaluation as can be seen in the GERBIL (Usbeck et al., 2015) evaluation framework, among others. In (Röder et al., 2014), Röder et al. developed N^3 , a collection of data sets for NERD, multiple data sets relying on the DBpedia or AKSW KGs, respectively. N^3 -RSS-500 is comprised of 500 English language documents, made up of a list of 1,457 RSS feeds from major worldwide newspapers including a variety of different topics, as compiled by Goldhahn et al. (Goldhahn et al., 2012). N^3 -Reuters-128 is based on the Reuters 21578 corpus, containing news articles relating to economics. 128 documents of relatively short length were chosen from these and compiled into the mentioned N^3 -Reuters-128 data set. DBpedia Spotlight (Mendes et al., 2011), an evaluation data set published along with the similar-named entity linking tool, is solely based on DBpedia as a KG. Alike aforementioned data sets, the vast majority of existing evaluation corpora are either based on DBpedia or Wikipedia – transformation from one to another being relatively trivial. Further data sets falling into this category are AQUAINT (Milne and Witten, 2008), a corpus of English News Texts; Derczynski (Derczynski et al., 2015),

a collection of microposts; MSNBC (Cucerzan, 2007) and finally, also the data set we base ourselves on for our extension, KORE 50 (Hoffart et al., 2012).

According to (Steinmetz et al., 2013), the KORE 50 data set is a subset of the larger AIDA corpus and mainly consists of mentions considered highly ambiguous, referring to a high count of potential candidate entities for each mention. It is made up of 50 single sentences from different domains such as music, sports and celebrities, each sentence representing one document¹. The high amount of forenames requires a NERD system to derive respective entities from given contexts. Surface forms representing forenames, such as *David* or *Steve*, can be associated with many different candidates in the KG. According to Waitelonis et al. (Waitelonis et al., 2016), this leads to an extremely high average likelihood of confusion. In particular, the data set contains 144 non-unique entities (mentions) that are linked to a total of 130 entities which mainly refer to persons (74 times) and organisations (28 times). Table 1 presents the most important features of the KORE 50 data set. For more details, (Steinmetz et al., 2013) provide a precise distribution of DBpedia types in the benchmark data set.

3. Approach

This section explains in detail the creation of KORE 50^{DYWC} evaluation data set. Section 3.1. provides information on the annotation of the data set using entities from different KGs. Section 3.2. analyzes the created data set using different statistical and qualitative measures.

3.1. Annotation Process

Text annotations are up-to-date with KG versions of YAGO, Wikidata and Crunchbase dating mid-November. In order to arrive at our goal of data set equivalents of KORE 50 for YAGO, Wikidata and Crunchbase, some work was necessary prior to the manual annotation of input documents. In particular, we had to prepare the KORE 50 data set for the annotation by extracting all sentences from the data set. This was done using regular expressions and resulted in 50 single text documents containing single sentences. Those documents were used as an input for WebAnno,² a web-based annotation tool which offers multiple users the opportunity to work on one project simultaneously and review annotations, among others for the sake of inter-annotator agreement (Yimam et al., 2013). Each document was manually annotated by searching for entities in the respective KG. Each of our investigated KGs provides a web-based search engine, allowing developers to explore them relatively detailed. The DBpedia pre-annotated KORE 50

¹Note that document 23 constitutes an exception and contains two sentences.

²<https://webanno.github.io/webanno/>

Prefix	URI
dbr	http://dbpedia.org/resource/
yago	http://yago-knowledge.org/resource/
wdr	https://www.wikidata.org/entity/
cbp	https://www.crunchbase.com/person/
cbo	https://www.crunchbase.com/organization/

Table 2: Used prefixes.

data set we base ourselves on can also be double-checked through use of site-integrated search engines. This helps identifying correct entities in the KG and makes comparison to DBpedia possible. Once the correct entity was identified in the respective KG, it was used to annotate the specific part in a document. If possible, we distinguished between plural and singular nouns, but in nearly any case the plural version was not available and we annotated the singular version instead. The only exception is the word *people* and the respective singular version *person*. The entity `:People` was available in Wikidata as well as YAGO. After we finished the annotation, the documents were exported using the *WebAnno TSV3* format, which contains the input sentence as well as the annotated entities including their annotation. In the following subsection, we outline peculiarities and difficulties by comparing the annotations that resulted from the different KGs and investigating inter-annotator agreement.

3.2. Annotation Peculiarities

During the annotation process, we observed some peculiarities in the KGs as well as in the KORE 50 data set. Some of these – or the combination hereof – may explain differences in terms of performance for some NERD systems to a certain degree. Table 2 provides an overview of our used URI prefixes.

`:Richard.Nixon` is available in YAGO, Wikidata and the gold standard DBpedia, but not in Crunchbase, likely due to Richard Nixon neither being a relevant business person nor an investor.

In Figure 1, an example annotation is provided showing a whole sentence (document 49 in KORE 50) as well as its annotations. Annotated words are underlined and available entities are shown in boxes. The word *arrival*, for instance, can only be annotated using YAGO.

Note that in document 15 of the KORE 50 data set, there is a spelling error (rather *prominent* than *prominent*) which might lead to biased results for some systems. The mistake was corrected for our KORE 50^{DYWC} version.

YAGO: YAGO offers a larger amount of resources to be annotated given our input sentences in comparison to DBpedia. Many entities originate from *WordNet*,³ a linguistic database integrated into the YAGO KG. Examples are *father*, *musical* or *partner*. In some cases, for example *year*, more than one entity is possible⁴. Overall, only one entity that was annotated with DBpedia was not found (`dbr:First_Lady_of_Argentina`). Interestingly, some resources were not available in YAGO although

one would expect them to be in the KG, for example *time* or *band*.

Wikidata: Wikidata provides information for a larger amount of mentions than DBpedia. For example the children of David Beckham (*Brooklyn*, *Romeo*, *Cruz*, *Harper Seven*), many nouns (*album*, *duet*, *song*, *career* etc.), articles (*the*), as well as pronouns (*he*, *she*, *it*, *his*, *their* etc.) can be found in Wikidata. Similar to YAGO, only one entity that has been annotated in the gold standard of DBpedia was not found (`dbr:First_Lady_of_Argentina`). The entity of Neil Armstrong (`wdr:Q1615`), the first human on the moon, was not available for a period of time, because his name was changed to a random concatenation of letters and the search function did not find him. Moreover, contrary to expectations, some resources, such as *groceries* and *archival*, were not available within Wikidata.

Crunchbase: In Crunchbase, less entities were found compared to DBpedia due to its tech-focused domain. Furthermore some famous people like *Heidi Klum*, *Paris Hilton* or *Richard Nixon* (former US president) are not included in Crunchbase. Since only parts of KORE 50 contain news articles on companies and people, Crunchbase fails to deliver information on the remaining documents. Moreover, the Crunchbase entity `cbp:frank.sinatra` has a picture of the real Frank Sinatra but describes him as the owner of `cbo:Sinatra`, a “web framework written in Ruby that consists of a web application library and a domain-specific language.”⁵

4. Data Set Format

As outlined in Section 3.1., we used the *WebAnno TSV3* format provided by the WebAnno annotation tool. The export files possess a convenient representation of the data which guarantees comfortable further processing. In Listing 1, an example output file is provided. At the very top, the input text is displayed. Each line consists of a word counter, followed by the character position of the respective word (using start and end offsets), the word itself and, if available, the annotation. The *tsv* files can easily be converted into the NLP Interchange Format (NIF) which allows interoperability between NLP tools, resources and annotations. As such, our data set can be used by a broad range of researchers and developers.

5. Data Analysis and Results

Table 3 shows the number of unique entities per KG in the KORE 50 data set. In parentheses, all non-unique entities (meaning multiple occurrences of the same entity are included) are displayed. The third column shows the average number of annotations per document. The original KORE 50 data set set has 130 unique entities from DBpedia (Steinmetz et al., 2013) which results in 2.60 entities per document. Obviously, most entities were found in Wikidata due to the broad range of topics and domains. On the contrary, Crunchbase contains only 45 unique entities in total, which equals 0.90 entities per document. This results from the strong focus on companies and people which allowed us to only annotate parts of the data set. Some documents

³<https://wordnet.princeton.edu>

⁴`yago:wordnet_year.115203791` or `yago:wordnet_year.115204297`

⁵<https://www.crunchbase.com/organization/sinatra>

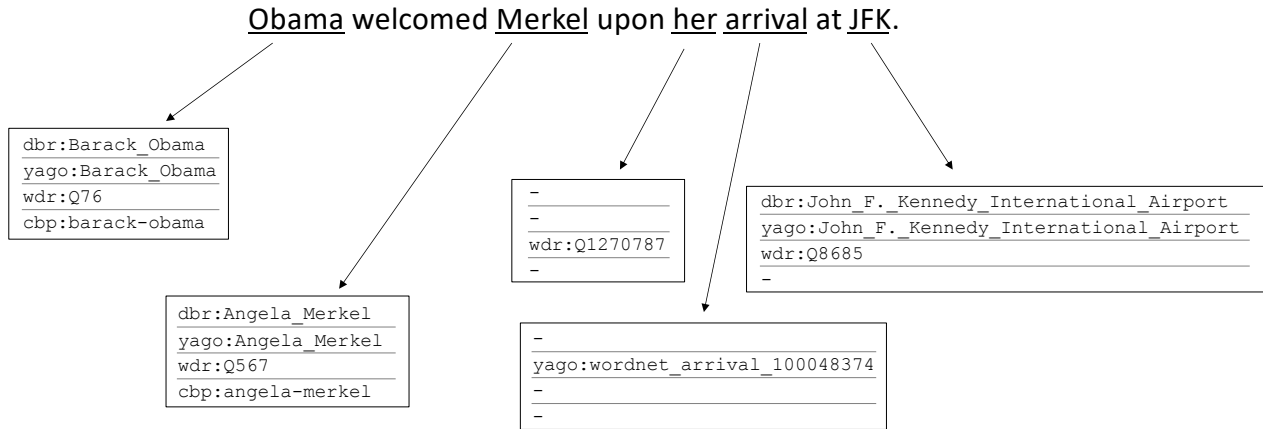


Figure 1: Example annotation showing the availability of entities in DBpedia, YAGO, Wikidata and Crunchbase.

```
#Text=Heidi and her husband Seal live in
Vegas.
1-1    0-5    Heidi    yago:Heidi_Klum
1-2    6-9    and      -
1-3    10-13   her      -
1-4    14-21   husband yago:
wordnet_husband_110193967
1-5    22-26   Seal     yago:Seal_(musician
)
1-6    27-31   live     -
1-7    32-34   in       -
1-8    35-40   Vegas    yago:Las_Vegas
1-9    40-41   .        -
```

Listing 1: Sample TSV3 output.

completely miss out on annotations from Crunchbase, especially when the text does not include information about companies and people relating to them, as is the case for example with sports-related news, such as “*Müller scored a hatrick against England.*” (document 34). A second example can be made of David Beckham’s children (*Brooklyn, Romeo, Cruz, Harper Seven*) who are not subject of the Crunchbase KG because they do not belong to its group of entrepreneurs or business people. However, they can be found in Wikidata. Another interesting fact that shows why Wikidata has such a high number of entities are pronouns like *he, she* or *it* and articles like *the* which are basic elements of each linguistic formulation. These also explain the huge gap between mentions and unique entities due to articles like *the* being used in a multitude of sentences.

KG	Unique (Non-Unique)	Entities per Doc.
DBpedia	130 (144)	2.60
YAGO	167 (217)	3.34
Wikidata	189 (309)	3.78
Crunchbase	45 (57)	0.90

Table 3: Number of entities and entities per document.

KG	Non-Unique Entities	Annotation Density
DBpedia	144	21.6%
YAGO	217	32.6%
Wikidata	309	46.4%
Crunchbase	57	8.6%

Table 4: Non-unique entities and their relative annotation density for each KG.

5.1. Annotation Differences based on KG

Two independent researchers annotated KORE 50 with the respective entities from YAGO, Wikidata and Crunchbase. Afterwards, results were compared and ambiguous entities discussed. This helped identify peculiarities in the underlying KGs and documents of the KORE 50^{DYWC}. In summary, the annotation results were similar to a large extent with exception of a number of ambiguous cases which we will be explaining in the following. We observed generally no verbs being included in any of the investigated KGs. On the flipside, using Wikidata proved to be difficult for the identification process of specific entities due to its sheer size. Especially pronouns such as *he, she* or *it* could be misleading. The third person singular *he*, for example, has two matching entities with the following descriptions: *third-person singular* or *third-person masculine singular personal pronoun*. The latter one was chosen due to it describing the entity in further detail. Another difficult task consists in entities referring to more than a single word, as is the case for *Mayall’s band*, referring to the entity `wdr:Q886072 (John.Mayall.and.the.Bluesbreakers)` and should not be labeled `wdr:Q316282 (John.Mayall)` and `wdr:Q215380 (band)` separately. In Crunchbase, the entity for `cbo: Arsenal` has a misleading description that characterizing the football club as a multimedia production company. Nevertheless, the logo as well as all links and ownership structures fit the football club. A similar issue occurred with the entity `cbp: Frank.Sinatra`, a famous American singer and actor. The picture clearly displays Frank Sinatra while the description tells us of him being the chairman of the board of `cbo: Sinatra`, a web framework that sells fan merchandising, among others.

	DBpedia	YAGO	Wikidata	Crunchbase
DBpedia (130/144)	-	-0.22 (-0.34)	-0.31 (-0.53)	1.89 (1.53)
YAGO (167/217)	0.28 (0.51)	-	-0.12 (-0.30)	2.71 (2.81)
Wikidata (189/309)	0.45 (1.15)	0.13 (0.42)	-	3.20 (4.42)
Crunchbase (45/57)	-0.65 (-0.60)	-0.73 (-0.74)	-0.76 (-0.82)	-

Table 5: Information gain of entities between KGs for unique (non-unique) entities. In brackets after the KG name, the number of unique/non-unique entities is shown. The value of 1.89 on the top right, for instance, indicates the information gain achieved using DBpedia in comparison to Crunchbase as KG for entity linking. (For *non-unique* entities in brackets)

5.2. Abstract Concepts

Phrases and idioms constitute hard disambiguation tasks due to their inexistence in all of our investigated KGs; such a case may be exemplified by document 2. Given “*David and Victoria added spice to their marriage.*”, we did not find the idiom “*to add spice*” in the underlying KGs and decided to not annotate the respective part. If it made sense in the context, specific parts of the idiom were annotated. Consequently, for the observed example, annotating *spice* as a flavour used to season food would not match the context and probably yield unexpected, as well as unsatisfactory results in regards to the NERD system. The same can be noted for document 3 (“*Tiger was lost in the woods when he got divorced from Elin.*”) related to “*being lost in the woods*” and for document 27 (“*the grave dancer*” refers to a nickname).

5.3. Issues with Original Data Set

In addition to the mentioned examples, we observed some inaccuracies in the ground truth data set of the KORE 50 which was annotated using DBpedia. In document 15, `dbr:Sony_Music_Entertainment` was labeled instead of using `dbr:Sony`. Sony music entertainment did not exist in the 1980s but was renamed by Sony in 1991 from CBS Records to the new brand (Reuters, 1990). In document 36 (“*Haug congratulated Red Bull.*”), *Red Bull* was labeled as `dbr:FC_Red_Bull_Salzburg`, but it is more likely to be `dbr:Red_Bull_Racing` due to *Norbert Haug* being the former vice president of Mercedes Benz Motorsport activity. Moreover, we could not find any news articles regarding Haug congratulating Red Bull Salzburg. By correcting these inaccuracies, the quality of the KORE 50^{DYWC} – and thus of NERD evaluation – could be improved.

5.4. Annotation Issues

Every document contains annotations from all used KGs except Crunchbase. With 48%, almost half of the documents from KORE 50 could not be annotated using Crunchbase. In NERD, empty data sets will lead to an increased false positive rate and thus a lower precision of the entity linking-system (Waitelonis et al., 2016). Therefore, documents that cannot be annotated, as is often the case with Crunchbase, should be excluded for a more realistic evaluation. Documents with extremely unbalanced annotations on different KGs should also be treated with caution. For some documents, Crunchbase was only able to provide a single annotation, whereas DBpedia managed to provide multiple. In document 16, the only entity available in

Crunchbase was `cbp:Rick_Rubin`. However, DBpedia provided two additional entities (`dbr:Johnny_Cash`, `dbr:American_Recordings_(album)`). With only one entity labeled in a whole sentence, this document is likely to not evaluate NERD systems properly. One solution to this problem would be to split the KORE 50^{DYWC} into smaller subsets by, for example, deleting sentences that can only be sparsely annotated or even completely miss out on annotations using Crunchbase.

5.5. Annotation Density

Table 4 shows the *annotation density* on each of the four underlying KGs. The annotation density (Waitelonis et al., 2016) is defined as the total number of unique annotations in relation to the data set word count. In total, the KORE 50 data set consists of 666 words⁶. Wikidata has the highest annotation density being able to annotate 46.4% which equals nearly half of the words included in the whole data set, followed by YAGO with 32.6%. The lowest density was observed using Crunchbase with 57 out of 666 words annotated to an equivalent of 8.6%. Due to the resulting ground truth for Crunchbase possessing a more restrained set of expected entities to be found, the task of context determination may prove harder for systems and therewith provides a different angle to test linking system quality with. On the other side of the spectrum, due to the large number of entities linked to Wikidata, the complexity for some systems may be greatly heightened, potentially yielding worse results, therewith testing a different qualitative aspect for the same base input documents. Combining these together allows for easier comparison of KG-agnostic approaches’ qualities, as well as opening the pathway for evaluating on systems allowing custom KG definitions, among others.

5.6. Information Gain

In order to identify the *information gain* (an asymmetric metric) achievable with one KG in comparison to a second one, we calculated *entity overlap*. The results are presented in Table 5. Entity overlap is defined as the difference between the unique (non-unique) entities in two KGs. Dividing the overlap by the subtracted KG in the basis then yields information gain. The result is a decimal number which can also be interpreted as a gain (or loss) in percent. The numbers in the first column right behind the KGs show the number of unique/non-unique entities used for calculation. The

⁶This number was derived using a Python script counting all words including articles and pronouns and equals exactly half of the amount calculated by (Röder et al., 2014) (1332).

entries in the table display information gain for unique entities and information gain for non-unique entities in brackets. For example, 1.89 on the top right explains how much information gain was achieved using DBpedia in comparison to Crunchbase. It is calculated by subtracting 45 from 130 and dividing it by 45. Calculations for non-unique entity information gain are analogous. In this example, an information gain of 1.53 is achieved by subtracting 57 from 144 and dividing it by 57. Negative values are also possible. On the bottom left, for example, an information loss of -0.65 for unique entities was achieved for Crunchbase in comparison to DBpedia. The greatest information gain was achieved using Wikidata instead of Crunchbase, with a value of 3.20 for unique entities and 4.42 for non-unique respectively. Of course, the greatest information loss was noticeable using Crunchbase instead of Wikidata with -0.76 for unique entities and -0.82 for non-unique. As such, from a general point of view, linking with Wikidata should yield the highest amount of information gained. Nevertheless, this does not imply for it to be the best applicable KG for all use cases, as linking with other KGs, such as Crunchbase, may be beneficial in order to provide further domain-specific knowledge and a potentially overall harder task for systems to master.

5.7. Disambiguation

In addition, we investigated the degree of *confusion* regarding KORE 50. In (Waitelonis et al., 2016), two definitions are presented both indicating the likelihood of confusion. First, the average number of surface forms per entity and second, the average number of entities per surface form. According to the author, the latter one is extraordinarily high on the ground truth data set with an average of 446 entities per surface form. This correlates to our observations that KORE 50 is highly ambiguous. From a qualitative perspective, confusion in the data set is relatively high. There are many words that can be labeled to a variety of entities, for example *Tiger* (Tiger Woods (Golf), Tiger (animal), Tiger (lastname)). Furthermore, forenames by themselves are highly ambiguous in terms of surface forms and could therefore be (erroneously) linked to a considerable amount of entities. For instance, *David* can refer to David Beckham (footballer), David Coulthard (Formula 1) or any other person named David which makes disambiguation extremely complicated.

6. Conclusion

In this paper, we presented an extension of the KORE 50 entity linking data set, called KORE 50^{DYWC}. By linking phrases in the text not only to DBpedia, but also to the cross-domain (and widely used) knowledge graphs YAGO, Wikidata, and Crunchbase, we are able to publish one of the first evaluation data sets for named entity recognition and disambiguation (NERD) which works for multiple knowledge graphs simultaneously. As a result, NERD frameworks claiming to use knowledge graph-agnostic methods can finally be evaluated as such, rather than being evaluated only on DBpedia/Wikipedia-based data sets. In terms of future work, besides incorporating annotations with additional knowledge graphs to our extended

KORE 50 data set, we aim to develop a NERD system that is knowledge graph agnostic, and thus can be evaluated by the proposed data set.

References

- Cucerzan, S. (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 2007, pages 708–716.
- Derczynski, L., Maynard, D., Rizzo, G., van Erp, M., Gorrell, G., Troncy, R., Petrak, J., and Bontcheva, K. (2015). Analysis of named entity recognition and linking for tweets. *Information Processing & Management*, 51(2):32–49.
- Doddington, G. R., Mitchell, A., Przybocki, M. A., Ramshaw, L. A., Strassel, S. M., and Weischedel, R. M. (2004). The Automatic Content Extraction (ACE) Program – Tasks, Data, and Evaluation. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, LREC’04.
- Färber, M. (2019). Linked Crunchbase: A Linked Data API and RDF Data Set About Innovative Companies. *arXiv preprint arXiv:1907.08671*.
- Ferragina, P. and Scaiella, U. (2010). TAGME: on-the-fly annotation of short text fragments (by wikipedia entities). In *Proceedings of the 19th ACM Conference on Information and Knowledge Management*, CIKM’10, pages 1625–1628.
- Goldhahn, D., Eckart, T., and Quasthoff, U. (2012). Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, LREC’12, pages 759–765.
- Hoffart, J., Seufert, S., Nguyen, D. B., Theobald, M., and Weikum, G. (2012). KORE: keyphrase overlap relatedness for entity disambiguation. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM’12, pages 545–554.
- Mendes, P. N., Jakob, M., García-Silva, A., and Bizer, C. (2011). DBpedia Spotlight: Shedding Light on the Web of Documents. In *Proceedings the 7th International Conference on Semantic Systems*, I-SEMANTICS’11, pages 1–8.
- Milne, D. N. and Witten, I. H. (2008). Learning to Link with Wikipedia. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, CIKM’08, pages 509–518.
- Moro, A., Cecconi, F., and Navigli, R. (2014a). Multilingual Word Sense Disambiguation and Entity Linking for Everybody. In *Proceedings of the 13th International Semantic Web Conference*, ISWC’14, pages 25–28.
- Moro, A., Raganato, A., and Navigli, R. (2014b). Entity linking meets word sense disambiguation: a unified approach. *TACL*, 2:231–244.
- Moussallem, D., Usbeck, R., Röder, M., and Ngomo, A. N. (2017). MAG: A Multilingual, Knowledge-base Agnostic and Deterministic Entity Linking Approach. In

- Proceedings of the Knowledge Capture Conference, K-CAP'17*, pages 9:1–9:8.
- Reuters. (1990). CBS Records Changes Name. <https://www.nytimes.com/1990/10/16/business/cbs-records-changes-name.html>. (last accessed: 2019-12-01).
- Rizzo, G., Basave, A. E. C., Pereira, B., and Varga, A. (2015). Making Sense of Microposts (#Microposts2015) Named Entity rEcognition and Linking (NEEL) Challenge. In *Proceedings of the the 5th Workshop on Making Sense of Microposts*, pages 44–53.
- Röder, M., Usbeck, R., Hellmann, S., Gerber, D., and Both, A. (2014). N³-A Collection of Datasets for Named Entity Recognition and Disambiguation in the NLP Interchange Format. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC'14*, pages 3529–3533.
- Sang, E. F. T. K. and Meulder, F. D. (2003). Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. In *Proceedings of the Seventh Conference on Natural Language Learning, CoNLL'03*, pages 142–147.
- Steinmetz, N., Knuth, M., and Sack, H. (2013). Statistical Analyses of Named Entity Disambiguation Benchmarks. In *Proceedings of the NLP & DBpedia Workshop, NLP&DBpedia'13*.
- Usbeck, R., Ngomo, A. N., Röder, M., Gerber, D., Coelho, S. A., Auer, S., and Both, A. (2014). AGDISTIS - Graph-Based Disambiguation of Named Entities Using Linked Data. In *Proceedings of the 13th International Semantic Web Conference, ISWC'14*, pages 457–471.
- Usbeck, R., Röder, M., Ngomo, A. N., Baron, C., Both, A., Brümmner, M., Ceccarelli, D., Cornolti, M., Cherix, D., Eickmann, B., Ferragina, P., Lemke, C., Moro, A., Navigli, R., Piccinno, F., Rizzo, G., Sack, H., Speck, R., Troncy, R., Waitelonis, J., and Wesemann, L. (2015). GERBIL: general entity annotator benchmarking framework. In *Proceedings of the 24th International Conference on World Wide Web, WWW'15*, pages 1133–1143.
- Waitelonis, J., Jürges, H., and Sack, H. (2016). Don't compare Apples to Oranges: Extending GERBIL for a fine grained NEL evaluation. In *Proceedings of the 12th International Conference on Semantic Systems, SEMANTICS'16*, pages 65–72.
- Yimam, S. M., Gurevych, I., Eckart de Castilho, R., and Biemann, C. (2013). WebAnno: A flexible, web-based and visually supported system for distributed annotations. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL'13*, pages 1–6, Sofia, Bulgaria.
- Zhang, L. and Rettinger, A. (2014). X-LiSA: cross-lingual semantic annotation. *Proceedings of the VLDB Endowment*, 7(13):1693–1696.