Multilingual Entity Linking: Comparing English and Spanish

Henry Rosales-Méndez, Barbara Poblete and Aidan Hogan

Center for Semantic Web Research, DCC, University of Chile
{hrosales,bpoblete,ahogan}@dcc.uchile.cl

Abstract. The Entity Linking (EL) task is concerned with linking entity mentions in a text collection with their corresponding knowledge-base entries. The majority of approaches have focused on EL over English text collections. However, some approaches propose language-independent or multilingual approaches to perform EL over texts in many languages. In this paper, our goal is to see how well EL systems perform outside of the primary language (often English). We first provide a survey of EL approaches that present evaluation over multiple languages. We then provide results of an initial study comparing selected entity linking APIs for equivalent documents and sentences in English and Spanish. Multilingual EL approaches fare best for Spanish, though all approaches still perform better for English text than the corresponding Spanish text. This indicates that there is an important gap between EL techniques for English in relation to Spanish (and possibly for many other languages) which has not been addressed yet. However, we leave investigation of the causes of this gap for future work, which could be due to many factors, for example, to differences in existing multilingual knowledge bases.

Keywords: multilingual, entity linking, information extraction

1 Introduction

Entity Linking is a task in Information Extraction that focuses on linking the entity mentions in a text collection with entity identifiers in a given knowledge base. Such a task has various applications, including semantic search, document classification, semantic annotation and text enrichment, as well as forming the basis for further Information Extraction processes.

Since EL is a challenging task with various applications, a wide range of EL approaches have been proposed in the literature. While many show good results in terms of recognizing and disambiguating entity mentions, most provide evaluation for EL exclusively over English texts (e.g., [1,2,3]). Other monolingual EL methods only consider a specific non-English language to link (e.g., [4,5,6,7]). Such approaches often use resources targetted at a specific language, such as recognition models trained over corpora of that language, or a knowledge base with labels only in that language. Such "monolingual" systems cannot be readily configured to perform EL over texts not expressed in the primary language.

However, other "multilingual" systems (e.g., [8,9,10,11,12,13,14,1,15]) allow for selecting from a variety of languages over which to apply EL. Such systems often either use generic language-agnostic components, or support an array of components for a specified list of languages that can be invoked based on the user-selected language. Likewise, they often employ knowledge bases, translation services or other resources that offer multilingual information as a reference.

In this paper, we wish to explore the state of the art in multilingual EL systems and techniques. In particular, our core research question is how EL systems perform in multilingual settings, where we wish to see, for example, if most systems perform better over English texts (as a primary language) versus other languages. Along these lines, we will first present some background on multilingual EL systems, providing a survey of the techniques they use, as well as the languages they support. Complimenting this survey, we present the results of some initial experiments testing a selection of EL systems over the same text collection in both English and Spanish. In particular, we are interested in the following questions: (1) How does EL performance differ between English and Spanish? (2) Do multilingual systems configured for the language perform much better for Spanish than monolingual systems not configured for that language? (3) What might be the possible reasons for the observed results?

To gain initial answers to these questions, we take an existing gold standard from the SemEval challenge with equivalent text in the Spanish and English languages, applying it for a selection of previously untested systems (we focus on systems with publicly available APIs). Though the focus of our experiments is on Spanish as the "second" language, the results do also provide insights more generally into the state of the art with respect to EL for multilingual settings.

We begin with some background on the EL process and multilingual resources that have emerged in the past few years (Section 2). We then present a survey of EL approaches that offer evaluation over multiple languages (Section 3). Next we present the results of our experiments comparing the performance of selected EL systems over equivalent texts in the English and Spanish languages (Section 4). We finish the paper by summarizing some main conclusions (Section 5).

2 Background

2.1 Entity Linking

Let E be a set of entities in a knowledge base and M the set of entity mentions in a given text collection. The EL process either associates (or links) each $m \in M$ with its corresponding entity $e \in E$ or concludes that there is no entity in E that corresponds to m. In the case that there are no entities in the knowledge base that correspond to a particular entity mention, then that mention is labeled *NIL* (Not In Lexicon) and is sometimes called an *unlinkable mention*.

In more detail, the EL process can be divided into two main phases:[1]

[1] In some works, EL is only considered to refer to the ED phase [16]. Here we see EL as being composed of both ED and ER.

Entity Recognition (ER) Entity mentions in the text are located. This problem directly relates to the traditional Information Extraction task of Named Entity Recognition (NER), where a variety of methods combining patterns, rules, lexicons and machine learning techniques have been applied. Such tools can be reused in an EL setting. However, traditional NER often focuses on recognition of entities from a standard selection of types – typically, persons, organizations, places, and other – where many EL scenarios involve knowledge bases with other types of entities. Hence, so-called End-to-End systems develop custom recognition tools that use the labels of the entities in the knowledge base during recognition (e.g., [17,18]).

Entity Disambiguation (ED) Entity mentions are associated with relevant knowledge base identifiers. This phase can be further divided as follows:

- *Candidate entity generation*: For each entity mention $m \in M$ this stage selects E_m: a candidate set $E_m \subseteq E$ that represents entities with a high probability of corresponding to m is selected, often based on matching m with entity labels in the knowledge base.
- *Candidate entity ranking*: Each entity $e_m \in E_m$ is ranked according to an estimated confidence that it is the referent of m. This can be performed considering a variety of features, such as the perceived "popularity" of e_m, its relation to candidates for nearby mentions, and so forth. The candidate in E_m with the best ranking may be selected as the link for m, possibly assuming it meets a certain threshold confidence.
- *Unlinkable mention prediction*: Some tools consider unlinkable mentions, where no entity in the knowledge base meets the required confidence for a match to a given entity mention m. Depending on the application scenario, these mentions may be simply ignored, or may be proposed as "emerging entities" that could be added to the knowledge base.

In many modern EL systems, the line between the EL and ED phases is blurred; these systems include the End-to-End approaches previously mentioned, but also other systems that apply ER and EL jointly in the same model [19,20,21]. Other systems assume that entity mentions have already been isolated (using some existing approach) and rather focus specifically on the ED phase [16].

2.2 Multilingual Resources

Multlingual EL frameworks typically rely on knowledge bases that contain multilingual information; for example, Wikipedia claims 296[2] supported languages, where such information can be leveraged for multilingual EL tasks. For instance, Wikipedia is used by Guo et al. [14] as a *Wiki-dictionary* to translate a Chinese title (entity) of Wikipedia to its corresponding English title. Likewise, knowledge bases built from Wikipedia – including DBpedia [18], Wikidata [22], YAGO [23], etc. – also often offer multilingual information, which in turn can be used to perform multilingual EL. One such example is DBpedia Spotlight [2], whose public API now allows to select from 10 different languages.[3]

[2] `https://en.wikipedia.org/wiki/List_of_Wikipedias`; July 1, 2017.
[3] `http://demo.dbpedia-spotlight.org/`; July 28, 2017.

However, the usage of Wikipedia and related resources in multilingual scenarios has limitations, in particular because the level of available information varies across languages, or because explicit cross-lingual links are not available. According to the 2013 study by Wang et al. [24], only 6%, 6% and 17% of English Wikipedia pages are linked, respectively, with their corresponding Chinese, Japanese and German Wikipedia articles. This issue is addressed by a variety of approaches to boost the level of cross-linking of Wikipedia articles [25,24]. In another direction, Navigli and Ponzetto [26] propose an extension of Wikipedia, called BabelNet, intended to fill this lack of information in resource-poor languages using Machine Translation and the lexical knowledge base WordNet.

Some conferences and challenges have included multilingual contests with the goal of increasingly drawing attention to EL for languages other than English. For instance, the Knowledge Base Population Track of the 2011 Text Analysis Conference[4] (TAC KBP) included a multilingual dataset for English and Chinese languages [27]; in the following years, they have centered their attention on the English, Chinese, and Spanish languages. More recently, the SemEval series of workshops has presented various tasks dedicated to multilingual EL. In fact, we will use the SemEval 2015 Task 13 dataset to perform our experiments; details will be provided later in Section 4. But first we provide a more detailed survey of the multilingual EL approaches that have emerged in the past few years.

3 A Survey of Multilingual EL Approaches

In Table 1, we provide a survey of the principal multilingual EL approaches published in the literature. Note that one could consider a variety of criteria to distinguish monolingual and multilingual systems, where indeed one could run monolingual tools over multiple languages and expect to link mentions, such as *Michael Jackson*. Thus we use the criterion that the paper explicitly performs experiments over multiple languages, presenting those languages in the table.

A high-level inspection of the table will reveal that English is by far the most popular language. Beyond that, most languages tackled are European languages, with Spanish, French, German and Dutch appearing frequently. Outside of these European languages, Chinese is the most commonly encountered.

Some of the approaches mentioned in this table do not actually address the multilingual problem directly. Rather they are developed as language-agnostic EL systems that rely on generic processing methods that can perform EL over a broad range of languages assuming a suitable knowledge-base with lexical forms (i.e., entity labels and aliases) in that language. Such systems include KIM [11], SDA [9], THD [10], TAGME [1,28] and AGDISTIS [12].

An example of a (largely) language-agnostic approach is DBpedia Spotlight. The first version of DBpedia Spotlight [2] only supports English language. However, a recent extension of DBpedia Spotlight was introduced in [8] which addresses multilingual EL using the variety of language versions now available for

[4] `https://tac.nist.gov/2011/`; July 28, 2017.

Table 1. Overview of multilingual EL approaches. The italicized approaches will be incorporated as part of our experiments.

Name	Year	Evaluated Languages	Demo	Src	API
KIM [11]	2004	English, French, *Spanish*	✓	✗	✓
SDA [9]	2011	English, French	✗	✗	✗
ualberta [14]	2012	English, Chinese	✗	✗	✗
HITS [25]	2012	English, *Spanish*, Chinese	✗	✗	✗
THD [10]	2012	English, German, Dutch	✓	✓	✓
DBpedia Spotlight [2,8]	2013	English, Italian, Russian, Dutch, French, German, *Spanish*, Hungarian, Danish	✓	✓	✓
TAGME [1,28]	2013	English, German, Dutch	✓	✗	✓
Wang-Tang [24]	2013	English, Chinese	✗	✗	✗
AGDISTIS [12]	2014	English, German	✓	✓	✓
Babelfy [13]	2014	English, *Spanish*, French German, Italian	✓	✗	✓
WikiME [29]	2016	English, *Spanish*, French, Italian, Chinese, German, Thai, Arabic, Turkish, Tamil, Tagalog, Urdu, Hebrew	✓	✗	✗
FEL [15]	2017	English, *Spanish*, Chinese	✗	✓	✗

DBpedia. In this multilingual version, DBpedia Spotlight identifies the entity mentions using Apache OpenNLP[5] and from the sequences of capitalized words. To perform ranking, the authors consider various (standard) features, including, for example, the probability that a mention could be a text anchor in Wikipedia.

While the previous systems assume a knowledge base in the same language as the text to analyze, a variety of tools rather support *cross-lingual EL*, where the goal is to link text in a language different to that from the given knowledge-base. Often the goal is to match text in a language other than English to a knowledge base with labels in English. This helps to address the aforementioned asymmetry in the structured information available in English versus other languages. Such cross-lingual approaches include ulberta [14], HITS [25], Babelfy [13], and those proposed by Wang and Tang [24], and Tsai and Roth [29]. We now discuss the two most recent cross-lingual EL systems in more detail.

Babelfy [13] is a graph-based approach for performing multi-lingual/cross-lingual EL over BabelNet. Viewed as a multigraph, each entity belonging to BabelNet is enriched with a set of related vertices (semantic signature) according to the Random Walk with Restart algorithm [30]. Additionally, each edge is

[5] `http://opennlp.apache.org/`

weighted by the number of directed cycles of length 3; thus the idea is that edges belonging to more dense areas are more highly weighted. The recognition stage uses a part-of-speech tagger, then identifying initial substring matches between mentions and entity labels. A new graph is built for all mentions and candidate entities. Finally, the entity belonging to the densest subgraph is selected.

WikiME [29] links a text collection in any language supported by Wikipedia to each corresponding entity in the English Wikipedia. In an initial stage, the skip-gram model of word2vec is applied to each language in Wikipedia separately. These foreign language embeddings are then projected with the English embeddings in a unified space. The entity mention recognition stage is based on a multi-class classification model and candidates are selected using analysis of the text anchor in that language's Wikipedia. To perform ranking, they use various features relating to mentions and their candidate entities, which are used for classification purposes using a Support Vector Machine.

4 Experiments

The motivating question for this paper is: *how well do available EL systems perform for languages other than English (as the most common primary language)*? In this section, we thus present some preliminary experiments to gain insights into this question. In particular, we perform experiments comparing EL over the same text expressed in English and Spanish for a variety of systems.

Metrics: We apply well-known metrics for measuring the performance of systems with respect to a gold standard. First, we count the *true positive* (tp) entity mentions that have been correctly identified by the ER phase, or correctly linked to the knowledge base by the ED phase, depending on the phase under evaluation. In contrast, the *false positive* (fp) measure counts those entities that are wrongly identified as entity mentions, or those entity mentions that have been linked to a wrong knowledge-base entity. Along similar lines, *true negatives* (tn) counts those mentions/links not detected by the system and not given by the gold standard, while *false negatives* (fn) counts mentions/links missed by system but given by the gold standard. EL-level metrics then combine the ER and ED phases, meaning that true positives are those mentions that are detected *and* correctly linked, false positives are those mentions that are not detected *or* not correctly linked, etc. Thereafter, the precision measure is computed by $P = \frac{tp}{tp+fp}$ and the recall by $R = \frac{tp}{tp+fn}$. Finally, our main metric is the F_1 score, which is the harmonic mean between both criteria: $F_1 = \frac{2P \times R}{P+R}$.

Dataset: A number of benchmarking frameworks have been proposed for Entity Linking systems, the most recent and comprehensive of which is GERBIL [31]; however, the system does not explicitly offer multi-lingual datasets. However, other comparative evaluations have looked at multiple languages. For example, Narducci et al. [28] perform comparison of a variety of approaches – TAGME, WikiMiner and DBpedia Spotlight – for German and Dutch text collections.

Table 2. Replicating results of available systems for the Spanish and English texts of the SemEval 2015 Task 13 (also adding novel Babelfy results)

	Spanish			English		
System	**P**	**R**	**F1**	**P**	**R**	**F1**
SUDOKU-Run2	0.607	0.525	0.563	0.640	0.609	0.625
SUDOKU-Run3	0.592	0.512	0.549	0.644	0.612	0.627
SUDOKU-Run1	0.601	0.490	0.540	0.501	0.488	0.494
LIMSI	0.535	0.440	0.483	0.694	0.608	0.648
EBL-Hope	0.525	0.446	0.482	0.490	0.429	0.457
Babelfy	0.586	0.427	0.493	0.642	0.574	0.606

Still, many of these evaluations use different texts in different languages with the goal of comparing across systems; our emphasis is rather on understanding how systems perform across languages. Hence, to facilitate such comparison, we wish to perform evaluation over the same text in multiple languages.

For this reason, our experiments are based on the SemEval 2015 Task 13 [32] dataset, which is divided into four documents with the same content in English, Italian and Spanish. In total, there are 137 sentences. For the moment, we focus on the English and Spanish languages. The goal is then to perform linking to BabelNet. In fact, a number of tools responded to the call for Task 13, and have reported results in these languages. To validate our evaluation process, we first reevaluate the annotations performed by the participants shown in Table 2 for which we could locate source code. In all the cases, we obtain the same results as reported in the contests except in the case of SUDOKU-Run1 for English, which was scored with 0.534 in SemEval 2015 Task 13, versus our result of 0.494. We also include in Table 2 some new results for Babelfy, which is the only other approach that links to BabelNet entries; in terms of F_1, the system falls behind the SUDOKU configurations but tends to fare better than other systems. We also note that with the exception of SUDOKU-Run1 and EBL-Hope, systems perform better for English than Spanish (and sometimes markedly so).

Systems: We now wish to extend the systems for which results are available on the selected SemEval 2015 Task 13 dataset. Given the wealth of EL systems proposed, in order to facilitate testing, we select systems based on four criteria: (1) details of the system must be published; (2) a public demo or API must be available for the system; (3) the system must be a complete EL system including both ER and ED phases; (4) the system must perform linking to Wikipedia or a related resource, such as DBpedia or YAGO. Hence, from the multilingual EL approaches selected in Table 1, these criteria mean we will test with THD, DBpedia Spotlight, TAGME, Babelfy and WikiME. KIM is excluded since it does not link to a Wikipedia resource; AGDISTIS is excluded since their APIs do not perform ER; other systems are excluded for not having a demo/API.

One may note that Spanish is not listed for THD and TAGME. We are still interested to see to what extent having explicit multilingual support is really

Table 3. ER-level evaluation of selected approaches for the SemEval 2015 Task 13 in Spanish and English. Approaches configured for Spanish are italicized.

	Spanish			English			
System	**P**	**R**	**F1**	**P**	**R**	**F1**	%
Sentence level							
Babelfy	0.727	0.540	0.620	0.820	0.644	0.721	85.99
DBpedia-Spotlight	0.298	0.607	0.400	0.556	0.554	0.555	72.07
WikiMe	0.737	0.018	0.036	0.656	0.028	0.053	67.92
TAGME	0.240	0.319	0.274	0.583	0.687	0.631	43.42
THD	0.281	0.061	0.100	0.587	0.080	0.142	70.42
AIDA	0.750	0.008	0.015	0.688	0.029	0.057	26.32
Document level							
Babelfy	0.765	0.581	0.661	0.864	0.704	0.776	85.18
DBpedia-Spotlight	0.300	0.612	0.403	0.555	0.549	0.552	73.01
WikiMe	0.783	0.023	0.045	0.621	0.024	0.046	97.83
TAGME	0.256	0.255	0.255	0.557	0.551	0.554	46.02
THD	0.277	0.060	0.098	0.587	0.080	0.142	69.01
AIDA	0.857	0.008	0.016	0.667	0.026	0.051	31.37

important for EL systems, and to compare systems that allow for explicitly selecting a given language such as Spanish versus those that do not allow for selecting a language and thus presume (e.g.) English text. We may consider, e.g., that *Michael Jackson* or *Chile* would be recognized/disambiguated by both systems, while *Irlanda* might not be recognized/disambigated by systems not configured for Spanish [33]. For the purposes of comparison, we thus test not only the THD and TAGME systems – multilingual systems without explicit support for Spanish – but also the AIDA system [3] – a monolingual system that does not allow for selecting a language, but that meets the other criteria.

Hence the final list of systems selected for evaluation are: *configurable for Spanish*: Babelfy, DBpedia Spotlight and WikiMe; *multilingual but not configurable for Spanish*: TAGME and THD; *monolingual*: AIDA. All systems are run with default configurations, except DBpedia Spotlight, which does not directly suggest defaults; we configured the system with *support* equal to 0 and *confidence* equal to 0.25 based on some initial experiments.

Results & Discussion: We evaluate approaches separately for ER and EL phases and for sentence-level and document-level texts. The evaluation results for ER and EL phases are presented in Table 3 and Table 4 respectively. Note that for quick reference, the % column presents the ratio of the F_1 measure for Spanish vs. English, directly comparing the performance for both languages. In constrast to the evaluation shown in Table 2 where we only take into account annotations over BabelNet, to facilitate comparison across all systems, these latter experiments only include annotations over Wikipedia, DBpedia and YAGO.

Table 4. Overall EL evaluation of selected approaches for the SemEval 2015 Task 13 in Spanish and English. Approaches configured for Spanish are italicized.

System	Spanish			English			
	P	**R**	**F1**	**P**	**R**	**F1**	%
Sentence level							
Babelfy	0.599	0.324	0.420	0.725	0.467	0.568	73.94
DBpedia-Spotlight	0.482	0.293	0.364	0.581	0.322	0.414	87.92
WikiMe	0.929	0.017	0.033	0.952	0.026	0.051	64.71
TAGME	0.371	0.118	0.179	0.568	0.391	0.463	38.66
THD	0.596	0.036	0.069	0.738	0.059	0.120	57.50
AIDA	0.667	0.005	0.010	0.773	0.022	0.044	22.73
Document level							
Babelfy	0.597	0.347	0.439	0.729	0.513	0.602	72.92
DBpedia-Spotlight	0.444	0.272	0.337	0.584	0.321	0.414	81.40
WikiMe	0.944	0.022	0.043	0.944	0.022	0.043	100.00
TAGME	0.327	0.083	0.133	0.555	0.306	0.395	33.67
THD	0.609	0.036	0.069	0.738	0.059	0.110	62.73
AIDA	0.667	0.005	0.010	0.900	0.023	0.046	21.74

From both tables, we can see that results for both EL and ER can vary significantly for Spanish and English, even for systems configurable for both languages. However, in general, those systems configurable for Spanish experienced much less of a gap across both languages when compared with the analogous results for systems not configured for that language.

The gap between Spanish and English performance is hardly surprising for tools not configured for Spanish: TAGME is based on the analysis of anchor text of the English Wikipedia pages; THD selects candidates using the Search API of English Wikipedia; AIDA is based on an English part-of-speech tagger. Clearly these approaches will not perform well for Spanish. The language gap for THD is not so pronounced; however, the F_1 scores in general are quite low, making it hard to draw conclusions: the performance for both languages is quite poor. In summary, such systems are likely to only be able to recognize/link entities that are also "valid" in English, such as *Michael Jackson* or *Chile*.

What is perhaps more interesting then, is the consistent gap between both languages for the three systems specifically configured for those languages. In particular, we propose that this result may be due to one (or more) of the following issues faced by multilingual systems:

– *The knowledge base contains different information for both languages.* In Wikipedia anyone can create or edit articles, but this is done separately for each language; thus, equivalent pages in both languages store different content; e.g., even though the label *Michael Jackson* does not change across languages, the content and links in the Spanish and English edition of Wikipedia involving that entity will change. Thus, the use of different

editions of Wikipedia to handle multilingual EL can introduce a gap in the performance for both languages. This issue may in particular affect DBpedia Spotlight, which performs the ranking stage of ER based on the occurrence of the text anchors for each specific-language Wikipedia pages. On the other hand, the EL model of WikiMe uses a transliteration model to avoid this issue. Likewise, Babelfy should not be as affected by this issue since BabelNet includes a Machine Translation process in its construction.
- *The models/techniques changes according to the target language.* Although using language-specific components will improve results for that specific language, it can also introduce another possible gap when considering performance across languages. For example, DBpedia-Spotlight's ER could be affected by this issue since the model to perform ER is selected according to the targeted language, where some models may be better than others; for example, for English and German, they use OpenNLP models, whereas for Dutch, they used a corpus of manually corrected entities. As another example, Babelfy bases the detection of candidate mentions during the ER phase on part-of-speech tagging, which requires language-specific knowledge, but such components may vary in performance across languages.
- *Variations in the languages themselves.* We must also consider that some languages are inherently more difficult for an EL process than others. As a simple example, many tools rely on capitalization as a feature for detecting entities, where Spanish tends to use less capitalization than English, including for months, languages, religions, personal titles, and titles of works. Likewise some tools consider a fixed-length window of words/tokens as potential candidate mentions, as well as simple noun phrases, whereas Spanish works tend to have longer titles with non-noun tokens, especially when translated from English (e.g., combining both issues, *Star Wars* translates as *La guerra de las galaxias*, which would be *much* more challenging for ER to recognize).

Due to such issues, even the approaches configured for Spanish do not perform as well as for English. Only in the EL/document-level experiment does WikiMe perform equivalently for Spanish and English, though it should be noted that again, the F_1 measure is quite low in both cases (due to low recall).

Summarizing other aspects of the experiments, in general, there are no substantial differences between the performance of the approaches for the document-level and sentence-level experiments (even though systems such as TAGME are specifically designed for short text collections). The gap between languages is not specifically a factor of precision or recall: the gap is implicit in both aspects of performance. The best system for both the ER and EL stages and for both the English and Spanish languages is consistently Babelfy.

5 Conclusion

There are a great many EL approaches in the literature, some of which support a variety of languages. In this work, we provide a short survey of the main

multilingual approaches found in the literature. We then performed experiments to compare the performance of a selection of EL approaches for equivalent texts in Spanish and English. We found that almost all approaches performed worse for Spanish than for English. This gap in performance was most pronounced for systems that did not support Spanish. However, even amongst those that do, the gap was quite significant. We proposed some potential explanations for this observed gap in multilingual settings.

Of course, these results currently involve one dataset, two languages, and a subset of systems one could consider, and hence should be considered as preliminary (but still we feel informative). In future work we plan to extend our experiments to consider more systems, more languages and more datasets to better understand state-of-the-art EL performance in multilingual settings. Likewise it would be interesting to perform experiments to specifically test our hypotheses as to why this gap between English and Spanish performance is observed.

Acknowledgements The work of Henry Rosales-Méndez was supported by CONICYT-PCHA/Doctorado Nacional/2016-21160017. The work was also supported by the Millennium Nucleus Center for Semantic Web Research under Grant NC120004.

References

1. Ferragina, P., Scaiella, U.: Tagme: on-the-fly annotation of short text fragments (by Wikipedia entities). In: CIKM, ACM (2010) 1625–1628
2. Mendes, P.N., Jakob, M., García-Silva, A., Bizer, C.: DBpedia spotlight: shedding light on the web of documents. In: I-SEMANTICS, ACM (2011) 1–8
3. 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: EMNLP, ACL (2011) 782–792
4. Södergren, A., Klang, M., Nugues, P.: Linking, searching, and visualizing entities for the Swedish Wikipedia. In: SLTC. (2016)
5. Van, D.K., Huynh, H.M., Nguyen, H.T., Vo, V.T.: Entity linking for Vietnamese Tweets. In: Knowledge and Systems Engineering. Springer (2015) 603–615
6. Xu, J., Gan, L., Zhou, B., Wu, Q.: An unsupervised method for linking entity mentions in Chinese text. In: APSCC, Springer (2016) 183–195
7. Nebhi, K.: Named entity disambiguation using Freebase and syntactic parsing. In: LD4IE, CEUR-WS. org (2013) 50–55
8. Daiber, J., Jakob, M., Hokamp, C., Mendes, P.N.: Improving efficiency and accuracy in multilingual entity extraction. In: I-SEMANTICS, ACM (2013) 121–124
9. Charton, E., Gagnon, M., Ozell, B.: Automatic semantic web annotation of named entities. In: Canadian Conference on Artificial Intelligence, Springer (2011) 74–85
10. Dojchinovski, M., Kliegr, T.: Recognizing, classifying and linking entities with Wikipedia and DBpedia. WIKT (2012) 41–44
11. Popov, B., Kiryakov, A., Ognyanoff, D., Manov, D., Kirilov, A.: KIM–a semantic platform for information extraction and retrieval. Natural Language Engineering **10**(3-4) (2004) 375–392
12. Usbeck, R., Ngomo, A.C.N., Röder, M., Gerber, D., Coelho, S.A., Auer, S., Both, A.: AGDISTIS-graph-based disambiguation of named entities using linked data. In: ISWC, Springer (2014) 457–471

13. Moro, A., Raganato, A., Navigli, R.: Entity linking meets word sense disambiguation: a unified approach. Trans. of the ACL **2** (2014) 231–244
14. Guo, Z., Xu, Y., de Sá Mesquita, F., Barbosa, D., Kondrak, G.: ualberta at TAC-KBP 2012: English and cross-lingual entity linking. In: TAC. (2012)
15. Pappu, A., Blanco, R., Mehdad, Y., Stent, A., Thadani, K.: Lightweight multilingual entity extraction and linking. In: WSDM, ACM (2017) 365–374
16. Plu, J., Rizzo, G., Troncy, R.: Enhancing entity linking by combining NER models. In: Semantic Web Evaluation Challenge, Springer (2016) 17–32
17. Ratinov, L., Roth, D., Downey, D., Anderson, M.: Local and global algorithms for disambiguation to Wikipedia. In: NAACL-HLT. (2011) 1375–1384
18. Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., Van Kleef, P., Auer, S., et al.: DBpedia–a large-scale, multilingual knowledge base extracted from Wikipedia. Semantic Web **6**(2) (2015) 167–195
19. Luo, G., Huang, X., Lin, C.Y., Nie, Z.: Joint named entity recognition and disambiguation. In: EMNLP. (2015)
20. Nguyen, D.B., Theobald, M., Weikum, G.: J-NERD: Joint named entity recognition and disambiguation with rich linguistic features. TACL **4** (2016) 215–229
21. Dalton, J., Dietz, L.: A neighborhood relevance model for entity linking. In: Open Research Areas in Information Retrieval. (2013) 149–156
22. Vrandečić, D., Krötzsch, M.: Wikidata: a free collaborative knowledgebase. Communications of the ACM **57**(10) (2014) 78–85
23. Suchanek, F.M., Kasneci, G., Weikum, G.: Yago: a core of semantic knowledge. In: WWW, ACM (2007) 697–706
24. Wang, Z., Li, J., Tang, J.: Boosting cross-lingual knowledge linking via concept annotation. In: IJCAI. (2013) 2733–2739
25. Fahrni, A., Göckel, T., Strube, M.: HITS' monolingual and cross-lingual entity linking system at TAC 2012: A joint approach. In: TAC, Citeseer (2012)
26. Navigli, R., Ponzetto, S.P.: BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Artificial Intelligence **193** (2012) 217–250
27. Heng, J., Grishman, R. Dang, H.: Overview of the TAC2011 knowledge base population track. In: TAC. (2011)
28. Narducci, F., Palmonari, M., Semeraro, G.: Cross-language semantic matching for discovering links to e-gov services in the LOD Cloud. KNOW@ LOD **992** (2013) 21–32
29. Tsai, C.T., Roth, D.: Cross-lingual wikification using multilingual embeddings. In: NAACL-HLT. (2016) 589–598
30. Tong, H., Faloutsos, C., Pan, J.Y.: Random walk with restart: fast solutions and applications. Knowledge and Information Systems **14**(3) (2008) 327–346
31. Usbeck R., Röder M., Ngomo A.N., Baron C., Both A., Brümmer 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., Wesemann L.: GERBIL: general entity annotator benchmarking framework. In: WWW. (2015) 1133–1143
32. Moro, A., Navigli, R.: SemEval-2015 Task 13: Multilingual all-words sense disambiguation and entity linking. In: SemEval@ NAACL-HLT. (2015) 288–297
33. Nadeau, D., Sekine, S.: A survey of named entity recognition and classification. Lingvisticae Investigationes **30**(1) (2007) 3–26