Machine Translation vs Multilingual Entity Linking

Henry Rosales-Méndez, Aidan Hogan and Barbara Poblete

Millenium Institute for Foundational Research on Data
Department of Computer Science, University of Chile
{hrosales,ahogan,bpoblete}@dcc.uchile.cl

Abstract. Entity Linking (EL) associates the entities mentioned in a given input text with their corresponding knowledge-base (KB) entries. There are several approaches that compute EL, but the majority of them deal with only one language. A recent trend is proposing multilingual approaches that incorporate more effort in order to handle several languages, while other authors address this problem taking advantage of the advances obtained in the machine translation. However, studies comparing both techniques are scarce in the literature, and we do not know with certainty which of both gets the best results. This work is an intent to address this gap in an experimental way, comparing the behavior of state-of-the-art approaches in both scenarios. The results show that both techniques obtain better results than the other depending on the system used and the dataset.

1   Introduction

Entity Linking (EL) is a task of Information Extraction that associates the entities mentioned in a given input text with their corresponding knowledge-base (KB) entries. For Instance, targeting Wikidata, we can process the text "*Michael Jackson was born in Gary, Indiana*" and highlight that the substring *Michael Jackson* is an entity mention that should be associated with the identifier `wd:Q2831`[1]. With this association (or annotation) we know that the text is about the US singer known as "King of Pop". However, there are several obstacles that make this task a challenge, one obstacle is that one entity mention can refer to several identifiers, for instance, `wd:Q2831` is not the only person in Wikidata that have the name Michael Jackson (e.g., `wd:Q167877`, `wd:Q6831554`, `wd:Q3856193`), so it is necessary to take into account the context of each entity mention to carry out the disambiguation process. On the other hand, the same entity can be referenced by more than one entity mention, for instance, either with "*Michael Jackson*", "*Michael J. Jackson*", or "*Jackson*" can refer to `wd:Q2831`.

This task has gained the attention of many researchers in the last decades due to current KBs have achieved a considerable amount of information (e.g., Wikipedia) and others have been freely published – and proposed – in different

---

[1] Throughout, we use prefixes according to `http://prefix.cc`

domains of knowledge (e.g., DBpedia, Wikidata, Freebase and BabelNet[2]). With EL we can take advantage of the valuable information contained by these KBs about real-life entities in many application scenarios such as semantic search, text enrichment, entity summarization, relation extraction and others.

The current state of the art of EL include several approaches, however the most of the cases only consider text input in English. A major challenge is that the same system can be configured so that the input text can be written not only in English but in other languages as well. Generally, multilingual EL approaches base their models in exploiting multilingual resources. For Instance, Babelfy [10] exploits the multilingual knowledge of BalbelNet to achieve the multilingual annotations, while WikiME [3] takes advantage of the different versions of Wikipedia written in different languages to achieve this same goal.

However, proposing a language-agnostic model to compute EL requires much more effort than a model that only deal with one single language. For this reason, some few authors [11,1] have been employed in machine translation to deal with several languages while the model support only one. In this way, to process a text written in an unsupported language, the automatic translation on the input text is applied to obtain a text written in a known language, and thus proceed according to the monolingual model. But is this really more efficient? So far, there is little research on machine translation in EL, and those, do not study beyond performance itself. In this work, we conduct some experiment in order to address this question. In our previous work [16], we propose a multilingual dataset benchmark and we perform an experiment in order to study the behavior of state-of-the-art approaches including machine translation only from English. Here, we complete our previous study incorporating machine translation from all the five languages supported by VoxEL.

2 Preliminaries

For a given Text input with a set of entities mentioned M, and a KB with a set of identifiers E, EL associate (or link) each $m \in M$ with a corresponding entity $e \in E$. This EL process is commonly divided into two main phases: Entity Recognition (ER) and Entity Disambiguation (ED). The first one has the goal of identifying the entities mentions in the text, process often performed by Name Entity Recognition (NER) tools such as Standard NER, Part-of-Speech tagger and other. These identified entity mentions are passed to the second phase where each of them is disambiguated following an End-to-End scheme. The ED phase also is performed commonly in three stages, described next:

Candidate entity generation: Once the mention of entity is identified it must be associated with an identifier in a KB, but making complex methods that consider all the entities of the targeted KB is not appropriate due to its large sizes. For this reason, the goal of this stage is to reduce that search space with simpler techniques such as keyword machine or a dictionary-based approach.

---

[2] `https://www.babelnet.org/`; June 1st, 2018.

Formally, this stage selects the set $E_m \subseteq E$ for each mention $m \in M$, where the KB entities in E_m are those that are most likely to be the corresponding to m.

Candidate entity ranking: This stage has the goal of ranking the KB identifiers of E_m for each entity mention m according to the probability that each $e \in E_m$ has to be the corresponding to it. Thus, the top-ranked identifiers are considered to be the target link in the annotations. This process is addressed by many techniques such as graph mining, string similarity, machine learning, function optimization and others.

Unlinkable mention prediction: Generally, it is expected that the targeted KB are large enough to link all the entity mentions, however this does not happen in reality. This stage looking for those entity mentions that does not have a corresponding entry in the target KB, and label them with NIL (Not In Lexicon). These are sometimes referred to as *emerging entities*, or *out-of-Wikipedia* when Wikipedia is the target KB.

Some authors conduct both phases at the same time following a joint scheme, commonly, based on the optimization of functions that simultaneously measure the feasibility of recognizing and annotating an entity mention. Others, only take into account the disambiguation process constrained the input text with tags that indicate where the entity mentions are.

3 Entity Linking and multilingualisms

Nowadays, there are several EL approaches in the literature, but only few are multilingual. In Table 1 we show an overview of those systems that have experimental results over more than one language. As we can expect, English is the language that is more supported by them, followed by the European ones. The majority of these approaches include in the same model the procedure to deal with different languages. Moro et al. propose Babelfy[13] a graph-based model that using the multilingual KB BabelNet, and they apply a densest subgraph heuristic to capture the most coherent interpretation of the entity mentions. THD uses the entity mentions as keywords and leverage of the Wikipedia Search API to obtain the more likely Wikipedia pages that describe them, taking into account properly version of Wikipedia that corresponds to the language of the input text. On the other hand, WikiME use word embedding and its projects across language to deal with multilingualism. For more explanations about these systems see [16].

Instead of adding complexity to monolingual models in order to deal with a multilingual environment, others authors take advantage of machine translation to deal with several languages. For instance, Monahan et al. [11] use Bing API[3] to translate the input text from Chinese to English, while Cassidy et al. [1] apply EL over Chinese using coreference resolution, various name translation techniques over the entity mentions to English, and machine translation to English. The use

---

[3] `http://www.microsofttranslator.com/dev/`; June 1st, 2018.

Table 1. Overview of multilingual EL approaches; the italicized approaches will be incorporated as part of our experiments. (Abbreviations: MT if the approach include machine translation)

Name	Year	Evaluated Languages	MT	API
KIM [15]	2004	English, French, Spanish		✓
TagME [7]	2010	English, German, Dutch		✓
SDA [2]	2011	English, French		
Monahan et al. [11]	2011	English, Chinese	✓	
ualberta [8]	2012	English, Chinese		
HITS [6]	2012	English, Spanish, Chinese		
THD [5]	2012	English, German, Dutch		✓
Cassidy et al. [1]	2012	English, Chinese	✓	
DBpedia Spotlight [4]	2013	English, Italian, Russian, Dutch, French, German, Spanish, Hungarian, Danish		✓
Wang-Tang [22]	2013	English, Chinese		
AGDISTIS [12]	2014	English, German, Spanish French, Italian, Japanese, Dutch		✓
Babelfy [13]	2014	English, Spanish, French German, Italian		✓
WikiME [20]	2016	English, Spanish, French, Italian, Chinese, German, Thai, Arabic, Turkish, Tamil, Tagalog, Urdu, Hebrew		
FREME [19]	2016	English, German		✓
FEL [14]	2017	English, Spanish, Chinese		

of machine translation brings advantages. In this way, monolingual systems deal with all the languages supported by the translator tool, which usually include the most popular languages in the world. Another advantage is that it is not necessary to deal with transliterations or cross-language links in cases when the target KB is available in a different language than the input text.

On the other hand, to evaluate EL systems we require benchmark datasets that allow us to study the behavior of those systems on multilingual scenarios of applications. So far, the majority of the datasets only concern English and others are proposed in the scope of conferences such as TAC KBP[4] and SemEval[5] that

---

[4] https://tac.nist.gov/2017/KBP/; June 1st, 2018.
[5] http://alt.qcri.org/semeval2018/; June 1st, 2018.

restring their availability. Only a few multilingual datasets are freely available, and that allows us the performing of Entity Linking evaluations in more than one language. One of them is MEANTIME [9], that consist on 120 English documents from WikiNews[6] automatically translated to other three languages, and also automatically annotated. Other multilingual dataset is DBpedia Abstracts that was recently created over the abstract of Wikipedia and taking the anchor text as annotations. For a more detailed explanation of benchmark datasets see [16].

Different to MANTIME and DBpedia Abstracts, in [16] was proposed VoxEL[7], which was manually annotated over a curated text extracted from VoxEurop[8]. This dataset contains 15 documents for each of the five supported languages: Germany, English, Spanish, French and Italian. Despite that the documents are news translated by human experts, in some cases, there are missing entities cross-language or small changes. For this reason, the document VoxEL was aligned to each other cross-language, adding missing entities and erasing the sentences that represent big changes. For Instance, it is very usual for the human expert the employment of pronouns instead of entity mentions cross-language in order to achieve a coherent translation. In this way, VoxEL contains the same number of sentences for each document cross-language – 94 sentences –, and in turn, the same number of annotations per sentence.

Another desirable property of VoxEL is that take into account that there are in the literature different point of view about what should entity linking link, see [18] for more details. Along these lines, VoxEL has two annotated version of the documents, one *strict* that include entities of person, places and organization, and other *relaxed* that include all the unambiguous pages of Wikipedia. In summary, VoxEL contains 237 and 635 sentences in the strict and relaxed version respectively.

4 Experiments

We conduct two experiments in order to compare the behavior of multilingual EL systems and their performance using machine translation. We select Babelfy, DBpedia Spotlight, FREME and TagME due to are the only ones that provide an available API and also target Wikipedia or DBpedia KBs. Despite THD fulfill these criteria, we leave it out of our selection because the API was not working at the time of the experimentations because the developer team were in the process of updating the server infrastructure.

As we mentioned in the previous section, there are some multilingual datasets that allow us to evaluate the quality of EL systems in multilingual environments. We select to perform our experiment the VoxEL datasets because was manually annotated over curated text and also aligned over the version to achieve the same number of sentences and annotations, guaranteeing the same scenario for each different language. Other works have been proposed to compare the performance

---

[6] `https://en.wikinews.org/`; June 1st, 2018.

[7] `https://users.dcc.uchile.cl/~hrosales/VoxEL.html`; June 1st, 2018.

[8] `http://www.voxeurop.eu`; June 1st, 2018.

of systems in several languages but generally on datasets that have different content or annotations between the different versions of the dataset [17].

Our first experiment consists of applying each system over the five version of VoxEL, which corresponds to each of the languages: Germany (DE), English (EN), Spanish (ES), French (FR) and Italian (IT). All the system were configured with their default parameters, except in the case of Babelfy. The system Babelfy handle a similar concept of VoxEL including parameters to specify if we expect to target only strict entities, or include other more relaxed. For this reason, we will study the performance of Babelfy with both strict and relaxed settings, detonate henceforth as Babelfy$_S$ and Babelfy$_R$ respectively.

The second experiment include the same previous systems, but this time we apply machine translation – Google Translator[9] – from each language of VoxEL to the other four. Each system is configured, either with the language of the entry text if there is no translation, or with the language of the translated text, as appropriate. Despite that there are some popular benchmark systems such as GERBIL [21] that provide a variate of facilities to compare the results of EL systems, we made our own evaluator script that fit the new problems that introduce the machine translation. For instance, with machine translation, we lose the positions of each label annotated in the text because in the translations the entity mentions can appear in different part of the translated text or even disappear. Therefore, some benchmarks systems like GERBIL does not work under this condition.

We show in Table 2 the result of both experiments. In the boxes are the results that are expected to be the best, because they are obtained by each language trough applying each system settings in that language, over the original version of VoxEL written in this same language. All the non-boxes results correspond to the applications of those systems over the translated version of VoxEL. We stress with bold the best result for each pair of system and language that represent the rows, split by Relax/Strict. With a simple inspection of this table, we can observe that there are not the best technique according to this experiment, rather than, some approaches have better value with machine translation and others do not. For Instance, for all the languages in both versions of the datasets $Babelfy_S$ and $Babeldy_R$ overcome their own performance. On the other hand, although for some language the rest of the systems obtain better punctuation with the automatic translation, in the majority of the cases they do not surpass their own result. For point of view of the individual languages, we can observe a similar behavior, where six of the ten columns of results indicate that machine translation is equal or better than without it.

5   Conclusions

In this work, we perform two experiments in order to study the behavior of state-of-the-art EL approaches that support more than one language, with the

---

[9] `https://translate.google.com`; June 1st, 2018.

Table 2. Evaluation of EL systems with F_1 measure. A value "–" indicates that the system does not support the corresponding language. The boxed results are those obtained for each system from the dataset without translation, while those results in bold are the best for that system comparing across the five languages (the best in each row, split by Relax/Strict). Each column $L \to_{-t}$ is corresponding to the results of the current systems over the dataset written in language L but translated to the language of the current row, split by Relax/Strict.

		Relaxed					Strict				
		DE→_t	EN→_t	ES→_t	FR→_t	IT→_t	DE→_t	EN→_t	ES→_t	FR→_t	IT→_t
Babelfy_R	DE	0.523	0.498	0.495	0.492	0.490	0.344	0.342	0.365	**0.369**	0.367
	EN	0.508	0.545	0.515	0.506	0.502	0.299	0.319	0.299	0.315	0.301
	ES	0.525	**0.558**	0.541	0.552	0.548	0.344	0.356	0.362	0.357	0.348
	FR	0.493	0.485	**0.502**	0.493	0.493	0.332	0.331	**0.342**	0.309	0.341
	IT	0.513	0.527	0.512	**0.533**	0.504	0.366	**0.379**	0.377	0.378	0.365
Babelfy_S	DE	0.279	0.271	0.275	**0.285**	0.273	0.572	0.584	0.589	**0.606**	0.588
	EN	0.312	0.308	0.309	**0.323**	0.304	0.518	0.567	0.523	0.559	0.533
	ES	0.318	0.327	0.325	0.334	**0.336**	0.577	0.607	0.611	0.610	0.590
	FR	0.301	0.299	**0.312**	0.290	0.310	0.574	0.601	**0.608**	0.583	0.606
	IT	0.306	0.319	0.318	**0.321**	0.311	0.604	0.634	**0.640**	0.638	0.616
DBpedia Spotlight	DE	**0.400**	0.139	0.177	0.155	0.166	**0.510**	0.220	0.292	0.248	0.280
	EN	0.442	**0.466**	0.454	0.465	0.450	0.697	0.707	0.695	0.722	**0.730**
	ES	0.159	0.121	**0.373**	0.130	0.199	0.292	0.209	**0.513**	0.234	0.350
	FR	0.176	0.177	0.181	**0.314**	0.180	0.245	0.252	0.252	**0.464**	0.255
	IT	0.184	0.163	0.221	0.158	**0.382**	0.272	0.219	0.335	0.223	**0.601**
FREME	DE	**0.282**	0.072	0.132	0.114	0.160	**0.483**	0.154	0.240	0.179	0.261
	EN	0.401	**0.407**	0.402	0.397	0.406	0.700	0.708	**0.715**	0.694	0.713
	ES	0.174	0.117	**0.302**	0.147	0.232	0.319	0.231	**0.583**	0.269	0.417
	FR	0.167	0.143	0.169	**0.268**	0.214	0.287	0.278	0.314	**0.483**	0.322
	IT	0.164	0.127	0.205	0.136	**0.373**	0.321	0.253	0.413	0.256	**0.726**
TagME	DE	**0.414**	0.100	0.127	0.119	0.124	**0.272**	0.122	0.153	0.137	0.152
	EN	0.432	**0.462**	0.450	0.442	0.440	0.331	0.327	0.334	0.321	**0.336**
	ES	-	-	-	-	-	-	-	-	-	-
	FR	-	-	-	-	-	-	-	-	-	-
	IT	-	-	-	-	-	-	-	-	-	-

goal of practically see if we must or not include machine translation in this scenario. The first experiment settings the systems to deal with multilingual datasets according their own model and the second one perform each system over automatically translated text. The results show that there is not a a best way because of both techniques of computing EL overcome the other in different cases. We observed that systems such as Babelfy respond well with the use of machine translation while others like FREME do not.

In future work, we will conduct deeper experiments in order to study the behavior of such systems, but a more reduced scenario where we can measure the suitability of applying machine translation independence of the entity type. Also, we expect to explain to details why each of these systems behaves such way.

References

1. Cassidy, T., Ji, H., Deng, H., Zheng, J., and Han, J. Analysis and refinement of cross-lingual entity linking. In CLEF (2012) 1–12
2. Charton, E., Gagnon, M., Ozell, B.: Automatic semantic web annotation of named entities. In: Canadian Conference on Artificial Intelligence, Springer (2011) 74–85
3. Chen-Tse Tsai and Dan Roth. Cross-lingual wikification using multi-lingual embeddings. In Proceedings of NAACL-HLT, pages 589–598, 2016.
4. Daiber, J., Jakob, M., Hokamp, C., Mendes, P.N.: Improving efficiency and accuracy in multilingual entity extraction. In: I-SEMANTICS, ACM (2013) 121–124
5. Dojchinovski, M., Kliegr, T.: Recognizing, classifying and linking entities with Wikipedia and DBpedia. WIKT (2012) 41–44
6. Fahrni, A., Göckel, T., Strube, M.: HITS' monolingual and cross-lingual entity linking system at TAC 2012: A joint approach. In: TAC, Citeseer (2012)
7. Ferragina, P., Scaiella, U.: Tagme: on-the-fly annotation of short text fragments (by Wikipedia entities). In: CIKM, ACM (2010) 1625–1628
8. 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)
9. Minard, A. L., Speranza, M., Urizar, R., Altuna, B., van Erp, M. G. J., Schoen, A. M., van Son, C. M. MEANTIME, the NewsReader multilingual event and time corpus. LREC-ELRA (2016)
10. Moro, A., Raganato, A., Navigli, R. Entity linking meets word sense disambiguation: a unified approach. TACL **2** (2014) 231–244
11. Monahan, S., Lehmann, J., Nyberg, T., Plymale, J., and Jung, A. Cross-Lingual Cross-Document Coreference with Entity Linking. In TAC (2011)
12. Moussallem, D., Usbeck, R., Röeder, M., Ngomo, A. C. N. MAG: A Multilingual, Knowledge-base Agnostic and Deterministic Entity Linking Approach. K-CAP 2017 , ACM, (2017) 9:1–9:8
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. Pappu, A., Blanco, R., Mehdad, Y., Stent, A., Thadani, K.: Lightweight multilingual entity extraction and linking. In: WSDM, ACM (2017) 365–374
15. 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
16. Rosales-Méndez, H., Hogan A., Poblete B. VoxEL: A Benchmark Dataset for Multilingual Entity Linking. In ISWC (2018) (to appear)
17. Rosales-Méndez, H., Poblete, B., and Hogan, A. Multilingual Entity Linking: Comparing English and Spanish. In LD4IE@ISWC (2017)
18. Rosales-Méndez, H, Poblete, B., and Hogan, A. What should Entity Linking link? In AMW (2018)
19. Sasaki, F., Dojchinovski, M., Nehring, J. Chainable and Extendable Knowledge Integration Web Services. In ISWC, (2016) 89–101
20. Tsai, C.T., Roth, D.: Cross-lingual wikification using multilingual embeddings. In: NAACL-HLT. (2016) 589–598

21. Usbeck, R. et al. GERBIL – General Entity Annotation Benchmark Framework. In 24th WWW conference, 2015
22. Wang, Z., Li, J., Tang, J.: Boosting cross-lingual knowledge linking via concept annotation. In: IJCAI. (2013) 2733–2739