

Machine Translation vs. Multilingual Approaches for Entity Linking

Henry Rosales-Méndez, Aidan Hogan and Barbara Poblete

IMFD Chile & Department of Computer Science, University of Chile

Abstract. Entity Linking (EL) associates the entities mentioned in a given input text with their corresponding knowledge-base (KB) entries. A recent EL trend is towards multilingual approaches. However, one may ask: are multilingual EL approaches necessary with recent advancements in machine translation? Could we not simply focus on supporting one language in the EL system and translate the input text to that language? We present experiments along these lines comparing multilingual EL systems with their results over machine translated text.

1 Introduction

Entity Linking (EL) associates the entities mentioned in a given input text with their corresponding knowledge-base (KB) identifiers; e.g., taking Wikidata as a target KB, for the input text “*Michael Jackson was born in Gary, Indiana*”, we can link *Michael Jackson* with the Wikidata identifier `wd:Q2831`. However, multiple KB entities may have the same label; e.g., `wd:Q167877`, `wd:Q6831554`, and `wd:Q3856193` are all identifiers for people called *Michael Jackson* in Wikidata. On the other hand, the same entity can be mentioned multiple ways, e.g., “*Michael J. Jackson*”, “*Jackson*”, “*King of Pop*”, etc., can refer to `wd:Q2831`.

Another practical challenge is being able to cope with input texts from various languages. While many EL approaches have been proposed down through the years, only recently have multilingual EL approaches – configurable for various input languages – become more popular (e.g., [2,1,4,7]). Despite this trend, there are few studies evaluating multilingual EL. Hence, in our paper accepted for the Resource Track at ISWC [5], we propose a multilingual EL benchmark and use it to perform experiments in order to study the behaviour of state-of-the-art multilingual approaches for five languages: English, French, German, Italian, and Spanish. We call our dataset VoxEL; a particular design goal of the dataset is to have (insofar as possible) the same text in different languages, and in particular, the same annotations per sentence across languages. Thus performance across languages – not just systems – can be compared directly. We also compared the results of multilingual EL systems with what would be possible using a state-of-the-art machine translation approach (Google Translate) to translate the text to English (the primary language supported by most tools). We refer the reader to our Resource Track paper [5] for more details.

In this poster, we present some additional results omitted from the full paper for reasons of space. More specifically, the poster focuses on the question of

how an *a priori* machine translation process compares with multilingual EL approaches, contributing novel results using VoxEL to evaluate EL performance using machine translation of the input to languages other than English. More generally, in the poster session, we would like to discuss with interested attendees the interplay between multilingual EL and machine translation.

2 Evaluating Multilingual Entity Linking Approaches

A multilingual EL system is characterised by being configurable for multiple input languages. In this work, we evaluate four multilingual EL systems with public APIs, namely *Babelify* [4], *DBpedia Spotlight* [1], *FREME* [7] and *TagME* [2]. For reasons of space, we refer to our previous work [6,5] for further details on these systems and other multilingual EL systems proposed in the literature.

Evaluating multilingual EL systems requires benchmark datasets with texts in various languages. To further compare the quality of EL results across languages – not just systems – we need (insofar as possible) the same text and annotations in the different languages. Only a few such datasets have been proposed: TAC KBP¹, SemEval², and MEANTIME [3]. However, MEANTIME [3] is the only publicly available dataset; SemEval is published by a third-party whereas the TAC KBP dataset we could not acquire. Furthermore, we found that these datasets exhibit differences in their annotations for different languages. For a more detailed explanation of multilingual benchmark datasets see [5] and for results comparing various EL systems over the SemEval dataset, see [6].

To support multilingual EL evaluation, in [5] we proposed VoxEL: a curated text extracted from the multilingual VoxEurop news site³ and manually annotated for EL benchmarking. This dataset contains 15 documents for each of the five supported languages: Germany, English, Spanish, French and Italian. To support comparison across languages, VoxEL was edited to ensure the same annotations per sentence across languages, normalising variances across languages. Given a lack of consensus on the definition of “entity”, VoxEL features two annotated version of the documents for each language: one *strict* that includes entities referring to people, places and organisations, and one *relaxed* that includes links to all unambiguous pages of Wikipedia. Per language, VoxEL contains 204 and 674 annotations in the strict and relaxed version respectively.

3 Experiments

We conduct experiments using VoxEL to compare the behaviour of the four aforementioned multilingual EL systems for the five different languages offered by the dataset: German (DE), English (EN), Spanish (ES), French (FR) and Italian (IT). All systems were configured with their default parameters, except

¹ <https://tac.nist.gov/2017/KBP/>; June 1st, 2018.

² <http://alt.qcri.org/semeval2018/>; June 1st, 2018.

³ <http://www.voxeurop.eu>; June 1st, 2018.

Table 1. Comparison of EL systems for native and translated texts (F_1 measure)

		Relaxed					Strict				
		DE →	EN →	ES →	FR →	IT →	DE →	EN →	ES →	FR →	IT →
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

Babelfy, which allows to select a more strict or more relaxed notion of entity; we study the performance of both, denoted henceforth as Babelfy_S and Babelfy_R respectively. Aside from testing EL over the text in its native language, we also include results for EL applying machine translation – namely Google Translate⁴ – from each language of VoxEL to the other four languages; the purpose of this approach is to simulate an EL approach supporting one language and see if EL performs competitively when input text is translated from other languages.

The results are given in Table 1, where we present the F_1 -measure for various configurations. On the left we present the system and language configured. At the top of the table we present the Relaxed and Strict versions of the dataset, where for each version, we present the language of the input text, which is machine translated to the configured language; for example, row → ES, column DE →, gives the result for a German input text translated to Spanish (DE → ES) and processed by the given EL systems configured for Spanish. Where input and translated languages coincide, we use the input text directly (such results are indicated with boxes). The best result per column for each dataset version and system is presented in bold. TagME supports English and German only.

⁴ <https://translate.google.com>; June 1st, 2018.

4 Discussion

In Table 1, we see that DBpedia Spotlight, FRENDE and TagMe often perform markedly better when the input text is either in English, or translated to English; the one exception to this trend is that FRENDE performs slightly better over the untranslated Italian text in the Strict version of the dataset than over the translated English text. On the other hand, Babelify generally performs best for (translated) Spanish texts in the Relaxed version, and (translated) Italian texts in the Strict version, though performance across languages is more balanced in general than for the former systems. These results suggest that prior machine translation makes little difference in the case of Babelify, but markedly improves the performance of other systems when dealing with non-English texts; the reasons for this may include the quality of language-specific components, the richness of KB information available for a particular language, etc.

It is important to highlight in such cases that the output of the EL process after translation is still in the translated language; e.g., if we process text in French by translating it to English and performing EL configured for English, we may get better results, but the output text is in English, not French. But we put forward that given (1) a high(er) quality annotation in the translated English text, (2) a sentence-to-sentence correspondence between the French and translated English text, and (3) cross-language links provided by KBs; it would not be difficult to “transfer” the annotations back to the original French text.

In any case, these results raise the question of what role machine translation should play for EL, and indeed, in what circumstances it makes sense to develop multilingual EL systems, and in what circumstances it makes sense to develop monolingual EL systems with *a priori* translation.

Acknowledgements Henry Rosales-Méndez was supported by CONICYT-PCHA/Doctorado Nacional/2016-21160017. The work was also supported by the Millennium Institute for Foundational Research on Data (IMFD) and by Fondecyt Grant No. 1181896.

References

1. Daiber, J., Jakob, M., Hokamp, C., Mendes, P.N.: Improving efficiency and accuracy in multilingual entity extraction. In: I-SEMANTICS, ACM (2013) 121–124
2. Ferragina, P., Scaiella, U.: Tagme: on-the-fly annotation of short text fragments (by Wikipedia entities). In: CIKM, ACM (2010) 1625–1628
3. Minard, A. L., et al. MEANTIME, the NewsReader multilingual event and time corpus. LREC-ELRA (2016)
4. Moro, A., Raganato, A., Navigli, R.: Entity linking meets word sense disambiguation: a unified approach. *Trans. of the ACL* **2** (2014) 231–244
5. Rosales-Méndez, H., Hogan A., Poblete B. VoxEL: A Benchmark Dataset for Multilingual Entity Linking. In ISWC (2018) (to appear)
6. Rosales-Méndez, H., Poblete, B., and Hogan, A. Multilingual Entity Linking: Comparing English and Spanish. In LD4IE@ISWC (2017)
7. Sasaki, F., Dojchinovski, M., Nehring, J. Chainable and Extendable Knowledge Integration Web Services. In ISWC, (2016) 89–101