Injecting Knowledge Base Information into End-to-End Joint Entity and Relation Extraction and Coreference Resolution

Severine Verlinden*, Klim Zaporojets*, Johannes Deleu, Thomas Demeester, Chris Develder

Ghent University – imec, IDLab
Ghent, Belgium

Key Ideas

1. End-to-end **document-level information extraction**.
2. Use **span-based** (Lee et al. 2017) architecture to connect each of the textual spans with **candidate entities**.
3. Use candidate entities to inject external knowledge from:
   - Knowledge graphs (Wikidata)
   - Hyperlinked knowledge bases (Wikipedia)
4. Research Wikipedia-derived **prior** and context-based **attention** schemes to weight candidate entities of each of the textual spans.

Introduction

Task: end-to-end named entity recognition (NER), relation extraction (RE), and coreference resolution.

1 Britain's Prince Harry is engaged to [..] partner Meghan Markle [..].
2 [..] the couple are to live in Kensington Palace.
3 [..] Harry's brother Prince William and Kate Middleton, congratulated the couple.
4. "We are very excited for Harry and Meghan"

Coreference Clusters: {"Meghan Markle", "Meghan"}, {"Britain"}, etc.
Entities: Britain (type:country), Harry (type:person,type:royalty,gender:male), etc.
Relations: <Kensington Palace *in* Britain>, <William *spouse_of* Kate Middleton>, etc.

Fig. 1: DWIE: some relations and entity types are not explicitly stated in the text.

Data: DWIE (Zaporojets et al. 2021) and DocRED (Yao et al. 2019).

1. **Document-level**: coreferent entity mentions spread across sentences.
2. Entity and relation annotations on cluster level (**entity-centric**).
3. Annotations (e.g., relations) are not always explicitly stated in the text: model can benefit from **external knowledge**.

Approach:

1. Inject **external knowledge** using **entity representations**.
2. Entity representations derived from **knowledge graph** (Wikidata) and from **hyperlinked textual knowledge base** (Wikipedia).
3. Explore **attention** and **prior**-based weighting of candidate entities for each of the textual spans.

Method

Main components of the proposed architecture:

1. **Text span** i: a span of text in the input document.
2. **Candidate entities** C associated to each of the spans (**EL dictionary**).
3. Wikipedia and Wikidata **KB representations** ξ of entities.
4. **Weighted combination** α of candidate entity representations.

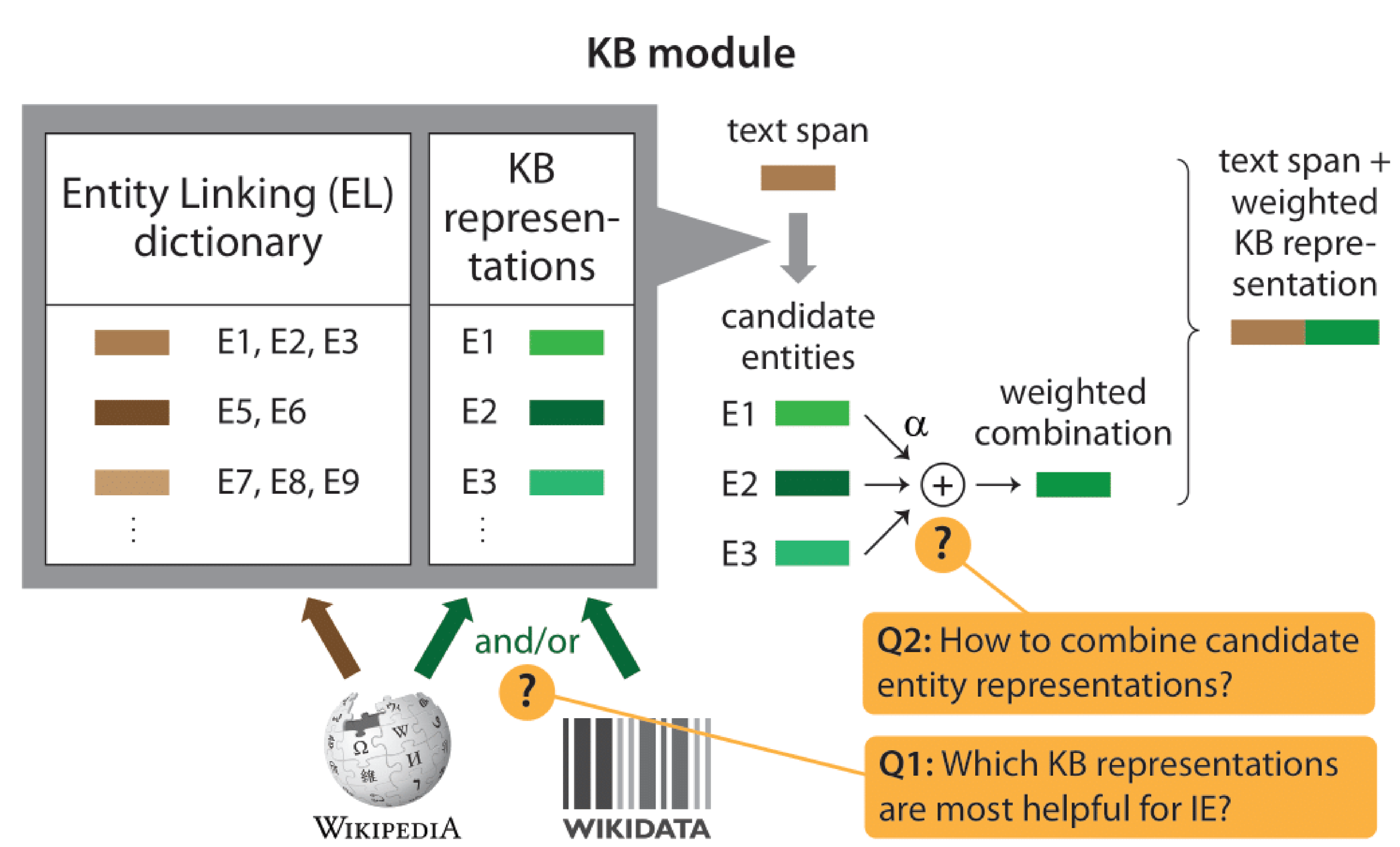


Fig. 2: Sketch of the proposed architecture.

Entity representation for span i:

$$\mathbf{e}_i^{\kappa} = \sum_{c_{ij} \in C_i} \alpha_{ij} \cdot \boldsymbol{\xi}_{\kappa}(c_{ij})$$

To answer Q1 (see Fig. 2) $\rightarrow$ sources of **external knowledge** κ:

1. Wikidata (**KB-graph**)
2. Wikipedia (**KB-text** - Yamada et al. 2016)
3. Concatenation of both (**KB-both**).

To answer Q2 (see Fig. 2) $\rightarrow$ weighted combination α for a span i:

1. **Prior** p_{ij} ($P(e_j|m_i)$ as per Yamada et al. 2016, §3): $\alpha_{ij} = p_{ij}$
2. **Uniform**: $\alpha_{ij} = 1/|C_i|$
3. **Attention**: $\alpha_{ij} = \mathcal{F}_A([\mathbf{g}_i; \boldsymbol{\xi}_{\kappa}(c_{ij})])$
4. **AttPrior**: $\alpha_{ij} = \mathcal{F}_{AP}([\mathbf{g}_i; \boldsymbol{\xi}_{\kappa}(c_{ij}); p_{ij}])$

$\mathcal{F}_*$ is a feed-forward neural network, $\mathbf{g}_i$ is the representation of span i.

Results

Ablation study:

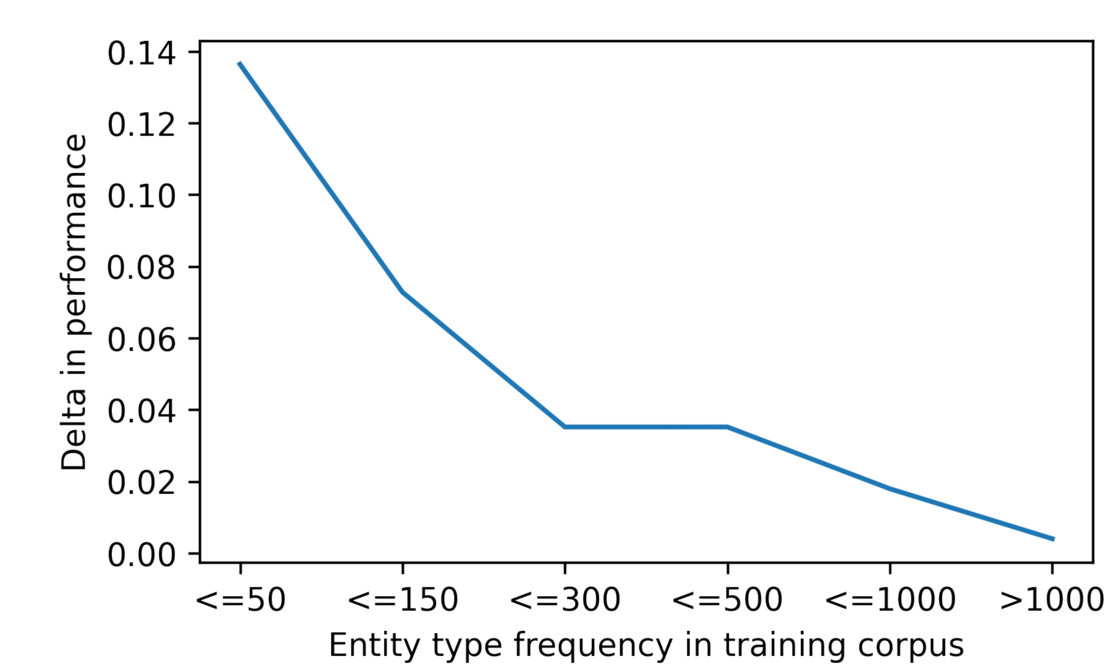
Model	Avg. F1	Δ
Baseline	63.77	-
External Knowledge		
+KB-text	65.55	+1.78
+KB-graph	66.08	+2.31
+both	**66.61**	**+2.84**
Weighting Scheme		
+Prior	65.65	+1.88
+Uniform	65.68	+1.91
+Attention	66.10	+2.33
+AttPrior	**66.61**	**+2.84**

Tab. 1:Average performance.

Q1: Complementarity of *both*, Wikidata and Wikipedia knowledge sources.

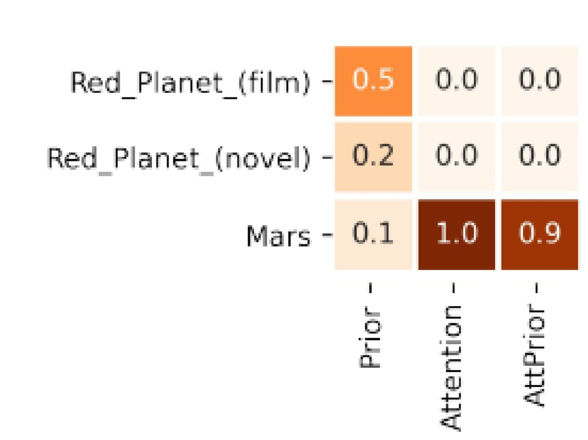
Q2: Best result for attention+prior (AttPrior) weighting scheme.

Performance on rare entity types: external knowledge boosts the performance for entities whose types appear less frequently in the corpus:



Qualitative analysis of weighting schemes $\rightarrow$ for text snippet:

"NASA's Mars rover, "Curiosity" will [...] continue exploring the surface of the Red Planet."



Attention-based schemes assign highest weight to correct entity (*Mars*).

Attention-based schemes are able to capture the textual context.

Contact Information

Severine Verlinden
severine.verlinden@outlook.be
Klim Zaporojets
klim.zaporojets@ugent.be
klimzaporojets.github.io

*Equal contribution

References

Lee, Kenton et al. (2017). "End-to-end Neural Coreference Resolution". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)*, pp. 188–197.
Yamada, Ikuya et al. (2016). "Joint Learning of the Embedding of Words and Entities for Named Entity Disambiguation". In: *Proceedings of The 2016 SIGNLL Conference on Computational Natural Language Learning (CoNLL 2016)*, pp. 250–259.
Yao, Yuan et al. (2019). "DocRED: A Large-Scale Document-Level Relation Extraction Dataset". In: *Proceedings of the 2019 Annual Meeting of the Association for Computational Linguistics (ACL 2019)*, pp. 764–777.
Zaporojets, Klim et al. (2021). "DWIE: An entity-centric dataset for multi-task document-level information extraction". In: *Information Processing & Management* 58.4, p. 102563.

Acknowledgements

Part of the research leading to these results has received funding from (i) the European Union's Horizon 2020 research and innovation programme under grant agreement no. 761488 for the CPN project,[a] and (ii) the Flemish Government under the programme "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen".

[a] https://www.projectcpn.eu/