Unified Examination of Entity Linking in Absence of Candidate Sets

Nicolas Ong, Hassan S. Shavarani, Anoop Sarkar
Simon Fraser University, 8888 University Dr., Burnaby, BC, Canada
{nong, sshavara, anoop}@sfu.ca

Contributions

- We benchmark end-to-end entity linking systems from the literature on the CoNLL/AIDA dataset [1] in a unified evaluation environment consisting of GERBIL [2] and `gerbil_connect` [11].
- We evaluate the resilience of these systems in the absence of pre-computed candidate sets using our unified evaluation environment. This environment enables future ablation studies.
- We examine the adaptability of these systems to unseen test data using the novel AIDA/testc dataset [11].

Candidate Set Ablations

	Micro-F1		
	testa	testb	testc
De Cao et al. (2021b) [7]	85.15	78.98	75.62
De Cao et al. (2021a) [8]	62.00	49.51	37.05
Zhang et al. (2022) [9]	86.81	84.30	72.55
Shavarani and Sarkar (2023) [11]	89.72	82.25	77.54
Poerner et al. (2020) [5]	22.81	18.81	17.56
Feng et al. (2022) [10]	35.00	32.58	27.48

Many ontologies lack high quality candidate sets for entity linking, hence this ablation study.

- Top: We used these models' candidate-set-independent setting.
- Bottom: We replaced their candidate sets with the entire in-domain mention vocabulary of AIDA (5598 entities).

Micro-F1 scores below 1.0 are excluded.

Unified Evaluation of Entity Linking Systems

	Micro-F1			Difference		Modifications
	testa	testb	testc	testa	testb	
Kolitsas et al. (2018) [3]	89.50	82.44	65.75	+0.10	+0.04	-
Peters et al. (2019) [4]						1, 2, 3
KnowBert-Wiki	76.74	71.68	54.12	-3.46	-2.72	
KnowBert-W+W	77.19	71.69	53.92	-4.91	-2.01	
Poerner et al. (2020) [5]	89.40	84.83	65.93	-1.40	-0.17	1, 2, 3, 5
van Hulst et al. (2020) [6]						-
Wiki 2014	83.30	82.53	71.69	-	-0.77	
Wiki 2019	79.64	80.10	73.54	-	-0.40	
De Cao et al. (2021b) [7]	90.09	82.78	75.60	-	-0.92	2, 4
De Cao et al. (2021a) [8]	87.29	85.65	47.54	-	+0.15	4
Zhang et al. (2022) [9]	86.81	84.30	72.55	-	-1.50	-
Feng et al. (2022) [10]	87.64	86.49	65.05	-	+0.19	1, 2, 3, 5
Shavarani and Sarkar (2023) [11]						-
large-500K (no cnds.)	89.72	82.25	77.54	+0.02	+0.05	
large-500K (Kb+Yago)	89.89	82.88	59.50	+0.09	+0.08	
large-500K (PPRforNED)	91.58	85.22	46.98	+0.08	+0.02	

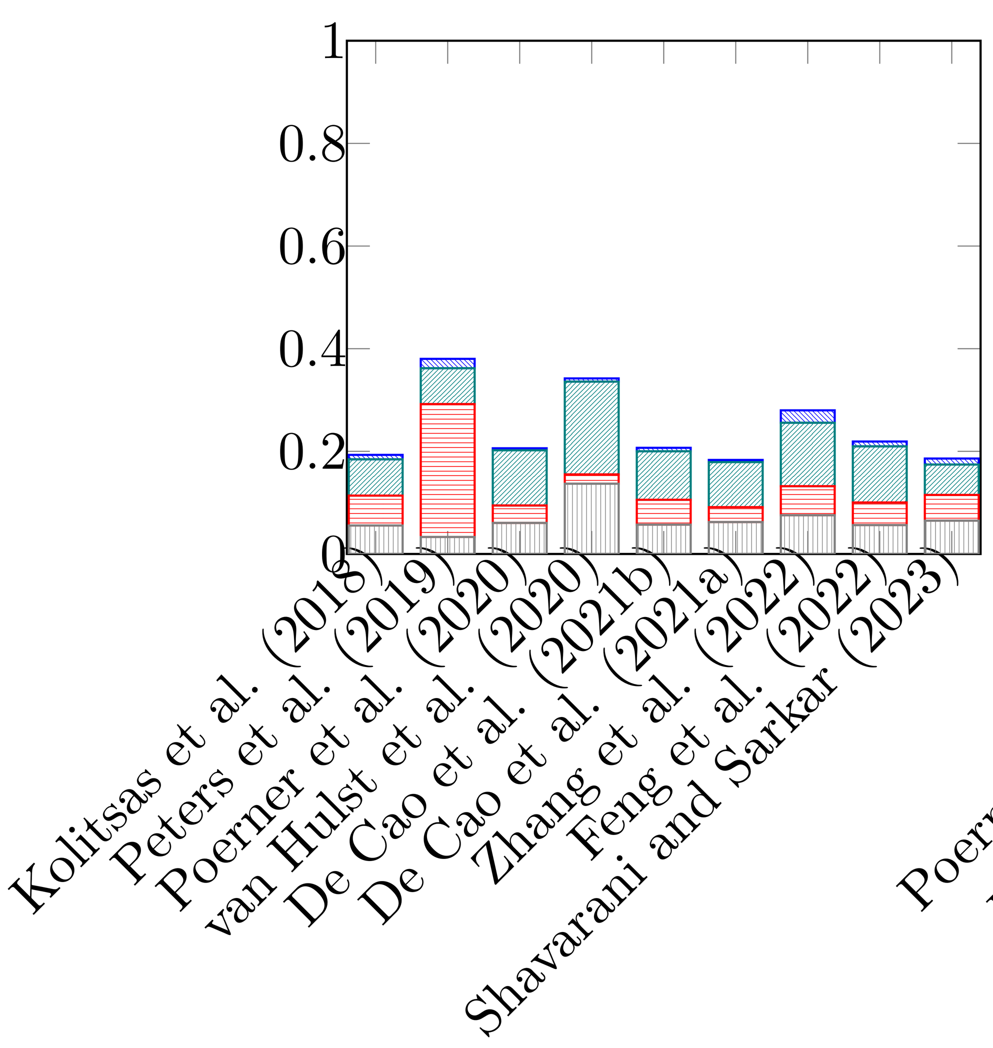
- Micro-F1: The systems' scores in our unified environment.
- Difference: The difference from their originally reported scores.
- Modifications: The changes made to each system to fit into our environment.

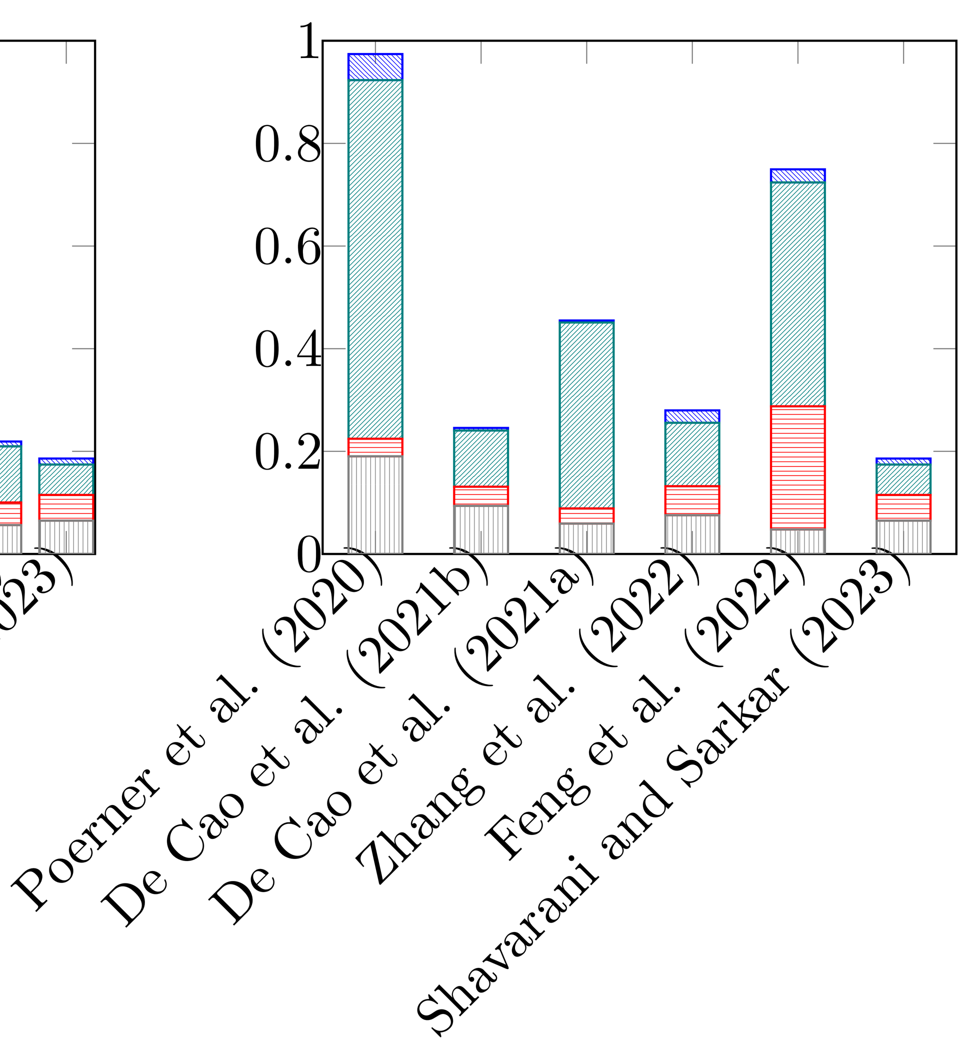
Major modifications:

1. Tokenization for input documents
2. Document splitting for input documents
3. Token-to-character resolution for output annotations
4. Creation of custom data
5. Training of a new model

Future authors should make their source code, trained models, and a `gerbil_connect` [11] integration with GERBIL [2] publicly available. At minimum, provide a function that accepts raw text and outputs a list of character-level annotations.

Results and Analysis





These graphs show error counts of four categories before (left) and after (right) the ablation study:

- over-generation (gray, vertical)
- under-generation (red, horizontal)
- incorrect entity (teal, north east)
- incorrect mention (blue, north west)

Candidate set ablation experiment conclusions:

- Current entity linkers are very dependent on candidate sets.
- Generative and structured prediction approaches are robust.
- The mention-entity similarity approach is less robust.
  - Creating the mention representation with the mention's word tokens fares worse than using its candidate entities.
- Large candidate sets increase run times for many systems.
- Candidate sets primarily decrease incorrect entity errors.

References

[1] J. Hoffart, M. Yosef, I. Bordino, H. Fürstenau, M. Pinkal, M. Spaniol, B. Taneva, S. Thater, and G. Weikum Robust disambiguation of named entities in text In *Proc. of EMNLP 2011*.
[2] M. Röder, R. Usbeck, and A. Ngomo GERBIL – Benchmarking Named Entity Recognition and Linking consistently *Semantic Web*.
[3] N. Kolitsas, O. Ganea, and T. Hofmann. End-to-end neural entity linking In *Proc. of CoNLL 2018*.
[4] M. Peters, M. Neumann, R. Logan, R. Schwartz, V. Joshi, S. Singh, and N. Smith Knowledge enhanced contextual word representations In *Proc. of EMNLP-IJCNLP 2019*.
[5] N. Poerner, U. Waltinger, and H. Schütze E-BERT: Efficient-yet-effective entity embeddings for BERT In *Proc. of EMNLP 2020*.
[6] J. van Hulst, F. Hasibi, K. Dercksen, K. Balog, and A. Vries Rel: An entity linker standing on the shoulders of giants. In *Proc. of SIGIR 2020*.
[7] N. De Cao, G. Izacard, S. Riedel, and F. Petroni Autoregressive entity retrieval In *Proc. of ICLR 2021*.
[8] N. De Cao, W. Aziz, and I. Titov Highly Parallel Autoregressive Entity Linking with Discriminative Correction In *Proc. of EMNLP 2021*.
[9] W. Zhang, W. Hua, and K. Stratos Entqa: Entity linking as question answering In *Proc. of ICLR 2022*.
[10] Y. Feng, A. Fayazi, A. Rastogi, and M. Okumura. Efficient entity embedding construction from type knowledge for BERT In *Findings of AACL-IJCNLP 2022*.
[11] H. Shavarani and A. Sarkar SpEL: Structured Prediction for Entity Linking In *Proc. of EMNLP 2023*.