LAVIS | Hochschule RheinMain

Neural Entity Linking on Technical Service Tickets

Nadja Kurz, Felix Hamann, Adrian Ulges

felix.hamann@hs-rm.de

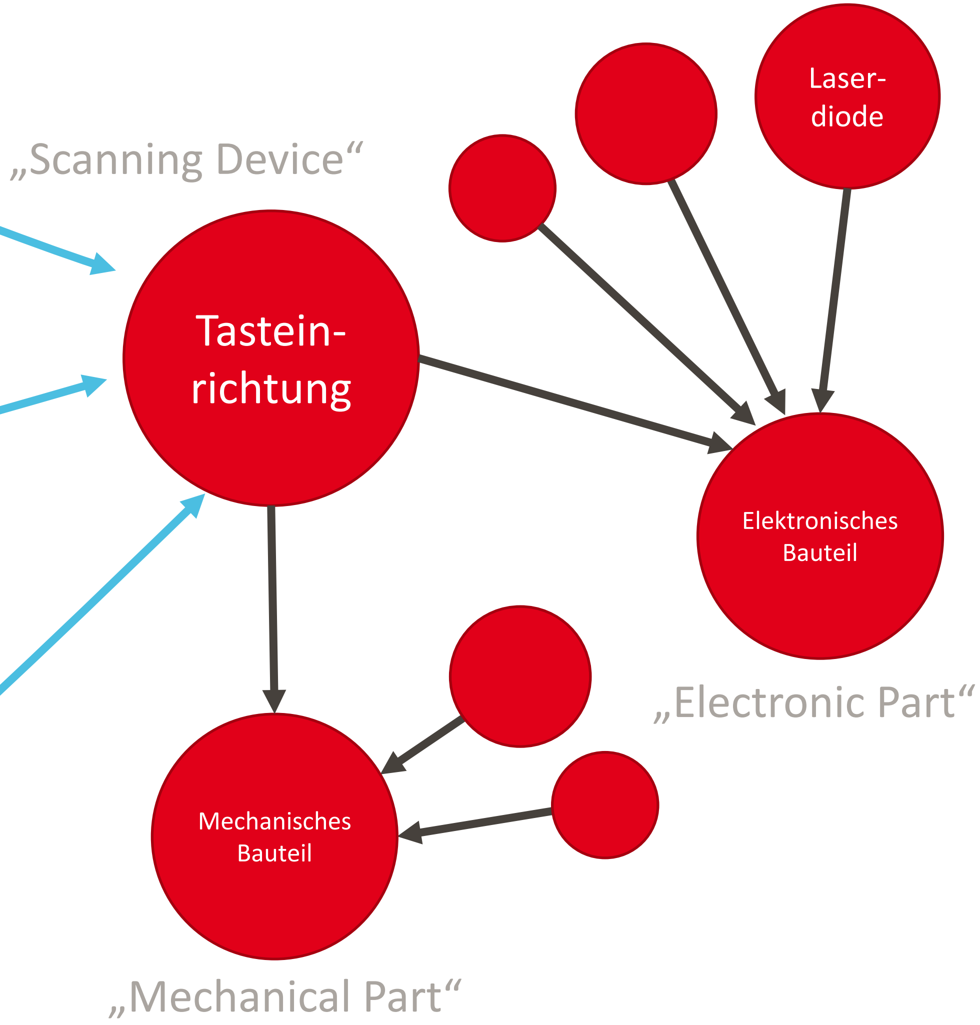
- Documentation is scarce and heterogeneous

Am Verleimteil ist keine Kantentastung im Moment dran. Der 6 mm Schlauch der Tastung hat sich oben in den Niederhalter geklemmt…

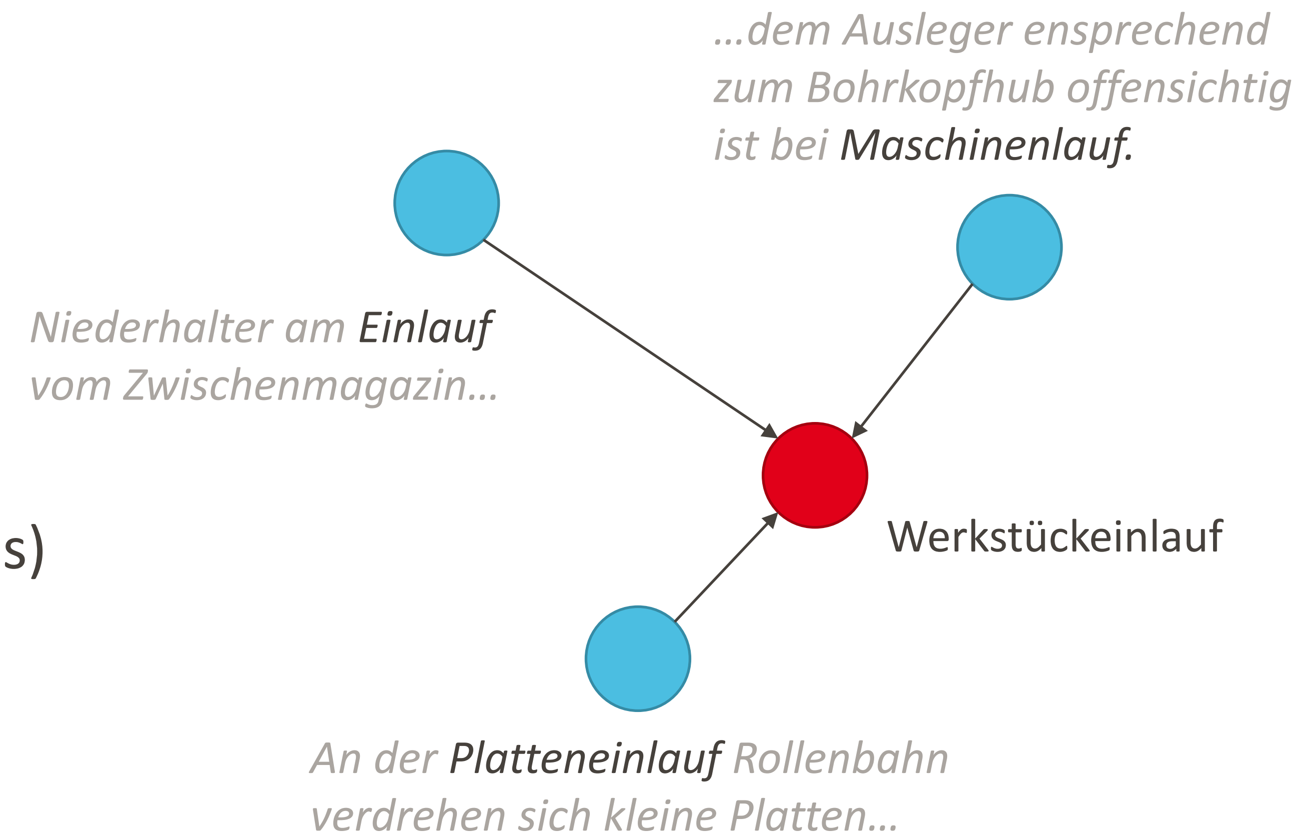
There is no **edge detection** on the gluing part at the moment. The 6 mm hose of the **probe** has clamped itself into the top of the hold-down…

Fehlerauswirkung: Die Vertikalachse (Y-Achse) meldet einen Schleppfehler (nur in Verbindung mit einem Stop der Direktverkettung), eventuell auch die Längstastung…

Error effect: The vertical axis (Y-axis) reports a contouring error (only in connection with a stop of the direct linkage), possibly also the **longitudinal scanner**…



- Task: Link a textual mention to a KB entity [1, 2]
- Easy
  - Spelling Errors
  - Abbreviations
  - Synonyms (general terms)
  - …
- Hard
  - Synonyms (domain specific terms)
  - Hyponyms/Hypernyms
  - Ambiguity
  - …

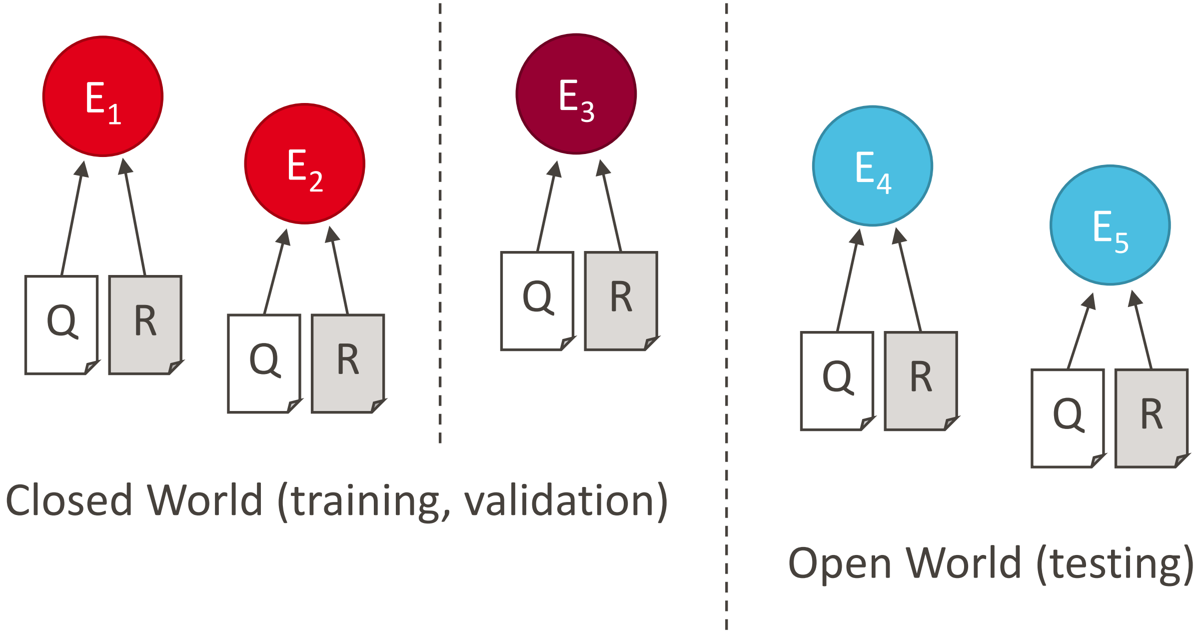


Current State of the Art

- SOTA tackles EL with Representation Learning [3, 4, 5]
  - Unsupervised pre-training on large out-of-domain data (language models)
  - Adaption on target task (transfer learning)
  - Evaluations usually on (high quality) Wikipedia data [6, 7]
- Industry mostly uses heuristics [8]
- Our contribution:
  - Working with low-quality (noisy) data
  - Deep learning in comparison to simple heuristics
  - Zero-shot setting [9, 10]

Data Setup

- Three Open-World datasets (zero-shot)
  - Wikipedia: MIXED, GERÄTE:
    - Mentions selected using the hyperlink structure
  - EMPOLIS: customer issues
    - Mentions selected on human annotated synonym lists



		Entities	Sentences
MIXED	Training	8331	107082
	Validation	1031	13560
	Testing	1027	12853
GERÄTE	Training	5717	65101
	Validation	3231	35823
	Testing	698	7680
EMPOLIS	Training	401	13587
	Validation	201	7680
	Testing	200	6601

Approach 1: Heuristics Pipeline

- Mentions and Entities are both transformed

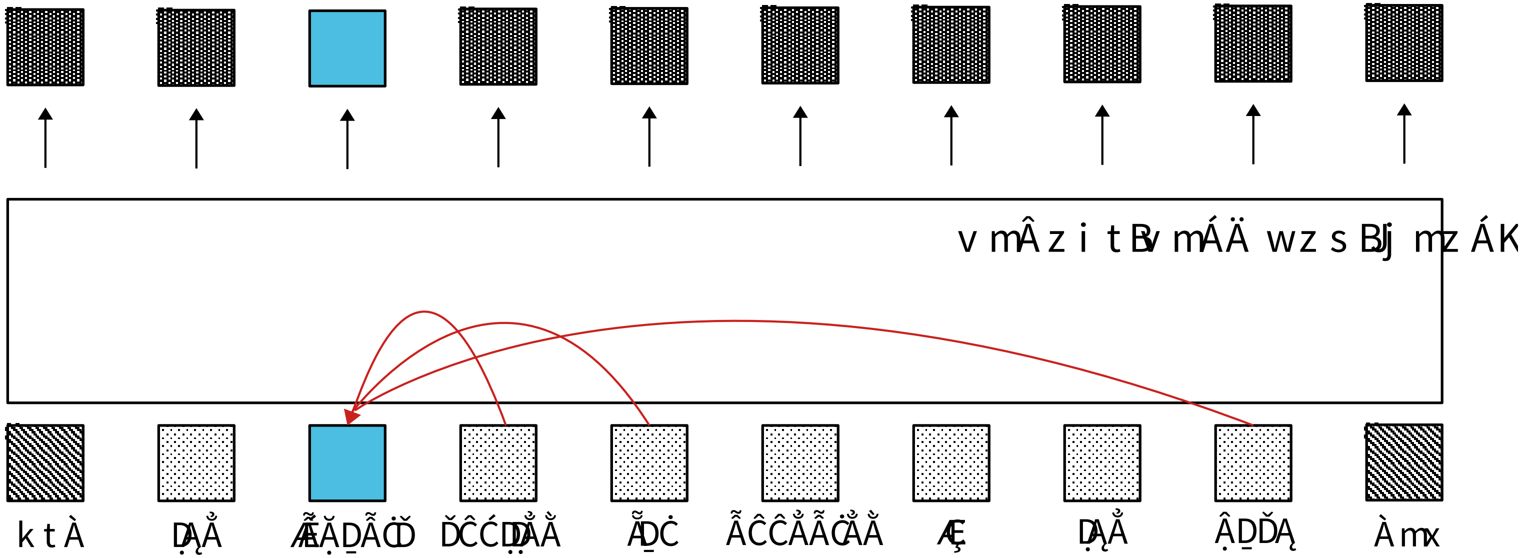
Heuristic	Before	After
Punctuation	CNC-Maschine	CNC Maschine
Corporate Forms	Empolis GmbH	Empolis
Lowercasing	Schwabbelscheibe	schwabbelscheibe
Stemming	astronomische einheit	astronom einheit
Stopword Removal	luren von brudevælte	luren brudevælte
Sorting	linde material handling	handling linde material
Abbreviations*	hohlschaftkegel	hsk

*both token- and compound-based

- Compare by edit distance
- The argmin is returned

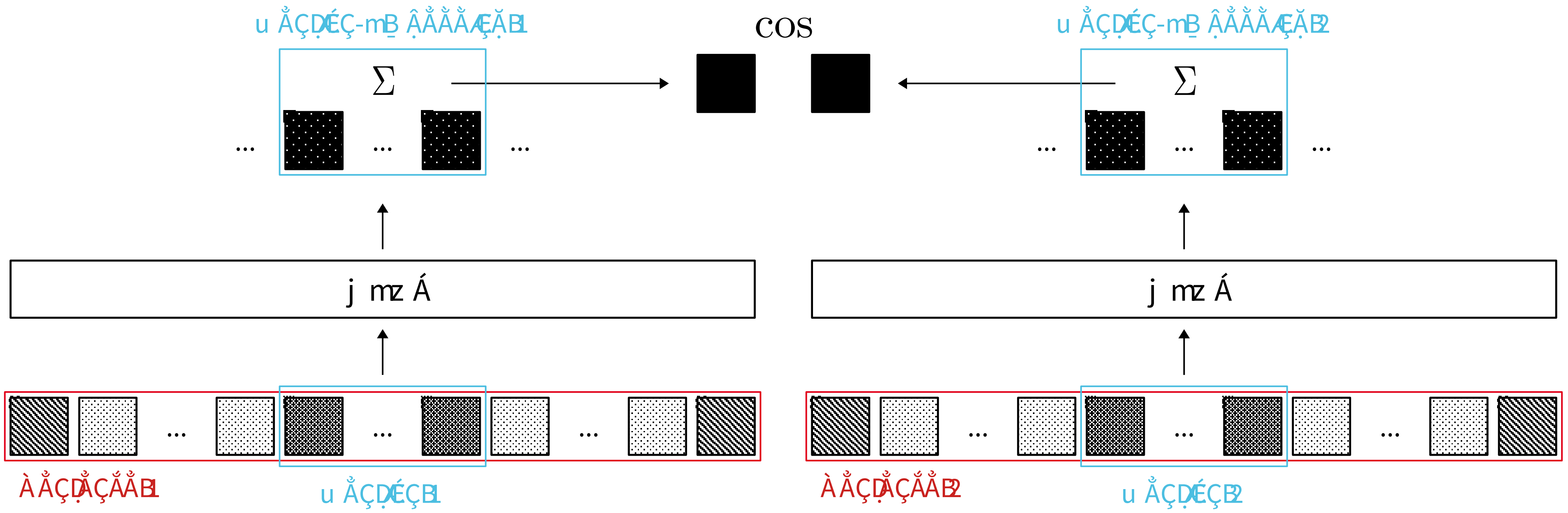
Approach 2: BERT

- Current SOTA: transformer models (self-attention) [11, 12, 13, 14]
- Large & deep models: fine-tuning and inference are expensive



Approach 2a: BERT Bi-Encoder

- Context sentences are successively transformed (caching possible)
- Domain adaption: max-margin loss with negative sampling
- Inference: minimum cosine distance



Evaluation: Heuristics vs. Neural Approach

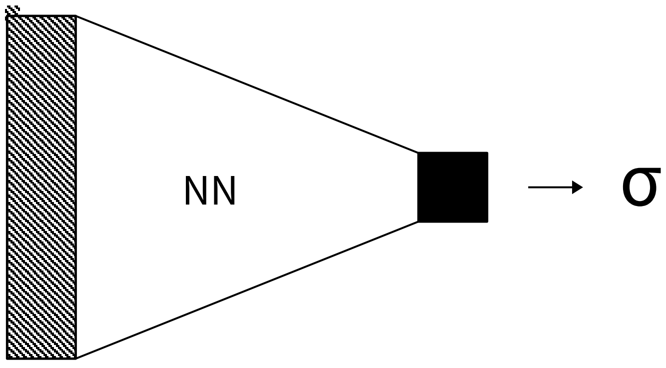
- Measured: Top-1 Accuracy

Classifier	Geräte	Mixed	Empolis
Heuristics	77.87	83.98	51.16
Bi-Encoder	93.30	95.93	40.06
Hybrid	**94.72**	**97.52**	**71.40**

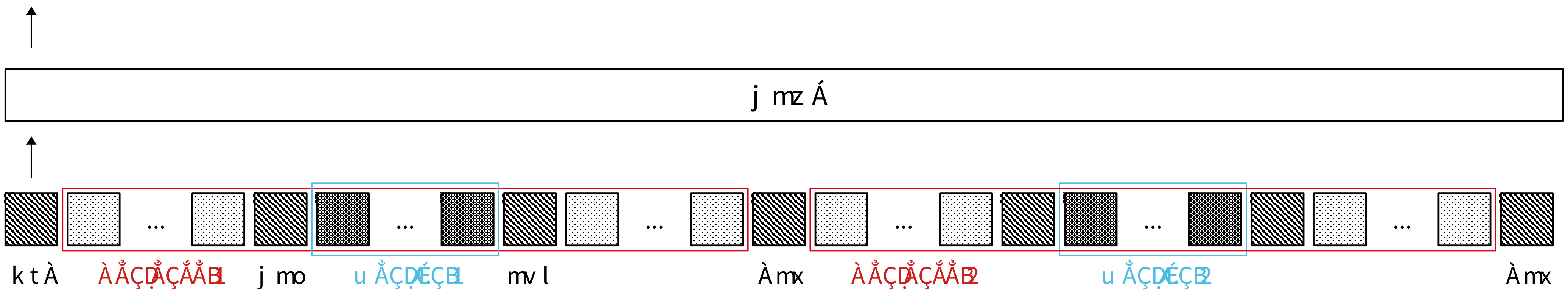
- Hybrid
  - If no suitable candidate was found, fallback to BERT
  - Greatly improves performance on Empolis

Approach 2b: BERT Cross-Encoder

- Context sentences are jointly transformed (no caching)
- Use CLS features for binary classification (feed forward network)



- Domain Adaption: binary cross entropy loss
- Inference: Brute-force search over all candidates



Evaluation: Cross Encoder

- Brute force approach too expensive: reduce number of queries
- Measured: Top-1 Accuracy

Classifier	Geräte	Mixed	Empolis
Bi-Encoder	89.68	93.09	51.53
Cross-Encoder	94.13	96.88	45.41
Hybrid Bi-Encoder	93.41	97.08	80.61
Hybrid Cross-Encoder	**96.42**	**98.05**	**81.63**

- Bi-Encoder with inverted index is much faster (multiple magnitudes)

Thank you!

[1] Heng Ji and Ralph Grishman. Knowledge base population: Successful approaches and challenges. In Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1, pages 1148–1158. Association for Computational Linguistics, 2011.

[2] Daniel Jurafsky and James H. Martin. Speech and Language Processing (2nd Edition). Prentice-Hall, Inc., USA, 2009. ISBN 0131873210.

[3] Kolitsas, Nikolaos, Octavian-Eugen Ganea, and Thomas Hofmann. "End-to-end neural entity linking." *arXiv preprint arXiv:1808.07699* (2018).

[4] Nitish Gupta, Sameer Singh, and Dan Roth. Entity linking via joint encoding of types, descriptions, and context. In Proc. EMNLP, pages 2681–2690, 2017.

[5] Thien Huu Nguyen, Nicolas R Fauceglia, Mariano Rodriguez Muro, Oktie Hassanzadeh, Alfio Gliozzo, and Mohammad Sadoghi. Joint learning of local and global features for entity linking via neural networks. In Proc. COLING, pages 2310–2320, 2016.

[6] Mihalcea, Rada, and Andras Csomai. "Wikify! Linking documents to encyclopedic knowledge." *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. 2007.

[7] Milne, David, and Ian H. Witten. "Learning to link with wikipedia." *Proceedings of the 17th ACM conference on Information and knowledge management*. 2008.

[8] Chiticariu, Laura, Yunyao Li, and Frederick Reiss. "Rule-based information extraction is dead! long live rule-based information extraction systems!." *Proceedings of the 2013 conference on empirical methods in natural language processing*. 2013.

[9] Lajanugen Logeswaran, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, Jacob Devlin, and Honglak Lee. Zero-shot entity linking by reading entity descriptions. arXiv preprint arXiv:1906.07348, 2019.

[10] Ledell Wu, Fabio Petroni, Martin Josifoski, Sebastian Riedel, and Luke Zettlemoyer. Zero-shot entity linking with dense entity retrieval. arXiv:1911.03814, 2019.

[11] Devlin, Jacob, et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.. 2019.

[12] Radford, Alec, et al. "Improving Language Understanding by Generative Pre-training." URL https://s3-us-west-2.amazonaws.com/openai-assets/researchcovers/languageunsupervised/language understanding paper.pdf (2018).

[13] Clark, Kevin, et al. "ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators." International Conference on Learning Representations. 2020.

[14] Kitaev, Nikita, Łukasz Kaiser, and Anselm Levskaya. "Reformer: The Efficient Transformer." arXiv preprint arXiv:2001.04451 (2020).