



Improving Few-Shot and Zero-Shot Entity Linking with Coarse-to-Fine Lexicon-Based Retriever

Shijue Huang^{1,2}, Bingbing Wang^{1,2}, Libo Qin³, Qin Zhao^{1,2},
and Ruifeng Xu^{1,2}✉

¹ Harbin Institute of Technology (Shenzhen), Shenzhen 518000, China
{22S051040,bingbing.wang}@stu.hit.edu.cn,
{zhaoqin,xurui Feng}@hit.edu.cn

² Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies,
Shenzhen 518000, China

³ School of Computer Science and Engineering, Central South University, Changsha,
China
lbqin@csu.edu.cn

Abstract. Few-shot and zero-shot entity linking focus on the tail and emerging entities, which are more challenging but closer to real-world scenarios. The mainstream method is the “retrieve and rerank” two-stage framework. In this paper, we propose a coarse-to-fine lexicon-based retriever to retrieve entity candidates in an effective manner, which operates in two layers. The first layer retrieves coarse-grained candidates by leveraging entity names, while the second layer narrows down the search to fine-grained candidates within the coarse-grained ones. In addition, this second layer utilizes entity descriptions to effectively disambiguate tail or new entities that share names with existing popular entities. Experimental results indicate that our approach can obtain superior performance without requiring extensive finetuning in the retrieval stage. Notably, our approach ranks the 1st in NLPCC 2023 Shared Task 6 on Chinese Few-shot and Zero-shot Entity Linking.

Keywords: Entity linking · Few-shot and zero-shot learning

1 Introduction

Entity linking is a crucial task in natural language processing that involves associating an ungrounded mention in text with its corresponding entity in a knowledge base, thereby facilitating language understanding. Entity linking serves as a fundamental component for various downstream applications, including question answering [4], knowledge base completion [13, 20], text generation [10] and end-to-end task-oriented dialogue system [11]. However, existing entity linking systems face challenges when dealing with newly created or tail entities that share names with popular entities. To address this challenge, the few-shot and zero-shot entity linking task has been proposed, aiming to enhance models’ ability to accurately link against the less popular and emerging entities [8].

Recent research on few-shot and zero-shot entity linking task has primarily employed a “retrieve and rerank” two-stage framework, owing to the vast number of entities present in knowledge bases. In this framework, the first stage involves selecting candidate entities for a given mention, while the second stage reranks these candidates and selects the most probable entity. Specifically, Logeswaran et al. [8] explore deep cross-attention within candidate rerank stage. Wu et al. [16] adopt a BERT architecture based two-stage approach for the zero-shot linking. And to leverage additional information from entity embeddings, Xu et al. [17] consider entities as input tokens and introduce a LUKE-based cross-encoder in the reranking stage. Moreover, Xu et al. [18] enhance the dual encoder model in the retrieve stage by incorporating the Wikidata type system as an auxiliary supervision task. And they also release a new benchmark in Chinese.

In this paper, we propose a coarse-to-fine lexicon-based retriever that improves the few-shot and zero-shot entity linking in an effective manner. Specifically, our proposed approach consists of two layers of lexicon-based retriever. In the first layer, two separate BM25 models [12], namely the AT-BM25 and KB-BM25, are employed to retrieve coarse-grained candidate entities based on the entity names from the alias table and the knowledge base, respectively. Subsequently, in the second layer, another BM25 model called *Description-BM25* is introduced to retrieve fine-grained candidates from the previously obtained coarse-grained candidates by leveraging the entity descriptions. This step utilizes the entity description as detailed context to disambiguate tail entities that share names with existing popular entities. To further refine the retrieved candidates, we employ a BERT-based dual-encoder for reranking. Finally, we propose an ensemble method to aggregate the results of the reranker with the outputs from the three BM25 models in the coarse-to-fine lexicon-based retriever. This ensemble approach ensures more robust predictions.

Experimental results demonstrate significant improvement of our coarse-to-fine lexicon-based retriever, even without extensive finetuning in retrieve stage. Notably, our approach achieves the 1st place in NLPCC 2023 Shared Task 6.

The main contributions of our work are summarized as follows:

- We propose a coarse-to-fine lexicon-based retriever based on BM25 to improve few-shot and zero-shot entity linking, which offers a solution to mitigate the computational burden while improving retrieve accuracy.
- We introduce an ensemble method to aggregate the results from both retrieve and rerank stage, resulting in more robust predictions.
- Empirical results and analyses indicate the effectiveness of our approach in few-shot and zero-shot entity linking. And our approach ranks the 1st in NLPCC 2023 Shared Task 6.

2 Related Work

Entity Linking. Entity linking (EL) refers to the task of associating mentions with entries in a designated database or dictionary of entities. In order to achieve

accurate and efficient identification of target entities from extensive knowledge bases (KBs), most entity linking systems adopt a two-stage approach known as “retrieve and rerank”. This contains a retriever that retrieves candidate entities from the knowledge base, followed by a reranker that reorders the candidates and selects the most probable entity. In traditional entity retrieval, existing methods typically rely on simple techniques such as frequency information [19] or sparse-based models [12] to quickly retrieve a small set of candidate entities. For the ranking stage, neural networks have been widely utilized to calculate the relevance score between mentions and entities [3, 7].

Recently, the emergence of pre-trained language models (PLMs) has led to their extensive adoption in both retrieve and rerank stages of entity linking. For instance, Logeswaran et al. [8] employ a cross-encoder architecture and introduce deep cross-attention in the candidate ranking stage, demonstrating significance of attention between the mention-context pairs and entity descriptions. Wu et al. [16] explore a BERT-based two-stage approach for zero-shot linking. Xu et al. [17] leverage additional information encoded in entity embeddings by considering entities as input tokens in the rerank stage, proposing an entity-enhanced cross-encoder based on LUKE. Wiatrak et al. [14] propose a proxy-based metric learning loss coupled with an adversarial regularizer, providing an efficient alternative to conduct hard negative sampling in candidate retrieve stage. In comparison to these existing works, our approach do not need large-scale fine-tuning for candidate retrieval, effectively reducing computational overhead while maintaining effectiveness.

Few-Shot and Zero-Shot Entity Linking. Entity linking on temporally evolving knowledge bases (KBs) presents a formidable challenge in zero-shot settings. To address this, Hoffart et al. [5] introduce a method for entity linking on emerging entities, particularly dealing with the ambiguity surrounding their names. More recently, Logeswaran et al. [8] propose the zero-shot entity linking task to evaluate the generalization capability of entity linking systems, making minimal assumptions. They also release an English zero-shot entity linking dataset. In the context of Chinese entity linking, Xu et al. [18] introduce a challenging multi-domain benchmark that utilizes Wikidata as the KB. In contrast to most of the existing works, our research focuses specifically on few-shot and zero-shot entity linking in the Chinese language.

3 Methodology

In this section, we introduce our approach for few-shot and zero-shot entity linking. Given a knowledge base E consisting of entities, each characterized by a name e_i and a description d_i . The objective of this task is to determine the most suitable entity for a given mention m_i within a document doc_i .

Our approach follows the “retrieve and rerank” two-stage framework. In the retrieve stage, we propose a coarse-to-fine lexicon-based retriever that efficiently

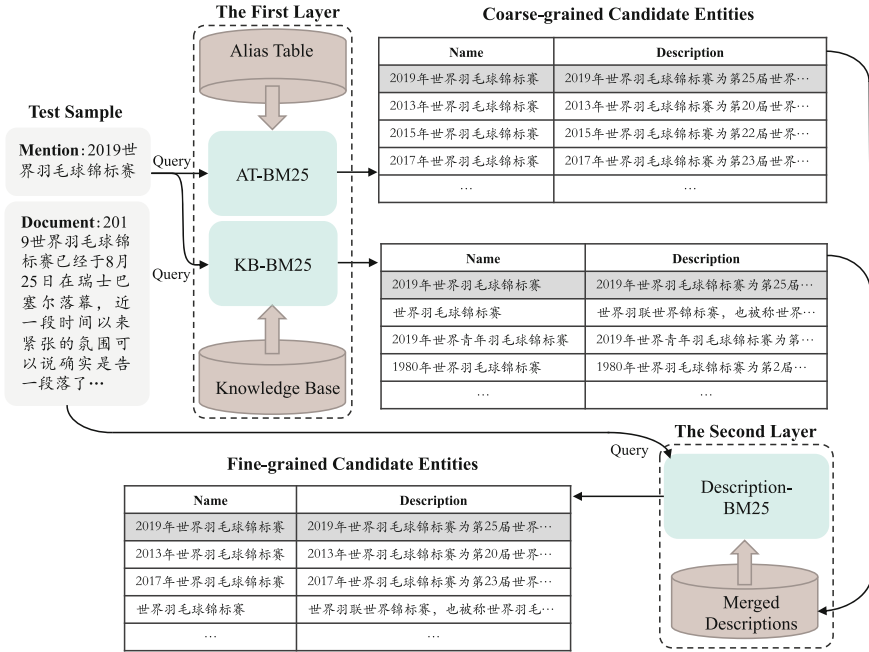


Fig. 1. The architecture of coarse-to-fine lexicon-based retriever, where the first layer contains AT-BM25 and AT-BM25 to retrieve coarse-grained candidate entities from alias table and knowledge base. And the second layer utilizes the Description-BM25 to obtain fine-grained candidate entities from coarse-grained ones.

retrieves candidate entities without the need for time-consuming large-scale fine-tuning (Sect. 3.1). In the rerank stage, we adopt a BERT-based dual-encoder to reevaluate the retrieved candidate entities and select the most appropriate entity (Sect. 3.2). Finally, we introduce an ensemble method to combine the results obtained from both stages, thereby producing more robust predictions (Sect. 3.3).

3.1 Coarse-to-Fine Lexicon-Based Retriever

The overall architecture of our proposed coarse-to-fine lexicon-based retriever is illustrated in Fig. 1. It contains two layers to retrieve candidate entities in a coarse-to-fine manner.

The First Layer (Coarse Stage). Given a text sample (m_i, doc_i) , an entity set $E = \{e_i, d_i\}_{i=1}^N$, and an alias table $AT = \{m_j, e_j\}_{j=1}^M$ that defines the probability of a mention m_j linking to an entity e_j , we construct the retriever in two layers. In the first layer, we employ two separate BM25 models, namely AT-BM25 and KB-BM25. The AT-BM25 model uses the entity names in the alias table as

the corpus, while the KB-BM25 model utilizes that in the knowledge base. By treating the mention m_i in the test sample as a query, we utilize these two models to retrieve coarse-grained candidate entities as follows:

$$D_{AT} = \{m_j, m_j \in AT\}, \quad (1)$$

$$D_{KB} = \{e_i, e_i \in E\}, \quad (2)$$

$$Cand_{AT} = AT\text{-BM25}(m_i, D_{AT}), \quad (3)$$

$$Cand_{KB} = KB\text{-BM25}(m_i, D_{KB}), \quad (4)$$

where D_{AT} and D_{KB} are the corpus to build AT-BM25 and KB-BM25, respectively; $Cand_{AT}$ and $Cand_{KB}$ are the retrieved coarse-grained candidate entities from alias table and knowledge base.

The Second Layer (Fine Stage). To further disambiguate the tail entities based on the detailed mention context, we treat the document doc_i of the test sample as query and merge the obtained coarse-grained candidate entities $Cand_{AT}$ and $Cand_{KB}$ into a non-repeated set $Cand_1$. We then introduce the second layer BM25 model, called `Description-BM25`, which utilizes the descriptions from the obtained coarse-grained candidate entities as the corpus. This step retrieves the fine-grained candidate entities as follows:

$$Cand_1 = Cand_{AT} \cup Cand_{KB}, \quad (5)$$

$$D_{des} = \{d_i, d_i \in Cand_1\}, \quad (6)$$

$$Cand_2 = \text{Description-BM25}(doc_i, D_{des}), \quad (7)$$

where $Cand_1$ are the merged coarse-grained candidate entities; D_{des} is the union of descriptions from $Cand_1$, which is used to build `Description-BM25`; $Cand_2$ is the obtained fine-grained candidate entities.

3.2 BERT-Based Dual Encoder

To rerank the obtained coarse-grained candidate entities and get the most proper entity, we follow previous works [1, 16] to train a BERT-based dual encoder. This approach offers scalability benefits, as the entity embeddings can be pre-computed and stored, allowing for fast retrieval and similarity score computation using dot product. Figure 2 illustrates the architecture of the BERT-based dual encoder. Given a document $\{x_1, x_2, \dots, x_n\}$ with n tokens, where a mention $m = \{x_i, x_{i+1}, \dots, x_j\}$ is present, the input sequence for the mention is constructed as follows:

$$X_1 = [\text{CLS}] x_1, \dots, x_{i-1} [\text{unused0}] x_i, \dots, x_j [\text{unused1}] x_{j+1}, \dots, x_n [\text{SEP}], \quad (8)$$

where `[unused0]` and `[unused1]` are special tokens stand of mention start and end, respectively. Following Humeau et al. [6], we feed the input sequence

of the mention into the BERT encoder [2] to obtain the representation of the [CLS] token in the last layer:

$$\mathbf{y}_m = \text{BERT}(X_1), \tag{9}$$

where \mathbf{y}_m is the representation of given mention.

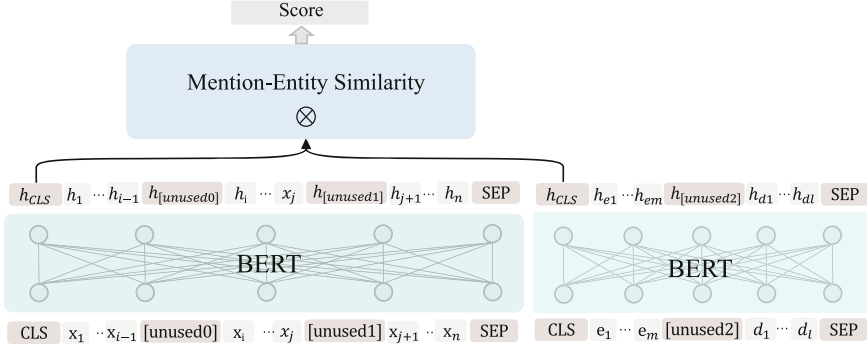


Fig. 2. The architecture of BERT-based dual-encoder. Two BERT encoders are employed to encode the mention within context and the entity-description pair separately. And the mention-entity similarity is computed by dot product.

For a given entity e , which includes the entity name $\{e_1, \dots, e_m\}$ and description $\{d_1, \dots, d_l\}$, the input sequence is created by concatenating the entity name and description:

$$X_2 = [\text{CLS}] e_1, \dots, e_m [\text{unused2}] d_1, \dots, d_l [\text{SEP}], \tag{10}$$

where [unused2] is a special token to separate the entity name and description. Similar to the mention encoding approach, we feed the input sequence of entity into the BERT encoder to obtain the representation of the entity:

$$\mathbf{y}_e = \text{BERT}(X_2), \tag{11}$$

where \mathbf{y}_e is the representation of entity.

Finally, the score of the mention and entity pair is calculated as the dot product of their representations:

$$s(m, e) = \mathbf{y}_m \cdot \mathbf{y}_e \tag{12}$$

Optimization. We apply the standard cross-entropy loss to train the BERT-based dual encoder:

$$\mathcal{L} = -\frac{1}{N} \sum_{n=1}^N y_n \cdot \log \hat{y}_n, \tag{13}$$

where N is the number of total samples, y_n is the golden label, and \hat{y}_n is the predicted distribution.

Inference. During the inference phase, we utilize the BERT-based dual encoder to rerank the combination of coarse-grained and fine-grained candidate entities ($Cand_1 \cup Cand_2$) in the retrieval stage to avoid missing retrievals and error propagation. Moreover, to enhance computational efficiency, we pre-compute \mathbf{y}_e for each $e \in E$ and store all the entity embeddings.

3.3 Ensemble Method

Due to we focus on entity linking in few-shot and zero-shot scenarios, where a majority of the entity mentions are either unseen or only a few of them are encountered during training. This could introduce a bias in the rerank stage, because there is a huge gap between the training set and the few-shot and zero-shot test set. To mitigate this bias, we propose an ensemble method that leverages information from both the retrieve and rerank stages, aiming to enhance the robustness of predictions.

To achieve this, we employ an ensemble strategy that combines four results obtained from the retrieve and rerank stages, include:

- The top-ranked result from the AT-BM25 retrieval process;
- The top-ranked result from the KB-BM25 retrieval process;
- The top-ranked result from the Description-BM25 retrieval process;
- The top-ranked prediction from the BERT-based dual encoder;

In our ensemble strategy, the final predicted result is determined by considering above four predicted results. If two or more results are identical, we select the same result as the final prediction. In cases where the four predicted results differ, we choose the output from the BERT-based dual encoder as the final prediction. Additionally, there may has a unique 2:2 situation arising during voting. In this case, we opt for the prediction that incorporates the result derived from the BERT-based dual encoder.

4 Experiments

4.1 Data

Hansel Dataset. Hansel [18] is a benchmark dataset for few-shot and zero-shot entity linking in simplified Chinese. The training set of Hansel is derived from Wikipedia and the test set consists of Few-Shot (FS) and Zero-Shot (ZS) settings. The FS setting focuses on tail entity linking, while the ZS setting aims to evaluate the zero-shot generalization to emerging and tail entities. The dataset statistics are presented in Table 1.

Knowledge Base. In order to reflect the realistic scenario of knowledge bases evolving over time, the Hansel dataset divides Wikidata entities into two sets, namely the Known and New sets, based on two historical dumps:

Table 1. Statistics of the Hansel dataset.

	# Mentions			# Documents			# Entities		
	In-KB	NIL	Total	In-KB	NIL	Total	E_{known}	E_{new}	Total
Train	9.89M	–	9.89M	1.05M	–	1.05M	541K	–	541K
Validation	9,677	–	9,677	1,000	–	1,000	6,323	–	6,323

- Known Entities (E_{known}) are Wikidata entities from the dump on August 13, 2018. And our models are trained using E_{known} as the knowledge base.
- New Entities (E_{new}) refer to Wikidata entities from the dump on March 15, 2021, which are not present in E_{known} . These entities are added to Wikidata between 2018 and 2021 and unseen during training on the 2018 data, representing a zero-shot setting.

Alias Table. For both E_{known} and E_{new} , Hansel constructs an alias table by extracting information from Wikipedia dated March 1, 2021. And this table is generated by parsing Wikipedia’s internal links, redirections, and page titles.

4.2 Experimental Settings

In the retrieve stage, we use rank-BM25¹ implementation of BM25 algorithm for the coarse-to-fine lexicon-based retriever. And the number of retrieved candidate entities for AT-BM25, KB-BM25 and Description-BM25 are all set 10.

In the rerank stage, our model is implemented using the Huggingface Library [15], with bert-base-chinese as backbone model for the BERT-based dual encoder. We use AdamW [9] to optimize the parameters of model. Due to the large amount of training data, the training epoch is set to 1 and the batch size is set 384. The learning rate is set $5e^{-5}$. And the max sequence length of mention and entity are both set 128. All experiments are conducted on Tesla V100 GPUs.

4.3 Main Results

The evaluation metric employed in is accuracy, and the evaluation results on test set are provided by the organizers. The main results in shown in Table 2, our approach achieves 1st place in NLPCC 2023 Shared Task 6 on Chinese Few-shot and Zero-shot Entity Linking.

It is evident from the results that our approach significantly outperforms the systems ranked second and third, with an accuracy advantage of 9.06% and 15.96%, respectively. This notable improvement can be attributed as follows:

- The proposed coarse-to-fine lexicon-based retriever can accurately retrieve candidate entities, which plays a crucial role in our overall approach and forms a solid foundation for subsequent processes.

¹ https://github.com/dorianbrown/rank_bm25.

Table 2. The results for NLPCC 2023 Shared Task 6.

System name	Accuracy
<i>Ours</i>	0.6915
ITNLP	0.6009
YNU-HPCC	0.5319

Table 3. The performance of three used lexicon-based retriever. The $r@1$, $r@5$ and $r@10$ represent the top-1 recall, top-5 recall and top-10 recall, respectively.

Retriever	$r@1$	$r@5$	$r@10$
AT-BM25	0.5232	0.8245	0.8626
KB-BM25	0.4846	0.6015	0.6614
Description-BM25	0.3891	0.6904	0.8341

- The utilization of a BERT-based dual encoder allows for effective reranking of candidate entities, thereby enhancing the final prediction result.
- Our proposed ensemble method effectively combines the information from both retrieve and rerank stages to disambiguate entities, which bridges the gap between training and few-shot or zero-shot test setting.

4.4 Performance of Retrieve Stage

In order to gain a deeper understanding of how our coarse-to-fine lexicon-based retriever improves few-shot and zero-shot entity linking, we present a comprehensive analysis of the retrieve stage’s performance. Specifically, we retrieve 10 candidate entities for all BM-25 models and report the recall at the top-1, top-5, and top-10 levels on the test set.

The results are illustrated in Table 3. We can have the following observations: Firstly, the AT-BM25 model achieves the best performance across all metrics, indicating that the prior knowledge contained in the alias table significantly benefits the few-shot and zero-shot entity linking. We speculate that this improvement stems from the fact that newly created or tail mentions in the test set may share similar names with certain aliases.

Secondly, the KB-BM25 model exhibits the poorest performance. This observation aligns with our expectations since most mentions do not strictly match entity names in the knowledge base. However, the results obtained from the KB-BM25 model still contribute because mentions link accurately when they match entity names in the knowledge base.

Lastly, the Description-BM25 model does not exhibit significant disambiguation as expected. We attribute this to the limited maximum length of the context, which is imposed by computational resource constraints, thereby impacting the disambiguation of tail entities. Nonetheless, Description-BM25

model provides an alternative perspective by retrieving candidate entities based on descriptions, thus also enhancing the final prediction.

Table 4. The ablation study.

System name	Accuracy
<i>Ours</i>	0.6915
w/o Ensemble	0.6791
w/o AT-BM25	0.6713
w/o KB-BM25	0.6791
w/o Description-BM25	0.6769

4.5 Ablation Study

In order to validate the effectiveness of our proposed coarse-to-fine lexicon-based retriever, we conducted an ablation study by individually removing the ensemble method and the AT-BM25, KB-BM25, and Description-BM25.

The results are presented in Table 4. We can observe that when we remove the ensemble method, the accuracy drops 1.24%. This suggest that the ensemble method can aggregate useful information form both retrieve and rerank stages, which results in more robust prediction.

And it can be observed that upon removing AT-BM25, KB-BM25, and Description-BM25, the accuracy decreased by 2.02%, 1.24%, and 1.46%, respectively. This observation provides evidence that all the employed BM25 models in our coarse-to-fine lexicon-based retriever contribute positively to the task of few-shot and zero-shot entity linking. Furthermore, the combination of these models leads to the retrieval of more accurate candidate entities.

5 Conclusion

In this paper, we present a novel approach for improving few-shot and zero-shot entity linking through a coarse-to-fine lexicon-based retriever. Our proposed method adopts the widely-used “retrieve and rerank” framework, consisting of two stages: the coarse-to-fine lexicon-based retriever for retrieving candidate entities, and a BERT-based dual encoder for reranking the candidate entities. Moreover, we address the learning bias during the training phase of rerank stage by employing an ensemble method that combines information from both retrieve and rerank stages. The experimental results and analyses verify the effectiveness of our approach, which achieves the 1st ranking in the NLPCC 2023 Shared Task 6 on Chinese Few-shot and Zero-shot Entity Linking.

Acknowledgements. This research was supported in part by the National Natural Science Foundation of China(62006062, 62176076), the Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies(2022B1212010005), Natural Science Foundation of Guangdong(2023A1515012922), and Key Technologies Research and Development Program of Shenzhen JSGG20210802154400001.

References

1. Botha, J.A., Shan, Z., Gillick, D.: Entity linking in 100 languages. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 7833–7845. Association for Computational Linguistics, Online (2020). <https://doi.org/10.18653/v1/2020.emnlp-main.630>. <https://aclanthology.org/2020.emnlp-main.630>
2. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, Minnesota (Volume 1: Long and Short Papers), pp. 4171–4186. Association for Computational Linguistics (2019). <https://doi.org/10.18653/v1/N19-1423>. <https://aclanthology.org/N19-1423>
3. Fang, Z., Cao, Y., Li, Q., Zhang, D., Zhang, Z., Liu, Y.: Joint entity linking with deep reinforcement learning. In: The World Wide Web Conference, WWW 2019, pp. 438–447. Association for Computing Machinery, New York (2019). <https://doi.org/10.1145/3308558.3313517>
4. Févry, T., Baldini Soares, L., FitzGerald, N., Choi, E., Kwiatkowski, T.: Entities as experts: sparse memory access with entity supervision. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 4937–4951. Association for Computational Linguistics, Online (2020). <https://doi.org/10.18653/v1/2020.emnlp-main.400>. <https://aclanthology.org/2020.emnlp-main.400>
5. Hoffart, J., Altun, Y., Weikum, G.: Discovering emerging entities with ambiguous names. In: Proceedings of the 23rd International Conference on World Wide Web, WWW 2014, pp. 385–396. Association for Computing Machinery, New York (2014). <https://doi.org/10.1145/2566486.2568003>
6. Humeau, S., Shuster, K., Lachaux, M.A., Weston, J.: Poly-encoders: architectures and pre-training strategies for fast and accurate multi-sentence scoring. In: International Conference on Learning Representations (2020). <https://openreview.net/forum?id=SkxgnnNFvH>
7. Kolitsas, N., Ganea, O.E., Hofmann, T.: End-to-end neural entity linking. In: Proceedings of the 22nd Conference on Computational Natural Language Learning, Brussels, Belgium, pp. 519–529. Association for Computational Linguistics (2018). <https://doi.org/10.18653/v1/K18-1050>. <https://aclanthology.org/K18-1050>
8. Logeswaran, L., Chang, M.W., Lee, K., Toutanova, K., Devlin, J., Lee, H.: Zero-shot entity linking by reading entity descriptions. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, pp. 3449–3460. Association for Computational Linguistics (2019). <https://doi.org/10.18653/v1/P19-1335>. <https://aclanthology.org/P19-1335>
9. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019). <https://openreview.net/forum?id=Bkg6RiCqY7>

10. Puduppully, R., Dong, L., Lapata, M.: Data-to-text generation with entity modeling. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, pp. 2023–2035. Association for Computational Linguistics (2019). <https://doi.org/10.18653/v1/P19-1195>. <https://aclanthology.org/P19-1195>
11. Qin, L., Xu, X., Che, W., Zhang, Y., Liu, T.: Dynamic fusion network for multi-domain end-to-end task-oriented dialog. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 6344–6354. Association for Computational Linguistics, Online (2020). <https://doi.org/10.18653/v1/2020.acl-main.565>. <https://aclanthology.org/2020.acl-main.565>
12. Robertson, S., Zaragoza, H.: The probabilistic relevance framework: BM25 and beyond. *Found. Trends® Inf. Retrieval* **3**(4), 333–389 (2009). <https://doi.org/10.1561/1500000019>
13. Shen, W., Wang, J., Han, J.: Entity linking with a knowledge base: issues, techniques, and solutions. *IEEE Trans. Knowl. Data Eng.* **27**(2), 443–460 (2015). <https://doi.org/10.1109/TKDE.2014.2327028>
14. Wiatrak, M., Arvaniti, E., Brayne, A., Vetterle, J., Sim, A.: Proxy-based zero-shot entity linking by effective candidate retrieval. In: Proceedings of the 13th International Workshop on Health Text Mining and Information Analysis (LOUHI), Abu Dhabi, United Arab Emirates (Hybrid), pp. 87–99. Association for Computational Linguistics (2022). <https://aclanthology.org/2022.louhi-1.11>
15. Wolf, T., et al.: Huggingface’s transformers: state-of-the-art natural language processing (2020)
16. Wu, L., Petroni, F., Josifoski, M., Riedel, S., Zettlemoyer, L.: Scalable zero-shot entity linking with dense entity retrieval. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 6397–6407. Association for Computational Linguistics, Online (2020). <https://doi.org/10.18653/v1/2020.emnlp-main.519>. <https://aclanthology.org/2020.emnlp-main.519>
17. Xu, Z., Chen, Y., Shi, S., Hu, B.: Enhancing entity linking with contextualized entity embeddings. In: Lu, W., Huang, S., Hong, Y., Zhou, X. (eds.) *NLPCC 2022*. LNCS, vol. 13552, pp. 228–239. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-17189-5_19
18. Xu, Z., Shan, Z., Li, Y., Hu, B., Qin, B.: Hansel: a Chinese few-shot and zero-shot entity linking benchmark, pp. 832–840. Association for Computing Machinery, New York (2023). <https://doi.org/10.1145/3539597.3570418>
19. Yamada, I., Shindo, H., Takeda, H., Takefuji, Y.: Joint learning of the embedding of words and entities for named entity disambiguation. In: Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning, Berlin, Germany, pp. 250–259. Association for Computational Linguistics (2016). <https://doi.org/10.18653/v1/K16-1025>. <https://aclanthology.org/K16-1025>
20. Zhang, C., et al.: Feature engineering for knowledge base construction. *IEEE Data Eng. Bull.* (2014)