

Machine Reading Comprehension Model in Domain-Transfer Task

I. S. Rozhkov^{1*} and N. V. Loukachevitch^{1**}

(Submitted by E. E. Tyrtysnikov)

¹*Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow, 119991 Russia*

Received March 15, 2023; revised April 17, 2023; accepted April 25, 2023

Abstract—The paper studies domain-transfer capabilities of the machine reading comprehension model (MRC) in named entity recognition task. In such a task, the model is required to retrieve some knowledge from texts with prompts in order to extract named entities. Usually, this requires training on a large dataset. However in low-resource or domain-specific scenarios creating training dataset could be hard and non-effective. This is where domain transfer from available datasets can be useful. We check prompt design for the machine reading comprehension model to transfer knowledge between domains. We study the approach in transfer from general dataset NEREL to biomedical NEREL-BIO, which share some set of common name entities.

DOI: 10.1134/S1995080223080504

Keywords and phrases: *transfer learning, named entity recognition, machine reading comprehension, prompt learning.*

1. INTRODUCTION

Named Entity Recognition task (NER) has been popular throughout many years in plenty of fields. In this task, a model is required to retrieve named entities that denote some real (or abstract) world objects and phenomena. In the general domain of processing news texts, the performance of NER models in extracting entities of such general types as PERSON, ORGANIZATION, LOCATION achieves high results exceeding 90% in precision and recall. But creating a NER extractor in a new domain requires constructing a new dataset annotated with domain-specific entities. General entities also can be met in these domain-specific texts but their contexts can significantly change, which leads to the degradation of NER results for models trained on general texts [1]. This problem became a basis for domain transfer task settings that is how to improve the results of a model trained on one dataset but applied to another dataset.

One of current state-of-the-art NER trends are models treating the NER task as a question-answering task, in which a question is constructed for each target entity. In such approaches, the performance significantly depends on the type of question (prompt). The current paper is devoted to the study of prompts in Machine Reading comprehension model (MRC) in transfer between two Russian NER datasets: from the general dataset NEREL [2] to biomedical dataset NEREL-BIO [3]. These two datasets have common entity types, which are used for our evaluation¹⁾. In previous studies the MRC model obtained the best results of named entity recognition in experiments on both datasets [2].

Both datasets are annotated with nested named entities, which means that a named entity can contain another named entity. This enhances the coverage of annotated named entities, provides a better basis for establishing relations between entities and linking entities with a knowledge base. At the same time, this feature of datasets needs application of specialized models that can work with nested structures. The MRC model is one of the best of such models but it is very resource-consuming because

*E-mail: fulstocky@gmail.com

**E-mail: louk_nat@mail.ru

¹⁾<https://github.com/nerel-ds>

it asks questions about each entity type to each sentence. The training and application of the MRC model requires high-performance capacities for computations.

We study the following research questions of domain transfer between two datasets with nested named entities:

- what prompt for the MRC model is the best for the model transfer between datasets without additional training,
- what prompt is the best for transfer and additional training on the target dataset,
- if it is useful to change a prompt from general one used in model training on the general NEREL dataset to a domain-specific prompt for application to NEREL-BIO.

2. RELATED WORK

Many different options were designed for named entity recognition task. Among them there are classic machine-learning methods like Support Vector Machines [4, 5], Hidden Markov Models [6, 7] and deep learning approaches such as recurrent neural networks and long-short term memory [8, 9], transformer models [10]. Recently, BERT [11] has become popular as a basic model for many natural language processing tasks, and named entity recognition has been no exception. It has proved to be efficient in many domains [12, 13].

Biomedical field of the named entity recognition is studied thoroughly. Among most common issues addressed we have data sparsity [14, 15], nestedness of biomedical named entities [16], automatic labeling of rare biomedical objects [17] and others.

Named Entity Recognition task in such narrow fields is commonly addressed as a few-shot or cross-domain task because of named entity class rareness in comparison to "usual" ones. Labeling such entities is crucial. There are different approaches such as utilization of large pretrained models [19], prototype-based methods [20], cross-domain approaches [21] and transfer learning.

One of the most popular ideas in transfer learning is prompt learning [18, 22, 23]. These methods "help" some pretrained model to retrieve such rare entities based on "prompts" or "hints" — sentences in natural language that somehow describe named entities. Among them Machine Reading Comprehension model [25] is utilizing both pretrained model and prompt design for effective nested named entity recognition.

3. DATASETS

Both NEREL and NEREL-BIO datasets are annotated with nested named entities intended for further annotation with relations and entity links to corresponding knowledge bases. NEREL entities are linked to Wikidata knowledge graph; NEREL-BIO entities are connected to UMLS entities.

3.1. NEREL

NEREL is the first Russian dataset annotated simultaneously with nested entities, relations between those entities, and knowledge base links [2, 26].

Nested entities enhance coverage of annotated entities, as well as relations between entities and knowledge base links. For example, extracting only single ORGANIZATION entity from *Lomonosov Moscow State University* leads to the loss of two internal named entities: CITY (*Moscow*) and PERSON (*Lomonosov*). In such cases, NEREL employs the nested annotation *//Lomonosov/PERSON [Moscow]/CITY State University/ORGANIZATION*.

Annotation of internal entities allows 'local' relations between nested entities. In the above example, it allows for establishing a relation between the university and its headquarters (*Moscow*). Annotated nested entities enable wider coverage of entity linking into a knowledge base. For example, entity *Mayor of Novosibirsk* is absent in Wikidata (which is a reference knowledge base for NEREL), but nesting permits linking of the internal entities *Mayor* and *Novosibirsk*.

At the time of this writing, NEREL contains 56K named entities and 39K relations annotated in 900+ person-oriented news articles. 29 entity types in NEREL can be categorized as follows:

- basic entity types: PERSON, ORGANIZATION, LOCATION, FACILITY;
- geopolitical entities subdivided into COUNTRY, CITY, DISTRICT and STATE_OR_PROVINCE entities;
- numerical entities: NUMBER, ORDINAL, DATE, TIME, PERCENT, MONEY, AGE;
- socio-political entities: NATIONALITY, RELIGION, IDEOLOGY, FAMILY, LANGUAGE;
- law-related entities: LAW, CRIME, PENALTY;
- work-related entities: PROFESSION, WORK_OF_ART, PRODUCT, AWARD;
- DISEASE for labeling various health disorders and symptoms;
- EVENT, which is used for labeling so called "news events" as opposed to everyday or regular activity.

3.2. NEREL-BIO

In construction of NEREL-BIO, some general entity types from NEREL were annotated to be able to describe the most important social relations of biomedical entities. The DISEASE entity type is most relevant to the biomedical domain. Also it was decided to use in the NEREL-BIO labeling the following NEREL general entity types:

- 4 basic types (PERSON, ORGANIZATION, LOCATION, FACILITY),
- 4 geopolitical types (COUNTRY, CITY, DISTRICT and STATE_OR_PROVINCE) entity types,
- 7 numerical entities (NUMBER, ORDINAL, DATE, TIME, PERCENT, MONEY, AGE),
- tags for characterizing persons (NATIONALITY, PROFESSION, and FAMILY),
- PRODUCT and EVENT.

EVENT entity is used for labeling such situations as epidemics, military conflicts, tsunamis, etc, mentioned in connection with the spread of diseases or the need for additional medical care. The EVENT entity mainly corresponds to Environmental event and Traumatic event concepts in UMLS.

In total, 20 general-domain entity types are available for annotating biomedical texts.

Besides, NEREL-BIO entity types includes 17 specialized biomedical entities such as ANATOMY, CHEM (chemical substances), procedural entities (MEDPROC, LABPROC, SCIPROC), etc. [3] but in the current study we work only with entities, which are common for both datasets.

Some principles of annotation employed in the general domain were changed in NEREL-BIO. In particular, in the general domain, mainly capitalized mentions were annotated as named entities. In the biomedical domain, the same entity types can also appear as lower-cased mentions:

- any humans or groups can be annotated with the label PERSON such as *patient, control group, population with low income*;
- ORGANIZATION tag is used not only for tagging specific organizations but organization types such as *hospital, medical institution, rehabilitation center*.
- location-related tags (LOCATION, COUNTRY, CITY, STATE_OR_PROVINCE, DISTRICT, FACILITY) are also used in both cases: *rural settlement, low-income countries, coastal areas, Brasilia, Vietnam*.

These changes in annotation principles reflect a typical situation in transfer between domains: the names and general meaning of entity types may be preserved, but the principles of their markup are modified to a greater or lesser extent.

NEREL dataset consists of 933 texts, and NEREL-BIO consists of 766 texts in total.

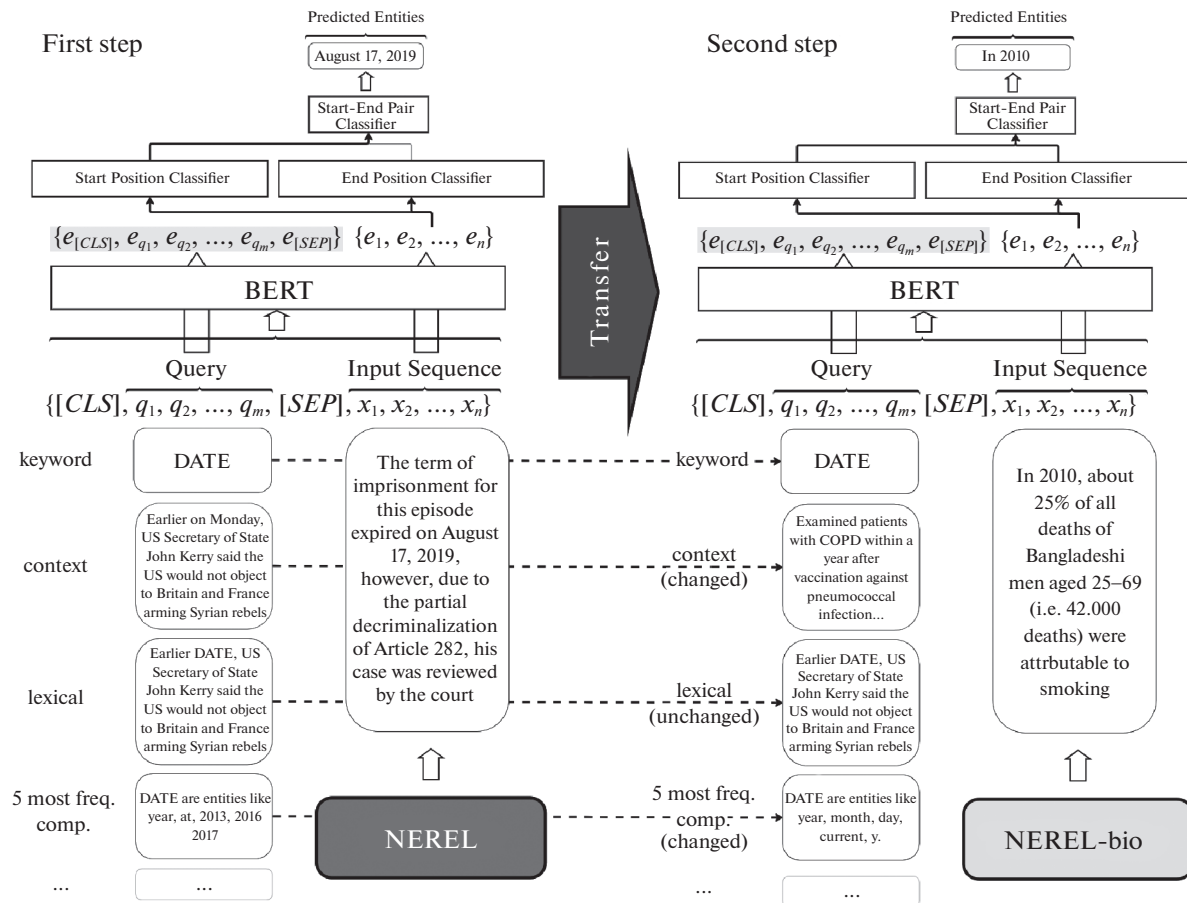


Fig. 1. Machine Reading Comprehension Model in Transfer Task. The transfer consists of two steps: the first one is to train our approaches on the NEREL dataset, and the second step is to use trained weights to train the model further on the NEREL-bio dataset. Different approaches to prompts design in both steps are given, with prompt and data examples.

4. MACHINE READING COMPREHENSION MODEL

MRC task is formulated in the following way [25]: for the given context X and question Q the model should obtain answer A with some function F defined as $A = F(X, Q)$. In the named entity recognition task, X would be the given sentence/paragraph; Q is some generated or selected query sentence for a given named entity type; A is the subsequence of the context X that denotes the named entity; F is the retrieving model itself.

For the MRC model, we employed three binary classifiers based on the output of the last hidden layer from the RuBERT model (Russian BERT) [24]. The first classifier determines the starting position of the named entity. The second classifier determines the ending position of a named entity (possibly different) of the same class. The third classifier decides, whether chosen start-end pairs represent a single named entity of such class. These classifiers are trained for each class (type) separately.

The outline of our model and studied approaches are given in the Fig. 1, combined with several examples.

We compare several question variants for the MRC model.

Keyword: the question consists of a single entity tag such as DISO or ANATOMY [25].

Component-based: 2-5-10 most frequent lemmatized components of a given entity are used for formulating a query, for example "DISO are entities such as a tumor, complication, disorder, disease, illness" (5-component example). Previous experiments with the general NEREL dataset showed that component-based questions outperformed other variants [27].

Contextual: a sentence from the training sample containing a named entity of a given type without explicit or implicit labeling used for this entity in the sentence. For example, a question for DISO entity type can be as follows: “60 patients in the most acute period of hemispheric ischemic stroke were examined.”

Lexical: as in the contextual variant, a sentence from the training corpus is used as a question; additionally, the entity of a given type is masked with its label [28].

We used the so-called full lexical approach, when all entities in a sentence of a given type are substituted with masks. An example of a masking sentence with several mentions of an entity looks as follows. The initial sentence contains three mentions of DISO: “The addition of gout contributes to endothelial dysfunction and worsens the course of hypertension”. The corresponding lexical question is: “The addition of DISO contributes to DISO and worsens the course of DISO”. If a longer entity contains a shorter entity of the same type, the longer entity is preferred (so called outmost variant). The selection of a sentence for contextual or lexical questions is carried out in the following way:

- The most frequent entity for a given entity type is selected;
- The first sentence in the training set that contains the selected entity is extracted to be used as a question. By “first” we imply here the lexicographic order of the filenames of the original dataset.

5. EXPERIMENTS AND EVALUATION

We study transfer abilities of the MRC model from the NEREL to NEREL-BIO dataset. Experiments are carried out as follows:

1. We train the MRC model on the NEREL dataset.
2. We apply trained model directly to the NEREL-BIO dataset: train it further on NEREL-BIO data and then test it.
3. We collect obtained results.

These steps are visualized in Fig. 1.

Moreover, we evaluated two other types of experiments.

The first one repeats the same algorithm, but here model was not trained further on NEREL-BIO dataset, only tested directly after the training on general NEREL domain. We motivate this as direct transfer task, in order to compare trained and untrained results of transfer on NEREL-BIO dataset.

The second one was usual training and testing of our approaches on NEREL-bio dataset, with no transfer or NEREL utilization at all. It was conducted for comparison purposes too, whether transfer training or usual training would obtain better performance in this task.

Thus, we get three types of experiments, by which we group all evaluations and obtained results.

For training we use pre-defined divisions of the datasets to training, validation and test sets. We train our models on training sets of NEREL and/or NEREL-BIO and test them on the test set of NEREL-BIO. The NEREL dataset is divided into the train, dev, and test sets – 746/94/93 texts, respectively. NEREL-BIO is split into train/dev/test subsets as follows: 612/77/77 documents.

Given that both these datasets contain common named entity classes, we can study different approaches to the transfer:

- Train on NEREL common classes, test and train again on NEREL-BIO common classes.
- Train on all NEREL classes, but test and train again only on common NEREL-BIO classes.
- Train on all NEREL classes, test and train again on all NEREL-BIO classes too.

Moreover, as aforementioned, the MRC model uses different prompt design strategies. Some of them require part of the training subset to be included and transformed. Hence, we get the option, which subset – original or new one – to use during transfer. Therefore, we get two approaches:

- “Swap”. The prompt is recreated on the target dataset, where the trained model is transferred too.
- “No swap”. The prompt is kept same.

In total we get six different scenarios.

Table 1. Results of MRC model in the common-to-common variants

Prompt	micro F1, %					macro F1, %				
	Transfer		Transfer, trained		Usual	Transfer		Transfer, trained		Usual
	swap	noswap	swap	noswap		swap	noswap	swap	noswap	
context	11.53	49.14	84.77	76.81	83.67	7.37	54.92	76.97	68.30	72.06
keyword	51.57	51.57	84.06	84.06	83.87	59.56	59.56	74.39	74.39	72.88
lexical	12.84	60.78	84.04	84.07	84.10	18.58	62.46	75.52	76.81	73.83
mfc2	56.57	61.18	84.46	83.82	83.78	48.73	60.31	75.33	74.13	71.33
mfc5	56.08	60.46	83.54	83.74	83.81	55.07	60.81	70.19	72.54	73.13
mfc10	47.62	59.39	84.39	83.94	83.86	52.65	59.19	74.08	70.30	71.61

Table 2. Results of MRC model in the full-to-common variants

Prompt	micro F1, %					macro F1, %				
	Transfer		Transfer, trained		Usual	Transfer		Transfer, trained		Usual
	swap	noswap	swap	noswap		swap	noswap	swap	noswap	
context	11.02	48.53	84.30	76.45	83.67	6.15	54.60	75.41	68.93	72.06
keyword	48.79	48.79	84.71	84.71	83.87	59.53	59.53	73.10	73.10	72.88
lexical	7.25	59.08	84.06	84.16	84.10	18.71	62.56	75.36	74.56	73.83
mfc2	57.07	60.09	84.46	84.46	83.78	53.12	60.74	72.88	74.54	71.33
mfc5	57.01	60.42	84.24	84.44	83.81	55.78	60.93	72.56	73.78	73.13
mfc10	50.52	60.19	83.89	84.40	83.86	55.90	61.12	74.25	74.08	71.61

Table 3. Results of MRC model in the full-to-full variants

Prompt	micro F1, %				macro F1, %		
	Transfer, trained		Usual		Transfer, trained		Usual
	swap	noswap			swap	noswap	
context	84.40	77.54	83.67		73.36	70.18	72.06
lexical	84.08	83.94	84.10		75.01	70.92	73.83
mfc5	84.53	71.02	83.81		74.21	67.36	73.13

6. RESULTS

We implemented different options and evaluated the results.

Batch size in MRC training was set to 16 with maximum length of the sequence to be 192 tokens. Model was trained during 16 epochs on 4 Tesla V100 GPUs with 3 randomly initialized starts in all cases and their results were averaged across different runs. Other parameters are set to default values after [25].

The results are given, respectively, in Tables 1–3. Abbreviations mfc2, mfc5, mfc10 mean 2-, 5-, 10-most frequent component prompts respectively.

All tables have similar structure. The “Transfer” column contains results of model training on the NEREL dataset and direct application of the model to NEREL-BIO dataset. The “Transfer, trained”

column contains the results of the MRC model pre-trained on NEREL dataset and further trained on the NEREL-BIO dataset. Their subcolumns “swap” and “noswap” match to two aforementioned settings of prompts change. The column “Usual” is the same in all tables and presents the results of MRC model trained on the training set of NEREL-BIO without any transfer approaches. These three columns correspond to the aforementioned three types of our experiments. All these options are divided into two groups by final metric used for evaluation performance: micro and macro F1. The best result in this scenario is obtained with a lexical prompt (73.83 Macro F-measure).

Table 1 shows the results of MRC model trained on common entities in NEREL dataset. As for the transfer with no change of prompts, we see that the worst results in the direct transfer are obtained via the contextual prompt, which is a sentence from source (general) NEREL dataset. The best and similar results are obtained with lexical, mfc2 and mfc5 prompts. After additional training on NEREL-BIO, variants with keyword, lexical and mfc2 prompts improve their Macro F results if compared with usual NEREL-BIO training.

When prompt is changed, we see that only the keyword prompt is not changed and the corresponding variant preserves the same results. In direct transfer, the results for all prompts with the change of prompt are worse for all variants. Sentence-based prompt variants (context and lexical) obtained extremely low results. But after additional training on NEREL-BIO, sentence-based variants greatly improve the results, context-based results achieves the best results on Micro and Macro F-measures.

Table 2 considers the same variants of preserving or changing prompts but the MRC model is trained on entities of all types in the NEREL dataset. Comparing this table with previous Table 1, we can see that the results are mainly similar for corresponding prompt variants, the best achieved results became worse. This means that recognition of entities of common types in the target dataset requires training only on these types without consideration all types.

Table 3 presents the results of training models on all entities of NEREL as pre-training and further training on NEREL-BIO. We see that the results became worse in noswap variant, which uses general prompts from NEREL in the MRC model. The results became better using pre-training in swap variant if compared with usual training on NEREL-BIO dataset.

From our experiments, we can make the following conclusions. Firstly, we can conclude that trained transfer option outperforms the performance of the standard training procedure. In most cases and for most prompts we get better results. Thus, we can state that using knowledge from different domains is beneficial.

Secondly, swapping option happened to be crucial in our studies. For example, for lexical and context prompts we can see that for the performance of swapped models vastly outperforms unswapped ones. But for “direct” transfer, on the contrary, they make much more mistakes and do poorly in general, while unswapped ones get decent results.

Thirdly, most frequent component prompts achieve similar results both in swapped and unswapped approaches, while differences for sentence-based prompts (context and lexical) are much more significant. We can conclude that most popular entities in different domains for same classes, in practice, do not change too much, and the MRC model can effectively retrieve such knowledge.

Fourthly, we see that changing initial training from only common classes to all entity classes does not enhance the performance vastly, it stays almost at the same level of performance. This means that for such transfer purposes we do not need other, non-common named entity classes to perform the task, saving computational time and resources.

7. CONCLUSIONS

In this paper we studied the role of prompts in Machine Reading comprehension model (MRC) in transfer between two datasets with common entity types. With this aim, we utilized Russian NER datasets: the general dataset NEREL and biomedical dataset NEREL-BIO. We experimented with such prompts as keyword, sentence-based prompts (contextual and lexical), and several variants of most frequent component prompts. In particular, we considered swap and noswap options, where swap option means change of the prompt in transfer from the general dataset to the domain-specific dataset.

The best results in direct transfer from NEREL to NEREL-BIO on common entities were achieved with lexical prompts, in noswap variant (without prompt changing): 62.46 Macro F-measure. The best performance in NEREL-BIO training with NEREL-based pretraining were achieved with sentence-based prompts with prompt changing about 76 Macro F-measure, which is significantly better than training only on NEREL-BIO.

FUNDING

The work is supported by the Russian Science Foundation, grant no. 20-11-20166. The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University.

REFERENCES

1. M. M. Tikhomirov, N. V. Loukachevitch, and B. V. Dobrov, "Recognizing named entities in specific domain," *Lobachevskii J. Math.* **41**, 1591–1602 (2020).
2. N. Loukachevitch, E. Artemova, T. Batura, P. Braslavski, I. Denisov, V. Ivanov, S. Manandhar, A. Pugachev, and E. Tutubalina, "NEREL: A russian dataset with nested named entities, relations and events," in *Proceedings of the International Conference on Recent Advances in Natural Language Processing* (2021), pp. 876–885.
3. N. Loukachevitch, S. Manandhar, E. Baral, I. Rozhkov, P. Braslavski, V. Ivanov, T. Batura, and E. Tutubalina, "NEREL-BIO: A dataset of biomedical abstracts annotated with nested named entities," arXiv: 2210.11913 (2022).
4. Z. Ju, J. Wang, and F. Zhu, "Named entity recognition from biomedical text using SVM," in *Proceedings of the 5th International Conference on Bioinformatics and Biomedical Engineering, Wuhan, China* (2011), pp. 1–4. <https://doi.org/10.1109/icbbe.2011.5779984>
5. H. Isozaki and H. Kazawa, "Efficient support vector classifiers for named entity recognition," in *COLING 2002, Proceedings of the 19th International Conference on Computational Linguistics* (2002).
6. R. Malouf, "Markov models for language-independent named entity recognition," in *COLING-02: Proceedings of the 6th Conference on Natural Language Learning* (2002).
7. S. Morwal, N. Jahan, and D. Chopra, "Named entity recognition using hidden Markov model (HMM)," *Int. J. Natl. Language Comput.* **1** (2012).
8. C. Lyu, B. Chen, Y. Ren, and J. Donghong, "Long short-term memory RNN for biomedical named entity recognition," *BMC Bioinform.* **18**, 462 (2017). <https://doi.org/10.1186/s12859-017-1868-5>
9. J. Hammerton, "Named entity recognition with long short-term memory," in *Proceedings of the 7th Conference on Natural Language Learning at HLT-NAACL 2003* (2003), pp. 172–175.
10. J. Yu, J. Jing, L. Lyang, and R. Xia, "Improving multimodal named entity recognition via entity span detection with unified multimodal transformer," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 3342–3352.
11. J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT* (2019), pp. 4171–4186.
12. Z. Ji, Q. Wei, and H. Xu, "Bert-based ranking for biomedical entity normalization," in *AMIA Summits on Translational Science Proceedings* (2020), p. 269.
13. L. Gessler and N. Schneider, "BERT has uncommon sense: Similarity ranking for word sense BERTology," in *Proceedings of the 4th BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP* (2021), pp. 539–547.
14. L. Weber, J. Münchmeyer, T. Rocktäschel, M. Habibi, and U. Leser, "HUNER: Improving biomedical NER with pretraining," *Bioinformatics* **36**, 295–302 (2020).
15. D. Piliouras, I. Korkontzelos, A. Dowsey, and S. Ananiadou, "Dealing with data sparsity in drug named entity recognition," in *Proceedings of the 2013 IEEE International Conference on Healthcare Informatics* (2013), pp. 14–21.
16. B. Alex, B. Haddow, and C. Grover, "Recognising nested named entities in biomedical text," in *Biological, Translational, and Clinical Language Processing, Anthology* (2007), pp. 65–72.
17. T. Munkhdalai, M. Li, K. Batsuren, H. A. Park, N. H. Choi, and K. H. Ryu, "Incorporating domain knowledge in chemical and biomedical named entity recognition with word representations," *J. Cheminform.* **7**, 1–8 (2015).
18. R. Ma, X. Zhou, T. Gui, Y. Tan, L. Li, Q. Zhang, and X. J. Huang, "Template-free prompt tuning for few-shot NER," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2022), pp. 5721–5732.
19. M. Khalifa, M. Abdul-Mageed, and K. Shaalan, "Self-training pre-trained language models for zero-and few-shot multi-dialectal arabic sequence labeling," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume* (2021), pp. 769–782.
20. M. Tong, S. Wang, B. Xu, Y. Cao, M. Liu, L. Hou, and J. Li, "Learning from miscellaneous other-class words for few-shot named entity recognition," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, Vol. 1: *Long Papers* (2021), pp. 6236–6247.

21. C. Jia, X. Liang, and Y. Zhang, "Cross-domain NER using cross-domain language modeling," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (2019), pp. 2464–2474.
22. Y. Huang, K. He, Y. Wang, X. Zhang, T. Gong, R. Mao, and C. Li, "Copner: Contrastive learning with prompt guiding for few-shot named entity recognition," in *Proceedings of the 29th International Conference on Computational Linguistics* (2022), pp. 2515–2527.
23. P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing," *ACM Comput. Surv.* **55** (9), 1–35 (2023).
24. Y. Kuratov and M. Arkhipov, "Adaptation of deep bidirectional multilingual transformers for russian language," arXiv: 1905.07213 (2019).
25. X. Li, J. Feng, Y. Meng, Q. Han, F. Wu, and J. Li, "A unified MRC framework for named entity recognition," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (2020), pp. 5849–5859.
26. N. Loukachevitch, P. Braslavski, V. Ivanov, T. Batura, S. Manandhar, A. Shelmanov, and E. Tutubalina, "Entity linking over nested named entities for russian," in *Proceedings of the 13th Language Resources and Evaluation Conference* (2022), pp. 4458–4466.
27. I. Rozhkov and N. Loukachevitch, "Machine reading comprehension model in RuNNE competition," in *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference Dialog* (2022).
28. W. Zhou and C. Muhao, "An improved baseline for sentence-level relation extraction," in *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing* (2022).