

Prompts in Few-Shot Named Entity Recognition

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

^a *Research Computing Center, Lomonosov Moscow State University, Moscow, 119991 Russian Federation*

^{*}*e-mail: fulstocky@gmail.com*

^{**}*e-mail: louk_nat@mail.ru*

Abstract—The article explores various methods for generating questions (prompts) in question-answer approaches to named entity recognition. In such methods, recognition is understood as answering questions about named entities. The system must return valid named entities of the required type that are referenced in the sentence. In particular, we study the effect of prompts on the nested named entity recognition in few-shot setting. We test different prompt-based approaches against data from the RuNNE-2022 nested named entity extraction competition.

Keywords: nested named entities, few-shot task, machine reading comprehension

DOI: 10.1134/S1054661823020104

INTRODUCTION

Among many tasks of automatic processing of texts in natural language, the task of named entity recognition (NER) remains one of the main ones. This task involves looking up named entities of various types to extract information from unstructured texts. It is known that approaches based on neural networks are considered the most effective [3, 14, 21, 27] in the tasks of extracting information from texts. Large pre-trained language models such as BERT [4] and ELMo [22] have improved the performance of many automatic natural language processing problems, including named entity extraction.

However, in most datasets on which the extraction of named entities is studied, and neural network models are trained, an important property of named entities is missing—their possible nesting. Nesting means that some entities can contain others. For example, the named entity “Lomonosov Moscow State University” contains a mention of the entities “Lomonosov” and “Moscow.” Most named entity recognition datasets contain named entity markup without specifying internal named entities, which can result in missing important named entities.

Successful training of neural network-based models in the named entity recognition problem requires large datasets with labeling containing many examples of each type of entity. However, in practice, large annotated datasets may not be available in many cases [7, 9, 28]. For low-resource languages and rare entities, a small number of examples in training datasets leads to poor quality results. Therefore, we can formulate the so-called few-shot problem of named entity

recognition task, that is, the problem of training models on a small amount of data. Thus, two subtasks can be formulated for the task of extracting named entities—a “general” task and a few-shot task.

In this paper, we consider one of the most effective methods for recognizing nested named entities, the MRC method (Machine Reading Comprehension, machine understanding of texts) [19] in its application to the recognition of nested named entities for the Russian language in a general task and in a few-shot task, where the training set for some entities includes a specially limited number of examples. The MRC method transforms the named entity recognition problem into a question-answer problem. We explore different wordings of questions (prompts) to identify the best options for the general and few-shot task of extracting named entities. In this paper, we propose several types of new prompts for MRC model. We analyze various methods for filling out templates and creating prompts that are used as queries for the MRC model.

The contribution of this work is as follows.

- We offer various approaches to the formation of prompts, in much the same ways as the techniques described in the methods of relation extraction [29].
- We study modifications of prompts taking into account the nesting of entities.
- We evaluate various combinations of such prompts.

Various prompts in the MRC method are tested against the data of the RuNNE-2022 [1] competition on nested named entity recognition based on the NEREL [20] dataset annotated with 29 named entity types. We extend the study of MRC prompts for the few-shot named entity recognition problem described in [23].

The structure of the article is as follows. In Section 1, we review related works. Section 2 describes the machine reading comprehension (MRC) model. Section 3 presents various methods for generating prompts. Section 4 describes the NEREL dataset and its modification used in testing RuNNE-2022. Section 5 reviews the results achieved.

1. RELATED WORKS

1.1. Named Entity Recognition Approaches

In the recognition of named entities, algorithms for marking up sequences (sequence models) are widely used, since in selecting a named entity it is important to take into account its context in the sentence [2, 10, 14, 21]. For the most part, each word of the sequence corresponds to one entity label (the so-called “flat” approaches), which means that the approaches cannot work with nested named entities [6, 24, 26]. Some approaches combine “flat” methods, using them on each nesting layer of named entities [11, 12]. So their idea is to aggregate the output of layer models to get internal (or external) entities. The authors of [24] achieve this goal by applying the “flat” named entity extraction algorithm recursively on each new internal subsequence of the original sentence. Another approach is a multiclass labeling of a sequence with concatenated labels from each of the layers [26].

1.2. Few-Shot Named Entity Recognition Task

In few-shot named entity extraction problems, many standard approaches are not effective because of the lack of training data. Therefore, new methods are being developed to solve this problem. Two methods are mainly proposed: transfer learning and meta-learning strategies.

Transfer training. Transfer training (or learning) is a method of “transferring” knowledge—from one solved problem to another unsolved one. There are several types of this approach. The first is preliminary training on large unlabeled data with further training on labeled data (the so-called semi-supervised learning). For example, the BERT model [4] was pre-trained on a large unlabeled collection of texts using a masking technique, which allowed the model to successfully determine information about masking words from the context of a sentence. Thus, this model can be used in specific tasks on labeled datasets, where it has already been trained the contextual and semantic relationships of words in a sentence. Secondly, neural models associated with some knowledge base (for example, a dictionary) provide more information about rare entities [5]. Finally, this group of methods includes training the model on some larger dataset (usually with other types of classes) [28]. This trained model is then used on a smaller dataset. The main goal of such a method is to help the model efficiently extract the hidden patterns of named entities in gen-

eral in order to obtain the missing semantic information. Note that this approach differs from the pre-trained model approach, where there is no preset task during training, but the so-called fine-tuning is used.

Metalearning. Metalearning techniques are designed to create a model that learns by effectively recognizing entity types as such, rather than by learning from individual examples from a dataset. In a few-shot setting, one of the most common ideas is metric learning or prototype-based learning [25]. Each entity is represented as a “prototype” in a given metric space in which the similarity between entities is defined. The model should then be trained not on the markup of the original dataset, but on the distances between entity prototypes in that metric space. Thus, this approach provides a more generalized pattern for model training.

Prompt-based learning. Recently, approaches based on questions or prompts have been developed in various problems of automatic text processing [8, 17, 18]. The key idea here is to create a more efficient process for fitting a pretrained model on a small data problem. These prompts are intended to help the model “get” useful knowledge from the pretrained representations in order to successfully be trained.

There are different methods for generating prompts. For example, we have manual selection, template filling, paraphrasing, etc. One approach is called prompt augmentation or demonstration learning [8, 15]. These prompts are designed using templates filled with examples taken from the training data. Therefore, the model should better predict tasks with masked slots given such samples.

2. MODEL OF MACHINE READING COMPREHENSION

The idea behind the machine reading comprehension (MRC) is the ability of a computer system to “understand” information from a given text and answer specific “questions” about the content of that text. Therefore, this method is determined on the basis of a question-answer task in which it is necessary to obtain answers to some questions.

Initially, the task of machine reading comprehension is formulated as follows: for a given context X and question Q , the model should deduce a response A with some function F , defined as $A = F(X, Q)$. For example, in the problem of named entity recognition, X will be a given sentence, a sequence of tokens; Q is some generated or selected query sentence for a given entity type; A is the subsequence of the context X denoting the named entity; and F is the extraction model itself.

Li et al. [19] presented an MRC model for extracting nested named entities (see Fig. 1).

The model consists of three binary classifiers based on the output of the last hidden layer from the BERT

[4] model, which is given a query (following the MRC formulation) and an input sentence. 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. Thus, two sets are created: probable starting and ending positions of some entity. This approach makes it possible to get nesting of any depth. The third classifier then decides whether the selected start–end pairs represent a single named entity of the given class. These classifiers are trained for each class type.

The main problem for the task of machine reading comprehension is the careful choice of query types. We call these questions “prompts” motivated by the idea that the model uses different semantic information from these sequences to improve the quality of information extraction. Next, we will look at our various techniques for generating prompts that can be used as queries for the MRC model. We will analyze several well-known techniques and propose several new types of prompts.

3. PROMPTS FOR MODEL OF MACHINE READING COMPREHENSION

3.1. Basic Approaches

In the MRC model, each entity type, as mentioned above, needs to be associated with a corresponding query. Several approaches to generation of prompts have been proposed.

First, Li et al. [19] proposed some of them, in particular, approaches based on keywords. In [23], prompts were proposed in the form of definitions and the N most frequent examples of entities. In this section, we will also look at a few simple approaches to generation of prompts, namely, contextual and lexical. We call them basic because they are not new, unlike the new approaches that we will describe later.

Keyword. This approach maps each entity class type to its as-is label (or its interpretation), creating one-word queries for each class. For example, for the class *ORG* it could be “*Organization*”; for *PER*, it could be “*Person*”; for *AGE*, it can be itself “*AGE*.”

Definitions. Prompts are created on the basis of definitions (interpretations) from dictionaries for words corresponding to the meaning of each named entity type. Definitions have been carefully selected from various vocabularies for each entity class. For example, for an entity class *AGE*, the query will be like this: “*Age is the period of time when someone was alive or something existed.*”

Components. There are two different options for component prompts.

- **N most frequent examples of entities.** Used as prompts are the N most frequent examples of entities of this class from the training dataset. Then a template is built over them, where these entities are listed in descending order of their frequency.

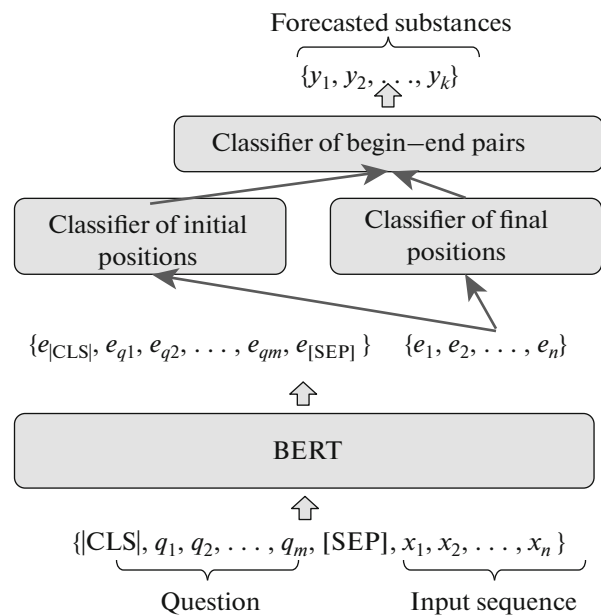


Fig. 1. Schematic of the MRC model.

- **N most frequent components of entities.** This is the same as in the previous approach, but all entities are divided into words (called entity components) and lemmatized, and only then are the most frequently occurring words selected from the training set. The query is formed in the same way.

Concatenation. In addition, it is also suggested as prompts to merge definitions and the N most common entity components.

In Table 1, we provide several examples of such prompts.

Contextual prompt. Some sentence (taken from the training sample) containing a named entity of a given type is taken as a prompt, but for this entity, there is neither explicit nor implicit marking in the sentence. This type of prompting was used in a demonstration approach to prompt-based learning [16].

Table 1. Examples of different types of prompts for the DATE entity [23]

Type of hint	Example
Definitions	DATE is the number of day in month, frequently combined with the name of day, month, and year
Most frequent examples	DATE are such substances as Monday, Tuesday, today, 2011 year, 2004 year
Most frequent components	DATE are such substances as year, 2016, 2013, 2017

The choice of sentence from the training dataset is carried out in the following stages.

1. As in the most frequent entities [23], we generate a set of all entities of this class type and count their number in the training set.

2. The most frequently occurring entity in the given class is selected.

3. The first sentence in the training subset that contains the selected entity is retrieved. By “first” we mean here the lexicographic order of the filenames of the original dataset. This sentence is the resulting context.

Let us take an example. Suppose we have two texts.

• File *c.txt*: *On Friday, June 21, US President Barack Obama nominated Republican James Comey to head the FBI, NBC News reported. If the Congressional Senate approves the candidacy of James Comey, he will replace the director of the FBI, 68-year-old Robert Mueller, who took up this post a week before the September 11, 2001, attacks.*

• File *a.txt*: *James Comey is 52 years old. His track record includes work as the United States Attorney for Manhattan and Deputy Attorney General in the administration of George W. Bush. Since leaving public service, he has worked as a lawyer for hedge fund Bridgewater Associates and military corporation Lockheed Martin, and is on the board of directors of HSBC Bank. He is also a Senior Fellow at Columbia University.*

Here we see the named entities of the class *PERSON*, namely, *Barack Obama*, *James Comey*, again *James Comey*, *Robert Mueller*, *James Comey*, and *George W. Bush*. To get a contextual query for an entity class *PERSON*, we now perform the following steps.

1. Count the number of occurrences for all entities of a class *PERSON*. Here we get (sorted in descending order) {*James Comey*: 3, *Barack Obama*: 1, *Robert Mueller*: 1, *George W. Bush*: 1}.

2. Select the most frequent entity. Thus, we get *James Comey* with the highest frequency equal to three.

3. We find the first sentence containing this entity.

We first parse the file *a.txt* and only then look at the file *c.txt*, because *a.txt* is lexicographically earlier in the given dataset than *c.txt*. Thus, we look over the sentences *a.txt*. We get the first sentence containing the entity *James Comey*: *James Comey is 52 years old. His track record includes United States Attorney for Manhattan and Deputy Attorney General in the George W. Bush administration.* This sentence becomes the search query for the contextual prompt of entity class *PERSON*.

Lexical prompts. Like contextual prompts, these prompts use sentences from the labeled corpus, but additionally mask the desired entity with its label in order to implicitly “demonstrate” the essence of a given entity type of model being trained. There are several variations of this approach.

• **Monolexical** approach. Only one named entity is replaced by its keyword.

• **Full lexical** approach. In the sentence, not only one entity of a given class, but all such entities are replaced by a keyword. Owing to the fact that the problem of nested named entities is being considered, with this approach, they can intersect; therefore, collisions are possible if they are nested in each other. In other words, this approach can be used only in the case of flat layout. In this regard, two options were considered:

- replacing all the most external named objects;
- replacing all the most internal named objects.

By outermost here, we mean a named entity that is not included in any other named entity. By innermost, we also mean a named entity that does not encompass any other named entity.

For example, in a sentence “*Lomonosov Moscow State is considered one of the best universities in the Russian Federation*,” there are entities *Russian Federation*, *Lomonosov Moscow State University*, *Moscow*, and *Lomonosov*. The innermost entities are *Lomonosov* and *Moscow*. The entity *Lomonosov Moscow State University* is the outermost entity, and the entity *Russian Federation* considered as both the outermost and the innermost entity. A similar idea of extracting entities using masks was used earlier in the problem of relation extraction [29].

Next, we consider new types of prompts, namely, structural and multiprompt.

3.2. Structural Prompts

For such prompts, we store both the entity and its class in the sentence, forming some structure out of them. Several options are considered.

• **Frame** approach. A combination of contextual and lexical approaches, in which the occurrence of a given named entity is replaced by a structure of the form *[Entity | Class]*.

• **Marker** approach. Same as frame, but the form of structure changed to *<Class> entity </Class>*.

• **Typed marker**. Same as in the marker approach, but here we take a structure of the form *@ CLASS @ entity*.

• **Full structural** approach. It is similar to the full lexical approach in relation to the structural approach: all occurrences of named entities are replaced by structures of a given form. But unlike lexical ones, they can be nested into each other. In this case, as in the case of the full lexical prompt, all entities can be “flat-tened.” Thus, there are three options:

- replacing all nested named entities;
- replacing all the most external named entities;
- replacing all the most internal named entities.

Similar approaches to representing entities using entity markers and entity type markers have previously

been used in relational extraction methods [29]. Tables 2 and 3 provide examples of “single-entity” (including contextual and lexical) and “multientity” prompts, respectively.

3.3. Multiprompt Approach

Here we propose the idea of multiple types of prompts at once, so that the model can get more semantic information about a given entity from different “points of view.” As for the MRC model, which is trained using the one-sentence—one-query method, this approach explicitly expands the data multiplied by the number of prompt types used.

The three-outermost full lexical prompt. Here we use the outermost full lexical approach three times in a row, but different contexts are chosen for each of them. As mentioned above in the context approach, where we chose a context in the lexicographic order of filenames and took the first one, here we take the first, second, and third such contexts. Each prompt itself is generated in the same way as was done originally. Here, the outermost full lexical approach was chosen as the one that showed the best results on the studied dataset.

Following the context help example above in Section 4.2 with two texts, we would get the second sentence “*On Friday June 21, US President Barack Obama named Republican James Comey to head the FBI, NBC News reported.*” and the third sentence “*If the Congressional Senate approves the candidacy of James Comey, he will replace 68-year-old Robert Mueller as director of the FBI, who took this post a week before the September 11, 2001, attacks.*”

So this query looks like “*PERSON is 52 years old. His track record includes work as the United States Attorney for Manhattan, Deputy Attorney General in the administration of George W. Bush. On Friday June 21, US President Barack Obama nominated Republican PERSON to head the FBI, NBC News reported. If the Congressional Senate approves candidacy of PERSON, he will replace 68-year-old Robert Mueller as director of the FBI, who took this position a week before the September 11, 2001, attacks.*”

Outermost full lexical + two most frequently components of entities + definitions. Here we use these three types of prompts that have achieved the best performance (each separately). Therefore, three queries are generated for each sentence. Each prompt is obtained in exactly the same way as was originally done.

Outermost full lexical + outermost full typed marker. Here we mix another great approach with the outermost full lexical approach to explore the generalization between the two.

Outermost full lexical + outermost full typed marker + contextual. We add another promising prompt to the previous method to analyze the quality of the results obtained.

Table 2. Examples of different types of prompts with the same entity

Type of hint	Examples
Context	In 1964, Towns together with Nikolai Basov and Aleksandr Prokhorov received the Nobel Prize in Physics. Vladimir Putin has already accepted his resignation, having appointed the first vice-governor Sergei Tsivilev to be the interim
One-lexical	In 1964, Towns together with Nikolai Basov and Aleksandr Prokhorov received the AWARD in Physics. PERSON has already accepted his resignation, having appointed the first vice-governor Sergei Tsivilev to be the interim
Framing	In 1964, Towns together with Nikolai Basov and Aleksandr Prokhorov received the [Nobel Prize AWARD] in Physics. [Vladimir Putin PERSON] has already accepted his resignation, having appointed the first vice-governor Sergei Tsivilev to be the interim
Labeling	In 1964, Towns together with Nikolai Basov and Aleksandr Prokhorov received the <AWARD> Nobel Prize </AWARD> in Physics. <PERSON> Vladimir Putin </PERSON> has already accepted his resignation, having appointed the first vice-governor Sergei Tsivilev to be the interim
Type. Labeling	In 1964, Towns together with Nikolai Basov and Aleksandr Prokhorov received the @AWARD@ Nobel Prize in Physics. @PERSON@ Vladimir Putin has already accepted his resignation, having appointed the first vice-governor Sergei Tsivilev to be the interim

4. DATASET

The NEREL dataset [20] was used for evaluation. NEREL was compiled from WikiNews texts in Russian. The dataset contains 29 different named entity types with a maximum nesting of six levels; the total number of annotated entities is more than 33000.

In the RuNNE competition [1], three classes were selected and reduced in number for the training sample, namely, *DISEASE*, *WORK_OF_ART*, *PENALTY*. For these entities, only 30–32 examples are given in the training sample.

To assess the quality of models, macro-averaged accuracy, recall, and F1-measure are used both for

Table 3. Examples of different types of prompts, with many entities

Type of hint	Examples
The outermost complete lexical	In the Research Institute of Pediatric Surgery and Traumatology, a kid was diagnosed with “DISEASE”: We have a girl 16 years old, admitted in the morning of October 11; the doctors made the diagnosis “DISEASE”
The innermost complete lexical	In the Research Institute of Pediatric Surgery and Traumatology, a kid was diagnosed with “DISEASE of bosom and belly”: We have a girl 16 years old, admitted in the morning of October 11; the doctors made the diagnosis “DISEASE of bosom”
The outermost complete framing	In the Research Institute of Pediatric Surgery and Traumatology, a kid was diagnosed with “[impact injury of bosom and belly DISEASE]”: We have a girl 16 years old, admitted in the morning of October 11; the doctors made the diagnosis “[impact injury of bosom DISEASE]”
The innermost complete framing	In the Research Institute of Pediatric Surgery and Traumatology, a kid was diagnosed with “[impact injury DISEASE] of bosom and belly”: We have a girl 16 years old, admitted in the morning of October 11; the doctors made the diagnosis “[impact injury DISEASE] of bosom”
Embedded complete framing	In the Research Institute of Pediatric Surgery and Traumatology, a kid was diagnosed with “[impact injury DISEASE] of bosom and belly DISEASE”: We have a girl 16 years old, admitted in the morning of October 11; the doctors made the diagnosis “[impact injury DISEASE] of bosom DISEASE”
The outermost complete labeling	In the Research Institute of Pediatric Surgery and Traumatology, a kid was diagnosed with “<DISEASE> impact injury of bosom and belly </DISEASE>”: We have a girl 16 years old, admitted in the morning of October 11; the doctors made the diagnosis “<DISEASE> impact injury of bosom </DISEASE>”
The innermost complete labeling	In the Research Institute of Pediatric Surgery and Traumatology, a kid was diagnosed with “<DISEASE> impact injury </DISEASE> of bosom and belly”: We have a girl 16 years old, admitted in the morning of October 11; the doctors made the diagnosis “<DISEASE> impact injury </DISEASE> of bosom”
Embedded complete labeling	In the Research Institute of Pediatric Surgery and Traumatology, a kid was diagnosed with “<DISEASE> <DISEASE> impact injury </DISEASE> of bosom and belly”: We have a girl 16 years old, admitted in the morning of October 11; the doctors made the diagnosis “<DISEASE> impact injury </DISEASE> of bosom </DISEASE>”
The outermost complete, type. Labeling	In the Research Institute of Pediatric Surgery and Traumatology, a kid was diagnosed with “@ DISEASE @ impact injury of bosom and belly”: We have a girl 16 years old, admitted in the morning of October 11; the doctors made the diagnosis “@ DISEASE @ impact injury of bosom”
The innermost complete, type. Labeling	In the Research Institute of Pediatric Surgery and Traumatology, a kid was diagnosed with “@ DISEASE @ impact injury of bosom and belly”: We have a girl 16 years old, admitted in the morning of October 11; the doctors made the diagnosis “@ DISEASE @ impact injury of bosom”
Embedded complete, type. Labeling	In the Research Institute of Pediatric Surgery and Traumatology, a kid was diagnosed with “@ DISEASE @ DISEASE @ impact injury of bosom and belly”: We have a girl 16 years old, admitted in the morning of October 11; the doctors made the diagnosis “@ DISEASE @ DISEASE @ impact injury of bosom”

Table 4. Results (macro-averaging, %) of different types of MRC prompts in the RuNNE dataset

Type of hint	General task			Few-shot task		
	precision	completeness	F1	precision	completeness	F1
Keyword	78.61	72.35	74.28	86.03	47.66	60.46
Definitions	78.76	72.44	74.31	80.62	50.77	61.21
Two most frequent examples	78.59	72.19	74.17	84.32	45.03	57.98
Five most frequent examples	79.23	71.58	73.89	84.76	47.29	58.98
Ten most frequent examples	78.13	70.64	73.09	81.60	45.56	56.96
Two most frequent components	78.65	73.05	74.63	86.15	49.07	60.80
Five most frequent components	78.54	72.77	74.62	83.39	48.35	60.30
Ten most frequent components	78.04	71.82	73.76	83.62	47.82	59.52
Contextual	75.97	67.05	69.39	86.74	49.83	62.53
One-lexical	79.94	71.69	74.30	87.81	49.05	62.34
Framing	80.69	71.47	74.43	89.12	47.85	61.83
Labeling	80.49	71.52	74.31	88.35	46.82	60.26
Type. Labeling	79.73	71.95	74.25	87.47	47.66	60.41
Outer complete lexical	80.18	72.11	74.54	89.94	50.33	63.88
Inner complete lexical	80.80	71.37	74.37	89.96	47.70	61.80
Outer complete framing	80.16	71.60	74.07	85.88	46.19	59.08
Inner complete framing	80.35	72.02	74.40	86.93	47.41	60.43
Embedded complete framing	80.74	71.93	74.59	91.03	48.27	62.40
Outer complete labeling	80.09	70.99	73.87	87.04	47.08	60.41
Inner complete labeling	79.86	71.92	74.26	89.16	48.01	61.15
Embedded complete labeling	79.30	71.85	73.95	87.89	44.72	58.22
Outer complete, type. Labeling	80.05	71.70	74.33	88.20	50.06	63.24
Inner complete, type. Labeling	80.36	71.63	74.34	90.48	46.74	61.00
Embedded complete, type. Labeling	79.09	71.42	73.90	86.11	46.75	59.76
The outermost complete Lexical with 3 contexts	79.62	71.75	74.15	85.96	46.87	59.93
The outermost complete Lexical + 2 components + definitions	79.53	73.14	75.00	85.09	50.99	62.46
The outermost complete Lexical + type. Labeling	79.73	71.37	73.74	88.75	48.56	62.02
The outermost complete Lexical + type. Labeling + contextual	79.32	70.76	73.39	85.67	50.37	62.22

new (few-shot NER task) and for all (general NER task) entity types.

5. EXPERIMENTS AND RESULTS

The RuBERT representation [13] is used as the basis for the MRC model. RuBERT was obtained by retraining the original BERT model [4] of the data from the Russian Wikipedia and the Russian-language news corpus. For the Russian language, using the Russian-language BERT model is usually more efficient than using its multilingual version.

The batch size was set to 16 and the maximum sequence length was 192 words. We train an MRC model for 16 epochs on eight GPUs against RuNNE data using various prompts.

All of the above approaches have been tried. In addition, as a baseline, we are testing the keyword approach of the MRC model, because this query is the easiest to generate and use when training the model, while it provides minimal information about the entity class. Table 4 presents the results obtained. In addition, Fig. 2 shows a chart of results for the top and base prompts.

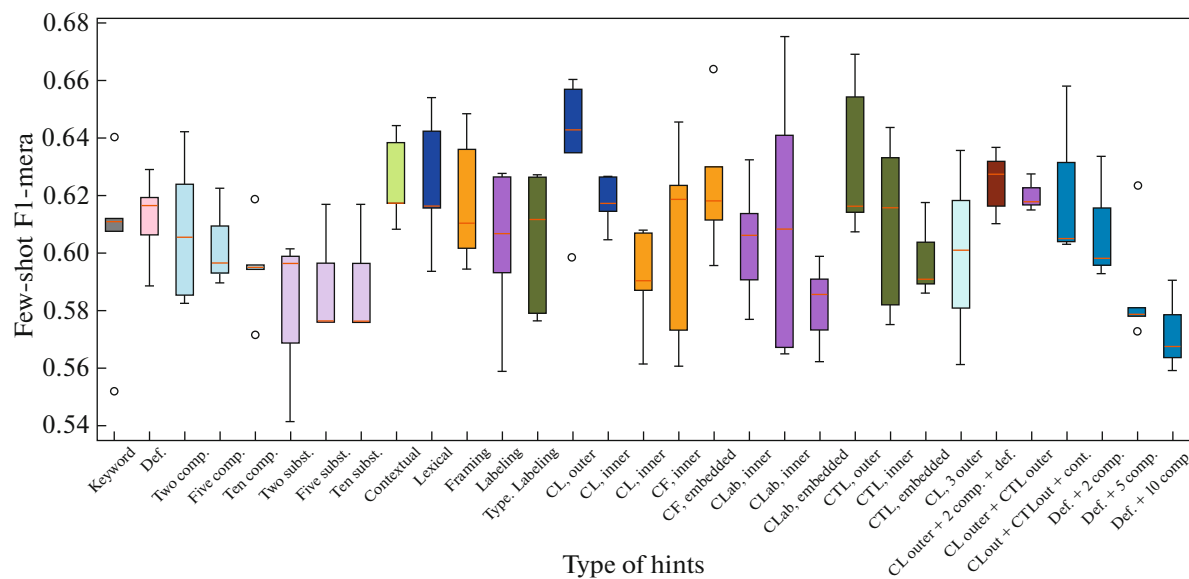


Fig. 2. The results of testing the basic and best prompts on the range chart. Here FL stands for full lexical prompt, FF stands for full frame prompt, FM stands for full marker prompt, FTM stands for full typed marker prompt, respectively.

On the basis of the presented results, the following conclusions can be drawn.

First, it can be seen that the simplest keyword prompt (in terms of generation) performs, on average, the same way as other prompts. It can be concluded that, in order to create prompts quickly and easily, but at the same time with decent quality, a keyword approach can be used.

Second, definitions help improve the quality of extraction of entities.

Third, among the component prompts, the best option was the two most frequent components prompt, which performs better than the keyword approach on both tasks. Adding more examples worsens the results.

Fourth, among the “single-entity” approaches, contextual prompt outperforms other methods in the few-shot problem, while producing the worst overall precision, recall, and F1-measures. Thus, contextual prompt helps the model predict rare entities, while it is not effective for a general task.

Fifth, among the “multientity” prompts, the outermost full lexical approach outperforms any other tested approaches, and the innermost lexical approach is worse than the single-lexical approach. Moreover, nesting for both labeling and typed labeling renders the model inefficient, in contrast to the nesting of the frame approach, where, conversely, it improves retrieval results. This is because the nested framing approach does not mix or shuffle entity tokens, unlike both labeling approaches, although the most external approach of full typed labeling comes second in terms of quality.

Finally, the “many prompts” approaches did not provide significant improvement in results, and in the few-shot problem, they were no better than other approaches. For example, the full lexical approach with the top three contexts performs worse than the original prompt. Thus, “overlay” of different contexts does not help the model to get more semantic information. Other multiple prompts that try to look from “other points of view” by combining different prompts do improve quality, though still underperforming compared to the outermost full lexical approach. Just the prompt *Outermost full lexical + 2 most frequent object components + definitions* gave some improvement in the results in the general problem; i.e., this approach of “different views” may be useful for achieving better results in future research.

CONCLUSIONS

The article presents various methods for generating questions (prompts) in named entity recognition systems based on questions, in particular, in a model of machine reading comprehension. We study the effect of prompts on the recognition of nested named entities in general and few-shot named entity recognition task. We test various approaches against the RuNNE-2022 test data on extracting nested named entities.

In the general task, many approaches, such as keywords, most frequent components, lexical and structural prompts, were comparable. The best result was achieved with the proposed combination of several prompts.

In the few-shot task, the proposed full lexical and full marker prompts turned out to be the best, taking

into account the nesting of named entities that mask or mark all external entities.

FUNDING

The project was supported by the Russian Science Foundation, grant no. 20-11-20166.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

1. E. Artemova, M. Zmeev, N. Loukachevitch, I. Rozhkov, T. Batura, P. Braslavski, V. Ivanov, and E. Tutubalina, "RuNNE-2022 shared task: Recognizing nested named entities," in *Computational Linguistics and Intellectual Technologies: Proc. Int. Conf. Dialogue 2022* (Moscow, 2022). <https://doi.org/10.28995/2075-7182-2022-21-33-41>
2. J. P. C. Chiu, and E. Nichols: "Named entity recognition with bidirectional LSTM-CNNs," *Trans. Assoc. Comput. Linguist.* **4**, 357–370 (2016). https://doi.org/10.1162/tacl_a_00104
3. R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. P. Kuksa: "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.* **12**, 2493–2537 (2011). arXiv:1103.0398 [cs.LG]
4. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova: "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proc. 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1: *Long and Short Papers*, Minneapolis, 2019 (Association for Computational Linguistics, 2019), pp. 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
5. M. Fan, L. Zhang, S. Xiao, and Yu. Liang, "Few-shot multi-hop question answering over knowledge base," *Wireless Commun. Mobile Comput.* **2022**, 8045535 (2022). <https://doi.org/10.1155/2022/8045535>
6. J. R. Finkel and Ch. D. Manning, "Nested named entity recognition," in *Proc. 2009 Conf. on Empirical Methods in Natural Language Processing*, Ed. by Ph. Koehn and R. Mihalcea (Association for Computational Linguistics, Singapore, 2009), pp. 141–150. <https://aclanthology.org/D09-1015>
7. A. Fritzler, V. Logacheva, and M. Kretov: "Few-shot classification in named entity recognition task," in *Proc. 34th ACM/SIGAPP Symp. on Applied Computing, Limassol, Cyprus, 2019* (Association for Computing Machinery, New York, 2019), pp. 993–1000. <https://doi.org/10.1145/3297280.3297378>
8. T. Gao, A. Fisch, and D. Chen, "Making pre-trained language models better few-shot learners," in *Proc. 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int. Joint Conf. on Natural Language Processing*, Vol. 1: *Long Papers*, Ed. by Ch. Zong, F. Xia, W. Li, and R. Navigli (Association for Computational Linguistics, 2021), pp. 3816–3830. <https://doi.org/10.18653/v1/2021.acl-long.295>
9. M. Hofer, A. Kormilitzin, P. Goldberg, and A. J. Nevado-Holgado, "Few-shot learning for named entity recognition in medical text," (2018). arXiv:1811.05468 [cs.CL]
10. Z. Huang, W. Xu, and K. Yu: "Bidirectional LSTM-CRF models for sequence tagging," (2015). arXiv:1508.01991 [cs.CL]
11. M. Ju, M. Miwa, and S. Ananiadou, "A neural layered model for nested named entity recognition," in *Proc. 2018 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1: *Long Papers*, Ed. by M. Walker, H. Ji, and S. Stent (Association for Computational Linguistics, New Orleans, 2018), pp. 1446–1459. <https://doi.org/10.18653/v1/N18-1131>
12. W. Jue, L. Shou, K. Chen, and G. Chen, "Pyramid: A layered model for nested named entity recognition," in *Proc. 58th Annu. Meeting of the Association for Computational Linguistics*, Ed. by J. Wang, L. Shou, K. Chen, and G. Chen (Association for Computational Linguistics, 2020), pp. 5918–5928. <https://doi.org/10.18653/v1/2020.acl-main.525>
13. Y. Kuratov and M. Arkhipov, "Adaptation of deep bidirectional multilingual transformers for Russian language," in *Computational Linguistics and Intellectual Technologies: Proc. Int. Conf. Dialogue 2019* (Moscow, 2019), pp. 333–339.
14. G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and Ch. Dyer, "Neural architectures for named entity recognition," in *Proc. 2016 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Ed. by K. Knight, A. Nenkova, and O. Rambow (Association for Computational Linguistics, San Diego, Calif., 2016), pp. 260–270. <https://doi.org/10.18653/v1/N16-1030>
15. D.-H. Lee, A. Kadakia, K. Tan, M. Agarwal, X. Feng, T. Shibuya, R. Mitani, T. Sekiya, J. Pujara, and X. Ren, "Good examples make a faster learner: Simple demonstration-based learning for low-resource NER," in *Proc. 60th Annu. Meeting of the Association for Computational Linguistics*, Vol. 1: *Long Papers* (Association for Computational Linguistics, Dublin, 2022), pp. 2687–2700. <https://doi.org/10.18653/v1/2022.acl-long.192>
16. D.-H. Lee, A. Kadakia, K. Tan, M. Agarwal, X. Feng, T. Shibuya, R. Mitani, T. Sekiya, J. Pujara, and X. Ren, "Good examples make a faster learner: Simple demonstration-based learning for low-resource NER," in *Proc. 60th Annu. Meeting of Association for Computational Linguistics*, Vol. 1: *Long Papers* (Association for Computational Linguistics, Dublin, 2022), pp. 2687–2700. <https://doi.org/10.18653/v1/2022.acl-long.192>
17. B. Lester, R. Al-Rfou, and N. Constant: "The power of scale for parameter-efficient prompt tuning," in *Proc. 2021 Conf. on Empirical Methods in Natural Language Processing*, Ed. by M.-F. Moens, X. Huang, L. Specia, and S. W. Yih (Association for Computational Linguistics, 2021), pp. 3045–3059. <https://doi.org/10.18653/v1/2021.emnlp-main.243>

18. X. L. Li and P. Liang, "Prefix-tuning: Optimizing continuous prompts for generation," in *Proc. 59th Annu. Meeting of the Association for Computational Linguistics and the 11th Int. Joint Conf. on Natural Language Processing*, Vol. 1: *Long Papers*, Ed. by Ch. Zong, F. Xia, W. Li, and R. Navigli (Association for Computational Linguistics, 2021), pp. 4582–4597. <https://doi.org/10.18653/v1/2021.acl-long.353>
19. X. Li, J. Feng, Y. Meng, Q. Han, F. Wu, and J. Li, "A unified MRC framework for named entity recognition," in *Proc. 58th Annu. Meeting of the Association for Computational Linguistics*, Ed. by D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault (Association for Computational Linguistics, 2020), pp. 5849–5859. <https://doi.org/10.18653/v1/2020.acl-main.519>
20. 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 *Proc. Int. Conf. on Recent Advances in Natural Language Processing (RANLP 2021)*, Ed. by R. Mitkov and G. Angelova (INCOMA, 2021), pp. 876–885.
21. X. Ma and E. Hovy: "End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF," in *Proc. 54th Annu. Meeting of the Association for Computational Linguistics*, Vol. 1: *Long Papers* (Association for Computational Linguistics, Berlin, 2016), pp. 1064–1070. <https://doi.org/10.18653/v1/P16-1101>
22. M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," (2018). arXiv:1802.05365 [cs.CL]
23. I. Rozhkov and N. Loukachevitch, "Machine reading comprehension model in RuNNE competition," in *Computational Linguistics and Intellectual Technologies: Proc. Int. Conf. Dialogue 2022* (Moscow, 2022). <https://doi.org/10.28995/2075-7182-2022-21-488-496>
24. T. Shibuya and E. Hovy, "Nested named entity recognition via second-best sequence learning and decoding," *Trans. Assoc. Comput. Linguist.* **8**, 605–620 (2020). https://doi.org/10.1162/tacl_a_00334
25. J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," *Adv. Neural Inf. Process. Syst.* **30** (2017). <https://proceedings.neurips.cc/paper/2017/file/cb8da6767461f2812ae4290eac7cbc42-Paper.pdf>
26. Ja. Straková, M. Straka, Ja. Hajic, "Neural architectures for nested ner through linearization," in *Proc. 57th Annu. Meeting of the Association for Computational Linguistics*, Ed. by A. Korhonen, D. Traum, and L. Màrquez (Association for Computational Linguistics, Florence, 2019), pp. 5326–5331. <https://doi.org/10.18653/v1/P19-1527>
27. X. Wang, Yo. Jiang, N. Bach, T. Wang, Zh. Huang, F. Huang, and K. Tu, "Automated concatenation of embeddings for structured prediction," in *Proc. 59th Annu. Meeting of the Association for Computational Linguistics and the 11th Int. Joint Conf. on Natural Language Processing*, Vol. 1: *Long Papers* (Association for Computational Linguistics, 2021), pp. 2643–2660. <https://doi.org/10.18653/v1/2021.acl-long.206>
28. Yi Yang and A. Katiyar, "Simple and effective few-shot named entity recognition with structured nearest neighbor learning," in *Proc. 2020 Conf. on Empirical Methods in Natural Language Processing*, Ed. by B. Webber, T. Cohn, Yu. He, and Ya. Liu (Association for Computational Linguistics, 2020), pp. 6365–6375. <https://doi.org/10.18653/v1/2020.emnlp-main.516>
29. W. Zhou and M. Chen, "An improved baseline for sentence-level relation extraction," in *Proc. 2nd Conf. Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th Int. Joint Conf. on Natural Language Processing*, Vol. 2: *Short Papers*, Ed. by Yu. He, H. Ji, S. Li, Ya. Liu, and Ch.-H. Chang (Association for Computational Linguistics, 2022), pp. 161–168.



Igor Sergeevich Rozhkov (born 2000) is a master's student at the Faculty of Computational Mathematics and Cybernetics of Moscow State University. Has three publications. Research interests: automatic text processing, named entity recognition, deep learning, automatic analysis of biomedical texts.



Natalia Valentinovna Loukachevitch (born 1964) graduated from the Faculty of Computational Mathematics and Cybernetics of Moscow State University, defended her dissertation for the degree of Doctor of Engineering Sciences in 2016. Currently, she is a leading researcher at the Research Computing Center of Moscow State University. She delivers lectures on automatic text processing and information retrieval at the Faculty of Computational Mathematics and Cybernetics and the Faculty of Philology of Moscow State University, as well as at the Bauman Moscow State Technical University. She has more than 300 publications in the field of automatic word processing, information retrieval, and knowledge representation. Research interests: automatic text processing, information retrieval, ontologies.