Machine Reading Comprehension Model in RuNNE Competition

Igor Rozhkov
Lomonosov Moscow State University,
Moscow, Russia
fulstocky@gmail.com

Natalia Loukachevitch
Lomonosov Moscow State University,
Moscow, Russia
louk_nat@mail.ru

Abstract

The paper studies machine reading comprehension model (MRC) (Li et al., 2020) in its application to extracting nested named entities (nested NER) in the RuNNE-2022 evaluation (Artemova et al., 2022). The model transforms named entity recognition tasks to a question-answering task. In this paper we compare several approaches to formulating "questions" for the MRC model such as entity type names (keywords), entity type definitions, most frequent examples for the train set, combinations of definitions and examples. We found that using two most frequent examples from the training set is comparable in quality of nested NER with gathering qualitative definitions from different dictionaries, which is much more complicated. In the RuNNE evaluation, the MRC model obtained the best results among models without any manual work (rules or additional manual annotation of texts).

Keywords: Nested named entities, RuNNE evaluation, Machine reading comprehension

Модель машинного понимания текстов (MRC) в тестировании RuNNE

Рожков И. С.
МГУ имени М.В. Ломоносова,
Москва, Россия
fulstocky@gmail.com

Лукашевич Н. В.
МГУ имени М.В. Ломоносова,
Москва, Россия
louk_nat@mail.ru

Аннотация

В статье исследуется модель машинного чтения (MRC) (Li et al., 2020) в ее применении для извлечения вложенных именованных сущностей в тестировании RuNNE-2022 (Artemova et al., 2022). Модель преобразует задачи распознавания именованных сущностей в задачи ответы на вопросы. В данной работе мы изучаем несколько подходов к формулированию «вопросов» для модели MRC. В тестировании RuNNE модель MRC показала лучшие результаты среди моделей, применяемых без какой-либо ручной работы (правил или дополнительной ручной аннотации текстов).

Ключевые слова: Вложенные именованные сущности, RuNNE, Модель машинного чтения

1 Introduction

Named entity recognition (NER) is one of the known task in natural language processing. Traditionally, NER task setting and datasets are devoted to extraction of so-called flat named entities, which presumes that a named entity cannot contain another named entity. For example, only one external ORGANIZATION entity should be extracted in *Lomonosov Moscow State University*, which leads to the loss of two internal named entity. During last years, due to the development of neural network models, the task of extracting nested named entities became much more frequent. Nested named entities allow for enhancing the coverage of found named entities, which is useful for such tasks as relation extraction, entity linking, knowledge graph population, etc. Specialized datasets are annotated with 2-6 levels of nestedness (Ringland et al., 2019; Plank et al., 2020; Loukachevitch et al., 2021). New NER methods specially devoted

to extracting nested named entities have been developed and significantly improved the performance in nested NER tasks (Shibuya and Hovy, 2020; Jue et al., 2020; Yu et al., 2020).

For Russian, two datasets annotated with nested named entity exist. The first dataset, FactRuEval (Starostin et al., 2016), is quite small for training machine learning models. Recently, new dataset NEREL (Loukachevitch et al., 2021) with nestedness up to 6 levels has been created. The NEREL dataset became a basis for organization of RuNNE-2022 evaluation (Artemova et al., 2022), devoted to recognition of nested named entities and also few-shot setting of nested NER.

In this paper we describe an approach applied to the RuNNE tasks, which is based on machine reading comprehension model (MRC) (Rajpurkar et al., 2016; Li et al., 2020). The model transforms NER tasks to question-answering tasks and achieve state-of-the art results on various NER datasets. We compare several approaches to formulating "questions" for the MRC model such as entity type names (keywords), entity type definitions, most frequent examples for the train set, combinations of definitions and examples. We found that using two most frequent examples from the training set is comparable in quality of nested NER with gathering of qualitative definitions from different dictionaries, which is much more complicated. In the RuNNE evaluation, the MRC model obtained the best results among models without any manual work (rules or additional manual annotation of texts).

2   Related Work

Early works regarding nested NER involved mainly hybrid methods that combined rules with supervised learning algorithms (Shen et al., 2003; Zhang et al., 2004). Another approach to the nested NER task relies on hand-crafted features (Alex et al., 2007; Muis and Lu, 2018). These methods mostly failed to take advantage of the dependencies among nested entities.

Later, LSTM-based models were developed to process nested named entities. LSTM-CRF model (Ju et al., 2018) was already able to capture context representation of input sequences and globally decode predicted labels for nested entities even of the same entity type. Dynamically stacked multiple layers recognize outer entities by taking full advantage of information encoded in their corresponding inner entities. Straková et al. (Straková et al., 2019) identify nested named entities by a seq2seq model exploring combinations of different context-based embeddings (ELMo, BERT, Flair). Sohrab and Miwa (Sohrab and Miwa, 2018) proposed to concatenate the LSTMs outputs for the start and end positions of spans and then calculate a score for each span. In Biaffine model, Yu et al. (Yu et al., 2020) demonstrated that the model provides a global view on the input and performs better results – the model scores pairs of start and end tokens to form a named entity. Pyramid model (Jue et al., 2020) consists of a stack of inter-connected layers. Each layer l predicts whether a l-gram is a complete entity mention. The Second-best Sequence Learning coupled with Decoding (Second Best) model (Shibuya and Hovy, 2020) uses the Conditional Random Field output layer. The model treats the tag sequence for nested entities as the second best path within the span of their parent entity. In addition, the decoding method for inference extracts entities iteratively from outermost ones to inner ones in an outside-to-inside way.

Machine Reading Comprehension (MRC) (Rajpurkar et al., 2016) treats the nested NER as a question-answering task (Li et al., 2020), when for each named entity type, a specialized question is created. The model should find answers to the questions in a sentence, which is equivalent to extracting corresponding named entities. In (Loukachevitch et al., 2021), several models (Biaffine, MRC, Pyramid) were studied for extracting nested named entities in Russian. The best results were obtained by the MRC model.

3   Machine Reading Comprehension Model

The MRC model treats the NER task as extracting answer spans to specialised questions, each entity type is associated with a specific question. The dataset sentences are converted into triples (Question Q, Answer A, Context C). Question Q is either generated or selected supplementary sequence (described below); the Answer A is the annotated named entity, the subsequence of the given sentence; the Context C is the given sentence. The MRC model is constructed over the BERT (Devlin et al., 2018) model, which obtains the following string as an input:

$$\{[CLS], q_1, q_2..., q_m, [SEP], t_1, t_2, ...t_n\}$$

where q_i are words of the question sequence, t_i are words of the given sentence, $[CLS]$ and $[SEP]$ are special tokens of the BERT model. The MRC model should extract a continuous span A in the context C:

$$A = \{t_i..., t_{i+k}, 1 \leqslant i \leqslant i + k \leqslant n\}$$

such that A is now a retrieved named entity.

The model backbone is as follows. BERT, given aforementioned input, outputs a context representation matrix $E \in \mathbb{R}^{n \times d}$, where d is the size of last layer dimension of BERT. The query part of output is dropped.

Next, given matrix E, model first predicts the probabilities of each word to be start index, to be end index and then probability of each start-end indices pair to be matched onto one named entity.

In more detail: model first predicts two values, P_{start} and P_{end} as follows:

$$P_{start} = softmax_{eachrow}(E \cdot T_{start}) \in \mathbb{R}^{n \times 2}$$

$$P_{end} = softmax_{eachrow}(E \cdot T_{end}) \in \mathbb{R}^{n \times 2}$$

where $T_{start}, T_{end} \in \mathbb{R}^{n \times 2}$ are the weights learned. Then for $\hat{I}_{start}$ and $\hat{I}_{end}$ sets

$$\hat{I}_{start} = \{i \mid argmax(P_{start}^{(i)}) = 1, \ i = \overline{1, n}\}$$

$$\hat{I}_{end} = \{j \mid argmax(P_{end}^{(j)}) = 1, \ j = \overline{1, n}\}$$

where superscripts (i) and (j) denote i-th and j-th row of a matrix respectively binary classification model is trained to predict value of matching probability:

$$P_{i_{start}, j_{end}} = sigmoid(M \cdot concat(E_{i_{start}}, E_{j_{end}}))$$

where $M \in \mathbb{R}^{1 \times 2d}$ is weights learned. Now this value predicts whether each occured span i_{start}, j_{end} in the context C is a desired answer A, i.e. named entity of given type.

There are different approaches to creating questions. (Li et al., 2020) proposed several of them:

- **Position index:** question is generated based on the position index of given tag, i.e. "first", "second", etc. or "one", "two", etc.
- **Keyword usage:** question is given or generated keyword describing tag, e.g. "*profession*", *"person"*.
- **Rule-based template filling:** generates a sequence from given template, e.g. "*Find named entities of type "person" in the given sentence.*".
- **Wikipedia definition retrieval:** question is generated with Wikipedia definition of a given tag, e.g. "*An organization is an entity comprising multiple people, such as an institution or an association.*".
- **Synonyms:** words that have the same or close meaning to the original tag, e.g. for tag "*profession*" that would be "*occupation*", "*job*", etc.
- **Concatenation of keyword and synonyms:** question is constructed from both keywords and synonyms, e.g. "*profession, occupation, job*".
- **Annotation guideline notes:** the guidelines of labeled entities provided by the dataset builder, e.g. for *location* it could be "*Find locations in the text including nongeographical locations, mountain ranges and bodies of water*".

The last approach achieves best results in the original work.

4   RuNNE task and data

RuNNE competition (Artemova et al., 2022) sets the few-shot version of the nested named entity recognition task. While most of the entities have considerable number of examples in the training set, several others occur much less frequently: the amount of such entities is limited in the training set. Dev and test

NE type	Number of mentions	
	train	test
PROFESSION	4566	848
PERSON	4517	961
ORGANIZATION	4049	675
EVENT	2850	683
COUNTRY	2521	456
DATE	2268	523
CITY	1101	239
NUMBER	1026	230
ORDINAL	565	107
AGE	554	138
NATIONALITY	394	66
LAW	389	61
FACILITY	371	63
STATE_OR_PROVINCE	343	112
AWARD	322	119
IDEOLOGY	300	43
LOCATION	270	62
PRODUCT	237	53
CRIME	180	35
MONEY	171	43
TIME	154	47
DISTRICT	98	25
RELIGION	94	24
PERCENT	82	7
LANGUAGE	43	8
DISEASE	32	57
PENALTY	32	17
WORK_OF_ART	30	88
FAMILY	17	14
Total amount	**27576**	**5804**

Table 1: Number of entities in the RuNNE training and test sets.

sets both contain the usual (non-limited) amount. Therefore, the main goal of the competition is to create models capable of retrieving both common and uncommon named entity types.

The dataset of RuNNE evaluation was created from the NEREL dataset (Loukachevitch et al., 2021). This data was collected from WikiNews texts in Russian language, manually labeled by the annotators of the NEREL dataset using the *brat* annotation tool (Stenetorp et al., 2012). After that the initial dataset was mixed and split into train, dev and test sets. The dataset contains 29 different named entity types, with maximum nestedness of 6 levels.

For the few-shot task formulation, three classes were chosen and decreased in the amount for the training set, namely *disease*, *work_of_art* and *penalty*. Table 1 shows the amount of each labeled entity type both in the training and test sets.

As a result, we can see that the classes are not balanced, and there are no more than 32 mentions of the aforementioned entity types in the training set. Moreover, other types have similar amount of the mentions, e.g. *language* has 43 mentions, while *family* has even less - 17.

For evaluation on the RuNNE dataset, the macro-average precision, recall and F1 both for only new (few-shot NER task) and all (general NER task) entity types are used.

5   Approach and Results

In this work we study what approaches to question generation help the MRC model in few-shot and general NER tasks [1].

Though annotation guidelines allows achieving the best results in the original work, they do not always exist for some dataset. Sometimes it is quite difficult to retrieve or generate such. In our case, the RuNNE dataset was not provided with annotation guidelines, and thus we cannot employ this approach. Therefore, aside from previously described approaches, this work proposes few new techniques for question generation:

- **Definition selection.** The questions are definitions of entity types, carefully selected from multiple dictionaries.
- N **most frequent examples.** The N most frequent examples of entities are obtained from the given training set and questions are generated. Number N is pre-defined from the start.
- N **most frequent entity components**. Each entity example in the training set is split into single words and then lemmatized. After that the N most frequent words are retrieved, which then compose the question.
- **Concatenation of definitions and most frequent examples**.

Examples of dictionary definitions are as follows (translated from Russian): "***Age*** *is the period of time when someone was alive or something existed.*"; "*A **city** is a place where many people live, with many houses, shops, businesses, etc.*"

Question examples of N most frequent examples (here we presume $N = 5$) are as follows (translated from Russian): "***Date*** *is an entity such as in Monday, in Tuesday, today, in year 2011, in year 2004.*". "***Law*** *is an entity such as Constitution, CC (Criminal Code), CC of RF, Yarovaya package, constitution.*"

Question examples of N most frequent entity components ($N = 5$) are as follows (translated from Russian): "***Disease*** *is an entity such as cancer, hurt, heart, heart as adjective, pain*". "***City*** *is an entity such as moscow, moskovsky, london, new-york, kyiv.*"

In this work, we study the results of utilizing following aforementioned approaches to question generation:

- **Keyword usage**
- **Definition selection**
- N **most frequent examples, for** $N = 2, 5, 10$.
- N **most frequent entity components, for** $N = 2, 5, 10$.
- **Concatenation of definitions and most frequent examples**

As baseline we use following models:

- **RuBERT** (Kuratov and Arkhipov, 2019): Baseline model of the RuNNE competition.
- **2nd-best-path-RuBERT** (Shibuya and Hovy, 2020): treats the tag sequence as the second best path within in the span of their parent entity based on RuBERT.

We utilize RuBERT (Kuratov and Arkhipov, 2019) as a basis for MRC model. We use the batch size of 32 and learn the MRC model for 16 epochs on the 8 GPUs over the RuNNE data. We use AdamW optimizer (Loshchilov and Hutter, 2017), with $\beta_1 = 0.9, \beta_2 = 0.98, \varepsilon = 10^{-8}$. RuBERT configuration was set to default values after (Kuratov and Arkhipov, 2019). We use OneCycleLR learning rate scheduler (Smith and Topin, 2019) with maximum learning rate of $2 * 10^{-5}$, final div factor $= 10^4$, linear anneal strategy. Weight decay was set to 0.01. Other hyperparameters were set to default values.

Table 2 shows experimental results on the RuNNE dataset. We can see that using two most frequent entity components from the training set is even slightly better than using well-constructed entity definitions. For few-shot setting, the results based on definitions are slightly better, but the extracting frequent entity components is much simpler than gathering well-written definitions from various dictionaries. The increase in the number of components leads to a decrease in quality of name extraction. Furthermore, we can see that using original entity examples and not split ones shows lower results. Though this approach is even simpler than previous one, it acts poorer. Also combinations of definitions and examples do not

---

[1]https://github.com/fulstock/mrc_nested_ner_ru

improve name recognition in both settings. Moreover, we indicate original results in the RuNNE competition, ours among the others. The resulting score is different due to the randomness of some model's parts, which generated different score.

Table 3 shows examples chosen as most frequent ones for each named entity type. Moreover, we provide the amount of these examples in the test set. As we can see, for the best approach (2 entity components) this amount is significantly low. Hence we can conclude that the MRC model does not memorize these components during learning procedure.
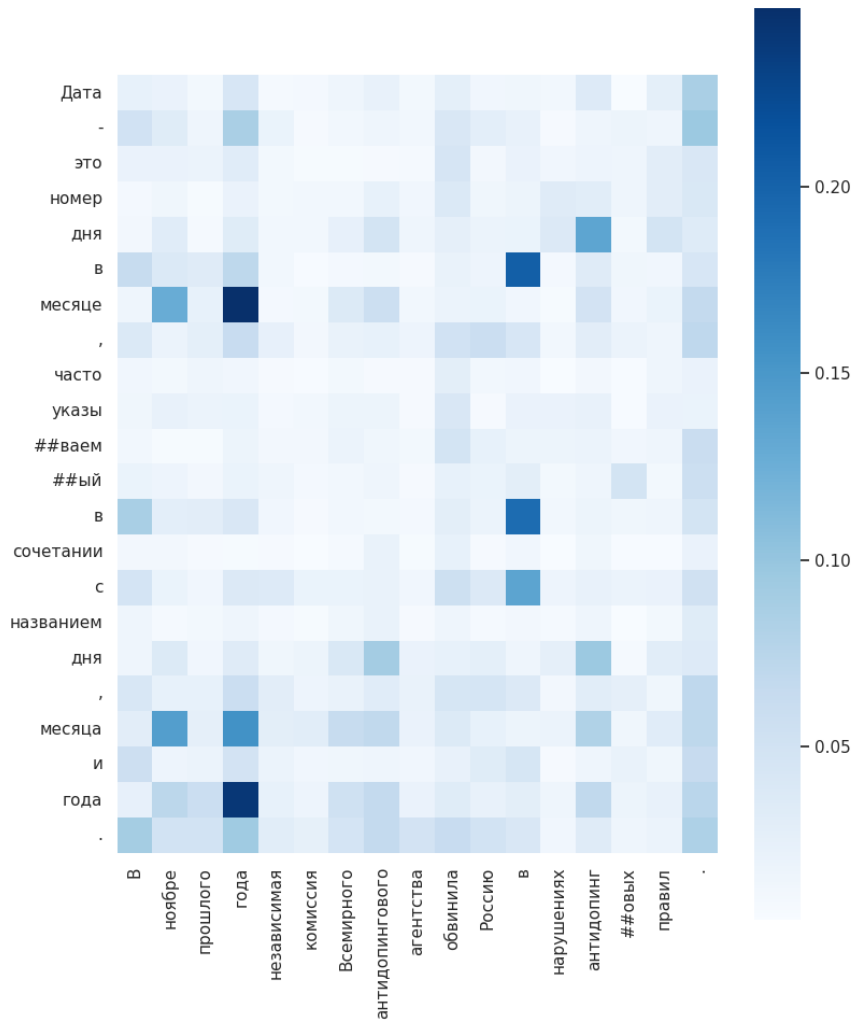

Figure 1: Attention layers inside RuBERT.

In addition, we visualize the attention layers inside RuBERT of the MRC model (Figure 1). The higher the score, the more semantically similar (in terms of attention) the words are. We see that similarities between definition and context words can be captured in the attention matrices, therefore model is able to gain knowledge of the given entity type.

6   Conclusion

In this paper, we studied machine reading comprehension model (MRC) in its application to extracting nested named entities in RuNNE-2022 evaluation. The model was applied in general nested NER

Model and approach		General Task			Few-shot Task		
		Precision	Recall	F1	Precision	Recall	F1
RuBERT-Tagger		-	-	67.44	-	-	44.66
2nd-best-path-RuBERT		74.83	61.78	67.68	77.61	09.77	17.36
MRC	Keyword	78.27	71.92	73.79	**88.09**	45.22	59.02
	Definitions	78.76	72.44	74.31	80.62	**50.77**	**61.21**
	2 most frequent examples	78.59	72.19	74.17	84.32	45.03	57.98
	5 most frequent examples	**79.23**	71.58	73.89	84.76	47.29	58.98
	10 most frequent examples	78.13	70.64	73.09	81.60	45.56	56.96
	2 most fr. entity components	78.65	**73.05**	**74.63**	86.15	49.07	60.80
	5 most fr. entity components	78.54	72.77	74.62	83.39	48.35	60.30
	10 most fr. entity components	78.04	71.82	73.76	83.62	47.82	59.52
	Def. + 2 most frequent ex.	78.37	71.74	73.96	80.21	49.89	60.83
	Def. + 5 most frequent ex.	77.83	72.62	74.26	78.47	48.71	58.69
	Def. + 10 most frequent ex.	77.60	71.36	73.22	82.50	45.68	57.24
RuNNE	pullenti	-	-	81.12	-	-	71.03
	MSU-RCC (ours)	-	-	74.93	-	-	60.39
	SibNN	-	-	74.25	-	-	40.37
	user:abrosimov_kirill	-	-	74.08	-	-	64.41

Table 2: Results (macro-averaged, %), compared with other models of the RuNNE competition.

task and the few-shot setting. The MRC model transforms named entity recognition tasks to question-answering tasks. We compared several approaches to formulating "questions" for the MRC model such as entity type names (keywords), entity type definitions, most frequent entity components and most frequent examples for the training set, combinations of definitions and examples. We found that using two most frequent entity components from the training set is even slightly better than using well constructed entity definitions. For few-shot setting, the results based on definitions are slightly better, but the extracting frequent entity components is much simpler than gathering well-written definitions from dictionaries.

In the RuNNE evaluation, the MRC model utilizing definitions as questions obtained the best results among machine learning models used without additional manual annotation of training texts.

Acknowledgments

References

Beatrice Alex, Barry Haddow, and Claire Grover. 2007. Recognising nested named entities in biomedical text. // *Biological, translational, and clinical language processing*, P 65–72.

Ekaterina Artemova, Maksim Zmeev, Natalia Loukachevitch, Igor Rozhkov, Tatiana Batura, Pavel Braslavski, Vladimir Ivanov, and Elena Tutubalina. 2022. RuNNE-2022 Shared Task: Recognizing Nested Named Entities. *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog"*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Meizhi Ju, Makoto Miwa, and Sophia Ananiadou. 2018. A neural layered model for nested named entity recognition. // *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, P 1446–1459.

Wang Jue, Lidan Shou, Ke Chen, and Gang Chen. 2020. Pyramid: A layered model for nested named entity recognition. // *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, P 5918–5928.

Yuri Kuratov and Mikhail Arkhipov. 2019. Adaptation of deep bidirectional multilingual transformers for russian language.

Xiaoya Li, Jingrong Feng, Yuxian Meng, Qinghong Han, Fei Wu, and Jiwei Li. 2020. A unified MRC framework for named entity recognition. // *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, P 5849–5859.

Ilya Loshchilov and Frank Hutter. 2017. Fixing weight decay regularization in adam. *ArXiv*, abs/1711.05101.

Natalia Loukachevitch, Ekaterina Artemova, Tatiana Batura, Pavel Braslavski, Ilia Denisov, Vladimir Ivanov, Suresh Manandhar, Alexander Pugachev, and Elena Tutubalina. 2021. Nerel: A russian dataset with nested named entities, relations and events. // *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, P 876–885.

Aldrian Obaja Muis and Wei Lu. 2018. Labeling gaps between words: Recognizing overlapping mentions with mention separators. *arXiv preprint arXiv:1810.09073*.

Barbara Plank, Kristian Nørgaard Jensen, and Rob van der Goot. 2020. Dan+: Danish nested named entities and lexical normalization. // *Proceedings of the 28th International Conference on Computational Linguistics*, P 6649–6662.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. Squad: 100, 000+ questions for machine comprehension of text. *CoRR*, abs/1606.05250.

Nicky Ringland, Xiang Dai, Ben Hachey, Sarvnaz Karimi, Cecile Paris, and James R Curran. 2019. Nne: A dataset for nested named entity recognition in english newswire. // *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, P 5176–5181.

Dan Shen, Jie Zhang, Guodong Zhou, Jian Su, and Chew Lim Tan. 2003. Effective adaptation of hidden markov model-based named entity recognizer for biomedical domain. // *Proceedings of the ACL 2003 workshop on Natural language processing in biomedicine*, P 49–56.

Takashi Shibuya and Eduard Hovy. 2020. Nested named entity recognition via second-best sequence learning and decoding. *Transactions of the Association for Computational Linguistics*, 8:605–620.

Leslie N. Smith and Nicholay Topin. 2019. Super-convergence: very fast training of neural networks using large learning rates. // *Defense + Commercial Sensing*.

Mohammad Golam Sohrab and Makoto Miwa. 2018. Deep exhaustive model for nested named entity recognition. // *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, P 2843–2849.

Anatoly S Starostin, Victor V Bocharov, Svetlana V Alexeeva, Anastasiya A Bodrova, Alexander S Chuchunkov, SS Dzhumaev, Irina V Efimenko, Dmitry V Granovsky, Viktor F Khoroshevsky, Irina V Krylova, et al. 2016. Factrueval 2016: evaluation of named entity recognition and fact extraction systems for russian. *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog"*, P 702–720.

Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. brat: a web-based tool for NLP-assisted text annotation. // *Proceedings of the Demonstrations Session at EACL 2012*, Avignon, France, April. Association for Computational Linguistics.

Jana Straková, Milan Straka, and Jan Hajič. 2019. Neural architectures for nested ner through linearization. *arXiv preprint arXiv:1908.06926*.

Juntao Yu, Bernd Bohnet, and Massimo Poesio. 2020. Named entity recognition as dependency parsing. // *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, P 6470–6476.

Jie Zhang, Dan Shen, Guodong Zhou, Jian Su, and Chew-Lim Tan. 2004. Enhancing hmm-based biomedical named entity recognition by studying special phenomena. *Journal of biomedical informatics*, 37(6):411–422.

NE type	10 most frequent entity components	Mentions in the test, %		
		2 ex.	5 ex.	10 ex.
AGE	год, 42-летний, 40-летний, 31-летний, 55-летний, 60-летний, 25-летний, 62-летний, 50, годовщина	20.81	22.84	27.41
AWARD	премия, нобелевский, мир, чемпион, медаль, чемпионка, золотой, за, год, олимпийский	11.82	23.92	30.55
CITY	москва, московский, лондон, нью-йорк, киев, санкт-петербург, римский, доха, бостон, столица	11.57	16.12	20.66
COUNTRY	россия, сша, российский, рф, украина, американский, израиль, великобритания, федерация, япония	20.34	29.56	39.83
CRIME	убийство, коррупция, домогательство, преступление, сексуальный, нарушение, незаконный, связь, насилие, шпионаж	08.43	09.64	15.66
DATE	год, в, 2016, 2013, 2017, декабрь, октябрь, июнь, 2012, день	28.91	32.37	39.82
DISEASE	рак, ушибить, сердце, сердечный, боль, брюшной, полость, отравление, грудной, клетка	07.45	14.89	17.02
DISTRICT	район, округ, военный, северокавказский, московский, федеральный, химкинский, косовский, скфо, сибирский	18.18	18.18	18.18
EVENT	выборы, отставка, назначить, погибнуть, чемпионат, родиться, пост, задержать, человек, арестовать	01.58	04.73	08.41
FACILITY	памятник, гора, храмовый, собор, улица, дом, здание, аэропорт, площадь, дворец	00.00	01.95	09.09
FAMILY	семья, королевский, романов, дом, бекхэм, abanyiginya, хантсмен, монарший, обама	20.00	20.00	20.00
IDEOLOGY	демократ, республиканец, демократический, оппозиционный, террорист, консерватор, коммунистический, социалистический, оппозиция, левый	06.12	08.16	30.61
LANGUAGE	английский, русский, арабский, французский, испанский, немецкий, итальянский, чувашский, марийский, татарский	25.00	37.50	50.00
LAW	закон, о, конституция, кодекс, ук, статья, ., рф, «, федеральный	06.02	17.67	34.96
LOCATION	европа, европейский, западный, море, запад, северокавказский, берег, америка, иордан, южный	12.86	14.29	17.14
MONEY	доллар, млн, рубль, миллиард, миллион, млрд, сша, тысяча, евро, ..	22.39	31.34	49.25
NATIONALITY	россиянин, гражданин, американец, российский, американский, серб, русский, канадец, сша, афроамериканец	04.48	23.88	28.36
NUMBER	два, тысяча, четыре, один, около, три, двое, шесть, 1, трое	08.98	16.02	24.22
ORDINAL	первый, второй, третий, пятый, ii, четвёртый, xvi, v, 1, 2	47.27	55.45	63.64
ORGANIZATION	правительство, россия, совет, полиция, парламент, партия, рф, мвд, госдума, комитет	03.13	06.19	09.25
PERCENT	%, процент, 1, 30, 90, 50, 20, 75, 49, 24	40.00	46.67	46.67
PERSON	владимир, путин, сергей, александр, дмитрий, обама, медведев, виктор, кастро, андрей	02.28	06.56	07.90
PENALTY	казнь, штраф, тюрьма, заключение, год, смертный, 20, въезд, пожизненный, денежный	17.65	29.41	29.41
PRODUCT	интернет, твиттер, facebook, сеть, википедия, youtube, союз, twitter, tumblr, як-18т	09.52	20.63	26.98
PROFESSION	президент, глава, министр, премьер-министр, губернатор, россия, директор, депутат, председатель, «	07.28	11.32	15.95
RELIGION	мусульманин, исламский, православный, мусульманский, ислам, католический, католик, христианин, баптистский, шиитский	25.00	58.33	58.33
STATE_OR_PROVINCE	область, край, каталония, техас, чечня, калининградский, архангельский, крым, калифорния, массачусетс	07.04	08.45	16.20
TIME	час, минута, год, время, вечером, ночь, в, утром, около, местный	08.21	18.66	30.60
WORK_OF_ART	рим, друг, список, шиндлера, старый, спасатель, малиб, проповедь, падение, le	00.00	00.00	00.00

Table 3: Most frequent components of each entity type in the training set (in decreasing order), and their corresponding amount in the test set.