Записки научных
семинаров ПОМИ
Том 499, 2021 г.

D. Mazitov, I. Alimova, E. Tutubalina

NAMED ENTITY RECOGNITION IN RUSSIAN USING MULTI-TASK LSTM-CRF

Abstract. Named entity recognition (NER) is aimed at obtaining the important information from the unstructured data presented in the form of natural language texts. In this paper, we investigate the efficiency of modern multi-task NER approach on Russian corpora by employing several different NER datasets and a dataset of part-of-speech (POS) tags. We apply a state-of-the-art neural architecture based on bidirectional LSTMs and conditional random fields. Convolutional neural networks were utilized to learn character-level features. We carry out an extensive experimental evaluation over three standard datasets of news written in Russian. The proposed multi-task model achieve states-of-the-art results with an F1 score of 88.04% on Gareev's dataset and an F1 score of 99.49% on Person-1000 dataset.

§1. Introduction

Named entity recognition (NER) is one of the essential tasks in the field of natural language processing (NLP). The main goal of NER is to extract and classify important named entities for a particular task. Named entities are words or phrases denoting a specific object, for example, human names, locations, organizations, facilities, products, dates, geopolitical entities, holidays. NER is applied in many areas related to NLP and extraction information, such as for information retrieval, question answering systems, text classification, relation extraction and etc. Extracted named entities help to understand the subject of the text and find keywords. The continued growth of unstructured information represented with text in a natural language and increasing requirement for extracting structured information from them make NER task is actual and important.

Different approaches were proposed for NER varying from simple lexicon-based approaches to more complicated neural network models. The aim of this work is an evaluation of a multi-task approach for NER task

Key words and phrases: named entity recognition and NER and LSTM and CRF and multi-task learning.

This work was supported by the Russian Science Foundation grant no. 20-11-20166.

on the Russian language. In this paper, we applied model based on state-of-the-art approach presented by Lample et al. [21]. The model is based on a combination of Bidirectional Long Short Term Memory (BiLSTM) and Conditional Random Field (CRF). The model has obtained the best results on corpora of various domains and languages [8, 12, 22, 24, 28, 31, 35, 37]. In order to evaluate the model efficiency, we conducted extensive experiments on three existing Russian annotated corpora collected from different resources and compared the results with the existing models.

The rest of the paper is structured as follows. In Section 2, we discuss related work. In Section 3, we describe our model for NER with char-level embedding features. Section 4 provides experiments and evaluation results. Section 5 concludes this paper.

§2. Related work

Most of the early works for the Russian language NER describe systems based on linguistic resources that have been carefully worked out by hand: dictionaries, templates, and rules [7, 10, 27]. Popov et al. describes the adaptation of the vocabulary approach for the Russian language [27]. Craidlin introduced the TagLite program, which aims to distinguish named groups consist of three types of proper names: persons, organizations and geographical objects [10]. The system includes the following dictionaries: proper names, generic concepts of investigated entity types (director, river, office) and other auxiliary words that can be part of target noun groups. In order to resolve the ambiguity and process words that are not encountered in dictionaries, the rule-based вЂњpredictorвЂќ module is applied. The module predicts the label of a named group based on the analysis of others found in the same document. The authors evaluated the quality of the system on their own annotated corpus. TagLite obtained 85.8% of F-measure for all categories of named groups. Brykina et al. proposed a system that recognizes named entities based on lists of terms from the input ontology and resolving polysemy with a set of manually developed rules and dictionaries of context words [7]. The authors evaluated the efficiency of a system on their own corpus, considering only entities included in the ontology. The system obtained F-measure varying from 91% to 98% for different types of entities. Both systems were evaluated on closed corpora, which makes it difficult to conduct a comparative analysis of the achievements in this area.

In the later works devoted to NER, machine learning methods on Russian texts started to apply [5,13,26]. The most commonly utilized method was Conditional Random Fields (CRF) [20]. Antonova et al. applied a CRF model to their own annotated corpora consisting of news feed texts [5]. There were five types of annotations: person names, geographical objects, organization, products, and events. The authors also evaluated different types of optimizers for CRF. The highest F-measure obtained by this approach was 87.18%. Podobryaev applied CRF model to person recognition and used information from ontology as one of the features [26]. The quality of the proposed approach was evaluated on a manually annotated corpora. Gareev et al. developed annotated corpus of Russian-language texts for evaluating NER methods and compared the effectiveness of two approaches [13]. The first approach is based on dictionaries of names and rules, which analyze the context of a named entity and compare the set of references to the same entity in a document. The second is based on the CRF model with various features. The developed corpus is publicly available and contains two types of annotations: persons and organizations. The results of experiments showed that CRF based approach outperformed knowledge-based approach on 13% of F-measure. Mozharova and Lukashevich investigated the features for the CRF model in NER task [25]. The authors found out the large contribution of the context bigram feature and the low contribution of knowledge-based features.

Recent works in the field of NER task devoted to utilizing neural network models. In particular, different neural network architectures are applied to English language data sets and are achieved a state-of-the-art results [2, 16, 21, 22]. There is only a few works devoted to Russian language [3,4,18,19]. Anh et al. investigated the quality of three models starting from Bi-directional Long Short Term Memory (Bi-LSTM) then supplementing it with CRF and finally adding external word embeddings [3]. The models were evaluated on three Russian language data sets Gareev's, Person-1000 and FactRuEval 2016. The results showed that the combination of Bi-LSTM and CRF models significantly increased the quality of entity recognition and external word embeddings allowed to achieve state-of-the-art for the Russian NER task.

An overview of previous work shows that methods for NER are actively developed and improved. Our study aims to fill a gap in applying modern neural network models for NER task in Russian language and evaluate the effectiveness of the multi-task approach.

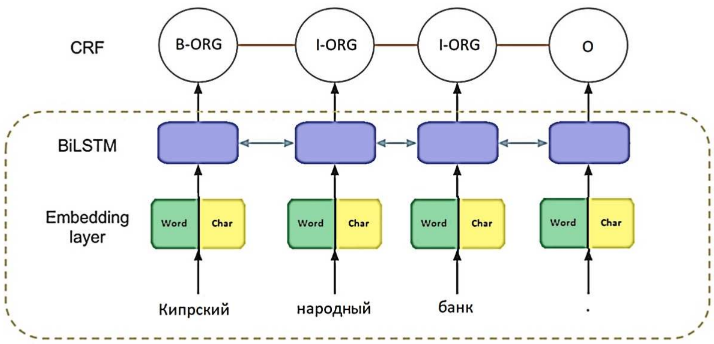

Figure 1. The overall architecture of NER model.

§3. Model architecture

In the following section, we describe the architecture of the proposed model. The model consists of three main parts: input layers, bidirectional LSTM (BiLSTM) for input representation and CRF for label prediction. The overall architecture of the model is presented in Fig. 1. We implemented our model with TensorFlow library [1]. The source code of our model can be found in a repository[1].

3.1. Model Input. The model takes as an input concatenation of word and character level embeddings for each word. Word embedding vectors were obtained from word2vec model trained on a publicly available corpus of unlabeled Russian news [4]. The corpus consists of 635,000 news from Russian online news resource lenta.ru[2]. The corpus contains around 46 million words. The vocabulary of word vector representation model is 376,000 words, vector dimension is 100.

The character level embeddings are obtained with Convolutional Neural Network (CNN) implemented from [9]. The CNN extracts character features for each word. The input of the network is word represented with randomly initialized vectors of length 25 concatenated with vectors of length

[1]`https://bitbucket.org/cyberdan7/ner/src/master/`
[2]`https://github.com/yutkin/lenta.ru-news-dataset`

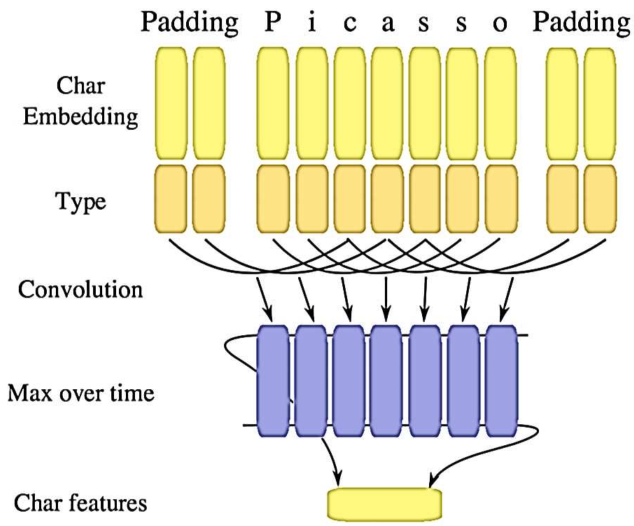

Figure 2. The architecture of CNN for char embedding.

10 identifying a type of the letter (capital or not) for each letter. We applied 4 filters with 3, 4, 5, and 7 sizes of kernels. The model's output vector length is 50. The overview of the model is presented in Fig. 2.

3.2. Bidirectional LSTM. LSTM is a kind of Recurrent Neural Network (RNN). RNNs are naturally used for sequence learning, where both input and output are word and label sequences. RNN has recurrent hidden states, which aim to simulate memory, i.e., the activation of a hidden state at every time step depends on the previously hidden state. The recurrent unit computes a weighted sum of the input signal. Training RNNs to capture long-term dependencies is difficult due to the effect of vanishing gradients, so the most widely used modification of RNN units is the Long Short-Term Memory (LSTM) [15] that provides the "constant error carousel" and does not preclude free gradient flow. The most common LSTM architecture contains three gates: an input gate, a forget gate, and an output gate, together with a recurrent cell. LSTM cells are usually organized in a chain, with outputs of previous LSTMs connected to the inputs of subsequent LSTMs. An important modification of the basic LSTM architecture is bidirectional LSTM, where the past and the future context is available at every time step. Bidirectional LSTMs, developed by Graves and Schmidhuber [14], contain two chains of LSTM cells flowing in both forward and backward direction, and the final representation is either a linear combination or simply concatenation of their states.

In our model, the aim of the BiLSTM layer is to make a presentation of input data and create the input features to the Conditional Random Fields part. The dimension of BiLSTM layer is 100. We applied 0.7 dropout rate for regularization and Adam optimizer [11].

3.3. Conditional Random Fields. CRF [20] is one of the state-of-the-art methods that takes a sequence of tokens as input, estimates the probabilities of labels (from a predefined set), and returns the best scoring label sequence. The CRF is defined by a graph whose vertices are feature vectors for each input token and edge weights is indicate a relationship between vertices. A linear-chain CRF is a CRF with a simple chain as the graph.

3.4. Multi-Task. Multi-task learning approach involves neural network training simultaneously on several similar tasks. This approach allows the network to improve results by using additional knowledge from other tasks. A neural network model for multi-task learning is implemented as follows: there are several input layers for different tasks, then common layers for all tasks and output layers, which number equal to the number of input layers. In addition, there can be more specific layers for each task before or after common layers.

We applied this approach to train our model to NER task (i) on three datasets simultaneously and (ii) on three NER datasets along with a dataset for POS tagging. The network shares all layers between tasks except input and output. The overall scheme of multi-task architecture is presented in Fig. 3.

§4. Experiments

In this section, we describe our evaluation metrics, datasets, baselines, and experimental results.

4.1. Metrics. To assure that our results are comparable to previously published ones, we apply the standard evaluation metric introduced for CoNLL shared tasks[3]. Additionally, we used a partial match scheme for

[3] `https://github.com/newsreader/evaluation/tree/master/nerc-evaluation`

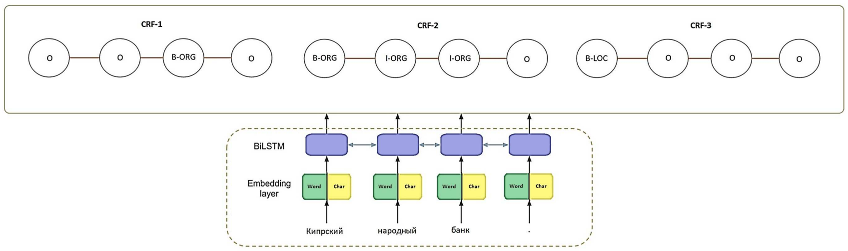

Figure 3. The overall architecture of multi-task model.

Study	**Evaluation Scheme**
Gareev et al. [13]	CoNLL script with 5-fold cross-validation
Malykh et al. [23]	CoNLL script with 5-fold cross-validation
Trofimov [34]	Manual evaluation per tag
Rubaylo et al. [30]	FactRuEval 2016 scripts
Sysoev et al. [33]	FactRuEval 2016 scripts
Ivanitsky et al. [29]	FactRuEval 2016 scripts
Mozharova et al. [25]	Part. matching with 4-fold cross-validation
Anh et al. [6]	CoNLL script

Table 1. Evaluation metrics adopted in previous studies.

FactRuEval 2016 dataset [32]. The comparison of evaluation metrics in related work is presented in Table 1.

4.2. Datasets. We use three Russian datasets for NER task: proposed by Gareev et al. [13], FactRuEval 2016 [32] and Person-1000 [25, 36] and one data set for multi-task learning with POS tags [17]. The overall statistic is presented in Table 2.

Gareev's et al. dataset contains 97 documents collected from ten top "Business" feeds in Yandex вЂќNewsвЂќ web directory. Titles and HTML tags were cleared from texts. Named entities of two types were manually annotated: persons and organizations. Only explicit mentions have been annotated without anaphoric references represented by pronouns or common nouns. The tagging scheme is IOB.

FactRuEval corpora were developed for the competition of extracting information from texts in Russian. The corpus consists of 255 news and

Set	**Gareev's**	**Persons-1000**	**FactRuEval**	**Kaggle POS**
Train	34765	414608	30940	719090
Test	8445	102152	59382	179772
Documents	97	1000	255	-
Tag scheme	IOB	IOB	IOB	-

Table 2. Overall statistics of the corpora.

analytical documents related to social and political issues. Entities of the following three types were annotated: person, organization and location (place names). The corpus is split into train and test parts. The tagging scheme is also IOB.

Person-1000 is a Russian news corpus with an annotated person named entities. This corpus consists of 1000 documents from the Russian online news services. The following entities were annotated: organization, person, location, geopolitician, media. The tagging scheme is IOB.

Kaggle dataset is developed for POS tagging task. The texts are already split into sentences and tokens. The data set is divided into train and test parts.

Since Gareev's dataset and Persons-1000 dataset do not have a ready split for train/test data, we used a 80/20 split.

4.3. Baseline models. Gareev et al. introduced two approaches for NER task: knowledge-based and statistical [13]. In this paper we focused on the statistical-based method, which is realized with the CRF model with a rich set of features: current token presented with one-hot embedding vector, number normalization, bag of words for 5-token window centered on the current token, bag of words for shape window centered on the current token, current token prefixes and suffixes up to 6 characters in length, LDA cluster labels of the current token, 3-window for LDA cluster labels centered on the current token, Brown cluster label, Clark cluster label, extended context and extended context over cluster labels.

Malykh et al. proposed character-aware neural network trained on corpus itself only [23]. The network consists of LSTM units which take as an input a current character and try to predict the next character and a markup label for the current character.

Rubaylo et al. applied Tomita-parser tool. The tool was developed by Yandex company and utilized in it services [30]. The system's pipeline of

Models	Gareev's dataset			Persons-1000			FactRuEval 2016		
	P	R	F	P	R	F	P	R	F
STM NER	75.89	84.70	80.05	97.33	97.36	97.34	46.83	36.50	41.02
MTM NER	87.70	**88.39**	**88.04**	**99.39**	**99.60**	**99.49**	73.04	**64.18**	**67.75**
MTM NER+POS	**89.37**	86.54	87.94	98.50	98.30	98.40	**73.89**	61.42	67.12

Table 3. Results of proposed model for single and multi-task modes.

information extraction consists of the following steps: tokenization, morphological analysis, lexical analysis, syntactic parse, an output of results. Tomita-parser utilizes a rule-based approach to NER.

Sysoev et al. method is based on sequential traversal and classification of text tokens [33]. In order to classify input tokens the SVM model with following features was utilized: token affixes of lengths from one to four; token text, part-of-speech tag, lemma and digit normalized token form, predicates, first letter case, containing characters of the same class, being constructed only from characters of the same class, token position in sentence, labels assigned to up to three previous tokens, dictionary-based and word2vec representation.

Ivanitsky et al. also applied SVM with named entity distributed vectors representation [29]. Vectors obtained by training the wor2vec model on four data sets: Russian subcorpus of Multilingual UN Parallel Text 2000–2009, Europarl, News, FactRuEval. The size of vectors are 200.

Anh et al. [3] proposed model consists of Bi-directional Long Short Term Memory (Bi-LSTM) and CRF. The model takes as an input external pre-trained word embedding vectors.

4.4. Results. We conducted three types of experiments with different groups of models: (i) single-task models (STM NER) on each of NER datasets; (ii) multi-task models (MTM NER) on a combination of NER datasets; (iii) multi-task models on NER and POS datasets. At the first set of experiments, we trained and evaluated our model separately on each dataset. After that, we trained the model in multi-task mode simultaneously on three NER data sets. We utilized learning rate 0.003 for target dataset and 0.002 for two others. In the last set of experiments, we train the model again in multi-task mode simultaneously on each NER dataset and POS-tag Kaggle dataset. The results of these experiments are presented in Table 3. The results show that multi-task NER (MTM NER)

Models	Gareev's dataset			FactRuEval 2016		
	P	R	F	P	R	F
Gareev et al. [13]	84.10	67.98	75.05	-	-	-
Malykh et al. [23]	59.65	65.70	62.49	-	-	-
Rubaylo et al. [30]	-	-	-	77.70	78.50	78.13
Sysoev et al. [33]	-	-	-	88.19	64.75	74.67
Ivanitsky et al. [29]	-	-	-	-	-	87.88
Anh et al. [3]	**89.57**	84.89	87.17	**83.88**	**80.84**	**82.10**
Multi-task NER	87.70	**88.39**	**88.04**	73.04	64.18	67.75

Table 4. Comparison results for Gareev's and FactRuEval 2016 datasets.

Models	Persons-1000 PER			Persons-1000 All		
	P	R	F	P	R	F
Trofimov [34]	97.26	93.92	95.57	-	-	-
Mozharova et al. [25]	-	-	97.21	-	-	-
Anh et al. [3]	**99.43**	99.09	99.26	**-**	**-**	**-**
Multi-task NER	99.39	**99.60**	**99.49**	**98.74**	**98.88**	**98.81**

Table 5. Comparison results for Persons-1000 PER and Persons-1000 All corpora

Dataset	PER			ORG			LOC		
	P	R	F	P	R	F	P	R	F
Gareev's	90.74	95.15	92.89	86.50	85.87	86.18	-	-	-
Persons-1000	99.11	99.60	99.35	98.55	98.08	98.31	97.15	97.08	97.11
FactRuEval 2016	84.54	76.60	80.37	62.90	46.86	53.71	74.32	74.42	74.37
FactRuEval 2016 (partial)	84.96	78.66	81.69	65.49	57.90	61.46	76.16	78.50	77.31

Table 6. Results of Multi-task NER model for PER, ORG and LOC tags

outperformed single-task NER (STM NER) on all datasets. The most significant increase was obtained on FactRuEval 2016 corpora. This result can be explained by a small amount of training data, which is not enough for training the model, in the initial splitting of FactRuEval dataset. The multi-task NER with POS tagging showed close results to multi-task NER and outperformed it in the terms of precision on Gareev's dataset and FactRuEval datasets.

Dataset	GEOPOLIT			MEDIA		
	P	R	F	P	R	F
Persons-1000	99.31	99.56	99.43	98.90	99.63	99.43

Table 7. Results of Multi-task NER model for GEOPOLIT and MEDIA tags

On the second step, we compared the results of multi-task NER model, which obtained the highest results in the first set of experiments, with baseline methods. The results are presented in Tables 4 and 5. Due to previous models were applied for identification only person entities on Person-1000 corpora, we compared the results only on these part of corpus. The highest results from baseline methods are obtained by Anh et al. approach. The other baseline models show significantly lower results compared with Anh et al. approach on Gareev's and FactRuEval 2016 datasets and comparable results on Person-1000 dataset. The proposed in this work model outperformed Anh et al. approach in terms of recall and F-measure on Gareev's dataset and Persons-1000 for person tag and showed comparable results in terms of precision. The increase on results for Gareev's corpus is 3.5% of recall, 0.87% of F-measure and for Person-1000 is 0.51% of recall and 0.23% of F-measure. Due to the results on Person-1000 are close to one hundred, such value of increase can be considered as significant. However, on the FactRuEval 2016 dataset, the results of proposed model are significantly lower than other baseline models. Our model obtained 14.35% less than Anh et al. approach. In addition, we evaluated our model for all types of entities in Person-1000 dataset. The results are presented in Table 5. Our model obtained 98.81% F-measure on this dataset.

We also evaluate the quality of our model for each type of entity. The results are presented in Tables 6 and 7. As it can be seen from the tables, entities with the person type are best recognized. The most difficult to recognize is organizations. Results for geopolitical and media types of entities are close to hundred percent (99.43% of F-measure for both types of entities).

§5. Conclusion

In this work, we investigated the efficiency of multi-task approach for NER in the Russian language. We conducted experiments on three NER datasets with different types of entities and a POS dataset. The results show the improvement of evaluated metrics in case of use multi-task mode. The comparison of obtained results with previous approaches shows the increase of metrics on two datasets. We also evaluated the efficiency of multi-task for NER and POS tagging tasks. This approach didn't give the increase in results. Further work will be devoted to improving CNN and LSTM parts of the proposed model and using attention mechanisms.

References

1. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., *Tensorflow: a system for large-scale machine learning*, OSDI, vol. 16, 2016, pp. 265–283.
2. C. Adak, B. B. Chaudhuri, and M. Blumenstein, *Named entity recognition from unstructured handwritten document images*, 12th IAPR Workshop on Document Analysis Systems (DAS), 2016, pp. 375–380.
3. L. T. Anh, M. Y. Arkhipov, M. S. Burtsev, *Application of a hybrid bi-LSTM-CRF model to the task of Russian named entity recognition*, Communications in Computer and Information Science book series – CCIS, vol. 789, 2017.
4. L.T. Anh, M. Y. Arkhipov, M. S. Burtsev, *Application of a hybrid bi-LSTM-CRF model to the task of Russian named entity recognition*, arXiv preprint arXiv:1709.09686 (2017).
5. A. Y. Antonova, A. N. Soloviev, *Conditional random field models for the processing of Russian*, Communications of the ACM 56(6), 2013.
6. M. Y. Arkhipov, M. S. Burtsev, L. T. Anh, *Application of a hybrid bi-LSTM-CRF model to the task of Russian named entity recognition*, Conference on Artificial Intelligence and Natural Language. Springer, Cham, 2017.
7. M. M. Brykina, A. V. Faynveyts, S. Yu. Toldova, *Dictionary-based ambiguity resolution in Russian named entities recognition*, International Workshop on Computational Linguistics and its Applications, ed. A. Narin'yani, vol. 1, 2013.
8. R. Chalapathy, E. Z. Borzeshi, M. Piccardi, *Bidirectional LSTM-CRF for clinical concept extraction*, arXiv preprint arXiv:1611.08373 (2016).
9. J.P.C. Chiu, E. Nichols, *Named entity recognition with bidirectional LSTM-cnns*, Transactions of the Association for Computational Linguistics **4** (2016), 357–370.
10. L.G. Craidlin, *Program of allocation of Russian individualized nominal groups taglite*, Computational linguistics and intellectual technologies Dialog, 2005.
11. D. Kingma, J. Ba *Adam: A method for stochastic optimization*, 3rd International Conference for Learning Representations, San Diego, 2014.
12. C. Dong, J. Zhang, C. Zong, M. Hattori, H. Di, *Character-based LSTM-CRF with radical-level features for chinese named entity recognition*, Natural Language Understanding and Intelligent Applications, Springer, 2016, pp. 239–250.
13. R. Gareev, M. Tkachenko, V. Solovyev, A. Simanovsky, V. Ivanov, *Introducing baselines for Russian named entity recognition*, Computational Linguistics and Intelligent Text Processing, 2013.

14. A. Graves, S. Fernández, J. Schmidhuber, *Bidirectional LSTM networks for improved phoneme classification and recognition*, Artificial Neural Networks: Formal Models and Their Applications – ICANN 2005, 2005.
15. K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, J. Schmidhuber, *LSTM: A search space odyssey*, IEEE Trans Neural Netw Learn Syst. doi: 10.1109/TNNLS.2016.2582924, 2016.
16. Z. Huang, W. Xu, K. Yu, *Bidirectional LSTM-CRF models for sequence tagging*, arXiv preprint arXiv:1508.01991 (2015).
17. Kaggle, *Predict Russian universal dependencies POS tags*, 2017.
18. G. Konoplich, E. Putin, A. Filchenkov, R. Rybka, *Named entity recognition in Russian with word representation learned by a bidirectional language model*, AINL, 2018.
19. G. Konoplich, E. Putin, A. Filchenkov, R. Rybka, *Named entity recognition in Russian with word representation learned by a bidirectional language model*, Conference on Artificial Intelligence and Natural Language, Springer, 2018, pp. 48–58.
20. J. Lafferty, A. McCallum, F. Pereira, *Conditional random fields: Probabilistic models for segmenting and labeling sequence data*, Proc. 18th International Conference on Machine Learning, 2001.
21. G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, C. Dyer, *Neural architectures for named entity recognition*, Proc. 2016 NAACL (2016), 260–270.
22. X. Ma, E. Hovy, *End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF*, arXiv preprint arXiv:1603.01354 (2016).
23. V. Malykh, A. Ozerin, *Reproducing Russian ner baseline quality without additional data.*, CDUD@ CLA, 2016, pp. 54–59.
24. S. Misawa, M. Taniguchi, Y. Miura, T. Ohkuma, *Character-based bidirectional LSTM-CRF with words and characters for japanese named entity recognition*, Proc. 1st Workshop on Subword and Character Level Models in NLP, 2017, pp. 97–102.
25. V. Mozharova, N. Loukachevitch, *Two-stage approach in Russian named entity recognition*, Proc. 2016 International FRUCT Conference on Intelligence, Social Media and Web (ISMW FRUCT), IEEE, 2016, pp. 1–6.
26. A. V. Podobryaev, *Searching for person memories in news texts with the use of a model of conditional random fields*, RCDL, 2013.
27. B. Popov, A. Kiryakov, D. Ognyanoff, D. Manov, A. Kirilov, *Kim — a semantic platform for information extraction and retrieval*, Journal of Natural Language Engineering 10, 2004.
28. R. M. Zavala, P. Martinez, I. Segura-Bedmar, *A hybrid bi-LSTM-CRF model for knowledge recognition from ehealth documents*, TASS 2018: Workshop on Semantic Analysis at SEPLN, 2018, pp. 65–70.
29. R. Ivanitskiy, A. Shipilo, L. Kovriguina, *Russian named entities recognition and classification using distributed word and phrase representations*, SIMBig, 2016.
30. A. V. Rubaylo, M. Y. Kosenko, *Software utilities for natural language information retrievial*, Almanac of modern science and education, Volume 12, 2016.
31. E. Sheng, S. Miller, J.S. Ambite, P. Natarajan, *A neural named entity recognition approach to biological entity identification*, Proc. BioCreative VI Workshop, 2017, pp. 24–27.

32. A. S. Starostin, V. V. Bocharov, S. V. Alexeeva, A. Bodrova, A. S. Chuchunkov, S. S. Dzhumaev, M. A. Nikolaeva, *Evaluation of named entity recognition and fact extraction systems for Russian*, Annual International Conference Dialogue, 2016.
33. A.A. Sysoev, I.A. Andrianov, *Named entity recognition in Russian: the power of wiki-based approach*, Proc. International Conference Dialogue, 2016, pp. 746–755.
34. I. V. Trofimov, *Person name recognition in news articles based on the persons-1000/1111-f collections*, 16th All-Russian Scientific Conference Digital Libraries: Advanced Methods and Technologies, Digital Collections, RCDL, 2014, pp. 217–221.
35. E. Tutubalina, S. Nikolenko, *Combination of deep recurrent neural networks and conditional random fields for extracting adverse drug reactions from user reviews*, Journal of Healthcare Engineering, vol. 2017, Article ID 9451342 (2017).
36. N. A. Vlasova, E. A. Suleymanova, I. V. Trofimov, *Report on Russian corpus for personal name retrieval*, Proceedings of computational and cognitive linguistics TEL, 2014.
37. Q. Wei, T. Chen, R. Xu, Y. He, L. Gui, *Disease named entity recognition by combining conditional random fields and bidirectional recurrent neural networks*, Database **2016** (2016).

Kazan Federal University,
Kazan, Russia, 420008
E-mail: daniyarttt@gmail.com

Поступило 14 января 2019 г.

Kazan Federal University,
Kazan, Russia, 420008
E-mail: alimovailseyar@gmail.com

Lomonosov Moscow State University,
Kolmogorova str., 1, Moscow 119991, Russia;
Kazan Federal University,
Kazan, Russia, 420008
E-mail: tutubalinaev@gmail.com