Comparison of Named Entity Recognition Tools Applied to News Articles

Conference Paper · December 2019

DOI: 10.1109/ISPRAS47671.2019.00017

2 authors:

Sergey Vychegzhanin
Vyatka State University

Evgeny Kotelnikov
Vyatka State University

Comparison of Named Entity Recognition Tools Applied to News Articles

Sergey Vychegzhanin
Department of Applied Matematics and Computer Science
Vyatka State University
Kirov, Russia
vychegzhaninsv@gmail.com

Evgeny Kotelnikov
Department of Applied Matematics and Computer Science
Vyatka State University
Kirov, Russia
kotelnikov.ev@gmail.com

***Abstract*—Named Entity Recognition in texts is an important natural language processing task. There are many systems to solve this problem. These systems differ in targeting domains, processing methodologies, supported languages and recognized entity types. The presence of a large number of aspects creates difficulties for the user when choosing the appropriate tool for solving a specific problem. The aim of this work is a comparative study of seven publicly available and well-known libraries that can elicit named entities: Stanford NER, spaCy, NLTK, Polyglot, Flair, GATE and DeepPavlov. The article consists of seven sections. The introduction lists the areas of application for the Named Entity Recognition task and the approaches used to solve it. The second section is devoted to a review of works in which comparative studies of existing tools are presented. In the third section, the characteristics of the four text corpora that were used during the experiments are given. The fourth section contains a brief description of the tools selected for research. The fifth section describes the metrics used to evaluate tool performance. The sixth section presents the results of the experiments and their discussion. In conclusion the results of the work are summarized. The results of the study show that for the English language close values of the F1-score for the problem of Named Entities Recognition have the Flair and DeepPavlov libraries. For the Russian language the first place is taken by the DeepPavlov library, significantly surpassing other tools in quality.**

Keywords—natural language processing; named entity recognition; machine learning

I. INTRODUCTION

In the field of natural language processing there is the Information Extraction task, which consists in elicitation of data relevant to a particular topic from non-structured texts. One of the subtasks of Information Extraction is the recognition and extraction of named entities (Named Entity Recognition, NER). The results of recognition and classification of proper nouns in a text document are widely used in information retrieval, machine translation, question answering and automatic summarization [1]. Depending on the specifics of the problem being solved the names of people (person, PER), names of organizations (ORG), geographical names (location, LOC), time indicators (TIM), which are widely used in news articles [2], and various measures (percent, weight, money), email addresses can be used as recoverable named entities. There can also be domain specific entity types such as medical drug names, disease symptoms, etc. [3].

To solve the problems of extracting information from texts two main approaches are used: rule-based and machine learning. Rule-based systems may achieve high degree of precision, but the development process is time-consuming and requires the transfer of these rules from one domain to another. Machine learning systems require large amount of annotated documents. Its performance depends on the quality of training data.

This article presents the results of a comparative study of well-known tools for natural language processing in relation to news articles. Although the examined tools could be trained with specific corpora, we used available pre-trained models in experiments. This situation is especially common for users that do not have enough experience, time or data for training the tools for a specific purpose. This study may be useful to developers and researchers who need a tool for named entity recognition.

II. REVIEW OF PREVIOUS WORK

The community of software developers has created many systems that can extract named entities from unstructured texts. Systems are distributed in the form of source codes, executables or web services and differ in licenses, targeting domains, supported languages, processing methodologies, recognized entity types, input and output formats.

There are works in which comparative studies of various Named Entity Recognition systems are presented. The authors of [4] compare the tools of Stanford NER, Illinois NET, OpenCalais NER WS and Alias-i LingPipe for bibliographic texts. For evaluation of tools performance a corpus of 247 articles taken from Wikipedia was formed and manually annotated. According to the results of the experiment, the tools are arranged in the following order in decreasing performance: Stanford NER, LingPipe, Illinois NET, OpenCalais.

In the work [5] the OpenNLP, Stanford NER, AlchemyAPI, and OpenCalais Named Entity Recognition Tools are studied for extracting entities from the output of an optical character recognition workflow. Wiener Library (WL) and the King's College London's Serving Soldier archive (KCL) are used as sources for text documents. Material WL consists of four documents containing a total of seventeen typescript pages. KCL material consists of 33 newsletters. The experimental results show that Stanford NER on average for two datasets gives the best performance and it is the most effective tool for the PER and LOC types. AlchemyAPI shows the best results for the ORG type.

A study [6] presents a comparison of four tools: Stanford NER, spaCy, Alias-i LingPipe, Natural Language Toolkit (NLTK). As a text corpus WikiGold is used, consisting of 145 labelled articles taken from Wikipedia. Tools are evaluated in relation to the entity types PER, LOC, ORG. The assessment

978-1-7281-6055-9/19/$31.00 ©2019 IEEE
DOI 10.1109/ISPRAS47671.2019.00017

is subject to exact and partial matching. As a result of the experiments, the following order of tools was obtained in decreasing value of the F1-score: Stanford NER, spaCy, NLTK, LingPipe. The maximum value of the F1-score obtained with the help of Stanford NER, on average for the three types with partial matching, was 0.7609.

The authors of [7] divide the tools into two groups: standard and social network-oriented. The first group includes tools designed to work with no specific kind of texts: NLTK, Apache OpenNLP, Stanford CoreNLP, Pattern. The second group includes tools designed to work with short messages from social networks: TwitterNLP, TweetNLP, TweetIE. The experiments use the corpora of CoNLL-2003, Alan Ritter's Twitter corpus and MSM2013 (Making Sense of micro post 2013 Concept Extraction Challenge). Alan Ritter's and MSM2013 corpora contain short and simple sentences; CoNLL-2003 contains longer and more complex sentences. As a result, the highest values of F1-score for Alan Ritter's and MSM2013 corpora were obtained using TwitterNLP and were equal to 0.95 and 0.91, respectively. For CoNLL-2003, OpenNLP is in the first place with F1-score of 0.87.

The article [8] explores GATE, Stanford NER, Stanford-Twitter, NERD-ML, DBpedia Spotlight, Lupedia, TextRazor, Zemanta. Tweet corpora from Alan Ritter, UMBC (University of Maryland, Baltimore County) and MSM2013 were used as text corpora. According to the results of experiments for the Alan Ritter's and MSM2013 corpora, the highest values of the F1-score were obtained using NERD-ML and were equal to 0.5149 and 0.7718, respectively. For the UMBC Stanford NER is the best with F1-score of 0.4027.

A study [9] presents F1-score estimates for the NER-Tagger, Stanford NER, Edinburgh Geoparser, spaCy, and Polyglot tools. Estimates were obtained on two corpora of historical texts: Mary Hamilton Papers and Samuel Hartlib Papers. The Hamilton corpus contains 161 texts in which the entities of the PER and LOC types are labelled, in the Hartlib corpus – 50 texts in which the entities of the LOC type are labelled. The best results on the Hartlib corpus are shown by Stanford NER (F1 = 0.708), on the Hamilton corpus – by Polyglot (F1 = 0.616).

The authors of the article [10] provide F1-scores for the DeepPavlov and spaCy tools obtained on the OntoNotes Release 5.0 corpus. The average value of the F1-score for 18 types of named entities for DeepPavlov is 0.8707, for spaCy – 0.8585.

An analysis of previous works allows us to conclude that when extracting named entities of the classes PER, ORG and LOC from English texts, most often the highest average F1-score is achieved using Stanford NER. Moreover, its quality is higher than the quality of other tools in each of the three classes. However, there are very few works that conduct comparative studies of tools based on modern neural network approaches. The aim of this work is a comparative study of tools, among which there are tools with a long history of development (Stanford NER and GATE) and relatively new and actively developing (DeepPavlov and spaCy). The differences between this work and existing ones are: (1) the assessment of tools for news articles in English and Russian, while most of the existing works use corpora containing only English texts; (2) the evaluation of processing time for tools, while existing works evaluate only F1-score.

III. TEXT CORPORA

Evaluation of the effectiveness of tools for Named Entity Recognition was performed using four publicly available text corpora containing news articles:

1. Kaggle-2016[1] – English-language corpus annotated for Named Entity Recognition using GMB (Groningen Meaning Bank) corpus. Corpus was published on Kaggle platform and was labeled into eight entity types: person (PER), organization (ORG), geographical entity (GEO), geopolitical entity (GPE), time indicator (TIM), artifact (ART), event (EVE), natural phenomenon (NAT) [11].

2. CoNLL-2003[2] – collection of English-language news articles from the Reuters Corpus, used in 2003 at the Conference on Computational Natural Language Learning (CoNLL) for performance evaluation of NER methods. The named entities were manually annotated at the University of Antwerp using four entity types, namely person (PER), organization (ORG), location (LOC) and miscellaneous (MISC) [2].

3. Named_Entities_3[3] – Russian-language corpus based on Person-1000 collection, created by Artificial Intelligence Research Center of the Institute of Program Systems of the Russian Academy of Sciences. In the Persons-1000 the mentions of person names (PER) in texts were annotated. In the Named_Entities_3 corpus the markup of person names from the Persons-1000 was supplemented by marking up the organizations (ORG) and locations (LOC) [12].

4. FactRuEval-2016[4] – Russian-language corpus, which consists of news and analytical texts on socio-political topic. In 2016 this corpus was used in the named entity recognition and fact extraction competition at the conference Dialogue. Corpus was annotated by volunteers using the OpenCorpora.org platform and contains markup of three entity types: person (PER), organization (ORG) and location (LOC) [13].

All texts in corpora are labelled in BIO format (Begin Inside Out). The characteristics of the corpora are presented in Table I. The statistical distribution of named entities by the types used in experiments is presented in Table II.

TABLE I. CHARACTERISTICS OF TEXT CORPORA

Corpus	Text language	Types of named entities	Number of texts	Average text length, tokens
Kaggle-2016	English	PER, ORG, GEO, GPE, TIM, ART, EVE, NAT	47,959	23
CoNLL-2003	English	PER, ORG, LOC, MISC	1,627	30
Named_ Entities_3	Russian	PER, ORG, LOC	1,000	273
FactRuEval-2016	Russian	PER, ORG, LOC	132	463

[1] https://www.kaggle.com/abhinavwalia95/entity-annotated-corpus
[2] https://www.clips.uantwerpen.be/conll2003/ner
[3] https://labinform.ru/pub/named_entities/descr_ne.htm
[4] https://github.com/dialogue-evaluation/factRuEval-2016

TABLE II. STATISTICAL DISTRIBUTION OF NAMED ENTITIES

Type of named entity	Corpus			
	Kaggle-2016	**CoNLL-2003**	**Named_ Entities_3**	**FactRuEval-2016**
PER	16,990	1,617	10,623	1,388
ORG	20,143	1,661	8,542	2,034
LOC (GEO+GPE)	53,514	1,668	7,244	1,574
TIM	20,333	–	–	–

IV. TOOLS

The selection of tools for Named Entity Recognition was carried out in accordance with the following criteria:

- free license;
- existence of local version;
- independence from targeting domain;
- ability to recognize basic entity types: person, organization, location, time indicators;
- support for the English or the Russian languages.

Based on these criteria, the following tools were selected:

1. Stanford NER [14] – a tool for Named Entity Recognition developed by the Stanford Natural Language Processing Group. Implementation is done in Java programming language. The NER component uses the Conditional Random Field model. There are models pre-trained on the CoNLL-2003, MUC-6 and MUC-7 corpora.

2. spaCy[5] – a library for the efficient natural language processing. Developed by Explosion AI. The NER module is based on a model that uses Bloom embeddings and a residual convolutional neural network. For the English language there are models that were pre-trained on the OntoNotes Release 5.0 package, which contains labelled texts, the sources of which are telephone conversations, news, blogs. For the Russian language there are models pre-trained on Wikipedia.

3. NLTK (Natural Language Toolkit) [15] – a library for symbol and statistical processing of natural language, implemented in the Python programming language. Developed at the University of Pennsylvania. It allows performing various text processing tasks: segmentation, tokenization, part-of-speech tagging, etc. It is usually used in the process of teaching students and conducting research. Module NER uses Maximum Entropy Classifier trained on ACE (Automatic Content Extraction) corpus.

4. Polyglot [16] – a library for solving various problems of natural language processing, including language detection, tokenization, part-of-speech tagging, sentiment analysis. The developer is Rami Al-Rfou. The library is written in Python. NER uses word embeddings as features. A simple neural network with one hidden layer is used as a classifier. The models used for Named Entity Recognition are pre-trained on Wikipedia.

5. GATE (General Architecture for Text Engineering) [17] – a framework for NER in natural language texts written in Java. Developed at the University of Sheffield. Named Entity Recognition is done using the ANNIE component (A Nearly-New Information Extraction), which uses finite state machines and rules in the Jape language.

6. Flair [18] – a library containing effective models for solving natural language processing tasks, such as named entity recognition, part-of-speech tagging, sense disambiguation and classification. Developed by the Zalando Research team. It is written in the Python programming language. The NER module uses contextual embeddings obtained using the bidirectional character language model. The BiLSTM-CRF model is used for Named Entity Recognition.

7. DeepPavlov [10] – a framework for chatbots and virtual assistants development, written in Python. Developed at the Laboratory of Neural Systems and Deep Learning at Moscow Institute of Physics and Technology (MIPT). For Named Entity Recognition two pre-trained neural network models are used: the BERT (Bidirectional Encoder Representations from Transformers) model and the Bi-LSTM-CRF model.

Table III shows characteristics of named entity recognition tools.

TABLE III. CHARACTERISTICS OF NAMED ENTITY RECOGNITION TOOLS

Tool	Programming language	License	Method	Model	Training corpus
Stanford NER	Java	GPL	Conditional Random Field	english_conll_4class	CoNLL-2003
				english_muc_7class	MUC-6, MUC-7
spaCy	Python	MIT	Bloom embeddings and a residual convolutional neural network	en_core_web_sm	OntoNotes
				en_core_web_md	OntoNotes, Common Crawl
				en_core_web_lg	OntoNotes, Common Crawl
				xx_ent_wiki_sm	WikiNER
				ru2	–
NLTK	Python	Apache License v2.0	Maximum Entropy	–	ACE
Polyglot	Python	GPLv3	Feedforward neural network	–	Wikipedia
Flair	Python	MIT	BiLSTM-CRF	–	CoNLL-2003
GATE	Java	LGPL	Finite state machines and rules in the Jape language	–	–
DeepPavlov	Python	Apache License v2.0	BERT	ner_conll2003_bert	CoNLL-2003
				ner_rus_bert	Wikipedia, news data
				ner_ontonotes_bert_mult	OntoNotes

[5] https://spacy.io

V. Evaluation Metrics

There are different approaches for the evaluation of NER systems [1]. Performance evaluation characterizes the ability of a tool to find the boundaries of a named entity and correctly determine its type. The score can be computed for exact and partial matching. The exact matching assumes the exact coincidence of the boundaries of the predicted and true named entity. Under such conditions the systems of the participants in the competition held as part of the CoNLL-2003 conference [2] were evaluated. However, in some cases the exact matching of boundaries is not as important as the identification of a major part of a named entity. For example, the phrases "The United States" and "United States" are almost similar and differ only in the presence of the article. At the MUC (Message Understanding Conference) [19], metrics were used to evaluate the systems, taking into account the partial overlap of the predicted and true named entities.

In this paper systems are evaluated under two conditions:

1) Exact matching of boundaries and types of predicted and true entities.

2) Partial matching of the predicted and true entities when the types coincide and the boundaries of the predicted entity are inside the boundaries of the true one, or vice versa.

To evaluate the performance of NER the following metrics are used, calculated with respect to type *i*:

- precision – the proportion of correctly classified entities (True Positives) among all entities assigned by the classifier to type *i* (True Positives and False Positives)

$$P_i = \frac{TP_i}{TP_i + FP_i}; \tag{1}$$

- recall – the proportion of correctly classified entities (True Positives) among all entities belonging to type *i* (True Positives and False Negatives)

$$R_i = \frac{TP_i}{TP_i + FN_i}; \tag{2}$$

- F1-score – harmonic mean of precision and recall

$$F1_i = \frac{2 \times P_i \times R_i}{P_i + R_i}. \tag{3}$$

The final values of the metrics are obtained by the method of macro-averaging for all types by the formulas:

$$P^M = \frac{\sum_{i=1}^{|C|} P_i}{|C|}, \tag{4}$$

$$R^M = \frac{\sum_{i=1}^{|C|} R_i}{|C|}, \tag{5}$$

$$F1^M = \frac{\sum_{i=1}^{|C|} F1_i}{|C|}, \tag{6}$$

where $|C|$ – total count of types.

VI. Results

Table IV shows the values of the F1-score obtained for the English language when classifying named entities into 3 types (PER, ORG, LOC) and 4 types (PER, ORG, LOC, TIM). Table V shows the values of the F1-score obtained for the Russian language when classifying named entities into 3 types (PER, ORG, LOC).

Analyzing the obtained results, the following conclusions can be drawn. Under the condition of partial matching all tools show higher performance than with exact matching. This means that there are cases when the tool correctly defines the entity type, but incorrectly defines its boundaries. Such cases are:

1) The presence of articles or words preceding a named entity. For example, "Prime Minister John Howard" and "John Howard", "The New York Times" and "New York Times", "in November" and "November".

2) Excessive or incomplete extraction of the entity. Examples for the English language: "Business Policy the National University of Singapore" and "National University of Singapore", "the Neolithic period" and "Neolithic". Examples for the Russian language: "Internal Affairs Directorate of Tomsk Region" and "Internal Affairs Directorate", "Danilovsky District of Moscow" and "Danilovsky District".

TABLE IV. Results of named entity recognition experiments for English

Tool	Model	F1-score					
		Exact matching			Partial matching		
		Kaggle-2016		CoNLL-2003	Kaggle-2016		CoNLL-2003
		3 types	4 types	3 types	3 types	4 types	3 types
Stanford NER	english_conll_4class	0.554	–	0.860	0.663	–	0.886
	english_muc_7class	0.511	**0.486**	0.611	0.650	0.613	0.683
spaCy	en_core_web_sm	0.483	0.452	0.521	0.649	0.619	0.608
	en_core_web_md	0.503	0.468	0.570	0.665	0.631	0.659
	en_core_web_lg	0.496	0.463	0.597	0.660	0.629	0.697
	xx_ent_wiki_sm	0.501	–	0.597	0.637	–	0.695
NLTK	–	0.476	–	0.467	0.616	–	0.555
Polyglot	–	0.476	–	0.467	0.650	–	0.595
Flair	–	**0.584**	–	**0.887**	**0.691**	–	**0.904**
GATE	–	0.460	0.448	0.528	0.575	0.554	0.598
DeepPavlov	ner_conll2003_bert	0.576	–	0.860	**0.691**	–	0.901
	ner_ontonotes_bert_mult	0.523	0.475	0.687	0.685	**0.637**	0.741

TABLE V. RESULTS OF NAMED ENTITY RECOGNITION EXPERIMENTS FOR RUSSIAN

Tool	Model	F1-score			
		Exact matching		Partial matching	
		Named_Entities_3	FactRuEval-2016	Named_Entities_3	FactRuEval-2016
		3 types		3 types	
spaCy	xx_ent_wiki_sm	0.454	0.418	0.681	0.559
	ru2	0.214	0.210	0.361	0.307
Polyglot	–	0.499	0.429	0.674	0.589
GATE	–	0.299	0.268	0.370	0.342
DeepPavlov	ner_rus_bert	**0.945**	**0.622**	**0.973**	**0.752**
	ner_ontonotes_bert_mult	0.688	0.556	0.816	0.679

TABLE VI. NAMED ENTITY RECOGNITION PROCESSING TIME FOR TOOLS

Tool	Model	Processing time, sec	
		Kaggle-2016	Named_Entities_3
Stanford NER	english_conll_4class	76,935	–
	english_muc_7class	75,723	–
spaCy	en_core_web_sm	298	–
	en_core_web_md	319	–
	en_core_web_lg	327	–
	xx_ent_wiki_sm	164	13
	ru2	–	136
NLTK	–	475	–
Polyglot	–	107	113
Flair	–	31,711	–
GATE	–	87	21
DeepPavlov	ner_conll2003_bert	2,793	–
	ner_ontonotes_bert_mult	2,759	497
	ner_rus_bert	–	465

For the English language when recognizing three types of named entities Flair shows the best results, and DeepPavlov is slightly poorer. When recognizing four types of entities the leaders are Stanford NER (exact matching) and DeepPavlov (partial matching). For the Russian language DeepPavlov takes the first place significantly surpassing the performance of other tools.

The most difficult to recognize are the types ORG and TIM, the simplest is the type PER. For example, for DeepPavlov on the Kaggle-2016 corpus subject to partial matching the following scores were obtained for types: $F1_{PER} = 0.813$, $F1_{ORG} = 0.540$, $F1_{LOC} = 0.703$, $F1_{TIM} = 0.493$.

Table VI shows the processing time of tools for Named Entity Recognition for the English language on the Kaggle-2016 corpus and for the Russian language on the Named_Entities_3 corpus. The fastest tools are spaCy, NLTK, Polyglot and GATE, while DeepPavlov works an order of magnitude slower. Stanford NER and Flair worked slower by two orders of magnitude. It should be noted that for all libraries except DeepPavlov the runtime was measured on the CPU. For DeepPavlov, the runtime was measured on the GPU, since this library supports working with the GPU. When working on the CPU the processing time of DeepPavlov is comparable with Flair.

VII. CONCLUSION

A comparative study of seven tools for Named Entity Recognition from natural language texts was conducted: Stanford NER, spaCy, NLTK, Polyglot, Flair, GATE, DeepPavlov. The experiments were held on two corpora for English (Kaggle-2016 and CoNLL-2003) and two corpora for the Russian language (Named_Entities_3 and FactRuEval-2016). When recognizing three main types of named entities (PER, ORG and LOC), Flair turned out to be the leader in performance for the English language, and DeepPavlov showed very close results to it. The first place for the Russian language was given to DeepPavlov with a performance significantly superior to other tools.

ACKNOWLEDGMENT

The reported study was funded by the Ministry of Education and Science of the Russian Federation according to the research project No. 34.2092.2017/4.6 ‘Research and Development of Sentiment Lexicons for Text Sentiment Analysis’.

REFERENCES

[1] D. Nadeau and S. Sekine, “A survey of named entity recognition and classification”, in Lingvisticae Investigationes, 30(1), 2007, pp. 3–26.

[2] E. F. T. K. Sang and F. D. Meulder, “Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition”, in Proceedings of the Conference on Natural Language Learning, 2003, pp. 142–147.

[3] A. B. Abacha and P. Zweigenbaum, “Medical Entity Recognition: A Comparison of Semantic and Statistical Methods”, in Proceedings of the 2011 Workshop on Biomedical Natural Language Processing, 2011, pp. 56–64.

[4] S. Atdag and V. Labatut, “A comparison of named entity recognition tools applied to biographical texts”, in 2nd International Conference on Systems and Computer Science (ICSCS), 2013.

[5] K. J. Rodriquez, M. Bryant, T. Blanke and M. Luszczynska, “Comparison of Named Entity Recognition tools for raw OCR text”, in Proceedings of KONVENS 2012 (LThist 2012 workshop), 2012, pp. 410–414.

[6] R. Jiang, R. E. Banchs and H. Li, “Evaluating and combining named entity recognition systems”, in Proceedings of the Sixth Named Entity Workshop, Joint with 54th Association for Computational Linguistics, 2016, pp. 21–27.

[7] A. Pinto, H. G. Oliveira and A. O. “Alves Comparing the Performance of Different NLP Toolkits in Formal and Social Media Text”, in Proceedings of the 5th Symposium on Languages, Applications and Technologies (SLATE’16), 2016, pp. 1–16.

[8] L. Derczynski, D. Maynard, G. Rizzo, M. Erp, G. Gorrell, R. Troncy, J. Petrak and K. Bontcheva, “Analysis of named entity recognition and linking for tweets”, in Information Processing & Management, vol. 51, no. 2, 2015, pp. 32–49.

[9] M. Won, P. Murrieta-Flores and B. Martins, “Ensemble Named Entity Recognition (NER): Evaluating NER Tools in the Identification of Place Names in Historical Corpora”, in Frontiers in Digital Humanities 5, vol. 5, 2018, pp. 1–12.

[10] M. Burtsev, A. Seliverstov, R. Airapetyan, M. Arkhipov, D. Baymurzina, N. Bushkov, O. Gureenkova, T. Khakhulin, Y. Kuratov, D. Kuznetsov, A. Litinsky, V. Logacheva, A. Lymar, V. Malykh, M. Petrov, V. Polulyakh, L. Pugachev, A. Sorokin,

M. Vikhreva, M. Zaynutdinov, “DeepPavlov: Open-Source Library for Dialogue Systems”, in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics-System Demonstrations, 2018, pp. 122–127.

[11] J. Bos, V. Basile, K. Evang, N. Venhuizen and J. Bjerva, “The Groningen Meaning Bank”, in Handbook of Linguistic Annotation, vol. 2, 2017, pp. 463–496.

[12] V. Mozharova and N. Loukachevitch, “Two-stage approach in Russian named entity recognition”, in Proceedings of the International FRUCT Conference on Intelligence, Social Media and Web (ISMW-FRUCT), 2016, pp. 43-48.

[13] A. S. Starostin, V. V. Bocharov, S. V. Alexeeva, A. A. Bodrova, A. S. Chuchunkov, S. S. Dzhumaev and M. A. Nikolaeva, “FactRuEval 2016: Evaluation of Named Entity Recognition and Fact Extraction Systems for Russian”, in Proceedings of the International Conference “Dialogue 2016”, no. 15, 2016, pp. 702 – 720.

[14] J. R. Finkel, T. Grenager and C. Manning, “Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling”, in 43rd Annual Meeting on ACL, 2005, pp. 363–370.

[15] S. Bird, E. Klein and E. Loper, “Natural Language Processing with Python”, 2009.

[16] R. Al-Rfou, V. Kulkarni, B. Perozzi and S. Skiena, “Polyglot-NER: Massive multilingual named entity recognition”, in Proceedings of the 2015 SIAM International Conference on Data Mining, 2015.

[17] H. Cunningham, D. Maynard, K. Bontcheva and V. Tablan, “GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications”, in Proceedings of the Meeting of the Association for Computational Linguistics, 2002.

[18] A. Akbik, T. Bergmann, D. Blythe, K. Rasul, S. Schweter and R. Vollgraf, “FLAIR: An Easy-to-Use Framework for State-of-the-Art NLP”, in Proceedings of NAACL-HLT 2019: Demonstrations, 2019, pp. 54–59.

[19] R. Grishman and B. Sundheim, “Message Understanding Conference - 6: A Brief History”, in Proceedings of the International Conference on Computational Linguistics, 1996, pp. 466–471.