



Document-Level Relation Extraction in Russian

Alexey V. Yandutov^(✉) and Natalia V. Loukachevitch

Lomonosov Moscow State University, Moscow, Russia
`wynand.enterprises@gmail.com`

Abstract. We explore automatic relation extraction at the document level for Russian texts, considering nested entities. Using state-of-the-art models, we conducted first-of-its-kind experiments on Russian texts and made necessary model adjustments to implement experiments on the NEREL dataset, which is annotated with nested named entities and document-level relations between them. The results validate the efficiency of the proposed techniques in relation extraction from Russian texts, including nested entities.

Keywords: Document level relation extraction · Nested named entities · Russian language

1 Introduction

The present era of rapid advancements in artificial intelligence (AI) marks the prominence of Natural Language Processing (NLP), a field that is particularly affected by the burgeoning amounts of digitally-stored data. With the increasing demand for automatic extraction of information from text, vital to numerous NLP applications such as question-answer systems, information retrieval, and augmentation of knowledge databases, the role of automatic extraction of relationships or relations between entities within a text, a process termed Relation Extraction (RE), is increasingly pivotal.

Entities within textual contexts often establish distinct relationships that can be harnessed in a multitude of applications, including the analysis of news articles, scholarly publications, and social media posts. RE thereby has far-reaching implications, significantly impacting industries beyond academia, particularly the business sector.

Despite substantial strides since the 1990s, RE as a field is complex and underdeveloped relative to Named Entity Recognition (NER). Significant improvement in RE has been observed with the introduction of transformer architectures, but traditional RE methods, largely confined to intra-sentence relationships, fall short when it comes to untangling complex relations that span sentences in a document. This highlights the need for extraction at the document level, a need that is particularly pressing given that most existing research in the area is largely focused on English texts.

Our work addresses this gap, embarking on the exploration of document-level relation extraction in Russian texts, a domain that remains largely uncharted. Until recently, available datasets for relation extraction in the Russian language were scarce. The advent of the NEREL dataset, annotated with nested named entities and their relationships, offers an unprecedented opportunity to delve into relation extraction methodologies, encompassing nested named entities and document-level relationships.

Within this framework, we investigate the automatic extraction of relations from Russian texts at the document level using advanced transformer models. The long-tail nature of these texts, often exceeding 1024 tokens, necessitates modifications to existing models to ensure accurate and efficient processing.

Document-level RE is inherently computationally intensive and poses challenges with regard to resource allocation. To address this, our research includes strategies to balance the quantity of negative relationships in the training set, thus optimizing memory utilization and accelerating model convergence, while maintaining the high quality of the models.

Through these experiments, our research aims to further the capabilities of RE in Russian language texts and to provide a robust foundation for subsequent research in the field.

2 Related Work

In sentence-level relation extraction approaches, the primary focus lies on extracting relations between two entities within a sentence. However, many real-world relations can only be extracted by reading across several sentences, leading to the more complex task of document-level relation extraction. For this reason, document-level relation extraction still attracts the attention of many researchers [10]. The document-level relation extraction cannot be reduced to cross-sentence relation-extraction. The main modern approaches to document-level relation extraction include graph-based models and transformer-based models.

One of the earliest works was the paper by Yao et al. [16], where the authors proposed the DocRED dataset and as a basic solution, adapted CNN, LSTM, BiLSTM, context-aware LSTM [7] models for document-level relation extraction. Graph-based models are widely used in relation extraction due to their efficiency and ability to justify predictions. In the study [4], a model that combines representations learned across various textual areas in the document and across the hierarchy of sub-relations was proposed. In the work [1], a graph-based model with directed edges (EoG) was proposed for document-level relation extraction. Explicit graphical justification can bridge the gap between entities appearing in different sentences, mitigating the dependence on long distances, and achieving promising performance.

Contrary to this, given the transformer architecture which can implicitly model long-distance dependencies, some researchers directly apply pre-trained language models without creating document graphs. In the work [12], a two-step

approach to training in DocRED was proposed using BERT vector representations. In the study [18], a novel transformer-based model (ATLOP) with an adaptive threshold and a localised context pool based on BERT was proposed.

Subsequent approaches such as DocUNET [17], based on semantic segmentation, were proposed in line with the idea of segmentation in computer vision. At present, the state-of-the-art approach in the DocRED dataset is the DREEM model proposed in the paper [6], which uses an attention mechanism to guide the model to the most relevant statements in the document.

Significantly, there has been a lack of experimentation with document-level relation extraction in the Russian language. Although there are several Russian datasets annotated with relations [2, 3, 8], there is a noticeable absence of research focus on this area. In 2020, the RuReBus evaluation event was organized, specifically dedicated to the advancement of relation extraction in Russian [3]. This event produced promising outcomes with the R-BERT model, which showed superior results [13]. But in recent RuReBus and RuRed datasets, relations were annotated on the sentence level. In particular, the NEREL dataset [5] is the first Russian resource that includes cross-sentence relation annotations suitable for document-level relation extraction research. In [5] some initial experiments in cross-sentence relation extraction based on NEREL were described, which showed complexity and low performance of extracting such relations.

While the DeepPavlov framework¹ incorporates model for document-level RE (ATLOP model [18]), known for its adaptive thresholding and localized context pooling features, it is worth noting that our study considers more recent advancements. Specifically, the DREEM model (2023) [6], which we include, can be seen as an extension or enhancement of the ATLOP model. This allows us to explore newer approaches in document-level relation extraction. Also it is important to highlight that DeepPavlov’s model does not offer a straightforward way to be retrained on new datasets. Additionally, the relationships predicted by DeepPavlov are based on the RuRed dataset [2], which differs from the NEREL dataset used in our study and mainly contains relation annotations on the sentence level. This presents limitations in terms of adaptability and scope.

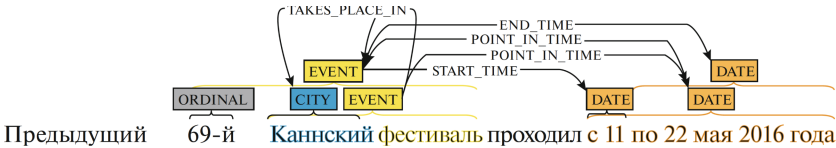


Fig. 1. Example of nested entities and relations between them in a sentence “The previous 69th Cannes Film Festival took place from 11 to 22 May 2016”.

¹ <https://docs.deeppavlov.ai/en/master/features/models/re.html>.

3 Dataset

Supervised models for RE require adequate amounts of annotated data for their training. It is time-consuming to manually label large-scale training data.

NEREL [5] is a dataset in Russian for named entity recognition and relation extraction based on Russian Wikinews documents. NEREL is much larger than the existing Russian datasets: today it contains 933 documents annotated with 29 entity and 49 relations types. NEREL also contains an event annotation involving named objects and their roles in the events.

An important distinction of NEREL from the previous datasets is the annotation of nested named entities (up to 6 levels of nestedness), as well as relations within nested entities and at sentence and the document level. NEREL allows the development of new models that can extract relations between nested named entities as well as relations at both the sentence and document levels. Figure 1 shows an example of nested entities and relations between them on sentence-level.

All the annotated relations can be subdivided into cross-sentence (24% of the total number of relations) in which the subject and object are in different sentences and within sentence relations (76% of the total number of relations).

Figure 2 shows an example of cross-sentence relations between entities. In NEREL there are also such relations as *alternative_name* and *abbreviation*, which allow connecting mentions of the same entities in the text. In Fig. 2 such relations are established between “Konstantin Ryabinov” and “Kuzia UO” mentions. A special document-level representation can be generated from the available NEREL annotation, which lists all mentioned entities and relations between them on the document level.

The average text length is 264 words. However, the maximum length is significantly larger at 1991 words, notably more than the standard DOCRED dataset for English, where the text length does not exceed 1024 tokens. This necessitates handling long sequences. Moreover, this task demonstrates a quadratic dependency on the number of entities in a document. It is crucial to note the large number of entities in a document: the 99th percentile equals 89 entities, while the 95th percentile equals 69 entities.

4 Approaches

A document can contain multiple mentions of an entity and its relations. The task of the document-level relation extraction is as follows: having a list of extracted entities (not mentions) to determine relations between them.

The DocRED dataset is one of the most popular and widely used benchmarks for document-level relation extraction. To identify the most effective models, we analysed the state-of-the-art models for DocRED and ReDocRED [11] listed on <https://paperswithcode.com/> in February-March 2023. From this list, we selected different top-performing approaches that not only demonstrated superior performance, but also had publicly available working code and could be

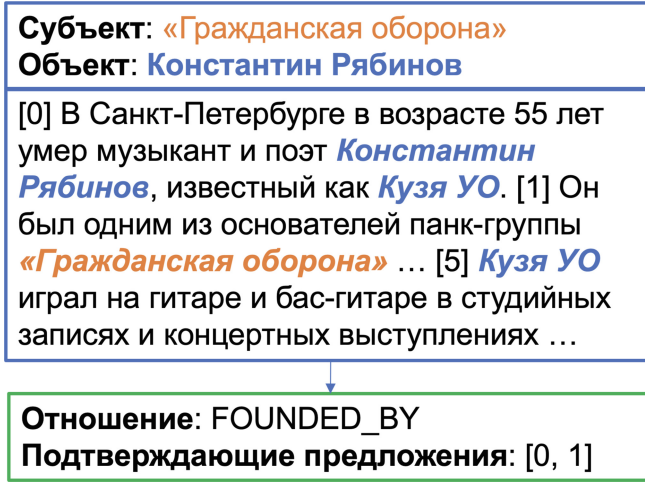


Fig. 2. Example of document level relation in text: “[0] In St. Petersburg, musician and poet Konstantin Ryabinov, known as Kuzia UO, has died at the age of 55. [1] He was one of the founders of the punk band “Grazhdanskaya Oborona” ... [5] Kuzia UO played guitar and bass in studio recordings and concert performances ...”.

adopted to reproduce on a Russian dataset. In addition, we considered it essential to establish a robust baseline for our experiments. To this end, we adopted the model recommended as a baseline by the authors of the DocRED paper. This baseline serves as our initial point of comparison, enabling us to evaluate the performance of the other models and iteratively improve our approaches.

4.1 BiLSTM and Context-Aware LSTM

For document-level relation extraction, the creators of the DocRED dataset proposed an adaptation of BiLSTM and context-aware LSTM models as a baseline solution. This implementation is described as follows:

- Document Encoding: BiLSTM is deployed to transcribe a document of n words into a sequence of corresponding hidden state vectors, denoted as h_i , with i spanning from 1 to n .
- Input Features: For every word within the document, the input features are constituted by the concatenation of GloVe word embeddings, entity-type vectors, and coreference vectors.
- Entity Representation: The representation of each named entity mention is determined as the mean of the hidden states for the words encompassed within that particular mention.
- Relation Prediction: The prediction of relations between pairs of entities is formulated as a multi-class classification task, employing a bilinear function for this purpose.

4.2 DocuNET

DocuNET [17] models relation extraction by predicting entity-level relation matrices akin to semantic segmentation in computer vision. The proposed model is constituted of:

Encoder Module. This module utilises a pre-trained language model to yield vector representations of tokens. To manage extensive documents, a dynamic window is used, and entity vector representations are derived through Logsum exp pooling.

Relation Matrix Calculation Module Using entity vector representations, an entity relation matrix is computed, employing both similarity-based and context-based features for entity pairs.

U-shaped Segmentation Module Employing a U-Net architecture, this module treats the entity-level relation matrix as a D-channel input image. It integrates both local and global information in predicting entity relationships via two down-sampling and up-sampling blocks with skip connections.

Classification Module Using a bidirectional linear function, this module predicts relationship probabilities between entity pairs. A balanced softmax method, inspired by circle loss in computer vision, addresses imbalanced distribution of relations.

4.3 DREEAM

Research indicates that evidence sentences, defined as those containing information about the relationship between a pair of entities, can aid DocRE systems in concentrating on the relevant text, thereby improving relation extraction. DREEAM [6] enhances the ATLOP [18] model by directing attention modules to assign higher weights to evidence-containing sentences without introducing any additional learnable parameters. Nevertheless, evidence prediction (ER) in DocRE encounters two principal challenges: high memory consumption and limited availability of evidence annotations. The model can be used for both supervised learning and self-training to compensate for the lack of manually annotated evidence, employing the same architecture with various training signals. A self-training ER process is proposed, comprising four main stages:

- Training a teacher model on human-annotated data, with labeled relations and supporting sentences;
- Applying the trained teacher model for evidence prediction on weakly-labeled data;
- Training a student model on weakly-labeled data with ER from earlier automatically generated evidence;
- Fine-tuning the student model on manually annotated data for model refinement.

Inference Procedure At this stage, it computes the importance of sentences based on an attention threshold, selecting sentences with importance exceeding a predefined threshold as evidence. A pseudo-document is constructed from the sentences deemed important. A single-parameter overlay layer, represented by τ , which represents the threshold value, is used to aggregate predictions from the pseudo-document and the entire document. Each triplet relation is chosen as the final prediction only if the sum of its aggregated values across the entire document and pseudo-documents exceeds τ . The τ parameter is tuned to minimize binary cross-entropy losses in relation extraction on the validation set.

Experiments on DocRED demonstrate that DREEAM currently achieves state-of-the-art results on the tasks of relation extraction and evidence detection, including through weakly-supervised learning on data obtained as a result of distant relation prediction on a large dataset.

5 Experiments

5.1 Data Processing

Initially, NEREL was annotated in the format of the brat annotation tool [9], in which mentions of entities and relations between them were established. For the inference models, we converted the NEREL dataset into the DocRED format, which lists entities, relations between them and supporting sentences. In adapting to the requirements of our task, we converted the entity mentions from character-based indices in the text to token-based indices in the sentence, subsequently converting this data into the DocRED format.

Initially, for baseline experiments with BiLSTM we excluded large documents exceeding 512 words from the training dataset, as tokenized using the Natasha library. This action affected less than 16 out of the total 746 documents. These particular documents were not only lengthy but also possessed a high volume of entities, causing a significant surge in GPU memory usage in certain models. For experiments involving transformer models, only a few documents were removed that had an extremely high number of entities, constituting the 95th percentile of this number across the entire dataset.

The authors of the original dataset provided a split of 80:10:10 for training, validation, and test sets. The metrics reported in the results tables have been calculated based on the validation set (also referred to as the dev set) over numerous runs. Furthermore, we have verified that the performance on the test set yields similar metric values.

We should also note the sequence of entities in the relationships in the data. In the DocRED dataset, the head and tail entities are sorted based on their order of appearance in the text, with the head entity always preceding the tail entity. However, in the NEREL dataset, the order is determined by the semantic relation, which might introduce extra complexities for the model when discerning which entity is the head and which is the tail. We conducted our experiments without altering the order of entities in these relationships.

BiLSTM and Context-Aware LSTM. To construct a basic solution for this task, models from the DocRED paper were employed.

- Distilled GLOVE embeddings, specifically the `navec_news_v1_1B_250K_300d_100q` from the Navec library of the Natasha project, were adopted as pre-trained vector representations for words. These representations were selected because:

- a) Their dimensionality is 100, similar to that used for DocRED, which required fewer modifications in the original model. Moreover, in comparison to the vectors from RusVectores, these vectors take up about 10 times less space (50 MB).

- b) These embeddings were trained on news datasets: `lenta`, `ria taiga_fontanka`, `buriy_news`, `buriy_webhose`, `ods_gazeta`, `ods_interfax`, which align well with the nature of the NEREL dataset.

- Metadata were also prepared: mappings of entity and relation titles to indices (`ner2id.json`, `rel2id.json`), a word embedding dictionary (`vocab.json`).

- The model implementation required a parameter for the maximum number of relations in a document. An estimate of the upper bound was taken as 93^2 , representing the square of the maximum number of entities in a document (excluding mentions) in the training set after preliminary filtering.

- One of the main problems encountered in solving this task is the high computational resource requirements of the algorithms. Specifically, a significant amount of memory is required for storing and processing information about entities and their relations at the document level. Moreover, the complexity of the algorithms often has an order of n^2 , where n is the number of entities in the document. This stems from the necessity to determine the presence and type of relation between every pair of entities, leading to an exponential growth in the amount of computations as the number of entities increases.

Furthermore, in the dataset under consideration, the number of negative relations can exceed the number of positive ones by up to 80 times, leading to a severe imbalance during training.

One approach to undersampling negative relations involves limiting the number of negative relations in a document to a certain boundary, for instance, `negative_bound = K * {number of positive examples}`, where K is the undersampling coefficient, $K = [15, 30, 50]$. This method enables a reduction in the number of negative relations involved in training without a significant loss in model quality, as well as lowering memory consumption and accelerating algorithm convergence.

Undersampling negative relations can reduce memory load and expedite algorithm convergence, which in turn can shorten training time and simplify the process of model tuning.

Experiments demonstrate that randomly eliminating negative relations from the training set leads to reduced memory usage and faster algorithm convergence (Fig. 3), without greatly impacting the final quality of the models. This indicates that methods for undersampling negative relations can be beneficial in the context of document-level relation extraction, as they optimise the training process and enhance the efficient utilisation of computational resources.

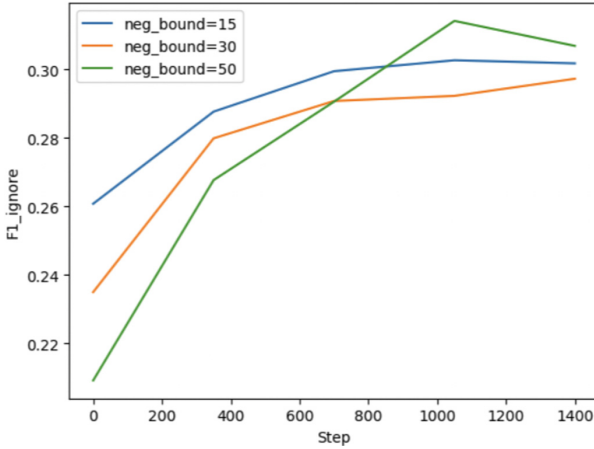


Fig. 3. F1_ignore score progression during training for different undersampling rates and their effect on convergence speed.

5.2 DREEAM and DocuNET Model

Handling Long Sequences. The document-level relation extraction task involves processing the entire text. However, transformer neural network architectures, such as BERT and RoBERTa, have limitations on the length of the input text, with the maximum number of tokens capped at 512.

In the applied models, this issue is addressed by dividing the input text into chunks, each of which is processed separately by the encoder. In the model implementations, documents containing more than 512 tokens (but fewer than 1024 tokens) are split into two parts: one for the first 512 tokens and another for the last ones. In other words, tokens at the very beginning/end of the document are encoded only once as they appear only in the first/second half, whereas tokens between them may be encoded twice as they can appear in both halves. The final vector representation of each token is calculated as the average of its vector representations in the first and second halves.

In the current implementation, the maximum number of chunks is 2, implying that the input text should not exceed 1024 tokens, which suffices for the DocRED dataset. However, in the NEREL dataset, some documents are longer, leading to incomplete processing of such documents. Despite this, the length on the filtered dataset does not exceed 1500 tokens.

Moreover, the model does not implement a mechanism for truncating long texts to 1024 tokens. This restriction precluded resolving the problem of long texts in such a way and resulted in errors. It is important to note that truncating documents necessitates additional adjustments to the training data (document-level relations), which adds complexity to the implementation of this feature.

To apply these models to the NEREL dataset, we made changes to the software implementation of the model. We added the capability to include

additional chunks, which allowed for more accurate processing of sequences up to 1536 tokens. This required alterations in handling long sequences and adjusting the attention indices of input tokens for proper chunk division. No significant changes were noticed in the resulting metrics when comparing the results before and after the fix. This can be attributed to the relatively small number of documents exceeding 1024 tokens in length. However, this enhancement helped avoid critical errors on long texts.

Incorporating Nested Entities

NEREL contains nested entities, for which relationships with other entities are also annotated. Working with texts with nested entities can introduce additional problems for many models due to the peculiarities of input processing and entity overlaps. Therefore, they are often ignored, leading to information loss and lower metrics.

To account for nested entities, we have chosen models that use entity markers for entity representation. According to work [15], this helps avoid the described issues. Specifically, for the BERT encoder, the entity classes are denoted by specific tokens, while in the RoBERTa model, distinct symbols are used. This strategy facilitates these models' ability to extract document-level relations, which is particularly beneficial when handling nested entities, a capability evident in the outcome of our model predictions.

Employed Encoder Variations

The transformer models mentioned in this study make use of several pretrained transformer models tailored for the Russian language. These function as encoders and include models such as xlm-RoBERTa-large, DeepPavlov/ruBert, ruBert-base, ruBert-large, and ruRoBERTa-large, the last three being offerings from Sber.

Our selection of these models was influenced by their performance on the Russian SuperGLUE benchmark. We considered the suitability of their architecture and their demands on GPU resources. Although we also attempted to incorporate RuLeanALBERT into our system, it proved somewhat incompatible with the current software implementation of our models. Therefore, despite its potential, its application was sidelined pending further optimization work.

It is also worth noting that NEREL does not contain weakly labelled data, which means an approach using student-teacher learning methods cannot be applied in this case.

For evaluation, we used F1 and F1_ignore as the evaluation metrics. F1_ignore denotes F1 score excluding relational facts shared by the training and development/test sets. Our experimental outcomes, as detailed in Table 1, reveal a variable performance across models, contingent on the encoders in use. Notably, the DREEM + Inference-stage Fusion model [14], in combination with the ruRoBERTa-large encoder, delivered the highest F1 scores. These findings underscore the critical role that the encoder selection plays in determining the ulti-

mate effectiveness of a model in performing document-level relation extraction (Table 2).

The code for experiments, along with additional supplementary material, is available on GitHub. Interested readers can access the repository at https://github.com/iden-alex/ru_relation_extraction.

Table 1. Document-level relation extraction results on NEREL

Model	Encoder	F1_ignore	F1
BILSTM	Russian Navec glove vectors	31.42	32.24
ContextAware	Russian Navec glove vectors	30.88	31.73
DocuNET	DeepPavlov/ruBert	53.55	54.04
DocuNET	ruBert-base	53.92	54.4
DocuNET	ruBert-large	52.15	52.64
DocuNET	xlm-Roberta-large	57.22	57.67
DocuNET	ruRoberta-large	53.21	53.72
DREEAM	ruBert-base	51.7	52.78
DREEAM	ruBert-large	51.77	52.78
DREEAM	xlm-Roberta-large	62.126	62.99
DREEAM	ruRoberta-large	67.07	67.73
DREEAM + Inference-stage Fusion	ruRoberta-large	67.87	68.53

Table 2. Evidence sentence prediction results on NEREL

Model	Encoder	F1_evidence
BILSTM	Russian Navec glove vectors	–
ContextAware	Russian Navec glove vectors	–
DocuNET	DeepPavlov/ruBert	–
DocuNET	ruBert-base	–
DocuNET	ruBert-large	–
DocuNET	xlm-Roberta-large	–
DocuNET	ruRoberta-large	–
DREEAM	ruBert-base	55.73
DREEAM	ruBert-large	56.13
DREEAM	xlm-Roberta-large	58.98
DREEAM	ruRoberta-large	58.74
DREEAM + Inference-stage Fusion	ruRoberta-large	–

6 Conclusion

We have undertaken the first exploration of the task of document-level relation extraction for the Russian language. Using state-of-the-art models that achieve the best results on the main English dataset for doc-level RE, a series of novel experiments were conducted with Russian texts at the document level. To enable the accurate execution of these experiments, the model was refined to process texts exceeding 1024 tokens in length.

The task of document-level relation extraction is computationally demanding, necessitating substantial computational resources. Consequently, the research included the application of undersampling the number of negative relations in the training sample for certain models. This approach optimizes memory usage and expedites convergence, without significantly deteriorating the final quality of the models.

As a result of the conducted research and experiments, we demonstrated that the proposed methods and models can successfully extract relationships from Russian texts at the document level, inclusive of nested entities. This pioneering work in the field of Natural Language Processing signifies a considerable contribution to the advancement of research on document-level relation extraction for underrepresented languages, specifically Russian.

References

1. Christopoulou, F., Miwa, M., Ananiadou, S.: Connecting the dots: Document-level neural relation extraction with edge-oriented graphs. CoRR **abs/1909.00228** (2019). <http://arxiv.org/abs/1909.00228>
2. Gordeev D.I., Davletov A.A., R.A.A.G.R.G.G.A.: Relation extraction dataset for the Russian (2020)
3. Ivanin, V., et al.: Rurebus-2020 shared task: Russian relation extraction for business, pp. 416–431 (01 2020). <https://doi.org/10.28995/2075-7182-2020-19-416-431>
4. Jia, R., Wong, C., Poon, H.: Document-level n-ary relation extraction with multiscale representation learning. CoRR **abs/1904.02347** (2019). <http://arxiv.org/abs/1904.02347>
5. Loukachevitch, N.V., et al.: NEREL: A Russian dataset with nested named entities and relations. CoRR **abs/2108.13112** (2021). <https://arxiv.org/abs/2108.13112>
6. Ma, Y., Wang, A., Okazaki, N.: Dreeam: Guiding attention with evidence for improving document-level relation extraction (2023)
7. Sorokin, D., Gurevych, I.: Context-aware representations for knowledge base relation extraction. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 1784–1789. Association for Computational Linguistics, Copenhagen, Denmark (Sep 2017). <https://doi.org/10.18653/v1/D17-1188>, <https://aclanthology.org/D17-1188>
8. Starostin, A., et al.: Factrueval 2016: evaluation of named entity recognition and fact extraction systems for Russian (2016)
9. Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., Tsujii, J.: brat: a web-based tool for NLP-assisted text annotation. In: Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics, pp. 102–107. Association for Computational Linguistics, Avignon, France (Apr 2012). <https://www.aclweb.org/anthology/E12-2021>

10. Tan, Q., Xu, L., Bing, L., Ng, H.T., Aljunied, S.M.: Revisiting docred – addressing the false negative problem in relation extraction (2022)
11. Tan, Q., Xu, L., Bing, L., Ng, H.T., Aljunied, S.M.: Revisiting docred – addressing the false negative problem in relation extraction (2023)
12. Wang, H., Focke, C., Sylvester, R., Mishra, N., Wang, W.Y.: Fine-tune BERT for docred with two-step process. CoRR abs/1909.11898 (2019). <http://arxiv.org/abs/1909.11898>
13. Wu, S., He, Y.: Enriching pre-trained language model with entity information for relation classification. CoRR abs/1905.08284 (2019). <http://arxiv.org/abs/1905.08284>
14. Xie, Y., Shen, J., Li, S., Mao, Y., Han, J.: Eider: Empowering document-level relation extraction with efficient evidence extraction and inference-stage fusion (2022)
15. Yandutov, A.V., Loukachevitch, N.V.: Approaches to relation extraction for nested named entities. Lobachevskii J. Math. **44**, 249–258 (2023). <https://link.springer.com/article/10.1134/S1995080223010456>
16. Yao, Y., et al.: DocRED: A large-scale document-level relation extraction dataset. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 764–777. Association for Computational Linguistics, Florence, Italy (Jul 2019). <https://doi.org/10.18653/v1/P19-1074>, <https://www.aclweb.org/anthology/P19-1074>
17. Zhang, N., et al.: Document-level relation extraction as semantic segmentation. In: Zhou, Z.H. (ed.) Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, pp. 3999–4006. International Joint Conferences on Artificial Intelligence Organization (8 2021). <https://doi.org/10.24963/ijcai.2021/551>, <https://doi.org/10.24963/ijcai.2021/551>, main Track
18. Zhou, W., Huang, K., Ma, T., Huang, J.: Document-level relation extraction with adaptive thresholding and localized context pooling. CoRR abs/2010.11304 (2020). <https://arxiv.org/abs/2010.11304>