



Article

Extraction of the Relations among Significant Pharmacological Entities in Russian-Language Reviews of Internet Users on Medications

Alexander Sboev^{1,2,*}, Anton Selivanov¹, Ivan Moloshnikov¹, Roman Rybka¹, Artem Gryaznov¹, Sanna Sboeva¹ and Gleb Rylkov¹

¹ National Research Centre “Kurchatov Institute”, 123182 Moscow, Russia; Selivanov_AA@nrcki.ru (A.S.); Moloshnikov_IA@nrcki.ru (I.M.); Rybka_RB@nrcki.ru (R.R.); Gryaznov_AV@nrcki.ru (A.G.); Sboeva_SG@nrcki.ru (S.S.); Rylkov_GV@nrcki.ru (G.R.)

² Moscow Engineering Physics Institute, National Research Nuclear University, 115409 Moscow, Russia

* Correspondence: Sboev_AG@nrcki.ru

Abstract: Nowadays, the analysis of digital media aimed at prediction of the society’s reaction to particular events and processes is a task of a great significance. Internet sources contain a large amount of meaningful information for a set of domains, such as marketing, author profiling, social situation analysis, healthcare, etc. In the case of healthcare, this information is useful for the pharmacovigilance purposes, including re-profiling of medications. The analysis of the mentioned sources requires the development of automatic natural language processing methods. These methods, in turn, require text datasets with complex annotation including information about named entities and relations between them. As the relevant literature analysis shows, there is a scarcity of datasets in the Russian language with annotated entity relations, and none have existed so far in the medical domain. This paper presents the first Russian-language textual corpus where entities have labels of different contexts within a single text, so that related entities share a common context. therefore this corpus is suitable for the task of belonging to the medical domain. Our second contribution is a method for the automated extraction of entity relations in Russian-language texts using the XLM-RoBERTa language model preliminarily trained on Russian drug review texts. A comparison with other machine learning methods is performed to estimate the efficiency of the proposed method. The method yields state-of-the-art accuracy of extracting the following relationship types: ADR–Drugname, Drugname–Diseasename, Drugname–SourceInfoDrug, Diseasename–Indication. As shown on the presented subcorpus from the Russian Drug Review Corpus, the method developed achieves a mean F1-score of 80.4% (estimated with cross-validation, averaged over the four relationship types). This result is 3.6% higher compared to the existing language model RuBERT, and 21.77% higher compared to basic ML classifiers.

Keywords: pharmacological text corpus; automatic relation extraction; natural language processing; deep learning



Citation: Sboev, A.; Selivanov, A.; Moloshnikov, I.; Rybka, R.; Gryaznov, A.; Sboeva, S.; Rylkov, G. Extraction of the Relations among Significant Pharmacological Entities in Russian-Language Reviews of Internet Users on Medications. *Big Data Cogn. Comput.* **2022**, *6*, 10. <https://doi.org/10.3390/bdcc6010010>

Academic Editor: Min Chen

Received: 31 October 2021

Accepted: 27 December 2021

Published: 17 January 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The developing ecosystem of social networks and other special Internet platforms expands the possibility of discussion of a broad set of topics in textual format. These texts often contain people’s publicly available opinions on various subjects. One of the topics of special interest is Internet reviews on medications, including information about their positive and adverse effects, qualities, manufacturers, administration regime etc. Such information could be useful for comprehensive analysis for the purposes of pharmacovigilance [1] and potential medicine re-profiling.

Analysing such a large amount of information is a time-consuming task, therefore requiring methods for automated extraction of pharmacologically-meaningful data. In

turn, these methods require textual corpora with annotation of pharmacological entities and their relations.

There is a wide variety of English-language datasets in literature sources, for example Drug–Drug Interaction (DDI) and Adverse Drug Event (ADE). These corpora contain pharmaceutically relevant entities of different types as well as relationships between them. A more detailed analysis of the corpora is presented in Section 2. However, currently there is only one large domain-oriented dataset in the Russian language: Russian Drug Review Corpus of Internet User Reviews with Complex NER labeling (RDRS), which was presented by our group [2,3]. Now, we present (in Section 3.1) an extension of this corpus that includes annotation of relationships among the named entities that are most relevant for the potential studies of drug efficiency.

The automation of the process of extracting meaningful information from a review written in a natural language requires solving the following tasks: text segmentation, Named Entity Recognition (NER), Relation Extraction (RE), structuring of the extracted information, and evaluation of the results. In this paper, we focus on the relation extraction task. The formulation of the relations extraction task in natural language processing is as follows: given a text and two entities from it, determine if there is a relation of a certain type between the entities. For example, in the text “Antiviral syrup for children Orvirem—we have an allergy to it!” with the entities “Orvirem” and “allergy”, the task is to determine that the allergy is mentioned as the adverse effect of the “Orvirem” medication.

The relation extraction task can be solved by two approaches: the sequential (cascade) approach of solving the named entity recognition and relation extraction tasks separately, or the combined approach of solving these tasks simultaneously (the combined approach, called “joint” or “end-to-end” in the literature). The sequential solution allows estimating the accuracy of solving each task separately, thus leading to more thorough analysis of the task complexity; therefore, the scope of our research is to analyze the relation extraction model within the sequential approach, applied to entities already extracted.

Our review (see Section 2) showed that the most promising technology that can be utilized to solve the relation extraction task is deep learning. This paper uses a model based on the XLM-RoBERTa language model, pre-trained on a huge unlabelled corpus of drug reviews. Section 3 contains the details of the model configuration and setting.

Based on this model, a set of computational experiments are performed (Section 5) on different parts of the RDRS corpus. In Section 5.1, the optimal model parameters and text representation are obtained using a part of the corpus that includes texts with ADR–Drugname relations. Section 5.2 presents evaluations on a subset of the corpus containing reviews with multiple contexts. This experiment is aimed at obtaining the state-of-the-art results for the task of relation extraction for the following four relation types: ADR–Drugname, Drugname–Diseasename, Drugname–SourceInfoDrug, Diseasename–Indication. The results of the proposed model are compared with the results of the existing language model RuBERT, as well as a set of baseline methods: multinomial naive bayes classifier, linear support vector machine, and dummy classifiers.

The main contributions of our work include:

- The relation extraction method is proposed, in which the task of determining the presence of a relation is formulated using multi-context annotation: entities belonging to the same context are considered to be related. The method is based on a language model fine-tuned to classify entity pairs by the presence of relations.
- Several variations of the text representation used to present the entities under consideration to the language model are compared, and the optimal representation is shown to be the one that includes the text of target entities along with the whole review text, concatenated with special tokens;
- The method based on a language model trained on a large corpus of unlabeled Russian drug review texts and fine-tuned on an annotated corpus of Russian drug reviews is shown to be applicable to the task of determining the relations among

pharmaceutically-relevant entities of the newly-created corpus. The accuracy estimations are obtained for this task for Russian language;

- The same proposed model, pre-trained on Russian drug reviews, is shown to achieve relation extraction results comparable to the state of the art on the DDI corpus.

2. Related Works

Currently, efficient natural language processing (NLP) methods are mostly based on neural network algorithms [4,5]. There is a wide variety of text analysis tasks that neural networks can solve with high accuracy, such as part of speech tagging [6], machine translation [7], authorship attribution [8], named entity recognition [9–13], entity relation extraction [14–16] and so on. In this paper, our research is aimed at the relation extraction task for the pharmacological domain.

There are two main approaches to the extraction of the relations between entities from a text:

- cascade approach: sequential text analysis, where the tasks of named entity recognition and relation extraction are performed separately. At the first stage, named entities are extracted from the text, either by expert annotation or using a machine learning model [3,9,13,17,18]. At the second stage, the entities extracted are evaluated in terms of their possible relations [14,15,19,20]. This approach allows one to control the learning process of each model, which in turn gives the opportunity for a more thorough choice of methods and hyperparameters.
- “joint” approach, where a single model is used to solve both tasks simultaneously [16,21–23]. The most commonly used neural network topology for this model has two parts (one for each task) that learn jointly using combined loss function.

This work is focused on the separate analysis of the relation extraction method which could be used as a part of a cascade model.

The development of tools for textual data analysis depends on the annotated data necessary for tuning the algorithms and assessing their performance. There are a number of corpora of textual data in English language with a markup of pharmacologically-relevant entities and relationships. These corpora differ by the types of texts (online reviews, tweets, clinical discharges etc.) and by the types of the named entities and relationships annotated, which are specified with varying levels of detail. Some studies provided the achievable accuracy of extracting relationships between pharmacologically-relevant entities using methods developed on base of these corpora.

Among the datasets on biomedical topics with markup suitable for solving the problem of identifying relationships between named entities, it is necessary to mention the corpora of the i2b2 Competition Corps Workshop on Natural Language Processing Challenges for Clinical Records. The competition is organized by the Department of Biomedical Informatics (DBMI) at Harvard Medical School. This organization provides datasets called n2c2, that consist of full texts of clinical records in English. The data annotation is enriched at every new competition as the scope of the competition expands and changes.

The task of extracting relationships between the named entities was considered in the 2009 [24], 2010 [25] and 2018 [26] corpora from the above-mentioned competition. The best-performing models used additional sources of information, handcrafted and engineered features, which facilitated better classification of the entities and relations in the medical discharges in English. On the Drug-Drug Interaction (DDI) dataset [27], which contains excerpts about drug interactions from the DrugName and MedLine databases, a model [28] based on a BERT-type language model SciBERT [29] was used to solve the task of classifying the sentences for relationships between the selected drugs. The model is, in fact, the BERT language model, additionally trained on scientific texts for better representation of the thematic. The model showed a result of 84.08% by the f1-micro metric.

On the Adverse Drug Event (ADE) dataset [30], which contains sentences from the abstracts of PubMed scientific articles with relations between medical drugs and their adverse reactions, the performance of a BERT-based model SpERT [16] was presented. The

model solves the entity extraction and relation extraction problems sequentially. To solve the problem of identifying named entities, all possible consecutive sets of words in the text (of limited length) are generated and then classified by the model according to the type of the entity. The entities obtained as the result of the classification are then filtered, forming pairs of entities for which the model determines the presence of a relationship. The model uses BERT to represent words as vectors. Word vectors are then processed by a fully-connected layer with softmax activation function and with the size equal to the number of the named entity types. This layer identifies named entities from the input word combinations. The second part of the model is a fully-connected neural network layer with the sigmoid activation function, the input of which is the concatenation of the entity vectors with the vector representation of words between the entities. Relations are determined by applying the threshold to the output activities. Such model achieves the f1-macro metric of 79.24% on the ADE dataset.

Regarding the Russian language, a fairly limited set of corpora for relation extraction tasks are publicly available. However, these corpora facilitate the a priori assessment of the accuracy in the extraction of the relationships between named entities of different types, not related to pharmaceuticals: RuSERRC [31]—80 manually annotated texts with entities from the computer science subjects (software, database, programming languages, etc.). RuREBus [32]—300 annotated texts of strategic programs of the Ministry of Economic Development of the Russian Federation, containing various relations between the entities of the following types: Social Objects, Actions, Goals, Tasks; RURED [33]—a corpus of 536 annotated texts about economics, containing the entity types of Geographic Objects, Names, Age, Currencies, etc., as well as the relationships of various types between them; Factrueval [19]—255 annotated texts with entities of type Persons, Locations and Organizations, and also relations: Ownership, Occupation, Meeting, and Deal; NEREL [34]—933 annotated documents with the markup of a large number of entities, including: Persons, Organizations, geopolitical entities, numbers, dates, time, money, age, etc., as well as relations between them.

On the RuSERC corpus (split by sentences), a BERT-based architecture, R-BERT [14], was used to obtain the result of 67% by the macro-f1 metric. On the RuRBus corpus (split by documents), the R-BERT architecture [14] was used to obtain the result of 44% by the micro-f1 metric.

The R-BERT model uses the BERT model to represent words as vectors. A vector representing a named entity is the average vector of its words. A concatenated pair of entities is presented as the input for the fully-connected layer, and the activities of the layer are the input for the softmax layer which is used to determine the type of a relation between the entities.

On the RURED dataset (split by sentences), the span-BERT architecture achieved 78% accuracy by the f1 metric (the method of aggregating f1 across different classes was not specified). On the Factrueval dataset (split by documents), that method achieved 66% accuracy on the fact extraction task (extracting relationships among multiple entities). On the NEREL dataset (split by documents), the RuBERT model achieved the f1-score value of 51% (the method of aggregating f1 across different classes was not specified).

As for the Russian-language corpora annotated to extract the relationships between pharmacologically significant entities, the only corpus of this type is the Russian Drug Review Corpus (RDRS 2800 reviews), which is considered in this paper. Therefore, the accuracy demonstrated in the works above with other types of texts is only an estimate of the possible accuracy of determining the relationships between pharmacological entities, which is an additional motivation to perform the present work.

Summarizing the above, it can be concluded that the current trend in identifying relationships between named entities is the use of models with transformer architecture pre-trained on large datasets. Further in this work, we develop this approach based on the XLM-RoBERTa language model [35] using the Russian Drug Review Corpus (RDRS) [3] described in Section 3.1 and available at the Sagteam project website (<https://sagteam.ru/u>).

3. Materials and Methods

3.1. Datasets

This paper uses the Russian Drug Review Corpus (RDRS) [3], which contains 2800 texts of drug reviews written by Internet users. The corpus contains markup for 18 types of named entities, which can be divided into 3 groups:

- Medication—this group includes everything related to the mentions of drugs and drugs manufacturers, including: Drug name, Drug class, Drug form, Route (how to use the drug), Dosage, SourceInfoDrug (source of the consumer’s information about the drug) etc.;
- Disease—this group contains entities related to the diseases or reasons for using the drug (disease name, indications or symptoms), as well as mentions of the effects achieved (NegatedADE—the drug was inefficient, Worse—some deterioration was observed, BNE-POS—the condition improved) etc.
- ADR—mentions of occurring adverse reactions.

In a subset of the corpus containing 1590 review texts, entities were marked up into “lines of meaning”—“contexts”, so that each context contains entities that describe the usage of some medication by one person for the treatment of one condition. Different contexts arise in a text, in particular, when describing the use of multiple drugs in the treatment, or different effects following the use of a single drug for different conditions, or when the review describes the use of a drug by different people. In terms of the relation extraction problem, entities that occur in the same context are related, while entities from different contexts are considered unrelated.

An example of the context annotation is shown in Figure 1. The main (1st) context of the review is about the drug “orvirem” which caused an allergy. This context includes the following mentions (denoted on the figure with a number 1 above them): “antiviral” (Drugclass), “syrup” (Drugform) “orvirem” (Drugname), multiple mentions of “allergy” (ADR), “red spots” (ADR), “swelling on the face” (ADR), “1 day” (Duration). There are other contexts in the review:

- 2nd context: “allergy” (Diseasename), “red spots” (Indication), “zyrtek” (Drugname), “the situation did not improve” (NegatedADE), “it seems to have gotten even worse” (Worse).
- 3d context: “allergy” (Diseasename), “red spots” (Indication), “doctor” (SourceInfoDrug), “On her recommendation” (SourceInfoDrug), “smecta” (Drugname), “the situation did not improve” (NegatedADE), “it seems to have gotten even worse” (Worse).
- 4th context: “allergy” (Diseasename), “red spots” (Indication), “doctor” (SourceInfoDrug), “On her recommendation” (SourceInfoDrug), “suprastin” (Drugname), “the situation did not improve” (NegatedADE), “it seems to have gotten even worse” (Worse), “Injected” (Route), “The redness seems to pass” (BNE-POS), “swelling on the face still remains” (NegatedADE).
- 5th context: “allergy” (Diseasename), “red spots” (Indication), “doctor” (SourceInfoDrug), “prednisone” (Drugname), “Injected” (Route), “The redness seems to pass” (BNE-POS), “swelling on the face still remains” (NegatedADE).

In Tables 1–4 the quantitative characteristics of the corpus with contextual markup are presented.

Table 1. The number of texts that contain the corresponding number of contexts.

Contexts Count	1	2	3	>3
Texts Count	682	559	218	131

Table 2. Average lengths of contexts in the corpus.

	Average Mentions Count	Average Tokens Count
Main context	19.9	38.9
Other contexts	3.7	6.6

Table 3. Statistics on the part of RDRS dataset that is comprised of ADR–Drugname relations.

Number of	Train	Test
Texts	502	126
Sentences	4016	1008
Words	82,425	20,961
“ADR” type entities	1461	356
“Drugname” type entities	1416	368
Relations	3444	845
Avg. numbers of relations per text	6.9	6.7

Table 4. Statistics on the types of relations in the RDRS corpus with 908 multi-context reviews.

Relation	ADR & Drugname		Drugname & Diseasename		Drugname & SourceInfoDrug		Diseasename & Indication	
	pos.	neg.	pos.	neg.	pos.	neg.	pos.	neg.
Relations count	1913	917	4277	2153	2700	1232	2588	701
Text fraction	0.273	0.204	0.634	0.514	0.598	0.457	0.416	0.148

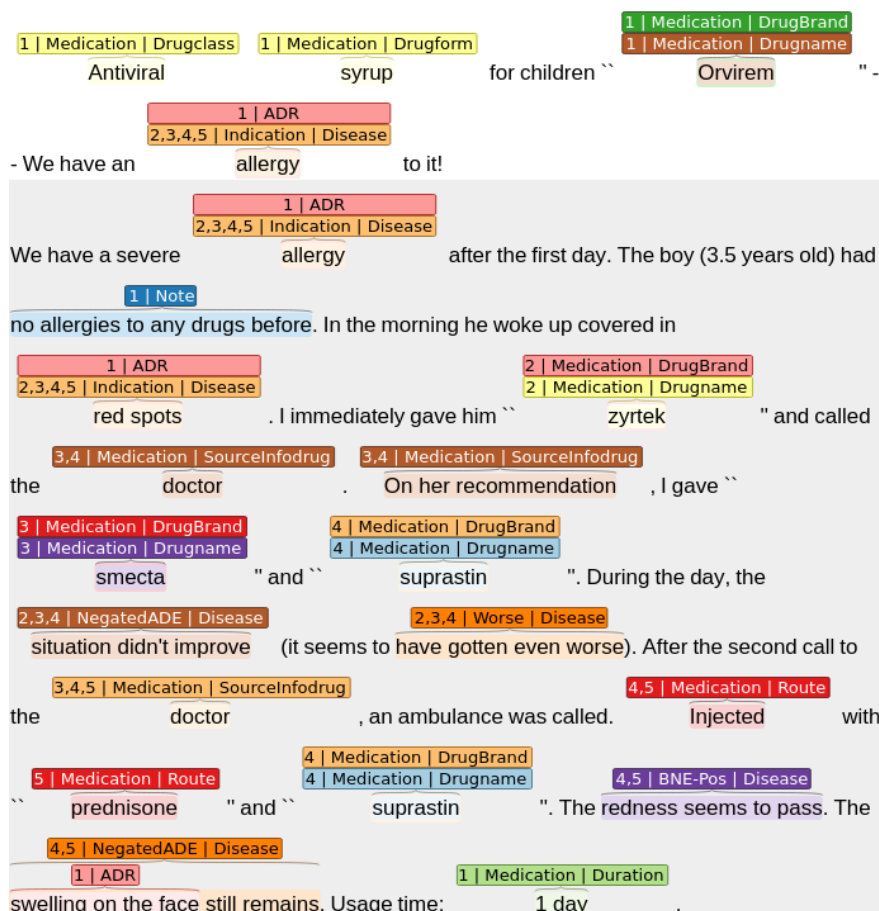


Figure 1. Example of an annotated review. The labels contain, from left to right: context number, entity type, attribute within the entity type.

In this paper, the following pairs of entities are chosen as the most interesting to analyze from the practical point of view:

- ADR–Drugname—the relationship between the drug and its side effects;
- Drugname–SourceInfodrug—the relationship between the medication and the source of information about it (e.g., “was advised at the pharmacy”, “the doctor recommended it”);
- Drugname–Diseasename—the relationship between the drug and the disease;
- Diseasename–Indication—the connection between the illness and its symptoms (e.g., “cough”, “fever 39 degrees”).

Two subsets of the original corpus have been compiled for the experiments:

1. The first one includes 628 texts containing ADR and Drugname entity pairs. The experiments on this part are aimed at selecting the most effective combinations of the input feature representations and hyper-parameters of the methods used. The texts of the RDRS corpus that contain ADR and Drugname entities were divided into training and test parts, the composition of which is presented in Table 3.
2. The second part includes texts that contain multiple contexts. The total number of such texts is 908. Statistics on the types of relationships are presented in Table 4. This corpus is used to establish the current level of accuracy in determining the relationships between pharmacologically-significant entities in Russian-language review texts.

Experiments with these subsets are described further in Section 4.

3.2. Methods

3.2.1. Deep Learning Methods

Language Models

In this work the XLM-RoBERTa-sag model [3] was used. The original XLM-RoBERTa [35] is a multilingual language model based on the transformer [36] architecture, consisting of multihead attention layers which create vector representations of the input data parts (words in case of NLP) that encode the information about their context. XLM-RoBERTa is trained on a large multilingual corpus from the CommonCrawl project that contains 2.5 TB of texts. XLM-RoBERTa-sag is a result of additional training of XLM-RoBERTa on a dataset of unlabeled internet texts about medicines (~1.65 M texts).

During the adjustment experiments, we used two versions of the model:

- XLM-RoBERTa-base-sag—12 Transformer blocks, 768 hidden neurons, 8 Attention Heads, 125 millions of parameters, 2 epochs of additional training on Russian texts about medications;
- XLM-RoBERTa-large-sag—24 Transformer blocks, 1024 hidden neurons, 16 Attention Heads, 355 millions of parameters, 1 epoch of additional training on Russian texts about medications;

Text preprocessing includes splitting it into words or word parts—“tokens”. For XLM-RoBERTa-sag, as well as for the original XLM-RoBERTa, such splitting is performed using the SentencePiece tokenizer [37].

Input Text Pre-Processing

To solve the classification task, transformer-based language models use a special token [CLS] added to the input sequence. During training, the loss function is aimed at class prediction based on the vector of the [CLS] token. That way, the model learns to create such a vector representation of the [CLS] token that accumulates information about the text as a whole and is informative in terms of the current task being solved.

In the approach proposed in this work, the classification is performed on the basis of the information about a pair of entities for which the existence of a relationship is determined, and the text that mentions this pair. Figure 2 shows the conceptual scheme of our approach to solving the relation extraction task using a language model.

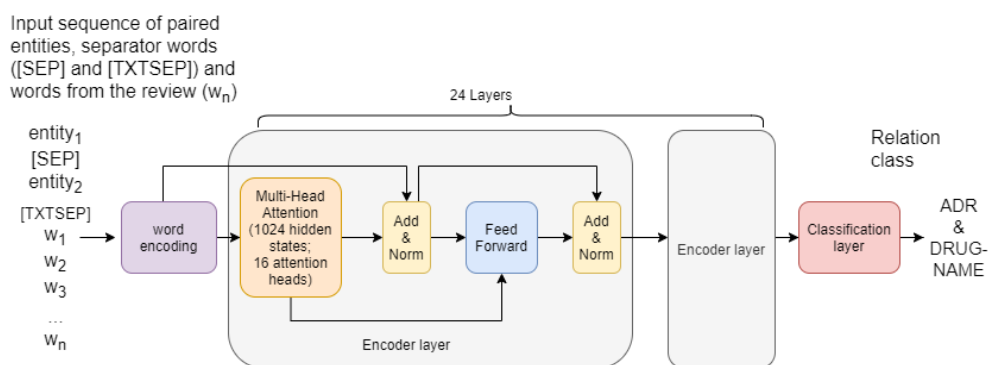


Figure 2. Conceptual scheme of our approach to relation extraction based on a language model.

For providing the language model with the information about which entities are of interest, several text representation variants are considered in our experiments:

1. The whole text—the tokenized input text that the language model receives at its input is the whole drug review text, in which target entities are highlighted using special start and end tokens, e. g. [T_ADR] and [\T_ADR] for an entity of type ADR:
 «[CLS]Antiviral syrup for children [T_DRUG]“Orvirem”[\T_DRUG] - We have an allergy to it! We have a severe allergy after the first day of taking it. Moreover, the boy (3.5 years old) had no allergies to any drugs before. In the morning he woke up covered in [T_ADR]red spots[\T_ADR]. I immediately gave him zyrtek... »
2. The text of target entities only—only the mentions of the target entities are used as the input text;
3. The text of the target entities and the text between the mentions of the target entities;
4. The text of the target entities concatenated with the whole text:
 [CLS]«text of first target entity»[SEP]«text of second target entity»[TXTSEP]«whole text of the drug review».

Here, the token [SEP] is placed between the two target entities, and the token [TXTSEP] separates the pair of entities from the whole text.

Potentially, this way of organizing the input data makes it possible to build a more informative vector representation due to the Attention mechanism inside the Transformer layers, and facilitates solving the problem in a classification formulation. The effectiveness of such a text representation was demonstrated previously [38].

As mentioned before, there are many degrees of freedom in such models that require consideration in order to achieve higher accuracy. Within the scope of the current research, the following options have been analyzed:

- Maximum input sequence length (in tokens);
- Learning rate;
- Batch size;
- Maximum learning epoch number;
- Learning rate decay technique [39];
- Early stopping technique [40].

3.2.2. Other Machine Learning Methods

Basic machine learning methods perform quite effectively in many applications [41–43]. These methods are highly efficient in terms of computational complexity, due to this fact, it is possible to search for the optimal set in an extensive space of hyperparameters and to test hypotheses relatively quickly.

The first goal of using basic machine learning methods was to obtain a reasonable baseline for the relation extraction task in the pharmacological domain in the Russian language, exceeding the random guess of the “Dummy” models’ results, for the purpose of comparison with the deep learning models described in the previous section.

As a textual data representation for the baseline methods, concatenation of frequency features (term frequency-inverse document frequency, TF-IDF) of the character n -grams of the target entities was used. The size of the n -gram n and the frequency filter of the tf-idf method were considered as the hyperparameters to tune during the experiments.

The second goal of using basic machine learning methods was to check if the information about the entities' text is sufficient to achieve competitive accuracy for the task.

The following methods were considered in the experiments as basic machine learning methods:

- Logistic regression [44]—a basic linear model for text classification using a logistic function to estimate the probability of an example to belong to a certain class;
- Support vector machine [45]—a linear model based on building a hyperplane that maximizes the margin between two classes;
- Multinomial Naive Bayes model [46]—a popular solution for baselines in such text analysis tasks as spam filtering or text classification. It performs text classification based on words' or n -grams' co-occurrence probability;
- Gradient Boosting [47]—a strong decision tree-based ensemble model, which iteratively “boosts” the result of each tree by building a next tree that should classify examples that the previous tree fails to classify correctly.

In addition, for comparison, the RuBERT [48] language model was considered, which is a BERT [49] model with 12 layers, 768 hidden neurons each, 12 attention heads, 180 M parameters. RuBERT was trained on the Russian part of Wikipedia and news data. When solving the problem, the language model is used to form a vector representation of the text, which is fed into the linear layer. The output activities of the linear layer are used to determine if there is a relationship between the pair of entities fed to the input.

3.2.3. Dummy Models

“Dummy” models were considered to be the low-level baseline. Such models generate labels randomly or according to some simple principle. The following methods were checked as methods for “dummy” classification:

- most frequent class labeling—every pair of entities is assigned to the most frequent class in the dataset (in case of extraction of ADR–DrugName relations in the RDRS dataset, thus classifier considers every pair to have a relation);
- uniform random labeling—labels are predicted randomly according to a uniform probability distribution, without taking into account any characteristics of the input dataset;
- stratified random labeling—labels are predicted randomly but from the distribution corresponding to that of the input data: the probability of an input example to belong to a class is proportional to the portion of examples of such class in the dataset.

The accuracy of the “dummy” methods based on the random label generation was averaged over 100 launches in order to operate with more stable results and to prevent possible occurrence of random outliers.

4. Experiments

4.1. Accuracy Metric

The performance of a model on the relation extraction task is estimated by the f1-macro metric, in which the f1 score is calculated separately for each class:

$$\text{f1score} = \frac{2 \cdot P \cdot R}{P + R}, \quad (1)$$

$$P = \frac{TP}{TP + FP}, \quad (2)$$

$$R = \frac{TP}{TP + FN}. \quad (3)$$

Here P is precision, the proportion of correctly predicted objects of the class A under consideration to the number of objects that the model assigned to class A ; R is recall, the proportion of correctly predicted items of class A to the real number of items in class A ; TP is the number of *true positive* instances, the number of relations of class A correctly identified by the model; FP is the number of *false positive* examples, the number of relations assigned to class A while actually having a different class; FN is *false negative*, the number of relations that actually have class A while being incorrectly assigned to a different class by the model.

The overall performance of the model is estimated by averaging the f1-score over the two classes. This method of averaging allows for the uneven numbers of relations in the different classes.

4.2. Selection of the Model Features and Hyperparameters

In these experiments we use a subset of RDRS that contains texts with the ADR and Drugname entities only. The following experimental setup is used:

- Fixed stratified split into training (80%) and testing (20%) sets; In order to avoid overfitting, entity pairs from each review all go either to the training set or to the testing set, but no review is split between the sets;
- Hyperparameters of the language model's fine-tuning process are searched manually so that to maximize the accuracy (by the f1-macro metric) on the validation part of the training set, without taking into account the testing set;
- The language model involves early stopping and learning rate decay (Experiments show the positive effect of such techniques on the model accuracy);

The experiments on language models have been carried out using a computing cluster node with the following configuration: CPU Intel® Xeon™ E5-2650v2 (2.6 GHz) × 8, 128 Gb RAM, NVIDIA Tesla V100 (16 Gb).

4.3. An Estimation of Efficiency of Selected Methods

After finding the optimal model parameters, the efficiency of the methods has been assessed on a part of the RDRS containing review texts with multiple contexts. Accuracy is measured by the f1-score metric with cross-validation over 5 splits: the data is divided into 5 equal parts, and at each iteration of the cross-validation 80% of the texts are used for fine-tuning the model and 20% for testing.

For a more complete assessment, we compare the proposed method to other machine learning methods different in terms of complexity and type, as well as to a "Dummy" classifier based on the probability distribution of positive and negative examples of the pairs of entities in question (Stratified random labeling).

"Dummy" models and basic machine learning method experiments have been carried out on a local machine with the following configuration: CPU Intel® Core™ i5-7400 @ 3.00 GHz × 4, 16 Gb RAM. The experiments with language models were performed on the same equipment as the experiments in the previous section.

The programming language python 3.8 and software libraries numpy [50], sklearn [51], pytorch [52] and simpletransformers [53] were used for software implementation of the described method. As part of a series of experiments, the parameters of the python random number generator, as well as the random number generators of numpy, sklearn, and pytorch libraries were fixed to ensure repeatability of the experiments.

5. Results

5.1. Comparison of the Model Features and Hyperparameters

This section compares the results of experiments on the identification of the entity relations using XLM-RoBERTa-large-sag and XLM-RoBERTa-sag with different input representations described in Section 3.2.1.

Table 5. Accuracies (by the f1-macro metric) of XLM-RoBERTa-base-sag (denoted “LM-Base”) and XLM-RoBERTa-large-sag (“LM-Large”) language models with different methods of text representation.

Text Representation	LM-Base	LM-Large
Text of target entities only	0.75	0.76
Whole text with highlighting target entities	0.78	0.82
Text of target entities and text between them	0.81	0.80
Text of target entities and the whole text	0.91	0.95

The comparison shows that the language model should receive both the target entities separated from the text and the entire text in order to achieve high accuracy and to outperform basic machine learning methods. The f1-macro achieved for ADR–Drugname relations from the RDRS dataset is 95% (see Table 5). This estimation is 41% higher than random class prediction, and 20% higher than basic machine learning models, even if the hyperparameters of the latter are tuned on a test set.

The optimal hyperparameter values found for XLM-RoBERTa-base-sag are:

- maximum input length—512;
- early stopping—active;
- learning rate—0.00005;
- batch size—32;
- maximum epochs—10;
- learning rate decay—active;

The resulting hyperparameter values for XLM-RoBERTa-large-sag are:

- maximum input length— 512;
- early stopping—active;
- learning rate—0.00001;
- batch size—8 (there was not enough memory for bigger batch size with XLM-RoBERTa-large);
- maximum epochs—10;
- learning rate decay—active;

5.2. Estimation of the Relation Extraction Efficiency

As a result of the experiments conducted on the 908 reviews from the RDRS corpus that have multi-context annotation, accuracy has been estimated for the task of determining the relationships between pharmacologically-significant entities using the method developed on base of the XLM-RoBERTa language model. The accuracy of the proposed method in comparison with the baseline classifiers is given in Table 6. Accuracy is measured by the f1-score metric averaged over five cross-validation splits and is presented separately for the positive (relation present) class and the negative (no relation) class. The results for the baseline machine learning methods are obtained with input represented by target entity pairs encoded with tf-idf of n-grams of 3–8 characters.

As follows from this table, the proposed model determines the four relations under consideration with the following accuracy (according to the f1-score metric for the positive class): between adverse drug reactions and drugs (ADR–Drugname) 92.7%, between drugs and diseases (Drugname–Diseasename) 89.9%, between a drug and its source of information (Drugname–SourceInfoDrug) 92.9%, between diseases and symptoms (Diseasename–Indication) 87.1%. This is 43.5%, 40%, 41.5%, 38.2% higher than the accuracy of the dummy classifier and higher than the accuracy of RuBERT by 3.9%, 3.8%, 3.5%, 2.1% respectively. At the same time, for the class without the relation between entities (negative class), the accuracy is more volatile, taking the values of 91.1%, 76.2%, 82.7%, 31%. However, they exceed the Dummy Classifier accuracy by 59.3%, 42.9%, 49.8%, 9% and RuBERT by 14.9%, 10.0%, 20.1%, 3.3% respectively. On average, the developed model outperforms RuBERT by 3.6%, achieving the f1-score of 80.4%.

Table 6. Accuracy of predicting relations (pos) and absense of relations (neg) between entity pairs of different types in multicontext reviews from the RDRS dataset.

Methods	ADR–Drugname		Drugname–Diseasename		Drugname–Source Info Drug		Diseasename–Indication	
	pos	neg	pos	neg	pos	neg	pos	neg
Proposed model	92.7	91.1	89.9	76.2	92.9	82.7	87.1	31
	91.9	83.05	87.8	59				
RuBERT	88.8	76.2	86.1	66.2	89.4	72.6	85.7	27.7
	82.5	76.15	81	56.7				
LinearSVM	72.8	45.0	75.6	44.9	77.9	45.2	83.2	24.4
	58.9	60.25	61.55	53.8				
Multinomial Naive Bayes	66.3	33.8	68.8	26.1	73.4	14.3	80.2	5.4
	50.05	47.45	43.85	42.8				
Stratified Random Labeling	66.5	31.8	66.5	33.3	69.8	32.9	77.8	22.0
	49.15	49.9	51.35	49.9				

5.3. Applying the Proposed Approach to the DDI Dataset

In order to test the applicability of the proposed model to the texts in another language, we estimated our model on the well-known Drug-Drug Interaction (DDI) dataset [27], used on the SemEval-2013 challenge as a dataset for biomedical relation extraction.

The DDI dataset is a manually annotated corpus consisting of 792 texts selected from the DrugBank database and other 233 Medline abstracts. The dataset has been annotated with a total of 18,502 pharmacological substances and 5028 relations. The dataset includes named entities of the following types:

- Drug—used to annotate those human medicines known by a generic name;
- Brand—drugs described by a trade or brand name;
- Group—drug interaction descriptions often include groups of drugs, that were separated to “group” entity type;
- Drug_n— active substances that weren’t approved for human use, such as toxins or pesticides.

The relations annotated in the dataset are four types of drug-drug interactions (DDIs):

- Mechanism— this type is used to annotate DDIs that are described by their pharmacokinetic mechanism;
- Effect—this type is used to annotate DDIs describing an effect or a pharmacodynamic mechanism;
- Advice—This type is used when a recommendation or advice regarding a drug interaction is given;
- Int—This type is used when a DDI appears in the text without providing any additional information.

When applying the proposed model to the DDI dataset, the model has been fine-tuned to DDI, but pre-training on Russian texts has remained the same. For the fine-tuning and testing on DDI, the data split is the same as in the BLURB project [18]—624/90/191 documents for train/validation/test sets respectively.

Experiments with the text representation “target entities and the whole text” (described in Section 3.2.1) yield the micro-averaged f1-score value of 71.2%. We have therefore modified the input text representation for higher accuracy on the DDI dataset. Inspired by the entity screening technique from the literature [18], we have employed both highlighting the target entity mentions with tags and concatenating target entity mentions with the texts of the whole reviews. For example, the text: “Cytochrome P-450 inducers, such as phenytoin, carbamazepine and phenobarbital, induce clonazepam metabolism, causing an approximately

30% decrease in plasma clonazepam levels.” was represented as: “phenytoin [SEP] carbamazepine [TXTSEP] Cytochrome P-450 inducers, such as [T]phenytoin/[T], [T]carbamazepine/[T] and phenobarbital, induce clonazepam metabolism, causing an approximately 30% decrease in plasma clonazepam levels.”

The resulting accuracy by f1-metric, micro-averaged over the four relation classes, is 81.46%, which is comparable to the accuracy other language model-based approaches [18,20,54] achieve for determining relations between entities extracted from this dataset, the state of the art being 84.05% [20].

6. Discussion

Table 6 shows that there is a volatility between different relation types in terms of accuracy. Preliminary analysis of the relations of different types shows that the DiseaseName–Indication relation has the following distinctive features: low number of the negative class samples (pairs of entities of the desired type that have no relation), high fraction of the unique pairs of entities (approx. 65%); high fraction (approx. 35%) of the unique relations that are represented with different classes in different texts (the same entity pair that has a relation in one text may have no relation in the other text). All these factors—unevenness of classes and the ambiguity of the relation existence between mentions of the same entities in different texts—make the classification task more difficult for the machine learning model. As a solution, we consider conducting further research of the data structure and the classification results, as well as extending our dataset by more relations that have lower representation in the corpus.

Overall, the developed approach shows the highest accuracy out of a group of methods considered: the language model RuBERT, trained on the Russian Wikipedia and news, classic machine learning algorithms (LinearSVM and Multinomial Naive Bayes) and the baseline “dummy” method of Stratified Random Labeling.

Though accuracy is the key performance value of the machine learning models, another important metric is their computational complexity. XLM-RoBERTa is the most resource-intensive model among those considered—it has approximately 550 million parameters, while RuBERT has approximately 110 million parameters. A limitation of XLM-RoBERTa-large is that it requires a GPU to work efficiently.

Another limitation of the transformer language models related to the computational complexity is a limit on the input sequence length—the input of the base BERT model cannot have a size larger than 512 tokens, while RoBERTa-large has this limitation set to 1024 tokens. In the case of longer texts, special approaches are needed in order to work efficiently and to use information about the whole text.

It is worth mentioning that this work considers the relation extraction task based on the ground truth named entity annotation, therefore, further research is required to determine the method’s efficiency when the named entities are predicted by another model.

7. Conclusions

The research conducted shows the strong dependency of the accuracy of the entity relation identification on the structure of the input text representation when using pre-trained language models based on the Transformer topology. The highest accuracy is obtained with our proposed model XLM-RoBERTa-large-sag with texts represented in the following form: the text of the first entity of the potential relation, followed by the text of the second entity of the potential relation and the whole input text. The information contained inside the text between the target entities proved to be insufficient to achieve the same accuracy with the same model, presenting the entire review text to the language model is thus necessary.

The average f1-score obtained over 4 relation types is 80.4%. At the same time, the RuBERT model yields a 3.6% lower f1-score, Linear SVM—21.77% lower, baseline stratified random labeling method—30.4% lower.

On the DDI dataset the same model achieves 81.46%, which is comparable to the state-of-the-art 84.05% obtained on that dataset by other language models trained on pre-defined NER annotation.

Another important observation is the volatility of the accuracy across the relation types, which could be explained by the imbalance in the number of relations of different types, in the number of unique representatives of entity mentions, and the distribution of the relations of particular type between training and test subsets. The issue could be corrected with enlarging the corresponding parts of the context-labeled dataset and with balancing the numbers mentioned above.

Overall, the results obtained provide the state-of-the-art accuracy level for the task of pharmacological entity relation identification in Russian-language reviews and could be positioned as a basis for the future tasks of automated analysis of medical reviews.

Author Contributions: Conceptualization, A.S. (Alexander Sboev) and R.R.; methodology, A.S. (Alexander Sboev) and I.M.; software, A.S. (Anton Selivanov), G.R., I.M. and A.G.; validation, G.R. and A.S. (Anton Selivanov); investigation, A.S. (Alexander Sboev), A.S. (Anton Selivanov), G.R. and I.M.; resources, A.S. (Alexander Sboev) and R.R.; data curation, A.S. (Alexander Sboev), S.S. and A.G.; writing—original draft preparation, R.R., A.S. (Anton Selivanov) and A.S. (Alexander Sboev); writing—review and editing, A.S. (Alexander Sboev), R.R. and A.S. (Anton Selivanov); visualization, A.G. and G.R.; supervision, A.S. (Alexander Sboev); project administration, R.R.; funding acquisition, A.S. (Alexander Sboev). All authors have read and agreed to the published version of the manuscript.

Funding: This work has been supported by the Russian Science Foundation grant No. 20-11-20246.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data can be obtained through sending a request from the website of our project: <https://sagteam.ru/en/med-corpus/> (accessed on 30 October 2021). Trained models are presented on the page of our team on the huggingface repository: <https://huggingface.co/sagteam> (accessed on 30 October 2021). The code is available at https://github.com/sag111/Relation_Extraction (accessed on 30 October 2021).

Acknowledgments: This work has been carried out using computing resources of the federal collective usage center Complex for Simulation and Data Processing for Mega-science Facilities at NRC “Kurchatov Institute”, <http://ckp.nrcki.ru/> (accessed on 30 October 2021)

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Segura-Bedmar, I.; Martínez, P. Pharmacovigilance through the development of text mining and natural language processing techniques. *J. Biomed. Inform.* **2015**, *58*, 288–291. [[CrossRef](#)] [[PubMed](#)]
2. Sboev, A.; Sboeva, S.; Gryaznov, A.; Evteeva, A.; Rybka, R.; Silin, M. A neural network algorithm for extracting pharmacological information from russian-language internet reviews on drugs. *J. Phys. Conf. Ser.* **2020**, *1686*, 012037. [[CrossRef](#)]
3. Sboev, A.; Sboeva, S.; Moloshnikov, I.; Gryaznov, A.; Rybka, R.; Naumov, A.; Selivanov, A.; Rylkov, G.; Ilyin, V. An analysis of full-size Russian complexly NER labelled corpus of Internet user reviews on the drugs based on deep learning and language neural nets. *arXiv* **2021**, arXiv:cs.CL/2105.00059.
4. Oliveira, A.; Braga, H. Artificial Intelligence: Learning and Limitations. *Wseas Trans. Adv. Eng. Educ.* **2020**, *17*, 80–86. [[CrossRef](#)]
5. Al-Haija, Q.A.; Jebri, N. A Systemic Study of Pattern Recognition System Using Feedback Neural Networks. *Wseas Trans. Comput.* **2020**, *19*, 115–121. [[CrossRef](#)]
6. Ganesh, P.; Rawal, B.; Peter, A.; Giri, A. POS-Tagging based Neural Machine Translation System for European Languages using Transformers". *Wseas Trans. Inf. Sci. Appl.* **2021**, *18*, 26–33. [[CrossRef](#)]
7. Xu, H.; Van Durme, B.; Murray, K. BERT, mBERT, or BiBERT? A Study on Contextualized Embeddings for Neural Machine Translation. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Punta Cana, Dominican Republic, 7–11 November 2021; pp. 6663–6675.
8. Ge, Z.; Sun, Y.; Smith, M. Authorship attribution using a neural network language model. In Proceedings of the AAI Conference on Artificial Intelligence, Burlingame, CA, USA, 8–12 October 2016; Volume 30.

9. Peters, M.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K. Deep contextualized word representations. In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online, 6–11 June 2021; Volume 1.
10. Luong, M.T.; Pham, H.; Manning, C.D. Effective approaches to attention-based neural machine translation. *arXiv* **2015**, arXiv:1508.04025.
11. Portelli, B.; Passabi, D.; Serra, G.; Santus, E.; Chersoni, E. Improving Adverse Drug Event Extraction with SpanBERT on Different Text Typologies. In Proceedings of the 5th International Workshop on Health Intelligence (W3PHIAI-21), Palo Alto, CA, USA, 8–9 February 2021.
12. Yan, H.; Gui, T.; Dai, J.; Guo, Q.; Zhang, Z.; Qiu, X. A Unified Generative Framework for Various NER Subtasks. *arXiv* **2021**, arXiv:2106.01223.
13. Ge, S.; Wu, F.; Wu, C.; Qi, T.; Huang, Y.; Xie, X. FedNER: Privacy-Preserving Medical Named Entity Recognition with Federated Learning. Available online: <https://arxiv.org/abs/2003.09288> (accessed on 30 October 2021).
14. Wu, S.; He, Y. Enriching pre-trained language model with entity information for relation classification. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3–7 November 2019; pp. 2361–2364.
15. Giorgi, J.; Wang, X.; Sahar, N.; Shin, W.Y.; Bader, G.D.; Wang, B. End-to-end named entity recognition and relation extraction using pre-trained language models. *arXiv* **2019**, arXiv:1912.13415.
16. Eberts, M.; Ulges, A. Span-Based Joint Entity and Relation Extraction with Transformer Pre-Training. In *ECAI 2020*; IOS Press: Amsterdam, Netherlands, 2020; pp. 2006–2013.
17. Lee, J.; Yoon, W.; Kim, S.; Kim, D.; Kim, S.; So, C.H.; Kang, J. BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* **2019**, *36*, 1234–1240. [[CrossRef](#)]
18. Gu, Y.; Tinn, R.; Cheng, H.; Lucas, M.; Usuyama, N.; Liu, X.; Naumann, T.; Gao, J.; Poon, H. Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. *arXiv* **2020**, arXiv:2007.15779.
19. Gordeev, D.; Davletov, A.; Rey, A.; Akzhigitova, G.; Geymbukh, G. Relation extraction dataset for the russian language. In *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog” [Komp’iuternaia Lingvistika i Intel’lectual’nye Tehnologii: Trudy Mezhdunarodnoj Konferentsii “Dialog”]*; Russian State University For The Humanities: Moscow, Russia, 2020.
20. Naseem, U.; Dunn, A.G.; Khushi, M.; Kim, J. Benchmarking for biomedical natural language processing tasks with a domain specific albert. *arXiv* **2021**, arXiv:2107.04374.
21. Ju, M.; Nguyen, N.T.; Miwa, M.; Ananiadou, S. An ensemble of neural models for nested adverse drug events and medication extraction with subwords. *J. Am. Med. Inform. Assoc.* **2020**, *27*, 22–30. [[CrossRef](#)]
22. Joshi, M.; Chen, D.; Liu, Y.; Weld, D.S.; Zettlemoyer, L.; Levy, O. Spanbert: Improving pre-training by representing and predicting spans. *Trans. Assoc. Comput. Linguist.* **2020**, *8*, 64–77. [[CrossRef](#)]
23. Wang, J.; Lu, W. Two Are Better than One: Joint Entity and Relation Extraction with Table-Sequence Encoders. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 16–20 November 2020; pp. 1706–1721.
24. Patrick, J.; Li, M. High accuracy information extraction of medication information from clinical notes: 2009 i2b2 medication extraction challenge. *J. Am. Med. Inform. Assoc.* **2010**, *17*, 524–527. [[CrossRef](#)]
25. Anick, P.; Hong, P.; Xue, N.; Anick, D. I2B2 2010 challenge: Machine learning for information extraction from patient records. In *Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data*; Boston, MA, USA, 12 November 2010.
26. Henry, S.; Buchan, K.; Filannino, M.; Stubbs, A.; Uzuner, O. 2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records. *J. Am. Med. Inform. Assoc.* **2019**, *27*, 3–12. [[CrossRef](#)]
27. Herrero-Zazo, M.; Segura-Bedmar, I.; Martínez, P.; Declerck, T. The DDI corpus: An annotated corpus with pharmacological substances and drug–drug interactions. *J. Biomed. Inform.* **2013**, *46*, 914–920. [[CrossRef](#)]
28. Asada, M.; Miwa, M.; Sasaki, Y. Using Drug Descriptions and Molecular Structures for Drug-Drug Interaction Extraction from Literature. *Bioinformatics* **2020**, *37*, 1739–1746. [[CrossRef](#)]
29. Beltagy, I.; Lo, K.; Cohan, A. SciBERT: Pretrained Language Model for Scientific Text. *arXiv* **2019**, arXiv:1903.10676.
30. Gurulingappa, H.; Rajput, A.M.; Roberts, A.; Fluck, J.; Hofmann-Apitius, M.; Toldo, L. Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. *J. Biomed. Inform.* **2012**, *45*, 885–892. [[CrossRef](#)]
31. Bruches, E.; Pauls, A.; Batura, T.; Isachenko, V. Entity Recognition and Relation Extraction from Scientific and Technical Texts in Russian. In Proceedings of the 2020 Science and Artificial Intelligence Conference (SAI Ence), Novosibirsk, Russia, 14–15 November 2020; pp. 41–45.
32. Ivanin, V.; Artemova, E.; Batura, T.; Ivanov, V.; Sarkisyan, V.; Tutubalina, E.; Smurov, I. Rurebus-2020 shared task: Russian relation extraction for business. In *Computational Linguistics and Intellectual Technologies*; Russian State University for the Humanities: Moscow, Russia, 2020; pp. 416–431.
33. Bondarenko, I.; Berezin, S.; Pauls, A.; Batura, T.; Rubtsova, Y.; Tuchinov, B. Using Few-Shot Learning Techniques for Named Entity Recognition and Relation Extraction. In Proceedings of the 2020 Science and Artificial Intelligence Conference (SAI Ence), Novosibirsk, Russia, 14–15 November 2020; pp. 58–65.

34. Loukachevitch, N.; Artemova, E.; Batura, T.; Braslavski, P.; Denisov, I.; Ivanov, V.; Manandhar, S.; Pugachev, A.; Tutubalina, E. NEREL: A Russian Dataset with Nested Named Entities and Relations. *arXiv* **2021**, arXiv:2108.13112.
35. Conneau, A.; Khandelwal, K.; Goyal, N.; Chaudhary, V.; Wenzek, G.; Guzmán, F.; Grave, E.; Ott, M.; Zettlemoyer, L.; Stoyanov, V. Unsupervised cross-lingual representation learning at scale. *arXiv* **2019**, arXiv:1911.02116.
36. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.U.; Polosukhin, I. Attention is All you Need. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30, pp. 5998–6008.
37. Kudo, T.; Richardson, J. Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. *arXiv* **2018**, arXiv:1808.06226.
38. Sboev, A.; Selivanov, A.; Rybka, R.; Moloshnikov, I.; Rylkov, G. Evaluation of Machine Learning Methods for Relation Extraction Between Drug Adverse Effects and Medications in Russian Texts of Internet User Reviews. Available online: <https://pos.sissa.it/410/006/pdf> (accessed on: 30 October 2021).
39. Smith, L.N. Cyclical learning rates for training neural networks. In Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, 24–31 March 2017; pp. 464–472.
40. Caruana, R.; Lawrence, S.; Giles, L. Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. *Adv. Neural Inf. Process. Syst.* **2000**, *13*, 402–408.
41. Sahoo, K.S.; Tripathy, B.K.; Naik, K.; Ramasubbareddy, S.; Balusamy, B.; Khari, M.; Burgos, D. An evolutionary SVM model for DDOS attack detection in software defined networks. *IEEE Access* **2020**, *8*, 132502–132513. [[CrossRef](#)]
42. Chun, P.J.; Izumi, S.; Yamane, T. Automatic detection method of cracks from concrete surface imagery using two-step light gradient boosting machine. *Comput.-Aided Civil Infrastruct. Eng.* **2021**, *36*, 61–72. [[CrossRef](#)]
43. Xu, F.; Pan, Z.; Xia, R. E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework. *Inf. Process. Manag.* **2020**, *57*, 102221. [[CrossRef](#)]
44. Hosmer, D.W., Jr.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*; John Wiley & Sons: Hoboken, NJ, USA, 2013; Volume 398.
45. Suykens, J.A.; Vandewalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300. [[CrossRef](#)]
46. Rish, I. An empirical study of the naïve Bayes classifier. In Proceedings of the IJCAI 2001 workshop on empirical methods in artificial intelligence, Seattle, WA, USA, 4 August 2001; Volume 3, pp. 41–46.
47. Mason, L.; Baxter, J.; Bartlett, P.; Frean, M. Boosting algorithms as gradient descent in function space. In Proceedings of the NIPS, Denver, CO, USA, 29 November–4 December 1999; Volume 12, pp. 512–518.
48. Kuratov, Y.; Arkhipov, M. Adaptation of deep bidirectional multilingual transformers for Russian language. In *Komp'juternaja Lingvistika i Intellektual'nye Tehnologii*; Russian State University For The Humanities: Moscow, Russia, 2019; pp. 333–339.
49. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
50. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. [[CrossRef](#)]
51. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
52. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. In Proceedings of the 33rd Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Volume 32, pp. 8026–8037.
53. Rajapakse, T.C. Simple Transformers. 2019. Available online: <https://github.com/ThilinaRajapakse/simpletransformers> (accessed on 30 October 2021).
54. raj Kanakarajan, K.; Kundumani, B.; Sankarasubbu, M. BioELECTRA: Pretrained Biomedical text Encoder using Discriminators. In Proceedings of the 20th Workshop on Biomedical Language Processing, Online, 11 June 2021; pp. 143–154.