Time Expressions Identification without Human-labeled Corpus for Clinical Text Mining in Russian

Anastasia A. Funkner[1[0000-0002-4596-6293]] and Sergey V. Kovalchuk[1[0000-0001-8828-4615]]

[1] ITMO University, Saint Petersburg, Russia
{funkner.anastasia, kovalchuk}@itmo.ru

Abstract. To obtain accurate predictive models in medicine, it is necessary to use complete relevant information about the patient. We propose an approach for extracting temporary expressions from unlabeled natural language texts. This approach can be used for the first analysis of the corpus, for data labeling as the first stage, or for obtaining linguistic constructions that can be used for a rule-based approach to retrieve information. Our method includes the sequential use of several machine learning and natural language processing methods: classification of sentences, the transformation of word bag frequencies, clustering of sentences with time expressions, classification of new data into clusters and construction of sentence profiles using feature importances. With this method, we derive the list of the most frequent time expressions and extract events and/or time events for 9801 sentences of anamnesis in Russian. The proposed approach is independent of the corpus language and can be used for other tasks, for example, extracting an experiencer of a disease.

Keywords: Time Expression Recognition, Natural Language Processing, Corpus Labeling, Clinical Text Mining, Machine Learning.

1 Introduction

Since hospitals began to use medical information systems, a large amount of data has accumulated in electronic form. Most often, information is stored in the form of an electronic medical record (EMR), which relates to a particular patient. Data can be stored in a structured form (tables with lab results), in semi-structured (filled fields in some form, for example, an operation protocol) and unstructured (free texts in natural language). Structured data usually requires a little processing and can be easily used to model medical and healthcare processes. In semi-structured data, it is necessary to extract features from natural language texts. However, the researcher has an idea about the topic or even the structure of the content in a field of the completed form and can extract features using regular expressions or linguistic patterns [1]. To work with unstructured data, it is necessary to use many methods and tools to find out specific facts about the patient.

The purpose of this study is to develop methods for extracting time expressions and related events. At the same time, the proposed methods should work with an unlabeled or minimally labeled input text corpus.

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_44

Processing of medical texts includes tasks which are often encountered for other texts: morphological and syntactical analysis, negation detection, temporal processing, etc. But there are specific tasks such as recovering family history from free texts [2]. The problem of time expression recognition relates to information retrieval from unstructured texts. Two approaches can be used to extract information [2]. The first approach includes rule-based methods. In this case, the necessary information is searched using regular expressions or linguistic patterns which are known in advance or are collected iteratively during multiple scanning of the text corpus [3, 4].

The second approach is based on machine learning methods. This approach does not require manually searching for the necessary constructions in the texts or building a knowledge base for each new corpus. The developed models are trained and detect patterns in the data automatically. However, for the training of any model, a labeled corpus is required [5]. However, for the Russian language, there is almost no labeled corpus of medical texts [6]. In [7, 8] only 120 EMRs were labeled for a specific hospital unit. State-of-the-art models for clinical temporal relation extraction are trained on thousands of labeled sentences [9]. Since we can not to mark up a sufficient amount of data to use the above methods, we try to develop methods for detecting time expressions and events using only 1k sentences, which are labelled as containing and not containing a time expression (TE).

The tasks of recognizing time expressions and extracting events can be solved simultaneously or in parallel using different models. The second one is called the clinical temporal relation extraction task. In this case, researchers usually train deep neural networks using a labeled corpus, which consists of sentences where events and temporal expressions are defined in advance [9, 10]. In this paper, we do not have an aim to extract both the event and the time expressions. However, Section 3 includes examples when the event is found with the timestamp using our approach.

In 2019, we started developing an application to process medical texts. Fig. 1 shows seven modules of this application. We have already developed modules for basic text processing (extracting abbreviations, searching for numerical data, lemmatisation, if necessary), correcting typos in texts [11], detection and removing negations [12], and topic segmentation of texts. Currently, two modules are being developed for determining the experiencer of a disease and a module for determining events with a timestamp [13]. This paper is about mining temporal data using machine learning. In the future, we planned to run these modules sequentially and automatically retrieve a set of patient's features. This application can be helpful for hospital staff as a navigation system in medical history because it allows quickly to find out what and when occurred with the patient in the past. Also, the application can be useful for scientists to build more accurate predictive models. As far as we know, there is no other application or system for the Russian language that could process the text with above-mentioned methods. Existing libraries and modules are trained with general Russian corpus and do not work accurately enough for medical texts [14, 15].

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_44

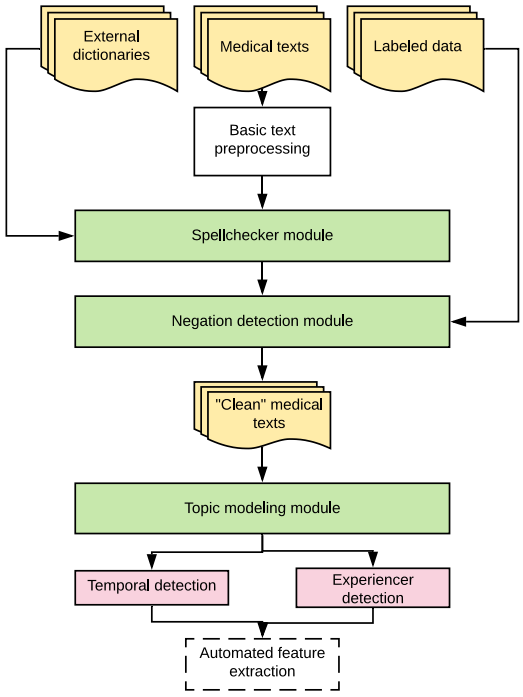

Fig. 1. An application for processing and extracting from medical texts. The green modules are finished and can be used [11, 12]. Pink modules are being developed now, and dotted ones will be created in the future.

2 Method

This section describes which and in what order machine learning methods are used to extract common expressions, including temporal ones. Fig. 2 shows the general scheme of the method with the flow of data and trained models between stages.

First, it is necessary to divide the data corpus into three groups, each of which is represented by a set of sentences. A group X consists of the smallest number of sentences and its size is defined by researchers so that they can label sentences in this group without spending too much time. A group Y contains a larger number of sentences and its size is limited by available computing power since $X + Y$ groups are used for clustering, in which a distance matrix is usually calculated (depends on the method). All remaining sentences form a group Z.

2.1 Manual labeling of the group X

The sentences of group X must be labeled as having and not having a time expression (TE and noTE). The temporal expression refers to any expression that indicates a period or a specific moment in time. According to the TimeML annotation system, temporal

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_44

expressions can be divided into five classes: date (in 2010), time (at 4 pm), duration (over the past 10 years), recurring events (once a month) and others [16]. Using this classification, we label sentences with the first three classes. Other classes of expressions were rare in our corpus. Besides, in [14] there is more specific for the Russian language classification of temporal expressions, of which we also used only a few classes for markup. At this stage of the method, the researchers determine what temporal structures which are important and necessary for their corpus (see Fig. 2(1)).

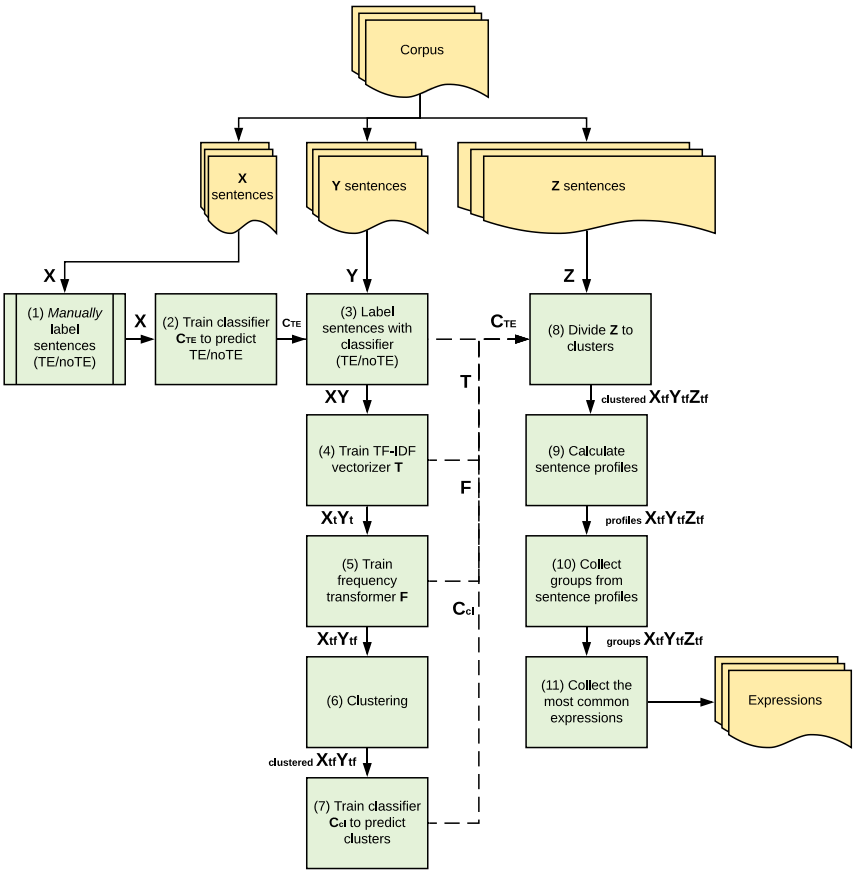

Fig. 2. Scheme of the method for obtaining common expressions.

2.2 Classifier to label the group Y

At this stage, it is necessary to select a classification model, train and test it with the labeled group X. The classification model helps to mark up an additional set of sentences (the group Y) without additional time. We use bag of words to present sentences for the classifier, but other representations are applied here (word embeddings, TF-IDF vectors, etc.). After training and testing, it is necessary to pay attention to sentences that

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_44

turn out to be false negative or false positive after testing. There may be sentences incorrectly marked manually. As a result of this stage, we have $\boldsymbol{C_{TE}}$ classifier and joined labeled groups X and Y (see Fig. 2(2,3)).

2.3 Frequency transformations

At stage (4) and (5), it is necessary to collect a bag of words with n-grams, if this has not been done before. The number n depends on the language of the corpus. The Russian language is characterized by long temporal expressions with compound prepositions, so we recommend using $n = 4$. Next, the transformer $\boldsymbol{T}$ is trained to obtain the term frequency (TF) and inverse document frequency (IDF). The transformer $\boldsymbol{T}$ remembers IDF of the training set and can be used later with new data. TF-IDF transformation helps to reduce the weight of frequent and non-specific words and words, which are rare and found in a small number of documents (see Fig. 2(4)). After calculating TF-IDF vectors and obtaining $X_t Y_t$, we compare the frequencies of sets with and without temporal constructions (S_{TE} and S_{noTE}: $X_t \cup Y_t = S_{TE} \cup S_{noTE}$). We assume that using frequencies of the S_{noTE} will reduce the weights of words that are not related to time, but also frequent in documents of S_{TE} (medical terms in our case). We simply subtract from each TF-ID vector the average vector of all TF-ID vectors from S_{noTE} and zero the negative components for S_{TE}. At the end, the function with this subtraction is saved in the transformer $\boldsymbol{F}$ and transformed vectors are called $X_{tf} Y_{tf}$ (see Fig. 2(5)).

2.4 Clustering and clusters' classifier

Obtained at the previous stage vectors are used for clustering. We assume that similar sentences can join in one cluster. Then, using the obtained clusters, another classifier is trained, and new sentences of the group Z are appended to the clusters. If the initial corpus is small and the group Z was not collected, then we can skip stages (7) and (8) (see Fig. 2). At the stage (8) all above mentioned trained models are used.

2.5 Expression retrieval from sentence profile

At the next stages of the method, we try to find the expressions inside the sentence that characterize it and determine its membership in the cluster. Using the classifier from the stage (7), we obtain feature importances for each cluster and summarize them for all components of the considered sentence. We propose using the classification one-vs-all strategy and tree-based models (decision trees, random forest, tree boosting) as basic classifiers. Such models make it easy to obtain feature importances. If the basic models are not interpretable, then model-agnostic interpretation methods can be used [17]. As a result, the sum vector is called a sentence profile and includes the groups of consequent words with positive importance (expressions). We can also adjust the threshold for more stringent selection of expressions. The resulting expressions are grouped by their syntax structure to define the most common ones (see Fig. 2 (9)-(10)).

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_44

3 Results

In this research, we used a set of anonymized 3434 EMRs of patients with the acute coronary syndrome (ACS) who admitted to Almazov National Medical Research Centre (Almazov Centre) during 2010-2015. Disease anamneses are one of the most unstructured records (free text without any tags; each physician writes as he/she prefers), and therefore we used them to demonstrate our approach. Totally, there are 34241 sentences in the set.

We divide the data set into 1k sentences (group X), 3k sentences (group Y) and the rest (group Z). We spent only 20 minutes to mark the group X. To label the group Y with TE/noTE, we train a decision tree (F1 score equals 0.95). Before TF-IDF transformation, enumerated temporal pointers were replaced by tags: years (2002, 1993 -> "#year"), months (January -> "#month"), dates (01/23/2010 -> "#fulldate"), etc. This helps not to divide temporal pointers to several components of the vector when forming a bag of words. Fig. 3 shows the difference in the average TF-IDF values for a set with and without TEs, as well as the top 10 words with positive and negative differences. Specific components to a set with TE are much greater than zero and are associated with temporal pointers (“year”, “#year”, “#month”). The typical words for a set without TE are much less than zero. These words are associated with negations of diseases (“not”), patient’s well-being (“normal”, “satisfactory”), medications (“take”) or a general description of the patient’s condition (“disease”, “chronic”, “age”). These topics are usually not accompanied by temporal pointers.

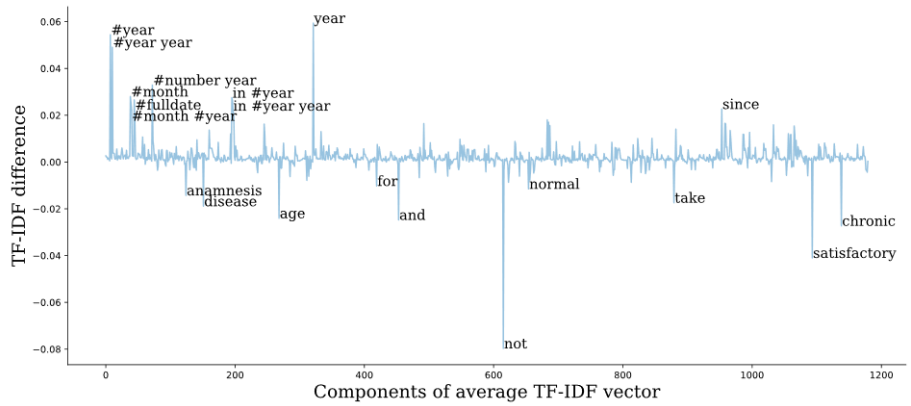

Fig. 3. The difference of average TF-IDF vectors for sets of sentences with TEs and without ones.

Hierarchical clustering is used to divide sentences with TEs. However, we perform clustering three times to divide large clusters into smaller ones. After the first clustering, 13 groups are identified so that 11 small ones include sentences with an almost identical structure. We suppose that the doctors of Almazov Centre used almost the same formulations when they write about an exact disease. For example, one of the clusters consists of almost identical phrases: "in #month #year, the patient had coronary angiography, no pathology." Large clusters gather unique sentences. Fig. 4 shows the

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_44

cluster identified after the first clustering and shows the thresholds with which the big clusters were then divided. As a result, 1145 sentences with TEs are divided into 23 clusters ranging in size from 14 to 237.

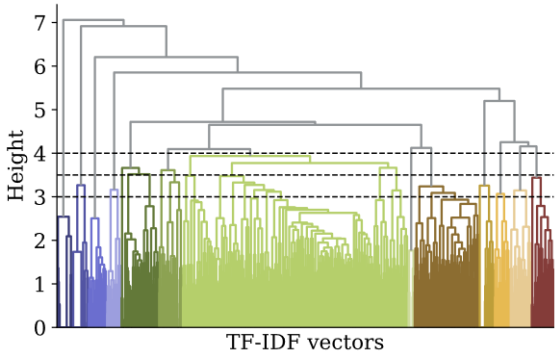

Fig. 4. Multiple hierarchy clustering of TF-IDF vectors (sentences) with TE.

Then, to divide the group Z into clusters, classifiers of tree boosting were trained using the classification one-vs-all strategy (F1 score varies from 0.83 to 0.96), and profiles were calculated for each sentence as described in Section 2.5. Fig. 5–7 show the profiles of sentences and defined groups of TF-IDF values. In Fig. 5 a simple sentence has two groups: temporal one ("#year year") and one with the event ("appear pressing pain behind the sternum"). In Fig. 6 we show the profile of a compound sentence, which should be divided into three simpler sentences. The event and TE of the first part were extracted: "treatment in the neurology department", "emergency treatment", and "in #year". Fig. 7 depicts an unsuccessful example of separation into expressions since the event and time remain in the same group.

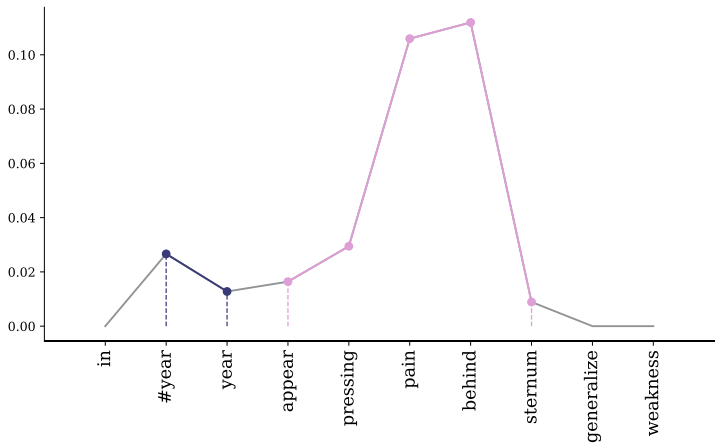

Fig. 5. A sentence profile for “In 2007 pressing pain behind the sternum and generalized weakness appeared” (the word order in x-labels is defined by Russian).

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_44

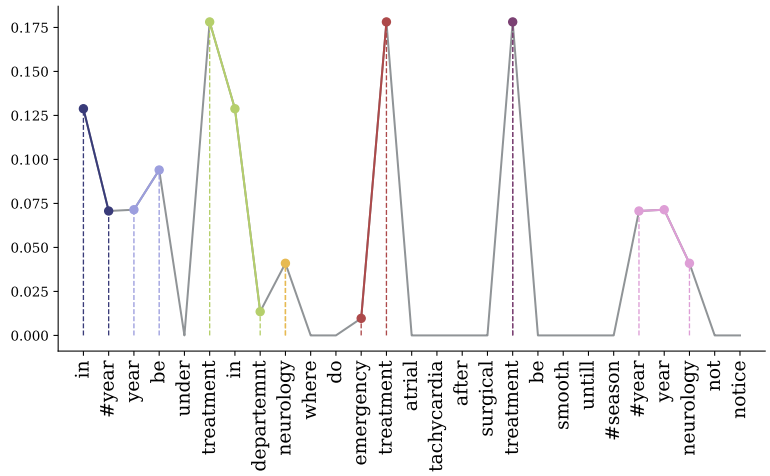

Fig. 6. A sentence profile for "In 2012 [the patient] was under treatment in the department of neurology where emergency treatment [because of] atrial tachycardia was done, after surgical treatment [the recover] was smooth, until the autumn of 2013 [the patient] did not notice neurology [problems]" (the word order in x-labels is defined by Russian).

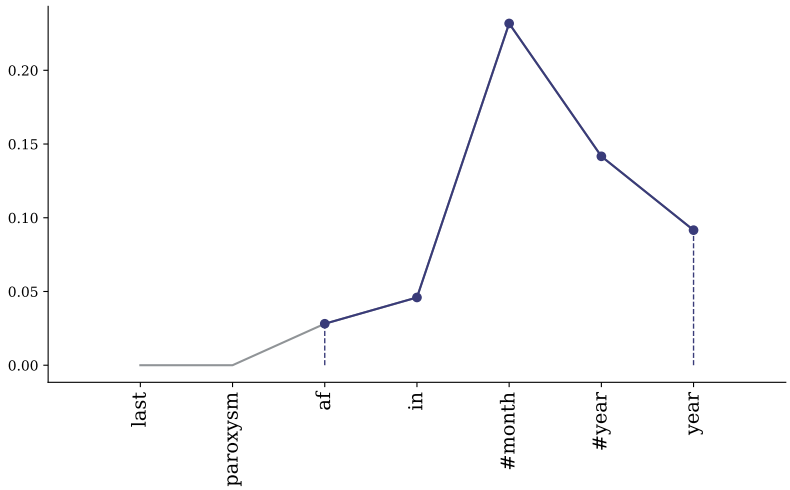

Fig. 7. A sentence profile for "Last paroxysm [of] atrial fibrillation (AF) [was] in March of 2008".

Table 1 contains the most frequent time expressions derived from sentence profiles. Almost all time expressions contain time tags ("#year", "#fulldate", etc.). However, there is a group in which time adverbs are contained. In addition to grouping by tags, we use morphological constructions obtained with the TreeTagger [18, 19]. The combination of morphological constructions and numerical regular expressions can be used to find temporal expressions in new sentences. Also, for each type of time expression,

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_44

we will specify a normalization function to define the exact date or period of the event [20, 21].

Table 1. The most popular time expressions among 9801 sentences with TE.

Time expression	Treetagger construction for Russian	Examples in Russian	Examples in English	Number
in #year years	Sp-l Mc---d Ncmsnnl	в 1983 году	in 1983	1827
from #year years	Sp-l Mc---d Ncmpgn	с 45 лет	from 45 years [old]	1428
#fulldate	Mc---d	4.03.2008	4.03.2008, 16.04.10	1030
about #number years	Sp-g Mc--g Ncmpgn	около 3-ёх лет	about 3 years	1030
in #month #year years	Sp-l Ncmsln Mc---d Ncmsgn	в апреле 2010 года	In April 2010 [years]	930
#year years	Mc---d Ncmsgn	2009 года	2009 [years]	830
adverb	R	неоднократно, после, около, вплоть, вновь, впервые, затем	repeatedly, after, about, until, again, for the first time, then	830
#year	Mc---d	2008	2008	565
#number years	Mc---d Ncmpgn	10 лет	10 years	565
about #number years ago	Sp-l Mc---d Ncmpgn R	около 10 лет назад	about 10 years ago	266
#number-#number	Mc---d	35-40, 2006-2009	35-40, 2006-2009	232
during #number	Sp-l Ncnsln Mc---d	в течение 5	during 5	199
from #fulldate to #fulldate	Sp-l Mc---d Sp-l Mc---d	с 2.03.2010 по 5.04.2010	from 2.03.2010 to 5.04.2010	133
from #fulldata	Sp-l Mc---d	от 07.05.09	from 07.05.09	100
#season #year year	Ncfsin Mc---d Ncmsgn	весной 2010 года	spring 2010 [year]	100
#time	Mc---d	10:30, 23-00	10:30, 23-00	66

Table 2 and Table 3 show the example of patient anamnesis with defined expressions in Russian and English. The anamnesis consists of seven sentences and includes different types of time expressions: years, dates, periods, a start of a period, repeated actions, etc. This example shows how complex a free clinical text can be: many abbreviations and partially written words, grammatical and syntax mistakes, compound sentences. However, time expressions are found in five sentences and events in four sentences. We hope that the recovery of syntax structure and abbreviations can help to improve the quality of extracted expressions.

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_44

Table 2. The example of anamnesis with extracted expressions in Russian.

Anamnesis in Russian
[Перенёс] 2 [Инфаркта миокарда: 2001] и [2006г].
Сегодня повторно вызвал бригаду [СМП,] которой был [доставлен в НИИ Кардиологии] для решения вопроса об имплантации ЭКС.
Считает себя больным в течение [20 лет,] когда впервые стал отмечает начало подъёма [АД до 190\110 мм.рт.ст.,] [адаптирован] к [цифрам АД 140\80 мм.рт.ст].
[03.10.10] вызывал СМП, однако в транспортировке в стационар больному было отказано.
[С 2001г появились давящие] загрудинные [боли,] возникающие [при физической нагрузке] выше [обычной,] купирующиеся после её прекращения [в течение 5-7] минут или приёмом нитроглицерина.
Нынешнее ухудшение началось примерно неделю назад, когда появилось урежение [ЧСС] до 34-36 в мин., слабость, [повторные эпизоды потери] сознания.
Последняя госпитализация [с 17.09.10 по 29.09.10г] [с диагнозом:] ИБС: прогрессирующая стенокардия ПИКС [2006г].

Table 3. The example of anamnesis with extracted expressions in English.

Anamnesis in English
[Have] 2 [Myocardial infarction: 2001] and [2006].
Today, he again called the [ambulance,] which [delivered to the Cardiology Institute] to resolve the issue of implantation of ECS.
He considers himself sick for [20 years], when he first began to note the beginning of the rise [blood pressure to 190\110 mm Hg,] [adapted] to [numbers of blood pressure 140\80 mm Hg].
[03.10.10] called the ambulance, but the patient was refused transportation to the hospital.
[Since 2001 there were pressing] sternal [pains] arising [during physical exertion] above [normal,] stopping after its cessation [within 5-7] minutes or by taking nitroglycerin.
The current deterioration began about a week ago, when there was a decrease in [heart rate] to 34-36 per minute, weakness, [repeated episodes of loss] of consciousness.
The last hospitalization [from 17.09.10 to 09.29.10] [with a diagnosis of] IHD: progressive angina pectoris PIKS [2006].

4 Discussion

The method described in Section 2 helps to get an idea about the set of texts in a natural language quickly. As a result, the list of final constructions turns out to be quite large and contains many unnecessary or incorrectly extracted expressions (see Fig. 7). We suppose these actions can help to improve quality of extracted expressions: complex sentences need to be divided into simple ones, acronyms and abbreviations of medical terms need to be deciphered, and important constructions inside sentences need to be found with shallow parsing [22]. However, to solve these problems, a model of syntax parsing needs to be trained with a specific medical corpus in Russian. So far, we did not find such an open access model. Also, in this paper, we work with anamnesis texts that describe the patient's events in the past. However, other texts in a natural language may contain information about future events and to solve this problem we can mark sentences as past, future and without TE.

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_44

In the future, to improve the quality and create more accurate models, we will have to mark up the words in sentences, for example, the word can belong to an event, TE, or none of this. However, there are many labeling systems: TimeML, BIO, TOMN, etc. [10]. We plan to compare what a system is best for our corpus and tasks.

5 Conclusion

In this paper, we propose a way to work with an unlabeled corpus of texts. This approach can be used for the first analysis of the corpus, for data labeling as the first stage, or for obtaining linguistic constructions that can be used for a rule-based approach to retrieve information [3, 4]. The proposed approach is independent of the corpus type and can be used for other tasks, for example, extracting an experiencer of disease [13].

To solve the problem of TEs and events recognition, we plan to train the model for syntax parsing sentences and label the corpus according to one of the known labeling systems, using the already obtained expressions [10].

When the development of the modules is completed (see Fig. 1), we plan to improve the already developed models for predicting diseases [23–25] and integrate the modules into the medical system of the Almazov Centre. It will help hospital staff to navigate the patient history and allows doctors to reduce the time when they need to get acquainted with patient data during the appointment.

Acknowledgements

This work is financially supported by National Center for Cognitive Research of ITMO University.

References

1. Jackson, P., Moulinier, I.: Natural language processing for online applications : text retrieval, extraction, and categorization. John Benjamins Publishing Company, Amsterdam (2002)
2. Dalianis, H.: Clinical text mining: Secondary use of electronic patient records. Springer (2018)
3. Riloff, E.: Automatically constructing a dictionary for information extraction tasks. Proc. Natl. Conf. Artif. Intell. 811–816 (1993)
4. Riloff, E., Jones, R.: Learning Dictionaries Ellen Riloff for Information Bootstrapping Extraction by Multi-Level. Proceeding AAAI '99/IAAI '99 Proc. Sixt. Natl. Conf. Artif. Intell. Elev. Innov. Appl. Artif. Intell. Conf. Innov. Appl. Artif. Intell. 474–479 (1999)
5. Shickel, B., Tighe, P.J., Bihorac, A., Rashidi, P.: Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis.
6. Kudinov M. S., Romanenko A. A., Piontkovskaja I. I.: Conditional Random Field in Segmentation and Noun Phrase Inclination Tasks for Russian. Компьютерная лингвистика и интеллектуальные технологии. 297–306 (2014)

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_44

7. Shelmanov, A.O., Smirnov, I. V., Vishneva, E.A.: Information extraction from clinical texts in Russian. Komp'juternaja Lingvistika i Intellektual'nye Tehnol. 1, 560–572 (2015)
8. Baranov, A., Namazova-Baranova, L., Smirnov, I., Devyatkin, D., Shelmanov, A., Vishneva, E., Antonova, E., Smirnov, V.: Technologies for complex intelligent clinical data analysis. Vestn. Ross. Akad. meditsinskikh Nauk. 71, 160–171 (2016). https://doi.org/https://doi.org/10.15690/vramn663
9. Lin, C., Miller, T., Dligach, D., Bethard, S., Savova, G.: A BERT-based Universal Model for Both Within- and Cross-sentence Clinical Temporal Relation Extraction. Proc. 2nd Clin. Nat. Lang. Process. Work. 2, 65–71 (2019)
10. Lin, C., Miller, T., Dligach, D., Bethard, S., Savova, G.: Representations of Time Expressions for Temporal Relation Extraction with Convolutional Neural Networks. 322–327 (2017). https://doi.org/10.18653/v1/w17-2341
11. Balabaeva, K., Funkner, A., Kovalchuk, S.: Automated Spelling Correction for Clinical Text Mining in Russian. (2020)
12. Funkner, A., Balabaeva, K., Kovalchuk, S.: Negation Detection for Clinical Text Mining in Russian, (2020)
13. Harkema, H., Dowling, J.N., Thornblade, T., Chapman, W.W.: ConText: An algorithm for determining negation, experiencer, and temporal status from clinical reports. J. Biomed. Inform. 42, 839–851 (2009). https://doi.org/10.1016/j.jbi.2009.05.002
14. Korobov, M.: Morphological analyzer and generator for Russian and Ukrainian languages. In: Communications in Computer and Information Science (2015)
15. Sorokin, A.A., Shavrina, T.O.: Automatic spelling correction for Russian social media texts. In: Proceedings of the International Conference "Dialog"(Moscow. pp. 688–701 (2016)
16. Ingria, R., Sauri, R., Pustejovsky, J., Gaizauskas, R., Setzer, A., Katz, G., Radev, D., Castano, J.: TimeML: Robust Specification of Event and Temporal Expressions in Text. New Dir. Quest. answering. 3, 28–34 (2003)
17. Molnar, C.: Interpretable machine learning. Lulu. com (2019)
18. Schmid, H.: Probabilistic Part-of-Speech Tagging Using Decision Trees. In: Proceedings of the International Conference on New Methods in Language Processing (1994)
19. Russian statistical taggers and parsers, http://corpus.leeds.ac.uk/mocky/
20. Negri, M., Marseglia, L.: Recognition and Normalization of Time Expressions: ITC-irst at TERN 2004. Rapp. interne, ITC-irst, Trento. (2004)
21. Zhao, X., Jin, P., Yue, L.: Automatic temporal expression normalization with reference time dynamic-choosing. In: Coling 2010 - 23rd International Conference on Computational Linguistics, Proceedings of the Conference (2010)
22. Korobkin, D.M., Vasiliev, S.S., Fomenkov, S.A., Lobeyko, V.I.: Extraction of structural elements of inventions from Russian-language patents. In: Multi Conference on Computer Science and Information Systems, MCCSIS 2019 - Proceedings of the International Conferences on Big Data Analytics, Data Mining and Computational Intelligence 2019 and Theory and Practice in Modern Computing 2019 (2019)
23. Funkner, A.A., Yakovlev, A.N., Kovalchuk, S. V.: Data-driven modeling of clinical pathways using electronic health records. Procedia Comput. Sci. 121, 835–842 (2017).

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_44

https://doi.org/10.1016/j.procs.2017.11.108
24. Derevitskii, I., Funkner, A., Metsker, O., Kovalchuk, S.: Graph-Based Predictive Modelling of Chronic Disease Development: Type 2 DM Case Study. Stud. Health Technol. Inform. 261, 150–155 (2019). https://doi.org/10.3233/978-1-61499-975-1-150
25. Balabaeva, K., Kovalchuk, S., Metsker, O.: Dynamic Features Impact on the Quality of Chronic Heart Failure Predictive Modelling. Stud. Health Technol. Inform. 261, 179–184 (2019)

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50423-6_44