ATLANTIS PRESS

Advances in Intelligent Systems Research, volume 174

Proceedings of the 8th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2020)

An Approach to Automated Extraction of Diagnostic Rules From the Text of Clinical Guidelines for Decision Support Systems

Ruslan Vafin*
The Faculty of Informatics and Robotics
Ufa State Aviation Technical University
Ufa, Russia
vafrus74@gmail.com

Rashit Nasyrov
The Faculty of Informatics and Robotics
Ufa State Aviation Technical University
Ufa, Russia
nrash@yandex.ru

Rustem Zulkarneev
Medical faculty
Bashkortostan State Medical University
Ufa, Russia
zrustem@ufanet.ru

***Abstract*—Currently, there is a large amount of accumulated medical knowledge about various diseases, formalized in the form of clinical guidelines. For general practitioners, it is difficult to remember several dozen documents, due to information overloaded. To solve this problem, medical decision support systems (DSS) are used. Those DSS based on digitized clinical guidelines, which are used to help doctors make more accurate diagnoses and prescribe treatment to patients. Periodic updating of clinical practice guidelines raises an additional problem with timely updates to the rules in the DSS. Natural language analysis methods were used to extract information from clinical guidelines. As a result of the analysis of natural language analysis methods, the dependency grammar method was chosen as the most suitable for the Russian language. In conclusion we has been built a prototype of a program for extracting information from the text of clinical guidelines. This program allows extracting of "if-then" rules in the "treatment" Chapter, whose performance has been tested on clinical guidelines for various diseases.**

Keywords—decision support system, medicine, clinical guidelines, natural language processing, syntactic analysis, dependency grammar, phrase structure grammar

I. Introduction

Currently, there is a large amount of accumulated medical knowledge about various diseases. Approximately every few years, working groups collect and verify new scientific research on various diseases, and formulate regulatory medical documents based on evidence-based medicine that contain better practices for the diagnosis, treatment and rehabilitation after various diseases. These documents are called clinical guidelines.

The volume of each document can reach hundreds of pages of text in A4 format. It is difficult for general practitioners, such as therapists, to remember dozens of documents containing prescriptions for each group of patients (men, women, pregnant women, smokers, etc.) and many additional conditions, due to information overloading [1]. In this regard, doctors can work by stereotypes, applying the same treatment rules for all types of patients. This reduces the effectiveness of treatment, since the individual requirements of the patient are not taken into account. To solve this problem in medicine, medical decision support systems (DSS) are used, which help doctors make more accurate diagnose and prescribe treatment for patients [2].

For the work of DSS knowledge is needed on the basis of which decisions will be made. Data sources can be explicit and implicit knowledge [3].

Implicit knowledge can be presented in the form of precedents accumulated in medical information systems (MIS), or collected in clinical data banks. A widely known example is IBM Watson for Oncology - an AI created by IBM, which has been trained on hundreds of thousands of various medical documents: medical histories, medical journals, and textbooks [4]. As a result, Watson has the ability to diagnose certain forms of cancer with a high degree probability and prescribe the most effective treatment according to the information in the medical record and the patient's current symptoms. A definite drawback of this system is the orientation toward the practice of treatment in only one medical institution – Memorial Sloan Kettering Cancer Center.

Explicit knowledge should be presented in the form of clinical pathways, clinical guidelines, special medical literature, scientific publications, etc [5]. Quite often, explicit knowledge can be formalized as sets of if-then rules. As a rule, knowledge is presented in this form in the clinical guidelines of the Ministry of Health of Russian Federation.

Clinical guidelines of the Ministry of Health on bronchial asthma and chronic obstructive pulmonary disease are based on the guidelines such as The Global Initiative for Asthma (GINA) [6] and Global Initiative for Chronic Obstructive Lung Disease (GOLD) [7] adapted to the environmental and organizational characteristics of Russian Federation. The GINA working group reviews the recommendations every six months; the GOLD working group reviews every year. The recommendations of the Ministry of Health on other diseases are also periodically updated (once every several years). In this regard, there is the problem of timely updating information in DSS.

The extraction of such knowledge for the subsequent formal presentation and use with the DSS is the main task. Since the clinical recommendations are written in a subset of the natural language, in this case in Russian, automatic extraction is a rather difficult task, since modern methods of processing the natural language do not allow achieving 100% accuracy for unstructured and poorly structured texts [8, 9].

In connection with these, the problem arises of the automated extraction of knowledge from the texts of clinical guidelines with the involvement of experts in each domain. The solution to this problem will reduce the time spent on formalizing constantly updated knowledge.

The aim of this paper is to develop system architecture for automated extraction of diagnostic rules from the text of clinical guidelines for DSS.

II. Medical DSS

DSS in medicine are used to solve a wide range of problems: support in the process of diagnostics, search for suitable cases (precedents), planning and monitoring of therapy, recognition and interpretation of images. An important function of the DSS is the dissemination of “best practices”, including international. Most often, DSS are used precisely to help with the diagnosis, treatment and, if necessary, the adjustment of the prescribed treatment [10].

In this paper, we consider DSS which are intended for use in the framework of a medical technological process. The medical technological process is a system of interconnected necessary and sufficient evidence-based treatment and diagnostic measures, the implementation of which allows the most rational way to conduct treatment and ensure that the maximum compliance of the scientifically predicted results with the real ones while minimizing costs [11]. In this case, the DSS helps doctors to follow the necessary measures, which are determined by the current standards in the field of health, as the clinical guidelines are considered in this paper.

In Russia, medical information systems are widespread [3]. MIS are analogous to ERP systems that are designed for medical purposes. As a rule, they do not provide the possibility of DSS, and DSS must be integrated into medical institutions separately.

A well-known analysis [12] of Russian DSS revealed that most of the destination systems are for special cases of diagnosis and treatment. Systems that would provide support for the medical technological process are few in number, and for the most part they either provide digitized versions of clinical guidelines or partially support decision making based on clinical guidelines.

At the moment, the process of formalizing clinical guidelines in Russian for use in DSS is almost entirely in manual mode. The open sources show that about the use of automated extraction of information from clinical guidelines for subsequent use in DSS is not reliably [10, 13, 14]. Foreign sources have precedents for using natural language processing methods to analyze the text of clinical guidelines [15, 16].

The idea of semiautomatic semantic markup of clinical guidelines is considered in [17]. To annotate the text, an approach is used that combines the methods of machine learning and recognition of entities from the text. During testing, it was revealed that the program can correctly mark out only 68% of the elements that were manually marked by experts, as well as that there are a large number of false positive entries. This program cannot replace the manual annotation of the text, but can reduce the time spent by experts on the manual annotation of the text. In this paper, the annotation of the text of clinical guidelines in German is considered.

The paper [18] describes the automated medical decision-making on colorectal cancer based on the analysis of electronic medical records. This paper uses natural language processing methods to extract information from the patient’s medical records and then make decisions based on clinical guidelines. Manual verification of the test data set confirmed that the program accurately enough retrieves data for all cases. Recommendations based on the extracted data corresponded to clinical guidelines in 99% of cases. This result is explained by the fact that the extract text is semi-structured.

A similar task of extracting information from medical notes for early diagnosis of peripheral artery disease is considered in [19]. In this paper, the key indicators of the presence and absence of peripheral artery disease were manually marked out in the extracts of 20 patients with a confirmed diagnosis and 20 healthy patients. Based on this sample, a model was trained that is designed to search for signs of the disease, based on the search for keywords in medical statements. The developed program made it possible to diagnose peripheral arterial disease earlier in time in 41% of cases compared to traditional diagnostic methods.

III. Clinical guidelines

Clinical guidelines - a document based on proven clinical experience. There are describes the actions of a doctor in the diagnosis, treatment, rehabilitation and prevention of diseases, helping him to make the right clinical decisions [20]. The documents define the types, scope and quality indicators of medical care for peoples in a specific disease, syndrome or clinical situation.

The structure of the clinical guidelines is specified in GOST R 56034-2014. A clinical guideline is a semi-structured text in which information is divided into sections and subsections that are specified in GOST. The text in the sections is unstructured text, since the standard prescribes only the structure of clinical guidelines, but does not contain formally defined structure for the entire sentences of the document.

Part of the information in the document has a formally defined structure. For example, information on the level of reliability of evidence and the level of credibility of recommendations is presented as follows: “Level of credibility of recommendations D (level of reliability of evidence IV)”. The level of reliability of evidence is the degree of confidence that the effect found from the use of medical intervention is true. The level of credibility of recommendations is the degree of confidence in the reliability of the effect of the intervention and that following the recommendations will do more good than harm in a particular situation.

Here is a quote from the text of the clinical guideline for influenza in the chapter “3.1 Conservative treatment” (translated from Russian):

“The appointment of non-steroidal anti-inflammatory drugs at a body temperature above 38 ° C is recommended in the treatment of influenza in outpatient and inpatient settings [52, 63, 187].

The level of credibility of recommendations C (level of evidence confidence - 4)

Comment: Ibuprofen is administered orally 200-400 mg 3-4 times a day for 3-10 days (maximum daily dose is 1200 mg). Paracetamol is prescribed orally for 1-2 tablets. (500-1000 mg) up to 4 times a day (maximum daily - 4000 mg.)"

From this text, the following rule for DSS can be distinguished: (If (body temperature) > 38.0C, then (prescribe non-steroidal anti-inflammatory drugs)).

The commentary contains information on drugs and prescribed dosages. This information can also be formalized as a record. This record consists of name of medication, minimum and maximum dosage, frequency, duration of use, method of use, maximum daily dosage, etc.

Consider another example in the Diagnosis chapter of a recommendation for chronic obstructive pulmonary disease (COPD) (translated from Russian):

“The main symptoms of COPD are shortness of breath during exercise, decreased exercise tolerance, and chronic cough [166].

The severity of dyspnea is recommended to be assessed using the modified mMRC scale (Appendix G1) [177].

The level of credibility of recommendations A (level of evidence is 1)"

In this case, the if-then rule cannot be explicitly identified. From this sentence, you can form the rules of the form (if (symptom), then the (disease)), and (check the symptom (symptom) by the method (method)).

For this passage of the text of the rule will take the form:

• if (shortness of breath during physical exertion), then COPD;

• if (a decrease in the tolerance of physical exertion), then COPD;

• if (chronic cough), then COPD;

• check the symptom of (dyspnea during physical exertion) by the method of (modified mMRC scale (Appendix G1))".

In the clinical guidelines for osteoporosis, you can identify the rules of the form (if (condition), then diagnose the method (method)). Here is a quote from the chapter “Diagnostics” (translated from Russian):

“RECOMMENDATION 1: Screening to identify groups with a high probability of fracture is recommended using the FRAX algorithm among all postmenopausal women and men over 50 years of age.

The level of credibility of recommendations A (level of evidence is 1).”

The following rule can be formed from this text: (if (man and age > 50) or (woman and postmenopause) then (screening using the FRAX algorithm)). Moreover, the purpose of this recommendation (to identify groups with a high probability of fractures) and commentary to this recommendation (contains a description of the FRAX algorithm) can be attributed to the rule as metadata.

IV. Proposed Solution

In almost all the works presented, a model for representing the syntactic structure of a sentence based on the phrase structure grammar is used. As will be further noted, this model is not fully suitable for presenting the sentence structure in Russian.

A typical natural language processing scheme is shown in Figure 1. Natural language processing is carried out in several stages.

The first step is text recognition of clinical guidelines. Since clinical guidelines are provided by the Ministry of Health of the Russian Federation in the form of electronic documents in PDF format [21], the technological task is to read the text from files containing complex formatting.

The main task of lexical analysis is to split the input text, consisting of a sequence of single characters, into a sequence of words, or tokens, that is, to select these words from a continuous sequence of characters [22, 23]. All characters of the input sequence from this point of view are divided into characters belonging to any tokens, and characters that separate tokens (delimiters). At this stage, problems may arise in the analysis of abbreviations and formulas.

Morphological analysis is a comparison of individual words and word forms in a dictionary and clarification of the grammatical characteristics of words [24]. One of the possibilities of morphological analysis is to find the normal form of the word. The main problem of this stage is the polysemy of words in Russian. In the analysis of clinical guidelines, this problem is less significant, since a limited subset of the Russian language is used in clinical guidelines, which reduces ambiguity.

Parsing is the construction of the grammatical structure of a sentence and the identification of syntactic relationships between words in a sentence [24]. The parser should build the syntactic structure of the sentence. In automatic parsing, as a rule, syntax trees are used that reflect the structure and/or relationships in the sentence [22].

Since there are a huge number of different medical abbreviations and abbreviations in clinical guidelines, parsing is generally difficult. In order to solve this problem, it is proposed to analyze the text in 2 stages: at the first stage, the lexical analyzer generates an array of tokens, at the second stage, the search for tokens containing the abbreviations and their definition. Further, all abbreviations are replaced by the found definitions for a more accurate operation of the parser.

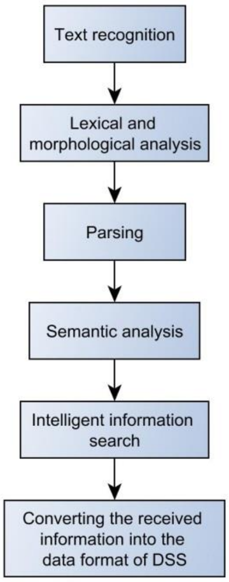

Fig. 1. Natural language processing workflow

Semantic analysis of the text is carried out on the basis of the constructed syntactic tree. The results of semantic analysis are used to find the necessary information.

Since it is impossible to completely avoid errors in the analysis of natural text, it is necessary to minimize false-negative errors. Since this will allow experts to discard incorrect assumptions of the program, reducing the time for independent search and formalization of information from clinical guidelines.

V. Text Parsing Methods

From the point of view of a formal approach to the analysis of linguistic constructions, parsing is the process of comparing the linear sequence of tokens (words, tokens) of a natural or formal language with its formal grammar. The result of the parsing is the syntactic structure of the sentence, presented in the form of a system of components (phrase structure tree) or a tree of dependencies (tree of subordination) [26].

The phase structure grammar is based on the postulate according to which every complex grammatical unit is composed of two simpler and not intersecting units, called its immediate components [27, 28]. The component that includes more than one word is called a group. The word that corresponds to the root node in the dependency tree that describes the group is called the root of the group. Representation of the syntactic structure of a sentence in the form of a hierarchy of immediate components in various versions is used in formal models of the languages, in particular in the generative linguistics of N. Chomsky. An example of a tree of direct components is shown in Figure 2.

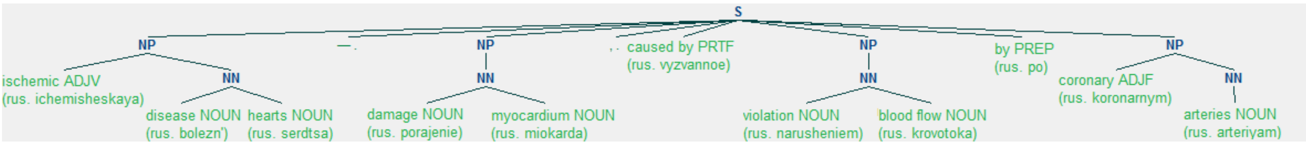

Fig. 2. Example of phrase structure grammar tree (transliterated Russian words are indicated in parentheses)

Tags are meaningfully interpreted as varieties of syntactic groups of words and phrases. The following types of syntactic groups (phrasal categories) are usually used:

- noun phrase (NP) - is headed by a noun;
- adjectival group (AP) - headed by an adjective;
- adverbial group (AdvP) - is headed by an adverb;
- prepositional group (PP) - is headed by a preposition;
- verb group (VP) - is headed by a verb;
- sentence (S).

One of the problems of the phase structure grammar is the removal of ambiguities (syntactic homonymy) [25]. In Russian, there is a relatively free word order in a sentence, so the grammar of the components cannot fully shows the sentence structure, since it implies that the blocks of the sentence part have a continuous word order. As a result, several phase structure trees can correspond to one sentence. Therefore, the grammar of the components for the Russian language does not fit very well, it is much more convenient to use the dependency grammar.

In dependency grammar, the order of words in a sentence is not important. The main thing to know is what word depends each word in the sentence and what type of connection this dependence is indicated.

Let x be an arbitrary nonempty chain in the dictionary V, X is the set of all elements of the chain x. An arbitrary binary relation → on X such that the graph 〈X, →〉 is a tree is called a dependency relation (or a syntactical subordination relation) for x. In other words, if words as elements of a chain, then the chain x is the sequence of words in the sentence, and X is the set of all words in the sentence. An example of a dependency tree is shown in Figure 3. The blue color indicates the type of communication and the green color indicates the part of speech.

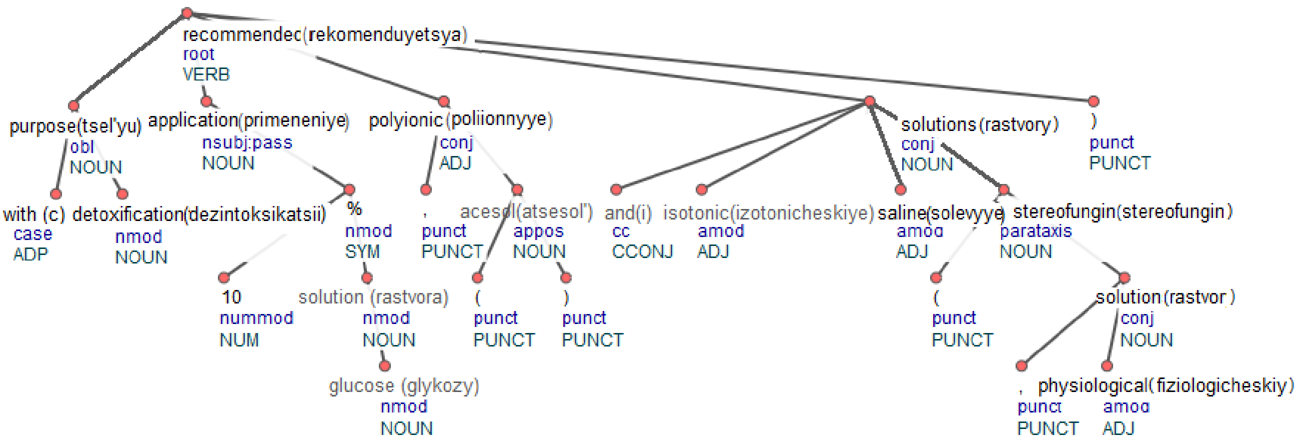

Fig. 3. Example of phrase structure grammar tree (transliterated Russian words are indicated in parentheses)

For the subsequent semantic analysis of the sentence in Russian, the use of dependency trees as a model for representing the syntactic structure of the sentence is more rational, since it allows you to remove the ambiguity of the trees, and also allows you to group semantic blocks in the sentence that are close in content, but separate in word order.

VI. Software Implementation of a Prototype Application

The objective of this paper is to develop a prototype application that would show the fundamental possibility of extracting the rules from the text of clinical guidelines.

In this paper, we used a lexical analyzer included in the Natural Language Toolkit (NLTK) software package [22] for lexical analysis. The NLTK software package allows you to use lexical, morphological and syntactic analysis of the English language. For this package to work with the Russian language, you must install the additional module NLTK Russian, which allows you to use lexical analysis of the Russian language.

Morphological analysis by means of NLKT is difficult, since this software package is designed for English, and there is no Russian language corpus. Instead, a pymorphy2 morphological analyzer was used [29]. The analyzer uses the OpenCorpora dictionary; hypotheses are built for unfamiliar words.

Currently, there are a large number of software tools for parsing a natural language. Table 1 shows the most popular libraries that support the parsing of sentences of the Russian language.

TABLE I. Text Analysis Software Comparison Chart

Library	Support a dependency tree	Compiled models for the Russian language	Highlighting key components and patterns	License	Using a third-party morphological analyzer
SyntaxNet [30]	+	-	-	Apache License 2.0	+
Stanford CoreNLP [31]	+	-	-	GPL v2	+
UDPipe [32, 33]	+	+	+	Mozilla Public License 2.0	+
Tomita-parser [34]	+	-	+	Mozilla Public License 2.0	+
NLTK[22]	+	-	+	Apache License 2.0	+

The main criterion for choosing a parser for building a prototype application was the availability of a trained model for the analysis of Russian-language text. From this table it can be seen that UDPipe is the most suitable. Tomita-parser and NLTK work on the basis of the rule base, which must be filled in manually or using machine learning methods. The UDPipe library has a ready-made trained model based on the Russian language shell SynTagRus. UDPipe also supports third-party morphological analyzers and has an API interface for use in modules implemented in python

The resulting text is processed to obtain a tree structure in python. To search for information, search templates are used to find the necessary information. These patterns are trees, nodes and arrows are regular expressions. This template is applied sequentially for each non-final node of the syntax tree. An example of a template for searching for offers, consisting of two parts, separated by the word “recommend” is shown in Figure 4.

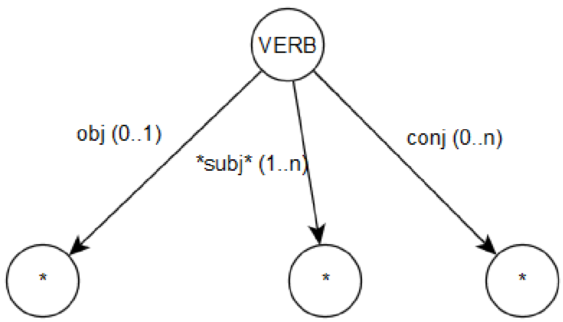

Fig. 4. Search template

As an example of a dependency tree, a tree obtained using this set of programs was introduced in Chapter 5. As a result of applying the template shown in Figure 4, the following information was extracted:

• Action: "recommend"

• Object: "For the purpose of detoxification"

• Subject: "use of a 10% glucose solution"

• Conjunct 1: "polyionic (Acesol)"

• Conjunct 2: "and isotonic saline solutions (stereofungin, saline)"

As can be seen from the results, the program was able to divide the sentence into meaningful parts, but some of the information was interpreted incorrectly, for example, the subjects in this sentence should have had one word "application", the child subtrees of which would be 3 conjuncts, which would represent 3 different medications.

VII. Conclusion

As a result of the relevant works analysis in the field of DSS, an urgent problem was identified – extracting knowledge from unstructured texts to formulate rules for diagnosis and treatment.

An analysis of the works on extracting knowledge from medical texts revealed that most of the works are related to the analysis of English-language clinical guidelines, and only a small part of the works is related to Russian clinical guidelines. After analyzing the work data, it was found that experts are used to analyze the text, or there is no information on the analysis methods.

The main difficulty in extracting knowledge from clinical guidelines is the rather high variability of texts in Russian, requiring the use of a sufficiently large number of rules for text analysis. This problem especially affects parsing. Models of syntactic structures that can be used for various languages were considered. The phase structure trees are suitable for languages with a strict word order, for example, for the English language. For the Russian language, the use of dependency trees will be more effective, in view of the more free word order.

A prototype program has been developed for extracting information from the text of clinical guidelines, which allows implementing the extraction of "if-then" rules in the "treatment" chapter, the operability of which is tested on clinical guidelines for various diseases.

Acknowledgment

This work was partly supported by RFBR grants № 17-48-020074, № 19-07-00780.

References

[1] A. F.Sokol, R. V. Shurupova, "Information fatigue Syndrome: ways and methods of its prevention" in Psychology of development and education (Russian journal of applied research), 2017, no. 1, pp. 116-118. (in Russian)

[2] O. U. Atkov, "Medical Decision Support System", in Vrach i informatsionnye tekhnologii, no. 6, 2013, pp. 67–75. (in Russian)

[3] V. L. Malykh, "Decision Support System in medicine", in Programmnye sistemy teoriia i prilozheniia, vol. 10, no. 2, 2019, pp.155–184. (in Russian)

[4] "IBM Watson for Oncology", available at: https://www.ibm.com/products/clinical-decision-support-oncology (accessed 20 Febrary 2020).

[5] O. Nomokonova, "Implicit knowledge in medical diagnostics" in "Obrazovatelnye resursy i tekhnologii", no. 1 (18), 2017, pp. 49-55. (in Russian)

[6] "Global Initiative for Asthma", available at: https://ginasthma.org/ (accessed 20 Febrary 2020).

[7] "Global Initiative for Chronic Obstructive Lung Disease", available at: https://goldcopd.org/ (accessed 20 Febrary 2020).

[8] A. N. Vinogradov, "Perspective directions of research in the field of clinical modeling, management and decision making" in Vrach i informatsionnye tekhnologii, no.5, 2014. (in Russian)

[9] R. R. Vafin, "Automatic extraction of information from clinical recommendations using text mining", proceedings of the "Mavlyutovsky readings: Materials of the XIII all Russian youth scientific conference", Ufa: USATU, 2019, vol. 4, pp. 74-77. (in Russian)

[10] I. V. Efimenko, V. F. Khoroshevsky, "Intelligent decision support systems in medicine: a retrospective review of the state of research and development and prospects.", 2017. (in Russian)

[11] G. I.Nazarenko, G. S. Osipov, "Basics of the theory of medical technological processes. Part 1", Moscow: Fizmatlit, 2005. p. 144. (in Russian)

[12] A.V. Gusev, "Review of Russian medical decision support systems / / Integrated medical information systems", available at: https://www.kmis.ru/blog/obzor-rossiiskikh-sistem-podderzhki-priniatiia-vrachebnykh-reshenii/ (accessed 29 Febrary 2020). (in Russian)

[13] E. G. Knyazev, A. A. Savchenko, K. V. Ryumkin, "Formation of a system for supporting medical decision- making based on digitization of clinical recommendations using formal logic tools", in Menedzhment kachestva v meditsine, 2019, no. 4, pp. 52-57. (in Russian)

[14] A. Névéol, H. Dalianis, S. Velupillai, G. Savova, and P. Zweigenbaum, "Clinical natural language processing in languages other than english: opportunities and challenges" in Journal of biomedical semantics., vol. 9, no. 1, 2018. p. 12.

[15] E. Ford, J. A. Carroll, H.E. Smith, D. Scott, J.A. Cassell, "Extracting information from the text of electronic medical records to improve case detection: a systematic review" in J Am Med Inform Assoc, 2016, vol. 23, no. 5, pp. 1007-1015.

[16] C. Friedman, L. Shagina, Y. Lussier, G. Hripcsak, "Automated encoding of clinical documents based on natural language processing" in J Am Med Inform Assoc, 2004, no. 5, vol. 11, pp. 392–402.

[17] M.Becker, B.Böckmann, "Semi-Automatic Mark-Up and UMLS Annotation of Clinical Guidelines" in Studies in health technology and informatics, vol. 245, 2017, pp. 294-297

[18] A. Karwa "Development of an Automated Algorithm to Generate Guideline-based Recommendations for Follow-up Colonoscopy" in Clinical Gastroenterology and Hepatology, 2019.

[19] N.Afzal "Surveillance of Peripheral Arterial Disease Cases Using Natural Language Processing of Clinical Notes" in AMIA Summits on Translational Science Proceedings, 2017, p. 28.

[20] "GOST R 56034-2014 Clinical guidelines (treatment protocols). General provisions", Moscow: STANDARTINFORM, 2015. (in Russian)

[21] "Categories of clinical guidelines", Ministry of Health of Russian Federation, available at: http://cr.rosminzdrav.ru/ (accessed 20 Febrary 2020). (in Russian)

[22] S.Bird, E.Klein, E.Loper, "Natural language processing with Python: analyzing text with the natural language toolkit", O'Reilly Media, 2009 .

[23] A. V. Anisimov, A. A. Marchenko, "Natural language word processing system" in Iskusstvennyi intellekt, Ukraine. 2002, no. 4, pp. 157–163. (in Russian)

[24] E. I. Bolshakova, "Automatic processing of texts in natural language and data analysis: Handbook", Moscow: Higher School of Economics National Research University, 2017, p.269. (in Russian)

[25] O. V.Mitrenina, I. S.Nikolaev, T. M. Lando, "Applied and computational linguistics ", Moscow: URSS, 2016, p. 320. (in Russian)

[26] U. D.Apresian, ”Ideas and methods of modern structural linguistics”, Moscow: FGUP Nauka, 1966, p. 302. (in Russian)

[27] T. V. Baturina, M. V.Charentseva, “Fundamentals of text information processing: a textbook”, 2016, p.45. (in Russian)

[28] A. V. Gladkii, “Natural language syntactic structures”, Moscow: URSS, 2007, p.152. (in Russian)

[29] M. Korobov, “Morphological analyzer and generator for Russian and Ukrainian languages”, proceedings of the “International Conference on Analysis of Images, Social Networks and Texts” in Springer, 2015. pp. 320–332.

[30] “SyntaxNet: A TensorFlow toolkit for deep learning powered natural language understanding”, available at: https://github.com/tensorflow/models/tree/master/research/syntaxnet (accessed 20 Febrary 2020).

[31] D.Chen, C. D. Manning, “A fast and accurate dependency parser using neural networks”, proceedings of the 2014 conference on empirical methods in natural language processing, 2014. pp. 740-750.

[32] “Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe”, available at: http://ufal.mff.cuni.cz/udpipe (accessed 20 Febrary 2020).

[33] M. Straka, J. Hajic, J. Straková, “UDPipe: trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, pos tagging and parsing” in Proceedings of the Tenth International Conference on Language Resources and Evaluation, 2016, pp. 4290–4297.

[34] “Tomita-parser”, available at: https://yandex.ru/dev/tomita/ (accessed 20 Febrary 2020). (in Russian)