Formalization of Medical Records Using an Ontology: Patient Complaints*

Klyshinsky E.[1[0000-0002-4020-488X]], Gribova V.[2[0000-0001-9393-351X]], Shakhgeldyan C.[3,4[0000-0002-4539-685X]], Shalfeeva E.[2[0000-0001-5536-2875]], Okun D.[2[0000-0002-6300-846X]], Geltser B.[4[0000-0002-9250-557X]], Gorbach T.[4[0000-0003-4380-6517]], and Karpik O.[5[0000-0002-0477-1502]]

[1] National Research University "Higher School of Economics" klyshinsky@hse.ru
[2] Institute of Automation and Control Processes, FEB, RAS {gribova, shalf}@iacp.dvo.ru, okdm@dvo.ru
[3] Vladivostok State University of Economics and Service carinashakh@gmail.com
[4] Far East Federal University boris.geltser@vvsu.ru, tagorbachdv@gmail.com
[5] Keldysh Institute of Applied Mathematics, RAS

Abstract. Medical records contain a textual description of such important information as patients' complaints, diseases progress and therapy. An extraction of this information could allow starting with processing information stored in medical databases. In this article we introduce a short description of a medical ontology storing information on patients' complaints. We also describe an algorithm that uses this ontology for extraction of claims from texts of medical records. The algorithm combines both syntactic properties, and peculiarities, of a text and connections between diseases' properties and their values. The algorithm corrects syntactical mistakes according to the hierarchical information from the ontology. The resulting algorithm was proved on 3000 clinical records of Department of Neurosurgery of FEFU.

Keywords: Medical record · Term extraction · Information retrieval.

1 Introduction

Modern medical information systems allow storing structured information on patient diseases, disease flow and outcome, therapy etc. The stored information could be divided into non-, weakly and strongly structured. Some of medical information systems divide a medical record into a set of formalized fields of numerical nature or defined by dictionaries; there couldn't be fields with text descriptions in free textual form. Such information could be processed using mathematical methods: statistics, machine and deep learning, time series analysis, etc. Weakly structured information - images, electroencephalograms, cardiograms, ultrasonography, MRI - could be also stored in such medical information

* The reported study was funded by RFBR according to the research project # 18-29-03131.

systems and mathematically processed. E.g. using Fourier and wavelet analysis, neural networks - for purposes of feature extraction and further analysis. That is why weakly and strongly structured medical information could be used as a source for data mining in such areas as causal relations among functional, metabolic and genetic status of patients, their way of living, therapy, morbidity and survival probability, etc.

However, the most part of medical records are stored in unstructured text form - fortunately, in data media. That is the case of patients' complaints, medical background, patient diary, etc. Russian medical information systems formalize the basic clinical records structure with the content of mentioned fields being tables or text in a free form. An extreme case of electronic medical records is a folder with documents stored by a doctor. Such information cannot be processed directly without preprocessing. The mentioned files or fields of information system should be processed by a fact extractor in order to construct a formalized representation.

This paper describes a new method for formalization of patient complaints written in a textual form. The method constructs a preliminary representation of a complaint text on the base of syntactic analysis; the final representation is refined using an ontology that describes relations among notions of a selected domain - neurological diseases.

2 Overview

In 1986 US National Library of Medicine started a project of Unified Medical Language System (UMLS) - a comprehensive publicly accessible collection of electronic dictionaries, thesauri and ontology. (For more historical information on UMLS consult [1].) It consists of Metathesaurus (hierarchy of terms collected from many vocabularies), Semantic Network (relationships among these terms and their categories), and SPECIALIST Lexicon and Lexical Tools (a large syntactic lexicon of biomedical and general English combined with natural language processing tools). The 2018AB Metathesaurus release (November 2018) contains approximately 3.82 million concepts and 14 million unique concept names from 207 source vocabularies [2]. Metathesaurus vocabulary (Medical Subject Headings - MeSH) was translated into 15 languages including Russian [3].

Currently, this collection is used in several big projects, e.g., MetaMap - a program for information extraction from medical texts [4]. The MetaMap method of a medical text processing consists of two stages: 1) natural language processing of the medical text and fact extraction, and 2) notions refinement. The first stage starts with tokenization and finishes with a syntactic analysis. It includes an acronym/abbreviation identification, multi-word terms extraction, and their identification in dictionaries. One word or phrase could have several entries in thesaurus or dictionaries. That's why MetaMap provides word and multi-word term sense disambiguation. It maps terms combinations and then filters improbable combinations out. However, some terms keep an unresolved ambiguity. The result of the medical text processing is a tagged text with a link to Metathesaurus.

MetaMap system allows extracting and indexing such terms as pharmacy names, their quantity, disease names, body parts, etc. Using these results, one could conduct such processing as a text clustering and classification, text indexing, search results ranking, word sense disambiguation, logical inference, historical information extraction and processing, etc. Authors claim that in 2014 the quality of Medical Text Indexer, based on MetaMap, was as high as 0.6 for precision, 0.56 for recall, and 0.58 for F1 [5].

The MetaMap system was adopted for the Russian language [6]. Authors use Exactus system for the natural language processing, Russian translation of UMLS dictionary, the State Register of Medicinal Remedies and some other local resources. The main purpose of this project is a logical inference for diagnosis of chronic diseases. Using of machine learning algorithms allows the authors to increase the precision of fact extraction up to 82% for a severity of disease and 99% for a flow of disease.

Another big system here is cTAKES [7]. Apart of UMLS, this system uses such extra corpora as SHARP and ShARe. This project also aimed in detection of body part and severity of disease. Authors use SVM method for increasing quality of extraction; however, the final quality is not much better than MetaMap [8].

Currently, the machine learning approach to fact extraction is very common in the medical text processing for different languages. Authors of MedInX system [9] proclaim about 95% for precision and recall in extraction of medical terms from Portuguese text. The same concept is used in TAKELAB system presented at SemEval 2015 devoted to medical texts processing [10]. The authors of [11] use the information extraction approach for a pictorial visualization of an electronic medical record. They extract names of disease and sick body parts, and draw this information on an abstract image of body.

In our project we could not use machine learning techniques [12] since the aim of the project is to find and connect terms stored in the Database of Terms and Observations, described below. That is why we are not using methods of Named Entity Recognition [13] but extracting terms from dictionary and then trying to find correct connections among them. We also could not use common sence onthology [14] or thesauri [15] since they do not containing successfull terminology.

3 Database of Terms and Observations

The main part of our system is the Database of Terms and Observations [16]. It is formed on the basis of the ontology with the same name, designed according to the best modern practice [17,18]. This ontology contains definitions of all concepts classes and consists of two main types of medical terms descriptions – symptoms and factors. Symptoms characterize the current functional state of a patient, and factors are used to describe the risks of various diseases. Symptoms and factors can be combined into logically related groups to make them easier to navigate. Symptoms can be simple or composite. The first ones are described by name and a set of qualitative, numeric, or interval values. Composite symptoms have a name and characteristics. Each characteristic is also described by its name and a set of possible values (qualitative, numerical or interval). Each medical term may have several synonyms. The Database consists of about 1500 Symptomes, about 1300 Features and more than 25500 Values.

The "Symptom" section of this database contains several groups of symptoms: “Complaints”, “Objective examination”, “Laboratory and instrumental examination”. In this article, we use only a group of symptoms of “Complaint”, describing the subjective feelings of the patient, characterizing its current functional status and the state of individual systems: digestive, respiratory, circulatory, nervous system, etc. This group contains a subgroup “General complaints”, which includes those that occur in many diseases (dizziness, weakness, nausea, sweating, etc.). The subgroup “Pain” is a part of the subgroup “General complaints”, it includes the symptoms: headache, back pain, neck pain, sore throat, etc. For most of composite symptoms of the group “Complaints” are used characteristics such as “localization”, “severity”, “cause", “time of occurrence”, “intensity”, “frequency”, etc. The characteristics of the group “Pain” also include the additional characteristics: “irradiation”, “increasing”, “increasing”, etc.

A fragment of the Database of Terms is presented at Fig. 1. The group of symptoms Pains includes Back Pain that has synonyms Spinal Pain and Lumbodynia. The symptom Back Pain has such characteristics as Localization (possible values are Lumbar Region and Lumbar Spine) and Amplification (possible values are Deep Breath and In a Strong Position).

We have also designed an ontology for description of a medical record. It includes personal information, patient complaints, disease flow, patient history, results of general examination, clinical diaries, and diagnosis. Patient complaints are described as symptoms and their values. When forming a medical record, the symptoms and their values defined in the Database of Terms and Observations are used.

The section of the Database in neurology was created according to 3000 anonymized medical records from the Department of Neurosurgery of Far Eastern Federal University. The information resources described above are stored in a heterogeneous repository developed by the authors [16].

Therefore, the aim of the current project is to create a software tool for information retrieval from patients’ medical records. The output is a fragment of the Database of Terms and Observations describing the current state of the patient according to the analysed text of complaints.

4 Algorithm of Term Extraction and Connection

The main idea of the algorithm for patients’ complaints extraction is to run a syntactic analysis and correct its results according to the hierarchy of the Database of Terms and Observations. During syntactic analysis, every extracted

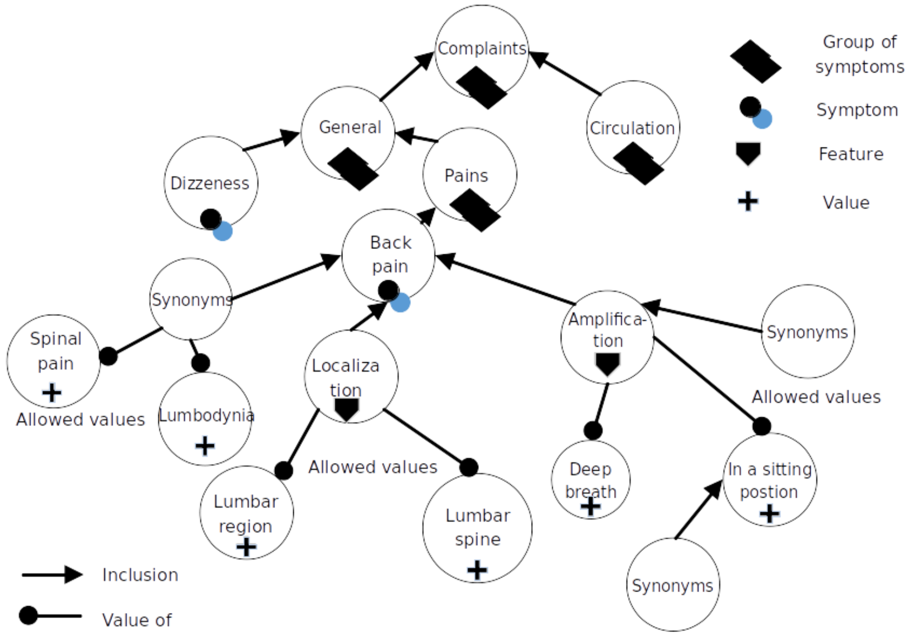

Fig. 1. A fragment of the Database of Terms and Observations

terms is considered as a single syntactic unit. The algorithm consists of two stages. The Stage 1 conducts the pre-syntactic analysis of a medical record and consists of patient's complaints extraction, tagging, and terms extraction. The Stage 2 conducts syntactic analysis of extracted terms and the whole text of the complaint, and correction of resulting dependency tree.

Now we will consider the algorithm in details starting with ***Stage 1***.

Step 1 – Patients' complaints extraction. There are several options here. The first is to load a text of a complaint from a specific field of a healthcare information system or a file in a specialized format. However, some medical records are stored in a plain text format, therefore, we should extract complaints from such text. Patient's complaints are usually placed close to the beginning of a document and start with such phrases as *Complaints, Chief Complaints* (*Жалобы, Жалобы при поступлении, На момент осмотра жалобы*) etc. In some cases a patient doesn't present any problems. Such situation is described by phrases like *No complaints, Doesn't present any problems, Doesn't present age-related problems* (*жалоб нет, жалоб не предъявляет, жалоб по возрасту не предъявляет, жалоб не предъявляет в связи с тяжестью*) etc. Thus, we should exclude paragraphs with the former formulae.

Step 2 – Complaint tagging. The extracted text of a complaint is tagged according to a selected dictionary. Here we consider a word as a sequence of Cyrillic characters. We used PyMorphy2 library [19] with OpenCorpora dictionary [20]. Thus, every token of a text is converted into its most frequent initial form.

Step 3 - Terms extraction. As it was mentioned before, the Database of Terms and Observations stores a hierarchy of such medical terms as Symptoms (P_i), Group of Symptoms (P_i), names of Features (C_{ik}) and their Values (H_{ijk}). The aim of Step 3 is to find such entities and their synonyms in a complaint.

An entity consists of one or several words. In order to find them in a text, the Database of Terms and Observations was converted into a prefix tree; all words are converted to the most frequent initial form using PyMorphy. Thus, the task of terms extraction could be reformulated as finding of the longest sequence of words from text in the prefix tree. Note that a term in a text can be ambiguous, i.e. it's stored in the Database several times in several branches. E.g., a *significant pain* could belong to quite any body part. Thus, such term should be connected to several branches in the Database. Depending on the extracted branch or branches, a word or a multi-word term will be marked as a Symptom, Feature, Value or their combination.

The result of Stage 1 is a sequence of words and word combinations, some of them are tagged as terms and store a list of connected entities from the Database of Terms and Observations. The aim of the ***Stage 2*** is to convert such sequence into a hierarchy according to our Database and to disambiguate the list of connections with branches in the Database.

The Russian language has a lot of peculiarities, therefore, there could be mistakes in connecting words using a parser. E.g., a value could be connected to a wrong feature or symptom since such syntactical connection is more probable than a correct one. The opposite problem is the syntactical incorrectness of some connections between features and their values. We cannot follow [4] and construct a Cartesian product of all possible connections filtering them out according to some common sense rules because of complexity of such calculations. Thus, at the Stage 2 we conduct a syntactic analysis and gradually correct its results according to the Database of Terms and Observations.

Step 1 – Replacing multiword expressions. Here we are going to reduce the complexity of a sentence by joining multiword terms into one token. Selected multiword terms are parsed by UDPipe parser [21]. The parser returns a dependency tree with a main word as a root. We copy all tags of this word to the new constructed word. It makes a sentence shorter, thus, the parser processes it faster and more correctly on the step 2. All non-term or one-word terms are also tagged by UDPipe by reasons of uniformity. All term nodes store a list of routes from the root node of the Database to an entity node with this term; because of ambiguous nature of terms, there could be several routes in the list.

Step 2 – Parsing. At this step we parse the shortened sentence consisting of one- and multiword terms, non-terms, and service words. The parsing allows us not to consider all possible connections, but syntactically reasonable ones only. However, the parser makes some mistakes in syntactic connections, thus we have to correct them at

Step 3 – Correction of the tree. The parser fails since a sentence could be syntactically ambiguous. The Russian language of medical records has a very specific structure; some sentences don't contain a verb in their structure but have a long sequence of patient complaints. Therefore, a constructed dependency tree could have mistakes in terms hierarchy; two connected Values, a Feature that is child of its Value etc. Thus, we have to correct the dependency tree constructed at Step 2 according to the Database of Terms and Observations. We state here that two terms could be connected in a dependency tree only if they have a direct path in the Database. Otherwise, a tree should be corrected according to the following rules.

Rule 1. If a parent node has a lower level in the Database hierarchy that its child node, we should swap these two nodes. E.g., if a Feature becomes a child of a Value, we should exchange them in the tree.

Rule 2. If there is no direct path from a child node to its parent node, we should move the child node to the parent's level of a dependency tree. This situation is possible if a Value was incorrectly subordinated to another Feature or Symptom.

Rule 3. Both parent and child nodes are Features but the child node has an upper level of hierarchy in the Database. In this case we should swap these two nodes.

Rule 4. There are two nodes connected to the same parent; one node has a lower level of hierarchy in the Database than the other one and there is a direct path between them. In this case we should subordinate the first node to the second one.

These four rules should be applied to a tree until the process converges. The rules move an incorrectly subordinated node to an upper level or subordinate it to a node with a higher level of hierarchy that could be a parent for this node. A node could be moved up several times until it finds a possible parent. Otherwise, it will be moved to the highest level and stays here. Such situation is possible if our Database is incomplete and this term should have at least one extra reference.

We can use the resulting tree to filter out ambiguous connections of terms with the Database of Terms and Observations. As it was mentioned above, a term could be placed in several nodes of the Database. We should leave a connection if it satisfies one of the following conditions.

Condition 1. The parent node has a route to a node in the Database that has a direct path to the connected node, i.e., the stored route starts from one of the routes of the parent node.

Condition 2. A child node has a route to a node in the Database that has a direct path to the connected node, i.e., the stored root starts on of the child's rout.

Condition 3. There is at least one neighbouring node with a route that coincides with the stored route.

These conditions state that we could leave only those routes which construct a correct hierarchy. All other routes should be eliminated.

The resulting dependency tree could be used to construct a subtree from the Database of Terms and Observations. According to Conditions 1-3, the resulting tree contains only such routes which can be connected: a parent contains a root (ideally, one root only) that points to a parent node for a child's route. The resulting subtree could be used as a formal description of a patient complaints.

5 An Example of Processing

Let us consider the following example, that was correctly analyzed: «*При поступлении жалобы на выраженную боль в поясничном отделе позвоночника, ограничение движений в пояснично-крестцовом отделе, нарушение ходьбы, нарушение функции тазовых органов*» (*At admission to hospital complains to an significant pain in lumbar spine, restraint of movement in lumbosacral spine, ambulation disorder, dysfunction of pelvic organs*). After selecting of terms, the sentence has the following structure: *При поступлении жалобы на [выраженную] [боль] в [поясничном отделе позвоночника], [ограничение движений] в [пояснично-крестцовом отделе], [нарушение ходьбы], [нарушение функции тазовых органов]*» (*At admission to hospital complains to an [significant] [pain] in [lumbar spine], [restraint of movement] in [lumbosacral spine], [ambulation disorder], [dysfunction of pelvic organs]*). Once terms are joined and parsed, the dependency tree looks like following (translations are given in parenthesis).

```
2 поступлении (admission to hospital)
  1 при (at)
    3 жалобы (complaints)
      6 боль (pain)
        4 на (to)
        5 выраженную (significant)
      8 поясничном отделе позвоночника (lumbar spine)
        7 в (in)
    10 ограничение движений (restraint of movement)
      12 пояснично-крестцовом отделе (lumbosacral spine)
        11 в (in)
      14 нарушение ходьбы (ambulation disorder)
        16 нарушение функции тазовых органов (dysfunction of pelvic organs)
```

Nodes 3, 6 and 8 have the following connected routes:

3 – Жалобы

6 – Жалобы|Общие|Боли

8 – Жалобы|Общие|Боли|Боль в спине|Составной признак|Локализация|Тип возможных значений|Качественные значения|поясничный отдел позвоночника

Nodes 3 and 8 are at the same level of the tree, route 8 starts with route 6. Thus, the node 8 should become a child of 3.

In the same way nodes 14 and 16 should be moved to one level up.

The resulting tree looks as below.

```
2 поступлении (admission to hospital)
  1 при (at)
    3 жалобы (complaints)
      6 боль (pain)
        4 на (to)
        5 выраженную (significant)
      8 поясничном отделе позвоночника (lumbar spine)
        7 в (in)
    10 ограничение движений (restraint of movement)
      12 пояснично-крестцовом отделе (lumbosacral spine)
        11 в (in)
      14 нарушение ходьбы (ambulation disorder)
        16 нарушение функции тазовых органов (dysfunction of pelvic organs)
```

The resulting subtree of the Database of Terms and Observations is following.

```
Жалобы
  Боли
    Боль в спине
      Выраженность|выраженная
      Локализация|поясничный отдел позвоночника
  Опорно-двигательная система
    Снижение подвижности сустава
      Локализация|пояснично-крестцовый отдел
  Нервная система
    Нарушение походки
    Нарушение функции тазовых органов
```

6 Results of Experiments

For our experiments, we have used 3000 of medical records describing a flow of such neurological deseases as microplasia, brain concussion, stenosis etc. Records were sampled and anonymized by neurologists of Department of Neurosurgery of FEFU. All of these records containing patient complaints part, however, some of them were short, did not contain or deny any complaints: "Вялость, отсутствие аппетита, *тошноту*", "Жалоб не предъявляет". An average size of a clinical record is about 1500 word tokens.

100 records of our collection were randomly sampled, processed by our algorithm, and manually checked. The selected records contain 1610 word tokens of patient complaints, 1093 of these tokens were recognized as 711 terms. The examination of selected medical records demonstrates that precision of our algorithm for terms extraction is 0.96 while recall is 0.82; resulting f1-measure is 0.79. The precision was calculated as a relation of number of extracted terms to number of correctly extracted terms; recall was calculated as the ratio of the number of extracted terms to the number of manually tagged terms in a text.

The resulting precision for terms connection was as low as about 0.5. By precision of terms connection we understand the ratio of the number of correctly connected terms (syntactycally connected in text, hyerarcically connected in the Database, with a term type and meaning being properly define) to the number of connections in the resulting graphs. Such low results could be equally explained by incompleteness of the Database and the set of Conditions changing an hierarchy of nodes in a tree. Some terms could be presented in the Database, however they are not liked to any possible position in the Database. E.g Localization characteristics could be connected to any kind of pain; entering a new kind of pain a doctor could forget to link this new pain to theLocalization.

For example, let us consider a sentence "жалобы: на боль в поясничном отделе позвоночника, иррадиирущую в левую ногу, онемение левой стопы, усиление боли при физической нагрузке" which has a misspelling "иррадиирущую", that is why one of the word wasn't properly detected as a term. The resulting graph is presented below.

```
жалобы
 боль
  на
  поясничном отделе позвоночника
   в
  иррадиирущую
   левую ногу
    в
  усиление
   боли
  при физической нагрузке
```

The term "при физической нагрузке" should be connected with the token "боли" but not "боль". Moreover, the term "усиление" was not disambiguated since the parent token have no semantic tags. Thus, there are 4 correctly attributed tokens out of 7 connections in the sentence (we are not considering prepositions here), i.e. the precision of connections is 4/7. In some cases our algorithm connects a node to a wrong parent if it's ambiguous. Thus, our algorithm for building subtree of the Database needs further improvements.

References

1. Catherine R. Selden, Betsy L. Humphreys: Current Bibliographies in Medicine. https://www.nlm.nih.gov/archive/20040831/pubs/cbm/umlscbm.html. Last accessed 25 Apr 2019
2. 2018AB UMLS Release Notes and Bugs Page https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/release/notes.html. Last accessed 25 Apr 2019
3. MSHRUS (MeSH Russian) - Statistics https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/MSHRUS/stats.html. Last accessed 25 Apr 2019
4. Aronson A.R., Lang F.M.: An overview of MetaMap: historical perspective and recent advances. Journal of the American Medical Informatics Association. 2010;17(3):229–236.

5. Mork J., Aronson A., and Demner-Fushman D.: 12 Years On – Is the NLM Medical Text Indexer Still Useful and Relevant? Journal of Biomedical Semantics 2017 8:8.
6. Shelmanov A.O., Smirnov I.V., Vishneva E.A.: Information Extraction from Clinical Texts in Russian. In: Computational Linguistics and Intellectual Technologies. International Conference «Dialogue» (2015). 13; 2015. p. 560–572.
7. Dligach D, Bethard S, Becker L, Miller TA, Savova GK. Discovering body site and severity modifiers in clinical texts. Journal of the American Medical Informatics Association (JAMIA). 2014; p. 448–454.
8. Reatugui R., Ratte S. Comparision of MetaMap and cTAKES for entity extraction in clinical notes. BMC Medical Informatics and Decision Making. 2018. 18 (Suppl 3): 74
9. Ferreira L.S.: Medical information extraction in European Portuguese - Abstract of PhD Thesis. Aveiro: Universidade de Aveiro: Departamento de Electrónica, Telecomunicações e Informática. (2011) https://www.researchgate.net/publication/220009587_Medical_Information_Extraction_in_European_Portuguese.
10. Glavaš G.: TAKELAB: Medical Information Extraction and Linking with MINERAL. In: Proc. of of the 9th International Workshop on Semantic Evaluation (SemEval 2015).
11. Ruan W., Appasani N., Kim K., Vincelli J, Kim H., Lee W.-S. Pictorial Visualization of EMR Summary Interface and Medical Information Extraction of Clinical Notes. In Proc. of 2018 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA). DOI: 10.1109/CIVEMSA.2018.8439958
12. Tutubalina E., Nikolenko S. Combination of deep recurrent neural networks and conditional random fields for extracting adverse drug reactions from user reviews. Journal of healthcare engineering. – 2017. – Т. 2017.
13. Giorgi J., Bader G. Towards reliable named entity recognition in the biomedical domain. BioRxiv. – 2019. – С. 526244.
14. Fellbaum C. WordNet: An Electronic Lexical Database. Cambridge: MIT Press.
15. Loukachevitch N., Dobrov B. RuThes Linguistic Ontology vs. Russian Wordnets. Proceedings of Global WordNet Conference GWC-2014, Tartu
16. Грибова В. В., Москаленко Ф. М., Шахгельдян К.И. и др.: Концепция гетерогенного хранилища биомедицинской информации. In: Информационные технологии, Том 25, №2, М.: "Новые технологии", 2019, сс. 97-106.
17. Sowa J. F. Building, Sharing, and Merging Ontologies. URL: http://www.jfsowa.com/ontology/ontoshar.htm Last modified: 01.18.2009
18. Лукашевич Н. В., Б. В. Добров. "Проектирование лингвистических онтологий для информационных систем в широких предметных областях." Онтология проектирования 5.1 (15)
19. Korobov M.: Morphological Analyzer and Generator for Russian and Ukrainian Languages. In: Analysis of Images, Social Networks and Texts, pp 320-332.
20. Бочаров В. В., Алексеева С. В., Грановский Д. В., Остапук Н. А., Степанова М. Е., Суриков А. В.: Сегментация текста в проекте “Открытый корпус”. In Proc. of "Компьютерная лингвистика и интеллектуальные технологии" Материалы конференции «Диалог» 2012, Вып. 11, Т. 1, сс. 51-60.
21. Milan Straka and Jana Straková. Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. In Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, Vancouver, Canada, August 2017.