ATLANTIS PRESS
Advances in Intelligent Systems Research, volume 166
7th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2019)

Infrastructure of the Electronic Health Record Data Management for Digital Patient Phenotype Creating

Alexander Zakharov
Institute of Mathematics and Computer Science
Tyumen State University
Tyumen, Russia
a.a.zakharov@utmn.ru

Alexander Potapov
Telemedicine Service Group
Tyumen Clinical Hospital No. 1
Tyumen, Russia
dr.potapov@gmail.com

Irina Zakharova
Institute of Mathematics and Computer Science
Tyumen State University
Tyumen, Russia
ei.g.zakharova@utmn.ru

Alexander Kotelnikov
Institute of Mathematics and Computer Science
Tyumen State University
Tyumen, Russia
aakotelnikov72@gmail.com

Dmitriy Panfilenko
Institute of Mathematics and Computer Science
Tyumen State University
Tyumen, Russia
dimjke-72@mail.ru

***Abstract*—improving the health care system requires the effective use of digitized biomedical data already accumulated and constantly updated due to the widespread introduction of applied information systems. One of the priorities here is the creation of a patient's digital phenotype based on data from an electronic health record (EHR). To solve this problem, it is necessary to rely on an ecosystem that provides secure storage, processing and analysis of large volumes of heterogeneous information (tables, images, texts in natural language). The main goal of the article is to study the possibilities of Big Data methods and technologies for the reuse of digitized biomedical data. We have considered designing and implementing of the EHR data management infrastructure that provides tools for digital patient phenotype creating. We designed the data lake prototype "5P Medicine-Big Data" based on the original "Big Data to Smart Data" (BD2SD) multi-layer approach. We have developed an information repository based on the generalized NoSQL approach and the "document repository" model, as well as the system of services. These services provide solutions for the secure data transfer from medical information systems, data storage, validation of information, preliminary data analysis and visualization, data extraction from unstructured documents, and sampling for the machine and deep learning methods. We offered methods for analyzing a significant amount of EHR based on machine learning and Big Data technologies. We applied these methods to extract valid information from unstructured EHR data (first of all, patient examination protocols) and identify characteristic patient categories. We constructed digital patient phenotype, represented by exactly those features extracted from the EHR data that are key in the context of a specific medical problem. Digital phenotypes were formed based on more than 30,000 medical texts for about 2,000 patients. We have shown the effectiveness of the proposed approach by examining the problem of identifying patterns in the patient's visits to a doctor before and after the cardiovascular diseases (angina pectoris, myocardial infarction, and ischemic heart disease) appeared in the electronic health record.**

Keywords*—medical information system, electronic health record, digital phenotype, information storage, Big Data, NoSQL, data extraction, data analysis, machine learning

I. INTRODUCTION

One of the generally recognized consequences of the wide distribution of medical information systems (MIS) is the accumulation of significant amounts of digitized biomedical data (BMD): laboratory results, indicators and images of electrocardiograms, descriptions and images of ultrasound, patient inspection protocols, etc. However, using only the accounting functionality of MIS is unjustified from the point of view of all stakeholders - patients, doctors and researchers, health care managers and medical insurance companies. Therefore, the reuse of BMD, both already accumulated and constantly replenished, acts as one of the promising areas of research. A special place here is occupied by the data of the EHR, determining the patient digital footprint. This is the main basis to identify and study patterns in changing health status, features of the course of diseases, quality of medical care, etc.

At the same time, new knowledge can be extracted from this data by considering a significant number of EHRs and relying on Data Mining methodology (in particular, machine and deep learning methods) and Big Data technology. We mean clustering, classification of patients for various reasons and on various grounds, as well as for risk prediction the emergence and development of various diseases, etc. One of

the priority tasks here is the creation of a "smart" digital patient phenotype (Smart DPP), represented by exactly those features extracted from the EHR data that are key in the context of a specific medical problem.

We assume that in context of this problem it is necessary to rely on a specialized ecosystem that provides secure storage, processing and analysis of a large amount of heterogeneous information (tables, images, texts in natural language). The aim of this research is to study the possibilities of using Big Data methods and technologies for the reuse of digitized biomedical data on the example of the design and implementation of EHR data management infrastructure, which provides tools for creating a digital phenotype of the patient.

II. Big Biomedical Data Mining: Related Work

In recent years, a lot of research has been dedicated to the processing of large volumes of biomedical data. However, at present there is no idea about the general structure of storing and processing this data. This is due both to the peculiarities of BMD, and to the specifics of the health care system in various countries. In their articles, D.V. Belyshev, E.V. Kochurov [1, 2] described the research results aimed at studying the characteristics of data presentation, storage and processing in MIS. They paid special attention to the specific constraints for these systems, which determine the requirements for distributed storage organization.

Such researchers as V.L. Malykh, A.N. Kalinin, A.E. Mikheev, et al. [3,4] analyzed possible structures for BMD repositories. They suggested the development of a relational repository structure by integrating the relational approach with the object and non-relational ones, as well as using process data model and Big Data technologies taking into account the specifics of the clinical and diagnostic process and the nature of the relevant information flows. S. Rea, J. Pathak, K.R. Bailey, et al. presented the solution to the problem of normalization and standardization of EHR data supplied by various MIS in the SHARP project (The Strategic Health IT Advanced Research Projects) [5,6]. They emphasize the fundamental importance of the reuse of accumulated BMD. These data are important not only to improve the quality of treatment, but also to conduct biomedical research and epidemiological monitoring at the level of the national health care system. The authors define the standardization of BMD as the most important problem and propose a platform architecture that provides data exchange services between various organizations, storing, extracting structured data from EHR texts of various formats using Apache tools (Apache clinical Text Analysis and Knowledge Extraction System, cTAKES), as well as standard terminology.

W. Raghupathi, V. Raghupathi [7], G. Manogaran, D. Lopez [8] show that the nature of BMD (large volume, different format and unstructured) plays a special role in the formulation and solution of research problems and opens up wide field for Big Data technology. However, as noted by N. Peek, J.H. Holmes, J. Sun [9], there are a number of problems associated with the features of the transition to new methods of data analysis (requirements for secure distributed data storage and processing). G. Hripcsak and D.J. Albers emphasize that the possible deviations and incorrectness of the original BMD ultimately determine the accuracy of the classification or prediction models and other conclusions [10].

In addition, J. Luo, M. Wu, et al. in their review [11] draw attention to the fact that the application of artificial intelligence technologies and Big Data to solve health problems based on EHC has its own specifics and differs from such areas biomedical informatics, as computer modeling of organs, image processing and analysis, genome research. In this context, the researchers such as L. Luo, L. Li, J. Hu, et al. [12]; K. Kreimeyer, M. Foster, A. Pandey, et al. [13]; A. Névéol, H. Dalianis, S. Velupillai, et al. [14] highlight the problem of extracting objective information from medical texts included in EHR. They note the objectivity of the fact that there are many natural language processing systems (NLP) for English medical texts processing. Therefore, there is the possibility of incorporating this experience when developing similar systems for texts in other languages.

S.M. Meystre, C. Lovis, T. Bürkle, et al. [15]; Y. Wang, L. Wang, M. Rastegar-Mojarad, et al. [16] showed another fundamental feature of the EHR data extraction. This internal structural complexity determines the relationship between survey data and medical prescriptions (procedures, medications), taking into account the chronology of events, etc. As emphasized by Y. Luo [17], L. Wang, Y. Wang, F. Shen, et al. [18], it determines the relevance of machine and deep learning methods already at the stage of extraction and preliminary analysis of EHR data.

Finally, a system of ontologies characterizing various groups of diseases, represented by M. A. Haendel, C.G. Chute, P. N. Robinson [19], plays a crucial role for the correct extraction of data and the use of fundamental objective (biomedical) relationships between entities in interpreting BMD, their classification and building predictive models.

III. Proposed Methodology

Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar. We performed this study as part of the interdisciplinary pilot research project "Smart Patient Digital Phenotype". The project aims to improve the efficiency of health care within the framework of the 5P Medicine concept (prevention, prediction, personalization, participation, practicality) based on the analysis of data from medical information systems using machine learning and Big Data methods and techniques.

The choice of information storage architecture is key to data management. We proceeded from the need to combine data of different types (and, possibly, streaming) from different sources, the possibilities not only for intellectual data analysis, but also for further system scaling. Based on the peculiarities of BMD reuse tasks and following the ideas of G. Manogaran, C. Thota, D. Lopez, et al. [20], E. Maini, B. Venkateswarlu, A. Gupta [21], we developed information storage as a Data Lake.

We propose the original "Big Data to Smart Data" (BD2SD) approach for developing an information storage architecture. Based on these preconditions, we designed the multi-layered information storage as the basis of the system

architecture. The first layer provides distributed storage of BMD in its original form (distributed raw data, DRD), which is generally different for different MIS. These data are loaded with concomitant processing (metadata extraction, depersonalization, etc.) into the next layer (pre-processed data, PPD).

The design and implementation of the PPD layer requires certainty when choosing a relational or not only a relational (Not only SQL, NoSQL) nature of relations between entities. The features of digitized data presented in EHR, namely, their heterogeneity, poor structuring (the presence of various tables and non-patterned texts in natural language) and, finally, significant amounts, suggest large biomedical data (BBMD, Big Biomedical Data [12]). Based on the results of research carried out by A.B.M. Moniruzzaman, S.A. Hossain [22]; S. Sharma, R. Shandilya, S. Patnaik, et al. [23]; S.M. Freire, D. Teodoro, F. Wei-Kleiner, et al. [24], we used the NoSQL approach to develop the PPD information storage layer. In our reasoning, we also started from the fact that there is another advantage of the NoSQL, which is very important for the scalability of the entire system. This means that it is possible to form a data structure in accordance with the usage scenario.

Further, SD layer services extract certain data necessary for solving a specific task from the PPD layer data and transfer it to a flexible Smart Data (SD) storage layer in a tabular form. Special SD layer services perform correctness checking, preliminary analysis and visualization, and form data sets for the medical problems research using machine and deep learning methods. The SD layer also stores problem-oriented knowledge bases (criteria, rules, medical guidelines and recommendations) that are replenished and / or adjusted for solving certain tasks using the NoSQL query method.

IV. Results

A. Infrastructure of the EHR Data Management

We used more than 30,000 medical texts from EHR for about 2,000 patients stored in the MIS SAP for Healthcare[1] and 1C: Medicine. Polyclinic[2] to create a prototype of the data lake "5P Medicine-Big Data".

In the future, this repository will be expanded, since significant amounts of data are collected in these systems. In 2013, in order to provide information support for specialized medical electronic document management in the healthcare sector in the Tyumen Region, the Regional Segment of the Unified State Information System in the Healthcare sector was created. In addition to the information required for the generation of statistical reports, the information system contains information on personalized accounting for the provision of medical services, including medical examination protocols for about a million patients from polyclinics and hospitals in Tyumen and the south of the Tyumen Region.

In accordance with the proposed BD2SD multi-layer approach, the information repository includes three layers. To create the first layer (DRD), which ensures the secure exchange of distributed BBMD in its original form, when unloading data from the MIS, the DRD layer service performs the depersonalization of personal information and attributive encryption for transmitting key information.

To create a PPD layer, we used the advantages of the NoSQL approach, which simplifies the solution of storage scalability and provides opportunities for coordinated distributed storage of key metadata of individual records (record creation date, document structure, doctor ID) and the EHR as a whole (patient ID, birth date, gender, etc.). Since the individual EHR records are presented in the form of XML files, the structure of documents is automatically determined. The services of this layer extract atomic metadata (gender, date of birth) and check their correctness using parallel search methods in hierarchical data structures (trees).

Each patient corresponds to the directory in the PPD layer. This directory stores medical records of various formats (laboratory results, descriptions of ultrasound examinations and electrocardiograms, examination protocols, etc.) in separate XML files. Static information about the patient (ID, gender, date of birth) extracted from the EHR is also associated with each directory. The data of the PPD layer are represented by object files, which can be divided into 3 types: tables, partially structured text, and unstructured text.

Table-type files have a clear structure with certain field names. These files are formed, as a rule, according to the results of laboratory tests (blood tests, urine tests, etc.). Therefore, the files are conveniently stored in a relational database that allows the use of an XML data type with the ability to extract data, but in the context of the NoSQL methodology. In our case, we used the PostgreSQL DBMS.

Partially structured text files contain textual conclusions in an arbitrary form, as well as unambiguously identifiable and mandatory fields (for example, electrocardiogram indicators, ultrasound examination of a specific organ, etc.).

Finally, files of the third type are formed form the patient examination records. Therefore, these files contain an unstructured text, which, along with the doctor’s conclusion and recommendations, may contain additional patient features (height, weight, smoking, drinking strong alcoholic drinks, etc.). In this case, the diagnosis can also be recorded not only in accordance with the International Classification of Diseases ICD-10 (http://mkb-10.com), but also rather arbitrarily (for example, I25, coronary artery disease, chronic ischemic heart disease). In general, the overall structure of such files will be different for different MIS.

The described nature of the source data determined the choice of a flexible data storage structure (such as “document repository”) and its integration with a scalable data extraction service. The invariant component of the service provides both direct and indirect extraction of the most important information - the date of the visit to the doctor. We carried out a preliminary processing of the

[1] https://www.sap.com/cis/industries/healthcare.html
[2] https://solutions.1c.ru/catalog/clinic

examination records of the patient, and as a result, we found out that in more than 20% of the records the corresponding field is either missing, not filled, or filled out with a format violation. These results justify the need for this component.

Since the SD level involves the extraction of data to solve a specific problem, we have developed a service that extracts the diagnosis from the medical text. The diagnosis refers to a specific group of cardiovascular diseases. To extract valid information related to specific named entities (in this case, diagnosis) from unstructured EHR data, we previously performed normalization using NLP methods, formed a dictionary of synonyms, and then replaced the diagnosis text with ICD-10 code.

B. Detection of the Patient Behavior Patterns

From the position of preventive medicine in cardiology, not only early diagnosis is very important, but also personalized prediction of the risk of cardiovascular diseases (CVD). Such prediction allows to reduce the number of cases when a full diagnosis occurs after the occurrence of adverse cardiac events. Methods of artificial intelligence help to identify implicit hierarchical relationships and dependencies of indicators that directly determine the patient's condition, which are used in predictive models. Integrating the predictive logic of expert rules with machine learning approaches can improve the quality of diagnosis and risk prediction of CVD.

The results of mass screening [25] also indicate the need to use machine learning methods for improving the accuracy of predictive models of CVD risk. They confirmed the importance of taking into account age, gender, nature of physical activity to clarify the threshold values of ECG indicators (limits of the norm), which can vary significantly even in healthy people.

In the context of this study, we proceed from the fact that it is possible to increase the effectiveness and reliability of the personalized prediction of the CVD risk if we rely on predictive models that will be based on the Smart DPP. Under Smart DPP, we understand the optimal data structure that combines the indicators that are key to predicting the CVD risk from the point of view of criteria (rules) adopted in cardiology, with features identified by Data Mining methods.

Traditional medical criteria (Reynolds Risk Score, Heart Score [26]) use objective information (height, weight, gender, age, laboratory data, blood pressure, ECG values), as well as subjective data provided by the patient (smoking, exercise etc.). We proceed from the fact that Smart CVD-DPP (that is, a set of characteristics determining the risk of CVD), in the context of 5P Medicine, should include additional data characterizing both the patient and the doctor. We are talking about dynamic features that show patterns in the attitude to the health of the patient himself in dynamics (visiting the doctor, timeliness of testing and passing additional examinations etc.). We assume that identifying patterns of patient behavior will provide an opportunity to form a set of personalized Smart CVD-DPP for more accurate prediction of CVD risk.

To validate the EHR data management infrastructure, we examined the identification of patterns in the dynamics of patient visits to the doctor before and after mentioning the diagnosis of categories I20-I25 (angina, myocardial infarction, coronary heart disease in accordance with ICD-10) in EHR.

The source (raw) dataset contains medical texts (laboratory tests, examination protocols, descriptions of ECG and ultrasound) from various MIS for 2016-2018: 33,285 records (of which 20,256 are protocols), total number of patients - 1823.

Preprocessed dataset included data on patients who visited the doctor more than three times: 19168 records (only protocols); the number of patients - 1029 (male - 350, female - 679); the number of protocols with a diagnosis of category I20-I25 - 6138; the number of patients diagnosed with category I20-I25 - 333 (male - 120, female - 213).

According to the dates of visiting the local doctor, cardiologist, emergency admission in the clinic and doctor's calls for home consultations, the service of the SD layer calculates the appropriate intervals, as well as the maximum and minimum values for each patient. We believe that the results shown in fig. 1 and fig. 2 are of interest for CVD risk prediction models. The graphs show the features of the density distribution function for the maximum values of the intervals before (OK) and after the diagnosis of CVD (I20-I25) for women (Fig. 1) and for men (Fig. 2).

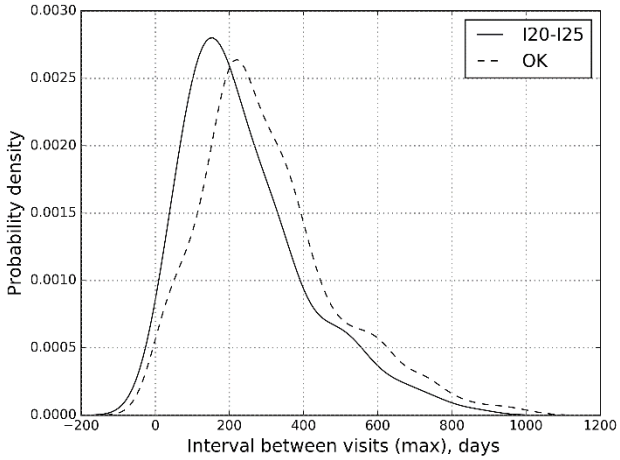

Fig. 1. The distribution of the maximum intervals between visits for women, before (OK) and after the diagnosis of CVD (I20-I25)

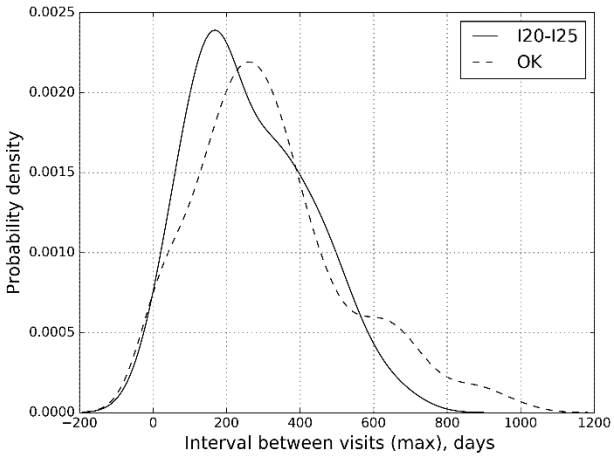

Fig. 2. The distribution of the maximum intervals between visits for men, before (OK) and after the diagnosis of CVD (I20-I25)

Visualization of the data processing results, extracted from the EHR, shows the presence of at least four patterns of patient behavior. Further segmentation by age allows clarifying the features of the shown distributions

In addition to such generalized dependencies, the special service of the SD layer allows visualizing (for a general practitioner, cardiologist, insurance company, etc.) the dynamics of visiting a doctor by a certain patient with a diagnosis marked on the visit date (Fig. 3).

This new knowledge, such as the ones presented, determines the possibilities of developing the SD layer of the system, since they are not only important in forming attributes for machine learning methods. Another equally important goal is to improve the database of personalized medical recommendations.

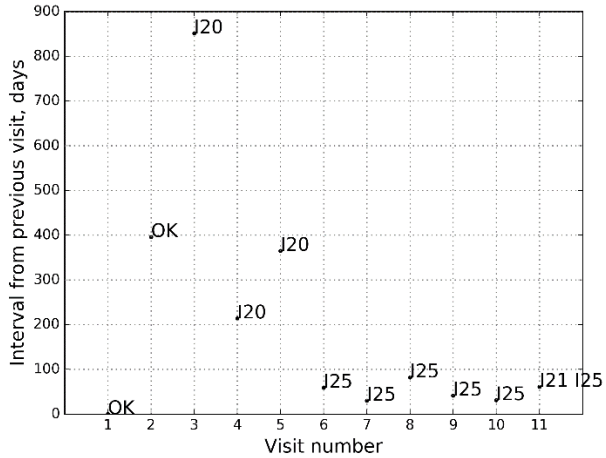

Fig. 3. Patient-specific visit schedule

V. Conclusion

Main results of the study are as follows:

- Designing a prototype data lake "5P Medicine-Big Data" based on original BD2SD multi-layer approach.
- Creating an EHR data management infrastructure based on the NoSQL approach and services for data extracting and phasing data processing.
- Validating proposed approach and created infrastructure on the example of identifying patterns in the dynamics of visits to the doctor by patients in the presence and absence of a diagnosis of categories I20-I25 (angina pectoris, myocardial infarction, coronary heart disease) in accordance with ICD-10 based on the Smart digital patient phenotype.

References

[1] D. V. Belyshev, E. V.Kochurov, “Analysis of data storage methods in modern medical information systems”, Software systems: theory and applications. 2016. Vol. 7. № 2. P. 85-103. (in Russ.)

[2] D. V. Belyshev, E. V.Kochurov, “Promising methods of working with data in medical information systems”, Software systems: theory and applications. 2016. Vol. 7. № 3. P. 79-97. (in Russ.)

[3] V. L. Malykh, A. N. Kalinin, T. Sh.Yusufov, “Object-relational approach to building a data warehouse”, Software systems: theory and applications. 2018. Vol. 8. № 3. P. 180-187. (in Russ.)

[4] V. L. Malykh, A. E. Mikheev, S. V. Rudetsky, “Problem-oriented model of a clinical data bank”, Software systems: theory and applications. 2018. Vo. 9. № 4. P. 219-237. (in Russ.)

[5] S. Rea, , J.Pathak, , Savova, G. Oniki, L.Westberg, C. E. Beebe, C. G.Chute, “Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: the SHARPn project”, Journal of biomedical informatics. 2012. Vol. 45. № 4. P. 763-771.

[6] P. Jyotishman, K. R. Bailey, C. E. Beebe, S. Bethard, D. S Carrell, P. J. Chen, D. Dligach et al. “Normalization and standardization of electronic health records for high-throughput phenotyping: the SHARPn consortium”, Journal of the American Medical Informatics Association. 2013. Vol. 20. № 2. P. 341-348.

[7] W. Raghupathi, V. Raghupathi, “Big data analytics in healthcare: promise and potential”,Health information science and systems. - 2014. №2-3. URL: https://link.springer.com/content/pdf/10.1186/2047-2501-2-3.pdf (date views 30.01.2019).

[8] G. Manogaran, D. Lopez, “A survey of big data architectures and machine learning algorithms in healthcare”, International Journal of Biomedical Engineering and Technology. 2017. Vol. 25. № 2-4. P. 182-211.

[9] N. Peek, J. H. Holmes, J. Sun, “Technical challenges for big data in biomedicine and health: data sources, infrastructure, and analytics”, Yearbook of medical informatics. 2014. Vol. 9. № 1. P. 42-47.

[10] G. Hripcsak, D. J. Albers, “High-fidelity phenotyping: richness and freedom from bias”, Journal of the American Medical Informatics Association. 2017. Vol. 25. № 3. P. 289-294.

[11] J. Luo, M. Wu, D. Gopukumar, Y. Zhao, “Big data application in biomedical research and health care: a literature review”, Biomedical informatics insights. – 2016. - № 8-2. URL:https://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC4720168&blobtype=pdf (date views 30.01.2019).

[12] L. Luo, L. Li, J. Hu, X. Wang, B.Hou, et al. “A hybrid solution for extracting structured medical information from unstructured data in medical records via a double-reading/entry system”, BMC medical informatics and decision making. – 2016. - № 16-1. URL:https://bmcmedinformdecismak.biomedcentral.com/track/pdf/10.1186/s12911-016-0357-5 (date views 30.01.2019).

[13] K. Kreimeyer, M. Foster, A. Pandey, N. Arya, G. Halford, S. Jones, “Natural language processing systems for capturing and standardizing unstructured clinical information”, A systematic review” Journal of biomedical informatics. 2017. Vol. 73. P. 14-29.

[14] A. Névéol, H. Dalianis, S. Velupillai, G. Savova, P. Zweigenbaum, “Clinical natural language processing in languages other than english: opportunities and challenges”, Journal of biomedical semantics. – 2018. - № 9-1. URL: https://jbiomedsem.biomedcentral.com/articles/10.1186/s13326-018-0179-8 (date views 30.01.2019).

[15] S. M. Meystre, C. Lovis, T. Bürkle, G. Tognola, A C. Budrionis, U. Lehmann. “Clinical data reuse or secondary use: current status and potential future progress”, Yearbook of medical informatics. 2017. Vol. 26. № 1. P. 38-52.

[16] Wang, Y., Wang, L., Rastegar-Mojarad, M., Moon, S., Shen, F. Feichen S., Naveed A., “Clinical information extraction applications: a literature review”, Journal of biomedical informatics. 2018. Vol. 77. P. 34-49.

[17] Y.Luo, “Recurrent neural networks for classifying relations in clinical notes”, Journal of biomedical informatics. 2017. Vol. 72. P. 85-95.

[18] L. Wang, Y. Wang, F. Shen, M. Rastegar-Mojarad, H. Liu, “Predicting “Practice Setting Using Topic Modeling”, Proc. Of the 6th IEEE International Conference on Healthcare Informatics Workshops, ICHI-W 2018. USA, NY, New York City. Institute of Electrical and Electronics Engineers Inc. 2018. P. 62-63.

[19] M. A. Haendel, C. G. Chute, P. N Robinson, “Classification, ontology, and precision medicine“, New England Journal of Medicine. 2018. Vol. 379. № 15. P. 1452-1462.

[20] G. Manogaran, C.Thota, D. Lopez, V. Vijayakumar, K.M. Abbas, , et al. “Big data knowledge system in healthcar”, In Internet of things and big data technologies for next generation healthcare. 2017. Springer, Cham. P. 133-157.

[21] E. Maini, B. Venkateswarlu, A. Gupta “Data Lake - An Optimum Solution for Storage and Analytics of Big Data in Cardiovascular Disease Prediction System”, International Journal of Computational

Engineering & Management. 2018. Vol. 21. № 6. URL:http://ijcem.org/papers112018/ijcem_112018_05.pdf (date views 30.01.2019).

[22] A. B Moniruzzaman,. S.A. Hossain "NoSQL database: New era of databases for big data analytics-classification, characteristics and comparison", arXiv preprint arXiv:1307.0191. 2013. URL:https://arxiv.org/ftp/arxiv/papers/1307/1307.0191.pdf (date views 30.01.2019).

[23] S. Sharma, R. Shandilya, R., S. Patnaik, A.Mahapatra? "Leading NoSQL models for handling Big Data", a brief review International Journal of Business Information System. 2016. Vol. 22. № 1. P. 1-25.

[24] S. M. Freire, D. Teodoro, F. Wei-Kleiner, E. Sundvall, D, Karlsson, P. Lambrix, "Comparing the performance of NoSQL approaches for managing archetype-based electronic health record data", PloS one. 2016. Vol. 11. № 3. URL:https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0150069 (date views 30.01.2019).

[25] P.R.Rijnbeek, G.Herpen, Michiel L. Bots, S. Man, N.Verweij, A.Hofman et al. "Normal values of the electrocardiogram for ages 16–90 years", Journal of Electrocardiology. 2014. V. 47. № 6. P. 914-921.

[26] M. Petretta, A.Cuocolo. "Prediction models for risk classification in cardiovascular disease", European journal of nuclear medicine and molecular imaging. 2012. Vol. 39. № 12. P. 1959-1969.