

Cloud Infrastructure for Creation of Interpretable Diagnostic Knowledge Bases of Diseases Regardless their Etiology

Valeriya Gribova

*Institute of Automation and Control
Processes Far Eastern Branch of
Russian Academy of Sciences
Vladivostok, Russia
gribova@iacp.dvo.ru*

Elena Shalfeeva

*Institute of Automation and Control
Processes Far Eastern Branch of
Russian Academy of Sciences
Vladivostok, Russia
shalfe@iacp.dvo.ru*

Margaret Petryaeva

*Institute of Automation and Control
Processes Far Eastern Branch of
Russian Academy of Sciences
Vladivostok, Russia
margaret@iacp.dvo.ru*

Abstract—The paper presents a cloud infrastructure for creation of interpretable diagnostic knowledge bases for intelligent decision support systems in medicine. An ontology of medical diagnostics allows us to form diseases regardless their etiology as a multivariate developing internal process. The relations types being defined in the ontology allow users to create a knowledge base sufficient for search of hypotheses of possible single diagnoses, for detection of combined and complicated pathologies, and for reduction of a set of hypotheses of diagnosis. The offered infrastructure includes the following tools: the knowledge base editor being generated on ontology, the terminology database editor, the tool for estimation of knowledge base quality and some instruments of automation of development and improvement of knowledge base. The infrastructure is implemented on the IACPaaS platform and used by experts to create knowledge bases in various fields of medicine. The article is useful for developers of medical diagnostics systems.

Keywords—diagnostics, disease, ontology, domain terminology base, knowledge base, intelligent system, knowledge-based system.

I. INTRODUCTION

The ongoing computerization of medicine has already made it possible to achieve success in organization of workflow, formation of statistical reports. It also made more convenient interaction of patients with institutions and individual specialists. However, the era of automation of support for intellectual work of physicians has not yet arrived. One of the problems is related to the peculiarity of the knowledge used by doctors: it's based on the most complicated causal relationships, their volume is extremely large, in some areas it increases exponentially [1, 2].

Intelligent systems for diagnostics and prediction based on neural networks are already achieved high enough percentage of accuracy. But trusting their results in practice is unacceptable, because the decision must be clear to a doctor, and such systems do not provide an detailed explanation [3, 4].

Intelligent systems based on precedents do not give a very convincing result, since a set of patient indicators is almost always unique and not necessarily complete. In addition, work with time series that reflect the patient's disease development process is practically not supported. Nevertheless, it is possible to obtain results that are useful to

a physician making or analyzing the decision for diagnostic and prognosis tasks [5].

For the knowledge-based systems that are able to provide a reasoned advice or decision, it is important to have up-to-date knowledge bases [6, 7]. But physicians or experts have not yet been offered the means of forming knowledge in familiar terms and structure to describe disease symptoms, syndromes, factors, nosological forms, etiologies, pathogenesis, clinical manifestations for different categories of patients, and to use this knowledge by decision support systems [4, 7, 8].

For several decades, it has been established and tested that the model of knowledge about diseases (ontology) does not depend (invariant) on their etiology [9]. To date, a paradigm has been proposed and a cloud-based information editing technology has been implemented on the basis of the explicitly described structure of this information (ontology), the domain of used terminology and the limitations of interpretation as well as the technology for creating software services to support and automate the work of specialists in processing this information [10]. The aim of the article is to present a cloud infrastructure for the formation of interpretable bases of diagnostic knowledge as components of decision support systems (DSS) in medicine.

II. ONTOLOGY OF MEDICAL KNOWLEDGE

One of the medical diagnostics ontologies, corresponding to the real concepts of medicine and a wide range of diseases, is the Ontology of medical diagnosis of diseases [9]. It contains a set of relationships between concepts to describe the clinical picture of diseases in dynamics of the pathological process, taking into account an influence of treatment and other events on diseases' manifestations (Fig.1).

The ontology of medical diagnosis defines types of statements such as <disease-k, symptom-j, range of values-kj of symptom-j>; the necessary condition for the existence of the disease, the variant of the symptom complex for the disease; <variation-m of the process of changing values-kj of symptom-j for some symptom complex>, conditions necessary for emergence of the disease; response to the impact of an event, etc. By means of statements of such types it becomes possible to build knowledge base which sufficient to find hypotheses about possible diagnoses, to identify combined and complicated pathologies, and to reduce the set of hypotheses about the diagnosis. (The latter is solved by

searching for such additional information about the patient's condition, the answer to which will refute one of the previously alleged diagnoses.)

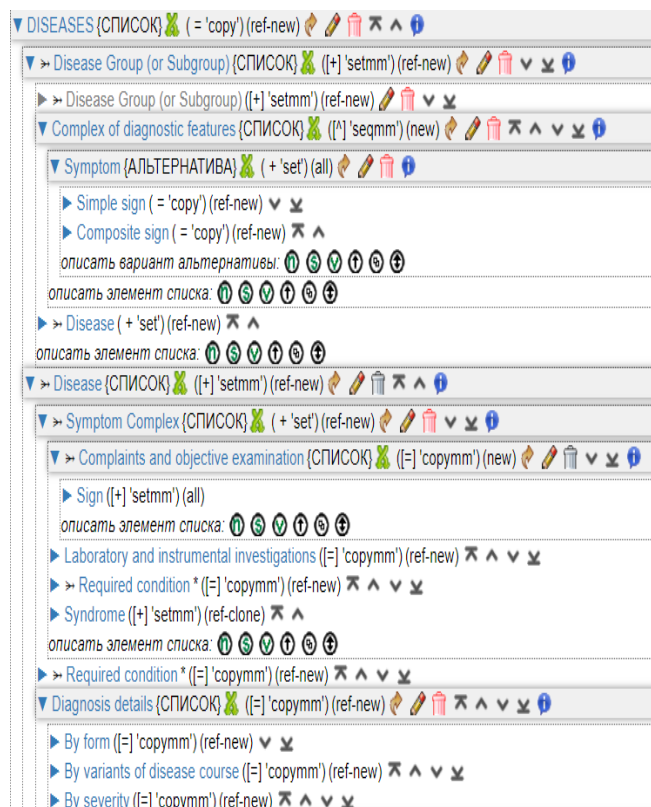


Fig. 1. The fragment of the ontology of diagnostic on the IACPaaS platform

The ontology of medical diagnosis has the following characteristics.

- The possibility of disease symptom complexes formation taking into account the categories of patients using reference ranges instead of average values for laboratory and instrumental examinations.
- The possibility of alternative symptom complexes formation with different approaches to the identification of reliable signs of the disease in order to select the most gentle, quick or inexpensive method of diagnosis.
- The possibility of specifying diagnoses taking into account the etiology, pathogenesis, case scenario, etc., for differential diagnosis of diseases and the choice of appropriate treatment methods.
- Uniform formalization of stages of chronic and periods of acute diseases.
- The possibility of specifying a measure of an indistinct symptom in a certain period of dynamics: the modality, the values of which (necessary, characteristic, possible) are sufficient for recommendations on the confirmation or refutation of the hypothesis; the probability scale of the conformity of the signs to the diagnosis hypothesis.
- Inclusion of the values of characteristics and symptoms, modified by influence of events. The

presence of such element of cause-effect relationships allows us to take into account external effects on the patients at different stages of the disease.

- Usage different options for dynamics of the values of symptoms. Such elements of knowledge allow us to take into account (on the basis of medical experience) the variety of a course of the same diseases for different patients.

A symptom of a disease can be simple or composite, its values are presented by periods of the dynamics of the development of a symptom or disease in general; Necessary conditions for a symptom can be specified. Each period of the dynamics is characterized by an upper and lower limit of the period duration, the unit of measurement. The composite symptom contains a description of the sets of its period-varying characteristics with modality.

III. BASIC REQUIREMENTS FOR FORMATION OF INTERPRETABLE KNOWLEDGE BASES TOOLS

After The creation of interpretable knowledge bases is possible when they are formed on the basis of an ontology, which sets a rules for generating information and restricting of its interpretation.

An ontology becomes structural basis of editing tools. Accordingly, an environment for creating interpretable knowledge bases should include an ontology editor, which will allow knowledge engineers (cognitive experts) to make changes in the ontology, if necessary.

Such environment should be provided by a knowledge base editor, controlled by an ontology and oriented to domain experts. Considering the different level of experts, different structure of knowledge being formed, the knowledge base editor must have advanced tools for front-end user support: the ability to view with varying degrees of detail and in different forms fragments of the knowledge base, control the correctness of the entered information, taking into account the limitations defined in the ontology, completeness of the knowledge base formation.

Knowledge bases should be formed in uniform terms of the domain to provide their unambiguous interpretation by both software services and various expert communities. To meet this requirement, the infrastructure for building knowledge bases should include a glossary of terms, which, like the knowledge base, should be formed on the basis of an ontology using a term editor. In this case, when forming statements in the knowledge base, the terms (symptoms and factors, their values) should not be defined in the knowledge base, but should be selected from the glossary of terms. The ontology oriented to experts and the glossary of terms allows us to provide an important property of knowledge bases — their interpretability.

The importance of quality of a knowledge bases requires tools for estimation their quality on representative archives of solved problems.

The set of the ontology editor, the knowledge base editor, the tool for estimation of knowledge base quality on representative archives of solved problems, the term editor form the infrastructure for the formation of interpretable diagnostic knowledge bases.

The importance of a knowledge base development, tools are needed for the continuous development (improvement) of knowledge bases not only by domain experts, but automatically (with subsequent verification by experts) based on new data (case records). This function is provided by tools of inductive formation of knowledge bases or knowledge base fragments. At the same time, each new version of a knowledge base will require use of a quality control tool (to confirm that previously solved tasks are solved no worse). Adding to the above-noted infrastructure tools the tool for inductive forming knowledge bases transforms the infrastructure for forming interpretable diagnostic knowledge bases into a specialized environment for their development.

IV. CHARACTERISTIC OF IACPaaS TOOLS FOR KNOWLEDGE FORMATION

The IACPaaS platform [4] implements a set of tools for processing information resources (data and knowledge bases), which are formed according to a clearly represented ontology. Ontologies, data and knowledge bases are represented by hierarchical semantic networks. Hierarchical semantic networks of IACPaaS platform are sufficient to describe cause-effect relationships (between symptoms, ranges of their values, etc.).

The IACPaaS platform has the capabilities to expand tools for processing information resources into a full-scale specialized environment - infrastructure for the formation and development of interpretable diagnostic knowledge bases.

The ontology editor provides a simple and convenient creation of ontologies on a metalanguage [11]. The knowledge base editor generator of the IACPaaS platform is automatically generates the knowledge base editor. Three types of user interfaces are supported for editing and viewing a knowledge base. The generated knowledge base editor provides the process of formation knowledge bases using an ontology, checking restrictions during knowledge base formation, and control of its integrity and completeness. The remaining tools are added to the platform as soon as they are ready.

To implement the process of monotonous improvement of knowledge bases (based on the method of inductive formation of fragments of knowledge bases), a tool to support the selection of precedents (correctly solved problems) is required.

Thus, the complex of IACPaaS tools for the formation of knowledge bases and their development is as follows: the editor of ontologies, the generator of knowledge base editors with the choice of three types of user interface, the editor of knowledge bases, controlled by the ontology with a self-adapting interface (when changing the ontology), the term editor, the tool for estimation of knowledge base quality on representative archives of solved problems, the tool for inductive forming knowledge bases, and the tool for selecting precedents for improving knowledge bases.

V. FORMATION OF KNOWLEDGE WITH THE SUPPORT OF THEIR UPDATE TOOLS IACPaaS

Knowledge bases on the diagnosis of diseases from various fields of medicine were formed (Fig.2): group of

diseases of the gallbladder, biliary tract and pancreas; metabolic disorders of glucosaminoglycans; hemorrhagic fever; respiratory diseases, and ophthalmic diseases [12-15].

Diseases are represented by several symptom complexes, corresponding to different case scenario of the disease in different age groups of patients or different diagnostic methods. In the complexes of complaints, objective, laboratory and instrumental examinations, a number of symptoms are presented, changes in the values of which are symptoms of the disease. Detailed diagnosis provides an additional set of symptoms that allows us to make the appropriate clarification to the main diagnosis, taking into account the etiology, pathogenesis, case scenario, stage, etc.

After the formalization of knowledge about each disease on the basis of expert knowledge, their verification with reference solutions from the medical archive is carried out. To do this, we use a software service that analyzes each symptom in the reference solution with the knowledge base and provides a detailed report.

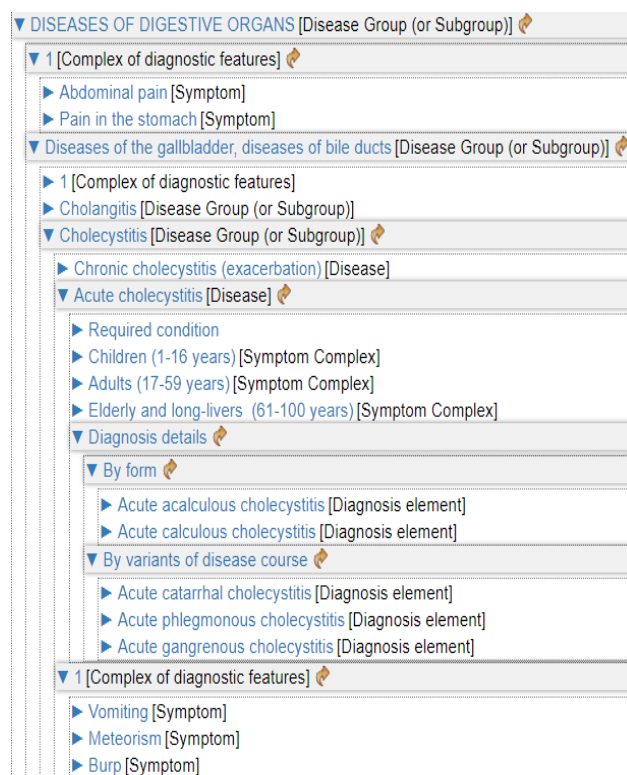


Fig. 2. The fragment of the disease diagnosis knowledge base on the IACPaaS platform

If you need to update the knowledge base in connection with obtaining precedents that are not consistent with the knowledge base, then it is effective to automatically (using inductive methods) form a new version of the knowledge base and then carry out its verification .

According to the principle of monotonous improvement of knowledge bases, adaptation is required more often than it is advisable to use induction. In obtaining new knowledge (statements), as well as in obtaining individual facts from practice, indicating an update of the clinical picture of diseases, it is necessary to modify the knowledge base manually. This process requires a knowledge base editor and a the tool for estimation of knowledge base quality. The

availability and operability of tools for the implementation of the process of monotonous improvement of knowledge bases (based on the methods of adaptation and inductive formation) is provided by the property “updatibility of knowledge”.

The complex of IACPaaS tools has an ontology editor. Adding new terms or new relationships between terms to the ontology ensures the integrity of the knowledge base, and the user interface of the domain expert is adapted automatically.

VI. CONCLUSION

The proposed infrastructure for the formation of interpretable medical diagnostic knowledge bases includes the following tools: the editor of ontologies, the editor of knowledge bases, which is generated by the ontology with a self-adaptable user interface, the term editor, the tool for estimation of knowledge base quality, the tool for inductive forming of the new version of the knowledge base (currently in development stage), and the tool for selecting precedents for improving knowledge bases.

The ontology, on the basis of which knowledge is formed, allows us to formalize any disease as a multivariate developing internal process.

The infrastructure is implemented on the IACPaaS platform (<http://iacpaas.dvo.ru/>) is already used by experts to create knowledge bases in various fields of medicine. Experts interested in the accumulation, improvement and using knowledge about diagnosis have the opportunity to join this process (both for their own bases and for shared resources).

ACKNOWLEDGMENT

The research is carried out with the partial financial support of the Russian Foundation for Basic Research (projects 18-29-03131 and 19-29-01077).

REFERENCES

- [1] Rybina G.V., Demidov D.V., Fontalina E.S. Automated acquisition, presentation and processing of temporal knowledge in dynamic integrated expert systems // Sixteenth National Conference on artificial intelligence. Conference proceedings. Vol.1. Voronovo, Moscow: National Research University Higher School of Economics, 2018. P. 61-69.
- [2] Le N.V., Kamaev V.A., Panchenko D.P., Trushkina O.A. Model of knowledge representation when creating a medical expert system for differential diagnostics // Bulletin of Volgograd State Technical University, 2014. № 6. P. 42-50.
- [3] Tang Z. et al. Interpretable classification of Alzheimer’s disease pathologies with a convolutional neural network pipeline // Nature Communications, 2019. № 10. P.1-14.
- [4] Ambilwade R.P., Manza R.R. et al Medical expert systems for diabetes diagnosis: a survey // International Journal of Advanced Research in Computer Science and Software Engineering, 2014. 4
- [5] Aderonke A., Babajide A., Kayode A. An Integrated Knowledge Base System Architecture for Histopathological Diagnosis of Breast Diseases // International Journal of Information Technology and Computer Science. 2013. № 1. P. 74-84.
- [6] Martinelli M., Moroni D. et al. Knowledge-based Infrastructure for the Management of Diagnostic Imaging Procedures in the Heart Failure Domain // Trans. Mass-Data Analysis of Images and Signals 2010. № 2. P. 3-18.
- [7] Chen J. et al. A Disease Diagnosis and Treatment Recommendation System Based on Big Data Mining and Cloud Computing // Information Sciences, 2018. № 435. P. 124–149.
- [8] Tu S., Hrabak K., Campbell J. R. et al. Use of Declarative Statements in Creating and Maintaining Computer-Interpretable Knowledge Bases for Guideline-Based Care // AMIA Annual Symposium proceedings. 2006. P. 784–788.
- [9] Gribova V., Okun D., Petryaeva M., Shalfeeva E., Tarasov A. (2018) Ontology for Differential Diagnosis of Acute and Chronic Diseases. In: Kuznetsov S., Osipov G., Stefanuk V. (eds) Artificial Intelligence. RCAI 2018. Communications in Computer and Information Science, vol 934. Springer, Cham. P. 152-163.
- [10] V. Gribova, A. Kleshev, Ph.Moskalenko, V. Timchenko, L. Fedorischev, E. Shalfeeva. The IACPaaS Cloud Platform: Features and Perspectives // 2017 Second Russia and Pacific Conference on Computer Technology and Applications (RPC) (Vladivostok, Russia, 25-29 sept. 2017). IEEE. 2017. P. 80-84.
- [11] Gribova V.V., Kleshchev A.S., Moskalenko F.M., Timchenko V.A. A Model for Generation of Directed Graphs of Information by the Directed Graph of Metainformation for a Two_Level Model of Information Units with a Complex Structure // Automatic Documentation and Mathematical Linguistics. 2015. Vol. 49, №6. C. 221-231.
- [12] Kobrinskii B.A., Demikova N.S., Blagosklonov N.A. (2018) Expertise Systems in Hereditary Diseases. In: Kuznetsov S., Osipov G., Stefanuk V. (eds) Artificial Intelligence. RCAI 2018. Communications in Computer and Information Science, vol 934. Springer, Cham.
- [13] Kompanets GG, Iunikhina OV, Petryaeva MV, Shalfeeva Ye.A., Okun DB New approaches to the diagnosis of hemorrhagic fever with renal syndrome: the development of a remote diagnostics technology on a cloud platform // Global Scientific Potential, 2018. № 12. P. 80-83.
- [14] Moskalenko F.M., Chernyakhovskaya M.Yu. Formal description of the Chronic pancreatitis disease // Informatics and control systems. 2012. №4. P. 97-106.
- [15] Moskalenko FM, Chernyakhovskaya M.Yu. Base of diseases "conjunctivitis" for a computer knowledge bank. Informatics and control systems. 2011. № 2 (28). P. 50-58.