Authors' Information

Nadezhda Kiselyova – A.A.Baikov Institute of Metallurgy and Materials Science of Russian Academy of Sciences, P.O.Box: 119991 GSP-1, 49, Leninskii Prospect, Moscow, Russia, e-mail: kis@ultra.imet.ac.ru

Andrey Stolyarenko –- A.A.Baikov Institute of Metallurgy and Materials Science of Russian Academy of Sciences, P.O.Box: 119991 GSP-1, 49, Leninskii Prospect, Moscow, Russia, e-mail: stol-drew@yandex.ru

Vladimir Ryazanov – A.A.Dorodnicyn Computer Center of Russian Academy of Sciences, e-mail: riazanov@ccas.ru

Vadim Podbel'skii – Moscow Institute of Electronics and Mathematics (Technical University), P.O.Box: 109028, B.Trehsvjatitelsky per. 3/12, Moscow, Russia, e-mail: vvp@mitme.ru, vpodbelskiy@hse.ru

AN IDEA OF A COMPUTER KNOWLEDGE BANK ON MEDICAL DIAGNOSTICS

Mery Chernyakhovskaya, Alexander Kleschev, Filip Moskalenko

***Abstract:** The paper is a description of information and software content of a computer knowledge bank on medical diagnostics. The classes of its users and the tasks which they can solve are described. The information content of the bank contains three ontologies: an ontology of observations in the field of medical diagnostics, an ontology of knowledge base (diseases) in medical diagnostics and an ontology of case records, and also it contains three classes of information resources for every division of medicine – observation bases, knowledge bases, and data bases (with data about patients), that correspond to these ontologies. Software content consists of editors for information of different kinds (ontologies, bases of observations, knowledge and data), and also of a program which performs medical diagnostics.*

***Keywords**: Medical Diagnostics, ontology model, parallel computing, knowledge bank.*

***ACM Classification Keywords**: I.2.1 Applications and Expert Systems, J.3 Life and Medical Sciences.*

Introduction

Computer systems for medical diagnostics are one of applications of AI systems. They can help doctors to improve the quality of their work. The task of such systems is to recognize diseases (one or several), with which a patient is ill, basing on the results of patient's observations. The important components of such systems are a confidence subsystem which can show to the doctors the knowledge base of the system and an explanation subsystem which can show to the doctors the information and reasoning way which were used to produce the result.

Two classes of the systems for solving the task of medical diagnostics have been developed by now, which differ by methods that lie in their base. The systems of the first class are based on statistical and other mathematical models – their bases are mathematical algorithms that perform the search of usually a partial correspondence between the symptoms of the current patient and the symptoms of previous patients for whom the diagnoses are known [1 – 4]. However such systems lack the confidence and explanation subsystems.

The systems of the second class are based on expert knowledge. Their algorithms operate with the information about the patient and with the knowledge about diseases which are represented in a form that is more or less close to the concepts of doctors (and described by expert-doctors). That is achieved by an explicit or implicit usage of ontologies of medical diagnostics. In these systems it is possible to create the subsystems of confidence and explanation which is capable of giving a doctor the results of analysis of patient's state that led to the derived result.

The models of ontologies used in such systems take into account temporal changes of symptom values [1], connections between the symptoms and diseases (for example, using logical rules) [2, 5], division of observations into several groups (for example, clinical, laboratory, morphological data – [4]), representing the state of a patient as a multi-level model [6].

The algorithms in such systems try to imitate the way of reasoning of doctors [7, 8], to look for a correlation between the information about the patient and clinical findings of diseases that were described by experts [9], or to process the rules that describe the links between observations and diseases and are set by experts [5, 8, 10, 11].

An analysis of recent systems for medical diagnostics of the second type shows that the ontologies that they use are rather simple and do not combine such information (widely used by doctors) as: knowledge about the reasons of diseases; knowledge about different types of cause-effect relation between symptoms and diseases; knowledge about the influence of events on symptom values; knowledge about different variants of temporal changes of symptom values, that depend on anatomical-and-physiological features of patients.

In addition, one of the negative properties of modern systems is that they cannot be widely used. This is because they are either prototypes made for research purposes or they are developed for a particular medical institution and are not available outside its LAN. On the other hand, systems that allow a wide access to their resources though modern technologies (like Internet) do not let experts to broaden their knowledge bases, for example – DXplain [2] and Diagnostics of preeclampsia [12].

Thus, to develop a system for medical diagnostics a) which would be based on expert knowledge and on ontology model that takes into account all features of medical knowledge stated above, b) in which the model of knowledge is close to concepts of doctors and c) which allows not only to recognize the diagnosis but also to explain how it was obtained – is an urgent task. Such a system should realize the diagnostics in acceptable (for doctors) time in spite of the fact that a non-trivial ontology of medical knowledge lies in its base. What is more, such a system must provide an access to its resources for as many users as possible for both purposes – performing medical diagnostics and participating in gathering and improving medical knowledge about different diseases.

The goal of this paper is to describe a concept of a network resource on medical diagnostics that possesses all features stated above.

This paper was made according to the project of FEBRAS № 06-III-A-01-457 «Designing, implementing and developing the bank of medical knowledge in the Internet network» and according to the project of RFBR № 06-07 89071 «Research of possibilities of collective control in semantic web for information resources of different types of generality».

1. A theoretical background and general principles for developing the Bank of knowledge on medical diagnostics

Specialists in the field of AI and experts in medical diagnostics have developed several models of ontology for medical diagnostics in recent years. Some of these models were used for developing diagnostic systems that were based on expert knowledge. As it was stated in the introduction, each of them was somehow better and somehow worse than others [1, 2, 4, 5, 6]. In order to unite the advantages of those models and to create a model of ontology that would be close to concepts of experts in the field of medical diagnostics – an ontology and its model were developed in [13,14,15]. This ontology of medical diagnostics describes acute diseases; interaction of cause-and-effect relations of different types is taken into account. The ontology is close to real concepts of medical diagnostics in the Russian Federation and it defines combined and complicated pathology, development of pathological processes in time, influence of treatment and other events on diseases' manifestations. The model of this ontology includes definitions of terms of the knowledge model (parameters), the definitions of terms of the reality model (unknowns) and is an unenriched logical relationship system with parameters, which also describes of integrity constraints for unknowns, and parameters and relationships between them.

The relationships between unknowns and parameters can be divided into the following semantic groups:

1) the relationships between knowledge about cause-and-effect relations and cause-and-effect relations which take place in situations;

2) the relationships which determine cause-and-effect relations that are the reasons of values of each sign during its development intervals;

3) the relationships which determine the properties of borders of intervals of the time axis for each sign;

4) the relationships which determine the reason of each disease from the diagnosis.

In [16] a general problem of medical diagnostics is formulated based on the described model of the ontology: to recognize the possible diagnoses of a patient basing on the domain knowledge and observation data, which are values of symptoms (during the times of observations), values of anatomical-and-physiological features (constant in time) and values of events that took place with the patient (in the moments when they occur), and also for each diagnosis – to state its reason (some event or another disease) and its explanation (by indicating the reasons of observed values of symptoms).

Because the outlined model of the ontology takes into account a large amount of interrelations between the processes that take place in patient's body, it can be expected that any algorithm for solving the formulated above problem of medical diagnostics (which has to analyze all these relations) will be of an intense computing complexity. One of the ways to improve the effectiveness of this algorithm is to parallel it and then to execute it on a multiprocessor computing system. The maximum effect from paralleling can be achieved while solving some particular but nevertheless a practically important problem – when the observed symptoms can be analyzed independently and thus – simultaneously on the nodes of a multiprocessor computing system. Such problem comes out when several restrictions are applied to the used model of the ontology: patient can be either healthy or sick with only one disease; each disease has particularly one period of development.

Such a problem is specified in the terms of the restricted ontology. An algorithm for its solving is presented in [16]. In [17] both a parallel algorithm for solving that problem and an algorithm for disease database optimization (for reduction of the amount of hypothesizes to check) are presented. The results of an experimental research of time-complexity for the optimized algorithm for solving the particular medical diagnostics problem [17] are described in [18]. This research shows that:

a) usage of the optimized disease database remarkably reduces the time of diagnostics, especially when a great number of diseases is stored in database;

b) the maximum speed of diagnostics is achieved when all the nodes of a multiprocessor computing system are used and on each of them only a single copy of a client-part of the algorithm is executed;

c) the time of diagnostics on al test-patterns (while using all available nodes and using the optimized database) did not exceed several minutes.

The described ontology and the algorithm for medical diagnostics can be used for developing a system that will support the process of a coordinated solving the tasks of gathering, formalizing, translating to a computer-readable representation, engineering, storing, managing and processing data and knowledge in the field of medical diagnostics and will be a combination tool for all this information into a single resource with remote access for various users.

The system will be called a Bank of knowledge on medical diagnostics. In order to implement all the mentioned requirements the system should be built in accordance with three-tier software architecture which means the Bank should consist of the following parts:

- information content with some standard interface for access to the stored information and with unified format for storing that data;
- software content which is oriented on intellectual support for users of the Bank and which includes the following: tools for editing the data of the information content, tools for their processing (optimization of knowledge base about diseases and medical diagnostics of a patient) and an administrative subsystem;
- interfaces for access to software components.

2. The information content of the Bank of knowledge on medical diagnostics

The information content can be divided into several partitions – one for each section of medical diagnostics (ophthalmology, cardiology, etc.). Each partition includes three types of bases, all formed in accordance with the model of the ontology for medical diagnostics [13, 14, 15].

The model of ontology consists of three parts:

- an ontology model for observations,
- an ontology model for knowledge about the influence of diseases on symptom values (model of knowledge about diseases),
- an ontology model for patient (an ontology model for patient's case record).

The ontology model for observations describes the structure of observations and their values. Observations are the observed symptoms, events and anatomical-and-physiological features. The base of observations is built on the basis of this model and it contains the names of events, symptoms and features and also lists of their possible values.

The ontology model for patients' case records is built on the basis of the ontology model for observations and it sets the structure for describing the state of a patient as a function of time. Events can happen to a patient at different moments of time and can have different values. Symptoms also can have different values at different moments of time and the values of anatomical-and-physiological features of a patient are constant in time. Basing on this model of a patient those who perform the medical diagnostics form the base of patients' case records, that is create the records of patients in the information content which store the values of symptoms (observed at different moment of time), the values of events (that happen to a patient at different moments of time) and the values of anatomical-and-physiological features.

The ontology model for knowledge about diseases contains the descriptions of basic terms of knowledge (including the relations between them) in terms of the ontology model for observations. They include the descriptions of diseases (and their reasons) and of the normal state of the patient. This scheme is used by experts as a template for describing the particular diseases (symptom values during diseases are defined by clinical manifestations and clinical manifestations modified by event's influence), their reasons (a set of etiologies is described) and normal state of the patient (normal reactions and reactions to event's influence are described).

Figure 1 shows the general scheme of the information content for the Bank of knowledge on medical diagnostics. Arrows on the figure show how one component of the information content is used for forming another one.

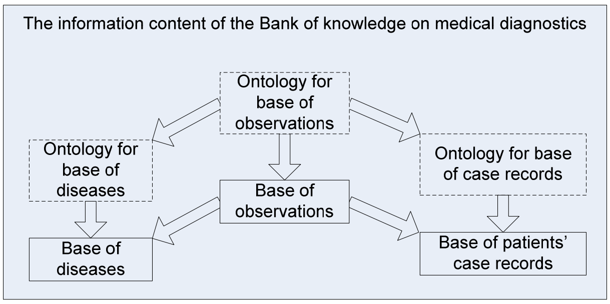

Figure 1. Information content of the Bank of knowledge on medical diagnostics.

3. Software for problems which can be solved by means of the Bank of knowledge on medical diagnostics by users of different types

As it was stated in section 1, the software content of the Bank consists of software components of three types:

- editors for information content,
- problem solvers,
- an administrative subsystem.

In accordance with the mentioned in section 2 parts of the information content of the Bank of knowledge on medical diagnostics the following editors for them have to be developed:

- an editor for observations bases;
- an editor for bases of knowledge about diseases;
- an editor for bases of patients' case records.

For solving the problem of medical diagnostics the above mentioned algorithm is used [17]. It takes into account all relations between knowledge and reality that are stored in the model of ontology [14,15].

The software tool that performs the optimizing transformation of the knowledge base with diseases' descriptions is an implementation of the algorithm described in [17]. It analyzes the information from the disease base and recognizes during which diseases each value of each symptom can be observed. Basing on that information the diagnostics algorithm forms a list of hypothesizes about diagnosis for checking which probably contains fewer diseases than the set of all the diseases.

The groups of users of Bank of knowledge on medical diagnostics are:

- servicing users (administrators),
- information mediums (experts),
- users (doctors and guests).

An administrator traces the functionality of the entire Bank of knowledge with the use of administrative subsystem, which lets him to perform two functions: working with user accounts and controlling the users' activities.

Experts edit the base of observations and the base of knowledge about diseases in accordance with the correspondent models of ontologies. There is a single base of observations in each partition of the information content of the Bank. As for the base of knowledge about diseases – for each user of this type a personal base is formed in a particular partition and he fills it with knowledge. When he achieves some final and essential results during his research, he sends to the administrator a special request for adding that information into the global knowledge base in the information content. If the administrator approves that – he expands the knowledge base with those new descriptions of diseases with the help of administrative subsystem.

Users are the doctors who work with the base of case records and perform the diagnostics and guests who are able only to view the data in the observations base and in the base with descriptions of diseases that are stored in information content.

4. An implementation of the Bank of knowledge on medical diagnostics

An implementation of the described system can be accomplished with using the Multipurpose Bank of Knowledge (MBK) which has been developed in the IACP FEB RAS [19,20]. It is meant to be used for supporting life-cycle of compatible systems for information processing.

In this case, the software content of the Bank of knowledge on medical diagnostics includes:

- an editor for information of different level of generality (IDLG), which is a part of MBK (this editor is used for editing the models of ontologies and information structures organized according to these models in IDLG language [21,22]);
- an administrative subsystem which is also a part of MBK and which is used for managing user accounts and authorities;
- a solver for the medical diagnostics problem;
- a transformer of the base with diseases' descriptions.

The diagnostics algorithm is implemented as a parallel application in C++ which runs at on a multi-processor computing machine (MPCM) under OS Linux. The knowledge bank has to have an access to that resource. The knowledge bank server contains a low-functional server-part of the solver, implemented in Java, which interacts with the user, executes the diagnostics on the MPCM, receives the results and sends them to user.

As the diagnostics algorithm has to have a rapid access to knowledge and data in order to perform the diagnostics as fast as possible – the data from the diseases knowledge base has to be sent to and stored at

MPCM before any diagnostics. This procedure is performed by a knowledge-base transformation subsystem. When the diagnostics is executed – only the patient's case record is transmitted to MPCM.

When the diagnostics ends all results (rejected and approved hypothesizes about diagnosis with their reasons and explanations for values of observed symptoms) are transmitted from MPCM by server-part of the solver to the user.

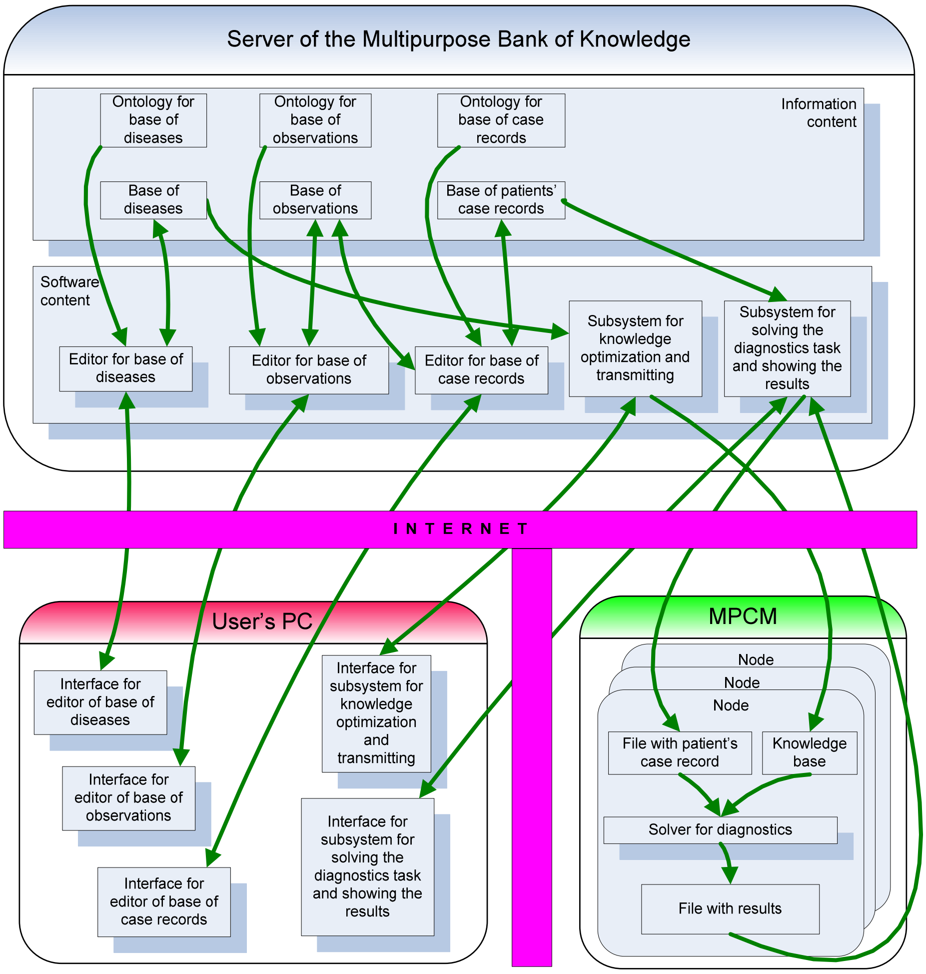

Figure 2. The architecture of the Bank of knowledge on medical diagnostics.

Figure 2 shows the general architecture of the Bank of knowledge on medical diagnostics in MBK/MPCM environment. As the scheme is too complex the administrative subsystem, its interface and user account database are not shown. Also the editors for ontology models (here – the editor of IDLG) are not shown as they are just used for initializing the correspondent bases and are not used afterwards. One should also understand that the standard (for MBK) editor for IDLG is used in the developed Bank for any base but it can be replaced with any other (more convenient) editor. In this case it has to be implemented and included into the software content.

Conclusion

A conception for the Bank of knowledge on medical diagnostics is presented in the paper. The information content of this Bank consists of three models of ontologies for the domain (model of observations, model of knowledge about diseases, model of patient's case record), knowledge of the domain (base of observations and base of knowledge about diseases) and data of reality (base of patients' case records). Which is more, the Bank contains the software content which consists of editors for models of ontologies, editors for data and knowledge, solver for the task of medical diagnostics, a software tool for transforming the knowledge base and administrative subsystem. The details of implementation of the Bank by means of the Multipurpose Bank of Knowledge (developed in IACP FEB RAS) are described.

Bibliography

1. Genkin A.A. About a sequential Bayes strategy and a mechanism for decision support in intellectual system OMIS. // Clinical laboratory diagnostics. - 1998. - №4.- pp. 42-49. (in Russian)
2. Detmer W.M., Shortliffe E.H. Using the Internet to Improve Knowledge Diffusion in Medicine. 1997. http://www-smi.stanford.edu/pubs/SMI_Reports/SMI-97-0658.pdf
3. Alekseev A.V. Using methods of informatics and computers for differential diagnostics of appendicitis and appendicular colic. http://arkadyal.chat.ru/rar/ddoaak_.rar (in Russian)
4. Burdaev V.P. Applied expert systems based on TECHEXP shell. 2002. http://uacm.kharkov.ua/eng/index.shtml?eexpert.htm (in Russian)
5. Expert system “Diagnostic of comatose states”. Admin LC. 1992. http://www.adminru.com/start_e.htm
6. Chabat F., Hansell D.M., Guang-Zhong Yang. Computerized Decision Support in Medical Imaging. 1997. http://www.doc.ic.ac.uk/~gzy/pub/chabat-decision-support.pdf
7. Gaskov A.P., Valivach M.N. The “Consultation” expert system for medical diagnostics. 2000. http://www.eksi.kz/consilium/librar/esmd_short.htm (in Russian)
8. Lhotska L., Vlcek T. Efficiency enhancement of rule-based expert systems. Proceedings of the 15th IEEE symposium on computer-based medical systems (CBMS 2002). pp.53-58.
9. Matsumoto T., Ueda Y., Kawaji Sh. A software system for giving clues of medical diagnosis to clinician. Proceedings of the 15th IEEE symposium on computer-based medical systems (CBMS 2002). pp. 65-58.
10. Filho M.M., Palombo C.R., Sabbatini R.M. TMJ Plus: A Knowledge Base and Expert System for Diagnosis and Therapeutics of the Temporomandibular Joint Disorders. 1994. http://www.epub.org.br/ojdom/vol03n03.htm
11. Sawar S.J. Diagnostic Decision Support System of POEMS. 1999. http://www.cbl.leeds.ac.uk/sawar/projects/poems/poems-in-detail.html
12. Zilber A.P., Shifman E.M., Pavlov A.G., Belousov S.E. The "Computer diagnostics of preeclampsia" Internet project. 1998. http://critical.onego.ru/critical/medlogic/ (in Russian)
13. Chernyakhovskaya M.Yu., Kleschev A.S., Moskalenko F.M. A metaontology for medical diagnostics of acute diseases. Part 1. An informal description and definitions of basic terms. INFOS 2008, Varna, Bulgaria. (to be published)
14. Chernyakhovskaya M.Yu., Kleschev A.S., Moskalenko F.M. A metaontology for medical diagnostics of acute diseases. Part 2. A formal description of cause-and-effect relations. INFOS 2008, Varna, Bulgaria. (to be published)
15. Chernyakhovskaya M.Yu., Kleschev A.S., Moskalenko F.M. A metaontology for medical diagnostics of acute diseases. Part 3. A formal description of the causes of signs’ values and of diseases. INFOS 2008. (to be published)
16. Moskalenko F.M. A task of medical diagnostics and an algorithm for its solving which can be paralleled // Informatics and control systems. – 2005. – № 2(10). – pp.52-63. (in Russian)
17. Moskalenko F.M. A parallel optimized algorithm for medical diagnostics // Informatics and control systems. – 2006. – № 1(11). – pp.87-98. (in Russian)
18. Moskalenko F.M. An experimental research of time-complexity for a parallel algorithm of medical diagnostics which is based on a real ontology // Informatics and control systems. – 2006. – № 2(12). – pp.42-53. (in Russian)
19. Orlov V.A., Kleschev A.S. Computer knowledge banks. Multipurpose bank of knowledge. // Information technologies – 2006. – №2. – pp.2-8. (in Russian)
20. Orlov V.A., Kleschev A.S. Computer knowledge banks. Requirements for the Multipurpose bank of knowledge. // Information technologies – 2006. – №4. – pp.21-28. (in Russian)
21. Orlov V.A., Kleschev A.S. Computer knowledge banks. A universal approach for solving the problem of information editing. // Information technologies – 2006. – №5. – pp.25-31. (in Russian)
22. Orlov V.A., Kleschev A.S. Computer knowledge banks. Модель процесса редактирования информационного наполнения. // Information technologies – 2006. – №7. – pp.11-16. (in Russian)

Information about authors

Chernyahovskaya M.Yu. - chernyah@iacp.dvo.ru

Kleschev A.S. - kleschev@iacp.dvo.ru

Moskalenko Ph.M. - philipmm@yahoo.com

Institute for automation and control processes, Far-eastern branch of Russian Academy of Sciences, Rusian Federation, Vladivostok, Radio st. - 5.