

Medical diagnostics with the classification task

Iaroslav Kardash

Center of Information Technologies, Lviv National University of Ivan Franko, Lviv, Ukraine

Email address:

ia.kardash@gmail.com

To cite this article:

Iaroslav Kardash. Medical Diagnostics with the Classification Task. *International Journal of Intelligent Information Systems*. Vol. 3, No. 3, 2014, pp. 19-26. doi: 10.11648/j.ijiis.20140303.11

Abstract: Modeling with reusable knowledge components on the basis of classification task for knowledge systems for medical diagnostics is considered. New problem solving method (method of feature set analysis) for classification task is proposed. It works faster than the known method of potential candidates pruning (from the *CommonKADS* methodology) for the domain with the number of attributes over 40.

Keywords: Knowledge Systems, Medical Diagnostics, Classification Task, Reusable Knowledge Components, Knowledge System for Determining Alloy Biocompatibility DentExp, Problem-Solving Method of Feature-Set Analysis

1. Introduction

Despite the global introduction of computer technologies in different branches of medicine, primarily, for audit and registration, visual interpretation of the examination and archived data, the need for computer program use in differential diagnostics is quite apparent. Half a century ago different methods of computational diagnostics became common (were introduced into common use): most salient out of the methods used was computer processing of the patient data for analysis of clinical signs (manifestations) of the disease.

Among computerized methods which are in common use at present are differential tables, discriminant analysis and the method that has become highly prospective recently, viz. indicator diagnostics. One of the latest methods of conducting medical diagnostics is the *Feature-set Analysis* method developed by the author in the process of solving the problem of dental alloy biocompatibility.

A significant obstacle on the way of information technologies development in medicine is irrational systematization of the clinical materials obtained over the last decades. Heuristic nature of medical knowledge (uncertainty), ambiguity, and presence of contradictions makes the process of detection, systematization and modelling of the clinical data rather complicated. Analytical work in this direction done by experts has a huge potential in the sense of improvement of the quality of diagnosis.

That is why the prospective and profitable task lies in

systematization and inordering (putting in order) of knowledge for medical diagnostics along with the development of new ones, particularly, reusable knowledge components. These are comparison and forming of relevant mappings that are of research interest: problem–task–PSM (problem solving method) and developing the necessary ontologies for common reuse and modelling and for knowledge systems construction as well.

Diagnostics is the core phase in the medical care because selection of the right treatment is based on it. But that knowledge is gained by medical professionals over decades. The line of the DentExp systems was developed for the problem of dental alloy biocompatibility in orthopedic dentistry, and reusable components and models developed during the work under these systems could be used for (medical) diagnostics within different problems and domains. Among the successful developments in the field of medical diagnostics the examples of classical systems MYCIN, INTERNIST and PUFF can be given, as well as a more up-to-day one – ESAGIL (esagil.org), on the terrains of Ukraine and former NIS — MODIS/МОДИС, CLASS/KJIACC [1], in the related domain with orthopedic dentistry the examples of the systems for diagnosis of the pain syndrome in temporomandibular joint dysfunction suggested by V. Onyshchenko, A. Mirza [2] and the diagnosis for dentitions by L. Leportskaya [3]. Original system to diagnose periodontal disease was presented by Brazilian researchers D. Deschamps, A. M. de Rocha Fernandes [4]. We also can provide the examples of a

computer program for audit and analysis of patient dentitions before and after treatment suggested by N. Dydik, Y. Zablotsky, M. Gzegotsky [5], computerized clinical DSS DART (J. Rudin [6]) for detection of oral pathologies and clinical DSS in dental implantology by A. Polášková et al. [7] in the Czech Republic.

This paper describes the model for medical diagnostics aimed at classification using reusable knowledge components (RKC) such as ontologies, problem-solving methods (PSMs) and generic tasks, which are used in the development of modern knowledge systems. We consider in it:

- 1) how medical diagnostics is modelled by the task of classification and what RKC in medical diagnostics with the task of classification are used for;
- 2) conceptual model of medical diagnostics, which became the basis for further work on knowledge system development;
- 3) what model of diagnostics is used for our modelling of medical diagnostics with the classification task;
- 4) generic model of RKC and PSM of *Feature-set Analysis*, which is the core reusable knowledge component and could be used further in solving different problems.

As the conclusion of the work on medical diagnostics for determining the biocompatibility of dental alloys we present the example of diagnostics in knowledge system Dent-Exp developed on the basis of RKC and the classification task.

2. Model Problem

As model problem in our case there was chosen the problem of determining dental alloy biocompatibility in orthopedic stomatology. This problem has high representativeness (the nature of diagnostics in the problem is interdisciplinary, and generalizations made on the basis of this problem could be applied for diagnostics in general) and, thus, provides wide room for conclusions.

The problem consists in selection of appropriate dental materials in prostheses, so that there was no intolerance between them and the patient. The problem includes 18 available (formulated by experts) diagnoses, among which 6 are disorders — primary diagnosis and the subject of examination, and 12 coexistent diseases form the subject of differential diagnostics with primary diagnoses-disorders.

For diagnosis determining values approximately 60 attributes (from the values of age and sex to the state of dentitions) are used. Statistical, direct preliminary and complete clinical diagnoses are consequently determined in the problem (for the data scope considered when a particular diagnosis is determined). For determining statistical diagnosis registration data are used (age, sex and the place of residence), for determining preliminary diagnosis the data of complaints, anamnesis and survey given in the form suitable for information processing are used. For determining clinical diagnosis 3-5 out of 14 allowed tests in

the problem domain also are used.

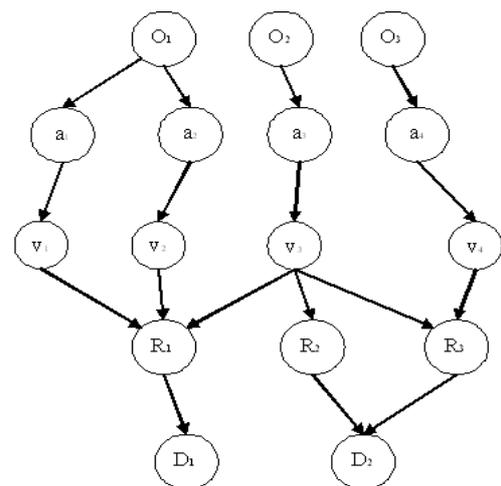
Researches were conducted on the basis of primary Ukrainian dental centers: in Kyiv, Lviv, Ternopil and Ivano-Frankivsk. Application of these methods and Dent-Exp knowledge system, presented in this paper, greatly helped to improve duration of examination and quality of decision-making, this being mentioned in section 5.

3. Medical Diagnostics with the Classification Task

Medical diagnostics can be carried out by different tasks, particularly by the task of classification.

Consider the problem of classification for this purpose. When medical diagnostics is modeled with classification the solution is determined by a certain correct class or classes (diseases) with a characteristic set of features of the object (the information that was identified or received about the object of investigation). Diseases are interpreted as independent from domain classes when independent from problem domain reusable knowledge components are used. Such way of solution is typical for professionals with a 5-10-year experience. Specifically for our task there was developed a solution method (described in detail below) which was called by the author as the *Feature-Set Analysis* method.

The input for such a problem is information about the object, while the output of it is the corresponding class or classes. The problem operates on the basis of the following terminology: *object*, a certain object for which the corresponding category has to be found, *class*, a certain class which consists of a group of objects with similar characteristics; *attribute*, the object's feature that can be traced or deduced; *feature*, a pair of attribute-value(s) characteristic of a certain object.



where O — a certain object (entity) of the domain, A — its attributes, V — attribute values, R — a certain relation in the domain (in this case they are modeled by the production rules) and — D the corresponding diagnosis

Figure 1. 5-component structure of medical diagnosis

For the problem of medical diagnostics the approach formulated by the author in 2002 has been applied in the work of our group at LNU at the stage of development of one of the early versions of ES DentExp [8] in orthopedic dentistry.

If the correct diagnosis d is from our set of diagnoses \mathcal{D} , then certain parameters-attributes can be traced out of the set of the ones allowed within our domain, each with the corresponding values. In the tasks of diagnostics at each séance (stage) any attribute (with some exceptions) acquires only one value (out of the set of possible ones), that is there is uniqueness and agreement of the data.

This model was applied during certain steps taken to pass the decision modeled by the classification tasks.

The problem operates the following sets: \mathcal{D} — possible diagnoses, \mathcal{A} — the set of all possible attributes, for each of them the set of their values is found, \mathcal{V} — all possible attribute values, \mathcal{F} — the data we operate at each séance of diagnostics within the set of all possible ones \mathcal{FA} — diagnoses-hypotheses (forecasted diagnoses), \mathcal{FH} , $\mathcal{FH}(\mathcal{D}_i)$ — all the forecasted findings and findings of certain diagnosis, \mathcal{FO} , $\mathcal{FO}(\mathcal{D}_i)$ — all the observed findings and the findings of certain diagnosis data, $\alpha(\mathcal{F})$ — all attributes of the séance and correspondingly forecasted $\alpha(\mathcal{FH})$ and observed $\alpha(\mathcal{FO})$ ones.

Diagnostics can be depicted in the form of a 5-component network (pentaplet), Fig.1 above.

3.1. Conceptual Model of Medical Diagnostics

Conceptual model was built in the process of analysis of domain and the problem. It is used with benefit on later stages of development. Figure 2 presents it, in the Table 1 below are only some of its components are given, which are core relations in our domain.

4. Reusable Knowledge Components and KSS Based on them

4.1. Modern KSS and Common KADS

Three core concepts in the field of knowledge engineering and modern knowledge systems absorbed almost a 40-year experience of development of KSS — *tasks, problem-solving methods (PSMs)* and *ontologies* of knowledge engineering and development of similar systems [12]. *Tasks* are corresponding to the generic strategy for solving problems *PSMs* are responsible for the decision-making process in the system, and *ontologies* make up the basis of static knowledge of the system, in particular knowledge bases.

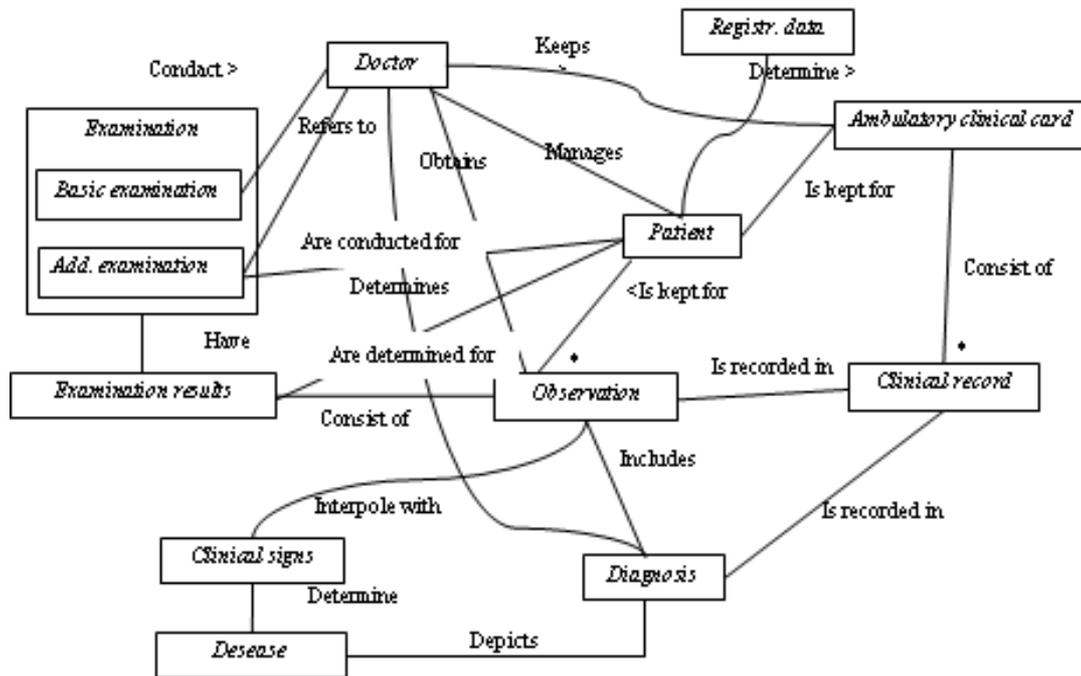


Figure 2. Conceptual model of medical diagnostics.

During the work on DentExp approaches to knowledge engineering have been analyzed in the research field (GT-TM [13,14], RLM [15], Protege [16], Expect [17], CommonKADS [4], D3 [18], MIKE [19], VITAL [20],

GDM [21]), analysis of PSMs (their characteristics and their other parameters have been studied) and ontologies (taking into account key projects, types and categories of ontologies) have been made as means of KSS development.

Table 1. Primary relations in the domain of medical diagnostics

Relation	Description
Doctor Conducts Examination	At the first visit of a patient to a dentist, the doctor conducts primary examination, where he/she determines the complaints, anamnesis and survey and during the subsequent visit, the dentist determines obtained additional tests results
Observation Consists of Examination Results	Examination results are coded appropriately to the values in their examination sections, the overall clinical observation consists of the all examination results
Observation is compared with Clinical Signs	Findings obtained as a result of examination are compared with clinical signs of each (out of 18th) diseases
Clinical Signs Describe Disease	Lately, diseases which has higher values of confidence are considered
Disease Defines Diagnosis	Selected diseases are described in the terms of IDC-10
Doctor Determines Diagnosis	Dentist has a possibility to select one primary (from several – till three) the most probable diseases, if the system could not make certain decision or if the dentist wants to provide a specific diagnosis of the patient

Under the methodology of knowledge engineering CommonKADS (which was the main methodology for our researches), the knowledge model of application is based on three main levels: 1) the level of knowledge of the domain (concepts, knowledge roles in domain, domain schemes, knowledge bases); 2) the inference level (a detailed process of decision-making expressed step by step: elementary logical inferences, transfer functions, static and dynamic knowledge roles); 3) the level of the task method: goals, tasks, decomposition into (sub)methods, general representation using elementary logical inferences and transfer functions.

The generic model of reusable knowledge components based on the domain knowledge, logical inference and task knowledge has also been formulated during the work on the next versions of DentExp [9] knowledge system. Reusable knowledge components have been developed for the task of medical diagnostics: domain knowledge and ontology, as well as logical inferences and PSM of *Feature-set Analysis* [10] for the classification task have been implemented. The reusable knowledge components have been filled in by the information specific for the domain of dental alloys biocompatibility. The components have been provided in the CML language of the *CommonKADS* methodology.

4.2. Generic Model of Reusable Knowledge Components

The level of domain knowledge is presented in the form of:

$$DK = \langle \langle C_{\leq C}, R_{\leq R} \rangle, RT, DS, KB \rangle \quad (1)$$

where $C_{\leq C} = \{c_i\}$ — a separate set of concepts of the domain with the corresponding hierarchy (partial order) in it; $R_{\leq R} = \{r_i\}$ — a separate set of relations with the corresponding hierarchy in it (a partial order). And if for two relations $R_1 \leq_R R_2$, that means that relation R_1 is of no more arity than R_2 :

$|\sigma(R_1)| \leq |\sigma(R_2)|$, and concepts and arity of relations satisfy the following condition: $\pi_i(\sigma(R_1)) \leq \pi_i(\sigma(R_2))$, i.e. that the domain of definition of the relation R_1 and its arity as well as those subrelations included into R_1 are covered with the relation R_2 ;

$RT = \{rt_i\}$ — the set of rules defining certain expressions between the concepts in $C_{\leq C}$; $KB = \{\{ds_i\}, \{kb_i\}\}$ — a set of

schemes of the domain ds_i and corresponding to them knowledge bases kb_i filled with actual examples, in particular $ran(ds_i) = kb_i$.

Knowledge of the level of logical inference consists of the following components:

$$IK = \langle I, TF, KR, DS', M \rangle \quad (2)$$

where $I = \{ei_1, \dots, ei_n\}$ — a separate set of elementary logical inferences: $ei_i(in-KR_b, out-KR_m)$ applied to solve the task with the corresponding input and output subsets of knowledge roles; $TF = \{tf_1, \dots, tf_k\}$ — a separate set of transfer functions in our task $tf_j(in-KR_p, out-KR_q)$ with the corresponding subsets of knowledge roles; $KR = \langle DKR, SKR \rangle$ — sets corresponding to dynamic and static knowledge roles applied in our task; $DS' \subset DS$ — a subset of domain schemes $\{ds_r\}$, applied in our task; $M = \{tdm_1, \dots, tdm_s\}$ — a set of “task–domain” mappings of relations for connecting our task knowledge roles with the schemes of our task: $tdm_s: KR \rightarrow DS'$.

Task knowledge are given in the following way:

$$TK = \langle T, TM, I, TF, T_{\leq T} \rangle \quad (3)$$

where T — the task for which the model is built; $TM = \{tm_1, \dots, tm_v\}$ — a set of methods to solve tasks and subtasks; $I = \{ei_1, \dots, ei_n\}$ — a set of the above elementary logical inferences of our task; $TF = \{tf_1, \dots, tf_k\}$ — a set of our transfer functions; $T_{\leq T}$ — a certain hierarchy or taxonomy imposed on the elements from the sets TM, I, TF .

In conclusion the model of modern knowledge systems can be presented in the form of the structure as follows:

$$KS = \langle \langle P, g; T, O_{P, g}, O_T \rangle, \langle PSM, T \rangle, IK; O_{PSM, T}, O_{IK} \rangle, \langle DK, O_D \rangle, AK \rangle \quad (4)$$

The first compound component determines the problem and goals of the application, the task with which the problem is solved and the corresponding ontologies; the second compound element defines the knowledge of PSMs and the knowledge of the task, knowledge of logical inferences and the corresponding ontologies; the third compound element determines the knowledge of the domain and its ontologies; the last component defines specific architectural decisions in the system design.

5. Medical Diagnostics and Reusable Knowledge Components in our Problem in Detailed View

5.1. Domain Knowledge

Further on there are the key components of ontology in the domain developed and represented in Protege (protege.stanford.edu):

Patient, registration data, complaints, anamnesis, survey, additional tests (examinations),

Examination — the main *relation* in the domain, it consists of the basic examination (complaints, anamnesis and observation) and some certain (mainly three to five)

additional examination/tests out of 14 possible ones,

Examination results: this is the data obtained as the result of examination (tests),

Observation: this is the findings doctors observed with the patient. Actually, these are the results of examination prepared for being compared with clinical signs,

Clinical symptoms: these are the signs of each of 18 (for our problem) types of diseases (disorders or coexisting diseases). They are divided into clinical symptoms under the main examination (complaints, anamnesis, observation) and additional examinations/tests,

Differential diagnostics: this is the process of outlining the most credible or primary diagnosis out of the several possible ones,

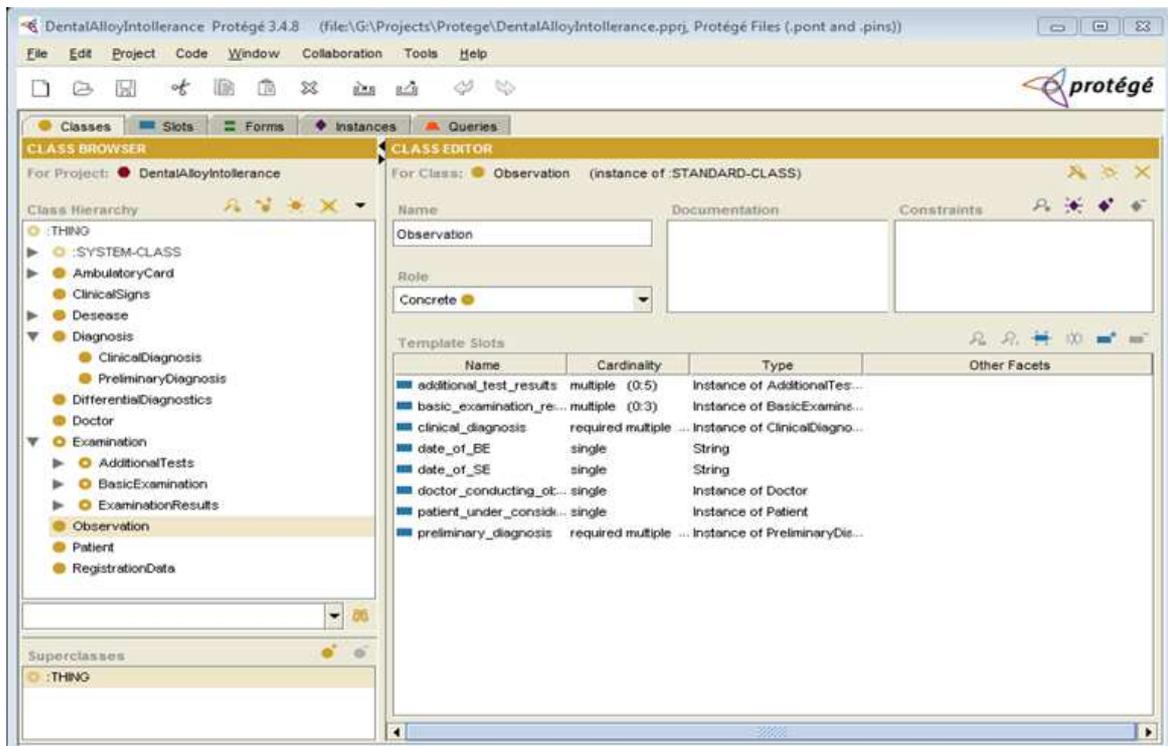


Figure 3. Domain ontology in Protege.

Disease: formulation of the disease in the terms of IDC-10. In our domain, 18 types of diagnosis should be distinguished, six of which are disorders (diseased conditions relating to usage of dental prostheses), while 12 co-existing diseases — complications (two for each disorder) constitute the subject of differential diagnostics with the main disease-disorder. The concomitant diseases are similar in their clinical symptoms of the basic examination to the disorder but are caused by other reasons,

Diagnosis which is divided into statistical (inference from registration data), (direct) preliminary (based on the results of the basic examination) and (complete) clinical one (diagnosis after differential diagnostics on the basis of the data of several, mainly three to five additional tests, out of 14 possible ones),

Ambulatory card: ambulatory card, or disease history consists of clinical records containing the data on the patients' condition,

Clinical record: the data about the patient based on the doctor's observations, his prescriptions and information on the treatment applied as well as its results in the ambulatory card.

5.2. Problem-Solving Method and Task Knowledge

Here we have the representation of the process of decision-making for medical diagnostics based on the classification task in CML-notation, used in CommonKADS [11]. This problem-solving method was called *Feature-Set Analysis* method.

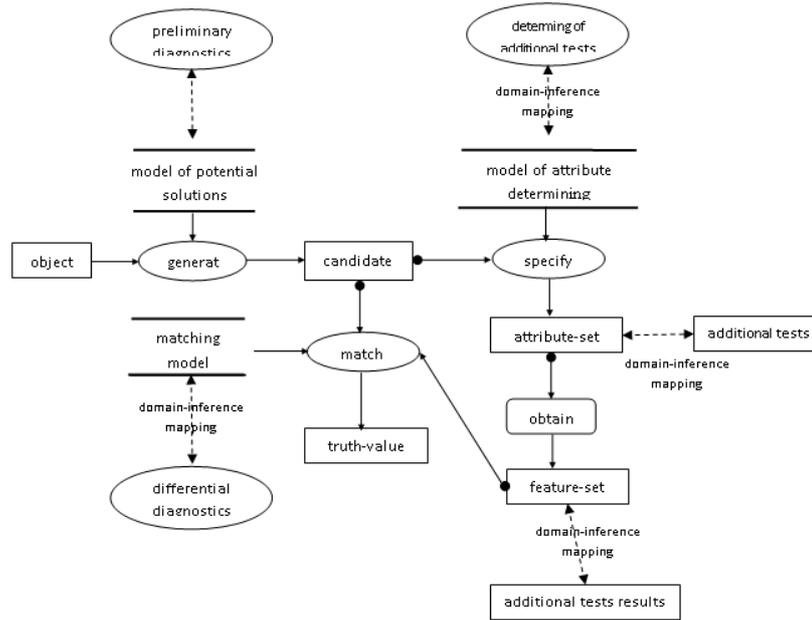


Figure 4. Feature-set analysis method with UML-like notation in Common KADS

5.3. Method Description

Search for solutions from the data is applied in our task of medical diagnostics that uses the classification task with application of the method of *Feature-Set Analysis*. That is at the first reasoning step GENERATE out of the input information we generate the list of possible candidate solutions. Candidate diagnoses, the quality measure of which exceeds 0.18, are obtained there.

At the next reasoning step SPECIFY for each element of the formed set of potential solutions we determine the set of attributes (corresponding with the appropriate potential classes- solutions) the value of which still has to be found out or specified. With this in view more information is gathered about the object (objects) of classification taking it additionally from the outside. In our domain additional tests are determined, for which the patient is referred to and the data of which we obtain within some time (transfer function OBTAIN). On the basis of the information obtained we apply the model of diagnostics again refuting the potential solutions which do not agree with the model, and accept the one in which the quality measure also exceeds a certain threshold set by the expert (reasoning step MATCH). In our case it made up 0.4.

As a result we get the corresponding set of solutions (which can be empty if the information obtained does not agree with any class description).

Now it is time to show how quality measures of belonging to each separate class are calculated.

We apply the approach which we elaborated in 2001-2002 for very first versions of DentExp and continue to calculate in later ones. We take into consideration the number of correct values of features for particular diagnosis:

$$\rho(\mathcal{D}_i) = |\mathcal{FO}(\mathcal{D}_i) \cup \mathcal{FH}(\mathcal{D}_i)| / |\mathcal{FH}(\mathcal{D}_i)|, \quad (5)$$

where $\mathcal{FO}(\mathcal{D}_i)$, $\mathcal{FH}(\mathcal{D}_i)$ are aforementioned characteristics in the section 3.

At every step of determining diagnosis in our PSM we calculate such a measure of confidence for particular diagnoses.

In the following section we illustrate diagnostics in DentExp with the last by one screen for displaying appropriate complete diagnoses for particular case.

6. Results: Diagnostics in DentExp. Discussion

In Fig. 5 above we provide the screen for the last stage of differential diagnostics in DentExp (in our case it is galvano-inflammation disorder).

Due to the introduction the system in the work of dental centers the values of certainty of concluded diagnoses were improved. Thus, after conducting this diagnostics for about 2,000 of patients in 4 dental centers, we can provide such results in average: the result of medico-statistical diagnosis was improved by 13%, the preliminary — by 17% and clinical — by 21%, while the usage of the system in the diagnostic process.

Also the duration of examinations was reduced by 20-25%.

As the result of improved quality of diagnosis, the costs spent on treatment have also been reduced too. For patients, who obtained appropriate diagnostics and treatment, the cost of treatment was reduced by 10-15%.

Time spent on the design and development of KS with the application of RKC is significantly reduced. According to

- [9] I. Kardash. "Java-version of DentExp intelligent medical system with binary logic application", in Materials of XIII International conference on automatic control (Automatics-2006), Vinnytsia, September 25-28, 2006, pp.392-396. / Кардаш Я.А. Java-версія інтелектуальної медичної системи DentExp з бінарною логікою. //Матеріали XIII Міжнародної конференції з автоматичного управління (Автоматика-2006), м. Вінниця, 25-28 вересня 2006 р., с.392-396.
- [10] I. Kardash. "Conceptual modelling and developing of reusable knowledge components for intelligent systems in medical diagnostics on the basis of CommonKADS approach", in Collection of Scientific Papers *Information Refinement and Search*, 25(101), 2006, pp.26-33 / Кардаш Я.А. Концептуальне моделювання та створення повторно вживаних компонент знань для інтелектуальних систем у медичній діагностиці на базі методології CommonKADS. Міжвідомчий збірник наукових праць „Відбір і обробка інформації”, випуск 25(101), 2006 р., с.26-33.
- [11] G. Schreiber et al. Knowledge Engineering and Management: The CommonKADS Methodology. Cambridge: MIT Press, 2000.
- [12] I. Kardash. Expert systems development for medical diagnostics on the basis of reusable knowledge components, unpublished. – Manuscript. An undefended yet work for a PhD sciences degree by speciality 05.13.06 – „Information technologies”. – Lviv Polytechnic National University, Lviv, Ukraine.
- [13] C. Duursma. Role-Limiting Methods withing CommonKADS Library, 1996, electronic publication http://arti.vub.ac.be/previous_projects/kads/CH/CH.html
- [14] W. Clancey. "Heuristic Classification". TR KSL-85-5, Stanford University, Department of Computer Science, 1985.