

A Case-Based Framework for Leveraging NutriGenomics Knowledge and Personalized Nutrition Counseling*

Daniel D. Wu¹, Rosina Weber¹, and Fredric D. Abramson²

¹ College of Information Science and Technology, Drexel University, Philadelphia, PA, USA

{Duanqing.Wu, Rosina.Weber}@drexel.edu

² AlphaGenics, Inc., 9700 Great Seneca Highway, Rockville, MD, USA

Abstract. NutriGenomics is the bioscience that links the way nutrients and other dietary components shape genetic activity. It builds on the success of Human Genome Project by applying systems biology methods to explain how the molecular components of food, supplements and pharmaceuticals dynamically influence and shape the activity of genomic subsystems, which in turn define how a person can stay healthy or become ill. Applying NutriGenomics knowledge is done through Directive Genomics, which develops purposeful dietary strategies that influence gene expression at the individual level with the goal of better genetic function and health. This paper proposes a case-based framework for leveraging nutrigenomics knowledge and Directive Genomics applications. The unique features of the proposed system include a self-maintained distributed case base structure and a CBR-based nutrition counseling module that can learn, adapt, and maintain its case base via the integrated distributed case bases as well as external resources.

1 Introduction

Significant advances in computing and laboratory technology enabled the Human Genome Project to successfully characterize the human genome. The resulting flood of data pertaining to gene identification and function creates new challenges to understand how the genome functions. The human genome has at least 30,000 different genes, which function as a robust and coherent system and carry each person from birth through the aging process and ultimately death. When viewed as a system, the scientific goal is to understand the relationships between genotypes and phenotypes, and to determine the interaction and regulation of the genome as a network.

An emerging research area with broad applications is NutriGenomics, which uses systems biology methods to deepen our understanding of how the molecular nutrition environment influences what the genome does [1, 2].

Traditional nutrition research uses epidemiological, clinical and physiological studies to identify how deficiencies in certain chemicals translate into compromised health. It builds on these findings to explore how dietary choices in general help explain differences in health from one population to another. Our knowledge of these re-

***To be presented in the workshop in CBR in Health Sciences, part of the workshop program of the 2004 European Conference in Case-based reasoning.**

lations, however, has been achieved from a traditional biomedical perspective, focusing more on group or population differences rather than individualized outcomes.

With the emergence of NutriGenomics research, the concept of personalized medicine can be realized. Molecular nutrients can alter molecular and cellular processes involved with gene expression, which in turn influences other cellular activities.

It is well known that people react differently to their environment. With pharmaceuticals, for example, which are well characterized molecules, one person may obtain a benefit while another may not, and a third may have an adverse reaction. And most, if not all, of this variation is directly connected to a person's genetic makeup, and what those genes are doing at the time the drug is taken. Similarly, each person's genetic variation appears to influence how that person responds to his or her diet.

NutriGenomics knowledge enables people to understand how their dietary lifestyle influences their health. This translates to greater demand for personalized dietary counseling with the goal of maintaining or even optimizing a person's health, and to delay or prevent the onset of diseases.

Human behaviors have been proven to be connected to individuals' gene expression. There is a potential application of NutriGenomics research and Directive Genomics in large organizations such as the military to maximize health benefits given scarcity of resources and control potential tendencies e.g. extreme violence.

Case-based reasoning (CBR) is a reasoning methodology that resembles the natural reasoning model of human beings when solving new problems by using previous experiences of similar problems [3]. It is an established and powerful methodology and has been used for developing many applications in a variety of domains [4].

In this paper, we propose a framework that uses CBR methodology to leverage NutriGenomics knowledge and provide NutriGenomics-based personalized nutrition counseling.

In Section 2, we present an introduction to NutriGenomics, followed by an introduction to case-based reasoning. In Section 4, we detail our proposed framework. Our plan for future work is presented in Section 5.

2 NutriGenomics

NutriGenomics expands the concept of nutrition by examining how the molecular components of food, supplements and pharmaceuticals influence and shape gene expression, and in turn help define how a person can stay healthy or become ill. NutriGenomics transforms the general dietary rules developed through years of epidemiologic observations into specific, individualized functions. It takes the nature versus nurture debate and makes it unified and precise, because each influences the other. One of its ambitious goals is to develop effective dietary intervention strategies based on a person's gene expression patterns, the "Expressitype", which target genes that can be successfully influenced by one or more molecular nutrients.

NutriGenomics focuses on that part of the chemical environment that comes from outside the body. While the main source of these chemicals is food, there are other sources such as dietary supplements, pharmaceuticals, cosmetics, water and air. There are four key propositions: (1) the genome functions as a coherent, robust system of

genes comprised of many sub-systems, each subsystem controlling specific life activities, (2) each sub-system is comprised of many genes that act in coordination, (3) the activity of the genes is influenced by the cell's chemical and physical environment, and (4) part of the chemical environment comes directly from what a person brings into the body and the remainder is manufactured by the cells.

From a NutriGenomics perspective, food is enormously complex. Over 350 different chemicals are documented in the ordinary garden tomato, for example. The NutriGenomics inquiry asks which of these chemicals are present in quantities that can influence genes, which genes are influenced, and in what way they are influenced.

NutriGenomics looks beyond one gene at a time because most of the major health conditions involve tens and hundreds of genes. Virtually every day, researchers identify new genes that contribute to health and disease. Conditions like breast cancer (109 genes) and asthma (27 genes) and diabetes (114 genes) highlight how broad these genetic factors will become.

At the same time, the new biotechnology methods are pinpointing how molecular nutrients shape genomic activity. For examples, Vitamin A changes activity in over 500 genes; calcium in over 145, zinc over 60. Cholesterol, which is made by the body, influences over 30 genes, which shows that molecular nutrients which might regulate cholesterol production are likely, in turn, to have important second order effects.

While NutriGenomics is new, research in related fields as nutrition science, molecular biology, genetics, and medicine, is generating an enormous amount of data containing valuable information that can be employed from a NutriGenomics perspective. The challenge is how to extract knowledge from these dispersed resources.

The development of technologies such as the high-density microarray (genes on a chip) makes it possible to examine tens of thousands of genes on an individual basis fairly quickly and inexpensively. This provides a platform to provide personalized dietary counseling based on both the body of NutriGenomics knowledge and each person's individual genome. Applying NutriGenomics knowledge is done through Directive Genomics that develops purposeful dietary strategies that influence the gene expression at the individual level to improve gene function and health.

3 Case-Based Reasoning

Our proposed framework uses a CBR module as its core reasoning methodology. The underlying assumptions of CBR are that: 1) similar problems have similar solutions; and 2) similar problems do recur [13]. In a CBR system, knowledge or past experiences are stored as cases in a case base or case bases. A case consists of the problem description, the solution to the problem, and its outcome. Given a new problem, a problem-solving CBR system retrieves similar cases from case bases and adapts the solutions of the retrieved similar cases to solve the new problem [4].

There are four basic steps in a typical CBR process [6]: retrieve, reuse, revise, and retain. There is an additional review step for applications including iterative learning [3], as illustrated in Figure 1.

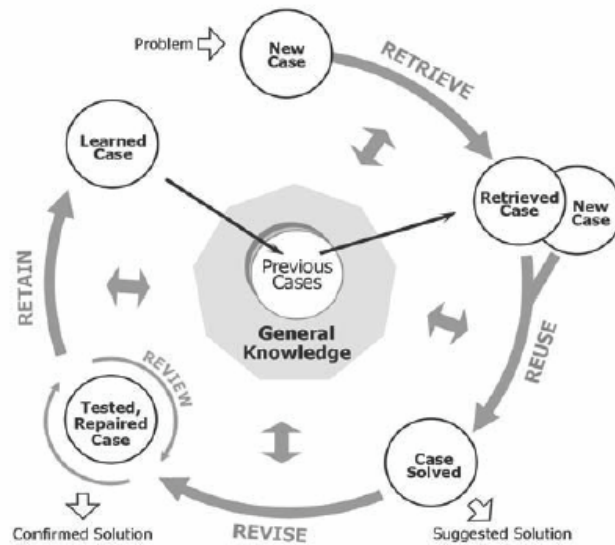


Fig. 1. The CBR Cycle

CBR is a reasoning methodology that is suitable for solving complex and dynamic problems. Given the use of experiences, CBR is particularly suitable for ill-defined areas in which there is no theory model and knowledge is incomplete.

CBR has been applied to a host of domains, including health related domains such as nutrition [7] and medicine [8, 24]. Knowledge in health related domains is complex and constantly evolving as the knowledge is incomplete. The practice in these domains also involves substantial amount of subjective knowledge and personal experience of the practitioners. A number of CBR systems have been successfully developed for prediction of medical studies [9, 10], nutritional consulting [7, 11], diagnosis of diseases [12], post-transplantation follow-up [14], care of Alzheimer patients [15], exercise planning for cardiac and pulmonary disease patients [16], and geriatric care [17]. CBR has also been applied to information retrieval and analysis of biological research data [18]. A CBR-based framework is proposed to provide personalized health maintenance information [19].

NutriGenomics represents another domain in which CBR can contribute. Firstly, NutriGenomics is a new field where our knowledge is still limited and constantly evolving. Secondly, the task of providing NutriGenomics-based personalized nutritional counseling is essentially a combination of diagnosis and planning. There is no well-defined theory model to guide this process. Thirdly, genes work in systems and subsystems that are remarkably coherent and robust, and have strong commonalities from one person to another. If two persons have very similar genotypes and Expression, they may as well have similar responses to dietary input.

4 Case-based Framework for Leveraging NutriGenomics Knowledge and Personalized Nutrition Counseling

The proposed framework is shown in Figure 2. It comprises 1) a knowledge management module that extracts and organizes knowledge from diverse resources; 2) a Case Builder module enabling automatic creation of cases from distributed case bases; and 3) a personalized nutrition counseling module performing CBR to provide NutriGenomics-based personalized nutrition counseling. The following sections discuss each of these modules.

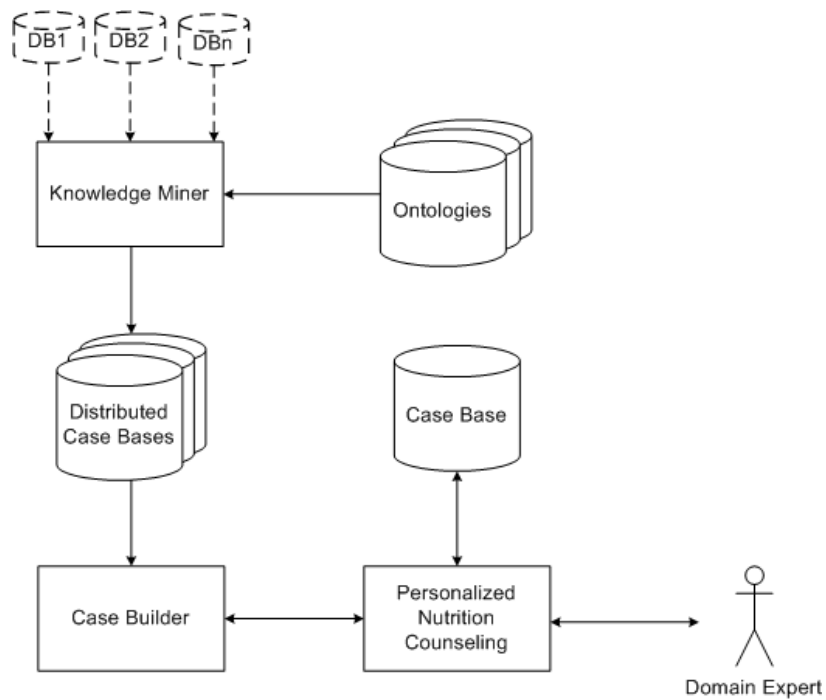


Fig. 2. The framework of CBR-based system for leveraging NutriGenomics knowledge and personalized nutrition counseling.

4.1 Knowledge Management (KM) Module

The knowledge management module includes a number of ontologies, a Knowledge Miner, a distributed case base structure, and a variety of external data sources.

4.1.1 Knowledge Miner

Knowledge Miner is the core player of the KM module. It is responsible for extracting NutriGenomics knowledge using data mining tools from the external heterogeneous data sources (DB_1, DB_2, \dots, DB_n). There are many publicly available resources that

can be used such as sequenced genes (e.g., EMBL), proteins (e.g., SWISS-PROT, PIR, BREMDA), transcription factors (e.g., TRANSFAC), biochemical reactions (e.g., KEGG), signal transduction pathways (e.g., TRANSPATH, GeneNet), and published literatures (e.g., PubMed). The extracted knowledge is organized and stored in a number of distributed case bases, each of which contains a subset of NutriGenomics knowledge (such as nutrients, genes, health conditions, diseases, and their relationships as revealed in published literatures or as results from data mining). The miner consists of a number of agents which run at specified frequencies to keep the knowledge bases updated.

An example of knowledge that may result from the data mining is the relationship between gene expression and molecular nutrients, as shown in Fig. 3, illustrating a first order interaction between Vitamin A and some of the genes whose transcription it affects.

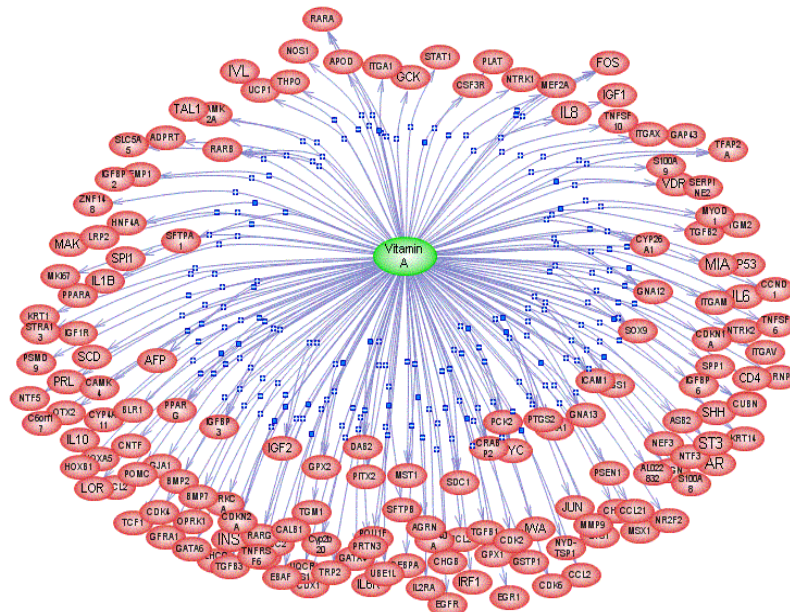


Fig. 3. Genes whose expression is affected by vitamin A¹.

4.1.2 Ontologies

The main purpose of using ontologies is to facilitate the representation of the domain knowledge in NutriGenomics. There are two types of ontologies used in our system as described below.

1.) Base ontology. This is the top level ontology containing basic definitions for terms that are used system-wide. It defines the semantics, the global constraints, and

¹ Picture was generated using PathWay Assist software with data from the ResNet database by Ariadne Genomics, Rockville, Maryland.

the architecture of our knowledge representation scheme for the entire NutriGenomics domain. It is designed to help resolve conceptual conflicts and maintain the homogeneous view and conceptual consistency of knowledge within the system.

2.) Individual sub-domain ontologies. An individual sub-domain ontology stores ontological information of a particular subset of knowledge within the NutriGenomics domain. There are three sub-domain ontologies defined in the system: a) nutrition ontology which provides ontological information of nutrition, including definitions and terms of all the nutrients; b) genomics ontology which contains information related to genomics, such as the definitions and terms used to describe genes and their functions; c) diseases ontology which stores information pertaining to specific diseases. While base ontology provides the overall structure and common vocabulary for the whole system, individual sub-domain ontologies define the structure and vocabulary for particular sub-domains.

4.1.3 Distributed Case Base Structure

The distributed case bases are the repository representing the collection of NutriGenomics knowledge. We define three distributed case bases to store knowledge as defined by the respective individual sub-domain ontologies, i.e. nutrition case base (NCB), genomics case base (GKB), and diseases case base (DKB).

In addition to the function as a knowledge repository, the distributed case bases also serve as resources that can be used by the Case Builder to automatically create new cases for personalized nutrition counseling.

4.2 Case Builder

The effectiveness of a CBR system depends largely on the qualitative and quantitative richness of its stored cases which are the knowledge repository of past experiences. Such knowledge repository typically is obtained by manually creating cases representing the experiences of domain experts according to pre-defined case format. In our proposed system, we keep such process to allow domain experts to add and modify the cases. In addition, to deal with the problem of lacking cases in case base and the efficiency issue of manual input, we employ a strategy, implemented in this system as Case Builder, to automatically create new cases.

Given a target case, if there is no case in case base or CBR fails to retrieve similar cases or CBR fails to produce a satisfactory solution, the Case Builder can be called upon to create a new solution based on the problem description of the target case.

The Case Builder constructs cases using the above discussed distributed case bases. This model is similar to the multi-case bases reasoning (MCBR) architecture proposed in [16]. The difference is that in our system, we isolate the task of drawing on external case bases from our main CBR module mainly because of the purposes of the Case Builder. The Case Builder can be used: 1) to initially populate the case base; and 2) to facilitate the CBR adaptation during regular system operation.

The Case Builder also uses rule-based logical inferences to create a new case. One example of such inferences may look like this:

IF molecular nutrient *A* can reduce the risk of disease *B* based on a published fact AND gent *C* is involved in disease *B* based on experimental facts, THEN treatment with nutrient *A* at concentration level *X* may decrease the expression of gene *X*, thus reducing the risk of disease *B*.

Based on this rule, a nutritional guidance for a person with higher than normal expression of gene *C* may include the nutrient *A*.

4.3 Personalized Nutrition Counseling

This is a standard CBR module with all cases stored in a case base. Its purpose is to provide personalized nutritional counseling based on personal information such as genotype, Expressitype, current and historical health conditions, and dietary lifestyle.

As mentioned above, traditional life science narrows things down to one gene, one cell, or one virus to find a single solution such as one drug or one vaccine, which is then applied to masses of people without distinction. NutriGenomics and Directive Genomics, on the contrary, attempt to configure a dietary strategy that works best for an individual given his or her genotype and Expressitype and eventually build upward from individualization to find well targeted population solutions. The goal of the CBR module is thus to deliver such a strategy.

4.3.1 Case Retrieval

Given a target case, the retrieval process retrieves similar cases using nearest-neighbor algorithm. If there is no case in case base or the retrieval process results in no similar cases, CBR module will call Case Builder to create a new case and store it in the case base with or without the intervention of a domain expert.

4.3.2 Case Adaptation

The solutions of the retrieved similar cases can be used directly or through adaptation to fit the problem of the target case. If the system cannot generate a satisfactory solution to a target problem, additional knowledge must be provided to the system. Two actions can be taken. The first choice is to allow domain experts to manually adapt the solutions from retrieved similar case(s). The second choice is to call the Case Builder to build a new case solution to the target problem.

5 Future Work

In this paper we propose a case-based framework that is applied to the domain of NutriGenomics and Directive Genomics applications. In this proposal, NutriGenomics knowledge is extracted from a variety of resources using data mining tools. The extracted knowledge is organized and stored in three knowledge bases facilitated by ontologies. The distributed case bases structure provides a mechanism in which a

Case Builder may automatically create new cases, when needed, for the personalized nutrition counseling, based on NutriGenomics knowledge represented by cases stored in its case base.

Much work remains ahead before the proposed framework becomes a fully functional system. First of all, we are working on implementing the knowledge management module, including the ontologies and the distributed case bases. Case mining proposed in [23] provides an interesting method that can be helpful in this regard. Research in [20, 21, 22] shed lights when dealing with distributed case bases and hierarchical case-based reasoning. Our future work will also include implementing the nutrition counseling module and building the interfaces to allow interactions between the system and domain experts and other end users.

References

1. van Ommen, B. and Stierum, R.: NutriGenomics: Exploiting Systems Biology in the Nutrition and Health Arena. *Curr. Opin. Biotechnol.* 13: 517-521 (2002)
2. Muller, M. and Kersten, S.: NutriGenomics: Goals and Strategies. *Nature Rev. Genet.*, 4: 315-322 (2003)
3. Aha, D. W.: The omnipresence of case-based reasoning in science and application. *Knowledge-Based Systems*, 11(5-6): 261-273 (1998)
4. Watson, I.: *Applying Case-Based Reasoning: Techniques for Enterprise Systems*. Morgan Kaufmann Publishers, Inc., San Francisco, California (1997)
5. van Ommen, B.: NutriGenomics: Exploiting Systems Biology in the Nutrition and Health Arenas. *Nutrition*, 20: 4-8 (2004)
6. Aamodt, A. and Plaza, E.: Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *Artificial Intelligence Communications*, 7 (1): 39-59 (1994)
7. Camargo, K.; Thé, M. A.; Weber, R.; Martins, A.; Barcia, R.: Designing Nutritional Programs with Case-Based Reasoning. In *Proceedings of the Sixth German Workshop on Case-Based Reasoning-Foundations, Systems, and Applications*, 141-147 (1998)
8. Alexanderini, F., Krechel, D., Maximini, K., von Wangenheim, A.: Integrating CBR into the Health Care Organization. In *Proceedings of the 16th IEEE Symposium on Computer-Based Medical Systems (CBMS 03)* (2003)
9. Schmidt, R., Montani, S., Bellazzi, R., Portinale, L., Gierl, L.: Case-Based Reasoning for Medical Knowledge-Based Systems. *Int. J. Med. Info.* 64: 355-367 (2001)
10. Seitz, A., Uhrmacher, A. M., Damm, D.: Case-Based Prediction in Experimental Medical Studies. *Artificial Intelligence in Medicine*, 15: 255-273 (1999)
11. Khan, A. S. and Hoffmann, A.: Building a Case-Based Diet Recommendation System without a Knowledge Engineer. *Artificial Intelligence in Medicine*, 27: 155-179 (2003)
12. Chang, C.-L., Cheng, B.-W., Su, J.-L.: Using Case-Based Reasoning to Establish a Continuing Care Information System of Discharge Planning. *Expert Systems with Applications*, 26: 601-613 (2004)
13. Leake, D. B.: CBR in Context: The Present and Future. In Leake, D., (ed.): *Case-Based Reasoning: Experiences, Lessons, and Future Directions*. Menlo Park: AAAI Press/MIT Press 1-35 (1996)
14. Bichindaritz I, Siadak MF, Jocom J, Moimpour C, Kansu E, Donaldson G, Bush N, Chapko M, Bradshaw JM, Sullivan KM. CARE-PARTNER: a computerized knowledge-support system for stem-cell post-transplant long-term follow-up on the World-Wide-Web. *Proc AMIA Symp.* 386-90 (1998)

15. Marling, C. and Whitehouse, P.: Case-Based Reasoning in the Care of Alzheimer's Disease Patients. In Proceedings of ICCBR-01, 702-715 (2001)
16. Evans-Romaine, K. and Marling, C.: Prescribing Exercise Regimens for Cardiac and Pulmonary Disease Patients with CBR. Workshop on CBR in the Health Sciences, ICCBR-03, (2003)
17. Marling, C. R., Whitehouse, P. J., Fioritto, P. A., Bendis, J. E.: Knowledge Sharing and Case-Based Reasoning in Geriatric Care. In AAAI-99 KM/CBR Workshop (1999)
18. Hofstadt, R. and Topel, T.: Case-Based Support of Information Retrieval and Analysis of Molecular Data. In Proceedings of the 15th IEEE Symposium on Computer-Based Medical Systems (CBMS 2002)
19. Abidi, S. S. R.: A Case Base Reasoning Framework to Author Personalized Health Maintenance Information. In Proceedings of the 15th IEEE Symposium on Computer-Based Medical Systems (CBMS 2002)
20. Leake, D. B. and Sooriamurthi R.: Managing Multiple Case Bases: Dimensions and Issues. In Proceedings of the 15th International Florida Artificial Intelligence Research Society (FLAIRS) Conference. AAAI Press, Menlo Park, 106-110 (2002)
21. Leake, D. and Wilson, D.: When Experience Is Wrong: Examining CBR for Changing Tasks and Environments. In Proceedings of the Third International Conference on Case-Based Reasoning, Springer-Verlag: Berlin 218-232 (1999)
22. Smyth, B., Keane, M. T., Cunningham, P.: Hierarchical Case-Based Reasoning Integrating Case-Based and Decompositional Problem-Solving Techniques for Plant-Control Software Design. IEEE Trans. Knowl. Data Eng. 13(5): 793-812 (2001)
23. Yang, Q. and Cheng, H: Case Mining from Large Databases. In Proceedings of the 5th International Conference on Case based Reasoning, Springer-Verlag: Berlin. 691-702 (2003)
24. Schmidt, R., Montani, S., Bellazzi, R., Portinale, L., Gierl, L.: Case-Based Reasoning for Medical Knowledge-based Systems. International Journal of Medical Informatics 64 (2-3): 355-367 (2001)