Submitted to MEDINFO 2010

A View on the Current State of the MedFrame/CADIAG-IV Project

Dieter Kopecky, Klaus-Peter Adlassnig, Thomas Vetterlein

Section on Medical Expert and Knowledge-Based Systems, Core Unit for Medical Statistics and Informatics, Medical University of Vienna, Vienna, Austria

Abstract

CADIAG-IV is going to be a data-driven fuzzy expert system for computer-supported consultation in internal medicine based on the PC-based medical expert system shell MedFrame. MedFrame provides a medical institution with a set of powerful tools for developing knowledge bases and inference mechanisms and applying them as expert systems in clinical routine. CADIAG-IV will be the first expert system completely based on MedFrame, significantly extending the usage of fuzzy concepts compared to its predecessor CADIAG-II/-III. After the implementation of the MedFrame core components, a high level inference engine for rule-based knowledge bases has been implemented and used for the realization of the CADIAG-II/-III inference process. In addition, the CADIAG-II/RHEUMA knowledge and patient data have been transferred from the original IBM host system to MedFrame. Currently, the realization of the CADIAG-IV inference, the integration of additional MedFrame components, and the implementation of the user interfaces is in progress. The results achieved so far confirm the applicability, correctness, and performance of the MedFrame concept and the CADIAG re-implementation.

Keywords:

Medical Expert System, MedFrame, CADIAG, Fuzzy Logic, Rheumatology.

Introduction

CADIAG-II and CADIAG-III [1–3] are data-driven fuzzy expert systems built for computer-supported consultation in internal medicine. They provide diagnostic hypotheses and propose further useful examinations in response to the input of a list of symptoms, signs, and laboratory test results seen in a patient. CADIAG-IV [4] is the successor of these two systems and extends the application of already applied fuzzy concepts. In contrast to CADIAG-II/-III, which were developed for an IBM host, CADIAG-IV is part of the PC-based medical expert system shell MedFrame [5,6].

Methods

MedFrame

Expert systems are required to contain at least three components: a knowledge base using a particular knowledge representation formalism, an inference engine, and a dialog component for communication between the system and the user. A knowledge acquisition component and a component for explaining the inferred results are desirable. This is especially true for the field of medical expert systems.

Expert system shells provide the user with a particular knowledge representation formalism and an adequate inference mechanism for building knowledge bases. Therefore, it is no longer necessary to re-implement or adapt many parts of an existing expert system when a new knowledge base is required. Such systems usually have the following drawbacks:

- only a single knowledge representation formalism (e.g., predicate logic) and a single inference mechanism are provided; therefore only a certain class of problems may be modelled;
- knowledge acquisition components are rarely targeted to the needs of domain experts, and therefore specially trained knowledge engineers are required;
- the implementation of tightly integrated—and therefore reusable—dialog and explanation components is usually not supported;
- the possibility to provide reference cases for an automatic validation of the knowledge base after a modification is commonly not available;
- the storage of external information such as patient data is usually not possible;
- extending the expert system shell is a difficult task.

These concerns prompted us to define a set of requirements that a modern expert system shell for use in the medical domain should fulfill. Accordingly, we decided to implement an expert system shell known as MedFrame that will include the following facilities:

- various knowledge representation formalisms to store medical knowledge and adequate inference mechanisms,
- interfaces to add further inference mechanisms,
- concepts for modeling and handling uncertainty in medical terminology and medical relations,
- mechanisms for storing patient data and history,
- a graphical user interface providing the four essential components of expert systems: (a) a knowledge acquisition component, (b) a component allowing the definition of a set of test patients with approved gold-standard diagnoses, (c) an interface for the input of patient administrative and examination data, and (d) a component for displaying the inferred diagnostic hypotheses and proposals for further examination,
- interfaces to adapt the components of the graphical user interface to the requirements of particular medical domains,
- various components for the maintenance of the expert system shell.

By offering these facilities, MedFrame provides the end user with a set of powerful tools for developing knowledge bases and also allows the application programmer to extend the possibilities of the expert system shell components, provided by the system programmer, by implementing a set of interfaces and using a collection of libraries. Therefore, MedFrame significantly reduces the time and cost of building new expert systems requiring specialized inference mechanisms, knowledge acquisition and reasoning components, or access strategies. A discussion of the core components of MedFrame, a flexible object model for storing both, clinical data and medical knowledge, can be found in [5], the further components of the expert system shell – like knowledge acquisition systems, inference mechanisms, … – are considered in [6,7]. A graphical view on the architecture and the components of MedFrame is presented in Figure 1.

CADIAG-IV

CADIAG-IV will be the first expert system completely based on MedFrame. Its predecessor CADIAG-II is a data-driven fuzzy medical expert system providing diagnostic hypotheses and proposing further useful examinations in response to the input of a list of symptoms, signs, and laboratory test results obtained for a patient. To deal with vagueness in medical terminology and medical relationships, CADIAG-II relies on the theory of fuzzy sets, particularly on the concepts of linguistic variables and fuzzy logic [8,9]. A comprehensive discussion of the concepts of CADIAG-II is given in [10,11].

As in CADIAG-II, a clear distinction between patient data on a detailed observational level, i.e., detailed history items, signs from physical examinations, quantitative laboratory test results, etc., and interpreted, abstract symptoms commonly used in diagnostic discourse is established in CADIAG-IV as well. In the beginning of a consultation, a transformation step known as data-to-symbol conversion, which abstracts and aggregates medical information measured or provided by the physician into this internal representation, is applied.

In CADIAG-II, this transformation process assigns a real number in [0, 1], a "degree of presence", to every symptom, where a value of 1 means that the corresponding symptom is proven, while values in (0, 1) mean that the symptom applies to the patient with some strength of confirmation. Symptoms that can be definitely excluded are assigned the value 0. The transformation process is formally defined by a set of linguistic variables and their corresponding membership functions [1–3].

Unfortunately, this methodology can only express total exclusion of a medical entity, but is unable to provide so-called negative evidence, thus indicating the absence of a particular medical entity only to a certain degree. To overcome this limitation, CADIAG-IV assigns two values to every medical entity, (a) strength of evidence and (b) strength of counterevidence. Both values are—in order to overcome the criticism of rather sharp and therefore not very fuzzy point values in CADIAG-II—fuzzy truth values in [0, 1]. The interpretation

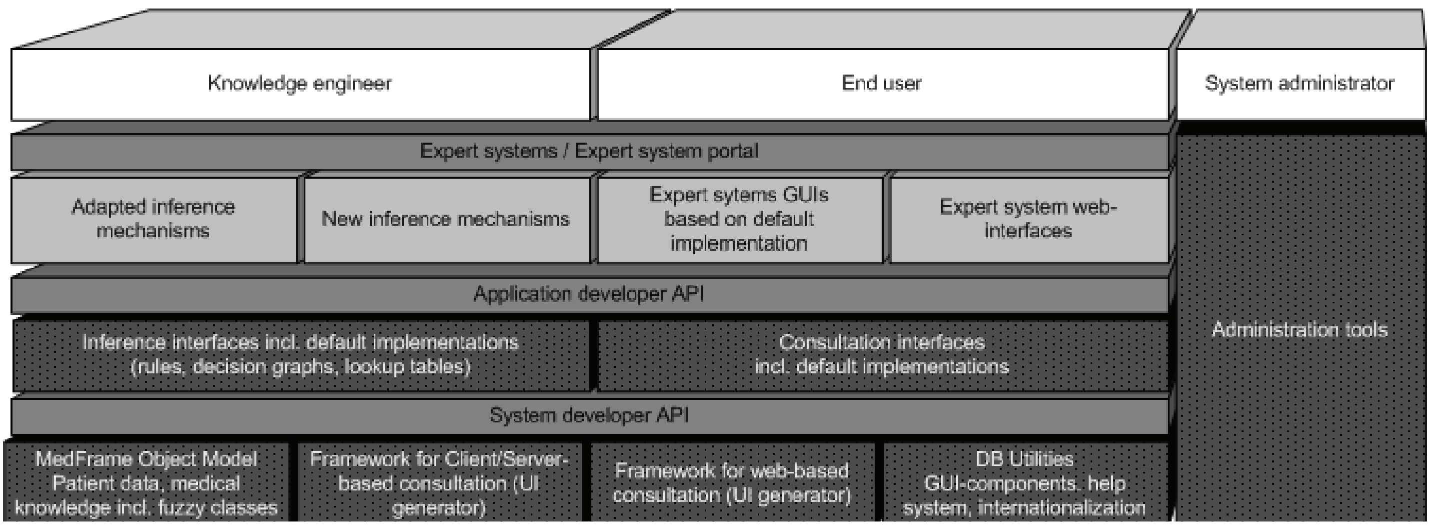


Figure 1 – Architecture View of MedFrame

of these values is as follows: a fuzzy truth value representing 0 means that we have no evidence (or counterevidence) regarding this medical entity, while a fuzzy truth value representing 1 is interpreted as evidence (or exclusion). Intermediate values denote evidence that is not sufficient to prove or exclude the entity in question. Therefore, the data-to-symbol conversion is adapted so that every symptom is assigned two fuzzy truth values instead of a single point value. In addition, the process has been enhanced to deal with context-sensitive as well as pathophysiologically interdependent data. A more detailed examination of the data-to-symbol conversion in CADIAG-IV may be found in [12].

As in CADIAG-II, relationships between medical entities are represented as rules being defined by (a) the strength of confirmation and (b) the frequency of occurrence in CADIAG-IV as well. The occurrence value describes the certainty with which the left-hand side of the rule will occur in patients already showing the right-hand side, and the confirmation value describes the certainty with which the consequence of the rule occurs in patients already showing the antecedent, that is, how much evidence the antecedent provides for the consequence. While in CADIAG-II a relation is described by a point value, in CADIAG-IV rules are classified into two categories: those giving evidence for and those giving evidence against the right-hand side; the strength of the relation is expressed by a fuzzy number in [0, 1].

CADIAG-II supports the following kinds of relations: symptom/symptom, symptom/diagnosis, diagnosis/diagnosis, and symptom-combination/diagnosis (a symptom-combination is a complex fuzzy logical combination of symptoms and diagnoses defining pathophysiological states). CADIAG-IV, in contrast, allows a wider range of combinations, also including therapies and medical treatments, both as antecedents and consequences of rules.

The basic concept on which the inference mechanisms of both systems rely is the compositional rule of fuzzy inference [10]. A comprehensive description of the inference in CADIAG-II is given in [1–3]. The main improvement in CADIAG-IV is its handling of the newly introduced concept of counterevidence and the computation of evidence and counterevidence with fuzzy numbers. Further improvements are the realization of patient-specific adaptation of the rule base during inference as well as the possibility to use different fuzzy operators for the evaluation of symptom combinations and the operators used for the compositional inference rule. The theoretical considerations regarding inference in CADIAG-IV may be found in [4].

Results

After the implementation of the components that extend MedFrame towards a fully-featured expert system shell, in a first step the knowledge base of CADIAG-II/RHEUMA including all available patient data from the relevant clinics has been transferred from the record-oriented representation of the IBM host system WAMIS to the object-oriented model of MedFrame. For this purpose, a set of converter components has been implemented, which contain two parser components, one

```
<Rule>                  ::=    <Symptom> :- <Symptom>  ,<SOC>, <FOO>, {+ | -}
                             | <Disease> :- <Symptom>  ,<SOC>, <FOO>, {+ | -}
                             | <Disease> :- <Disease>  ,<SOC>, <FOO>, {+ | -}
                             | <Disease> :- <SYC-Name> ,<SOC>, <FOO>, {+ | -}
                             | <Therapy> :- <Disease>  ,<SOC>, <FOO>, {+ | -}
                             | <Therapy> :- <Therapy>  ,<SOC>, <FOO>, {+ | -}
                             | <Symptom> :- <Disease>  ,<SOC>, <FOO>, {+ | -}
                             | <Symptom> :- <Therapy>  ,<SOC>, <FOO>, {+ | -}
                             | <Disease> :- <Therapy>  ,<SOC>, <FOO>, {+ | -}

<Symptom-Combination>   ::=  <SYC-Name> := <SYC-Expression>
<SYC-Expression>        ::=  <SYC-Term> {∨ <SYC-Term>}*
<SYC-Term>              ::=  <SYC-Factor> {∧ <SYC-Factor>}*
<SYC-Factor>            ::=  [¬] <SYC-Variable>
<SYC-Variable>          ::=    <Symptom>
                             | <Disease>
                             | <Therapy>
                             | <IC-Name>
                             | (<SYC-Expression>)
                             | <SYC-MIMA-Term>
<SYC-MIMA-Term>         ::=  {fzatleast | fzatmost}(<Part-Whole>, <SYC-Variable-List>)
<SYC-Variable-List>     ::=  <SYC-Expression> {, <SYC-Expression>}*

<SOC>                   ::=  <Fuzzy-Degree>
<FOO>                   ::=  <Fuzzy-Degree>

<Symptom>               ::=  S<Integer>
<Disease>               ::=  D<Integer>
<Therapy>               ::=  T<Integer>

<SYC-Name>              ::=  SYC<Integer>
<IC-Name>               ::=  IC<Integer>

<Fuzzy-Degree>          ::=    PI(<Degree>,<Degree>,<Degree>,<Degree>)
                             | Lambda(<Degree>,<Degree>,<Degree>)
                             | Singleton(<Degree>)
<PartWhole>             ::=  <Integer>/<Integer>
<Degree>                ::=  0.{<Digit>}+ | 1.0
<Integer>               ::=  {<Digit>}+
<Digit>                 ::=  0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9
```

Figure 2 – The Syntax of CADIAG-IV Rules in BNF

for CADIAG-II- and one for CADIAG-IV-rules, as an integral part. Therefore, both the syntax of CADIAG-II and CADIAG-IV-rules have been set up in BNF (as shown in Figure 2) and automatically converted into parser components by the Java parser generator JavaCC. As a result, two MedFrame modules – the equivalent to a knowledge base in MedFrame – were created: CADIAG-II/RHEUMA and CADIAG-IV/RHEUMA.

Building on the components having been developed for MedFrame, in a second step, the inference mechanism of CADIAG-II/-III has been re-implemented in an object-oriented fashion in Java. For this purpose, a very general set of inference components was developed inside of MedFrame, which are capable of dealing with any kind of rules, not only CADIAG-rules. On top of these components the CADIAG-II/-III-inference process has been implemented and compared to the results having been generated by the original CADIAG.

Discussion

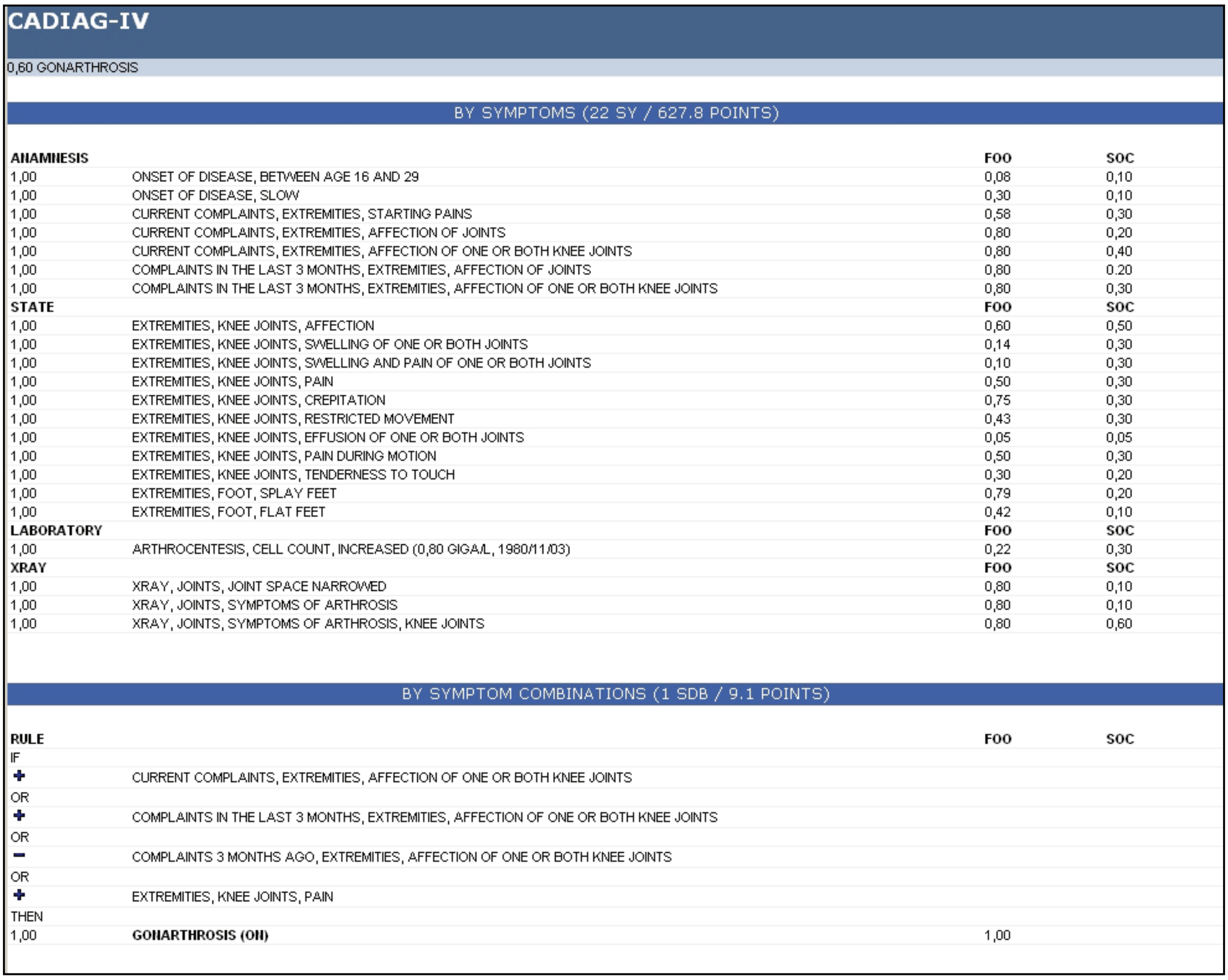
Three major decision support systems for internal medicine have been developed in the past: DXplain [13], Iliad [14], and QMR [15].

DXplain interprets clinical findings (signs, symptoms, laboratory results) to infer a sorted list of diagnoses which might explain (or be associated with) the clinical manifestations. The system applies a modified form of Bayesian logic to derive clinical interpretations and provides justifications for why each of these diseases might be considered.

Iliad uses Bayesian reasoning to calculate the posterior probabilities of various diagnoses under consideration, given the findings present in a case. Iliad is also part of an expert system shell that can be used to develop and evaluate knowledge bases.

In QMR, finally, the successor of INTERNIST-I, data input includes signs and symptoms, laboratory data, and other items of patient history. The inference in QMR is based on a ranking algorithm to produce a list of ranked diagnoses based on disease profiles. The heuristic rules of the system rely on a partitioning algorithm to create problem areas, and exclusion functions to eliminate diagnostic possibilities.

Although the overall design is very similar in all described systems, there is one major concept discriminating them: knowledge representation. While DXplain and Iliad apply Bayesian logic, QMR relies on hierarchical decision-tree logic.


CADIAG-IV

0,60 GONARTHROSIS

BY SYMPTOMS (22 SY / 627.8 POINTS)

ANAMNESIS		**FOO**	**SOC**
1,00	ONSET OF DISEASE, BETWEEN AGE 16 AND 29	0,08	0,10
1,00	ONSET OF DISEASE, SLOW	0,30	0,10
1,00	CURRENT COMPLAINTS, EXTREMITIES, STARTING PAINS	0,58	0,30
1,00	CURRENT COMPLAINTS, EXTREMITIES, AFFECTION OF JOINTS	0,80	0,20
1,00	CURRENT COMPLAINTS, EXTREMITIES, AFFECTION OF ONE OR BOTH KNEE JOINTS	0,80	0,40
1,00	COMPLAINTS IN THE LAST 3 MONTHS, EXTREMITIES, AFFECTION OF JOINTS	0,80	0.20
1,00	COMPLAINTS IN THE LAST 3 MONTHS, EXTREMITIES, AFFECTION OF ONE OR BOTH KNEE JOINTS	0,80	0,30
STATE		**FOO**	**SOC**
1,00	EXTREMITIES, KNEE JOINTS, AFFECTION	0,60	0,50
1,00	EXTREMITIES, KNEE JOINTS, SWELLING OF ONE OR BOTH JOINTS	0,14	0,30
1,00	EXTREMITIES, KNEE JOINTS, SWELLING AND PAIN OF ONE OR BOTH JOINTS	0,10	0,30
1,00	EXTREMITIES, KNEE JOINTS, PAIN	0,50	0,30
1,00	EXTREMITIES, KNEE JOINTS, CREPITATION	0,75	0,30
1,00	EXTREMITIES, KNEE JOINTS, RESTRICTED MOVEMENT	0,43	0,30
1,00	EXTREMITIES, KNEE JOINTS, EFFUSION OF ONE OR BOTH JOINTS	0,05	0,05
1,00	EXTREMITIES, KNEE JOINTS, PAIN DURING MOTION	0,50	0,30
1,00	EXTREMITIES, KNEE JOINTS, TENDERNESS TO TOUCH	0,30	0,20
1,00	EXTREMITIES, FOOT, SPLAY FEET	0,79	0,20
1,00	EXTREMITIES, FOOT, FLAT FEET	0,42	0,10
LABORATORY		**FOO**	**SOC**
1,00	ARTHROCENTESIS, CELL COUNT, INCREASED (0,80 GIGA/L, 1980/11/03)	0,22	0,30
XRAY		**FOO**	**SOC**
1,00	XRAY, JOINTS, JOINT SPACE NARROWED	0,80	0,10
1,00	XRAY, JOINTS, SYMPTOMS OF ARTHROSIS	0,80	0,10
1,00	XRAY, JOINTS, SYMPTOMS OF ARTHROSIS, KNEE JOINTS	0,80	0,60

BY SYMPTOM COMBINATIONS (1 SDB / 9.1 POINTS)

RULE		**FOO**	**SOC**
IF			
+	CURRENT COMPLAINTS, EXTREMITIES, AFFECTION OF ONE OR BOTH KNEE JOINTS		
OR			
+	COMPLAINTS IN THE LAST 3 MONTHS, EXTREMITIES, AFFECTION OF ONE OR BOTH KNEE JOINTS		
OR			
–	COMPLAINTS 3 MONTHS AGO, EXTREMITIES, AFFECTION OF ONE OR BOTH KNEE JOINTS		
OR			
+	EXTREMITIES, KNEE JOINTS, PAIN		
THEN			
1,00	**GONARTHROSIS (ON)**	1,00	

Figure 3 – Screenshot of the Explanation for an Inferred Diagnosis

In our opinion, systems based on Bayesian logic do not have the necessary power to explain to a physician in a comprehensive way why a particular diagnosis has been inferred. The hierarchical or taxonomic decision-tree logic of QMR on the other hand, which links each disease profile to only one "parent" disease class, misses the power to model all the complex cases of internal medicine.

The rule-based approach of the CADIAG systems combines the power to model complex cases and relations with the possibility to clearly explain the reasons for the proposed inference result. Using the concept of fuzzy logic, an additional level of expressiveness can be introduced into automated clinical decision making.

The tests of the newly developed CADIAG-II have confirmed the applicability, correctness, and performance of the MedFrame concept and the CADIAG re-implementation: the inference engine produces the same results as the original CADIAG-II/-III [16], a consultation only takes about a second and is, therefore, pretty fast, and MedFrame provides all functionalities required for implementing an expert system of the complexity of CADIAG-II.

Currently, the realization of the CADIAG-IV inference engine is in progress, so are the consultation user interfaces. For the latter, modern web technologies are applied, including Java Server Faces (JSF), the IceFaces component library and the SEAM application framework. Figure 3 gives an impression of how a typical user interface in MedFrame and CADIAG-IV looks like. It shows the explanation why CADIAG has inferred a particular diagnosis by showing all relevant medical information having led to this decision.

Conclusion

After the implementation of the rule-based inference engine, MedFrame already represents a solid framework for building medical expert systems. In addition, MedFrame knows two additional knowledge representation formalisms and inference mechanisms: decision graphs and lookup tables. For the beginning, these three formalisms form the basis of MedFrame.

The upcoming months will be dedicated to finalizing the CADIAG-IV system and finalizing the missing components of the expert system shell.

References

[1] Adlassnig KP, Kolarz G, Scheithauer W, and Grabner H. Approach to a Hospital-Based Application of a Medical Expert System. Med Inform. 1986;11:205-23.

[2] Adlassnig KP, Scheithauer W, and Kolarz G. Fuzzy Medical Diagnosis in a Hospital. In: Prade H and Negoita CV, eds., Fuzzy Logic in Knowledge Engineering. Köln: TÜV Rheinland, 1986; pp. 275-94.

[3] Vetterlein T and Ciabattoni A. On the (Fuzzy) Logical Content of CADIAG-2. Fuzzy Sets Syst. 2009; doi:10.1016/j.fss.2009.09.011.

[4] Brein L, Adlassnig KP, and Kolousek G. Rule Base and Inference Process of the Medical Expert System CADIAG-IV. In: Trappl R, ed., Cybernetics and System '98. Vienna: Austrian Society for Cybernetic Studies, 1998; pp. 155-9.

[5] Kolousek G. The System Architecture of an Integrated Medical Consultation System and Its Implementation Based on Fuzzy Technology. Vienna: Technical University of Vienna, 1997.

[6] Sageder B, Boegl K, Adlassnig KP, Kolousek G, and Trummer B. The Knowledge Model of MedFrame/CADIAG-IV. In: Pappas C, Maglaveras N, and Scherrer JR, eds., Medical Informatics Europe '97. Amsterdam: IOS Press, 1997; pp. 629-33.

[7] Bögl K, Adlassnig KP, Hayashi Y, Rothenfluh TE, and Leitich H. Knowledge Acquisition in the Fuzzy Knowledge Representation Framework of a Medical Consultation System. Artif Intell Med. 2004;30:1-26.

[8] Zadeh LA. Fuzzy Sets. Inf Control 1965;8:338-53.

[9] Zadeh LA. The Concept of a Linguistic Variable and Its Application to Approximate Reasoning-I. Inform Sci. 1975;8:199-249.

[10] Adlassnig KP. Representation and Semiautomatic Acquisition of Medical Knowledge in CADIAG-1 and CADIAG-2. Comp Biomed Res. 1986;19:63-79.

[11] Adlassnig KP. A Fuzzy Logical Model of Computer-Assisted Medical Diagnosis. Meth Inform Med. 1980;19:141-48.

[12] Bögl K, Leitich H, Kolousek G, Rothenfluh T, and Adlassnig KP. Clinical Data Interpretation in MedFrame/CADIAG-4 using Fuzzy Sets. Biomed Eng Appl Basis Comm. 1996;8:488-95.

[13] Barnett GO, Cimino JJ, Hupp JA, and Hoffer EP. Dxplain – An Evolving Diagnostic Decision-Support System. JAMA 1987;258:67-74.

[14] Warner HR and Bouhaddou O. Innovation Review: Iliad – A Medical Diagnostic Support Program. Top Health Inf Manage. 1994;14:51-8.

[15] Myers JD. The Background of INTERNIST-I and QMR. In: Blum BI and Duncan K, eds., A History of Medical Informatics. New York: ACM Press, 1990; pp. 427-33.

[16] Leitich H, Kiener HP, Kolarz G, Schuh C, Graninger W, and Adlassnig KP. A Prospective Evaluation of the Medical Consultation System CADIAG-II/RHEUMA in a Rheumatological Outpatient Clinic. Methods Inf Med. 2001;40:213-20.

Address for correspondence

Dieter Kopecky, E-mail: dieter.kopecky@ait.ac.at.