Computer Assisted Development of Diagnostic Expert Systems. A Domain-Independent Package (EMPTY) for Acquisition and Use of Expert's Medical Knowledge*

Gianpaolo Molino, Francesca Molino
Dipartimento di Fisiopatologia Clinica - Universita' di Torino (Italia)

Pietro Torasso, Luca Console, Liliana Ardissono, Paolo Pogliano[1]
Dipartimento di Informatica - Universita' di Torino (Italia)

*Supported in part by the Regione Piemonte and Consiglio Nazionale delle Ricerche
1 Current affiliation: CISE SpA, Milano

ABSTRACT

The package EMPTY has been developed with the aim of providing a tool based on well assessed methodological principles which can support both the organisation and use of medical knowledge for diagnostic and educational purposes. EMPTY is domain-independent and results in two interactive programs: ASK guides the acquisition of medical knowledge and RUN supports medical decision-making and provides facilities for medical education. The knowledge bases developed on ASK have a standard formal structure: they include taxonomies, definitions and descriptions of diseases, clinical findings and investigations; production rules for activation and refinement of diagnostic hypotheses; frame-like profiles of diseases; and quantitative criteria for scoring the clinical evidence on the base of available data.

METHODOLOGICAL BACKGROUND

Most authors agree that major diagnostic problems can be categorised as ill-structured problems, the solution of which is neither unique nor based on a well assessed set of indicants [1]. Thus formal reasoning and methodological aspects are of primary importance in medical decision-making.

Clinical diagnosis can be regarded as the result of a reasoning process based on the hypothetico-deductive approach, in which relevant clinical findings are used as cues for identifying suitable hypotheses, the evidence of which is evaluated and compared to be eventually confirmed or disregarded. According to this approach the final diagnosis results from an iterative process progressively refining the hypotheses [1,5].

Early activation of preliminary hypotheses is quite typical both in novices and experts, but experts are more prone to select the most relevant among them [5]. Experimental studies based on the method of "proposition analysis" [17] were performed in both experts and novices. They showed that reasoning from "particular" to "general" is quite typical in the former, while the latter usually reach their conclusions making much larger use of deduction. Moreover experts are much more able in recognising pertinent data, making inferences, using causal explanations and patterns. As a consequence, structured organisation of information and careful selection of reasoning strategies are of basic relevance for both medical decision-making and medical education.

MEDICAL KNOWLEDGE AND MEDICAL EXPERT SYSTEMS

The importance of capturing inside an expert system not only a significant portion of the medical knowledge about a given domain, but also the reasoning strategies that best fit the way human experts use domain knowledge has been recognised, so that in the last few years a significant amount of research has been devoted to analyse specific tasks in order to single out patterns of reasoning. Even if various tasks (diagnosis, treatment and monitoring) have been investigated, the diagnostic activity has received most of the attention. Among the most significant proposals it is worth mentioning "heuristic classification" [3], "establish-and-refine" (hierarchical classification) [2], and "cover-and- differentiate" [6]. Not only is such a kind of analysis quite useful to design knowledge-based systems with a reasoning mechanism which emulates at some degree the reasoning performed by human experts (in our case the clinicians), but also to provide a general framework useful for acquiring and structuring knowledge about a particular domain [15].

Apart from the very few large projects such as INTERNIST/CADUCEUS/QMR [7,10] where the knowledge base is intended to cover a significant portion of medical knowledge, most of the knowledge-based system address only a speciality and more often a sub-speciality. For this reason, in order to be able to develop and test knowledge based systems on more than one sub-speciality, some tool for reducing the amount of work in knowledge engineering is mandatory.

In the last few years some systems have been developed that are able to exploit the decomposition of the task to be performed (diagnosis, therapy, planning) in terms of subtasks and appropriate problem solving methods, in order to guide the domain expert in the development of the knowledge base. Among others, it is worth mentioning PROTEGE [16], MOLE [6], [8].

The present paper is concerned with the presentation of a package (EMPTY), which was designed to provide

0195-4210/92/$5.00 ©1993 AMIA, Inc.

methodological and technological support in developing diagnostic expert systems. The design of EMPTY takes into account our experience in developing knowledge-based systems supporting medical diagnosis in hepatology, such as LITO 1 [9], LITO 2 [4] and CHECK [12]. In particular, the design of EMPTY has benefited by the experience in developing LIED [14], a system able to support both medical decision-making and medical education in the same domain.

GENERAL ARCHITECTURE OF EMPTY

EMPTY has been designed as a package supporting the development of expert systems for clinical diagnosis. As depicted in Figure 1 it results in two interactive modules. The former (ASK) is a program for acquisition and organisation of medical knowledge. Knowledge bases implemented with ASK can be directly interfaced with the latter module (RUN), which contains an inference engine and, in its final version, is expected to provide several educational facilities. The evaluation of EMPTY is similar in some respect to other systems which contain an acquisition and a performance module. In EMPTY, as well as in those other systems, there is a strong connection between the two modules, because the knowledge acquisition is done accordingly to the kinds of knowledge and the reasoning mechanisms employed in the performance module, that are specific to the task at hand.

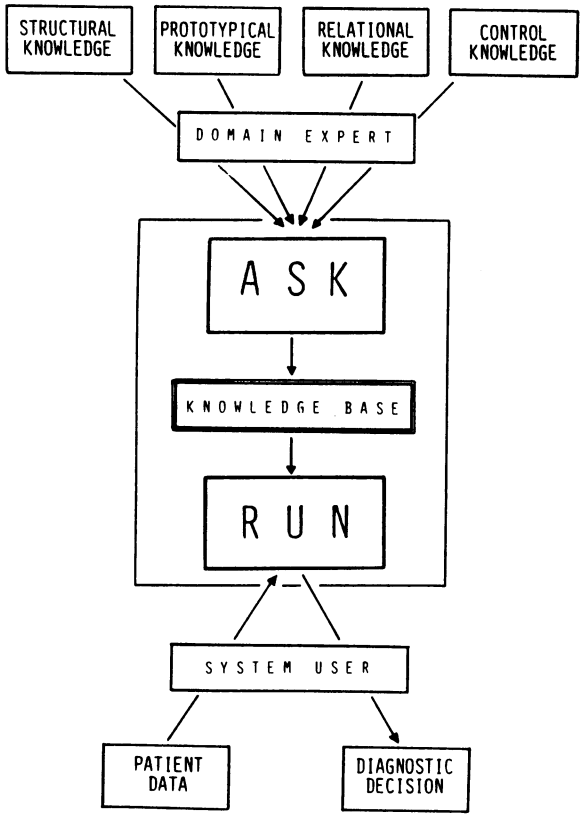

Figure 1: General organisation of the EMPTY system.

Because of these characteristics, a pre-requisite for using the ASK module in the proper way is that the domain knowledge is suitably organised by human experts (clinicians) with respect to both structure and function: structure is essentially based on multi-level taxonomies, definitions and descriptions of diagnostic classes and data, while functions are defined in terms of production rules, frame-like profiles of diseases, and criteria for the evaluation of clinical evidence. To this extent, before using ASK, clinicians responsible for developing or updating a knowledge base are requested to collect all available pieces of information related to the chosen domain from medical literature and personal experience, and to organise them according to a very precise framework. Such a preliminary task is very important since it makes easier the use of ASK and shortens significantly the time requested for implementation.

The methodological approach used in the RUN module can be roughly assimilated to the heuristic classification [3] and hierarchical classification [2] paradigms, even if there are some relevant departure, especially in the treatment of differential diagnosis. The system algorithm features an iterative hypothetico-deductive process, in which diagnostic conclusions (diseases) are considered as specialisations of more general diagnostic classes (syndromes). Hypotheses are activated at first, then refined, evaluated and eventually confirmed or rejected. However, the final diagnosis is never achieved automatically, since the conclusion results from the informed but responsible user choice.

KNOWLEDGE BASE FEATURES

The ASK module is able to guide knowledge organisation since the conceptual entities of the domain have a pre-defined syntax and semantics. In the following the steps that physicians are requested to perform on the knowledge base will be illustrated, with the double aim of describing the basic criteria used in EMPTY, and making available a description of what experts are expected to do.

1. Taxonomies

The starting point for the organisation of knowledge in the EMPTY framework is the definition of taxonomies for diagnoses, clinical data and investigations.

The sequential organisation of the diagnostic algorithm used in RUN implies the existence of a multi-level classification of the diagnostic classes considered. At the first level these are quite general (syndromes), while the subsequent levels include more specific nosological entities (diseases). Further details related to aetiology are considered only for differential diagnosis, which is based on the use of appropriate investigations.

Clinical findings are classified as a taxonomic tree, whose first-level branches are pre-defined in ASK (general data, history, present complaints, physical examination, routine investigations), but can be modified by domain experts whenever this appears to be more convenient. Experts are also requested to select from the

taxonomy of clinical data a limited set of relevant findings, which will be displayed in a menu and proposed by RUN to user attention at the beginning of the consultation (preliminary data).

The taxonomy of investigations includes biochemical, instrumental and histological data, which can significantly contribute to confirm or exclude the final diagnostic hypotheses, or to better define the aetiology. Such a taxonomy can be implemented either instantiating the taxonomic tree defined a-priori and then selecting the investigations related to each disease, or appending suitable sets of tests to each diagnostic hypothesis and labelling them as specialisation of first-level branches.

2. Textual descriptions

In order to provide a common terminology between the system and the user, descriptions of diagnostic classes, clinical findings and investigations must be provided in the form of free texts. Disease descriptions should include a brief definition, a comment on aetiology and pathophysiology, a concise presentation of relevant clinical findings, and relevant references. Clinical findings should be described with a free text including a definition, the description of possible values in the system, and a comment on pathophysiological meaning. Investigations should be defined and described in terms of diagnostic relevance. In addition, essential technical information should be provided, and reference values indicated.

3. System description of data

Since in most cases medical data can be defined in terms of quite different properties (e.g. site, intensity, duration, time-course), in the EMPTY system the author is requested to characterise clinical findings and investigations by means of an appropriate set of specifications. In particular, all clinical findings are characterised by a number of attributes. For each attribute a set of mutually exclusive linguistic values must be indicated. During the implementation attributes and admissible values must be specified immediately after the introduction of the finding name, at any time this may occur. For each investigation the set of admissible values must be defined by authors selecting them from a pre-defined list displayed by ASK.

4. Disease profiles

Each diagnostic class can be defined as a prototypical condition characterised in terms of either clinical findings or typical laboratory results. The assessment of such disease profiles is very important in developing EMPTY knowledge bases, since they support the crucial phase of the diagnostic process concerned with the evaluation of clinical evidence. This is accomplished by matching real patient data with those included in profiles based on clinical findings. For this purpose, authors are asked to select for each disease a set of typical findings. Each finding is associated with a "relevance" index, while each couple attribute/value is associated with a "compatibility" index. The relevance index (0 = irrelevant; 1 = essential) indicates the importance of evaluating that clinical finding in the considered disease. The compatibility index (0 = incompatible; 0.5 = neutral; 1 = expected) indicates how much that value of the attribute is compatible with the considered disease.

During the process of knowledge acquisition the system asks also for disease profiles based on the expected findings of pertinent investigations. Authors can achieve the preliminary assessment of the knowledge base in the same way as for data profiles: significant couples "relevant investigation / expected value" can be selected either considering separately all tests for each disease, or moving from the taxonomy of investigations. Whenever a specific value denoted as "typical result" is selected, a free text describing it must be entered together with appropriate images and drawings in the case of imaging techniques.

5. Control procedures

At the last stage of knowledge acquisition the author is requested to identify associated, alternative and default hypotheses, and to define the rules activating or refining the diagnostic hypotheses. Associated hypotheses are those which reflect possible implications of the current hypothesis. Alternative hypotheses indicate diagnostic classes having clinical manifestations similar to those of the current hypothesis, but which cannot co-exist with it. Default hypotheses are diseases which are so common or relevant that they are activated automatically whenever none of the activation rules for subordinate hypotheses is verified.

During knowledge acquisition all the above hypotheses can be directly selected from the disease taxonomy, under the heading of each diagnostic class.

Activation, confirmation and exclusion rules are suitable associations of clinical data (triplets finding / attribute / value) respectively evoking a particular hypothesis or modifying the clinical evidence previously gathered.

ASK guides the acquisition of rules for each diagnostic hypothesis by displaying the possible findings. The user builds the rules by selecting appropriate findings and attributes, providing values for the selected attributes, and defining a numeric value representing the importance of the rule. For confirmation and exclusion rules the conditions for rule application must be additionally specified, indicating the diagnostic hypothesis whose absence or presence respectively makes a confirmation or exclusion rule applicable.

ACQUISITION MODULE (ASK)

The ASK module has been conceived as an interface supporting a guided implementation of medical knowledge bases. Knowledge acquisition follows a quite complex scheme; this implies that the flux of information provided by experts is well structured and sequentially organised. Thus, as outlined above, it is strongly

recommended that available information is carefully selected and assessed by authors before implementation.

The program provides the following functions:

1. it guides the expert in implementing the knowledge base properly, according to the ASK framework; detailed helps and hints are available to this extent;
2. it ensures the structural and functional compatibility of implemented knowledge base with the RUN module;
3. it provides display and printing facilities to make easier the check of the implemented knowledge base;
4. it allows easy access to the different items of the knowledge base to modify or cancel them, whenever needed.

A pre-defined sequence of steps for knowledge acquisition is proposed by the program; however, the user can choose any alternative approach, provided that he can supply the information requested in the due time.

An example may illustrate how the process of knowledge acquisition works in ASK. Let us imagine that the system is used to implement a knowledge base for liver diagnosis. After entering the taxonomies of diseases and data, and characterising the latter in terms of attributes and admissible values, the user is requested to provide disease profiles and control procedures for each disease.

Let us suppose that the disease proposed by ASK is "alcoholic liver cirrhosis". First, the user is requested to select from the taxonomy of diseases (displayed) associated hypotheses and alternative hypotheses (e.g. hepatoma and primary biliary cirrhosis, respectively). The following operations are: (a) identification of an activation rule, by selecting from the taxonomy of data (displayed) appropriated associations of relevant findings (e.g. liver examination and ascites) evoking the hypothesis; (b) definition, for each selected finding, of the couple attribute/value characterising it in the rule (e.g. volume/increased for liver examination); and (c) indication of the relative importance of the rule. The process can be repeated as many times as necessary up to the completion of the rule set. Thereafter, the user is asked to define by a similar interactive procedure possible confirmation rules (e.g. spleen: volume/increased) and exclusion rules (e.g. AST: level/higher than 1000 IU and jaundice: time course/increasing); possible contexts of application whose absence (e.g. malignant infiltration for the above confirmation rule) or presence (e.g. acute hepatitis for the exclusion rule) makes the rule applicable can be directly selected from the taxonomy of diseases. It is interesting to notice that the system exploits the domain knowledge already entered by the user, in order to guide him in the continuation of the task. In particular, in the definition of confirmation rules for a diseases, only findings relevant to it are displayed. On the other hand, for exclusion rules all findings are considered as potential information to exclude the hypothesis.

OPERATIVE MODULE (RUN)

As suggested by the name, the RUN module has the task of solving specific diagnostic problems by using a proper knowledge base developed by means of ASK. Thus the graphic interface of RUN is user-oriented and very interactive. The diagnostic module follows the algorithm sketched in Figure 2.

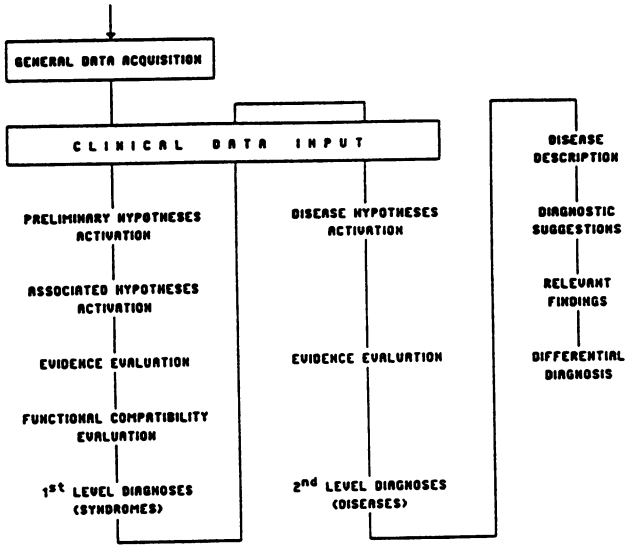

Figure 2: Sequential process featuring the diagnostic algorithm of RUN.

During the consultation, when the knowledge base is used to guide the diagnostic process in a particular patient, RUN asks for patient data requested for activating preliminary hypotheses.

Patient findings are used by means of production rules, which are considered by the program in a decreasing order of importance. This implies that unnecessary data are not requested, since as soon as a hypothesis is activated, the system proceeds to another one without inquiring about data included in less important rules.

Activated hypotheses are evaluated by matching patient data with those included in disease profiles; to this extent missing data are further requested by the program. During the matching, the degree of evidence is evaluated by adopting methods derived from previous studies [18]. Then the user is asked to define a threshold value under which the activated hypotheses will be disregarded following upon.

Thereafter the evidence scores of preliminary hypotheses are displayed and the system activates associated hypotheses (if any) reflecting possible complications of diseases still considered. Moreover, whenever a hypothesis is activated but gathers an evidence score lower than the defined threshold, the system activates and evaluates not yet activated alternative hypotheses corresponding to diseases having quite similar clinical features.

The most relevant first-level hypotheses are eventually refined by means of confirmation and exclusion rules, and the evidence scores re-evaluated accordingly and they are

considered as significant for the corresponding hypothesis only when higher than 0.5. Hypotheses gathering lower values are disregarded by the program unless the threshold has been modified by the user.

Two levels of specialisation can be defined according to the taxonomy of classes. After achieving its first-level hypotheses, the system enters the final diagnostic phase in which conclusive diagnostic hypotheses are activated as specialisations of surviving first-level hypotheses. All clinical data needed for this purpose are acquired. Then a procedure very similar to the one described above is applied to activate, evaluate and refine compatible second-level hypotheses. Whenever no specialisation is verified by existing rules, the systems activates the default hypothesis, if this was defined in the knowledge base.

At the end of the system evolution, the user is asked to select the most reasonable hypotheses on the base of displayed evidence scores and prevalence indexes. For each final hypothesis a set of suitable investigations is suggested to finalise the diagnosis, to support differential diagnosis and to specify possible etiologic subclasses.

The conclusions achieved by the program at any step of the process are illustrated on user request by detailed explanations. These include short texts summarising essentials about diseases and data, disease profiles describing diseases by suitable associations of clinical finding or laboratory investigations, and how-and-why explanations. For the latter purpose RUN displays the rules applied and the data collected, as well as their respective contribution to the evidence scores.

SYSTEM APPLICATIONS

The package EMPTY has been used to develop ICTERUS, an expert system supporting diagnosis in jaundiced patients according to the criteria and rules defined in the Euricterus Project (EC COMAC-BME) [11, 13]. This program was tested on 200 cases, providing satisfactory results. The diagnosis was "correct" (true diagnosis gathering the highest score) in 41% of the cases and approximate (true diagnosis achieved but not with the highest score) in the 35% of cases. A larger evaluation study is in progress.

Another prototype developed with EMPTY is IMMUNE, an expert system devoted to the diagnosis of malignancies of the immune system. Other decision support systems are under development on the diagnosis of vertigo syndromes, female virilism, and diarrhoea.

We are currently extending the RUN system to include all the educational facilities developed in LIED [14].

The current version of EMPTY is written in PROLOG-2 and runs under MS-DOS on 80386 personal computers.

REFERENCES

1 Barrows HS and Feltovich PJ - The Clinical Reasoning Process -Med Ed 1987, 21: 86-91

2 Bylander T and Mittal S - CSRL: A language for classificatory problem solving and uncertainty handling - AI Magazine 1986, 7: 66-77

3 Clancey WJ - Heuristic classification - Artificial Intelligence 1985, 25: 289-350

4 Cravetto C, Lesmo L, Molino G, Torasso P - LITO 2: a frame based expert system for medical diagnosis in hepatology - In: Artificial Intelligence in Medicine (De Lotto I and Stefanelli M, Eds), Elsevier Science Publishers, North Holland, 1985

5 Elstein AS, Schulman LS, Sprafka SA - Medical Problem Solving. An Analysis of Clinical Reasoning - Harvard University Press, Cambridge (Mass), 1978

6 Eshelman L, Ehret D, McDermott J, Tam M - MOLE: a tenacious knowledge acquisition tool - Int J Man-Machine Studies 1987, 26: 41-54

7 First MB, Soffer LJ, Miller RA - QUICK (Quick Index to Caduceus Knowledge): Using the INTERNIST-I/CADUCEUS Knowledge Base as an Electronic Textbook in Medicine - Computers and Biomedical Research 1985, 18: 816-822

8 Lanzola G, Stefanelli M. - A Knowledge Acquisition Tool for Medical Diagnostic Knowledge-based Systems - Proc. AIME 91, 273-282.

9 Lesmo L, Marzuoli M, Molino G, Torasso P - An expert system for the evaluation of liver functional assessment - J Med Systems 1984, 8: 87-101

10 Miller RA - INTERNIST-I/CADUCEUS: Problem Facing, Expert Consultation Programs - Meth Inform Med 1984, 23: 9-14

11 Molino G, Ballare` M, Aurucci PE, Ripa di Meana V - Application of artificial intelligence techniques to a well defined clinical problem: jaundice diagnosis - Int J Biomed Comput 1990, 12: 189-202

12 Molino G, Cravetto C, Torasso P, Console L - CHECK: a diagnostic expert system Combining HEuristic and Causal Knowledge - Biomed Meas Infor Contr 1986, 1: 182-193

13 Molino G, Molino F, Ballare` M - ICTERUS: expert system for medical decision-making using the Euricterus Knowledge Base - Internal report (Euricterus Project) 1992

14 Molino G, Ripa di Meana V, Torchio M, Console L, Torasso P - Educational application of a knowledge-based expert system for medical decision-making in hepatology - It J Gastroenterol 1990, 22: 97-104

15 Musen M - Automatic generation of model-based knowledge acquisition tool - Pitman, Boston 1989

16 Musen M - An Editor for the Conceptual Models of Interactive Knowledge-acquisition Tools - Int J Man-Machine Studies 1989, 31: 673-698

17 Patel VL, Groen GJ, Frederiksen CH - Differences between medical students and doctors in memory for clinical cases - Med Ed 1986, 20: 3-9

18 Torasso P, Console L - Approximate reasoning and prototypical knowledge - Int J of Approximate Reasoning 1989, 3: 157-178