HEARTVIEW - a knowledge base to support clinical research in cardiology

Beatriz de F. Leão¹, Teun Timmers², Johan van der Lei², Erik M. van Mulligen²

¹ Post-doctoral fellow from the Inst. of Cardiology RGS, PA, RS, Brazil
² Department of Medical Informatics, Erasmus University, Rotterdam

Abstract

This paper presents HEARTVIEW - a knowledge base (KB) that offers aids for clinical research in cardiology. This KB is an essential component of the medical workstation MW2000, now under development at the Department of Medical Informatics at the Erasmus University. HEARTVIEW integrates different types of knowledge: a conceptual model of the medical record in cardiology and knowledge on how to perform data analysis, according to the cardiological sub-domain. The design of HEARTVIEW is based on the assumption of a general structure in the medical record. The KB can be consulted by other modules in the MW2000 and by users through a medically-oriented graphical interface. The prototype is implemented on a Xerox 1100-series workstation. After evaluation, it will be transferred to the Unix environment, where the MW2000 is under development [1,2].

Introduction

In spite of the great advances in medical informatics in the last 20 years, clinical research still is an arduous, time consuming, not computer-supported task. In order to offer physicians support on how to perform the necessary steps in clinical researches (starting with the research protocol design, up to the selection of the appropriate graphics to present the results), a clinical research workstation should contain, besides a solution to the complex and fundamental issues of software and data integration, a knowledge base with both medical and statistical knowledge [1-5].

The source of information for clinical research remains the medical record, whether or not computerized. Medical records from different specializations and different sub-domains of the same specialization generally have different contents. For example, medical records of ischemic patients probably contain in their anamnesis an extensive and detailed description of coronary risk factors, while the record from a newborn with a congenital heart defect contains details on the patient's birth conditions. Furthermore, the type of analysis to be performed depends on the medical context. For example, long-term evaluation of cardiac surgery patients requires actuarial methods [6].

In spite of differences in each medical context, a general structure can be recognized in the medical record, expressed by its standard subdivisions such as "identification", "anamnesis", "physical examination", "investigation", "diagnosis", etc.. This structure has remained almost unchanged and is still taught to medical students as an adequate framework to acquire and represent information about a patient [7]. However, most of the computerized medical files do not use this framework for representing medical data. The potential of our approach is to formalize a general structure of the medical record, which retains some of the flexibility of the paper.

Intelligent processing of information stored in a database demands the integration of two different well-known technologies: a database management system (DBMS) and an expert system (ES). This integration is nowadays pointed out as one of the main issues for the development of real decision-support systems [8-11].

Two interrelated techniques have been described for encoding knowledge about data [12-13]. The first technique is to enhance the data model, incorporating a rule or frame-based mechanism with data-interrelationships, statistical knowledge, or causal relationships. The second technique (the one most frequently used) consists of building an external interpreter, apart from the DBMS, which contains knowledge about the data. Queries to the DBMS are performed through this "intelligent" module.

More recently, object-oriented (O-O) semantic models have been identified as a possible paradigm for integrating knowledge bases and database management systems [14-18]. Their main advantage over traditional hierarchical, network or relational approaches is based on their capability of explicitly representing abstractions, allowing for the database to be seen as a collection of abstract objects, rather than as a set of flat tables. Generalization or subtyping relationships and the possibility of defining methods locally, capturing integrity constraints easily and more efficiently, are other benefits of object-oriented approaches [16,19-23].

The purpose of this paper is to present HEARTVIEW - a prototype of a KB for clinical research in cardiology. HEARTVIEW is an independent module inside MW2000, now under development as described in [2]. This module can be consulted by other system modules or by users through a graphical interface. Its main objectives are to offer, (1) advice for setting up clinical research, e.g., suggesting the contents and structure of the database; (2) support of data-analysis, and (3) mapping of different types of databases to the central research database of the MW2000.

The following sections describe the knowledge representation employed, how the model can be consulted, and how it will be evaluated.
Knowledge Representation

The knowledge encoded in HEARTVIEW consists of (1) a conceptualization of the medical record, including description of medical concepts, data dictionaries for coded information, as well as database handling information (type of data, field-size, ranges) and (2) support for data analysis, suggesting tests that could be applied in specific research contexts, variables that should be present, and suggestions for possible graphical outputs.

The medical knowledge for this prototype is based on two well-defined research protocols now under evaluation in the Thorax Center at the Erasmus University: "Impetus" for myocardial infarction (ischemic heart diseases) [24] and "HCMFU" for hypertrophic cardiomyopathy late follow-up [25].

The HEARTVIEW KB is encoded in a five-layered structure. Each level corresponds to one medical concept represented by a conceptual object and instances of these entities. From top to bottom the conceptual objects, from now on mentioned as objects in the text, are: (1) "domain" describe a sub-domain in cardiology; (2) "hospital" describe the hospital department or ambulatory clinic where the information comes from; (3) "record" are formed by the different parts or chapters of the medical record, such as identification, history, physical examination, etc.; (4) "class" are chunks of medical findings that can be seen through a common medical concept, like "non invasive exams" consisting of electrocardiogram, echocardiogram, CXr, scintigraphy, etc.; and finally, (5) "findings" representing all elementary information from the domain, such as a description of a murmur, a lab test or a symptom like chest pain. Although not yet complete, all finding frames have "synonyms" and "thesauri" slots. The former gives additional medical names by which this same finding can be called (there are sometimes local peculiarities of describing specific signs in cardiology), while the latter intends to be a link with other medical nomenclatures such as UMLS [26-27], SNOMED [28] and QMR [29].

Relations were defined to link the different layers of the KB. These links were implemented in the O-O environment as properties of the objects. From top to bottom:

(1) "has_department", property of the object "domain", links "domain" instances to "hospital" instances, through a list of all hospital instances that might be present in a given domain;

(2) "has_record", property of the object "hospital", links "hospital" instances to "record" instances. For example, for the "amb_cong" instance, the property "has_record" could be: "ident_cong", "history_cong" and "diagn_cong" expressing a record with information on the patient identification, history and diagnosis;

(3) "has_class", property of the object "record", links "record" instances to "class" instances. For example, for the "investigation_isch" instance the property "has_class" contains the reference to "non_invasive_isch" and "invasive_isch". These latter two are instances of a "non_invasive" and "invasive" - and are represented in the third layer of the KB. Finally, the property:

(4) "has_finding" links "class" instances to "findings" instances. For example, "non_invasive_isch" is linked by this property to the findings instances "egc_isch", "echo_isch" and "scintigraphy_isch".

Figure 1. HEARTVIEW five-layered structure. Example for ischemic heart diseases.

Each sub-domain in Cardiology (ischemic, congenital, valvular etc...) is represented by specific instances in the different layers of the KB. For example, ischemic heart diseases (IHD) can be represented by (fig 1):

(1) one "isch_domain" instance in the top layer of the KB, linked to:

(2) instances of "hospital" objects, in the second layer of the KB, reflecting the hospital departments that are related with ischemic heart diseases, like "isch_amb" (ambulatory for ischemic patients) and "ccu" (coronary care unit), linked to:

(3) instances of "record" objects, in the third layer of the KB, specializing the different aspects of the medical record for IHD, like "isch_hist", for the patient's history "isch_investigation" for the patient's investigation and "isch_diagn" for the patient's diagnosis, linked to:

(4) instances of "classes" objects, in the fourth layer of the KB, like "non_invasive_isch", "risk_factors" and "chest_pain", linked to:

(5) instances of "findings" objects, in the fifth layer of the KB, describing the medical finding for IHD, like "egc", "smoking", and "angina_pectoris".
Knowledge for data-analysis is stored in two different types of objects, at the top layer of HEARTVIEW: instances of domain frames and instances of statistical frames. The former contain for each sub-domain in cardiology, a list of statistical tests that could be applied, lists of instances of finding frames that should be present, and advice on the type of graphs that could be used for the presentation of the data-analysis results. The statistical frames, not yet implemented in the prototype, will contain the description of statistical tests, their constraints and rules for interpreting the results. This description will be used for offering help to users not familiar with data analysis.

**Heartview Consultation**

HEARTVIEW communicates with the external world through a graphical interface, presenting the five-layer model to the user, allowing navigation with a mouse-pointing device. A major concern was to make the user interface medically-oriented. For example, if a user wants to specify a field for a new database, he is asked to indicate in the model which medical finding he wants in his environment rather than the name of the database attribute. Additionally, different types of users get different default help screens. For example, an epidemiologist, when ordering for help, gets a description of the medical concept he is pointing to with the mouse, while a physician receives a comment on how this information could be used for data analysis.

Creating or importing a database.

The process of creating a new database or mapping an already existing one to the MW2000 environment, starts by selecting which parts of the model the user wants. The selection is performed interactively in a top-down approach. Starting from the “domain” layer, the user selects the hospital departments that are present in his records. For each hospital department the user goes down in the model, specifying the parts of the medical record that are present and, finally, the medical findings that are present in his database. Figure 2 shows an example with the user’s selection for mapping a database for follow-up of patients with hypertrophic cardiomyopathy. For example, to map an electrocardiogram from the ambulatory, the user first selected at the hospital layer Ambulatory, and later selected ECG at the findings level. Each finding selected by the user in HEARTVIEW’s model, is internally represented as an instance of a conceptual object named “finding-selected”. These objects are similar in structure to the “findings” object however, they have two additional properties to define the attribute’s name in the external database - "ext_atr_name" and in the central research database - "centr_atr_name". The "ext_atr_name" contains the name of the attribute that represents this finding in the database that is being mapped to the MW2000 environment. The "centr_atr_name" is a unique name built by HEARTVIEW and contains the relation and attribute name of that finding for the central research database of the MW2000. The relation name is formed by the concatenation of the name of the hospital department instance selected with the selected finding name. Finally, a number is added that specifies the order of occurrence in that relation, in order to cope with repetitions of attributes in the same relation. For instance, in the example mentioned above of mapping an electrocardiogram from the Ambulatory to the MW2000 environment, the relation name will be AMB_NI, where “AMB” is the short name for Ambulatory, and "NI" is the short name for non-invasive, the class frame to which electrocardiogram (ECG) is connected (Fig 1). The attribute name of the first ECG from the Ambulatory will then be AMB_ECG_1.

Since we are dealing with research topics, discrepancies between the model and the user’s database are expected, especially when dealing with coded information. However, in order to avoid ambiguities in the KB, physicians will be able to modify the model only at the findings level, by adding or editing elements in a code list, or changing ranges for numerical values. These modifications will only be seen in each user’s specific environment, maintaining the integrity of the central model. Major changes are only allowed when introduced by the development and maintenance team.

![Figure 2. Example of selection from HEARTVIEW for building a follow-up database of patients with hypertrophic cardiomyopathy.](image)
Implementation and Evaluation

The prototype is implemented on a Xerox 1100-series workstation, using the Epitool development environment, version 3.0 [30]. The KB will have its first evaluation in the Epitool environment, before being transferred to the Unix environment and integrated with the other modules of MW2000.

The evaluation of the prototype will be performed in the Thorax Center at the Erasmus University. During the evaluation, HEARTVIEW will be used by epidemiologists and physicians of different cardiology sub-domains, for a period of two months. In this evaluation, the quality of the user interface (according to physicians and epidemiologists), the medical knowledge, the support of data analysis, and the overall gain in time when using the O-O model for manipulating clinical research databases in cardiology will be assessed.

HEARTVIEW automatically keeps track of each consultation, allowing the construction of a mean graph, for each domain-frame, expressing the most frequent paths employed in that context, at the end of the period of evaluation. The algorithm to calculate mean knowledge representation graphs has already been successfully employed for constructing diagnosis graphs for congenital heart diseases, based on the knowledge of several experts [31-32]. Based on the information of these mean graphs the design of the model will be updated for further evaluation.

Discussion

The difficulties in formalizing medical information are probably among the main aspects why, until now, medical decision support systems are not widely in use. This paper starts from the observation that the medical record remains the main source of information for clinical research. The basic assumption in the construction of HEARTVIEW was that there exists a common structure in the medical record, represented in HEARTVIEW by its highest three layers (i.e., the domain, hospital and record frames). Below this common structure, differences may occur in each finding's description.

The work described in this paper overlaps with other recent developments in medical informatics. Linnarsson et al. [8] argued for the necessity of a richer data-dictionary allowing for the integration of clinical databases and medical knowledge bases. Various groups aim at the implementation of a UMLS semantic network [26-29], including synonyms and thesauri.

However, our approach differs from those described from various points of view. Firstly, the structure of HEARTVIEW comprises not only the findings level of the medical record and groupings thereof, but higher levels as well. This allows for both flexibility at the lower levels, and a stable structure at the higher levels. Secondly, the model can be used for graphic database browsing, and it provides a framework for storing knowledge about support for data-analysis. In addition to this, the design that we have chosen fits into the general design of the MW2000, where the knowledge base will be consulted by other modules as well (e.g., to provide contexts for parts of the user interface). Finally, our development methodology intends to cover in depth many aspects of a limited, practical domain (cardiology) rather than a broad domain.

It is expected that this approach is feasible with a close connection with the daily practice of clinical research, allowing for rapid prototyping, evaluation and development.

Acknowledgments

The authors would like to express their gratitude to Mr. Mees Mosseveld, M.Sc. for his competent programming support.

This project is partially supported by the Brazilian Federal research Agency, grant nº 204034/89-2.

References


[24]. Fioretti P. Predischarge assessment of prognosis after myocardial infarction. Doctoral thesis, the Faculty of Medicine, Erasmus University Rotterdam, November 1986, pp.133.


Address for correspondence: Beatriz de F. Leão, MD
Department of Medical Informatics
Erasmus University Rotterdam
PO BOX 1738 3000 DR Rotterdam
The Netherlands
e-mail: LEAO@HROEUR51.bitnet