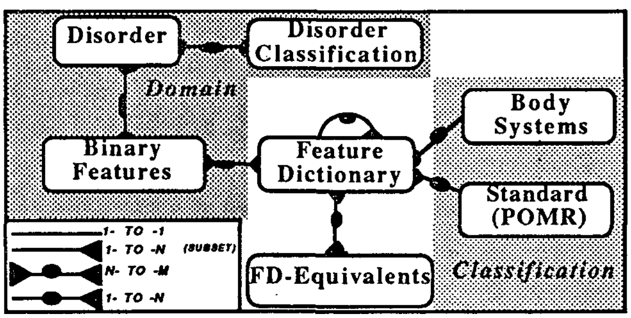
A Relational Database Design in Support of Standard Medical Terminology in Multi-domain Knowledge Bases

Frank Naeymi-Rad[1], Lowell Carmony[2], David Trace[1], Christine Georgakis[3], Max Harry Weil[1], Martha Evens[4]

1. *University of Health Sciences/The Chicago Medical School, North Chicago, Illinois 60064*
2. *Lake Forest College, Department of Mathematics and Computer Science, Lake Forest, Illinois 60045*
3. *Northeastern Illinois University, Department of Information Science, Chicago, Illinois 60625*
4. *Illinois Institute of Technology, Department of Computer Science, Chicago, Illinois 60616*

Relational database techniques have been used to create knowledge bases for a medical diagnostic consultant system. Known as MEDAS (Medical Emergency Decision Assistance System), this expert system, using disorder patterns consisting of features such as symptoms and laboratory results, is able to diagnose multiple disorders. Database technology has been used in MEDAS to develop knowledge engineering tools, called the TOOL BOX, which permit domain experts to create knowledge without the assistance of a knowledge engineer.

In the process of knowledge development with the TOOL BOX a standardization of terms was needed. This led us to design a Feature Dictionary and a grammar to support a standardized format for features. A common dictionary of features will allow us to merge knowledge bases, translate between multi-domain bases, and compare competing expert systems. In addition, standard terminology will assist communication across domains.



The ER-Diagram of MEDAS and the Feature Dictionary

The **Feature Dictionary** has the following attributes: *Long* forms of the feature name (White Blood Count) and *short* forms (WBC) as well as a three line description of the feature. The *type*, binary (Abdominal Pain), continuous-valued (WBC), or derived (pulse pressure = systolic - diastolic) is also kept for each feature.

For value features the appropriate *unit* (cc, kg, etc.) as well as *range* limits are stored so that these can be used as a form of quality control on input. The *permanence* (Y/N) of each feature is kept so it is possible to automatically include permanent features in future encounters. In addition, for each feature three separate *"cost"* parameters are kept: *Risk* measures the danger to the patient from no risk such as taking a blood pressure to highly invasive proceedings such as a liver biopsy. *Time* measures whether results can be expected in minutes, hours, or days. *Money* measures the actual cost to the patient. **FD-Equivalents** stores the synonyms and antonyms of each feature. These are used to translate between knowledge bases using different terminology.

Features were first **classified** in terms of a Problem Oriented Medical Record. We have added an anatomical reclassification in terms of body systems. Experts will be able to add new kinds of feature classifications.

MEDAS, a multi-membership Bayesian model, needs binary representations for its inference. These **Binary Features** are created by the expert physician in the given disorder patterns. For example, "WBC > 50,000", or "Age > 2 & Female & Hematocrit > 42" are binary features that might appear in a disorder pattern. Laboratory results often lead to a multiplicity of binary features (such as "WBC < 3,000", or 3,000 < WBC < 10,000, etc.). Our design allows the user to enter the value of such a feature and have the system set of all the corresponding binary features. This intelligent user interface is controlled by a grammar that allows us to parse the binary features and generate rules for them.

The knowledge base for a particular problem **domain** such as OB/GYN is organized as a collection of disorder patterns. Each of these is represented as a list of binary features and associated probabilities. The domain knowledge base contains only the features relevant to that domain.

Experience with the Feature Dictionary has convinced us that there are many advantages in using a DBMS to store the knowledge base for an expert system. The TOOL BOX, originally in ACCENT-R, was rewritten in dBase III for the PC. The knowledge bases created on the PC were then ported to the mainframe. As the number of domains supported by MEDAS grew, it became evident that we needed a DBMS that could function in both environments so we are in the process of converting to ORACLE.