MEDINFO 2007
HISA Ltd Australia
© 2007 The authors. All rights reserved.

A Data Structure For Decision Support Systems, Medical Expert Systems and Clinical Decision Making

Bassøe, Carl-Fredrik

The Norwegian Center for Electronic Medical Records, Institute of Neuroscience, Faculty of Medicine, University of Trondheim, the PROMED Institute, Bergen, Norway and the Department of Information Science and Media, Faculty of Social Sciences,University of Bergen, Norway

Abstract

The International Classifications of disease (ICD10) is normalized the into a relational containing tables for the etiology, body regions, organs, pathogenesis, cells, subcellular organelles, direction of change, time course and degree. Functions and morphology of different organs are abstracted into the general organ. The hierarchal structures of the etiology and anatomy are implemented. The database enables the construction of diagnoses according to the equation d:=e+o+p, where d, e, o and p each is a variable holding names of a diagnoses, etiology, organ and pathogenesis, respectively. Together, the data structure and the equation d:=e+o+p allow structured clinical decision making and a combinatorial classification.

Keywords:

classification, database normalization, clinical decision making, computer decision support system, medical expert system, education

Introduction

Computerized medical decision support systems (CDSS) and medical expert systems (MES) need to be based on a clear and versatile data structure. Both CDSS and MES require a knowledge base containing the etiology, organs and their parts, and the pathogenesis. However, etiology and pathogenesis are used differently in different settings and their meaning overlaps. A clear concept of a general organ as an object, that abstracts functions and morphologically different components, are also lacking.

The present study aims to extract etiology and pathogenesis from dependencies in ICD10 and ICF and normalize ICD10 into a database. The dependences were identified semiautomatically using an application crafted for the goal. The database is used to create a combinatorial classification and construct diagnoses.

Materials and methods

Hardware and software

The programs and databases were run on ordinary PC's. MS Access 97® was used as database. The applications were implemented in MS Visual Basic 6® (VB6) using ADO to connect to the database [1]. All work on the classification was done interactively.

Normalization of ICD10

An English version of ICD10 was downloaded from WHO [2], and a Norwegian version from KITH [3]. ICD10 was converted it into Boyce Codd Normal Form (BCNF) [4,5] by identifying the entities and attributes of diseases in ICD10. An Entity-Relationship model (ER) was built from the entities, their attributes and relationships [4,5], storing the entities in separate tables.

The lists containing codes and diagnoses in ICD10 were automatically transferred from the sequential text files into a database table with fields for tagging. A VB6 application was developed to identify words with standard medical prefixes and suffixes associated with the etiology, pathogenesis, morphology and direction of change. The collection of tables is called a combinatorial classification

Results

Functional dependencies in ICD10

ICD10 shown to contain dependencies such as 'due to use of', 'due to' and 'of presumed'. Altogether 76 different dependencies were identified. The dependencies were used to automatically parse and display dependenciesand transfer them to the appropriate tables combinatorial tables. The process iterated until the whole ICD10 had been examined for dependencies.

The etiology is defined as all agents that cause disorder and are external to the individual in time (heredity) or space. Nymerous etiological agents were detected in ICD10. Anatomical regions were identified >184 diagnoses. The direction of change is given by the prefixes a, hypo, normo, hyper, dys, meta and neo. A total of 173 diagnoses with the direction of change were identified. Names organs and organ parts were common in ICD10 diagnoses. The general organ is an abstraction of features common to all organs: the parenchyma, tube systems, cavities or slits, and the functions and morphology of these three parts. Mechanical, electrical, chemical and optic functions were described in ICD10. Thus, the entities eti-

ologi, region, direction of change, organ (part), general organ functions, degree, time course are required for a combinatorial classification created from ICD10. In addition, morphological deviations require a separate entity. Surprisingly, ICD10 contained only six diagnoses related to organelles, but this is sufficient to require an organelle entity.

The pathogenesis is defined as a relation between organs and between cells. The pathogenetic mechanisms were identified by suffixes, disorders of mesenchyme, intercellular metabolism and regulators immunological reactions, disturbances of coagulation and psychic reactions. These elements were allocated to the pathogenesis entity.

Diseases and dagnoses as ternary relations

Causal connectives between the etiology, organs or cells and the pathogenesis were identified. In particular infection can be seen as a ternary complex consisting of microorganisms (etiology), functional and/or morphological disorder in the primary affected organ or cell, and immunological reactions (pathogenesis). Allergy also becomes a ternary concept where the etiology is some allergen, and the organ and pathogenesis are the same as in infection. All the above entities give rise to the entity-relationship diagram (Fig. 1).

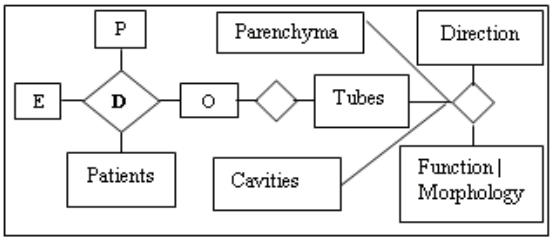

Figure 1 - Entity-relationship (ER) diagram modeling diagnoses or diseases (D), showing the relations of d to the etiology (E), organs (O) and pathogenesis (P). Organ is decomposed into the organ parts cells, tubes and cavities. These subcategories combine with the direction and function or morphology to make out a diagnosis. Diagnoses D are also related to patient

Since pre- and suffixes are recovered with the name of the elements in the entities they were used to construct diagnoses by the equation **d:=e+o+p** where d is the diagnosis, and e, o and p holds names from the etiology, organ and pathogenesis entities, respectively. A typical example is e:='bacterial ', o='endocard' and p='itis', which gives d='bacterial endocarditis'. An extended version of (1) generated diseases using all the entities described above allows the construction of about 10^{15} diagnoses (not shown).

Discussion

The present study shows that ICD10 can be normalized into a relational database. The entities (tables) can be used in a combinatorial classification and used to construct diagnoses according to the equation d:=e+o+p. This allows the construction of diagnoses simultaneously with the classification of a clinical case. The data structure is devoid of redundancy and seems well adapted to classification..

The clinical findings associated with different entities are largely disjunctive (not shown). Thereby, the prominent overlap in clinical findings between different diseases is eliminated. The combinatorial approach is therefore well suited to clinical decision making, decision support and medical expert systems.

In conclusion, ICD10 can be normalized into a relational database which allows a combinatorial classification that is easy to extend, fast and systematic, and allows the construction of diagnoses.

References

[1] Siler B & Spotts J: Using Visual Basic 6. Que, Indianapolis, 1998.
[2] http://www.who.int/classifications/icd/en/
[3] http://www.kith.no/
[4] Elmasri R & Navathe SB: Fundamentals of database systems. Benjamins Cummings Publishing, Redwood City, CA, 1989.
[5] Jackson GA. Relational database design with microcomputer applications. Prentice-Hall Inc, Englewood Cliffs, NJ, 1988, pp. 23.

.Acknowledgments

I thank Andreas Lothe Opdahl for valuable advice.