Medical Informatics Europe '96
J. Brender et al. (Eds.)
IOS Press, 1996

Medical Data Warehouse, an Investment for Better Medical Care

Niinimäki J [1], Selén G [2], Kailajärvi M [1], Grönroos P [1], Irjala K [1], Forsström JJ [1]

[1] *Central laboratory, Turku University Central Hospital, Kiinamyllynkatu 4-8, 20520 Turku, Finland*
[2] *Department of Computer Science, Åbo Akademi University, Lemminkäisenkatu 14, 20520 Turku, Finland*

Abstract. The concept of data warehouse has been utilised for decision support in business world for some years. Medical care has similar demands as business companies, since the quality of care is a central criterion for hospital efficiency. Information collected in hospitals is often scattered and cannot be efficiently used to improve the quality of care. A data warehouse, which is a central database separated from operational databases, was used to investigate how frequently drugs that may have affected the result of a common laboratory test were prescribed. Similarly, data warehouse was used to survey the frequency and types of potential drug interactions in hospital. Medical data warehouse comprising laboratory results, patient medications, diagnoses and medical procedures facilitates clinical research and consequently improves the quality of medical care.

1. Introduction

The data warehouse concept has played an important role in decision support systems of business companies for the last few years. The value of centralised information repository has been realised. In many ways modern hospitals resemble large business companies, since only high quality medical practice is profitable. Medical practise is mainly improved via clinical research, for example by comparing established methods with new experimental ones. Unfortunately, medical information systems have rarely been developed for clinical research, thus the information is not easily accessible for research purposes.

The need for storing medical data originates from specific interests, such as hospital administration or collecting disease statistics. Quite commonly, the recorded data are scattered in several separate non-uniform databases. In the current situation, in order to utilise the collected data, the researcher must first apply for a permission to access these data stores and then fill in search forms to ask computer department to execute the needed database queries. This convention is not only inefficient and waste of human and computer resources but also prone to errors.

The data warehouse provides a modern and sophisticated answer to this problem. Scattered information is collected in one central database where the researcher can personally make the queries from a PC workstation without consuming time and resources of the computer department.

In Turku University Central Hospital (TUCH), hospital and laboratory information systems have been in routine use for more than a decade. During the last few years, databases have been established for patient medications and laboratory test results too. The

most recent additions have been databases for drug-drug interactions and for drug effects on laboratory tests [1].

In the next section we describe the structure of the data warehouse in TUCH and in section 3 we will present our preliminary results on utilising the data warehouse.

2. Technical Description of the Data Warehouse

The data used daily in hospitals are generally accumulated in operative databases. The structure of an operative database enables easy insertion, update and retrieval of information concerning one single patient at a time. This kind of data management is usually called on line transaction processing. The need for executing extensive searches processing many patients and combining e.g. laboratory tests, medications and diagnoses, was not taken into consideration when the operative database was designed.

The vast amounts of data stored in operative databases are spread over many tables, files and computers, and stored in a variety of formats, that also change over time. In such an environment, a search program is inevitably difficult to construct and even more difficult to maintain, demanding regular involvement of computer specialists. The risk that the performance of the operative database drastically decreases during an extensive search is intolerable in a hospital. In the worst case, an error in the search program could also wipe out important information.

Our aim is to develop a procedure enabling a way for the researchers to conduct the search process themselves without having to trouble the already overworked computer department. Here, the concept of data warehouse is introduced. The data warehouse is sometimes called on-line analytical processing [2], sometimes data mining [3]. The concept of data warehouse is undergoing rapid development and is target for large discussions these days.

The main principle of a data warehouse is that various parts of the operative database are extracted, transported to the data warehouse, and arranged in a form that is tailored for search purposes. The data warehouse resides physically on a separated computer, isolated from the operative database, and sensitive data is left out, eliminating security risks. As we are only interested in clinical data (medications, diagnoses, test results and medical procedures), unnecessary and uninteresting data are not imported, reducing storage requirements.

The data warehouse is divided into three parts (see figure 1): the data transportation (loading the pool), the pool itself (the database engine, the repository and the search daemon) and the query tools (the front-end). Each part is described respectively:

2.1. Data Transportation

First the data warehouse must be filled with the data. This is done in two steps. The first step is to dump the interesting parts of the data from the operative database into files. The second step is to load these files into the data warehouse. This procedure is automatic and takes place each weekend, when the workload of the mainframe computers is low.

The dumping and loading programs use the repository that states what to dump, where to find it, in what format, etc. When the structure of the operative database is modified, only the specifications in the repository need to be altered, not the programs themselves, making maintenance easy.

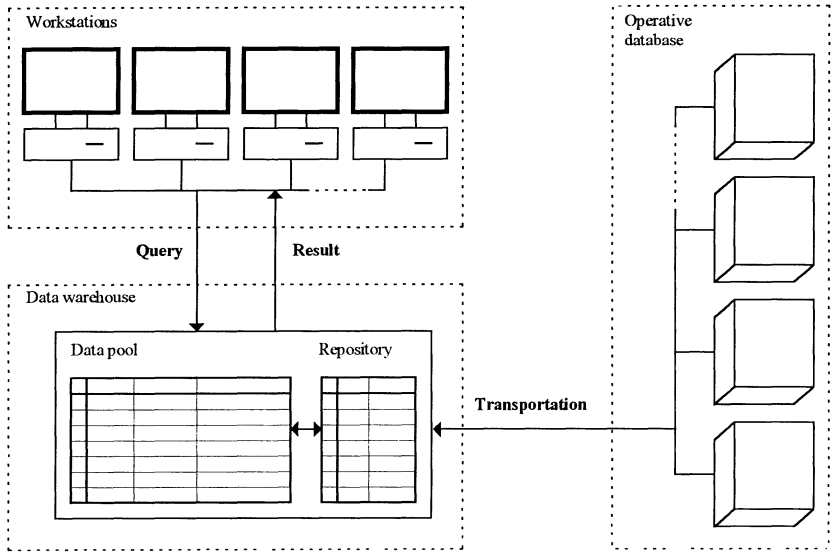

Figure 1. Architecture of data warehouse in relation to user query tools and operative database

2.2. Data Warehouse

The database engine and the repository (the meta data description), are implemented using a Microsoft SQL-server running on a Digital Alpha server under Windows NT. About 4 gigabytes of hard disk space is required for the moment, but the amount will grow steadily. The search daemon and the query front-end are written in the C-language.

2.3. Query Tools

A search is conducted in the following way. By using a query front-end, the researcher can fill in a query himself without having to trouble computer specialists. The query is transferred to the data warehouse. A search daemon recognises the incoming query, which is translated into SQL code. The query is executed as a background process, allowing many queries to be handled simultaneously. When the search is ready, the result can be transferred to the researcher's workstation. This search procedure is very rapid compared to the old methods. If the query is executed when the working day is at its end, the result is ready the next morning.

3. Two Examples Demonstrating the Research Potential of Data Warehouses

A large-scale project of building rule bases for drug-laboratory interferences and drug-drug interactions has been launched in our laboratory. Using these rule bases, the information on drug effects on laboratory tests as well as alarms on potentially serious drug interactions will be automatically available for clinicians. These databases, when included in the hospital's data warehouse, will bring new dimensions to the knowledge and data easily accessible for clinicians.

3.1. Frequency of Drugs Affecting Thyrotropin

The data on drug-laboratory interferences have been collected in some catalogues and databases [4,5]. However, the data have not been properly coded to be used automatically by a computer system. The need for a code was obvious and, therefore, the DLI (drug-laboratory interference) code was developed [6]. The code reveals the most important features of the interference. The actual frequency of medications affecting laboratory tests has seldom been studied epidemiologically. Thyrotropin (TSH) was chosen as an example of a common laboratory test affected by several drugs. By using data warehouse, it was simple to determine how often drugs affecting TSH were prescribed in TUCH during a period of 13 months. The patient population comprised patients treated mostly in internal medicine wards.

A total of 22508 drug events - a drug event is a doctor's decision to start a drug or modify the dosage - were recorded. 2614 (11.6%) of them could have affected TSH. They included 17 different types of drugs. In table 1, the ten most often prescribed drugs affecting TSH are shown.

Table 1. The most often prescribed drugs affecting TSH.

Drug	Direction of effect	No of drug events
Acetylsalicylic acid	Decreasing	1121
Corticosteroids	Decreasing	768
Valproic acid	Increasing	187
Metoclopramide	Increasing	130
Morphine	Decreasing	124
Haloperidol	Increasing	114
Heparin	Decreasing	49
Amiodarone	Increasing	40
Propranolol	Increasing	30
Oestrogens	Contradictory data	15

These results will be used as a basis for further development of the DLI code, e.g. a new character for drug form will be added.

3.2. Drug Interactions in Nephrological Patients

This study on the same material of 22508 drug events was carried out in order to determine how many reminders of drug interactions would actually take place when using the FASS coding for the interactions [7]. The number of potentially serious interactions in the medication of nephrological patients during the 13 months was counted with the help of the data warehouse. The result was 943 drug combination events that would have generated a reminder. When scrutinising the events, it became evident that some alterations to the code are needed before introducing the alarming system to everyday use (the results of further studies will be published in the near future).

4. Discussion

Business companies have adopted the concept of data warehouse as it supports informational processing by providing a solid platform of integrated, cumulative data for analysis. The data warehouse offers the possibility of integration in a world of unintegrated application systems. It also organises and stores the data needed for informational and analytical processing over a long time perspective.

The medical data warehouse has similar architecture than the business data warehouse, but some basic principles are different. Business companies are not usually involved with data protection whereas it has high priority in hospital environment. As a result the design and building of medical data warehouse must be done very carefully with the help of data protection experts and authorities.

Contrary to business applications, the medical data warehouse is not time-critical. Dumping data from operative databases can be made at longer intervals than in typical business applications. Similarly, the few hours' response time in queries is considered acceptable in the medical environment.

The medical data warehouse opens possibilities of e.g. detecting drug interactions, remindering of adverse drug reactions, quality assurance of medical care, better quality of medical information and possibilities of finding new drug effects and interactions [1]. Our examples demonstrate how clinically interesting information can be retrieved with reasonable efforts. Similarly, new technology generates completely new projects by creating new requirements for further development. Examples of new requirements are DLI coding and construction of drug interaction rule base.

According to Lusted [8], medical information can be made more useful by making it more usable. The improved accessibility of information achieved via the data warehouse will boost the research on several health care related fields and undoubtedly make the data more usable. Setting up a data warehouse consumes resources, but in the long run maintenance work will be minimal. The only major investments are the database engine and the hardware. In our opinion the data warehouse is very beneficial for research and it will pay back what is invested in it.

As mentioned earlier, improvements in medical care result from clinical research, and in our case we believe that the use of the data warehouse will speed up the process of adopting research results into practice.

References

[1] Grönroos P, Irjala K, Heiskanen J, Torniainen K, Forsström JJ. Using computerized individual medication data to detect drug effects on clinical laboratory tests. Scand J Clin Lab Invest 1995;55, Suppl. 222:31-36

[2] Codd EF, Codd SB. OLTP (On-Line Analytical Processing) with TM/1. Codd & Associates, 1994

[3] DeJesus EX, Hedberg SR, Watterson K, Krivda CD. Data Mining. Byte, 1995;10:81-102

[4] Young DS. Effects of drugs on clinical laboratory tests. 4nd ed., AACC Press, Washington DC: 1995

[5] Salway JG ed. Drug-Test Interactions Handbook. 1st ed. Chapman and Hall, London 1990

[6] Grönroos P, Irjala K, Forsström JJ. Coding drug effects on laboratory tests for health care information systems. Proc Annu Symp Comput Appl Med Care; 1995: 449-53

[7] Sjöqvist F. FASS `95 Interaction between drugs - systematic review. LINFO Drug Information LTD, Stockholm, 1995

[8] Lusted LB.The Clearing "Haze", a view from my window. Med Decis Making. 1991;2: 76-87