# Chapter 1.21
# Interactive Information Retrieval as a Step Towards Effective Knowledge Management in Healthcare

**Jörg Ontrup**
*Bielefeld University, Germany*

**Helge Ritter**
*Bielefeld University, Germany*

## ABSTRACT

The chapter shows how modern information retrieval methodologies can open up new possibilities to support knowledge management in healthcare. Recent advances in hospital information systems lead to the acquisition of huge quantities of data, often characterized by a high proportion of free narrative text embedded in the electronic health record. We point out how text mining techniques augmented by novel algorithms that combine artificial neural networks for the semantic organization of non-crisp data and hyperbolic geometry for an intuitive navigation in huge data sets can offer efficient tools to make medical knowledge in such data collections more accessible to the medical expert by providing context information and links to knowledge buried in medical literature databases.

## INTRODUCTION

During the last years the development of electronic products has lead to a steadily increasing penetration of computerized devices into our society. Personal computers, mobile computing platforms, and personal digital assistants are becoming omnipresent. With the advance of network aware medical devices and the wireless transmission of health monitoring systems pervasive computing has entered the healthcare domain (Bergeron, 2002).

Consequently, large amounts of electronic data are being gathered and become available online. In order to cope with the sheer quantity, data warehousing systems have been built to store and organize all imaginable kinds of clinical data (Smith & Nelson, 1999).

In addition to numerical or otherwise measurable information, clinical management involves administrative documents, technical documentation regarding diagnosis, surgical procedure, or care. Furthermore, vast amounts of documentation on drugs and their interactions, descriptions of clinical trials or practice guidelines plus a plethora of biomedical research literature are accumulated in various databases (Tange, Hasman, de Vries, Pieter, & Schouten, 1997; Mc Cray & Ide, 2000).

The storage of this exponentially growing amount of data is easy enough: From 1965—when Gordon Moore first formulated his law of exponential growth of transistors per integrated circuit—to today: "Moore's Law" is still valid. Since his law generalizes well to memory technologies, we up to now have been able to cope with the surge of data in terms of storage capacity quite well. The retrieval of data however, is inherently harder to solve. For structured data such as the computer-based patient record, tools for online analytical processing are becoming available. But when searching for medical information in freely formatted text documents, healthcare professionals easily drown in the wealth of information: When using standard search engine technologies to acquire new knowledge, thousands of "relevant" hits to the query string might turn up, whereas only a few ones are valuable within the individual context of the searcher. Efficient searching of literature is therefore a key skill for the practice of evidence-based medicine (Doig & Simpson, 2003).

Consequently, the main objective of this chapter is to discuss recent approaches and current trends of modern information retrieval, how to search very large databases effectively. To this end, we first take a look at the different sources of information a healthcare professional has access to. Since unstructured text documents introduce the most challenges for knowledge acquisition, we will go into more detail on properties of text databases considering MEDLINE as the premier example. We will show how machine learning techniques based on artificial neural networks with their inherent ability for dealing with vague data can be used to create structure on unstructured databases, therefore allowing a more natural way to interact with artificially context-enriched data.

## SOURCES OF INFORMATION IN HEALTHCARE

The advance of affordable mass storage devices encourages the accumulation of a vast amount of healthcare related information. From a technical point of view, medical data can be coarsely divided into structured and unstructured data, which will be illustrated in the following sections.

### Data Warehouses and Clinical Information Systems

One major driving force for the development of information processing systems in healthcare is the goal to establish the computer-based patient record (CPR). Associated with the CPR are patient registration information, such as name, gender, age or lab reports such as blood cell counts, just to name a few. This kind of data is commonly stored in relational databases that impose a high degree of structure on the stored information. Therefore, we call such data "structured." For evidence based medicine the access to clinical information is vital. Therefore, a lot of effort has been put into the goal to establish the computer-based patient record (CPR) as a standard technology in healthcare. In the USA, the Institute of Medicine (1991) defined the CPR as "an electronic patient record that resides in a system specifically designed to support users by