After reading this chapter, you should know the answers to these questions:

• What types of on-line content are available and useful to healthcare professionals?
• What are the three major steps in the information retrieval process?
• How do techniques differ for indexing bibliographic versus full-text information?
• How effectively do searchers utilize retrieval systems?
• What challenges do the Internet and World Wide Web pose for information-retrieval researchers?
• How will changes in technology affect the scientific, economic, and political aspects of medical publishing?

15.1 Evolution of Medical Information Retrieval

Information retrieval (IR) is the science and practice of identification and efficient use of recorded media. Although medical informatics has traditionally concentrated on the retrieval of text from the biomedical literature, the domain over which IR can be applied effectively has broadened considerably with the advent of multimedia publishing and vast storehouses of chemical structures, cartographic materials, gene and protein sequences, video clippings, and a wide range of other digital media of relevance to biomedical education, research, and patient care. As the ease with which information can be recorded and stored in digital form grows, the growth of biomedical knowledge, once spoken of primarily with reference to the biomedical literature, is now equally relevant to the retrieval of information from text within computer-based patient records, documents concerning the administration of medical care, supporting data for biomedical publications, and the literally tens of thousands of electronic-mail messages now sent and stored every day within personal and professional settings.

As has occurred in the the area covered by many chapters in this volume, IR has changed substantially since the first edition (Siegel et al., 1990). The name
change alone (this chapter was previously titled “Bibliographic-Retrieval Systems”) is telling, and the description of new content (full text and hypertext), methods (improved indexing, retrieval, and evaluation techniques), and technologies (CD-ROMs and the World Wide Web) shows that much progress has been made.

Although this chapter focuses on the use of computers to facilitate IR, methods for finding and retrieving information from medical sources have been in existence for over a century. In 1879, Dr. John Shaw Billings created *Index Medicus* to help medical professionals find relevant journal articles (DeBakey, 1991). Journal articles were indexed by author name and subject heading and then were aggregated in bound volumes. A scientist or practitioner seeking an article on a topic could manually search the index for the closest-matching subject heading and then be directed to citations of published articles.

The printed *Index Medicus* served as the main medical IR source until 1966, when the National Library of Medicine (NLM) developed an electronic version, the *Medical Literature Analysis and Retrieval System* (MEDLARS) (Miles, 1982). Because computing power and disk storage were tightly limited, MEDLARS, as well as the subsequent *MEDLARS Online* (Medline), stored only abstracted information from each article, such as author names, article title, journal source, and publication date. In addition, the NLM assigned to each article a number of terms from its Medical Subject Heading (MeSH) thesaurus. Searchers could then retrieve articles on a topic by first identifying the most appropriate MeSH term for the topic, and then retrieving all articles from Medline that were indexed with that term.

As computing power grew and disk storage became less expensive in the 1980s, full-text databases began to emerge. These new databases allowed searching of the entire text of medical documents. Although lacking graphics, images, and tables from the original source, these databases made it possible for users to retrieve the full text of important documents quickly, as well as from remote locations. Advanced information-science methods that had been developed in the 1960s, such as vector-space and probabilistic retrieval techniques, could now be implemented on a large scale (Salton, 1983).

In the early 1990s, the pace of change in the IR field quickened. The advent of the World Wide Web and the exponentially increasing power of computers and networks enabled a world where vast quantities of medical information from multiple sources with various media extensions were now available over the global Internet (Berners-Lee et al., 1994). The result of these dramatic changes is that the scope over which IR technologies are relevant has become almost as broad as the scope of published biomedical thought.

### 15.2 The Information-Retrieval Process

How do we develop a framework for understanding a field that encompasses virtually all biomedical information recorded in digital format? One such framework comes from a modification of ideas advanced by one of the earliest pio-