Data volumes in medical informatics and bioinformatics are being generated faster than researchers can handle using traditional methods of the past century. Thus, new methods, e.g., in data mining and text mining, and in other database searching methods are needed. In the medical informatics field, three methods of data or information retrieval stand out:

1. The **MedBlast system**, making use of BLAST (which will be defined and discussed in Chapter 15), allows researchers and clinicians to search for articles of interest among the plethora of research discovered each month.
2. The **HelpfulMed system** is used in medical informatics to retrieve documents from different databases, which are then clustered using a self-organizing map algorithm.
3. The **NLM’s Visible Human** produces three-dimensional representations of the normal male and female human bodies by obtaining transverse CT, MR, and cryosection images of representative male and female cadavers. The data provide a good test-bed for medical imaging and multimedia processing algorithms. The Visible Human project has been applied to various diagnostic, educational, and research uses.

Data mining has a valuable *predictive* power to enable clinicians to determine, with some measure of accuracy, the proper dosage or treatment protocol. Classification is the most widely used technique in medical data mining, using things like *decision trees* (see one of the Medical Informatics Tutorials included on the DVD accompanying this book for examples, such as Tutorial Z). It is difficult to select in advance which algorithms will be best for a particular study; thus, the authors of this book always recommend doing a “competitive evaluation” of data mining algorithms to see which performs best. For example, Dreiseitl et al. (2001) compare five classification algorithms for the diagnosis of pigmented skin lesions. Their results show that *logistic regression*, *artificial neural networks*, and a *Support Vector Machine* performed comparably, while *k-nearest neighbors* and *decision trees* performed worse. But with other situations, other diseases, decision trees can perform quite adequately (again, see one of the tutorials accompanying this book, on the DVD). Another example is Acir and Guzelis (2004), who applied a Support Vector Machine algorithm in automatic spike signal detection in electroencephalograms (EEGs), which can be used in diagnosing neurological disorders related to epilepsy.

And, finally, a third example is Kandaswamy et al. (2004), who used an artificial neural network to classify lung sound signals into six different categories to assist diagnosis.

The amount of information in medicine/bioinformatics is so vast that clinicians and researchers cannot begin to read all the literature but must have methods of seeking what they need. Google’s Page Rank algorithm can do well in finding popular web sites, but it just cannot meet the need of these clinicians/researchers to find very specific needed information. The following sections describe a couple of examples of document retrieval systems that medical informatics have been using in recent years.

**XplorMed**

XplorMed is a system developed by Perez-Iratxeta and colleagues (Perez-Iratxeta et al., 2001, 2002, 2003) for browsing the MEDLINE literature database. Given a set of MEDLINE