

Record 1	text	d.p.
Record 2	text	d.p.
Record 3	text	d.p.
•		
•		
•		
Record 9998	text	d.p.
Record 9999	text	d.p.
Record 10,000	text	d.p.

Figure 8.2
Sequential file organization.

contained in a document profile, the document is considered found and its text (the content of field 1) is included in the output. Then the program reads the record next in order, performing operations similar to those carried out with the previous record, and this goes on until the end of the file when the last (10,000th) record is read. Then the search is considered over, and the output formed is presented to the user. Whenever the system receives another request, the whole procedure repeats itself in the same manner: again, like in the previous case, the 10,000 records will be read one following another and the query formulation will be compared with the document profile 10,000 times.

Realization of this idea, incidentally, is a simple task for any programmer. For this, the programmer does not need even a cursory acquaintance with the IR system: a basic knowledge of data structures and file organization will be quite sufficient. However, this idea is never implemented in such a simple way because in practice this kind of search would be very slow. Now, before we describe the main methods of speeding up the process of the search, it would be helpful to review the reasons why this search takes so much time.

8.3

Primary and Secondary Storage

The analysis starts with another look at the sequential access search pattern, this time presenting it as a sum of two highly generalized processes. The first process is getting a record from storage (reading a record) and the second one is comparison of the document profile, as it was entered into this record, with the query formulation. The question of storing information has been mentioned several times. What is the actual (physical) place of file storage? The file

is kept in the system's memory or, more precisely, in the memory of a computer. The latter has two memories, however. One is *primary storage* and the other is *secondary storage*. Primary storage is part of the central processing unit (CPU), that is, part of the computer subsystem where an arithmetic or logical information-processing operation is carried out on program instructions. In our case (search), the comparison of document profile and query formulation (logical operation) is performed in the CPU. As a matter of fact, all program instructions are performed in that part of the CPU, which is called the arithmetic and logic unit (A&LU). The information processed according to the program comes into this part of the CPU from primary storage. In our case such information is a record consisting of the text (field 1) and the document profile (field 2). Although any record of any file arrives in the A&LU from primary storage, it is practically impossible to keep the files in the primary storage. In this case the storage of a file should be understood as a situation in which a large amount of "permanent" data needs to be saved for repeated use and updating. Also, the file must remain in memory even though the programs using it are idle, such as when the computer is shut down. We should stress here that a file cannot be kept in primary storage, not because it is difficult or expensive but because primary storage is intended for a different use. It has quite different functions, one of which is to keep ready the file records that are to be processed by the program next, that is, following the record being processed at a given moment. Primary storage is electronic memory consisting of special chip-based electronic circuits, and it is characterized by a high speed of operation and a high cost.

Now let us consider the following question. If a file is not kept in the primary storage, where is it kept? The files are stored in the secondary storage, a medium specifically created with the aim of storing files. Most commonly such storage is a magnetic disk or, simply, disk. The magnetic disk is the most important form of secondary storage today. Disks can take on many different forms, including floppy disks, cartridge disks, and fixed-head disks. The most common type used in mainframe computers is the hard, moving-head disk, and it is this type that we will concentrate on.

The special advantages of the magnetic disk include its ability to store files organized in any of the known methods. This means that when a sequential file is recorded, the information (records) from it will be read only sequentially, that is, from the first record to the last one in order. Whenever the files are organized in some other manner, it is possible to read only one record or a collection of records of that file in any order. In other words, the file organization methods that are different from the sequential method require random access. This is why the disk is often called the *random access device*.

Usually the file stored on the disk is processed as follows. In the program processing the records of this file (a search program in our case), there is a certain instruction (usually called READ) making it possible to read part of the file from the disk into primary storage. When saying "part of the file" we imply that the