of the user appear as input 2 of the IR system. We wrote "as a rule" because in some cases in an IR system, instead of search requests, by which we usually mean the expression by users of their IN in a natural language (see Chapter 2), other forms of retrieval requirements are used. For example, users have to formulate (for input into the system) the retrieval requirement in retrieval language, that is, the expression of an existing IN in an information retrieval language (IRL) different from natural language. The output in an IR system is the set of documents found during the search, which is usually called the *output*.

When we say that a system fulfills its function (or that it functions), we mean that the system (i.e., the complex of interacting elements) carries out a process, as a result of which the goal of creation of the system is achieved with the quality stipulated in the function. In order to see the process of functioning of an IR system, and thus the complex of interacting elements composing it, it is necessary to peek into the so-called black box illustrated in Figure 4.2. Naturally in the process of the functioning of the black box, the transformation of inputs 1 and 2 into the output takes place.

Documents and search requests enter the system through inputs 1 and 2, respectively. Documents entering at input 1 are translated (if this is necessary) into the language of the IR system, that is, into the IRL used for the retrieval. This operation is called *indexing of documents*. The result of the translation of the document into the IRL, which will be used for comparison and selection, is called a *document profile*. Search requests entering at input 2 are also translated into the IRL, and this process is called indexing, that is, *indexing of search requests*. The result obtained in this translation (in the retrieval language) is called *query formulation*. It is clear that the processes of indexing (of documents and search requests) precede the processes of retrieval (i.e., comparison of query formulations with the available document profiles, selection of the document profiles according to some criteria, and the formulation of the output). Thus, processes of indexing and the process of retrieval are integral parts of the functioning of the IR system as represented in Figure 4.2. Moreover, these processes are completely sufficient for realization of a function "at minimum." Therefore we "open" the black box and represent the structure of this system in Figure 4.3.

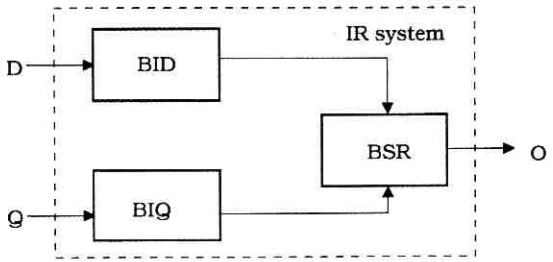

Figure 4.3
Enlarged block diagram of an IR system.

The illustrated structure contains three elements forming the system. The element BID (block of indexing of documents) is used to translate documents from the natural language into the IRL. The element BIQ (block of indexing of search requests) is used to translate search requests into IRL. The results of indexing (translations) enter the element BSR (block of storage and retrieval) in which a search for information is performed. The arrows indicate flows of information circulating in the system (interactions between elements).

One should note that in some cases this structure can be simplified. Although it seems that the diagram is already as simple as possible, it can be simplified even more, such as when any text of a document in a system is considered a document profile. It might seem that this takes place only when a natural language is used as the IRL (which is not feasible for IR systems today and which will be considered in more detail in Chapter 5). However, in practice, within the framework of the IRL of such systems, any text is considered the final set of words from the dictionary of the IRL, where, as a rule, the order of words in the text and punctuation marks are not taken into account. We call such systems *free text searching*. Simplification is also possible when the user sends a query formulation into the system instead of sending a search request. It is clear that in the first case there is no need for the indexing of documents, and in the second case there is no need for the indexing of search requests.

The structure illustrated in Figure 4.3 is realized in many actual systems and is well known in the scientific literature. However, as indicated previously, in the function of this system there are no explicit requirements for the quality of information retrieval. Clearly this does not mean that the creators of systems realizing this function do not care about the quality of satisfaction of the IN.

After the appearance of the first IR systems, the euphoria caused by the ability to use the computer in the process of information retrieval disappeared rather quickly. From the beginning there was a general tendency to improve the quality of the service to users. This led to the introduction of various parameters by which quality was evaluated, and improvement of these parameters became one of the most important concerns of investigators. Much work has been done in recent decades aimed at improving IRL, developing better methods of indexing documents and search requests, and finding more efficient methods of storage and information retrieval. This work provided positive results, and in due time we may expect further improvement in these various aspects. However, the general structure of an IR system as illustrated in Figure 4.3, in principle, is not able to give high-quality results in satisfying POIN, because it is unable to take into account the extremely important properties of POIN. In other words, it is not sufficient to create the high-quality elements of a system. It is necessary for the system's structure itself (i.e., the set of interrelated elements) to correspond to requirements of providing quality service to users. It is clear then that a function that takes into account the quality of satisfaction of POIN must be realized. Such a function was formulated earlier. The requirement contained in it for carrying out an optimal retrieval is the requirement of taking into account the