

due to the explosion of information science today we in fact have a set of methods and algorithms for all processes participating in retrieval. But the existence of these methods is still insufficient for creating an "optimal" IR system. It is also necessary to know how to choose the best of the available methods or set of methods. This problem was solved at the beginning of the 1980s when a simple criterion was developed for evaluating different methods in a search process. Hence, it is possible to create IR systems that realize optimal search. We can formulate the function of a documentary IR system as follows: *the function of a documentary information retrieval system is to fulfill an optimal, from the user's point of view, retrieval of information to satisfy this user's POIN with any information about the user's POIN given to the system* (Voiskunskii & Frants, 1974; Frants, 1986).

It should be pointed out that the requirement about "any information" about the user's POIN is a consideration of attribute 2 of POIN (Chapter 2); in other words, this means that regardless of how well the request is stated, the user must obtain optimal output from the system. Attribute 4 is also taken into consideration because the optimization will require a feedback process and the change in the user's POIN will be accounted for. The function also provides for the highest quality output because only the best output (out of all possible outputs of a system) is given to each user. In Chapter 9 we will consider in detail how this is carried out in the description of the mechanism realizing the optimization.

Having defined the function of a documentary IR system, we can consider its structure. In the following section we will determine what kind of system structure enables realization of the formulated function. In this connection, we will use the term "information" as it is accepted in the systems approach.

4.5

Structure of a Documentary IR System

In considering the general rules for creating systems, it was suggested that to define the structure of a system we in essence provide a set of elements necessary to fulfill the formulated function and establish an interconnection between them. From the definition of the concept "system" it follows that only those elements that are essential to fulfilling the given function must enter into this set. But how do we choose those elements that are needed for realization of the function? First, we note that all of the known functions of an IR system contain the requirement of realizing the process of information retrieval. Moreover, as indicated previously, a function formulated "at minimum" contains only this requirement. Therefore, at the beginning it is expedient to consider the structure of a system fulfilling a "minimal" function, that is, a system carrying out the retrieval.

A general scheme of any retrieval was considered in detail in Chapter 3

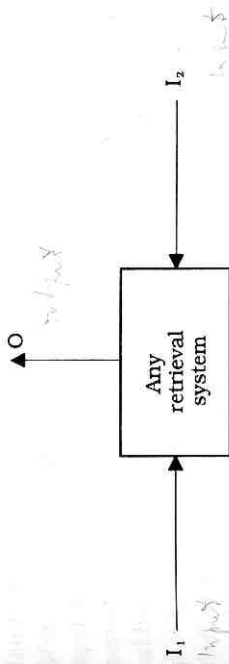


Figure 4.1
System for the retrieval of objects.

and is graphically represented in Figure 3.1. One should especially emphasize that, independent of what we are in reality seeking, in any retrieval system there must be two inputs from the external environment; in fact, the system must admit objects of retrieval (input 1) and the requirements for retrieval (input 2). The output of such a system consist of the objects found (chosen) in the process of retrieval. For example, in a retrieval system realized for selecting eggs of specific size on a bird farm, one input consists of the eggs collected at the farm (objects of retrieval) and the second input includes instructions on their size (retrieval requirements). The output in this system will be some set of selected eggs whose size corresponds to the given instructions. Thus, it is convenient to represent any retrieval system graphically in the following form (see Figure 4.1).

On the other hand, it is not difficult to see that this representation of a retrieval system on the whole is identical to the block diagram of the general process of retrieval given in Figure 3.1, and in essence it is itself a "black box," with two inputs and one output. Because in this book we are interested specifically in information retrieval, further discussions of this "black box" will exemplify only IR systems. Such a system is represented in Figure 4.2 in somewhat more standard (for a "black box") form.

We will consider Figure 4.2 in somewhat more detail. In analyzing the process of information retrieval (Chapter 3), we showed that the object of retrieval in an IR system is information contained in a document. Consequently, documents appear as input 1 of this system. We also stated that the requirement for such a retrieval is formulated by a user and represents information about the IN of the user as he or she determined it to be. Thus, as a rule, search requests

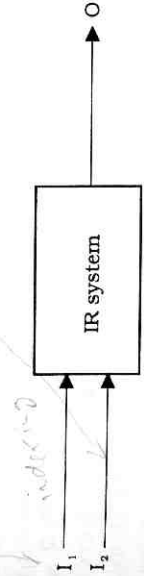


Figure 4.2
IR system in the form of a "black box."