4

Concept of an Information Retrieval System

4.1
Introduction

Chapter 3 showed that the current information crisis is caused by the problems associated with the search for information. As the text explained, these problems are not solvable within the framework of traditional library forms of retrieval (and this is the essence of the present crisis). The development of information science was society's reaction to the crisis occurring within the sphere of informational activity.

Practically at once with the rise of information science appeared a principally new form of information retrieval, the information retrieval system (IR system). Today, hopes for a successful solution to many of the existing problems associated with satisfying information need (IN) are connected with IR systems. Within the framework of information science, an extensive literature dedicated to various aspects of creating and using the IR system has appeared. Suffice it to say that in the opinion of many investigators, today about 30% of all scientific publications in information science are directly related to IR systems (Jarvelin & Vakkari, 1993). However, in spite of the fact that "IR system" has become a common and widely used term, a rigorous, unambiguous, and widely accepted concept of it does not exist, and in various publications this concept acquires different shades of content as it correspondingly changes its scope. Such a situation in some cases not only prevents mutual understanding among the developers of IR systems but also leads to some dubious solutions in their design. Therefore in this chapter we will introduce the concept of an IR system that will be used consistently in subsequent chapters.

4.2
IR System as Object

Do a camel and an ocean liner have anything common? The question seems to be a joke for which there is an original and witty answer. However, if

we place both of these objects into a certain class of transportation means, then their similarity not only does not seem strange, but it seems trivial and ordinary. A horse, a camel, an automobile, and an airplane all are clear and obvious examples of objects from this "transportation means" class. Moreover, we are certain that almost any reader can without difficulty continue this list.

Each object in the class "transportation means" has its own "name" (for example, "camel," "automobile"). If we tell someone that we used a transportation means and do not specify the name of this means, then it is not obvious to the listener which object of the class "transportation means" we used. In other words, the concept "transportation means" is collective and characterizes a whole set of objects.

This discussion of "transportation means" is important and useful for understanding some characteristics associated with the definition of the concept "IR system." The problem in understanding the meaning of the term "IR system" is based on the disagreement among different investigators: some suggest that this term denotes an object, whereas others suggest that it is intrinsic only to a class of objects, not to some concrete object.

Keep in mind the fact that the term "IR system" arose about 40 years ago, and the term was used to designate an object in which the process of seeking information through comparison and selection was carried forth without the direct participation of a person but instead was, as a rule, realized on a computer. Nevertheless some investigators rather quickly began to use the term "IR system" to designate a *class* of objects. Their position was formulated as follows: "Any system that is designed to facilitate literature searching activity may legitimately be called an information retrieval system. The subject catalog of a library is one type. So is a printed subject index" (Lancaster, 1979). In some cases the list of examples of an IR system is supplemented by such objects as a thesaurus, a telephone book, and the universal decimal classification. It is clear that under this approach an IR system appears in the role of "transportation means," that is, in the capacity of some class represented by objects created to facilitate the process of information retrieval.

Such an approach is completely legitimate, for example, in discussing the goal of different means used within the framework of informational activity. However, we note that the listed objects, for example, the traditional (manual) subject catalog in a library or a printed subject index, in themselves are not in a position to carry out a search, that is, to conduct a comparison and selection of information. They are only auxiliary means that a person might use in the process of performing comparison and selection. As a rule, these means are used when the user is attempting to retrieve the information on his or her own. In other words, *this form of the term "IR system" does not require that the system retrieve the information on its own*. In addition, it seems that objects included in the class of such IR systems and having the IR system's commonly accepted (and as a rule useful) name—such as a thesaurus—are called systems on the basis that they are