mary consideration in this intricate labyrinth of problems. To this end, let us look at what is to be evaluated; in other words, what will be our primary concentration when evaluating the results of information retrieval and the IR systems performing the search?

10.2

Aspects of Evaluating

We stated in Chapter 2, when considering the information need (IN), that any information process is generated by IN and is meaningful only in connection with its satisfaction. It is quite evident that information search is one of such processes, and an analysis of the present-day information crisis (see Chapter 3) shows that this process is of the utmost importance. Recall, that various types of IN exist and that they are characterized by a certain unique set of inherent properties (Frants & Brush, 1988). This is important in that, as stated earlier, to satisfy diverse types of IN, various types of systems must be created, respectively implementing different information search processes, that is, systems capable of taking into account sets of properties inherent to a particular IN type are needed. This is what determines both structural and design differences between IR systems to be created. However, is it necessary to use different means to evaluate searches carried out in various types of systems? In other words, do IN properties affect information search evaluation? As an example, let us consider IN types such as POIN and CIN, which we have singled out previously.

As discussed in Chapter 4, IR systems that have been developed to satisfy CIN are usually called factographic information retrieval systems, whereas IR systems created to satisfy POIN are called document retrieval systems (or sometimes text retrieval systems). As we also discussed earlier, in a case of a documentary IR system, for instance, such an important parameter as recall is used for search evaluation. However, it may be unused (superfluous) for evaluating factographic retrieval. Instead, we will use, for example, the "search success" concept: the search was successful if the required information was found; otherwise, the search was unsuccessful. (Generally, it is inessential in this case how many documents containing this information will be selected, whether one or a hundred documents. In our opinion, one document is even preferable.) This is apparent from attribute 3 for POIN and attribute 3 for CIN (see Figure 2.2). Thus, quite evidently, in information search evaluation one must take into account the type of IN to be satisfied by the search. Because in this book we consider only POIN satisfaction and only those IR systems that realize the information search for POIN satisfaction, we will discuss the evaluation of the information search for POIN satisfaction, that is, the document information search.

Regarding realization (organization) of the information search, it should be noted that it is also not quite singular. At present, there are two major ap-

10

Evaluation of Search Results

10.1

Introduction

As was shown in the previous chapter, an IR system construction is completed after development of adaptation and optimization mechanisms, that is, after creation of the last element of the IR system structure. It should be emphasized that the construction discussed in Chapters 6 to 9 implements the structure of the fully automated IR system performing the function defined for this system in Chapter 4. Now it would seem that the last stage is to implement the developed construction (in this case, algorithms), and this step is more technological than research oriented. However, implementation is not the last stage. When describing the system approach, we demonstrated that yet another stage exists that is very important for the creation of a system; that stage is evaluating what has been created. In other words, one must be able to evaluate how the construction of the developed system fulfills the defined function. This is important not only for better understanding the performance of the created system, but also for improving it in the future. Thus, it would be impossible to create and develop information retrieval systems and keep them in service unless evaluation is made both of the information search process itself and of the systems implementing this process.

This fact was recognized long ago in information science, as witnessed by the number of publications that concern these problems. These publications discuss a wide variety of topics related to evaluation, including the indexing languages used, the indexer's skill, indexing time, the mean number of terms assigned to documents and queries, query form, output form, file size, recall and precision of the search, query processing time, cost of the search, economic efficiency, functional efficiency, and the measurement precision of individual parameters. The variety of problems illustrates the complexity of the task as well as the lack of a uniform understanding of what must be evaluated and how the evaluation must be done. It is impractical, within the scope of this book, to discuss even briefly all facets of evaluation. Therefore, we must define our pri-