the constructive features of the system's every element, of its possibilities, of the range of its statistical characteristics, and so on. Also, the experience searchers have in the construction of query formulations (query indexing)—which is their main activity—is much broader and diverse than that gained by the users. Probably, at a certain stage of the development of IR systems, the advantages enjoyed by intermediary searchers outweighed all of the mentioned disadvantages. At any rate, these advantages can be viewed as the main components of the best results achieved by an intermediary searcher. In many instances, these "best results" did not differ much from the results obtained by users. Also, the available experimental data were known mostly to researchers, whereas users, dealing with concrete results of a search done without their involvement, were often dissatisfied with its quality.

The emergence of on-line systems substantially decreased the number of users dissatisfied with the system's performance. This decrease was due, among other things, to an access to the system gained by a user who became actively involved in the search process. In other words, the user could construct his or her own query formulation. The quality of the query formulations constructed by users began to climb steadily and by the late 1970s researchers noticed that in a number of instances the results achieved by the users exceeded the results attained by intermediary searchers. This was something to be expected, because the introduction of a dialogue brought certain conveniences into the query formulation procedure. Moreover, the user obtained access to some internal information, such as the occurrence frequency of a specific term in the collection of documents and information on the number of documents retrieved at different stages of the search. Methods of using a thesaurus in the dialogue mode were developed in a number of systems; at the user's request some thesaurus fragments, which can be useful for creating query formulations, were put on the screen. Also, many years of research provided the user with helpful information that enabled him or her to adapt to a specific type of IR system. This largely accounts for the fact that at the present time in most of the functioning IR systems, query formulations are constructed by users. What is more, users today tend to give thoughtful considerations to methods for improving query formulations; that is, they assume the function of researchers in information retrieval. Publications written by users and mainly addressed to other users have also appeared; in them, users use their own experience to give useful advice (methodological recommendations) on how to construct query formulations that favorably compare to those formulated by experts in information science. In 1992, for instance, Goldman, an engineer and manager at Bell Northern Research, published a book with an indicative title *Online Information Hunting*, which gives an account of his search for his own information needs (Goldman, 1992). He is certain that "the retrieval of information by an intermediary searcher is, in principle, impossible unless the search is controlled by the subject expert himself." Goldman continued, "It looks like the classical information gathering intermediary . . . is going to be included in the list of endangered species in the very

near future." Goldman's recommendations mainly replicated those contained in commonly known methodological manuals. He set forth a number of steps, such as how to organize a search strategy, input the first version, proceed with the initial on-line adaptation, and organize a final search version, citing examples of the query formulations he obtained. Yet Goldman did not go so far as to describe a concrete method of constructing a query formulation. As for the terms to be used in formulating disjunctive normal form, his major recommendation is to "think carefully" about which of the terms are likely to be more successful for the search. Incidentally, in each of the seven methodologies we examined, we found, in one way or another, the same recommendation.

Be that as it may, today many users can construct query formulations that are no worse than those constructed by intermediary searchers. However, the acuteness of the problem remains. Researchers think, like before, that it is one of the bottlenecks in the functioning of the IR system. Moreover, the widespread use of on-line IR systems prompted the developers of the systems to find methods of helping users to construct query formulations. This problem of interaction between IR systems and users is one of the most important in developing IR systems. It mainly consists of answering the following two questions:

1. What is the best way to help users construct better query formulations?
2. What is the best way to simplify the interaction between the user and the IR system in the process of constructing a query formulation (and in correcting it)?

It is important to observe that these two goals are at odds with each other. As mentioned earlier, to improve the quality of a query formulation the system typically has to provide the user with information about possible formulations, different statistics, descriptor relations, and so forth. This would force the user to spend a lot of time learning the system's internal characteristics and analyzing the information provided by the system on the screen. All this would clearly increase the complexity of the user's interaction with the IR system.

The problem of constructing a query formulation remains one of the most important problems in information retrieval, primarily because the process is performed by humans. Indeed, many papers dealing with this problem express this opinion. It is probably the reason for the growing interest in the automatic construction of query formulations. Next we focus our attention on this process.

7.3
Approaches to the Automatic Indexing of Search Requests

The creation of a method (algorithm) for the automatic indexing of queries is in some sense the creation of an automatic intermediary searcher. In other words, the creation of an algorithm is the creation of a model of the process