principle to automation. Moreover, the advantages of automatic methods are so evident today that there is practically no need to advocate them. So what is the problem? Perhaps, the main reason is as follows.

Even a cursory analysis is sufficient to show that the challenge encountered in constructing disjunctive normal forms or sets consisting of numerous unordered sets of descriptors is far greater than that of constructing only one unordered set of descriptors for each document (i.e., construction of the document profile). This challenge is also reflected in the existing approaches to automated indexing. Indeed, the idea of a word-for-word translation, which underlies the automatic indexing of documents, is not only simple but is also very intelligible to people with a different level of training.

On the other hand, the idea of the automatic construction of query formulations (i.e., the automatic indexing of queries) underlying the algorithm that will be described in this chapter is not always easy to comprehend (though we do not consider it complex), as we have often seen in practice. This conclusion is supported by the fact that the method used for the automatic indexing of documents, advanced for the first time by Luhn in the late 1950s (Luhn, 1957) was immediately accepted by researchers and was followed by a series of relevant publications in the early 1960s. But the method and the algorithm for the automatic indexing of queries for the Boolean search systems (i.e., the method of the automatic construction of query formulations in Boolean form) published for the first time in 1970 (see Frants, Voiskunski, & Frants, 1970) remained largely unnoticed for a quite a long time, despite the importance (stressed time and again by most researchers) of reducing labor intensiveness and raising the quality of the construction of query formulations (see, for example, Croft & Thompson, 1987; Lancaster, 1968; Meadow, 1967; and Spink, 1994).

Already in the 1960s, the construction of query formulations became one of the most important problems in designing IR systems. The importance of the problem was probably first emphasized by Lancaster (1968) who, while analyzing the reasons for an unsuccessful search in the IR system MEDLARS, concluded that the main reason for unsuccessful searches in IR systems is the low quality of query formulations. Despite this observation, far fewer publications deal with search request indexing than with document indexing. Though the situation has changed somewhat in the past few years (the researchers' interest has shifted toward the construction of query formulations), the automation of this process gets insufficient attention even on the part of researchers.

The point is that the automation of any intellectual activity—and it is precisely this kind of activity that we have to deal with when analyzing the existing practice of formulating disjunctive normal forms—is traditionally one of the most complex automation problems, and this complexity scares away many researchers. Some authors believe that without new important breakthroughs in the artificial intelligence area, it is pointless to try to automate the construction

of query formulations. But the solution to the problem requires intellectual creativity only as long as there is no algorithm to solve it.

For example, the problem of finding the greatest common divider required a substantial intellectual effort before Euclid found an algorithm (Euclidean algorithm). Similarly, some problems can be intellectually challenging only because there are no algorithms for them. It should be noted, however, that the created algorithms do not necessarily have to copy intellectual processes. It is sufficient to make an algorithm as a working model of this activity; whether the model is close to the original is of secondary importance. These considerations underlie the available solutions.

To better understand the problems that occur in constructing query formulations, we will first discuss the most common features of the existing methods for obtaining query formulations. In most of the systems known to us, query formulations are obtained in one of the following two ways:

1. By the user, if the user expresses his or her POIN in the IRL.
2. By the system, if the user expresses his or her POIN in some language other than the IRL and the system translates his or her expression into the IRL.

The second method prevailed until the mid-1970s, as query formulations were being obtained primarily by the most qualified experts of the system (commonly known as intermediary searchers). With the emergence of on-line search, that is, when a dialogue was introduced into the information retrieval process, the construction of query formulations in an ever-increasing number of functioning systems was entrusted to the users themselves, and today this method appears to be more popular. (In this case, too, breakthroughs in computer science have had a marked impact on IR system progress).

Earlier we mentioned the wide use of the manual construction of query formulations in functioning IR systems. Clearly, when developers presuppose that a manual method will be used (incidentally it makes no difference who is going to implement it—the user or an intermediary searcher), they relieve themselves of the fairly labor-intensive and complex job of either implementing some existing algorithm of query indexing or developing a new algorithm. In this case, the developers usually limit themselves to providing some methodological recommendations. According to some experts even this is not necessary because in their opinion the offered recommendations are intended to clarify the problem facing an inexperienced searcher rather than to help in the construction of query formulations. In any case, we can say that an experienced intermediary searcher does not need a methodological manual to construct a query formulation. In a certain sense, the available methodologies are similar to those used when translating from one natural language into some other natural language, from English into Russian, for example. The regulatory part in such