ously. It follows that only one realization of each procedure (participating in the creation of the document profiles) can be used; that is, if there are several realizations of a procedure, then the most preferred one should always be selected (unlike, say, a case in which there are several different creation schemes of a document search variant). Because fewer schemes are involved, they may all be used, thus eliminating the necessity for selection (for more detail, see, e.g., Voiskunskii, 1985). It is the necessity of selecting the preferred realization of a corresponding procedure that highlights the importance of its evaluation.

11.3
Averaging Values of Functional Effectiveness

At the beginning of the chapter, we stressed that our interest is in an evaluation of macroevaluated objects that could allow us to predict search results for a newly entered search request when such a search is executed with the help of or within the scope of the object under evaluation. Approaches to performing the required evaluation, as we also pointed out, are based primarily on averaging values of functional effectiveness attained in a specially organized search series. Different ways of averaging the mentioned values are known (see, e.g., Lakhuti, 1971; van Rijsbergen, 1979; and Salton, 1975). But we believe that in practice, within the scope of the evaluation concerned, it makes sense to apply only one of them, namely, the calculation of arithmetic means of existing values. This is dictated by three factors that favor the proposed approach: (1) the simplicity of such a method, (2) the reliability of such a method, and (3) a well-known "predictive" value of arithmetic means. Considering these advantages, as well as the fact that other ways of averaging values of functional effectiveness have been discussed in sufficient detail in information science literature (see, e.g., van Rijsbergen, 1979; and Salton, 1979), we believe there is little reason to address them in this paper once more. Therefore throughout this paper, where we are dealing with averaging values of functional effectiveness, calculation of their arithmetic means is assumed.

Recall that technically this method is performed in the following manner. Let F^i be the values of arbitrary search characteristic F obtained during series of searches performed in a certain collection of documents for each search request from a given set, and let m represent the number of search requests in such a set. Then

$$F^{ave} = \frac{\sum_{i=1}^{m} F^i}{m}.$$

Regarding the averaging of values of functional effectiveness, it would be useful to discuss the problem of organizing a series of searches from which the values

are obtained (we are interested in a series of searches that would allow reliable prediction based on the mentioned averaging of values). However, we will not consider this problem for the following reason: it has been discussed, as a rule, by any researcher dealing with issues of evaluating information retrieval systems (see, e.g., Cleverdon & Keen, 1966; Lakhuti, 1971; and Salton, 1968). As a result, there are established traditions in information science on how to organize such a series of searches. The researchers do not doubt that when following the established traditions the results obtained in performing evaluation can be safely trusted. Thus, instead of including a routine discussion of this problem (and obtaining the same conclusions once more!), it seems more appropriate in our opinion to simply present the established traditions in an explicit form.

These traditions assume that if the search whose results are predicted is executed in a retrospective monotopic search collection, then the series of searches should be performed in a random subcollection of this search collection containing 2000 to 4000 documents. As for the search request to be used in the series of searches, the search should be based on 20 to 50 search requests randomly chosen from a set of those search requests that arrived for performing the search in the mentioned search collection. Note that if a search whose results are predicted is performed in other search collections, then the established traditions are just a modification of the ones given here. For example, a "predictable" search is executed in the selective dissemination of information (SDI) mode in monotopic collections of current acquisitions. It is clear that each of these collections may be considered as a random sample from a retrospective search collection representing a set of collections of current acquisitions received over some period. Thus, the series of searches should be executed in a subcollection (of this retrospective collection), which is a union of current acquisitions collections (from the mentioned time period) that contains 2000 to 4000 documents. Further, in discussing problems related to the evaluation of macroevaluated objects, it will be assumed that the series of searches involved is organized in accordance with established traditions (which essentially have been presented here).

11.4
Requirements for Search Characteristics

Another problem related to performing an evaluation of macroevaluated objects is the lack of a clear view as to which search characteristics' averaged values should be used in the mentioned evaluation as well as when such an evaluation can be trusted and when it cannot. It is clear that we are dealing here only with those search characteristics that are used in evaluating the functional effectiveness of a document search and primarily with complex search characteristics. Recall from Chapter 10 that the functional effectiveness of a document