In cases of large retrospective collection, it is clearly unrealistic. Of course, a user can be "talked into" perusing a document collection of acceptable size. However, this can be done only in exceptional (experimental) situations. That is why development of methods to determine recall has long been underway within the information science framework with the aim of producing ways to attain adequately accurate values of this search characteristic with acceptable labor expense. Such methods have been suggested. We will discuss three varieties of these methods that, in our view, provide a full picture of the situation of recall determination in information science.

One method assumes that one can get an adequately accurate value of recall by determining the number of pertinent documents in a representative subcollection and by extrapolating this result to the entire search collection. Technically this can be done, for example, as follows. A random subcollection of, say, 3000 documents is given to the user for evaluation for pertinence. Let us assume that the user finds 20 pertinent documents. Then, if the entire collection contains 30,000 documents, it is assumed that the number of pertinent documents in the collection is 200.

Evidently, this method to determine recall employs the "selective approach" well known in mathematical statistics. Therefore, all the results needed to build the methods are borrowed from it, particularly the ways to specify the required size of the representative subcollection, ways of determining the precision of the results produced, and so forth. For this reason, we will not dwell on these problems, and we will note only that quite a large representative collection may be needed to obtain the required precision of recall. This, in its turn, may create problems resulting from the user's participation in this process.

Another method used to determine the recall is based on the assumption that if all document profiles in the collection are obtained by the same method, and the search is based on the same query formulation and uses same output criterion, the probability of getting a pertinent document in an output is equal for every pertinent document in the collection and has the same value as the recall. Some considerations on the justification of using such methods can be found in Frants, Voiskunskii, & Frants (1970). As an illustration of these methods, we will consider one method suggested in Frants, Voiskunskii, Frants.

Under this method, before the search based on some search request is conducted, the user requesting the search is asked to select and give to the system N_1 pertinent documents from, for example, the user's personal file. The document profiles for these documents are obtained by the method adopted in the IR system and are added to the collection, and then the search is performed in this expanded collection. Naturally, the pertinent documents given by the user are excluded from the produced output. Denote the number of excluded documents by n. Then the recall is determined (within the method under discussion) by the formula n/N_1.

It must be stressed that this application and similar methods for determining recall involve a number of problems, including theoretical ones that have yet

to be solved. The most urgent of them, in our opinion, are those for evaluating the precision of the methods under discussion and the problem of finding such value of N_1, which would ensure determination of the recall with an acceptable precision.

Finally, the third variety of the methods to determine recall is based on the assumption that if the search is performed using several query formulations constructed for the same search request and the pertinent documents are combined from all outputs, then the number of documents in this set will be close to the number of actual pertinent documents in the collection. The technical aspects of these methods are self-evident; however, it must be recognized that the theoretical side of applicability of this variety of methods has the least foundation.

10.6
Construction of Complex Search Characteristics

We have just discussed certain problems involving use of a CSC in evaluating the functional effectiveness of a document search. Here we will consider their construction. First, we should note that quite a few complex search characteristics are described in the literature (see, for example, Bollmann, 1977; Bollmann, Raghavan, Jung, & Shu, 1992; Borko, 1967; Cleverdon, 1970; Cooper, 1973; Kraft & Bookstein, 1978; Lancaster, 1979; Raghavan, Bollmann, & Jung, 1989; Salton, 1975; Saracevic, 1995; Shaw, 1995; Sparck-Jones, 1980; and Voiskunskii, 1984), and at present two basic approaches have emerged in information science for the construction of complex search characteristics. One of the approaches can be called, with reservations, the empirical one, and the other can be called the theoretical one.

In the case of the empirical approach, a researcher proposes some new CSC, relying on accumulated experience and intuition. Naturally, a thorough experiment check of the proposed CSC is necessary. If testing reveals that the characteristic is pragmatically justified, it should be recommended for use in the information service. Such a CSC may find a broad acceptance at a later date, as is the case for complex search characteristic $I_1 = R + P$.

The theoretical approach to CSC construction is usually reduced to building a formal model for presentation and evaluation of the search results. In this, it is proposed that the complex search characteristics are based on the tools that are used for search results evaluation within these models. Clearly, not every model of this kind would lead to the construction of a pragmatically justified CSC. In our view, the best chances of getting such a CSC are with the models whose content justification is sufficiently evident, at least in cases of document search where the point corresponding to the achieved pair "recall–precision" belongs to the determining square (see Figure 10.3). We will discuss some of these models that, in our view, are best suited to the preceding consideration.