

the respective CSC would make it possible to draw conclusions similar to those made by specialists in the course of by-content evaluation of this effectiveness. Because by-content evaluation is based on a position determined by the totality of the content criteria, one must use the same totality of the criteria in deciding on the justifiable use of the specific CSC in a concrete case. Note that different combinations of such criteria may lead to different solutions to this question with the same CSC and in the same situations. We will follow the totality of content criteria that has been formed in a search situation where the search is related to obtaining some new result (a new method, a new approach, a new solution, etc.) in the process of scientific activity. This is done primarily because this situation is one of the most common. (A detailed analysis of search situations can be found in Chapter 12.)

It is not realistic to provide the preceding criteria for all possible cases of evaluation, especially considering the fact that researchers are not always in agreement about all the details of these criteria. However, for a relatively broad range of cases determined by the preceding situation we will provide a number of such criteria, particularly those that, in our view, fit the concept of the discussed criteria on which a majority of experts agree. In fact, many experts have developed a pretty good idea (at least intuitively) of the content criteria evaluation, which is derived from the situation mentioned. These criteria are as follows:

1. The closer the search output is to the "ideal" (containing all and only pertinent documents from the collection), the higher is the functional effectiveness of document search.
2. Recall and precision of the search are equally important (they mutually complement each other), at least in cases of a search producing recall and precision levels of 0.5 and above.
3. When a document search produces recall and precision levels other than those stipulated by the previous criterion, such search characteristics may be found to be substantially unequal. For instance, when the search produces 0.5 or higher precision with a very low recall, the recall would be much more important.

The formulated criteria, in our opinion, provide a full enough picture of the position from which document search functional effectiveness is commonly evaluated. It is on these criteria that a person usually bases his or her by-content evaluation of functional effectiveness.

Further, we will note an important fact: the content criteria, in a great number of cases, leads to a negative resolution of the validity of using the CSC, thus imposing certain limitations on using the formal method of the functional effectiveness evaluation. To clarify this note, we will use the example of the most common search characteristic,  $I_1 = R + P$ . We will assume that three different

searches for the same search request in a 10,000 document collection with 100 pertinent documents produced outputs that included the following:

1. 50 pertinent and 50 nonpertinent documents
2. 40 pertinent and 27 nonpertinent documents
3. One and only one pertinent document

Clearly, all three outputs are admissible. It seems reasonable to consider the functional effectiveness of the searches resulting in the first and second outputs as admissible while considering that leading to the third output as inadmissible. However, if in these cases the functional effectiveness is evaluated on the basis of achieved values of the complex search characteristic  $I_1$ , we find that the effectiveness in all three cases is close ( $I_1 = 1$ ,  $I_2 \approx 0.997$  and  $I_3 = 1.01$ ), and, moreover, the highest functional effectiveness is reached in the third case. It means that whichever formal criterion (enabling a functional effectiveness evaluation based on the produced values of  $I_1$ ) is used in this situation, in these cases it will not ensure a pragmatically justified evaluation of the functional effectiveness. For example, if one uses a "natural" criterion, which presupposes that the functional effectiveness is acceptable with  $I_1$  values close to 1, then evaluating this effectiveness in cases of the searches producing the first and second outputs will be seen as pragmatically justified whereas in the case of the third output it will not be seen as such. (In the following we will assume that document search functional effectiveness is evaluated on the basis of "natural" formal criteria.)

The previous example illustrates the rather serious problem involved in the use of the formal method of functional effectiveness evaluation. The seriousness of the problem is clarified by the fact that the book's authors are not aware of complex search characteristics (probably because there are none and unlikely to appear) that can serve as the basis for a pragmatically justified evaluation of the functional effectiveness in every "admissible" situation, that is, when evaluation is being made of the functional effectiveness of a document search that resulted in admissible outputs.

## 10.4

### Limits of Applicability of Complex Search Characteristics

It follows that to apply any complex search characteristics to an evaluation of the functional effectiveness of a document search, one must have an idea for which set of outputs it is justifiable to use this characteristic. We will call this set of outputs the *domain of objects* for which the complex search characteristic is applicable. Determining the limits of such a domain is one of the important questions discussed in this book.