10.11 Conclusion

The problems that arise when evaluating information retrieval form a complex labyrinth, and in this chapter we analyzed our place in this labyrinth. As the result of this analysis, the following was determined:

- Only documentary search is being evaluated.
- The functional effectiveness of documentary search is being evaluated.
- As the result of this search, the search collection is divided into two subsets: the documents corresponding to the search request and the documents not corresponding to the search request.
- In analyzing documents, binary scale will be used (unless stated otherwise) describing whether the documents are pertinent or nonpertinent

The evaluation of the functional effectiveness of a documentary search assumes two methods: the formal method and the by-content method. Both methods were considered in the this chapter, but we stressed the first method. This method is based on using search characteristics of type $I_1 = R + P$, which were called complex search characteristics. For the complex search characteristic (CSC) I_1, we gave an example illustrating that this CSC would not allow a pragmatically justified evaluation of functional effectiveness in all situations. In extending this example, we formulated a more general result; the authors are not aware of any complex search characteristics that would allow for a pragmatically justified evaluation of the functional effectiveness in any "acceptable" situation, that is, that would allow one to evaluate the functional effectiveness of a documentary search, with any acceptable output (such search characteristics probably do not exist and are not likely to be developed).

That is why the problem of determining the boundaries of the applicability domain of complex search characteristics became one of our key concerns in this chapter. In this connection we presented special methods for solving this problem. Using these methods, domains of the applicability of complex search characteristics $I_1 = R + P$ and $I_2 = \sqrt{R \cdot P}$ were determined. Analysis of the domain of applicability of complex search characteristic I_2 led to a very important conclusion: evaluation of functional effectiveness based on this characteristic will be pragmatically justified in the case of any search whose precision is not less than 0.5, that is, in fact, in any search in contemporary information retrieval systems. At the same time, to determine in general if it is pragmatically justified to evaluate functional effectiveness in a given situation on the basis of some CSC, it is necessary to know recall level (in addition to precision level). Therefore, we also considered methods for determining recall level.

A large part of this chapter was devoted to constructing complex search characteristics. We considered several formal models that would allow one to construct such characteristics. One of these models led to a series of CSCs with the order preservation property. The existence of this property in some CSCs allows for a significant decrease in the required effort (to an acceptable level!) in comparison with a typical effort required for evaluating functional effectiveness. It was shown that, in particular, CSC $I_2 = \sqrt{R \cdot P}$ has an order preservation property. Taking this into account and using the analysis of the domain of applicability, it could be stated that the given CSC provides a pragmatically justified comparison of functional effectiveness with acceptable efforts required for this operation practically in any cases of search carried out in contemporary information retrieval systems. This conclusion has far-reaching consequences, in particular, the use of complex search characteristic I_2 allowed us (see Chapter 9) to solve one of the very important and complicated problems in information science—the problem of creating a selection mechanism that would allow the best search method to be selected for every search request (from a set of available methods).

The described results were obtained under the assumption that a binary scale of pertinence was used in evaluating documents. On the basis of these results, we also considered situations in which a "fuzzy" scale of pertinence is used. We constructed complex search characteristics for such cases. In addition, we adapted to this situation formulation of the order preservation property and showed that constructed CSCs have this property. Finally, we discussed some problems for which solutions are necessary in order to use these complex search characteristics. In concluding the chapter, we discussed the role of formulating the goal of documentary search in the context of the problem of evaluating the functional effectiveness of this search.

References

Avetisyan, D. O. (1975). *Statistical methods used in solving applied problems in information retrieval.* Ph.D. thesis, Nobvosibirsk.

Avetisyan, D. O. (1977). Evaluation of functional effectiveness of recognition systems. *Dokladi Academii Nauk of Armenian SSR, Applied Mathematics, 64,* (3).

Bollmann, P. (1977). A comparison of evaluation measures for document retrieval systems. *Journal of Informatics, 1,* 97–116.

Bollmann, P., Raghavan, V. V., Jung, G. S., & Shu, L. C. (1992). On probabilistic notions of precision as a function of recall. *Information Processing and Management, 28,* 291–315.

Borko, H. (1961). Research plan for evaluating effectiveness of various indexing systems. System Development Corporation, Report FN-5649. Santa Monica, CA: System Development Corporation, June 1961.

Cherniavsky, V. S., & Lakhuti, D. G. (1970). The problem of evaluating retrieval systems. *Nauchno-Teknicheskaya Informatsiya (NTI),* ser. 2, no. 1, 24–34.

Cleverdon, C. W. (1970). Evaluation of tests of information retrieval systems. *Journal of Documentation, 26,* 55–67.