

Evaluation of Macroevaluated Objects

11.1

Introduction

In the first section of Chapter 10 we emphasized that, along with the evaluation of the functional effectiveness of a search, we are also interested in the evaluation of objects with whose help or within the scope of which a search for documents proceeds. These objects include information retrieval systems as well as those subprocesses that involve document search (e.g., the translation of search requests from natural language to information retrieval language). In this chapter our concern will be such an evaluation of these objects, which could enable researchers to predict the search results for a newly entered search request when this search is performed with help or within the scope of the evaluated object. In Chapter 10 we described situations for which such an evaluation may be required. For example, when each subprocess of a document search, except for the subprocess under evaluation (for example, the translation of search requests from natural language to information retrieval language), proceeds using the same algorithm (for a corresponding subprocess) irrespective of the entered search request, then the formation of a set of "best" realizations of the subprocess under evaluation may be required for the purpose of using the set in selecting the most suitable realization for each search request entered; or, after forming such a set, it may be required to determine whether a newly proposed realization should be included in this set. In information science, as was pointed out earlier, different approaches for such an evaluation have been proposed. These approaches are based mainly on averaging functional effectiveness values obtained in a specially organized series of searches. By "functional effectiveness values" we mean the values of those search characteristics that are used in evaluating functional effectiveness. Note that the objects whose evaluation requires averaging of the original set of values we refer to as *macroevaluated*. This chapter deals with some of the problems related to evaluating macroevaluated objects.

11.2

Determination of Expediency for Evaluating Specific Macroevaluated Objects

Information science provides no clear picture as to which macroevaluated objects are expedient to evaluate and which are not. For example, in Cherniavsky & Lakhuti (1970) and Lakhuti (1971) the authors argue that it is expedient to evaluate not only retrieval services but also retrieval systems. In these references the notion of retrieval system is not defined formally but is introduced as a contraposition to the notion of retrieval service. In other words, on the one hand, any retrieval system is characterized by a language that includes the original alphabet or vocabulary and rules for constructing complex expressions, rules for indexing or translating texts from the natural language to a system language, and logic rules (which include comparison rules and possibly base relations between expressions of the system language). On the other hand, any functional retrieval system actually exists in the form of a particular collection of documents in a specific language derived from a specific source—personnel and hardware that realize search and indexing rules—and, finally, a particular circle of users serving as a source of search requests and in one way or another assessing search results. In the aggregate, all of these physical components are referred to as a retrieval service. The same retrieval system may be realized in different retrieval services and vice versa (the same retrieval service may use different retrieval systems).

In the view of Lakhuti (1971), basic considerations defining the expediency of evaluating retrieval systems are the need to study the behavior of a retrieval system for different classes of retrieval services realizing it to selecting a system optimal for a given service; the need to debug the system prior to its operation under the conditions of a specific retrieval service (pre-operation debugging); and the algorithmability of basic semantic processes of retrieval system operation (indexing and search), enabling one to consider the system as a model of human intellect.

From the standpoint of the authors of the mentioned articles, a retrieval service—in contrast to a retrieval system—includes or may include nonalgorithmized semantic processes such as search iteration based on a user's feedback; the preliminary screening of information noise by experts; and, especially, the content modification of queries and the formulation of subqueries or "search strategy." At the same time it is clear that the evaluation of retrieval services is at least as important as the evaluation of retrieval systems if for no other reason than the former should in general allow us to predict the functional effectiveness of a search based on user queries, whereas the existing evaluations of retrieval systems will permit such a prediction only in certain situations, if at all.

Hence, there are convincing arguments for the expediency of evaluating