Then the pertinent documents obtained from the user are added to those on the basis of which the best algorithm(s) constructed the best query formulation. After this, the best algorithm(s) constructs a new, *preliminary* query formulation, where occurrence frequencies of descriptors in document profiles of the new collection are used. Those original subrequests, which were selected before, are added to the obtained query formulation(s), and thus the final query formulation is constructed. Notice that all extraneous subrequests (a subrequest in a query formulation is considered extraneous if it contains another subrequest in the same query formulation) are removed from the final query formulation. Then a new search is performed using this new final query formulation and the new (obtained in the new collection of documents) output is given to the user. This is the end of first feedback iteration allowing us to change the system into better state. Beginning with the second output, the system realizes optimal service, because the best possible alternative was used for the search. Table 9.8 lists the basic steps of the feedback algorithm in an IR system performing optimal search in a dynamic collection of documents. We emphasize that an optimal search proceeds for an individual user and for a specific request regardless of how well the request was formulated, and this raises the quality of information service to the user.

Thus we have shown how an optimal search is realized in practice, that is, how an IR system is able to satisfy the function stated in Chapter 4.

Table 9.8
Basic Steps of the Algorithm for Selective Feedback for a Dynamic Collection of Documents

1. Pertinent documents from previous output are selected based on the user's evaluation of this output.
2. When a combined query formulation constructed by more than one algorithm was used in the previous search, an evaluation of available algorithms proceeds with the help of the criterion $\sqrt{R \cdot P}$. The highest value of this criterion determines the choice of best algorithm(s).
3. With the help of the same criterion $\sqrt{R \cdot P}$, the best original subrequests are determined among other than the best query formulations and those subrequests whose criterion values are not smaller than the highest value are selected for use in the search of the new collection of documents.
4. Pertinent documents found during the previous search are added to those on the basis of which the previous query formulation was constructed; that is, a more precise search request is obtained.
5. On the basis of the new marked set, a new preliminary query formulation is constructed by the best algorithm(s).
6. The final query formulation is constructed by adding the best original subrequests (see Step 3) to the preliminary query formulation.
7. A search is performed using the final query formulation, and a new output is given to the user.

9.11 Internal Control

Further improvement in the selective feedback algorithm in an IR system realizing optimal search can be achieved by *internal control,* which we describe next for both static and dynamic collections of documents. First, we look at a static case. After the initial search and the user's evaluation of the output, the best query formulation is determined and the values of r^2/n of all the outputs corresponding to query formulations used in the search are stored. After the first iteration, where the query formulation constructed by the best algorithm is used for the search, the value of r^2/n is computed to further compare the quality of the obtained output with other outputs. If the quality improved as compared to the previous output obtained by using the query formulation constructed by the best algorithm, then this indicates the pragmatism of using this algorithm further. However, if the quality decreased, then the obtained (lower) value is compared to the next highest value (corresponding to the second-best algorithm). If the obtained value in the new search is higher, then the next iteration uses the same algorithm as the previous iteration; otherwise, the next iteration uses the algorithm with the next highest value of r^2/n, and so on.

The additional control avoids a commitment to the algorithm, which was possibly not the best choice. Moreover, if the user's initial search request was not formulated well, then it is possible that at the initial step an algorithm gave better results precisely because the search request was not indicative of the user's information need (this was discussed previously). Then after obtaining additional information about POIN (the user's evaluation of the output), a different algorithm may give better search results. Hence, one of the main functions of internal control is to monitor the correctness of the selection algorithm.

There are other possible approaches to organizing internal control. For example, after selecting the best algorithm(s), other algorithms could still be used to construct query formulations for subsequent searches. The system gives the user only documents that were found using the query formulation constructed by the best algorithm. After receiving the user's evaluation of the output, the system could evaluate other query formulations to determine if they find the same pertinent documents with less noise. It is also possible for evaluating the documents found with other query formulations (and hence not given to the user) to introduce weights analogous to selective feedback for the dynamic collection of documents and to use these weights in selecting the best system's state. These and other approaches should be tested experimentally, but it is important to know that such approaches exist.

In the case of a dynamic collection of documents, the internal control could be implemented as follows. The initial output with the user's evaluation is the basis for the choice of the best algorithm for constructing query formula-