

because all of the documents that can be found by these subrequests were already seen by the user in the process of obtaining CM (pertinent documents contained in the outputs from the previous searches).

If NQ is not empty (i.e., if it contains at least one subrequest), a new search is performed, and if a set of obtained documents contains any new documents not appearing in the previous outputs, these documents (new output) are given to the user from which he or she selects pertinent documents. The system keeps the entire output for the following iterations. The pertinent documents in this output as chosen by the user are added to all other pertinent documents from previous searches, a new query formulation is constructed, and a new search is performed. This iterative process continues until one of the following conditions is satisfied:

1. An auxiliary query formulation (AQ) does not contain new subrequests.
2. A set of documents obtained as the result of a new search does not contain any new documents.
3. An output does not contain any new pertinent documents.
4. The user does not provide any evaluation of the documents from the output (leaves the system).

Table 9.1 provides a summary of one iteration in the feedback algorithm for static collection.

This finishes our discussion of the feedback algorithm for a static collection of documents. Now we will consider the dynamic collection of documents, where each feedback iteration is performed on a new collection of documents.

**Table 9.1**

Summary of the Steps in One Iteration of the Feedback Algorithm for Static Collection

1. Pertinent documents from the previous output (using the user's evaluation) that were obtained as the result of the search based on the previous query formulation (PQ) are selected.
2. A new marked set is constructed by combining the pertinent documents obtained in Step 1 with the documents used for the construction of the previous query formulation.
3. The auxiliary query formulation (AQ) is constructed using as input the marked set from Step 2. This step uses an algorithm described in Chapter 7.
4. The new query formulation (NQ) is constructed by removing from AQ all subrequests appearing in PQ and all subrequests that contain subrequests from PQ.
5. The user is given the output obtained in two steps:
  - (a) A search based on NQ
  - (b) The removal from the obtained set of documents of all documents from the previous outputs

## 9.5

### Feedback Algorithm for the Dynamic Collection of Documents

Recall that in realizing feedback for the dynamic collection of documents, the task of control is different from that in the static collection. In static collection, we try to obtain a new query formulation (at every new iteration) that does not intersect with any of the previous query formulations. But in dynamic collection, we only try to improve the previous query formulation. In other words, it is necessary to remove all subrequests that led to the unsuccessful search and to add new (original) subrequests that will presumably give us pertinent documents. Hence, in the case of dynamic collection the goal is not to find pertinent documents that were not found during a previous search but to have a more successful search of a new collection of documents; that is, in this case it is necessary to construct a new query formulation (by correcting the previous one) so that if we performed searches in the same collection using the previous query formulation and the new one, the latter search would give us better results.

The described feedback algorithm for the dynamic collection consists of two separate steps: (1) to increase the recall level (by adding new subrequests) and (2) to increase the precision (by removing "bad" subrequests). These steps will be called the "recall step" and the "precision step," respectively.

The input to the feedback algorithm consists of the user's evaluation of the system's output (a set of documents), which identifies pertinent (and hence nonpertinent) documents in this set. If there are no pertinent documents in the output, then the algorithm stops at this point (as is done with the static collection of documents), because there is no additional information about the user's actual POIN, and the search in the new collection is performed using the previous (unchanged) query formulation. If pertinent documents are found in the output, the recall step is executed. Then if the output does not contain nonpertinent documents, the feedback algorithm stops, and the search is performed using the query formulation obtained in this step. However, if nonpertinent documents are found in the output, then the precision step is executed. We now describe the recall and precision steps.

The recall step starts by adding to the set of pertinent documents all the documents used to construct the previous query formulation. This new marked set is used as an input to construct a new auxiliary query formulation (AQ) by applying the algorithm for the automatic construction of query formulations (described in Chapter 7). Then a new query formulation (denoted by NQ) is obtained by combining all subrequests from AQ and PQ (previous query formulation); that is,  $NQ = AQ \cup PQ$ . Notice that the application of the set union operator  $\cup$  will prevent NQ from having any extraneous subrequests. Recall that a subrequest S is extraneous in query formulation Q if there exists a sub-