VI. Relevance Feedback in an Information Retrieval System

W. Riddle, T. Horwitz, and R. Dietz

1. Introduction

An information retrieval system must be judged in the end by the user population, and therefore it should be designed primarily to satisfy the user's needs. However, the implementation of an efficient system, measured in terms of the amount of relevant material retrieved for the average user, is difficult to implement with presently available computing equipment because of their batch processing mode of operation. Typical turn-around times preclude efficient man-machine interaction, and a user must generally be satisfied with the results of a single search. However, the use of only a single search may not produce adequate results. The reduction of search time by the use of two-level searches, which match the query first against the centroid vectors of document clusters and then against the individual documents in highly correlated clusters, may cause some relevant documents to be lost.[2] Moreover, some indexing terms may be interpreted differently by users with different fields of interest. In fact, users unfamiliar with the indexing terms employed may formulate queries which, after translation to indexing concepts, may not adequately represent the user's requirements. In addition, the result of most correlation procedures presently used to match documents and search requests depends on the relative positions of the queries and documents in the n-space determined by the indexing terms; but the resulting correlations do not necessarily reflect the relevance of

the documents to a user's need.

Attempts have been made to improve the results obtained in a single search through a document collection by improving the query before a search is made, or by using a correlation function which better reflects relevance. Improvement of a query by an expansion done by the system prior to a search of the document set has been suggested.[4] This expansion is done either on the basis of statistically determined concept relations, or on the basis of a concept hierarchy, and causes concepts to be added to the query vector if they do not originally appear but are statistically correlated with or hierarchically related to concepts which do appear. It has also been suggested that the user himself reformulate his query prior to the search, and tests using the SMART System [3] indicate that improved results, in terms of the number of relevant documents retrieved, are obtained by this method. The reformulation is done before the query is processed, on the basis of a statistical analysis of the document set with respect to the index terms present in the original query. The improvement is effected by the elimination of those terms which have a high frequency in the document set (and are therefore not adequate differentiators), and reinforcement of those terms appearing infrequently in the document set (i.e. good differentiators). Maron and Kuhns have suggested a correlation technique using relevance numbers. These numbers are determined by probabilistic indexing, a method in which the indexer assigns a numerical value indicating the probabilistic value of that term to the document being indexed.[1] These methods, however, are not entirely adequate, since either they depend on a priori determination of relevance relationships which may not apply to the entire user

population, or they impose a waiting period on the user, after which yet another wait may be required for the processing of another reformulation of his request.

The advent of time-sharing computer systems allows the user to take a more active part in achieving the satisfaction of his request. In addition to providing a convenient means of man-machine communication, time-sharing appreciably reduces the time necessary for interaction between man and machine. With such a computing system, it is practical to propose an iterative method of information retrieval in which the user returns relevance judgments for the documents retrieved. This can be done after the abstracts, tables of contents, or full texts of the documents are read. The computing system, using this relevance feedback information, modifies the query last used in searching the document set by adding multiples of the relevant document vectors to the previous query. The system then performs a new search and retrieves a new set of documents on the basis of this modified query. The iterative process may be continued by the user until he feels that his needs have been adequately satisfied.[2,3]

This paper analyzes an information retrieval system which is based on an iterative query modification process, using relevance feedback information. Neither a time-shared computing system nor a user population were available for the study; but a FORTRAN program, run on a CDC 1604 computer was used to investigate various updating strategies as applied to a set of 82 documents and 34 queries. A priori judgments of the relevant documents for a given query were available for the document set and were used to simulate the user's relevance feedback information. The development of the various updating strategies is discussed in the next part; then the experimental

results are presented; finally, conclusions are drawn from the study.

2. Principal Methods

The relevance feedback information is used iteratively to perturb the query vector in the following manner:[2]

$$\underset{m\times 1}{\begin{bmatrix} Q' \end{bmatrix}} = \underset{m\times 1}{\begin{bmatrix} Q \end{bmatrix}} + \alpha \underset{m\times n}{\begin{bmatrix} R \end{bmatrix}} \cdot \underset{n\times 1}{\begin{bmatrix} W \end{bmatrix}}$$

where m = number of indexing terms

n = number of retrieved document

R = mxn matrix in which the ith column is the concept vector of the document of rank i

W = nxl vector of relevance weighting factors

It is assumed that query Q has caused a set, R, of n documents to be retrieved. Using the a priori judgments of the relevance of the retrieved documents, the relevance weighting factors are determined, and are used to define W, a vector of length n, so that the relevance weight of the document of rank j is the jth element. The relevance weight of a document reflects the relevance of that document to the query (the determination of the magnitude of these weights is discussed later). α is a multiplier controlling the strength of the perturbation to the query Q. The end result of this modification is that some linear combination of the kth elements of the retrieved documents is added to the kth element of the query vector Q, for all k of the indexing terms, thus producing a new query Q'.

In order to determine the effects of varying the parameters α, n, and the relevance weighting factors, sample runs are made using selected

queries. The results of these sample runs, presented in the following paragraphs, are used to develop strategies which are then applied to the entire set of queries.

A) Determination of the Number of Documents Retrieved

The number of documents, n, that are returned to the user is set at fifteen. The sample runs show that if this number is reduced to eight, the effectiveness of the updating process is diminished. In the case cited in Figure 1, returning fifteen documents leads to the retrieval of four relevant documents after three modifications are made, while returning only eight documents leads to the final retrieval of only two relevant documents after the same number of modifications. This implies the need to return initially as many relevant documents as possible so that more information can be used in the updating procedure. (The number of relevant documents initially retrieved also depends on the correlation function, as is discussed later.) Further, in determining the number of documents to be retrieved, a compromise must be made between the desirability of retrieving a large number of documents and the desirability of not imposing a large reading task on the user.

B) The Effect of the Correlation Function

The result of an iteration is a list of n documents ranked by their correlations with the query. These correlations are determined by one of the following correlation functions:

Cosine correlation function:[7]

$$c = \frac{\sum_{i=1}^{m} (q_i d_i)}{\sqrt{\left(\sum_{i=1}^{m} q_i q_i\right) \times \left(\sum_{i=1}^{m} d_i d_i\right)}}$$

Co-occurrence correlation function:[6]

$$c = \frac{\left(\sum_{i=1}^{m} q_i d_i \right)}{\left(\sum_{i=1}^{m} q_i q_i \right) + \left(\sum_{i=1}^{m} d_i d_i \right) - \left(\sum_{i=1}^{m} q_i d_i \right)}$$

Simple vector matching correlation function:*

$$c = \frac{\left(\sum_{i=1}^{m} q_i d_i \right)}{m}$$

where m = the number of indexing concepts

q_i = the ith concept weight of the query vector

d_i = the ith concept weight of the document vector

The effect of these different correlation functions on the relevance feedback process is not known, so the correlation function is included as another parameter in this investigation.

C) Determination of the Relevance Weighting Factors

In determining the relevance weighting factors the assumption is made that no information concerning the relative ranking of the relevant documents is available. That is, there is no way of knowing if one relevant document is more relevant than another. This is consistent with the proposed information retrieval system, in which the user returns only "relevant" or "non-relevant" judgments, without indicating the degree of relevance of each document retrieved. This implies that the numerical interpretation of the relevance information should be binary; therefore a weight of 1 is used as the relevance weighting factor of a relevant document

* The simple vector matching correlation function, as stated, is strictly suitable for use only with binary document and query vectors. Its use with other than binary vectors, without the addition of a normalization factor, does, however, preserve the relative rankings of the documents.

and 0 is used as the relevance weighting factor of a document for which the relevance is not known. However, the relevance weighting factors can be used to assign weights other than 1 to the relevant documents. In particular, the weights of the relevant documents can be defined as their correlations, thus giving the relevant documents with higher correlation more weight in the modification process. This method is also used to assign weights to the relevant documents in this investigation.

Negative relevance weighting factors can be employed in the perturbation of the query. Essentially, these negative factors indicate that the documents are "irrelevant" rather than the non-committal judgment that the relevance is not known. Since the relevance judgments available do not include any "irrelevant" indications, it is not feasible to simulate fully such judgments by a user. But some use of negative relevance weighting factors can be helpful when, after modification of the query, documents of unknown relevance are still retrieved with high correlations. Figures 3 and 4 show the effect of applying negative relevance weighting factors to a query for which no additional relevant documents are retrieved after the first modification. Documents 1102, 302, and 109 are ranked 2, 3, and 4 for both the original and updated queries but are of unknown relevance. Updating the modified query with negative relevance weighting factors yields a query which retrieves three relevant documents, a significant improvement. Further updating, using the normal 0 and 1 relevance weighting factors, eventually produces a query which retrieves all five relevant documents.

The improvement resulting from the use of negative relevance weighting factors suggests that the following heuristic method would prove useful.

When no relevant documents are retrieved, the two most highly correlated documents are given -1 relevance weighting factors. However, without "irrelevance" judgments the use of this heuristic method is at best an arbitrary procedure, and it should be noted that in some cases the negative relevance weighting factors are not helpful. In particular, the results shown in Figure 5 indicate that the use of too many negative relevance weighting factors overly perturbs the query vector, so that the modified query is unable to retrieve more relevant documents.

The use of negative relevance weighting factors implies that some of the concept weights in the query vector will become negative. Negative concept weights have no significance in a term-document matrix, since an indexer cannot indicate the non-appearance of a certain concept. It is plausible, however, to have negative weights in a query vector after it has been perturbed because the relevance feedback information can properly indicate that certain concepts are irrelevant.

D) Determination of the Value of α

The parameter α is necessary to control the manner and the strength of the perturbation to the query vector caused by the relevance feedback information. α may affect the modification in one of three ways. First, if α is increased with every iteration, the new relevance information will have the same weight as the sum of all the previously-gathered relevance informations. For example, if one relevant document with concept vector (10101) is retrieved initially by a query with concept vector (10100), the modified query (when α equals 1) will have a concept vector (20201). If this modified query, in turn, retrieves relevant documents having

concept vectors (10101) and (11100), then these vectors should have their concept weights multiplied by 2 to make their concept weights comparable to those of the modified query. Second, α can be kept at a constant value for all the iterations, which would cause all types of relevance information to be treated in the same manner, regardless of when they are obtained in the process. Finally, if α is decreased with every iteration, more emphasis will be placed on the relevance information obtained during the early iterations.

Figure 6 shows the results obtained in sample runs using various progressions and magnitudes of α. In general, the use of small fractional values of α yields, after two or three iterations, the results which can be obtained in one iteration by the use of a larger, integral value of α. For this reason, only integer values of α are used in further investigations.

The decreasing progression of α is also discarded because it is not significantly better than the other progressions. Also, a decreasing progression is illogical, since if the iteration process is converging on some area of n-space, then more weight should be given to relevance information obtained later in the process.

Thus two strategies involving α are used in the final investigations. In the first, α is set initially to 1, and is increased by 1 at each succeeding iteration. In the second, α is held constantly equal to 2 (the value 2 is used to represent a typical value rather than an optimal one). However, it should be noted that whenever the correlations of the relevant documents are used as the relevance weighting factors, α is held constantly equal to 1; this is denoted in the figures by the label "ALPHA = CORRELATIONS".

E) Termination of the Modification Process

Updating is terminated when all relevant documents have been retrieved (since the user's needs are then satisfied as fully as possible), or when at most three modifications have been made (since the results presented in Figure 6 indicate that, with the progressions of α used in the final investigations, the iteration process can safely be terminated in general after three modifications have been made).

3. Experimental Results

In general, the modification of a query using relevance feedback information leads to an improvement in both the number of relevant documents retrieved and in the ranks of all the relevant documents. The modification normally yields an increase in both precision and recall (as shown in Figure 2), regardless of how α is applied,* provided that the set of relevant documents lies in one basic cluster in n-space. If the relevant documents cluster in two separate regions in n-space (as a result of the indexing scheme used), the results are as shown in Figure 7.

When such a dual clustering of the relevant documents exists, Rocchio suggests the use of multiple queries.[2] This is good theoretically, when a priori relevance judgments, which list all the documents relevant to a given query, have been made. However, in a real system, the user is uncertain of the existence of other relevant documents and the technique is impossible to carry out. A possible solution is the use of a list that guarantees, for example, that whenever document X, Y, and Z are deemed relevant, then documents A, B, and C are also relevant and are returned

* The statement that the iterative retrieval process does not significantly depend on the particular strategy of applying α (for the progressions of α used in the final investigations) is supported by the data given in Figure 6, for the progressions used and the correlation function.

to the user. But this procedure assumes that what is relevant to one user is also relevant to another. Such an assumption is of doubtful value at best. Therefore, there does not appear at present to be any feasible means of returning documents to the user which are actually relevant but which, because of the indexing scheme, lie in a region of space far removed from the query and the other relevant documents.

Figures 9, 10, and 11 show the effect of using the negative relevance weighting factor heuristic method. Application of the heuristic method to two of the queries show results in a modified query which eventually retrieves the relevant documents. In the third case shown, query QB11, no definite conclusion can be drawn after three modifications have been made; however, it appears that the query vector is being modified correctly, and that it is moving toward the correct region of the document space.

The magnitude of the initial value of α affects the speed at which the modified query converges to a position in n-space where it is capable of retrieving all of the relevant documents. For the query shown in Figure 12, when α is initially equal to 1, the final ranks of the relevant documents are 1, 2, 3, 5, and 19. If, on the other hand, α is initially set equal to 2, the final ranks are 1, 16, 18, 28, and 34. This dependence on the initial value of α can be explained graphically (see Figure 13). Assume that documents A, B, C, and D are relevant to query q, and that q retrieves documents B, C, and D. Document D has comparitively large concept weights so that when q is modified using an α of 2, the resulting query q" is strongly biased toward D. Consequently, q" is unable to retrieve document B. With a milder modification using a smaller value of

α, the query q' is obtained and all the relevant documents are retrieved.

The final values of recall and precision depend on the number of relevant documents retrieved on the successive searches, since more information will obviously perturb the query to a greater extent. In particular, there is a dependence on the number of relevant documents retrieved initially, which is, in turn, dependent on the correlation function used. (In this investigation, the dependence is actually on the denominator of the correlation formula, since all of the functions tested possess the same numerator.) If only a few of the relevant documents are retrieved initially, then convergence is slow. In other words, given a query having three relevant documents, the probability of retrieving all three is higher if two of the documents are retrieved initially rather than only one. As shown in Figure 14, for query QA15 the cosine correlation function initially retrieves three relevant documents, while the co-occurrence and simple vector matching correlation functions retrieve two and four respectively. Since the simple vector matching case now includes more information concerning the concepts in the relevant documents, the final values of recall and precision achieved by the modification process are higher when simple vector matching is used as the correlation function, than when either of the other two functions is used. These results suggest that it is unwise to restrict the proposed retrieval system to the use of a single correlation function.

4. Conclusions

The implicit assumption underlying this investigation is that relevance feedback is a necessary part of the overall retrieval process. As the

feedback process is studied in detail, the validity of this assumption is demonstrated repeatedly; for practically every query tested, the recall is eventually increased.

The explicit assumption basic to this investigation is that the relevance feedback and query modification proccesses can be optimized in some sense. Although the analysis demonstrates that no uniformly best process exists, the investigation of the relevance weighting factors, the search correlation function, and the iteration strategy shows that some combinations of these parameters produce methods that work well in many cases.

Specifically, the following choices of the parameters constitute an optimal system since each optimizes that part of the modification process which it affects most directly. First, the simple vector matching correlation function is optimal in that it retrieves more relevant documents on the initial search than do either the cosine or co-occurrence correlation functions. Second, the iteration strategy in which α is increased with every iteration is optimal in the sense that the scale of the query updating information is effectively equal to the scale of the current query. Finally, the use of zero and one as relevance weighting factors, including the negative relevance weighting factor heuristic method, is optimal in the sense that it adequately describes the feedback information provided by the user. Therefore, this combination of parameters yields a retrieval system which is optimized for high precision and recall through the use of relevance feedback information. Of course, it can be argued that some of these choices for the parameters may lead to inefficiencies in the retrieval process, but this objection is of no consequence in this investigation

because the search time is not included in the investigation.

The investigation indicates two specific areas where more research is possible. Of primary importance, since it is now known that negative weights are useful in the relevance feedback process, is a more specific investigation of negative relevance weighting factors using relevance judgments given by an actual user population. Also, the number of iterations needed to obtain the best possible recall and precision should be studied.

In addition to its use in document retrieval, the relevance feedback process provides an efficient method for testing the efficiency of document indexing schemes, since external disturbances, such as incorrect formulations of queries, are minimized. Theoretically, one could possibly develop an indexing scheme which would eliminate the need to use relevance feedback information. However, the present investigation has shown how a rather simple use of relevance feedback information can greatly improve the recall achieved by the document retrieval system. Thus, it would seem more practical to use the best of the present indexing schemes, and direct further investigation to the area of retrieval improvement by the efficient and optimal employment of user relevance feedback information.

References

[1] M. E. Maron, and J. L. Kuhns, On Relevance, Probabilistic Indexing and Information Retrieval, Journal of the Association for Computing Machinery, Vol. 17, No. 3, July 1960, pp. 216-244.

[2] J. J. Rocchio, Harvard University Doctoral Thesis, Report ISR-10, to the National Science Foundation, Chapters 3 and 4.

[3] J. J. Rocchio and G. Salton, Search Optimization and Iterative Retrieval Techniques, Proceedings of the Fall Joint Computer Conference, Las Vegas, Nov. 1965.

[4] G. Salton, A Document Retrieval System for Man-Machine Interaction, Proceedings of the 19th National Conference, ACM Philadelphia, 1964, pp. L2.3-1 - L2.3-20.

[5] G. Salton, The Evaluation of Automatic Retrieval Procedures - Selected Test Results Using the SMART System, American Documentation, Vol. 16, No. 3, July 1965, pp. 209-222.

[6] G. Salton, Evaluation of Computer-Based Information Retrieval Systems, FID Congress, October 1965.

[7] G. Salton, Progress in Automatic Information Retrieval, IEEE Spectrum, Vol. 2, No. 8, August 1965.

APPENDIX A

FORTRAN Program Operating Instructions for CDC 1604 run

The program appears in two forms. The first form reads information concerning the document and query set from an input tape (tape number 138 as the program is presently written). The second form reads the document and query vectors from cards. The two programs operate identically, the only difference being that the tape version requires the input tape to be mounted on unit two, whereas the card input version must have a data deck appended to it. (The card input version qualifies as an 1S job.)

Both programs are driven by input data which should be organized as follows:

1. Relevant document information. One card per query, in the order in which the queries appear. Eleven three-column fields (using columns 1 through 33) specify the numbers of the relevant documents (these numbers are the sequential numbers of the documents as they appear in the document set, not the identification numbers). A single three-column field (columns 34 through 36) gives the count of the relevant documents for the query.
2. Factorial information for use in the evaluation formulas.
3. Concept number data deck if card input version is used. Each document and query have a set of cards which give the following:
    a) First card contains the alphameric identifier of the document or query in columns 1-16 and the total number of concepts in columns 20-21.
    b) The second through last card contain the concept numbers and total weights paired together (10 pairs to a card) in an I4 format, e.g.

column 1 4 7 . . . .
10 12 78 36 198 12

indicates that concept	has weight
10	12
78	36
198	12

4. Program driving information. The program can only set up the query-document correlations for a maximum of 17 query designations at a time, but these query designations need not be unique. Therefore the driving information should appear so that the desired analysis is grouped into packages of 17. Each package has the following format:

   a) First card contains in -

   columns

1-3	number of documents retrieved
4-6	correlation function indicator 1 → Simple Vector Matching 2 → Co-occurrence 3 → Cosine
7-9	number of analyses to be made (maximum of 17)

   b) Second card contains in -

   columns

1-3	number of the query used in 1st analysis
4-6	" " " " " " 2nd "
⋮	
49-51	" " " " " " 17th "

   c) Following cards, one for each analysis contain in

   columns -

1-3	initial value of α numerator
4-6	increment added to α numerator after each iteration
7-9	α denominator
10-12	total number of iteration modifications allowed

   e.g. if the progression of α desired is 3,5,7,... with only three modifications allowed, the card pertaining to this analysis would be:

   3 2 1 3

   whereas the progression 3/15, 7/15, 11/15,... with a maximum of 6 modifications would require:

   3 4 15 6

The program as it is presently written is for a document-query set containing 82 documents and 34 queries. DO loop indices, tables of factorials, and matrices are all set for this size set, and would have to be changed before the program could be used on a set of different size.

A data deck to perform a cosine correlation analysis on queries, 2,7, and 33, using the increasing α, and α equal to the correlations strategies for both 15 and 30 documents retrieved would appear as follows:

15 3 6	set number retrieved at 15
2 2 7 7 33 33	query numbers concerned
1 1 1 3 0 0 1 3	α strategy cards for query 2
1 1 1 3 0 0 1 3	α strategy cards for query 7
1 1 1 3 0 0 1 3	α strategy cards for query 33
30 3 6	change number retrieved to 30
2 2 7 7 33 33	query numbers concerned
1 1 1 3 0 0 1 3	α strategy cards for query 2
1 1 1 3 0 0 1 3	α strategy cards for query 7
1 1 1 3 0 0 1 3	α strategy cards for query 33

APPENDIX B

Evaluation of Relevance Feedback Methods.

E. M. Keen

Some of the results are here presented in summary form for searches employing relevance feedback, using averages over 22 search requests. Results are computed for three different correlation functions: cosine, co-occurrence and simple vector matching; and also for three different feedback strategies: increasing alpha (1,2,3), constant alpha (2,2,2) and alpha equal to the correlations of the relevant documents (c,c,c). Only 22 of the original 34 requests are averaged, since full results were not available for 12 of the 34. Nine of these 12 were not processed by all of the above procedures, some because the initial search result was very good and no iteration was needed, and the other three because each had only one relevant document, and averages were therefore not believed to be meaningful.

Tables B1 through B9 give average results using the measures of normalized recall, normalized precision and normalized overall. In Table B1 for example the cosine correlation function is used with the increasing alpha strategy, and the normalized measures indicate the improvement in performance that results from each update. Tables B2 and B3 also illustrate the use of the cosine correlation function, but the increasing alpha strategy is altered to constant alpha and alpha correlations respectively. Tables B4 through B6, and Tables B7 through B9 cover these same three alpha strategies but use the co-occurrence correlation function and the simple vector matching correlation function, respectively.

Comparing the three correlation functions alone, on the initial search

result the normalized overall measure is better for cosine than for co-occurrence, and better for co-occurrence than for simple vector matching. This can be seen in Table B10 where all searches are compared for each correlation function, with the correlation functions ranked in order of merit according to the normalized overall score achieved by each. The cosine correlation function works the best on the initial search and also on the updated searches using the alpha correlation strategy. The co-occurrence correlation function is the best for the updated searches using the increasing and constant alpha strategies. With one exception, the simple vector matching correlation function performs the worst on all searches.

Evaluation of the relevance feedback methods requires examination of the effectiveness of the updated searches, and the tables given show only two cases where updates result in a drop in performance. This occurs when the simple vector matching correlation function and alpha correlation strategy is used, and Table B9 shows that the second and third updates had a performance progressively worse than the first update. A single request contributed largely to the result, since in request QA9 the normalized overall measure dropped from 1.4677 to 0.6887 after the second update, and dropped further to 0.5989 after the third update. However all the combinations of correlation functions and alpha strategies resulted in a considerable improvement in performance with the updated searches compared to the initial search.

An order of merit of the nine combinations tested is given for the three updated searches in Table B11, where merit is based on the increase in the normalized overall measure achieved by each update compared with

the initial result. The co-occurrence correlation function with the constant alpha strategy achieves the greatest increase in the first and second updates, and the same correlation function, but with increasing alpha strategy, gives the best result for the third update. Combinations using the co-occurrence and cosine correlation functions together with the constant and increasing alpha strategies perform the best. Combinations using the simple vector matching correlation function and the alpha correlations strategy always have low merit.

Another method of displaying the increase in retrieval performance achieved by relevance feedback is the use of a plot of precision versus recall. Table B12 gives such a plot of the same results as those given in Table B1, in which the cosine correlation function and increasing alpha strategy is used. The results of the 22 requests are averaged by the use of a cut-off made after examination of m consecutive documents (m ranging from 1 to 20), and the precision and recall values are averaged over all requests at each cut-off point to obtain the nine average points for each curve. The four performance curves show the initial search and the three updated searches. The first update gives the biggest improvement in performance, and the updated searches achieve a very substantial improvement in performance at the high precision end of the curves. This improvement decreases as recall increases, and is almost lost when 0.7 recall is reached. If performance is compared at a cut-off of four documents, for example, the third updated search achieves an improvement over the initial search of more than 0.3 in both precision and recall.

To summarize, evaluation of the relevance feedback methods averaged over 22 search requests shows a very considerable improvement in performance

with all the updating methods used. The co-occurrence and cosine correlation functions with constant and increasing alpha strategies are a little superior to the simple vector matching correlation function and the alpha correlations strategy.

Search Type	Normalized Measures		
	Recall	Precision	Overall
Initial Search	0.7601	0.5566	1.3167
First Update	0.8083	0.6967	1.5050
Second Update	0.8234	0.7481	1.5715
Third Update	0.8267	0.7554	1.5821

Average Search Results for Cosine Function
and Increasing Alpha

Table B1

Search Type	Normalized Measures		
	Recall	Precision	Overall
Initial Search	0.7601	0.5566	1.3167
First Update	0.8100	0.6987	1.5087
Second Update	0.8193	0.7381	1.5574
Third Update	0.8189	0.7401	1.5590

Average Search Results for Cosine Function
and Constant Alpha

Table B2

Search Type	Normalized Measures		
	Recall	Precision	Overall
Initial Search	0.7601	0.5566	1.3167
First Update	0.8038	0.6873	1.4911
Second Update	0.8119	0.7138	1.5257
Third Update	0.8150	0.7153	1.5303

Average Search Results for Cosine Function
and Alpha Correlations.

Table B3

Search Type	Normalized Measures		
	Recall	Precision	Overall
Initial Search	0.7552	0.5545	1.3097
First Update	0.8129	0.7027	1.5156
Second Update	0.8377	0.7493	1.5830
Third Update	0.8385	0.7550	1.5935

Average Search Results for Co-Occurrence Function and Increasing Alpha

Table B4

Search Type	Normalized Measures		
	Recall	Precision	Overall
Initial Search	0.7552	0.5545	1.3097
First Update	0.8177	0.7098	1.5275
Second Update	0.8390	0.7490	1.5880
Third Update	0.8386	0.7514	1.5900

Average Search Results for Co-Occurrence Function and Constant Alpha

Table B5

Search Type	Normalized Measures		
	Recall	Precision	Overall
Initial Search	0.7552	0.5545	1.3097
First Update	0.7899	0.6590	1.4489
Second Update	0.8014	0.6898	1.4912
Third Update	0.8124	0.7048	1.5172

Average Search Results for Co-Occurrence Function and Alpha Correlations

Table B6

Search Type	Normalized Measures		
	Recall	Precision	Overall
Initial Search	0.7487	0.5356	1.2843
First Update	0.7958	0.6789	1.4747
Second Update	0.7994	0.6864	1.4858
Third Update	0.8001	0.6885	1.4886

Average Search Results for Simple Vector Matching Function and Increasing Alpha

Table B7

Search Type	Normalized Measures		
	Recall	Precision	Overall
Initial Search	0.7487	0.5356	1.2843
First Update	0.7968	0.6782	1.4750
Second Update	0.8029	0.6966	1.4995
Third Update	0.8033	0.6984	1.5017

Average Search Results for Simple Vector Matching Function and Constant Alpha

Table B8

Search Type	Normalized Measures		
	Recall	Precision	Overall
Initial Search	0.7487	0.5356	1.2843
First Update	0.7944	0.6712	1.4656
Second Update	0.7753	0.6566	1.4319
Third Update	0.7744	0.6547	1.4291

Average Search Results for Simple Vector Matching Function and Alpha Correlations

Table B9

Merit	Initial Search	First Update			Second Update			Third Update		
		INC.	CON.	CORR.	INC.	CON.	CORR.	INC.	CON.	CORR.
1	COS	CO-OC	CO-OC	COS	CO-OC	CO-OC	COS	CO-OC	CO-OC	COS
2	CO-OC	COS	COS	SVM	COS	COS	CO-OC	COS	COS	CO-OC
3	SVM	SVM	SVM	CO-OC	SVM	SVM	SVM	SVM	SVM	SVM

INC = alpha increasing
CON = alpha constant
CORR = alpha correlations

COS = cosine function
CO-OC = co-occurrence function
SVM = simple vector matching function

Order of Merit assigned by Normalized Overall for the Correlation Functions

Table B10

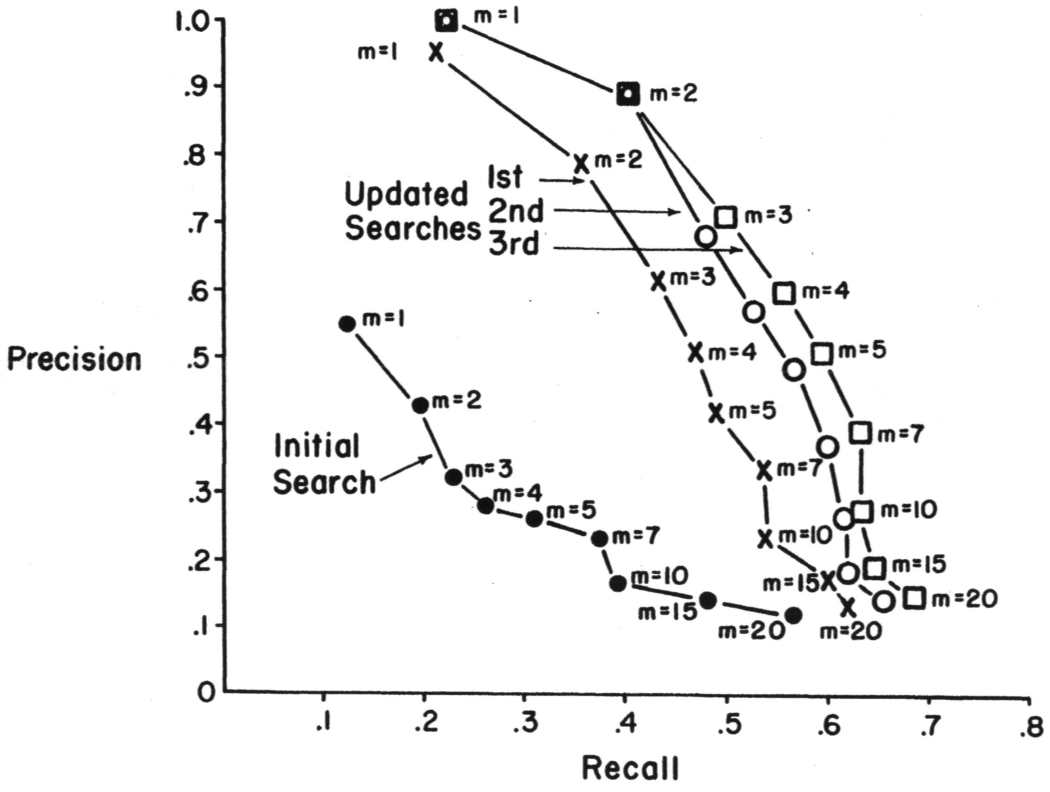
Merit	First Update		Second Update		Third Update	
1	CO-OC	CON	CO-OC	CON	CO-OC	INC
2	CO-OC	INC	CO-OC	INC	CO-OC	CON
3	COS	CON	COS	INC	COS	INC
4	SVM	CON	COS	CON	COS	CON
5	SVM	INC	SVM	CON	SVM	CON
6	COS	INC	COS	CORR	COS	CORR
7	SVM	CORR	SVM	INC	CO-OC	CORR
8	COS	CORR	CO-OC	CORR	SVM	INC
9	CO-OC	CORR	SVM	CORR	SVM	CORR

INC = alpha increasing
CON = alpha constant
CORR = alpha correlations

COS = cosine function
CO-OC = co-occurrence function
SVM = simple vector matching function

Order of Merit assigned by Normalized Overall for Different Functions and Alphas when the Increase in Performance of each Update Compared with the Initial Search is Considered

Table B11



Precision versus Recall for Initial Search and Updated Searches
Using Relevance Feedback (averaged over 22 search requests)

Table B12

INDEX OF FIGURES

		Page
Figure 1	Ranks of Relevant Documents After Each Update as a Function of the Number of Documents Returned in Answer to a Query	VI-30
Figure 2	Typical Results of Query Modification Using Relevance Feedback Information in the Case when the Relevant Documents are in One Region of N-space	VI-30
Figure 3	The Effect of Negative Weightings Applied to Non-Relevant Documents which Consistently Appear on the Output List with a Relatively High Correlation	VI-31
Figure 4	Successful Use of Negative Weighting	VI-32
Figure 5	Unsuccessful Use of Negative Weighting	VI-37
Figure 6	Results of Various Progressions of Alpha	VI-42
Figure 7	Typical Results of Query Modification Using Relevance Feedback Information in Case when the Relevant Documents Split into Two Regions in N-Space	VI-45
Figure 8	Shifting of Queries in N-Space Due to Query Modification Using Relevance Feedback Information	VI-46
Figure 9	Selected Results Showing Various Effects of the Negative Weighting Heuristic Method	VI-48

INDEX OF FIGURES (continued)

		Page
Figure 10	Use of Negative Weights	VI-49
Figure 11	Continued Use of Negative Weights when Use in the First Update Fails to Move Query into Correct Region of N-Space	VI-58
Figure 12	Comparison of Results After Three Iterations as a Function of the Value of Alpha used in First Update Cosine Correlation Function	VI-62
Figure 13	Effect of the Initial Value of α used in the Iteration Process	VI-70
Figure 14	The Result of the Query Modification by Use of Relevance Information as a Function of the Number of Relevant Documents Initially in the Fifteen Highest Correlated Documents	VI-71

	15 Documents Returned	8 Documents Returned
Initial Ranking of Relevant Documents	3, 12, 17, 33, 44	3, 12, 17, 33, 44
Ranking After First Update	1, 2, 3, 15, 30	1, 3, 10, 30, 34
Ranking After Second Update	1, 2, 3, 4, 19	1, 2, 11, 12, 29
Ranking After Third Update	1, 2, 3, 4, 19	1, 2, 11, 13, 30

Query QA10GROUP
Cosine Correlation Function
Increasing Alpha Strategy

Ranks of Relevant Documents After Each Update as a Function of the Number of Documents Returned in Answer to a Query

Figure 1

	Ranks of Relevant Documents	Normalized Recall	Normalized Precision	Rank Recall	Log Precision
Initial Results	1,2,19,27	.877	.739	.204	.458
Results After First Update	1,2, 6,33	.899	.805	.238	.531
Results After Second Update	1,2, 3,21	.946	.885	.370	.657
Results After Third Update	1,2, 3,18	.956	.895	.417	.679

Query QB10CHEMIST
Cosine Correlation Function
Increasing Alpha Strategy

Typical Results of Query Modification Using Relevance Feedback Information in the Case when the Relevant Documents are in One Region of N-Space

Figure 2

	Document Rank	Name	Relevant	Weighting Applied
Initial Results	1	307ROLE OF /	X	1
	2	1102CENTRALI		0
	3	302RECORDING		0
	4	109STATUS RE		0
	5	1212TOWARD A		0
	6	1201ACCREDIT		0
	7	1107THE AUTO		0
Results After First Update	1	307ROLE OF /	X	1
	2	1102CENTRALI		-1
	3	302RECORDING		-1
	4	109STATUS RE		-1
	5	1212TOWARD A		0
	6	1201ACCREDIT		0
	7	1107THE AUTO		0
Results After Second Update	1	307ROLE OF /	X	1
	2	902DOCUMENT		0
	3	214ELECTRONI		0
	4	201A SYSTEM		0
	5	209THE REDUC	X	1
	6	903THE USE O		0
	7	203NEW PHOTO	X	1
Results After Third Update	1	307ROLE OF /	X	
	2	209THE REDUC	X	
	3	203NEW PHOTO	X	
	4	214ELECTRONI		
	5	902DOCUMENT		
	6	201A SYATEM		
	7	211A COMPUTE	X	
	.			
	.			
	10	206ANALYSIS	X	

Query QA12JOURNAL
Cosine Correlation Function
Constant Alpha Strategy

The Effect of Negative Weightings Applied to Nonrelevant Documents which Consistently Appear on the Output List With a Relatively High Correlation

Figure 3

Results for Query No. 12 QA12JOURNAL

Document		Relevant	Correlation
No.	Name		
4	307ROLE OF /	X	.64956980
40	1102CENTRALl		.38254603
19	302RECORDING		.30096463
46	109STATUS RE		.29774567
60	1212TOWARD A		.24494897
10	1201ACCREDIT		.20412415
81	1107THE AUTO		.20412415
75	201A SYSTEM		.20134682
43	1206THE EDUC		.19658927
21	1209SCIENCE		.17234550
35	1004DENSITY		.16823165
13	1119RECENT A		.16439899
57	603COMPILATI		.16366342
45	1215GRADUATE		.15861032
38	303C. M. HER		.15713484

$$\text{Query}(1) = \text{Query}\ (0) + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT}(I)}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Alpha = 2

(a) Initial Results (Cosine Correlation)

Successful Use of Negative Weighting
(Five relevant documents in the collection)

Figure 4

Results for Query No. 12 QA12JOURNAL

Document		Relevant	Correlation
No.	Name		
4	307ROLE OF /	X	.99146269
40	1102CENTRALI		.34989921
19	302RECORDING		.30969005
46	109STATUS RE		.25531480
33	11017INFORMA		.22165517
17	301DOCUMENT		.22003279
60	1212TOWARD A		.21004201
81	1107THE AUTO		.19837301
22	914THE RAPID		.18670401
71	108SEARCHERS		.17389250
35	1004DENSITY		.16349177
32	902DOCUMENT		.16114324
37	306PROGRESS		.15403081
73	906MICROFILM		.15244319
69	813AUTOMATIC		.15105098

$$\text{Query(2)} = \text{Query (1)} + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT(I)}}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	1	-1	-1	-1	0	0	0	0	0	0	0	0	0	0	0

Alpha = 2

(b) After First Update (Cosine Correlation)

Figure 4 (continued)

Results for Query No. 12 QA12JOURNAL

Document			
No.	Name	Relevant	Correlation
4	307ROLE OF /	X	.53571541
32	902DOCUMENT		.10271470
63	214ELECTRONI		.05685047
75	201A SYSTEM		.02853186
25	209THE REDUC	X	.02711354
44	903THE USE O		.01753145
42	203NEW PHOTO	X	.01711912
74	305THE USE O		.01171571
69	813AUTOMATIC		.00936073
73	906MICROFILM		.00472350
14	518EDUCATION		0
35	1004DENSITY		-.01430356
33	1101/INFORMA		-.01569842
28	506THE ROLE		-.01592627
22	914THE RAPID		-.02314034

$$\text{Query(3)} = \text{Query(2)} + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT(I)}}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0

Alpha = 2

(c) After Second Update (Cosine Correlation)

Figure 4 (continued)

Results for Query No. 12 QA12JOURNAL

Document No.	Document Name	Relevant	Correlation
4	307ROLE OF /	X	.70658278
25	209THE REDUC	X	.44245670
42	203NEW PHOTO	X	.39855516
63	214ELECTRONI		.21694460
32	902DOCUMENT		.15644221
75	201A SYSTEM		.14526777
7	211A COMPUTE	X	.13650823
73	906MICROFILM		.11716347
17	301DOCUMENT		.11570746
9	206ANALYSIS	X	.10580301
69	813AUTOMATIC		.10387310
44	903THE USE O		.10070364
11	1117PLANNING		.08802188
71	108SEARCHERS		.08792684
33	1101/INFORMA		.08539239

$$\text{Query}(4) = \text{Query}(3) + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT}(I)}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0

Alpha = 2

(d) After Third Update (Cosine Correlation)

Figure 4 (continued)

Results for Query No. 12 QA12JOURNAL

Document		Relevant	Correlation
No.	Name		
4	307ROLE OF /	X	.67270472
25	209THE REDUC	X	.58764284
42	203NEW PHOTO	X	.56324980
7	211A COMPUTE	X	.42694230
9	206ANALYSIS	X	.37116251
63	214ELECTRONI		.24151839
75	201A SYSTEM		.23915765
32	902DOCUMENT		.20116064
73	906MICROFILM		.19796493
20	304PHOTOGRAP		.19125152
17	301DOCUMENT		.17945400
69	813AUTOMATIC		.17232484
22	914THE RAPID		.16768022
51	1115SOME OBS		.16539875
71	108SEARCHERS		.16145026

No Update for these Results

(e) After Fourth Update (Cosine Correlation)

Figure 4 (continued)

Results for Query No. 9 QA9ANALYSIS

Document		Relevant	Correlation
No.	Name		
39	517RESEARCH		.43301270
47	1116RETRIEVA		.42135049
11	1117PLANNING		.40209035
62	809A MECHANI		.39900373
12	1113THE MATE		.34815531
2	814A NEW EFF		.34299717
4	307ROLE OF /		.32478490
17	301DOCUMENT		.32075015
9	206ANALYSIS		.29329423
23	815A FACET A		.29329423
27	504IS RELEVA		.28867513
21	1209SCIENCE		.28724249
1	1104THE IBM		.28647316
71	108SEARCHERS		.28517591
70	601ENCODING		.28306926

$$\text{Query(1)} = \text{Query(0)} + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT(I)}}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Alpha = 2

(a) Initial Results (Cosine Correlation)

Unsuccessful Use of Negative Weighting
There are two relevant documents in the collection.

Figure 5

Results for Query No. 9 QA9ANALYSIS

Document		Relevant	Correlation
No.	Name		
31	207PRACTICAL		-.06676846
74	305THE USE O		-.08498545
57	603COMPILATI		-.08688523
15	610CHARACTER		-.09306588
32	902DOCUMENT		-.09431514
35	1004DENSITY		-.09719073
50	418ENGLISH-L	X	-.10714347
18	1122STATE-OF		-.10990208
28	506THE ROLE		-.11406536
54	901A SELECTE		-.12266150
63	214ELECTRONI		-.13492076
13	1119RECENT A		-.13540572
80	1001DESIGN O		-.13721113
67	716A NATIONA		-.13844392
65	505THE PLACE		-.13864335

$$\text{Query}(2) = \text{Query}(1) + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT}(I)}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

Alpha = 2

(b) After First Update (Cosine Correlation)

Figure 5 (continued)

Results for Query No. 9 QA9ANALYSIS

Document		Relevant	Correlation
No.	Name		
50	418ENGLISH-L	X	.2268128
31	207PRACTICAL		-.05874554
57	603COMPILATI		-.07965704
74	305THE USE O		-.08545551
15	610CHARACTER		-.09358063
32	902DOCUMENT		-.09483680
35	1004DENSITY		-.09508698
28	506THE ROLE		-.09926725
18	1122STATE-OF		-.11050995
54	901A SELECTE		-.11563119
65	505THE PLACE		-.13300052
80	1001DESIGN O		-.13500295
63	214ELECTRONI		-.13566701
13	1119RECENT A		-.13615465
67	716A NATIONA		-.13920965

$$\text{Query (3)} = \text{Query(2)} + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT(I)}}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	1	-1	-1	-1	0	0	0	0	0	0	0	0	0	0	-1

Alpha = 2

(c) After Second Update (Cosine Correlation)

Figure 5 (continued)

Results for Query No. 9 QA9ANALYSIS

Document		Relevant	Correlation
No.	Name		
50	418ENGLISH-L	X	.12680691
32	902DOCUMENT		-.09522233
18	1122STATE-OF		-.10680565
54	901A SELECTE		-.11023756
35	1004DENSITY		-.11934192
31	207PRACTICAL		-.13022519
14	518EDUCATION		-.13494266
63	214ELECTRONI		-.13599739
29	212A STATIST		-.13606583
15	610CHARACTER		-.13951793
28	506THE ROLE		-.14294774
44	903THE USE O		-.14716034
80	1001DESIGN O		-.15302047
25	209THE REDUC		-.15309977
38	303C. M. HER		-.15538235

$$\text{Query}(4) = \text{Query}(3) + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT}(I)}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	1	-1	-1	-1	0	0	0	0	0	0	0	0	0	0	0

Alpha = 2

(d) After Third Update (Cosine Correlation)

Figure 5 (continued)

Results for Query No. 9 QA9ANALYSIS

Document		Relevant	Correlation
No.	Name		
50	418ENGLISH-L	X	.22115640
35	1004DENSITY		-.10861615
28	506THE ROLE		-.12343199
29	212A STATIST		-.12639036
14	518EDUCATION		-.12844204
31	207PRACTICAL		-.13875402
15	610CHARACTER		-.14017633
65	505THE PLACE		-.14537614
80	1001DESIGN O		-.15220234
63	214ELECTRONI		-.15303001
44	903THE USE O		-.15481228
48	804COMPUTERI		-.15488673
25	209THE REDUC		-.15707111
58	424PRODUCING		-.16395750
13	1119RECENT A		-.17453190

No Update for these Results

(e) After Fourth Update (Cosine Correlation)

Figure 5 (continued)

Progression of Alpha	Correlation Function	Iteration	Ranks of the Relevant Documents Appearing in the Top Fifteen	Normalized Recall	Normalized Precision
1,1,1	Cosine	0	3,4,15	.706	.514
		1	1,2,5	.675	.612
		2	1,2,5	.667	.608
		3	1,2,5	.667	.608
1,1,1	Co-occurrence	0	2,7	.662	.457
		1	1,5,10	.647	.528
		2	1,3,7	.673	.580
		3	1,3,10	.671	.565
1,1,1	Simple Vector Matching	0	2,8,9,11	.725	.550
		1	2,3,4,14	.760	.643
		2	1,3,4,14	.766	.682
		3	1,3,4,14	.768	.683
2,2,2	Cosine	0	3,4,14	.706	.514
		1	1,2,5	.667	.608
		2	1,2,5	.665	.607
		3	1,2,5	.660	.605
3,3,3	Cosine	0	3,4,14	.706	.514
		1	1,2,5	.667	.608
		2	1,2,5	.661	.605
		3	1,2,5	.662	.606
4,4,4	Cosine	0	3,4,14	.706	.514
		1	1,2,5	.665	.608
		2	1,2,5	.662	.606
		3	1,2,5	.662	.606
1/15,1/15,1/15	Cosine	0	3,4,14	.706	.514
		1	2,5,8	.701	.532
		2	2,5,7	.690	.530
		3	2,3,7	.684	.549
2/15,2/15,2/15	Cosine	0	3,4,14	.706	.514
		1	2,3,5	.708	.576
		2	1,3,5	.699	.605
		3	1,2,5	.686	.618
3/15,3/15,3/15	Cosine	0	3,4,14	.706	.514
		1	1,3,5	.708	.609
		2	1,2,5	.686	.618
		3	1,2,5	.682	.615

Query QA15COST

Results of Various Progressions of Alpha

Figure 6

Progression of Alpha	Correlation Function	Iteration	Ranks of the Relevant Documents Appearing in the Top Fifteen	Normalized Recall	Normalized Precision
4/15,4/15,4/15	Cosine	0	3,4,14	.706	.514
		1	1,3,5	.697	.603
		2	1,2,5	.682	.616
		3	1,2,5	.677	.613
1,2,3	Cosine	0	3,4,14	.706	.514
		1	1,2,5	.675	.612
		2	1,2,5	.666	.608
		3	1,2,5	.660	.606
2,3,4	Cosine	0	3,4,14	.706	.514
		1	1,2,5	.667	.608
		2	1,2,5	.662	.606
		3	1,2,5	.662	.606
2,3,4	Co-occurrence	0	2,7	.662	.457
		1	1,6,9	.643	.522
		2	1,3,9	.662	.564
		3	1,3,12	.665	.555
2,3,4	Simple Vector Matching	0	2,8,9,11	.725	.550
		1	1,3,4,14	.766	.682
		2	1,2,4,14	.771	.705
		3	1,2,4,15	.773	.705
3,4,5	Cosine	0	3,4,14	.706	.514
		1	1,2,5	.667	.608
		2	1,2,5	.662	.606
		3	1,2,5	.662	.606
1/15,2/15,3/15	Cosine	0	3,4,14	.706	.514
		1	2,5,8	.701	.532
		2	1,3,6	.703	.598
		3	1,2,5	.686	.618
2/15,3/15,4/15	Cosine	0	3,4,14	.706	.514
		1	2,3,5	.708	.576
		2	1,3,5	.686	.598
		3	1,2,5	.682	.616
3/15,4/15,5/15	Cosine	0	3,4,14	.706	.514
		1	1,3,5	.708	.609
		2	1,2,5	.686	.618
		3	1,2,5	.677	.613
3,2,1	Cosine	0	3,4,14	.706	.514
		1	1,2,5	.667	.608
		2	1,2,5	.662	.606
		3	1,2,5	.660	.606

Figure 6 (continued)

Progression of Alpha	Correlation Function	Iteration	Ranks of the Relevant Documents Appearing in the Top Fifteen	Normalized Recall	Normalized Precision
4,3,2	Cosine	0	3,4,14	.706	.514
		1	1,2,5	.665	.607
		2	1,2,5	.662	.606
		3	1,2,5	.662	.606
5,4,3	Cosine	0	3,4,14	.706	.514
		1	1,2,5	.662	.606
		2	1,2,5	.662	.606
		3	1,2,5	.662	.606
3/15,2/15,1/15	Cosine	0	3,4,14	.706	.514
		1	1,3,5	.708	.609
		2	1,3,5	.686	.598
		3	1,3,5	.686	.618
4/15,3/15,2/15	Cosine	0	3,4,14	.706	.514
		1	1,3,5	.697	.603
		2	1,2,5	.686	.618
		3	1,2,5	.682	.616
5/15,4/15,3/15	Cosine	0	3,4,14	.706	.514
		1	1,2,5	.693	.620
		2	1,2,5	.680	.614
		3	1,2,5	.677	.613
Correlations	Cosine	0	3,4,14	.706	.514
		1	1,3,5	.677	.595
		2	1,2,5	.668	.609
		3	1,2,5	.662	.591

Figure 6 (continued)

	Ranks of Relevant Documents	Normalized Recall	Normalized Precision	Rank Recall	Log Precision
Initial Results	3 1117PLANNING 4 306PROGRESS 14 1110COMPUTER 19 1106A NEW CE 49 902DOCUMENT 68 716A NATIONA	.706	.512	.134	.407
Results After First Update	1 1117PLANNING 2 1110COMPUTER 5 306PROGRESS 42 1106A NEW CE 52 716A NATIONA 67 902DOCUMENT	.670	.613	.124	.464
Results After Second Update	1 1117PLANNING 2 1110COMPUTER 5 306PROGRESS 47 1106A NEW CE 51 716A NATIONA 68 902DOCUMENT	.669	.607	.121	.464
Results After Third Update	1 1117PLANNING 2 1110COMPUTER 5 306PROGRESS 49 1106A NEW CE 51 716A NATIONA 68 902DOCUMENT	.665	.605	.119	.459

Query QA15COST
Cosine Correlation Function
Increasing Alpha Strategy

Typical Results of Query Modification Using Relevance Feedback Information in Case when the Relevant Documents Split into Two Regions in N-Space

Figure 7

Initial Results		Results After First Update	
Rank	Name	Rank	Name
2	306PROGRESS	1	306PROGRESS
22	504IS RELEVA	15	1104THE IBM
24	1104THE IBM	39	505THE PLACE
45	518EDUCATION	45	1001DESIGN 0
49	505THE PLACE	46	518EDUCATION
51	1001DESIGN 0	50	504IS RELEVA

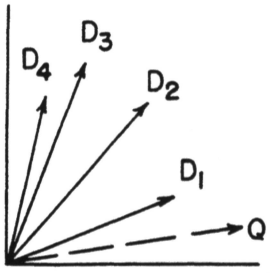
Results After Second Update		Results After Third Update	
Rank	Name	Rank	Name
1	1104THE IBM	1	1104THE IBM
2	306PROGRESS	2	306PROGRESS
46	5041IS RELEVA	45	504IS RELEVA
63	1001DESIGN 0	63	1001DESIGN 0
64	518EDUCATION	66	518EDUCATION
69	505THE PLACE	69	505THE PLACE

Query QA13EVALU
Cosine Correlation Function
Increasing Alpha Strategy

(a) Typical Results of Query Modification Using Relevance Feedback Information Showing the Shifting of the Ranks of the Relevant Documents

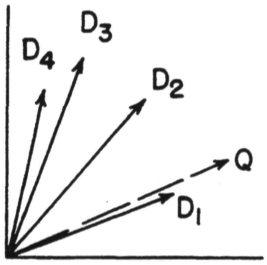
Shifting of Queries in N-Space Due to Query Modification Using Relevance Feedback Information

Figure 8



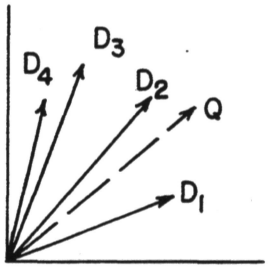
a) Initial Configuration

D_1, D_2, D_3, and D_4 are relevant documents, but only D_1 is ranked in the 15 highest correlated documents.



b) After First Update

Query has been shifted and D_1 and D_2 appear ranked in the top 15.



c) After Second Update

Query shifted further so that D_1, D_2, and D_3 now appear in top 15, but D_2 now appears above D_1.

(b) Two Dimensional Representation of Query Modification Showing the Shifting of the Ranks of the Relevant Documents

Figure 8 (continued)

	Relevant Document Ranks	Normalized Recall	Normalized Precision
A. Query QA9ANALYSIS			
Initial Results	33,50	.506	.172
Results After First Update	6,28	.506	.454
Results After Second Update	1,2	1.000	1.000
B. Query QB11INDEX			
Initial Results	21	.756	.309
Results After First Update	44	.476	.141
Results After Second Update	41	.512	.157
Results After Third Update	22	.744	.299
C. Query QB8COSTRET			
Initial Results	16	.817	.371
Results After First Update	65	.220	.053
Results After Second Update	37	.561	.181
Results After Third Update	7	.927	.558

Cosine Correlation Function and Increasing Alpha Strategy used for all Three Cases

Selected Results Showing Various Effects of the Negative Weighting Heuristic Method

Figure 9

Results for Query No. 9 QA9ANALYSIS

Document Rank	Document No.	Name	Relevant	Correlation
1	39	517RESEARCH		.43301270
2	47	1116RETRIEVA		.42135049
3	11	1117PLANNING		.40209035
4	62	809A MECHANI		.39900375
5	12	1113THE MATE		.34815531
6	2	814A NEW EFF		.34299717
7	4	307ROLE OF /		.32478490
8	17	301DOCUMENT		.32075015
9	9	206ANALYSIS		.29329425
10	23	815A FACET A		.29329423
11	27	504IS RELEVA		.28867513
12	21	1209SCIENCE		.28724249
13	1	1104THE IBM		.28647316
14	71	108SEARCHERS		.28517591
15	70	601ENCODING		.28306926

Ranking took .120000 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
33	82	308MACHINE R	X	.20701967
50	50	418ENGLISH-L	X	.12038585

Normalized Recall = .5061728 Normalized Precision = .1717605

Rank Recall = .0361 Log Precision = .0936

$$\text{Query}(1) = \text{Query}(0) + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT}(I)}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0

N = Number of retrieved documents

Alpha = 1

(a) Initial Results

Use of Negative Weights
Cosine Correlation Function Increasing Alpha Strategy

Figure 10

Results for Query No. 9 QA9ANALYSIS

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF /		.0540474
2	74	305THE USE U		.01273911
3	31	207PRACTICAL		0
4	48	804COMPUTERI		0
5	57	603COMPILATI		0
6	82	308MACHINE R	X	0
7	15	610CHARACTER		-.01836796
8	25	209THE REDUC		-.02948198
9	13	1119RECENT A		-.03039738
10	68	811THE RELAI		-.03125382
11	46	109STATUS RE		-.03303192
12	80	1001DESIGN O		-.03494283
13	75	201A SYSTEM		-.03722904
14	18	1122STATE-OF		-.03827795
15	28	506THE ROLE		-.03896433

Ranking took .166667 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
6	82	308MACHINE R	X	0
28	50	418ENGLISH-L	X	-.06677806

Normalized Recall = .8086420 Normalized Precision = .4535267

Rank Recall = .0882 Log Precision = .1353

$$\text{Query}(2) = \text{Query}(1) + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT}(I)}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0

N = Number of retrieved documents

Alpha = 2

(b) After First Update

Figure 10 (continued)

Results for Query No. 9 QA9ANALYSIS

Document Rank	Document No.	Name	Relevant	Correlation
1	82	308MACHINE R	X	.80137497
2	50	418ENGLISH-L	X	.15977647
3	4	307ROLE OF /		.06465833
4	77	213ADAPTIVE		.06020315
5	57	603COMPILATI		.05430365
6	61	516AN EXPERI		.04789131
7	68	811THE RELAI		.03738971
8	41	511ANALYSIS7		.02824476
9	54	901A SELECTE		.02715182
10	28	506THE ROLE		.02330699
11	75	201A SYSTEM		.02226901
12	5	1103FEASIBIL		.01279949
13	7	211A COMPUTE		.01236548
14	65	505THE PLACE		.01128809
15	74	305THE USE O		.00762006

Ranking took .166657 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
1	82	308MACHINE R	X	.80137497
2	50	418ENGLISH-L	X	.15977647

Normalized Recall = 1.0000000 Normalized Precision = 1.0000000

Rank Recall = 1.0000 Log Precision = 1.0000

No update for these results

(c) After Second Update

Figure 10 (continued)

Results for Query No. 9 QA9ANALYSIS

Document Rank	Document No.	Name	Relevant	Correlation
1	39	517RESEARCH		.43301270
2	47	1116RETRIEVA		.42135049
3	11	1117PLANNING		.40209035
4	62	809A MECHANI		.39900373
5	12	1113THE MATE		.34815531
6	2	814A NEW EFF		.34299717
7	4	3L7ROLE OF /		.3247849O
8	17	301DOCUMENT		.320750.5
9	9	206ANALYSIS		.29329423
10	23	815A FACET A		.29329423
11	27	504IS RELEVA		.28867513
12	21	1209SCIENCE		.28724249
13	1	1104THE IBM		.28647316
14	71	108SEARCHERS		.28517591
15	70	601ENCODING		.28306926

Ranking took .166657 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
33	82	308MACHINE R	X	.20701967
50	50	418ENGLISH-L	X	.12038585

Normalized Recall = .5061728 Normalized Precision = .1717605

Rank Recall = .0361 Log Precision = .0936

$$\text{Query}(1) = \text{Query}(0) + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT}(I)}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0

N = number of retrieved documents

Alpha = 2

(d) Initial Results

Use of Negative Weights
Cosine Correlation Function Constant Alpha Strategy

Figure 10 (continued)

Results for Query No. 9 QA9ANALYSIS

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF /		.00855386
2	74	305THE USE O		.00604851
3	31	207PRACTICAL		0
4	48	804COMPUTERI		0
5	57	603COMPILATI		0
6	15	610CHARACTER		-.01744216
7	82	308MACHINE R	X	-.02726146
8	68	811THE RELAI		-.02967852
9	13	1119RECENT A		-.03608157
10	18	1122STATE-OF		-.03634862
11	28	506THE ROLE		-.03700040
12	14	518EDUCATION		-.04082164
13	80	1001DESIGN O		-.04147700
14	25	209THE REDUC		-.04199400
15	7	211A COMPUTE		-.04907625

Ranking took .166657 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
7	82	308MACHINE R	X	-.02726146
28	50	418ENGLISH-L	X	-.07926529

Normalized Recall = .8024691 Normalized Precision = .4345146

Rank Recall = .0857 Log Precision = .1313

$$\text{Query}(2) = \text{Query}(1) + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT}(I)}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

N = Number of retrieved documents

Alpha = 2

(e) After First Update

Figure 10 (continued)

Results for Query No. 9 QA9ANALYSIS

Document Rank	Document No.	Name	Relevant	Correlation
1		308MACHINE R	X	.52017756
2		418ENGLISH-L	X	.06772233
3		603COMPILATI		.03682715
4		307ROLE OF /		.02923295
5		811THE RELAI		.01267830
6		305THE USE O		.00516770
7		506THE ROLE		0
8		207PRACTICAL		0
9		804COMPUTERI		0
10		213ADAPTIVE		-.00510350
11		901A SELECTE		-.00613786
12		610CHARACTER		-.01490215
13		1122STATE-OF		-.01552769
14		211A COMPUTE		-.01677182
15		201A SYSTEM		-.02265329

Ranking took .166557 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
1	82	308MACHINE R	X	.52017756
2	50	418ENGLISH-L	X	.06772233

Normalized Recall = 1.0000000 Normalized Precision = 1.0000000

Rank Recall = 1.0000 Log Precision = 1.0000

No Update for these results

(f) After Second Update

Figure 10 (continued)

Results for Query No. 9 QA9ANALYSIS

Document Rank	Document No.	Name	Relevant	Correlation
1	39	517RESEARCH		.43301270
2	47	1116RETRIEVA		.42135049
3	11	1117PLANNING		.40209035
4	62	889A MECHANI		.39900373
5	12	1113THE MATE		.34815531
6	2	814A NEW EFF		.34299717
7	4	307ROLE OF/		.32478490
8	17	301DOCUMENT		.32075015
9	9	206ANALYSIS		.29329423
10	23	815A FACET A		.29329423
11	27	5041S RELEVA		.28867513
12	21	1209SCIENCE		.28724249
13	1	1104THE IBM		.28647316
14	71	108SEARCHERS		.28517591
15	70	601ENCODING		.28306926

Ranking took .150000 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
33	82	308MACHINE R	X	.20701967
49	50	418ENGLISH-L	X	.12038585

Normalized Recall = .5123457 Normalized Precision = .1742522

Rank Recall = .0366 Log Precision = .0938

$$\text{Query}(1) = \text{Query}(0) + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT}(I)}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0

N = Number of retrieved documents

Alpha = Correlations

(g) Initial Results

Use of Negative Weights
Cosine Correlation Function Alpha = Correlations Strategy

Figure 10 (continued)

Results for Query No. 9 QA9ANALYSIS

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF/		.18272327
2	82	308MACHINE R	X	.0849256 4
3	75	201A SYSTEM		.03933260
4	74	305THE USE O		.03230139
5	46	109STATUS RE		.02442883
6	23	81 A FACET A		.01289122
7	25	209THE REDUC		.01245912
8	40	1102CENTRALI		.0074729 4
9	3	1210IN INFOR		0
10	10	1201ACCREDIT		0
11	31	207PRACTICAL		0
12	37	306PROGRESS		0
13	45	1215 GRADUATE		0
14	48	804COMPUTERI		0
15	57	603COMPILATI		0

Ranking took .166557 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
2	82	308MACHINE R	X	.0849256 4
26	50	418ENGLISH-L	X	-.02116537

Normalized Recall = .8456790 Normalized Precision = .5981638

Rank Recall = .1071 Log Precision = .1754

$$\text{Query}(2) = \text{Query}(1) + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT}(I)}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0

N = Number of retrieved documents

Alpha = Correlations

(h) After First Update

Figure 10 (continued)

Results for Query No. 9 QA9ANALYSIS

Document Rank	Document No.	Name	Relevant	Correlation
1	82	308MACHINE R	X	.22188008
2	4	307ROLE OF/		.17902224
3	75	201A SYSTEM		.05009643
4	74	305THE USE O		.03164696
5	46	109STATUS RE		.02393390
6	25	209THE REDUC		.01831005
7	50	418ENGLISH-L	X	.01382437
8	23	815A FACET A		.01263004
9	41	511ANALYSIS		.00977531
10	57	603COMPILATI		.00939705
11	40	1102CENTRALI		.00732154
12	77	213ADAPTIVE		.00260448
13	3	1210IN INFOR		0
14	10	1201ACCREDIT		0
15	31	2L7PRACTICAL		0

Ranking took .156667 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
1	82	308MACHINE R	X	.22188008
7	58	418ENGLISH-L	X	.01382437

Normalized Recall = .9691358 Normalized Precision = .8454909

Rank Recall = .3750 Log Precision = .3562

No Update for these results

(i) After Second Update

Figure 10 (continued)

Results for Query No. 28 QB11INDEX

Document Rank	Document No.	Name	Relevant	Correlation
1	15	610CHARACTER		.56195149
2	8	1109A PROGRA		.48112524
3	6	1106A NEW CE		.43643578
4	30	409RELATIONA		.41478068
5	65	505THE PLACE		.38490018
6	7	211A COMPUTE		.31622777
7	9	206ANALYSIS		.31108551
8	61	516AN EXPERI		.30618622
9	48	804COMPUTERI		.28867513
10	28	506THE ROLE		.26490647
11	20	304PHOTOGRAP		.25819889
12	57	603COMPILATI		.23145502
13	42	203NEW PHOTO		.18983160
14	51	1115SOME OBS		.18983160
15	80	1001DESIGN O		.18490007

Ranking took .056667 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
21	36	1110COMPUTER	X	.13801311

Normalized Recall = .7560976 Normalized Precision = .3091181

Rank Recall = .0476 Log Precision = 0

$$\text{Query}(1) = \text{Query}(0) + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT}(I)}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0

N = Number of retrieved documents

Alpha = 1

Update plus matching took 7.416667 seconds

(a) Initial Results (Increasing Alpha Strategy)

Continued Use of Negative Weights when Use in the First Update Fails to Move Query into Correct Region of N-Space
Cosine Correlation Function

Figure 11

Results for Query No. 28 QB11INDEX

Document Rank	Document No.	Name	Relevant	Correlation
1	5	1103FEASIBIL		0
2	52	1121SCOPE AN		0
3	55	907SATIRE 01		0
4	82	308MACHINE R		0
5	74	305THE USE O		-.00778240
6	35	1004DENSITY		-.00950143
7	50	418ENGLISH-L		-.01019877
8	77	213ADAPTIVE		-.01537145
9	43	1206THE EDUC		-.01665453
10	54	901A SELECTE		-.01848685
11	53	1118DOCUMENT		-.02109709
12	4	307ROLE OF/		-.02201194
13	39	517RESEARCH		-.02445580
14	23	815AFACET A		-.02484712
15	76	1214SOME HUM		-.02638095

Ranking took .183333 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
44	36	1110COMPUTER	X	-.08267569

Normalized Recall = .4756098 Normalized Precision = .1412683

Rank Recall = .0227 Log Precision = 0

$$\text{Query}(2) = \text{Query}(1) + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT}(I)}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0

N = Number of retrieved documents

Alpha = 2

Update plus matching took 8.716667 seconds

(b) Results after First Update (Increasing Alpha Strategy)

Figure 11 (continued)

Results for Query No. 28 QB11INDEX

Document Rank	Document No.	Name	Relevant	Correlation
1	50	418ENGLISH-L		-.00576985
2	77	213ADAPTIVE		-.00889623
3	54	901A SELECTE		-.01045874
4	29	212A STATIST		-.02426927
5	35	1004DENSITY		-.02687667
6	31	207PRACTICAL		-.03415810
7	18	1122STATE-OF		-.03968813
8	82	308MACHINE R		-.03968813
9	2	814A NEW EFF		-.04383773
10	38	303C. M. HER		-.04518692
11	74	305THE USE O		-.04843090
12	11	1117PLANNING		-.05139033
13	32	902DOCUMENT		-.05146744
14	14	518EDUCATION		-.05200079
15	64	507PARAGRAPH		-.05325330

Ranking took .150667 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
41	36	111COMPUTER	X	-.17773708

Normalized Recall = .5121951 Normalized Precision = .1572932

Rank Recall = .0244 Log Precision = 0

$$\text{Query}(3) = \text{Query}(2) + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT}(I)}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0

N = Number of retrieved documents

Alpha = 3

Update plus matching took 10.650000 seconds

(c) Results after Second Update (Increasing Alpha Strategy)

Figure 11 (continued)

Results for Query No. 28 QB11INDEX

Document Rank	Document No.	Name	Relevant	Correlation
1	35	1884DENSITY		-.02580081
2	32	902DOCUMENT		-.03087952
3	74	305THE USE O		-.03698248
4	18	1122STATE-OF		-.04762430
5	31	207PRACTICAL		-.05123559
6	4	307ROLE OF/		-.06360848
7	54	901A SELECTE		-.08127568
8	13	1119RECENT A		-.08509394
9	25	209THE REDUC		-.09170154
10	14	518EDUCATION		-.09805563
11	63	214ELECTRONI		-.10501212
12	38	303C. M. HER		-.10844530
13	81	1107THE AUTO		-.10956911
14	64	507PARAGRAPH		-.11821871
15	71	108SEARCHERS		-.12027386

Ranking took .183333 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
22	36	1110COMPUTER	X	-.13189557

Normalized Recall = .7439024 Normalized Precision = .2985615

Rank Recall = .0455 Log Precision = 0

No Update for these results

(d) Results after Third Update (Increasing Alpha Strategy)

Figure 11 (continued)

Results for Query No. 12 QA12JOURNAL

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF/	X	.64956980
2	40	1102CENTRALI		.38254603
3	19	302RECORDING		.30096463
4	46	109STATUS RE		.29774567
5	60	1212TOWARD A		.24494897
6	10	1201ACCREDIT		.20412415
7	81	1107THE AUTO		.20412415
8	75	201A SYSTEM		.20134682
9	43	1206THE EDUC		.19658927
10	21	1209SCIENCE		.17234550
11	35	1004DENSITY		.16823165
12	13	1119RECENT A		.16439899
13	57	603COMPILATI		.16306342
14	45	1215 GRADUATE		.15861032
15	38	303C. M. HER		.15713484

Ranking took .056557 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF/	X	.64956980
16	7	211A COMPUTE	X	.14907120
17	9	206ANALYSIS	X	.14664712
25	25	209THE REDUC	X	.10629880
33	42	203NEW PHOTO	X	.06711561

Normalized Recall = .8025641 Normalized Precision = .5599956

Rank Recall = .1630 Log Precision = .3886

$$\text{Query}(1) = \text{Query}(0) + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT}(I)}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

N = Number of retrieved documents

Alpha = 1

(a) Initial Results (Increasing Alpha Strategy)

Comparison of Results After Three Iterations as a Function of the Value of Alpha used in First Update Cosine Correlation Function

Figure 12

Results for Query No. 12 QA12JOURNAL

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF/	X	.97316392
2	40	1102CENTRALI		.36664099
3	19	302RECORDING		.31881481
4	46	109STATUS RE		.27034708
5	60	1212TOWARD A		.22240874
6	17	301DOCUMENT		.21357210
7	81	1107THE AUTO		.20593402
8	33	11017INFORMA		.20490176
9	22	914THE RAPID		.18534062
10	35	1004DENSITY		.16972328
11	71	108SEARCHERS		.16782759
12	32	902DOCUMENT		.16250565
13	9	206ANALYSIS	X	.15534474
14	74	305THE USE O		.15291796
15	69	813AUTOMATIC		.14994794

Ranking took .150000 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF/	X	.97316392
13	9	206ANALYSIS	X	.15534474
18	42	203NEW PHOTO	X	.14219245
27	7	211A COMPUTE	X	.11279471
32	25	209THE REDUC	X	.09651717

Normalized Recall = .8051282 Normalized Precision = .5660867

Rank Recall = .1648 Log Precision = .3919

$$\text{Query(2)} = \text{Query(1)} + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT(I)}}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0

N = Number of retrieved documents

Alpha = 2

(b) Results after First Update (Increasing Alpha Strategy)

Figure 12 (continued)

Results for Query No. 12 QA12JOURNAL

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF/	X	.88970072
2	9	206ANALYSIS	X	.57284742
3	40	1102CENTRALI		.40983981
4	19	302RECORDING		.27680043
5	46	109STATUS RE		.27678442
6	33	11017INFORMA		.27389171
7	60	1212TOWARD A		.26646284
8	7	211A COMPUTE	X	.25798837
9	75	201A SYSTEM		.24558033
10	17	301DOCUMENT		.24107313
11	22	914THE RAPID		.23551009
12	81	1107THE AUTO		.22878123
13	69	813AUTOMATIC		.22320053
14	73	906MICROFILM		.21701659
15	71	108SEARCHERS		.20838206

Ranking took .183333 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF/	X	.88970072
2	9	206ANALYSIS	X	.57284742
8	7	211A COMPUTE	X	.25798837
16	42	203NEW PHOTO	X	.20575649
24	25	209THE REDUC	X	.16819640

Normalized Recall = .9076923 Normalized Precision = .7701336

Rank Recall = .2941 Log Precision = .5488

$$\text{Query}(3) = \text{Query}(2) + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT}(I)}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0

N = Number of retrieved documents

Alpha = 3

(c) Results after Second Update (Increasing Alpha Strategy)

Figure 12 (continued)

Results for Query No. 12 QA12JOURNAL

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF/	X	.76756620
2	9	206ANALYSIS	X	.68759401
3	7	211A COMPUTE	X	.53283944
4	40	1102CENTRALI		.44774693
5	42	203NEW PHOTO	X	.31186705
6	75	201A SYSTEM		.28223262
7	33	11017INFORMA		.27951524
8	60	1212TOWARD A		.27811415
9	46	109STATUS RE		.27545574
10	22	914THE RAPID		.27181938
11	56	1112THE USE		.26324643
12	48	804COMPUTERI		.26323561
13	73	906MICROFILM		.25698269
14	81	1107THE AUTO		.25179059
15	20	304PHOTOGRAP		.24952061

Ranking took .183333 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF/	X	.76756620
2	9	206ANALYSIS	X	.68759401
3	7	211A COMPUTE	X	.53283944
5	42	203NEW PHOTO	X	.31186705
19	25	209THE REDUC	X	.22573736

Normalized Recall = .9615385 Normalized Precision = .9089968

Rank Recall = .5000 Log Precision = .7545

No Update for these results

(d) Results after Third Update (Increasing Alpha Strategy)

Figure 12 (continued)

Results for Query 12 QA12JOURNAL

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF/	X	.64956980
2	40	1102CENTRALI		.38254603
3	19	302RECORDING		.30096463
4	46	109STATUS RE		.29774567
5	60	1212TOWARD A		.24494897
6	10	1201ACCREDIT		.20412415
7	81	1107THE AUTO		.20412415
8	75	201A SYSTEM		.20134682
9	43	1206THE EDUC		.19658927
10	21	1209SCIENCE		.17234550
11	35	1004DENSITY		.16823165
12	13	1119RECENT A		.16439899
13	57	603COMPILATI		.16366342
14	45	1215 GRADUATE		.15861032
15	38	3030. M. HER		.1571384

Ranking took .083333 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF/	X	.64956980
16	7	211A COMPUTE	X	.14907120
17	9	206ANALYSIS	X	.14684712
25	25	209THE REDUC	X	.10629880
33	42	203NEW PHOTO	X	.06711561

Normalized Recall = .8025641 Normalized Precision = .5599956

Rank Recall = .1630 Log Precision = .3886

$$\text{Query(1)} = \text{Query(0)} + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT(I)}}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

N = Number of retrieved documents

Alpha = 2

(e) Initial Results (Constant Alpha Strategy)

Figure 12 (continued)

Results for Query No. 12 QA12JOURNAL

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF/	X	.99146269
2	40	1102CENTRALI		.34989921
3	19	302RECORDING		.30969005
4	46	109STATUS RE		.25531480
5	33	11017INFORMA		.22165517
6	17	301DOCUMENT		.22003279
7	60	1212TOWARD A		.21004201
8	81	1107THE AUTO		.19837301
9	22	914THE RAPID		.18670401
10	71	108SEARCHERS		.173689250
11	35	1004DENSITY		.16349177
12	32	902DOCUMENT		.16114324
13	37	306PROGRESS		.154C3081
14	73	906MICROFILM		.15244319
15	69	813AUTOMATIC		.15105098

Ranking took .150000 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF/	X	.99146269
16	9	206ANALYSIS	X	.15089864
19	42	203NEW PHOTO	X	.14963301
28	7	211A COMPUTE	X	.10226200
31	25	209THE REDUC	X	.09115047

Normalized Recall = .7948718 Normalized Precision = .5505320

Rank Recall = .1579 Log Precision = .3835

$$\text{Query}(2) = \text{Query}(1) + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT}(I)}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

N = Number of retrieved documents

Alpha = 2

(f) Results after First Update (Constant Alpha Strategy)

Figure 12 (continued)

Results for Query No. 12 QA12JOURNAL

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF/	X	.99758179
2	40	1102CENTRALI		.33804448
3	19	302RECORDING		.30263667
4	46	109STATUS RE		.24496349
5	33	11017INFORMA		.23067222
6	17	301DOCUMENT		.22284037
7	60	1212TOWARD A		.20152621
8	81	1107THE AUTO		.19281829
9	22	914THE RAPID		.1865934
10	71	108SEARCHERS		.17669065
11	63	214ELECTRONI		.1598643
12	32	902DOCUMENT		.15951821
13	35	1004DENSITY		.15891377
14	73	906MICROFILM		.15743547
15	37	306PROGRESS		.15674261

Ranking took .150000 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF/	X	.99758179
16	42	203NEW PHOTO	X	.15338289
18	9	206ANALYSIS	X	.14746183
28	7	211A COMPUTE	X	.09539051
32	25	209THE REDUC	X	.08745493

Normalized Recall = .7948718 Normalized Precision = .5518355

Rank Recall = .1579 Log Precision = .3842

$$\text{Query}(3) = \text{Query}(2) + \text{Alpha} \sum_{I=1}^{N} \frac{\text{WEIGHT}(I)}{(1)} * \text{Document of Rank I}$$

Document Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

N = Number of retrieved documents

Alpha = 2

(g) Results after Second Update (Constant Alpha Strategy)

Figure 12 (continued)

Results for Query No. 12 QA12JOURNAL

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF/	X	.99887910
2	40	1102CENTRALI		.3342788
3	19	302RECORDING		.29979583
4	46	109STATUS RE		.24097886
5	33	11017INFORMA		.23373192
6	17	301DOCUMENT		.22365318
7	60	1212TOWARD A		.19824818
8	81	1107THE AUTO		.19062321
9	22	914THE RAPID		.18638714
10	71	108SEARCHERS		.17754277
11	63	214ELECTRONI		.16331062
12	73	906MICROFILM		.15910194
13	32	902DOCUMENT		.15878053
14	37	306PROGRESS		.15758186
15	35	1004DENSITY		.15710467

Ranking took .133333 seconds

Relevant Document Ranks

Document Rank	Document No.	Name	Relevant	Correlation
1	4	307ROLE OF/	X	.99887910
16	42	203NEW PHOTO	X	.15460210
18	9	206ANALYSIS	X	.14607761
28	7	211A COMPUTE	X	.09280767
34	25	209THE REDUC	X	.08603237

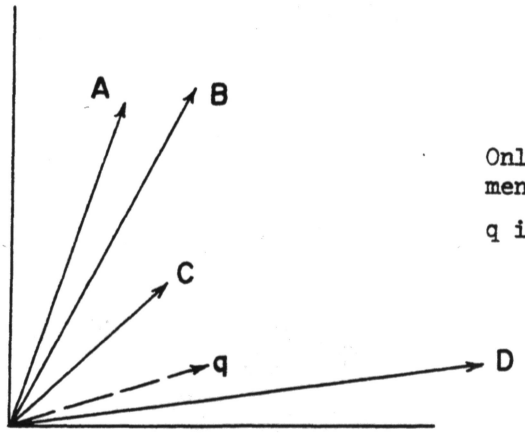
Normalized Recall = .7897436 Normalized Precision = .5482948

Rank Recall = .1546 Log Precision = .3823

No Update for these results

(h) Results after Third Update (Constant Alpha Strategy)

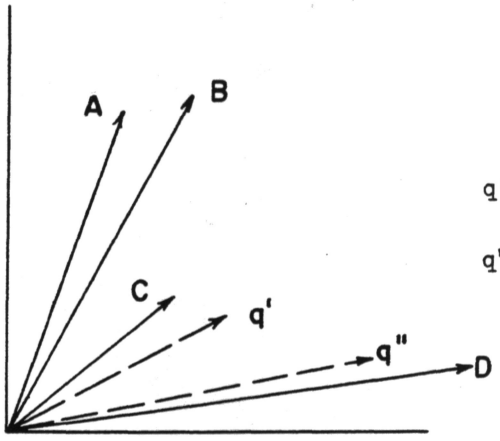
Figure 12 (continued)



Only relevant documents are shown.

q is the initial query.

A. Initial configuration



q' obtained using small value of α

q" obtained using larger value of α

B. Configuration after one modification

Effect of the Initial Value of α used in the Iteration Process

Figure 13

Correlation Function	Strategy	Iteration	Relevant Document Ranks	Normalized Recall	Normalized Precision
Cosine	Increasing Alpha	0	3,4,14,19,49,68	.706	.512
		1	1,2,5,42,52,67	.680	.613
		2	1,2,5,47,51,68	.669	.607
		3	1,2,5,49,51,68	.665	.605
	Constant Alpha	0	3,4,14,19,49,68	.706	.512
		1	1,2,5,47,51,68	.669	.607
		2	1,2,5,47,51,68	.669	.607
		3	1,2,5,49,51,68	.665	.605
	Correlations	0	3,4,14,19,49,68	.706	.512
		1	1,3,5,39,55,67	.677	.593
		2	1,2,5,47,48,70	.671	.609
		3	1,2,7,45,49,71	.667	.592
Co-occurrence	Increasing Alpha	0	2,7,25,30,48,65	.662	.455
		1	1,5,9,44,47,75	.654	.533
		2	1,3,9,34,50,74	.675	.570
		3	1,3,12,34,49,73	.673	.557
	Constant Alpha	0	2,7,25,30,48,65	.662	.455
		1	1,6,8,45,47,76	.650	.528
		2	1,3,8,36,49,74	.675	.574
		3	1,3,10,35,51,74	.669	.562
	Correlations	0	2,7,25,30,48,65	.662	.455
		1	1,3,26,39,57,60	.643	.513
		2	1,2,16,43,57,70	.636	.546
		3	1,5,9,43,50,75	.649	.531
Simple Vector Matching	Increasing Alpha	0	2,8,9,11,46,72	.725	.548
		1	2,3,4,13,32,77	.762	.646
		2	1,3,4,14,29,77	.768	.682
		3	1,2,4,14,27,78	.773	.706
	Constant Alpha	0	2,8,9,11,46,72	.725	.548
		1	1,3,4,13,30,77	.768	.684
		2	1,3,4,14,28,78	.768	.683
		3	1,2,4,14,27,78	.773	.706
	Correlations	0	2,8,9,11,46,72	.725	.548
		1	1,3,6,18,29,79	.751	.647
		2	1,2,7,24,65,80	.658	.604
		3	1,2,8,25,65,79	.656	.600

Query QA15COST

The Result of the Query Modification by Use of Relevance Information as a Function of the Number of Relevant Documents Initially in the Fifteen Highest Correlated Documents

Figure 14