XIII. The Use of Relevant Documents Instead of Queries
in Relevance Feedback

R. G. Crawford and H. Z. Melzer

Abstract

The use of relevance feedback in the automatic document retrieval process has numerous realizations. This paper deals with the use of a relevant document as the updated query for the subsequent iteration in a search. Experiments are also performed using a known relevant document as the original query submitted to the system. The experiments are performed within the structure of the SMART retrieval system, using the 200 document Cranfield collection. Recall and precision measures are taken to evaluate the results of the experiments and suggestions are made for possible further investigations.

1. Introduction

The improved retrieval yielded by user interaction with an automated information retrieval system has been well established by previous work in this area [2] and has justified further experiments with this aspect of the retrieval process. The problem dealt with in this paper is the implementation of relevance feedback information for query modification to maximize the effect of this interaction. Specifically, the method under consideration is that of using only the relevant documents retrieved as the updated query submitted for the subsequent iteration of the search procedure. This is an extension of the work done by Ide [2]

with the query update formula:

$$Q_{i+1} = \alpha Q_i + \beta Q_o + \gamma \sum_{i=1}^{N_1} R_i + \delta \sum_{i=1}^{N_2} N_i + \sum_{i=1}^{N_3} \omega_i d_i + \sum_{i=1}^{N_4} \nu_i c_i \ .$$

Experiments with this formula have been performed by Ide and by Riddle,

Horwitz and Dietz [4] with,

$$\alpha=1, \ \beta=0, \ \gamma=1, \ \delta=0, \ \omega_i=0, \ \nu_i=0 \quad \text{and}$$

$$\alpha=0, \ \beta=1, \ \gamma=1, \ \delta=0, \ \omega_i=0, \ \nu_i=0 \ .$$

All these approaches utilize some combination of the original query, the

previous query, the relevant documents retrieved, and the nonrelevant

documents retrieved. The current study considers the case of

$$\alpha=\beta=\delta=\omega_i=\nu_i=0$$

thus completely substituting the relevant documents for the original

query. The equation is then:

$$Q_{i+1} = \gamma \sum_{i=1}^{N_1} R_i \quad .$$

Specifically, it allows $\gamma=1$, $N_1=1$ to begin and proceeds with experiments

where $N_1>1$. Thus,

$$Q_2 = r_1$$

A natural extension of this approach is to consider the case when the user comes to the system with a known relevant document rather than a query. It is desired to discover whether there is an advantage in his submitting the actual vector of the known relevant document with the hope of retrieving more documents similar to it, or if he is better off constructing a query vector to retrieve documents dealing with the appropriate subject matter.

This problem of submitting a known relevant document as the original "query" divides into two cases:

1. The document is in the collection. In this case, it is possible to retrieve the document vector of this known relevant document and submit it as the query. It is apparent that such a procedure, in a real situation, would maximize the consistency of construction of the query vector and the document vectors and yield some gain in retrieval for this reason alone.

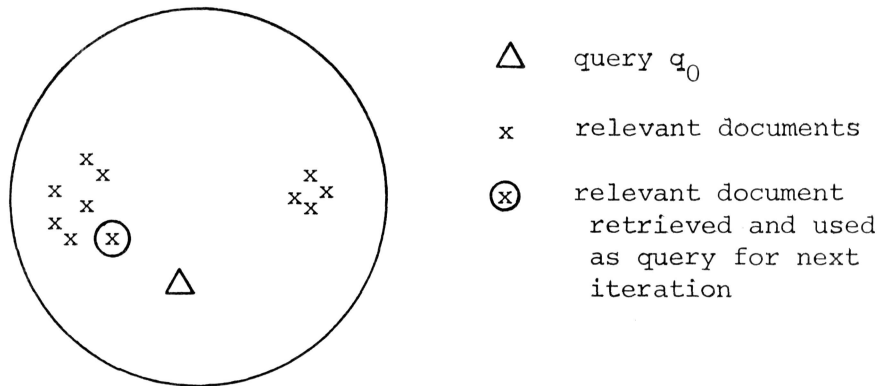2. The document is not in the collection, in which case a document vector must be constructed.

Further motivation for the strategy and implementation of the methods actually used in these experiments is contained in the next section.
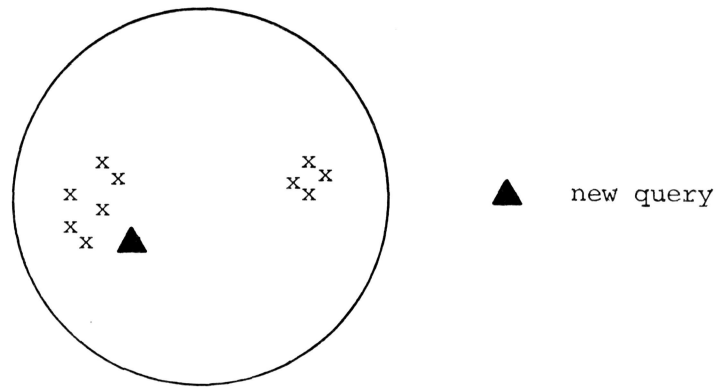
2. Motivation and Assumptions

The assumptions made both in the strategy used in carrying out this study and in the evaluation of the results of the experiments are crucial to an understanding of the work.

First, it is assumed that the user of an automatic document retrieval system is well versed in the particular area he is working in, and that he comes to the system with a fairly good idea of what type of document he is looking for. He constructs a query which he hopes will retrieve several, but at least one, document relevant to his needs. Once he is shown the documents retrieved by his query, he is able to judge whether any of these documents are relevant. Furthermore, he is able to judge whether one (or more) of these documents is a better expression of what he is searching for than his original query. In this case, substituting this "very relevant" document for the query brings him closer to other documents similar, in some way, to the retrieved one, and the next iteration retrieves these other documents for him. Of course, the procedure has the drawback of pulling the query vector away from possibly relevant documents which are distant in the document space from the "very relevant" document. An illustration is contained in Fig. 1. For this reason, the procedure must be restricted to the case where the retrieved document is "very relevant" rather than merely relevant. In the latter case, the standard feedback procedure [2] previously proven effective can be chosen. Thus, the crucial assumption here is that a user in a real system is able to determine whether a given document, in vector form, expresses his request better than his original query, in vector form.

Second, it is assumed that in a real situation, quite often the user knows of a document relevant to his needs before approaching the system. His request is for more documents similar to the known one. Thus, the problem arises of whether he should construct a query vector in an attempt to retrieve documents similar to the known relevant one, or simply submit the relevant

△   query q_0

x   relevant documents

⊗   relevant document
    retrieved and used
    as query for next
    iteration

(i)

▲   new query

(ii)

⊗   relevant documents
    retrieved by the
    new query
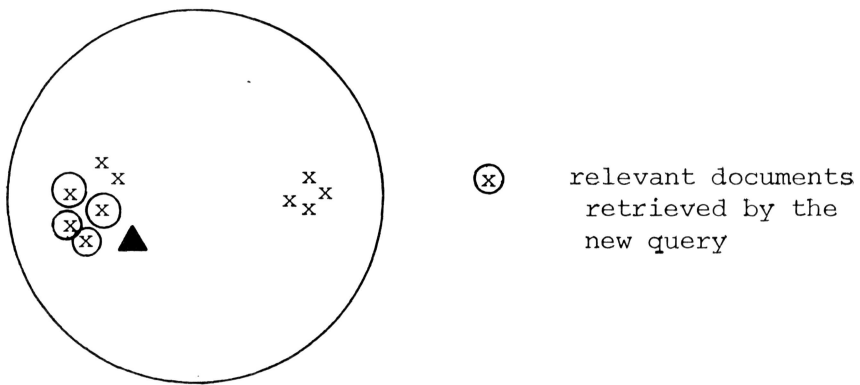
(iii)

Illustration of Document Feedback

Fig. 1

document as his query. This question is investigated in the second part of
this study.

Third, it is assumed that in a real situation, the document collection
is large enough so that the document known by the user may be expected to
be contained in the collection, and thus the gain due to consistency of
construction of the vectors referred to above is a realistic one.

Under these assumptions, the real situation being referred to is that
of the user of an actual document retrieval system as opposed to that of the
experimenter working in an experimental environment. The distinction is that
the actual system has a larger document collection which is more complete
in the sense of containing a large number of the documents in the area which
it serves. Taking these assumptions into consideration, the claims which
are made become somewhat clearer. The use of the feedback algorithm con-
sidered, which updates the query with a relevant document only, must be
limited to the case when the user has judged such a document "very relevant".
The use of a known relevant document in place of an original query is clearly
limited to the case when the user wants to retrieve documents of a nature
similar to the given one. The query is then immediately in the proximity
of other documents similar to the known relevant one, as in Fig. 1(b).

3. Implementation

The document collection used for this study is the Cranfield 200
collection, consisting of 200 documents in the field of aerodynamics, with
a collection of 42 associated queries. For the experiments dealing with

the feedback algorithm, the thesaurus form of the document and query

vectors is used. The experiment using a document as the original query

is performed using the word form or suffix 's' document and query vectors.

It is established [6] that these two forms yield close enough results to

be comparable.

The program which does the basic search and retrieval is a

FORTRAN program using the same methods and achieving the same retrieval

as that of the SMART retrieval system [7]. Modifications are made to

allow for variations in the parameters such as the number of relevant

documents used for the new query from one iteration to the next. The

data input is set so that all coefficients of the feedback formula are

zero except for the coefficient of the relevant documents which is

set to one. For the feedback experiment, the algorithm is as follows:

> Given a query,
>
> 1. Do a standard full search and display the top 5 or 10
>   correlated documents for the user as the first group
>   retrieved.
>
> 2. (i) If there are any relevant documents in the dis-
>   played set, let the top ranked relevant document be
>   used for the query modification, i.e. let this docu-
>   ment be used as the query for the next iteration.
>   (ii) If there are no relevant documents retrieved, do
>   a standard feedback iteration using the formula,
>
> $$Q_1 = Q_o - n_1$$
>
> where n_1 is the top correlated nonrelevant document.
>
> 3. Iterate this process, incrementing the number of
>   relevant documents used for updating by one for each
>   iteration (if these documents are available), as many
>   times as needed and practical.

For the experiment using documents as original queries, the procedure is simply to submit the appropriate document vector as the query, Q_o, and proceed as in the feedback experiment. The implementation of this part of the study merits some closer examination. It is not clear a priori what is to be used as relevance judgments for documents submitted as queries, since there is no study currently available for the Cranfield collection providing document-document relevant judgments. This makes it difficult to determine success of a technique in an experimental situation and to evaluate results. However, a special characteristic of the Cranfield collection allows such an experiment to be done with realistic evaluation. The 42 queries for the collection have been constructed by the authors of 42 documents in the larger document collection of which the Cranfield 200 is a subset. They are constructed so as to best describe this set of "source" documents written by these authors. Furthermore, the relevance judgments for the queries can now be considered as valid relevance judgments for these documents. The "source documents" are not included in the document collection under study but are used as the documents substituted as original queries. It is now apparent that the documents in the collection judged relevant to a given query are also relevant to the source document for that query. The comparison can be made of retrieval using a query versus retrieval using a known relevant document.

4. Results

A) Feedback Using Only Relevant Documents

The experiments performed using the query update formula

$$Q_{i+1} = \sum_{i=1}^{N} R_i$$

where n=1 on the first iteration, two on the second, and three on the third yield results which show significant improvement over the initial retrieval. Thus, it is found that the recall and precision measures, calculated over the 37 queries for which at least one document in the top ten retrieved is relevant, increase from one iteration to the next, as seen in Table 1. To study these measures more closely, two subsets of the queries are taken. The first subset consists of all the queries which retrieve no relevant document in the top ten retrieved by the initial query. In this case, negative feedback [3] is used to yield better retrieval for the next iteration. Once a relevant document is retrieved, it is submitted as the query for the subsequent iteration. Fig. 2 indicates the improvement obtained with this subset. The second subset consists of those queries which retrieve only one relevant document in the first 10 documents retrieved. This document is submitted as the query for the next iteration, and the process repeated, with the top two and three relevant documents added to comprise the queries for iterations two and three respectively. Fig. 3 indicates that improvement is significant from the initial search to the subsequent iteration but does not improve from the second to the third iterations.

In comparing the results of this experiment with those of previous experiments done with the query update formula

$$Q_{i+1} = Q_o + \sum_{i=1}^{N_1} R_i$$

n	Initial Search		First Iteration		Second Iteration		Third Iteration	
	P	R	P	R	P	R	P	R
5	.330	.372	.438	.794	.503	.567	.530	.597
10	.252	.567	.295	.665	.308	.695	.306	.689
15	.186	.628	.214	.726	.218	.738	.216	.732

Recall-Precision Measures After n Documents Retrieved
over 37 Queries

Table 1
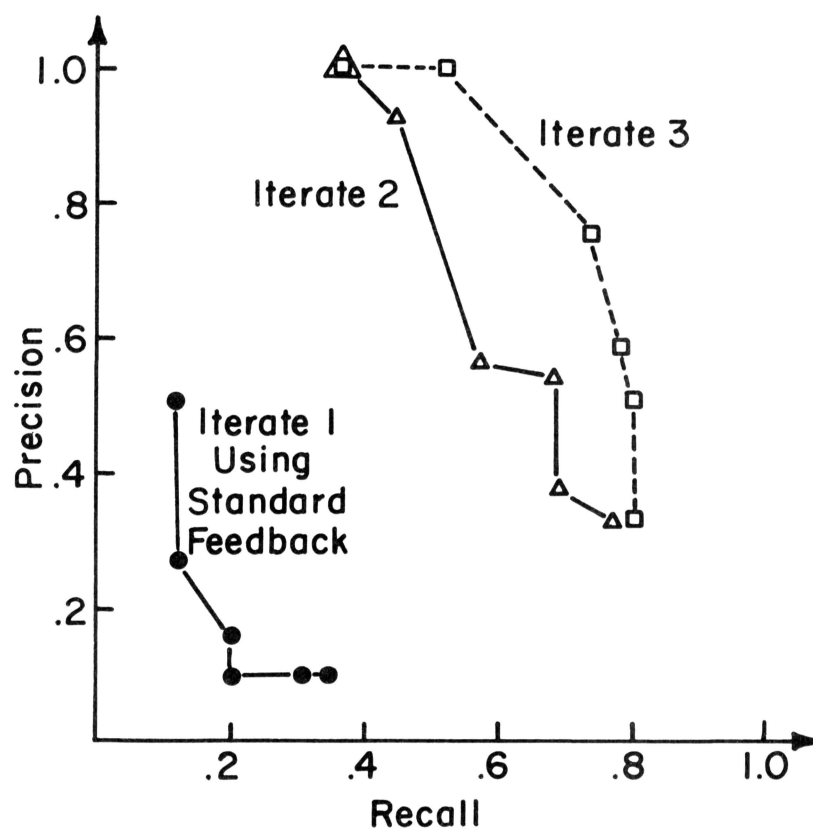
Recall-Precision Measures for Queries Retrieving
No Relevant Documents in the Top 10 Retrieved

Queries 1,19,34,41

Fig. 2

Recall-Precision Measures for Queries Retrieving
One Relevant Document in the Top 10 Retrieved

Queries 2,3,4,9,11,14,17,22,30,33

Fig. 3

it is found that there is not a significant difference in the recall
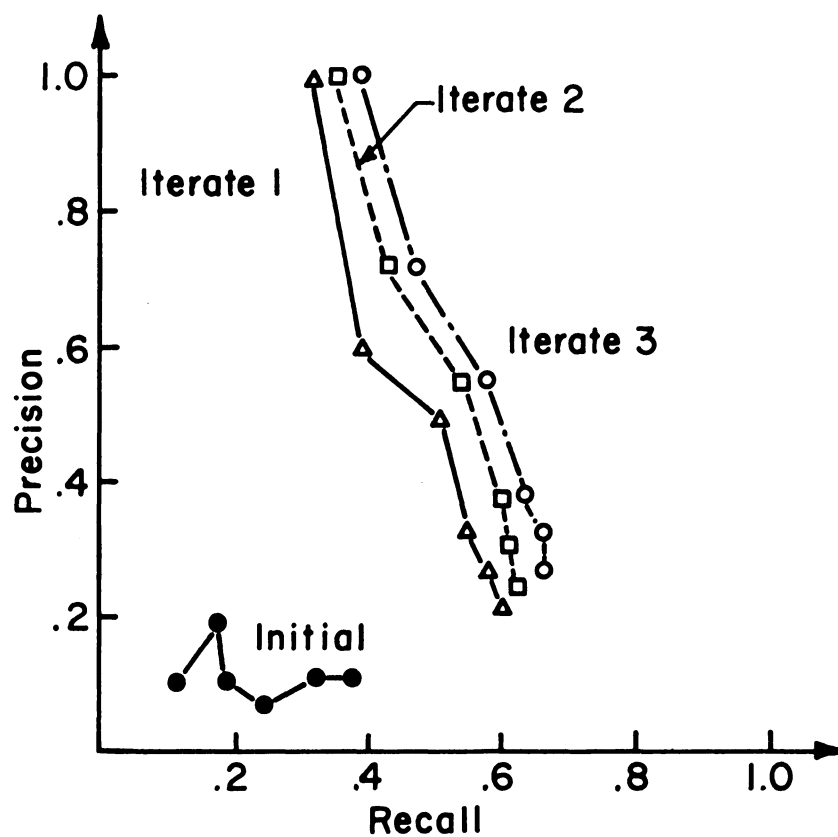and precision measures calculated over the 42 queries. These results
are shown in Table 2. Furthermore, the results judged with respect
to number of relevant documents retrieved and ranks of the relevant
documents retrieved show that in 21 of the 42 queries the results are
identical. The implications to be drawn from these results are that:

1. significant improvement in retrieval is obtained
    on an iteration using only the relevant documents
    as the query for the next iteration.

2. there is no significant difference between using
    only relevant documents and using relevant documents
    in combination with the original query, thus it seems
    that the contribution made by the original query is
    minimal.

Experiments which allow the user to see, and hence judge, only
the top five correlated documents indicate that this may not be a re-
commended procedure. Results different from those in which the user
considers ten documents retrieved are obtained only in the case of six
queries. These are the queries for which the initial search yields a
relevant document in the first ten retrieved but no relevant document
in the first five retrieved. In the latter event, negative feedback is
used but produces considerably poorer results than in the case where
the user is shown the top ten documents. This is seen in Table 5.

It is thus found that if the feedback algorithm suggested is to
be used, it is preferable for the user to be shown ten documents to be
judged rather than five since he then clearly has a better chance of
judging a document relevant.

	Relevant Documents Only			Relevant Documents and Original Query		
	5	10	15	5	10	15
Initial Search	.328	.500	.581	.328	.500	.581
Iteration 1	.452	.613	.677	.484	.624	.704
Iteration 2	.559	.677	.720	.554	.661	.704
Iteration 3	.591	.677	.715	.581	.672	.704

a) Recall

	Relevant Documents Only			Relevant Documents and Original Query		
	5	10	15	5	10	15
Initial Search	.300	.230	.180	.300	.230	.180
Iteration 1	.415	.285	.208	.445	.290	.217
Iteration 2	.515	.313	.222	.510	.305	.217
Iteration 3	.545	.313	.220	.535	.310	.217

b) Precision

Comparison of Recall-Precision Measures for Feedback Using
Only Relevant Documents versus Feedback Using Relevant
Documents and the Original Query, for 4 Iterations,
at Cutoffs of 5, 10, and 15 Documents Retrieved
(averages over 42 queries)

Table 2

	5 Retrieved		10 Retrieved		15 Retrieved	
	Query	Source Document	Query	Source Document	Query	Source Document
1	0	1	0	2	0	2
2	0	4	1	4	2	4
3	0	2	1	3	2	4
4	1	1	1	2	1	2
5	2	1	2	2	3	2
6	3	3	3	3	3	3
7	2	1	2	1	3	1
8	1	1	2	1	3	2
9	0	0	1	0	1	0
10	1	1	2	3	2	3
11	0	1	1	2	2	2
12	3	4	6	6	6	6
14	1	1	1	1	1	1
15	4	2	7	3	7	3
16	1	2	2	3	3	3
17	0	1	1	1	1	2
18	3	2	3	2	3	2
19	0	2	0	4	1	4
20	1	0	2	0	2	0
21	2	2	2	2	2	2
22	0	1	1	1	1	2
23	1	2	2	2	2	2
24	2	5	4	7	4	7
25	2	2	2	2	2	2

Number of Relevant Documents Retrieved
Out of 5, 10, or 15 Retrieved Items

Table 3

	Query	Source Document
1		3,9
2	9	1,2,3,4
3	7	1,3,8
4	2	3,7
5	1,2	3,7
6	1,3,5	1,2,3
7	1,3	2
8	3,9	2
9	8	
10	5,6	2,6,10
11	6,10	5,9
12	2,3,5,6,7,8	1,2,3,5,6,10
14	1	1
15	1,2,3	1,4,7
16	2,9	1,5,7
17	7	1
18	1,2,4	1,4
19		1,3,6,10
20	2,9	
21	1,3	3,4
22	9	3
23	5,10	1,2
24	1,3,6	1,2,3,4,5,6,8
25	1,2	1,2

Ranks of Relevant Documents Retrieved Using
a Cutoff of 10 Documents Retrieved

Table 4

Ranks higher using source documents - 18 instances
Ranks higher using original queries - 6 instances

Query Number	User Sees 10 Documents		User Sees 5 Documents	
	Initial Search	Iterate 1	Initial Search	Iterate 1
2	1	4	0	4
3	1	4	0	3
9	1	1	0	0
11	1	1	0	1
17	1	5	0	0
22	1	1	0	1

Number of relevant documents retrieved, in the case where
one document is retrieved in the top 10 where one document
is relevant, but none are retrieved in top 5.

Queries 2,3,9,11,17,22

Table 5

B) Source Document Used as Original Query

Experiments using source documents in place of original queries yield interesting and promising results. It is found that in more than 50% of the queries studied, there is significant improvement using source documents, both in the number of relevant documents retrieved at various cutoff points and in the ranks of the relevant documents retrieved. Data are contained in Tables 3 and 4. It is clear that use of the source document yields a higher correlation with the relevant documents and thus retrieves the relevant documents earlier. One of the explanations for this is given by the fact that the source document vector is significantly longer and thus consists of more concepts than the query vector. Thus, there is greater probability of high correlation with relevant documents. Throughout this analysis, it must be remembered that the assumption is made that the relevance judgments made for the query are applicable to the source document used.

The earlier and wider retrieval of relevant documents yields better recall and precision figures both for the initial retrieval and for the first feedback iteration, as seen in Figs. 4 and 5. The main focus of this experiment is to determine whether improvement occurs in the initial retrieval results by use of a source document instead of a query, and it is in this respect that the most significant improvement is found. Improvement obtained by further feedback iteration over the initial results is assumed, and it is not found to be significantly different from that obtained in feedback iterations where a query was used for initial retrieval.
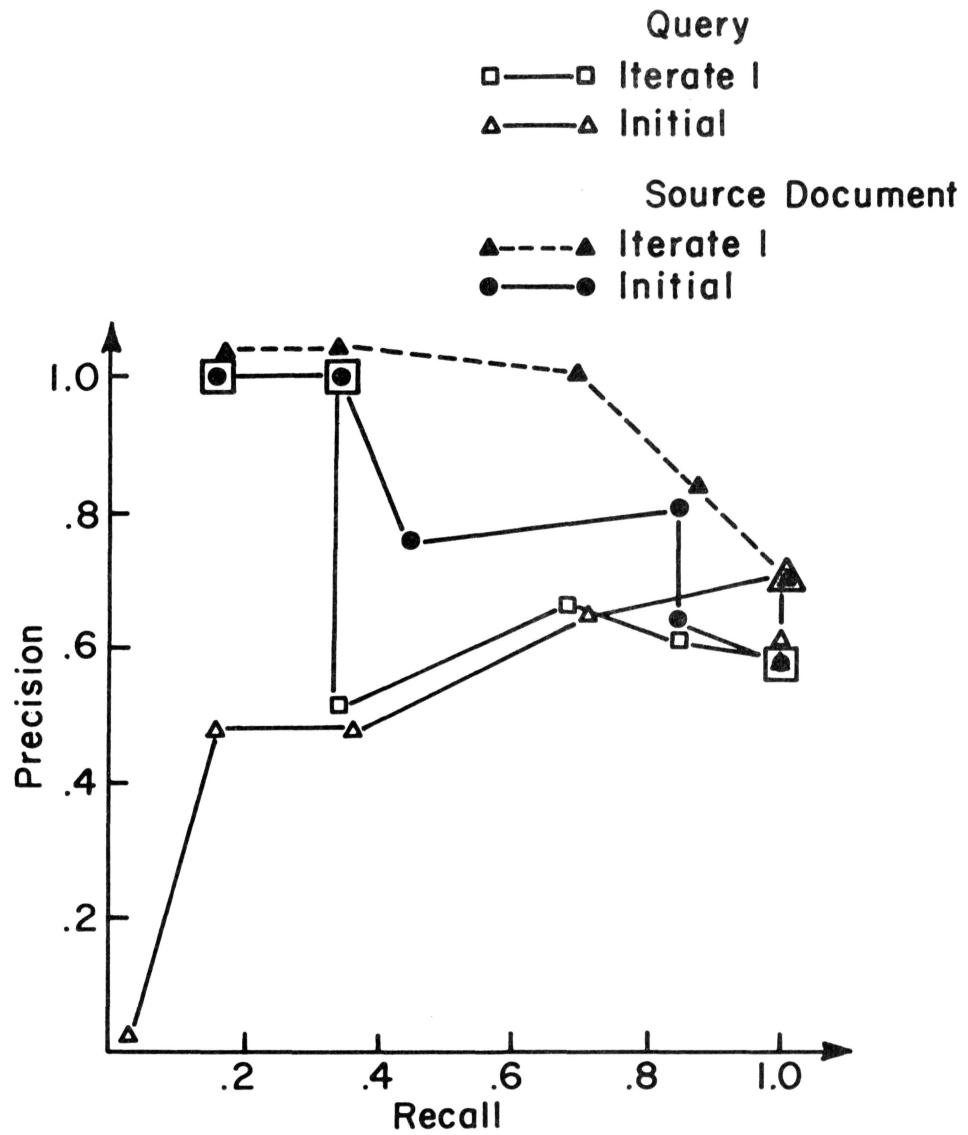
Recall-Precision Measures for Query 12 and Source Document 12

Fig. 4
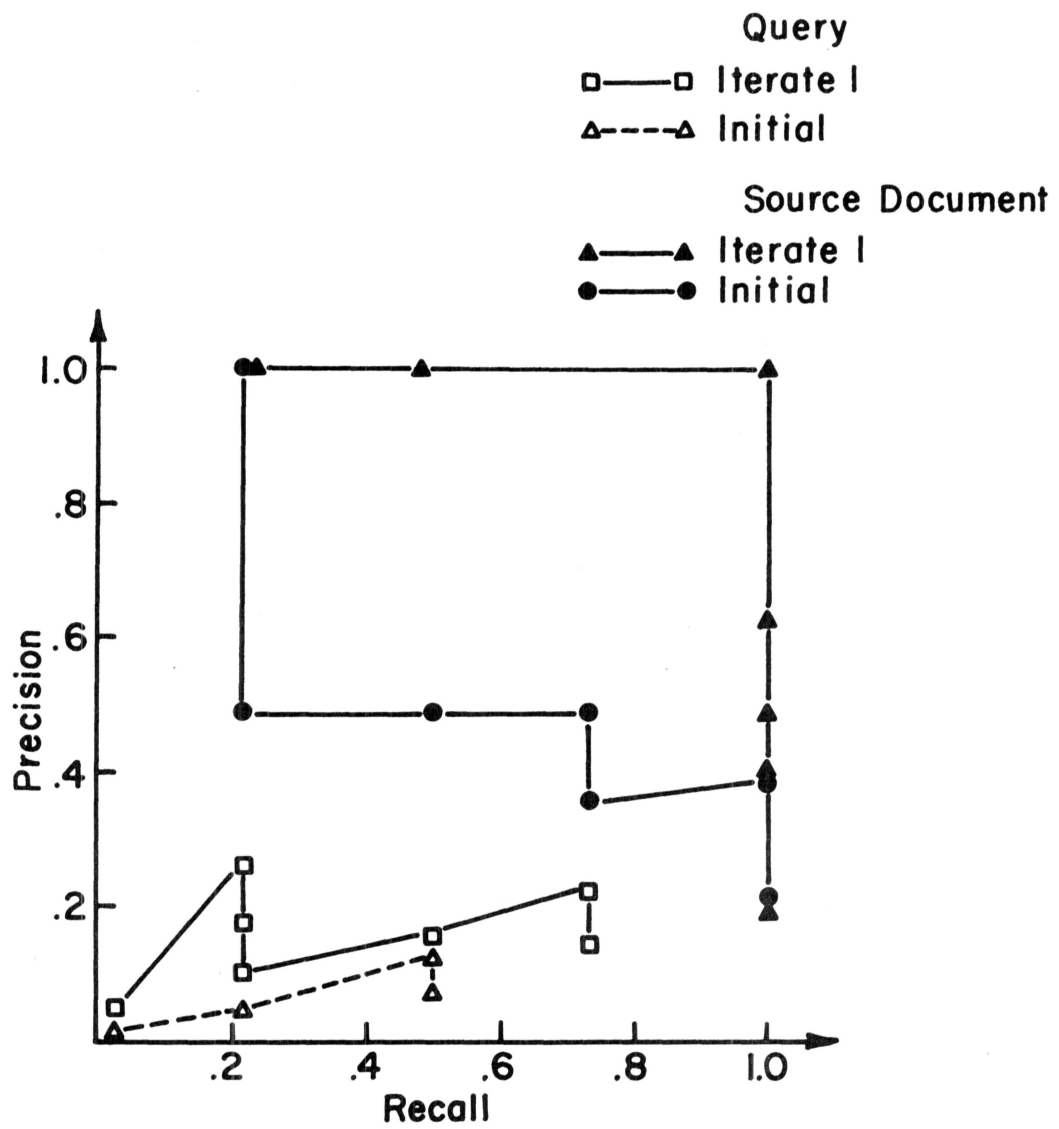
Recall-Precision Measures for Query 19 and Source Document 19

Fig. 5

Some typical examples of source documents retrieving relevant documents significantly faster than original queries are given by queries 1, 2, and 19 as shown in Fig. 6.

These results indicate that there is good reason to believe that a user approaching a system knowing a "very relevant" document might be advised to use the vector representation of that document rather than to construct a query vector which may or may not retrieve as many relevant documents as quickly.

5. Conclusions

It is apparent that the more significant results are obtained from the second experiment described. The approach of using documents as original queries is new and merits further investigation. It is found that a user may do signficantly better using a document which he knows to be relevant rather than a query to formulate his request. The first experiment leads to the conclusion that feedback using only relevant documents does not do much better or worse than the more standard feedback techniques when considered over an entire query collection. This experiment also reinforces the well-established idea that feedback does improve retrieval and must be considered a significant part of the retrieval process.

The relatively successful results of the second experiment may be attributed to two possible influences. First, as previously noted, the document vectors allow more concepts to be compared with the other document vectors than do query vectors. Second, the restricted nature of the subject matter of the document collection may have the tendency

Query 1		Source Document 1		Query 2		Source Document 2		Query 19		Source Document 19	
Rank		Rank		Rank		Rank		Rank		Rank	
1	125	1	176	1	180	1	107 R	1	98	1	123 R
2	12	2	141	2	9	2	105 R	2	86	2	102
3	159	3	1 R	3	16	3	108 R	3	102	3	122 R
4	64	4	30	4	197	4	106 R	4	94	4	81
5	66	5	13	5	61	5	137	5	25	5	94
6	123	6	27	6	115	6	176	6	81	6	125 R
7	67	7	73	7	160	7	141	7	66	7	86
8	65	8	96	8	28	8	180	8	58	8	195
9	24	9	21 R	9	105 R	9	13	9	109	9	193
10	29	10	31	10	181	10	167	10	18	10	124 R
11	190	11	49	11	14	11	139	11	64	11	191
12	25	12	156	12	108 R	12	16	12	111	12	167
13	127	13	173	13	56	13	58	13	143	13	163
14	136	14	59	14	191	14	172	14	196	14	103
15	85	15	29	15	128	15	29	15	123 R	15	80
16	59	16	32	16	62	16	160	16	99	16	64
17	88	17	157	17	7	17	166	17	114	17	20
18	4	18	60	18	172	18	197	18	122 R	18	109
19	160	19	171	19	190	19	56	19	93	19	111
20	5	20	140	20	3	20	191	20	160	20	58
27	0	24	0	37	106 R			21	125 R		
32	21 R	51	22 R	38	107 R			25	0		
33	22 R	112	0					62	124 R		
45	0										
89	1 R										

Ranks of Retrieved Documents
Original Query vs. Source Document

Fig. 6

to produce document vectors with similar concepts throughout. The first influence is inherent in using a document vector. The second influence can be overcome by testing the technique over different and wider document collections. It must be reiterated here that a real situation would allow a user to use a document as his query only when his request was best represented by just that document rather than a query.

The results of the first experiment must also be considered in light of the fact that in a real situation the user is advised to use the specific form of the update formula discussed here only when the relevant document retrieved expresses his request better than does his original query. In an experimental environment, it is difficult to simulate such a situation. However, the experiment indicates that, in general, the user does not do any worse using only a relevant document for updating the query. Thus, the assumptions and restrictions discussed earlier are crucial. The claim is made only that improvement occurs under the stated conditions.

Throughout this study, reference is made to suggested experiments to be performed.

1. Use the same methods with a variation in the number of documents seen by the user at each iteration. This number can either be fixed or dependent on when the first relevant document is retrieved.

2. Use the same methods with a variation in the number of relevant documents used for query modification.

Thus, if two relevant documents are retrieved either can be used
for query modification by simply adding the vectors, or another
concept for query modification may be introduced. One of the
possible variations suggested is to consider the two documents
retrieved and judged relevant by the user, correlate them with
each other, and if the correlation is "high", proceed with
addition of the two vectors. If the relevant documents do not
have a "high" correlation, it is asssumed that they belong to
different document clusters and thus a simple addition for the
updated query would take the query vector away from each rele-
vant document and probably not yield very successful retrieval.
A type of splitting procedure may be recommended here which
would take each relevant document and return it as the new query
on the next iteration. This, hopefully, would retrieve more
documents in either cluster. Such a procedure can be imple-
mented by use of the method described within the query splitting
procedure suggested by Borodin, Kerr and Lewis in their work
on query splitting [8].

3. Use the same methods, but only show the user documents
he has not previously seen and judged relevant or nonrelevant.

4. Modify the above methods in the event that a document retrieved
is judged relevant but not "very relevant". In this case, the
consideration of the relevant document along with negative
feedback is suggested. This allows the user to opt against
judging a document sufficiently relevant to substitute as a
query. Standard feedback query modification is used in this
case.

5. Use the same methods applied to different, preferably larger,
collections with all the variations previously suggested.

6. Perform a more substantial test of the effectiveness of using
a known relevant document as an original query when the user
specifically wants documents similar to the submitted one.

It is clear that numerous variations exist of the concept of relevance feedback and that many of these merit investigation, since it has been shown that feedback does produce better retrieval in a large number of cases. There is a need for further work on the problem of using a document rather than a query as a request in an automatic document retrieval system. This problem may become much more realistic in an actual retrieval system where a user is well-versed in the field and recognizes a document "very relevant" to his needs, which may be known to him before he approaches the system.

References

[1] H. A. Hall, N. Weiderman, The Evaluation Problem in Relevance Feedback Systems, Report No. ISR-12 to the National Science Foundation, Section XII, Department of Computer Science, Cornell University, June 1967.

[2] E. Ide, User Interaction with an Automated Retrieval System, Report No. ISR-12 to the National Science Foundation, Section XII, Department of Computer Science, Cornell University, June 1967.

[3] J. Kelly, Negative Response Relevance Feedback, Report No. ISR-12 to the National Science Foundation, Section IX, Department of Computer Science, Cornell University, June 1967.

[4] W. Riddle, T. Horwitz, R. Dietz, Relevance Feedback in an Information REtrieval System, Report No. ISR-11 to the National Science Foundation, Section VI, Department of Computer Science, Cornell University, June 1966.

[5] G. Salton, Automatic Information Retrieval, Class Notes, Computer Science 435, Cornell University.

[6] G. Salton and M. E. Lesk, Computer Evaluation of Indexing and Text Processing, Report No. ISR-12 to the National Science Foundation, Section III, Department of Computer Science, Cornell University, June 1967.

[7] G. Salton and M. E. Lesk, The SMART Automatic Document Retrieval System — An Illustration, Communications of the ACM, Vol. VIII, No. 6, June 1964.

[8] A. Borodin, L. Kerr, and F. Lewis, Query Splitting in Relevance Feedback Systems, Project Report Computer Science 435, Cornell University, May 1968.