AUTOMATIC PROCESSING OF CURRENT AFFAIRS QUERIES

G. Salton

TR 72-139

August 1972

Department of Computer Science
Cornell University
Ithaca, New York 14850

Automatic Processing of Current Affairs Queries

G. Salton+

Abstract:

The SMART system is used for the analysis, search, and retrieval of news stories appearing in Time magazine. A comparison is made between the automatic text processing methods incorporated into the SMART system and a manual search using the classified index to Time. The results indicate that equivalent retrieval results are obtainable when both the manual and the automatic searches are carried out in a feedback mode.

1. Introduction

It has been apparent for some time that simple, fully automatic text processing methods incorporated into a search and retrieval system produce retrieval results equivalent in effectiveness to those obtainable conventionally by manual indexing and search operations carried out under controlled conditions. [1,2,3] The results published to-date have, however, been obtained in each case with reasonably homogeneous collections covering certain specified, technical fields. Non-technical material of the type normally found in news stories has not so far been used for test purposes. Moreover, the queries processed until

---

+ Cornell University, Dept. of Computer Science, Ithaca, New York 14850

This study was supported in part by the National Science Foundation under grant GJ-314 and in part by the National Library of Medicine (NIH) under grant LM 00704.

now have generally been reference-retrieval oriented, in the sense that they covered reasonably broad areas of interest that might be answered by various documents in the stored collections.

The present study differs from the earlier ones in three important respects: the topic area is "world affairs" — a subject normally believed to involve a "soft" vocabulary; many of the queries are of the fact-retrieval type, being directed at a specific news story included in the stored collection; the documents, consisting of the full text of articles published during 1963 in Time magazine, are in general short, narrowly focused, and with a definite point of view.

A manual search process based on a classified index covering the collection may be particularly effective in this environment because each article is often easy to classify unambiguously under the few, applicable subject headings normally provided by a classified index. The manual search process may then be expected to produce high-precision results, including very few so-called "false drops", and the extra indexing depth (greater number of terms) and greater exhaustivity of indexing (more detail of subject coverage) normally obtainable by an automatic text analysis procedure may be of no importance.

The automatic process may also be at a disadvantage because of the narrow, factual queries that must be processed in the environment of news articles. Indeed, the lack of know-how regarding the syntactic and semantic properties of natural languages does not generally make it possible automatically to furnish direct answers to incoming queries.

On the other hand, an automatic analysis process which does not rely on the vagaries of human indexing may also exhibit some potential advantages. Some articles will always be misclassified in a manual index, either because the indexer misses the main point of the article, or because the indexing language does not provide the terms most appropriate for a given item. Second, it is sometimes difficult with a classified index to perform high-recall searches, that is, to retrieve a large proportion of the articles relevant to a given topic; in that case, the automatic indexing procedure usually affords more entry points, leading to a higher recall. Finally, a classified index generally contains only short abstracts of the referenced articles; the searcher must then engage in a guessing game in order to establish the relevance of a given index entry to a search request. This may lead both to recall failures where relevant articles are eventually not retrieved, and to precision failures where nonrelevant articles are produced by mistake.

In the remainder of this report, some earlier retrieval studies involving published indexes are reviewed, and an effort is made to assess the usefulness of automatic indexing methods in the search of news articles.

2. Subject Heading Indexes

Some typical excerpts from an issue of the Time Index are reproduced in Fig. 1. [4] It may be seen that this particular published index contains terms denoting people, organizations, countries, or activities. Many entries are listed with a "see reference" (for example, "Tunnels see Railroads") to indicate that the relevant items will be found elsewhere. Some "see also references" are used (for example, "United States see also John F. Kennedy"), but this feature is used sparingly in the Time Index to prevent excessively complicated manual searches. Finally, some index entries form a hierarchical classification (for example, "House of Representatives" is hierarchically inferior to "Congress", which is in turn inferior to "United States.") In the Time Index, very little classification of this type is provided.

It should be noted that for Time, the index is directly oriented to the collection which it covers. This is seen from the cross-references which often relate terms that appear not to be parallel (for example: "Unesco <u>see</u> Dams", "Ulm, Germany <u>see</u> Theater", or "Udall, Stewart <u>see</u> Mountain Climbing"); the same feature is also apparent from the abstracts identifying the items, which reflect in detail the point of view of the corresponding articles.

Various studies have been made of the effectiveness of classified indexes for information search and retrieval [5,6,7,8,9]. In general, an attempt is made to compare the usefulness of a manual search using a classified index with some more automatic procedure — for example, an automatic extraction of keywords from document titles, or a listing of certain text words in the surrounding context (KWIC index). The results obtainable from the various studies are not directly comparable, because of the variety of search tools actually used and the differing test conditions. Still, one invariably finds that for a large proportion of documents, ranging from 50 to 90 percent, a correlation (synonymity or other relationship) exists between the manually assigned subject headings, and words automatically extracted from document titles or texts. [2] While this type of evidence does not directly lead to usable retrieval procedures, it does suggest that automatic extraction and analysis procedures may prove to be competitive in retrieval effectiveness with specialized classified indexes.

3. "Time" Test Design

The principal collection and query characteristics are outlined in Table 1. The experimental collection consists of 425 articles published during 1963 in "The World" section of Time magazine. This comprises all articles included in that section from 42 out of the 52 issues of Time appearing in 1963. A set of 83 search requests in current events and international affairs was obtained from a variety of sources all appearing in 1963. Specifically, 44 queries were taken from section 4 of the Sunday New York Times (News of the Week in Review); an additional 27 queries originated in the "Current Affairs Test for Colleges" published by the Book and Educational Program of the New York Times; and the last 12 queries were from the

"Time Current Affairs Test" issued by the Time Educational Program, and from the "Ask Yourself" column in "Senior Scholastic".

Most of the questions — a total of 65 out of the 83 are highly specific "fact retrieval" type questions which might reasonably be answered by one or at most a few of the Time articles. A typical sample is included in the top half of Table 2. Another 18 queries are of a more general nature and would likely be covered by a greater variety of Time articles. Some examples are reproduced in the lower half of Table 2.

The searches of the published Time Index were conducted by a trained college student. A feedback mode was used in the sense that each query first gave rise to a number of primary index terms, and these in turn led to additional terms through the cross-references, and through entries appearing in the index under the primary terms. On the average, seven subject terms were obtained from the index for each query, and over 6,000 article entries were examined in response to the 83 queries; however, after reading the corresponding published abstracts, only 229 articles were eventually retrieved as potentially relevant to the queries.

An additional 11 potentially relevant articles were obtained through searches conducted internally by personnel of Time Incorporated, using a manual index card file, more complete than the published index but not accessible to the general public.

The relevance assessments were made independently of the searches and without utilization of the Time index following an exhaustive comparison of the full texts of all articles in the collection with the corresponding Time queries. A total of 324 articles were thus judged to be relevant to the 83 queries.

The automatic searches were performed by the SMART system using as input the full document and query texts. [10,11] Specifically, the following automatic procedures were utilized to obtain for each document and search request a set of concept vectors, consisting of weighted content identifiers:

a) The word form process produces terms representing words extracted from the original text, and weighted according to the frequency of occurrence in the documents or query texts; final "s" endings are cut off before incorporation into the concept vectors, so that words like "apple" and "apples" are reduced to a unique term.

b) A thesaurus, or synonym dictionary may be used to group related or synonymous words into word classes; each word is then replaced by the corresponding thesaurus class identifier, or concept number, before incorporation into the concept vectors; the concept numbers are weighted as before. For the Time study, a manually constructed thesaurus covering the world affairs area was utilized.

c) A word discrimination dictionary can be constructed fully automatically, in which the words occurring in a given document collection (or alternatively the thesaurus categories) are ranked in decreasing order of their effectiveness as document discriminators [12,13]; a word with a high discrimination value, which is useful in distinguishing one document from another, will then be preferred over one with a low discrimination value, the latter consisting mostly of common words or high frequency words included in most of the documents. The discrimination value attached to each word, or to each thesaurus category, can be used in addition to the normal frequency-oriented weight — for example by multiplying the standard term weight by the discrimination value for the corresponding term. This gives rise, respectively, to the <u>word form</u> * <u>discrimination</u> and the <u>thesaurus</u> * <u>discrimination</u> processes.

d) A user-controlled feedback process, known as <u>relevance feedback</u> is incorporated into the SMART system, which produces improved query formulations automatically, based on information furnished by the user to the system as a result of previous search operations. Specifically, when a previously retrieved document is designated as relevant to the user's information needs, a feedback query is constructed which is more similar to this retrieved item; the reverse is true when previously retrieved items are designated as nonrelevant. For the Time study, first and second iteration feedback queries are constructed for most computer runs.

In the next section a comparison is made between the retrieval performance of the Time Index and the automatic text processing capabilities incorporated into SMART. The various SMART procedures are also ranked in decreasing order of their retrieval effectiveness.

4. Retrieval Results

A) The Time Index

The performance of the Time Index is reflected in the output of Table 3, averaged over 83 queries and 425 documents. In each case, both recall and precision values are computed for each query. Specifically,

$$\text{Recall} = \frac{\text{Number of documents retrieved and relevant}}{\text{Total number of relevant documents in collection}} = \frac{a}{b}$$

$$\text{Precision} = \frac{\text{Number of documents retrieved and relevant}}{\text{Total number of documents retrieved}} = \frac{a}{c}$$

Two types of averages are generated to reflect the performance for the 83 queries. The per-query average is simply the arithmetic mean of the individual recall and precision performances averaged over all the queries. For k queries, the following formulas are used:

$$\text{Average (per-query) Recall} = \frac{1}{k} \sum_{i=1}^{k} \frac{a_i}{b_i}$$

$$\text{Average (per-query) Precision} = \frac{1}{k} \sum_{i=1}^{k} \frac{a_i}{c_i}$$

For the document-level averages, a single composite query is constructed from the 83 queries, whose relevant documents comprise the sum of the relevant for the 83 individual queries; an attempt is then made to retrieve as many relevant as possible for that composite query, regardless of the individual query performances:

$$\text{Average (document-level) Recall} = \frac{\sum_{i=1}^{k} a_i}{\sum_{i=1}^{k} b_i}$$

$$\text{Average (document-level) Precision} = \frac{\sum_{i=1}^{k} a_i}{\sum_{i=1}^{k} c_i}$$

To permit a comparison of the performances of the Time Index and the SMART runs, the SMART ranking feature, which produces ranked document output in decreasing order of the query-document similarity coefficients is used in such a way that for each query, the number of documents retrieved by SMART is exactly the same as that obtained with the Time Index (a total of 240 documents).

Consider first the performance of the Time Index itself. The average recall per query of 66 percent, and the precision of 77 percent, far exceed the performance to be expected from controlled indexing systems in operational retrieval environments. (The same is true of the overall document-level averages which produce somewhat lower recall but also somewhat higher precision). The Medlars system operating with medical documents at the National Library of Medicine was found in contrast to achieve only about 58% recall and 50% precision. [14] The much higher precision obtained in the present case with the Time Index is due unquestionably to the large number of low generality fact-retrieval questions for which the answers are contained in only a few documents. In these cases, a single subject term is normally sufficient to retrieve all relevant documents, and the Time Index more often than not provides a perfect performance. Examples are query 23 of Table 2 ("What countries have newly joined the United Nations?"), or query 51 ("Who is the successor to Premier Krushchev?"), where the terms "United Nations" and "Krushchev", respectively, retrieve all relevant items.

The Time Index does not, however, produce perfect results. Most of the precision failures, accounting for the 23 percent of nonrelevant items retrieved, are due to ambiguities or lack of specificity in the abstracts printed in the index. In these cases, the searcher made the wrong guesses based on the published abstracts, and produced documents which eventually turned out to be nonrelevant.

The recall failures, on the other hand, are generally due to mistakes in the classification of items in the Time Index. Three examples of this type are shown in Table 4 for queries 33, 46, and 9 respectively. The search terms correspond exactly to the query texts, and the text excerpts reproduced from the documents show that each document is indeed relevant to the corresponding query. Unfortunately, the indexer classified a document about the internal unrest in Belgium under "weather", and one dealing with de Gaulle's attitude toward the Common Market under "Ireland". Obviously, such relevant documents will not be retrievable with a normal search strategy, thereby causing the recall deficiency.

B) The SMART Runs

When the automatic indexing procedures used by SMART are compared with the manual use of the Time Index, it is generally the case that, for the same number of retrieved documents, SMART outperforms the Time Index for the broader reference-retrieval questions, whereas Time outperforms SMART for many of the narrow fact retrieval queries. This is illustrated in Table 2 for some fact-retrieval queries (top) and reference retrieval queries (bottom). Since 65 of the queries, or almost 80 percent, fall into the fact-retrieval category — each one having 5 or fewer relevant documents in the collection — it is not surprising that the overall average performance is somewhat lower for SMART than for the Time Index.

The figures of Table 3 indicate, however, that after two applications of the relevance feedback process, the SMART averages are within ten to fifteen percent of the Time output. For the underlined SMART averages, the statistical significance tests (sign test and Wilcoxon signed rank test) indicate that the performance differences between SMART and the Time Index are <u>not</u> statistically significant — that is, the probability that the two sets of performance figures might have been obtained from the same distribution is higher than five percent. This is the case notably for the word form discrimination process which involves no thesaurus or other word normalization tool, and is based strictly on word extraction and automatic term discrimination values. Thus, when the same number of documents is retrieved by SMART as by the Time

Index, the automatic feedback runs produce an output — even for the world affairs fact-retrieval environment — which is only slightly inferior to the controlled indexing used at Time; in each case, the performance differences are not statistically significant.

The restriction on the number of documents retrieved per run limits the effectiveness of the SMART feedback procedures. Since the Time Index retrieves a total of 240 documents, an average of only 3 previously retrieved documents can be used to construct each feedback query ($3 \times 83 = 249$). When this limitation is removed, and 5 retrieved documents are used to generate each feedback query in the automatic system (for a total of 415 documents overall), the output of Table 5 is produced.

It may be seen that on a query averaged basis, the SMART results are now only slightly below the corresponding Time figures, following the second feedback operation. When document-level averaging is used, the SMART results are actually somewhat superior. None of the differences between the two systems are statistically significant.

The various SMART runs are compared with each other in the recall-precision graphs of Fig. 2. These graphs are produced by plotting recall and precision values for various retrieval levels (that is, following the retrieval of 5, 10, 15 and so on, documents). The curve closest to the upper right-hand corner where both recall and precision are maximized represents the best performance in each case. It may be seen from the output of Fig. 2, that the relevance feedback method produces significant improvements in retrieval effectiveness. In every case, the second iteration feedback curves are at least 20 percent closer to the ideal performance region than the initial search curves, and all performance improvements are statistically significant at the 0.01 probability level (that is, the probability that the two sets of runs might in fact represent an identical performance is less than 0.01).

A comparison of three SMART procedures is shown in Fig. 3 for initial runs (Fig. 3(a)) and second iteration feedback runs (Fig. 3(b)). It may be seen that the three SMART processes produce performance figures which are fairly close, although the three feedback runs are superior to the corresponding initial runs. The best results are obtained with the word form * discrimination procedure which is, of course, fully automatic, and can be used without elaborate preparation or extraneous user inputs.

The statistical significance figures for the output of Fig. 3 is shown in Table 6. When the word form * discrimination method is compared with the standard thesaurus output, the former is preferable because the probabilities are very small that the two sets of results are indistinguishable. When the discrimination values are used in addition to either word form or thesaurus processes no differences in in performance are noticeable, the probabilities this time being close to 1 that the two runs are in fact identical. Obviously, the fully automatic word from extraction process supplemented by the automatic computation of word discrimination values is the preferred SMART process for the retrieval of current events news questions.

5. Conclusion

A comparison was made between the performance of a conventional, classified index and automatic word extraction and weighting procedures used for the retrieval of current affairs information. The automatic system provides a superior performance for the broad reference-retrieval type queries, and is least successful with the narrow fact-retrieval queries. When the automatic text processing methods are used in a feedback mode, whereby improved queries are automatically formulated based on previously retrieved documents, the average performance differences between manual and automatic systems are not statistically significant.

The automatic system thus appears to be competitive, even for non-technical subject areas, and narrow, fact-retrieval type material.

Acknowledgement

Several people from the Information Processing Department of Time Incorporated were instrumental in arranging for, or in carrying out, the searches using the private card index file, notably Mrs. Margaret T. Fischer (now at Xerox Education Group), Mrs. Margaret N. Kastner, and Mr. Peter Draz. Dr. Michael E. Lesk of Bell Laboratories and Mr. E.M. Keen of the School of Librarianship in Aberystwyth, Wales helped with the experimental text desing. Miss Wendy Bryce and

Mrs. Angie Rettig of the SMART staff carried out the searches in the published Time Index, and obtained the relevance assessments, respectively. Finally, the SMART computer runs were arranged by Mrs. Barbara Galaska and Mr. Joel Zumoff.

The writer gratefully acknowledges the help and assistance of all these individuals, and especially the cooperation of the Time personnel.

References

[1] G. Salton and M.E. Lesk, Computer Evaluation of Indexing and Text Processing, Journal of the ACM, Vol. 15, No. 1, January 1968, p. 8-36.

[2] G. Salton, Automatic Text Analysis, Science, Vol. 168, 17 April 1970, p. 335-343.

[3] G. Salton, A New Comparison between Conventional Indexing (Medlars) and Automatic Text Processing (SMART), Journal of the ASIS, Vol. 23, No. 2, March-April 1972, p. 75-84.

[4] Time Inc., Index to Volume LXXXI-LXXXII of Time Magazine, New York, 1963.

[5] J.A. Virgo, An Evaluation of Index Medicus and Medlars in the Field of Ophthalmology, Journal of the ASIS, Vol. 21, No. 4, July-August 1970, p. 254-263.

[6] N.E. Brodie, Evaluation of a KWIC Index for Library Literature, Journal of the ASIS, Vol. 21, No. 1, January-February 1970, p. 22-28.

[7] C. Montgomery and D.R. Swanson, Machine-like Indexing by People, American Documentation Vol. 13, 1962, p. 359-366.

[8] T.N. Shaw and H. Rothman, An Experiment in Indexing by Word Choosing, Journal of Documentation, Vol. 24, No. 3, Sept. 1968, p. 159-172.

[9] M.E. Stevens, Automatic Indexing: A State-of-the-Art Report, NBS Monograph 91, National Bureau of Standards, March 1965.

[10] G. Salton, Automatic Information Organization and Retrieval, McGraw Hill Book Col, New York 1968.

[11] G. Salton, The SMART Retrieval System — Experiments in Automatic Document Processing, Prentice Hall Inc., Englewood Cliffs, N.J., 1971

[12] K. Bonwit and J. Aste Tonsman, Negative Dictionaries, Report No. ISR-18, Section 6, Dept. of Computer Science, Cornell University, Ithaca, New York, October 1970.

[13] G. Salton, Experiments in Automatic Thesaurus Construction for Information Retrieval, Proc. IFIP Congress-71, North Holland Publishing Company, Amsterdam, 1972.

[14] F.W. Lancaster, Evaluation of the Medlars Demand Search Service, National Library of Medicine, Bethesda, Maryland, January 1968.

Query and Collection Characteristics	Comments
Document Collection	425 articles published in "The World" section of Time magazine in 1963 (42 issues of Time complete)
Search Requests	83 queries taken from 1963 issues N.Y. Times "News of the Week in Review", N.Y. Times "Current Affairs Test for Colleges", Time "Current Affairs Test", and Senior Scholastic "Ask Yourself" column.
Relevance Judgments	324 relevant articles obtained by exhaustive comparison of all document texts with each search request
Time Index Search	229 documents retrieved by using the published "Time Index" in a feedback mode;
	11 additional documents retrieved by using private Time card index

Collection and Query Characteristics

Table 1

Query Number	Query Text	Number Relevant	Time R	Time P	SMART R	SMART P
	<u>Narrow Fact Retrieval Type</u>					
3	Number of troops the U.S. has stationed in Vietnam compared with the number of U.S. troops in Germany.	4	.25	1.0	0	0
23	What countries have newly joined the United Nations?	1	1.0	1.0	0	0
48	What is the "hot line" proposal?	1	0	0	0	0
51	Who is the successor to Premier Krushchev?	3	1.0	1.0	.33	1.0
	<u>Broad Reference Retrieval Type</u>					
46	President de Gaulle's policy on Britain's entry into the Common Market.	18	.39	.87	.72	.65
49	Background on the new Chancellor of Germany, Ludwig Erhard.	8	.62	1.0	.50	1.0
63	President Nasser's ruling out Arab Union so long as present governing party in Syria remains in control.	11	.54	.75	.91	.62

Average Performance	Time Index		SMART Word Form	
	Recall	Precision	Recall	Precision
All Queries	.6600	.7726	.4170	.4851
18 General Queries	.5510 (-17%)	.8489 (+10%)	.4323 (+4%)	.6305 (+30
65 Specific Queries	.6924 (+ 5%)	.7500 (- 3%)	.4141 (-1%)	.4419 (- 9

Specific and General "Time" Queries

Table 2

Procedures	Recall		Precision	
	Initial Run	Second Feed-back Run	Initial Run	Second Fee back Run
Time Index	.6600	-	.7726	-
SMART Word Form	.4170 (-37%)	.5544 (-16%)	.4851 (-37%)	.6251 (-19
Thesaurus	.4215 (-36%)	.5746 (-13%)	.5040 (-35%)	.6610 (-14
WF.*DISC.	.4322 (-35%)	.5778 (-12%)	.5126 (-34%)	.6675 (-14
Thes.*DISC.	.4225 (-36%)	.5771 (-13%)	.5056 (-35%)	.6801 (-12

a) Per-Query Recall and Precision Averages

Procedures	Recall		Precision	
	Initial Run	Second Feed-back Run	Initial Run	Second Fee back Run
Time Index	.5864	-	.7916	-
SMART Word Form	.4197 (-28%)	.5246 (-11%)	.5666 (-28%)	.7083 (-11
Thesaurus	.4320 (-26%)	.5617 (- 4%)	.5833 (-26%)	.7583 (- 4
WF.*DISC.	.4444 (-24%)	.5493 (- 6%)	.6000 (-24%)	.7416 (- 6
Thes.*DISC.	.4444 (-24%)	.5648 (- 4%)	.6000 (-24%)	.7625 (- 4

b) Document-Level Recall and Precision Averages

SMART-Time Comparison
(averages for 425 queries, 83 documents; feedback
based on 3 retrieved documents per query)

Table 3

Query No.	Query Text	Search Terms	Indexed Under	Sentences from Descriptions in Index	Excerpt from Document Text
33	Government crisis produced by the controversy involving the Walloons and the Flemings	Walloons Flemings Belgium	Weather	Summer in Britain, France, Germany, Italy, Low countries yielded such irritants as heavy winds, rain, cool July, August snow, high prices, and blighted harvests.	.... but of greater concern to the Belgians than the meager harvest or the tempestuous weather was a new law that goes into effect this week, creating a formal language barrier across the land .....
46	De Gaulle's policy on British entry into the Common Market	De Gaulle France Com. Market Trade Britain	Ireland	Ireland's new mood of confidence is largely due to bold industrialization program of Prime Minister Sean Lemass thrusting the nation belatedly into the industrial revolution	.... seven of the thirteen members of Lemass' Cabinet are under 60 ... a relatively green age in Irish politics...Lemass' most bruising disappointment in office was Charles de Gaulle's rejection of British membership in the Common Market.
9	Opposition of Indonesia to the newly created Malaysia	Indonesia Malaysia Great Britain	Nuclear	Refusing to sign the test ban treaty, Red China charges Russia with surrender to U.S. imperialism. Peking's antitreaty stand is backed by Albania, North Korea and North Vietnam. Cuba remains silent on the treaty...France is the the only non-signer in the Western camp...	....but wary of angering the Chinese, other Asian nations have been slow to indicate their approval of the (nuclear test ban) pact. They include Nepal...., Ceylon and Cambodia,...and Indonesia, which is hopeful of support in any future action against the soon-to-be-born Malaysian federation.

Procedures	Recall		Precision	
	Initial Run	Second Feed-back Run	Initial Run	Second Feed-back Run
I Per Query Averages				
Time Index	.6600		.7726	
SMART Word Form *Discrimination	.4322 (-35%)	.6289 (-5%)	.5126 (-34%)	.7340 (-5%)
II Document-Level Averages				
Time Index	.5864		.7916	
SMART Word Form *Discrimination	.4444 (-24%)	.5895 (+1%)	.6000 (-24%)	.7958 (+1%)

SMART-Time Comparison

(425 documents, 83 queries; feedback based on 5 documents retrieved per query)

Table 5

SMART Procedure (testing for run A equivalent to B)	Probability Value	
	T-Test	Wilcoxon Signed Rank Test
A. Word Form * Discrimination		
B. Thesaurus Run		
initial run	.0077	.0000
second feedback run	.0776	.0000
A. Thesaurus * Discrimination		
B. Word Form * Discrimination		
initial run	.9368	.9975
second feedback run	.9827	.9953

Statistical Significance Testing for SMART Runs

Table 6

Typical Entries in Printed "Time" Index

(including "see" and "see also" references and subheadings)

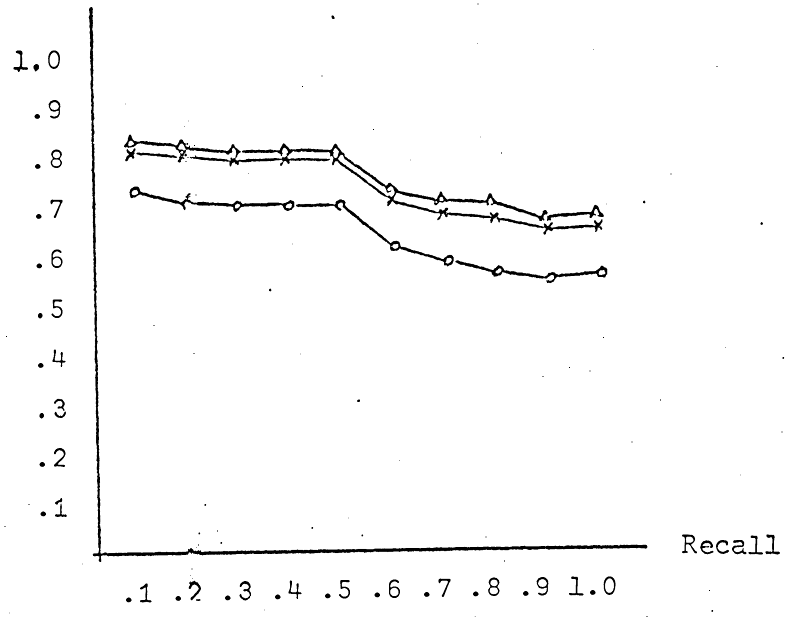
Fig. 1

o—o INITIAL SEARCH - WORDFORM

x—x 1st ITERATION FEEDBACK (TOP 3)

△—△ 2nd ITERATION FEEDBACK (TOP 3)

Re-call	Precision o—o	x—x	△—△
0.1	.7025	.7942	.8073
0.3	.6763	.7585	.7759
0.5	.6475	.7224	.7439
0.7	.5037	.6026	.6236
0.9	.4584	.5511	.5790



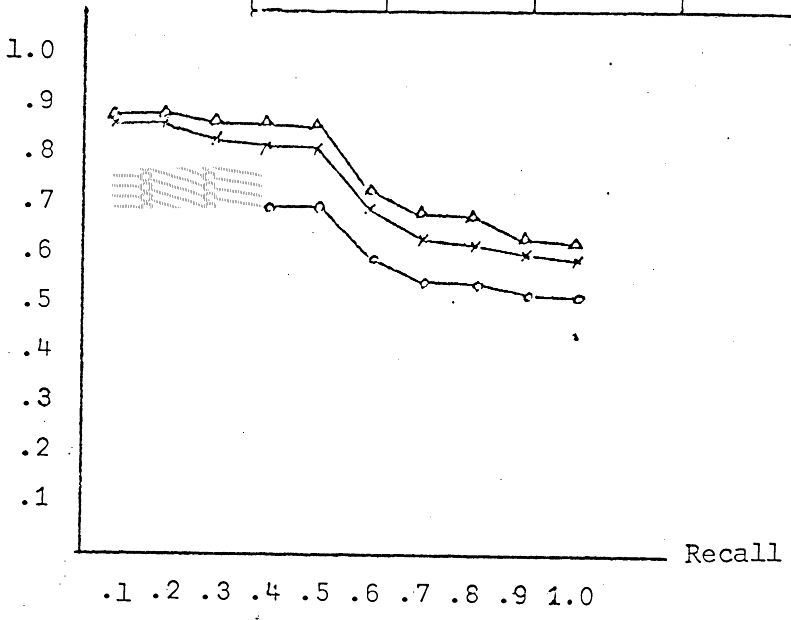
a) Word Form Process

o—o INITIAL SEARCH - THESAURUS

x—x 1st ITERATION FEEDBACK (TOP 3)

△—△ 2nd ITERATION FEEDBACK (TOP 3)

Re-call	Precision o—o	x—x	△—△
0.1	.7292	.8449	.8546
0.3	.6906	.8053	.8311
0.5	.6553	.7797	.8119
0.7	.4971	.5816	.6214
0.9	.4479	.5258	.5696



b) Thesaurus Process

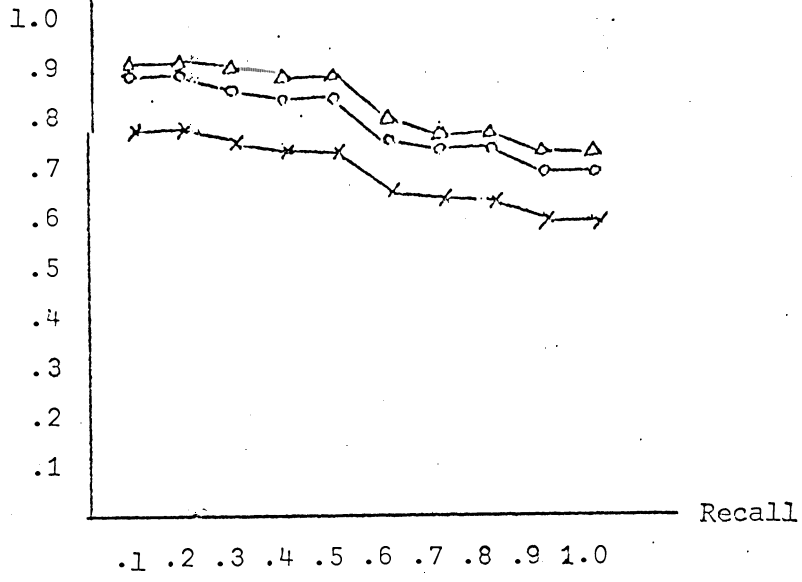
SMART Performance

INITIAL SEARCH (WORDFORM * DISCRIMINATION VALUE)

1st ITERATION FEEDBACK

2nd ITERATION FEEDBACK

Re-call	Precision (initial search)	Precision (1st iteration)	Precision (2nd iteration)
0.1	.7221	.8285	.8422
0.3	.6960	.7947	.8191
0.5	.6605	.7652	.7971
0.7	.5314	.6218	.6532
0.9	.4844	.5803	.6105



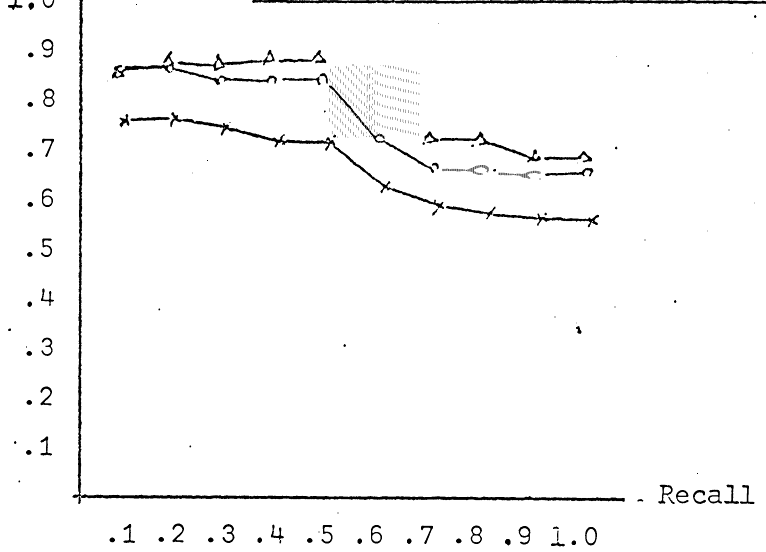
c) Word form * Discrimination Value

INITIAL SEARCH (THESAURUS * DISCRIMINATION VALUE)

1st ITERATION FEEDBACK

2nd ITERATION FEEDBACK

Re-call	Precision (initial search)	Precision (1st iteration)	Precision (2nd iteration)
0.1	.7313	.8411	.8393
0.3	.7038	.8069	.8258
0.5	.6670	.7855	.8125
0.7	.5015	.5861	.6356
0.9	.4464	.5346	.5838



d) Thesaurus * Discrimination Value

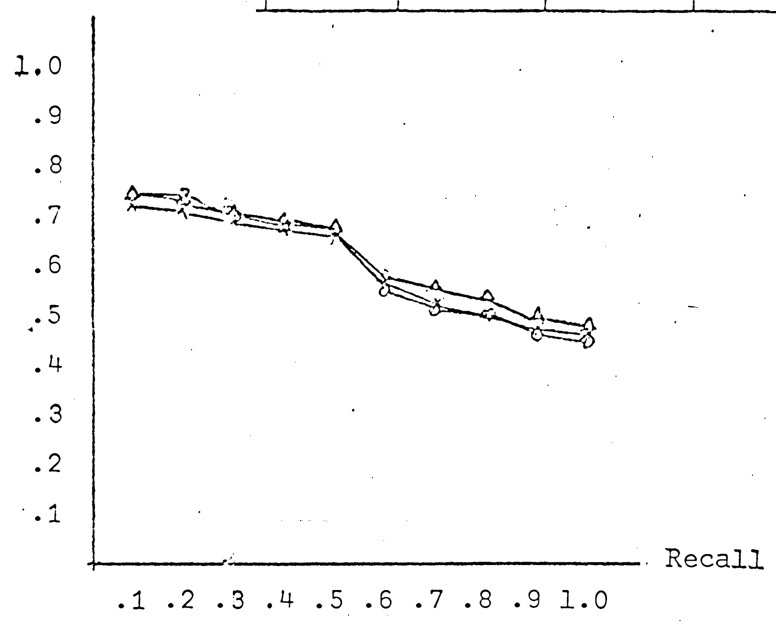
SMART Performance

WORD FORM - INITIAL SEARCH

THESAURUS - INITIAL SEARCH

WORDFORM * DISCRIMINATION VALUE-
INITIAL SEARCH

Recall	Precision x——x	Precision o——o	Precision △——△
0.1	.7025	.7292	.7221
0.3	.6763	.6906	.6960
0.5	.6475	.6553	.6605
0.7	.5037	.4971	.5314
0.9	.4584	.4479	.4844



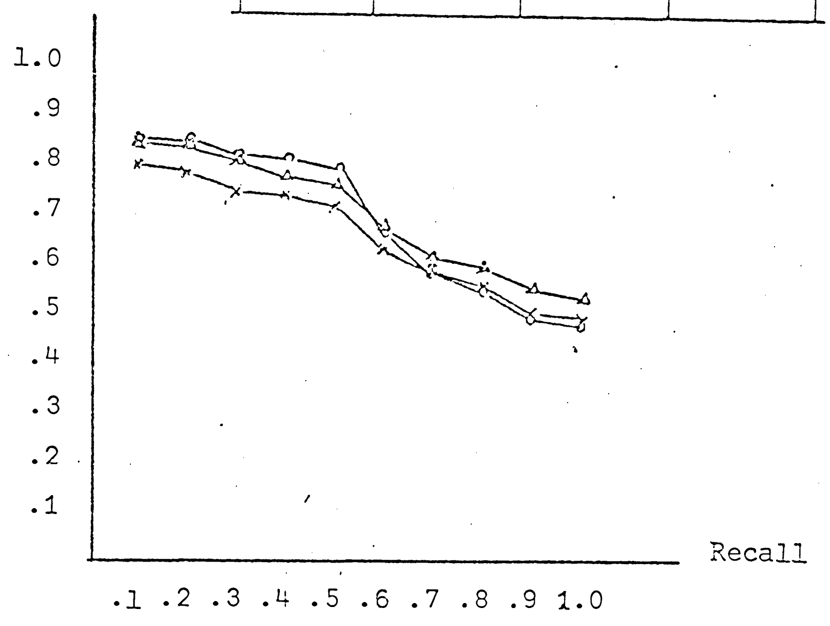
a) Initial Runs

WORDFORM - 2nd ITERATION FEEDBACK

THESAURUS - 2nd ITERATION FEEDBACK

WORDFORM * DISCRIMINATION VALUE -
2nd ITERATION FEEDBACK

Recall	Precision x——x	Precision o——o	Precision △——△
0.1	.8073	.8546	.8422
0.3	.7759	.8311	.8191
0.5	.7439	.8119	.7971
0.7	.6236	.6214	.6532
0.9	.5790	.5696	.6105



b) Second Iteration Feedback

Comparison of SMART Runs