III. Search Matching Functions

E. M. Keen

1. Introduction

It is the function of a document retrieval system to draw to a requestor's attention some or all of the documents in the store that have some probability of being relevant to the requestor's need. The processing of a search request is in part a matching operation, and procedures used in manual and mechanized systems are briefly described before automated systems are discussed. SMART test results are given, separating the results of the matching functions themselves from the weighting scheme that may be employed. The analyses show that one particular matching function used with the weighting scheme is superior to the others that have been tested. A suggested experiment is proposed to determine whether development of still better functions is possible.

2. Matching Procedures used in Manual, Mechanized and Automated Systems

A) Manual Systems

The task of matching a search request against a stored file of document representations is frequently an entirely manual process. After the search request has been received and "translated" into the vocabulary of the system, either words or code numbers, a searcher then seeks documents that match the coded request by referring to the index file, that is, the cards or the pages constituting the physical index. In searches made with dictionary or classified catalogs, the searcher normally proceeds by starting in one place initially, and then seeking other likely places through which to extend the search if necessary. The same procedure is followed in using a KWIC (Keyword-In-Context) index, which although mechanically produced is usually manually searched. An example of a search request and some of the strategies used to perform a search through various types of indexes is given in Figure 1. The main characteristic of these manual systems is that the indexes are designed to be entered by the searcher in one place at a time only. Thus the subject headings and classification numbers used must represent quite complex ideas in a single entry to cope with modern knowledge.

Another type of manually searched index that has gained widespread acceptance is the type that allows entry into several parts of the file simultaneously, and is designed to identify documents that are found in all of the places entered. These systems are known as co-ordinate systems, or better post-co-ordinate, since the documents retrieved are those which match the search terms of the request only if the terms are present in the documents in the required combinations. The processing of search requests in such systems requires not only a decision as to which vocabulary terms shall be used in the search, but also a statement of logical combinations of the terms, in terms of logical products (AND), logical sums (OR), and logical differences (NOT). An example of such a search formulation is given in Figure 2; although this example illustrates a mechanized system to be described, a similar search formulation could be used in a manually searched system.

In these manual systems described, each entry into the index produces a set of documents that match the search formulation, usually called the retrieval set; the remainder of the collection is considered to be not retrieved. User satisfaction is related both to the finding of relevant

<u>Request</u> Documents on cerebrospinal fluid oxygen concentrations.

a) Dictionary Catalogue using Library of Congress Subject Headings:

CEREBROSPINAL FLUID

OXYGEN IN THE BODY

ANOXEMIA

etc.

b) Classified Catalogue using Library of Congress Classifications:

RC 400-406	Diseases of the spinal cord
RB55	Spinal fluid (Examination)
RC104.A4	Deficiency of oxygen in the blood; anoxemia; etc.

c) KWIC Index based on natural language document titles:

CEREBROSPINAL

OXYGEN

OXYGENATION

FLUID

etc.

Hypothetical request showing some of the search terms that could be considered for use in three types of manually searched indexes.

Figure 1.

<u>Request Statement</u>: The relationship of blood and cerebrospinal fluid oxygen concentration or partial pressures. A method of interest is polarography. English [Language Documents] only.

<u>Medlars Search Formulation</u>

CEREBROSPINAL FLUID <u>and</u> ANOXEMIA <u>or</u> HYPOXEMIA <u>or</u> OXIMETRY <u>or</u>

[OXYGEN CONSUMPTION <u>or</u> ANOXIA <u>or</u> OXYGEN] <u>and</u> [BLOOD <u>or</u> BLOOD GAS ANALYSIS]

<u>References Retrieved</u> 11

<u>Postings of Search Terms</u>

Cerebrospinal fluid	1162	Oxygen	1730
Anoxemia	210	Blood	8050
Hypoxemia	129	Blood Gas Analysis	1452
Oximetry	607	(Partial Pressures	91)
Oxygen Consumption	652	(Polarography	471)
Anoxia	1302		

Data from a search request made in the Medlars System.

Figure 2.

documents in the file, and also to the prevention of an exhaustive examination of the file. A manual search may easily be modified as it proceeds to prevent retrieval of more documents than the user is willing to examine, the modification being carried out by controlling the choice of search terms, as well as the logical combinations demanded.

B) Mechanized Systems

The word "mechanized" is here used to describe systems in which certain parts only of the storage, processing, and retrieval stages are mechanized. The present discussion thus concerns systems in which the physical search or matching operation is mechanized, but the search formulation is manually constructed. Since mechanization of document retrieval systems has been based in the past almost invariably on manual systems of the post-co-ordinate type, search formulations for mechanized systems require both a selection of search terms and also a request statement in the form of logical combinations of these search terms. The formulation of one particular search request, together with the requestor's original statement, is given in Figure 2, taken from the Medlars system, in which a file of nearly half a million items is searched by means of a computer.

Systems which mechanize the matching process in this way have solved some of the problems associated with manual systems, but have introduced other problems which are unlikely to be solved by this approach. Three areas of difficulty may be outlined:

1. Both manual and mechanized systems of the type described require manual search formulation, consisting of a choice of terms on which to search, and a statement of the logic that is to isolate a set of retrieved documents. This process takes skill and is time-consuming; good results are not, however, obtained for every search.

2. Although mechanized systems are successful in saving some of the time and effort that is required to maintain the document file,

search time frequently becomes a problem, requiring the batching of search requests for reasons of economy. One specific part of the search time problem reflects on the use of the logical relations in the search formulations, since complex logic may tend to increase search time. This problem is circumvented to some extent in the NASA search system by the use of term weights to get the same results as logical formulations, with reduced machine effort. [1]

3. A third problem relates to the cut-off, which is rarely critical in manual systems where a searcher/system interaction may control the search as it proceeds, but in mechanized systems the total search formulation must be made in advance of the search and no changes are normally possible during the search. This problem is again linked to the logical relations in the formulation, since too few documents will be retrieved if the search logic is made too restrictive and 'tight'; contrariwise, if it is made too 'loose' more documents than the user is willing to examine may be retrieved.

An example of this latter point is given using an actual search made in the Medlars system. The request statement and search formulation are shown in Figure 2, and the search of the total document file of nearly half a million items eventually produced eleven references. The user was asked, before the search was carried out, to indicate the number of journal articles relevant to his request that he considered likely to have been published since January 1964; 6 - 20 was chosen as a likely range for the number of relevant items. It is surprising therefore that after receiving the 11 references retrieved in the search, the requestor did not seem to be satisfied, and the librarian who handled the request reports the requestor as saying, "In view of the small return ... Dr. X would be interested in any articles dealing with oxygen concentration in the cerebrospinal fluid." Clearly the system and the search formulation are not at fault in this case: it is just a case of a change in requestor need, presumably arising from the requestor's examination of some retrieved document. This changed need, in the Medlars system, requires a completely new search of the system.

Figure 2 also gives the postings of the nine Mesh terms used in the search formulation plus two more terms from the request not used in the search formulation. Assuming that the search formulation included all the terms likely to be connected with the requestor's need, it can be seen that, ignoring the logical combinations asked for in the search formulation, 8,000 to 10,000 documents in the file are associated with the search term chosen, or, more simply still, regarding "Cerbrospinal fluid" as an essential notion, 1162 associated items are in the file. The user's second need would almost certainly be satisfied by some of these 1162 items, but it might have been possible to have satisfied both the user's first and second needs in a single search in the following manner: with "cerbrospinal fluid" identified as a key notion, all the 1162 documents posted with that term would be retrieved, and would then be presented to the user in some "ranked" order of probable relevance. A satisfactory result for this process would be obtained if the 11 references found in the Medlars search were to appear at the front of this ranked list (the user identified 10 of these 11 as relevant). If such a system could be provided, the user's second need could have been satisfied merely by an examination of more than the first 11 documents in the search output.

A second example of problems of this type lies at the opposite extreme: a request on the effects of drugs and pesticides on the bone marrow of man and animals resulted in 1,235 retrieved documents. Since the set almost certainly contains too many items, a ranked output would again provide a solution, allowing the requestor to examine only as many documents as he desired. Searches in Medlars are frequently made in three matching strengths in an attempt to meet this latter problem.

C) Automated Systems

These illustrations are given to highlight some of the problems inherent in conventional systems, as an introduction to the "ranking" methods used with the SMART system. The objectives of introducing such methods in a retrieval situation are:

1. To avoid the need for manual search logic formulation;

2. To minimize search time not by using logical relation demands but by asking only for request/document term matches using, if necessary, weights associated with each term;

3. To present to the user an ordered list of documents arranged in decreasing order with the search request correlation (ranked output) so that the cut-off may be a user decision at the output stage.

Fully automated retrieval systems are characterized by the replacement of human intellectual effort where it can be as efficiently performed by a machine, and also on occasion by the provision of a man-system interaction that permits the human's irreplaceable contribution (decisions as to cut-off point, judgment of value of documents examined, etc.) to be entered as a control in the search process. The SMART system is investigating the design and evaluation of such automated systems; the use of algorithms that establish matching coefficients between search requests and documents is a part of such a task.

One question to be answered is whether it is necessary to use any human judgment during the search or matching process. In addition, techniques of "weighting" search terms according to some criterion of importance appear to be worth investigating, since they may easily be incorporated into automated systems. The results from the SMART experiments give insight into matching functions for automated systems, as the following results show.

3. SMART Test Results — Matching Functions

A) Description of Functions

Retrieval runs made on the SMART system have concentrated on the use of two optional matching functions, known as the overlap correlation coefficient and the cosine correlation coefficient. The task of matching search requests with the documents in the file is viewed in SMART as a vector similarity problem. The individual elements used in the document and request vectors are the individual content identifiers usually referred to as concepts or concept numbers. For tests comparing the two matching functions, binary vectors are used, in which concepts are either present or absent from a vector; if present, all exert equal weight in the functions.

Since search requests and documents are considered to be simply strings of concept numbers, with no logical relations of the type used in manually formulated searches linking the concepts, only three primary types of data may be incorporated in the matching function:

a) the number of concepts in the request;

b) the number of concepts in the document;

c) the number of concepts that are found both in the request and in the document, i.e. the matching concepts.

The number of matching concepts is used in matching functions of all types, with cosine using both the request and document concepts, and overlap either the request or document concepts, whichever has the smaller total number.

As an example illustrating the two functions, a document vector (b) represented by 18 concept numbers is to be matched against a request vector (a) represented by 8 concept numbers, where 5 concept numbers match:

$$\text{OVERLAP} = \frac{\sum \text{Min}(a\ ,\ b)}{\text{Min}(\sum a\ ,\ \sum b)} \tag{1}$$

$$= \frac{5}{\text{Min}(8\ ,\ 18)} = \frac{5}{8} = 0.63$$

$$\text{COSINE} = \frac{\sum a \cdot b}{\sqrt{\sum a^2 \cdot \sum b^2}} \tag{2}$$

$$= \frac{5}{\sqrt{8 \times 18}} = \frac{5}{12} = 0.42$$

Both functions are designed for use with weighted concept numbers, and their use in this manner is illustrated in part 4. Since in the tests carried out, the requests are generally shorter than the documents (except occasionally when title runs are being made), the overlap function in documentary terms measures the inclusion of the request terms in the document only. Thus, if a request with eight concepts matches five of them in several documents, all such documents will receive identical correlations with the request. The cosine function measures the similarity of the total request to the total document, and non-matching concepts in both requests and documents affect the final correlation. Thus, for a request that matches five out of eight concepts in several documents, the document that has the fewest number of non-matching concepts will receive the highest correlation. Cosine thus takes into account document length, following the principle that if two documents have equal request/document matching concepts, the shorter document has a higher probability of being useful to the requestor, since it will contain less extraneous material. In documentary terms this principle seems of doubtful validity since a requestor may be equally satisfied by treatment of the requested topic in a long document as in a short one.

The overlap correlation provides generally higher correlation coefficients than cosine, but this is of no direct importance, since the correlations are used only to order the documents into a ranked list in relation to each search request so that the positions taken up by the relevant documents may be determined. The correlation values could be displayed for the user to permit him to examine only those documents above a certain correlation; however, since a ranked output is provided, it seems more likely that users will examine the highest ranked documents anyway and continue to look at the ranked list until they are satisfied, or until they are unwilling to examine additional documents on the basis of the document titles or abstracts.

B) Retrieval Performance Results

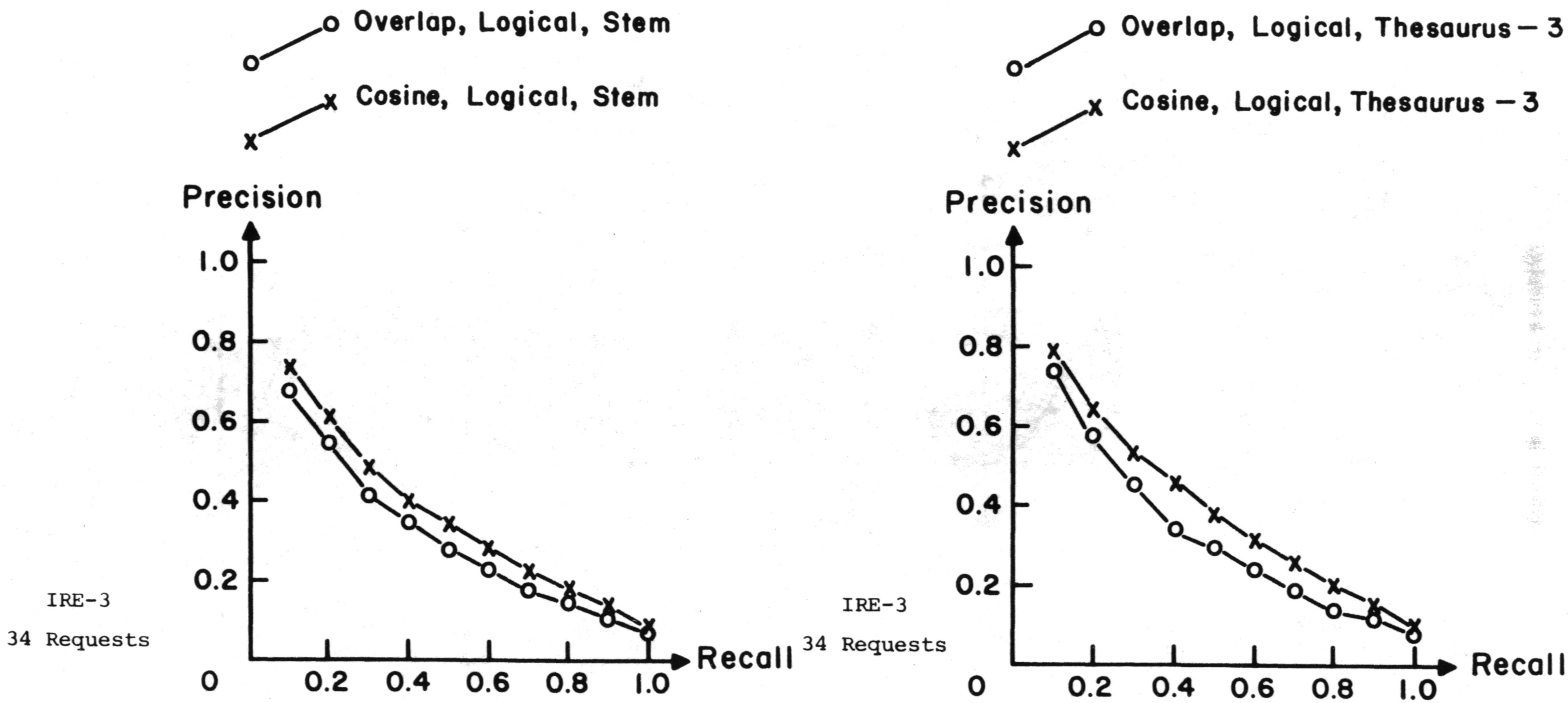
Retrieval runs are made on SMART comparing the overlap and cosine correlation coefficients, without weights (i.e. logical vectors), and keeping other variables such as document length and dictionary type constant. Twelve comparison runs on three collections are presented in Figure 3, and evaluated by normalized recall and normalized precision. In every case, the run with the cosine correlation gives a higher normalized recall and precision than the run with the overlap correlation. The ADI text thesaurus run shows an .043 increase in normalized recall, and the Cran-1 Abstract Stem run shows a normalized precision increase of .055, both increases in favor of cosine.

Figures 4, 5, 6, and 7 present precision versus recall graphs for the stem and thesaurus dictionaries on the IRE-3 collection (Figure 4), the Cran-1 collection (Figure 5), and the ADI collection using text (Figure 6) and abstracts (Figure 7). General merit still favors cosine, although the ADI results show very small differences in the curves, and overlap is superior to cosine in the low recall-high precision area on the stem dictionary, text, and abstract runs.

COLLECTION	INPUT AND DICTIONARY	EVALUATION MEASURE	OVERLAP CORRELATION	COSINE CORRELATION
IRE-3 34 Requests	Abstract, Suffix 's'	Normed Recall Normed Precision	.8408 .5611	.8707 .6134
	Abstract, Stem	Normed Recall Normed Precision	.8725 .5829	.8777 .6167
	Abstract, Thesaurus -3	Normed Recall Normed Precision	.8974 .6041	.9067 .6574
CRAN-1 42 Requests	Abstract, Stem	Normed Recall Normed Precision	.8237 .5830	.8397 .6377
	Abstract, Thesaurus -3	Normed Recall Normed Precision	.8535 .6251	.8729 .6936
	Title, Stem	Normed Recall Normed Precision	.8082 .5979	.8120 .6212
ADI 35 Requests	Text, Suffix 's'	Normed Recall Normed Precision	.7546 .5097	.7768 .5462
	Text, Stem	Normed Recall Normed Precision	.7434 .4978	.7695 .5248
	Text, Thesaurus-1	Normed Recall Normed Precision	.7386 .4350	.7819 .5092
	Test, Thesaurus-2 (Hastie)	Normed Recall Normed Precision	.6589 .3602	.6884 .4332
	Abstract, Stem	Normed Recall Normed Precision	.7423 .4904	.7546 .5221
	Abstract, Thesaurus -1	Normed Recall Normed Precision	.7830 .5257	.8043 .5823

Performance results comparing cosine and overlap correlation coefficients, using logical vectors, for twelve options on three collections, using normalized recall and precision

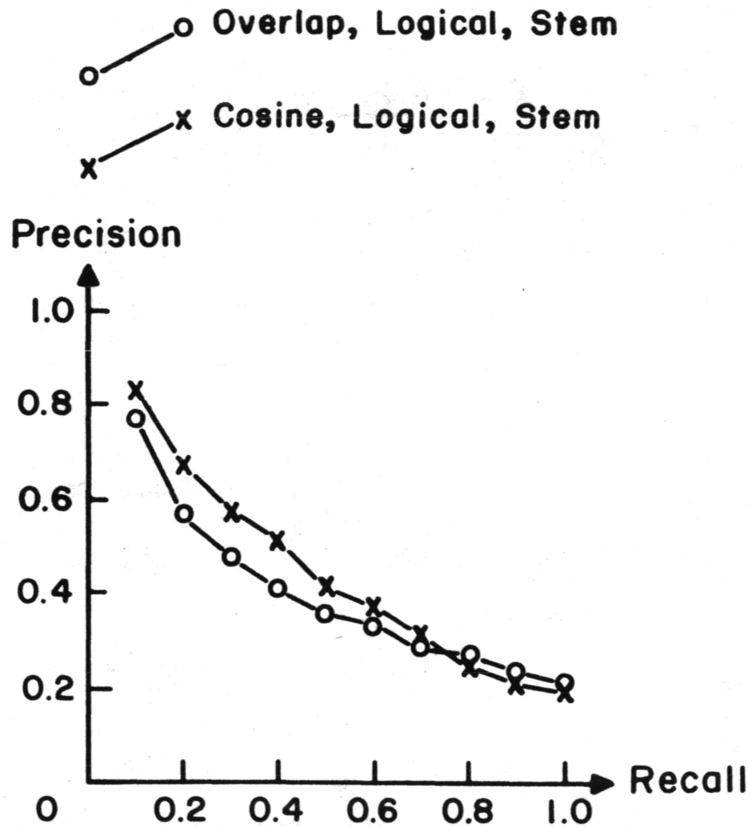
Figure 3.



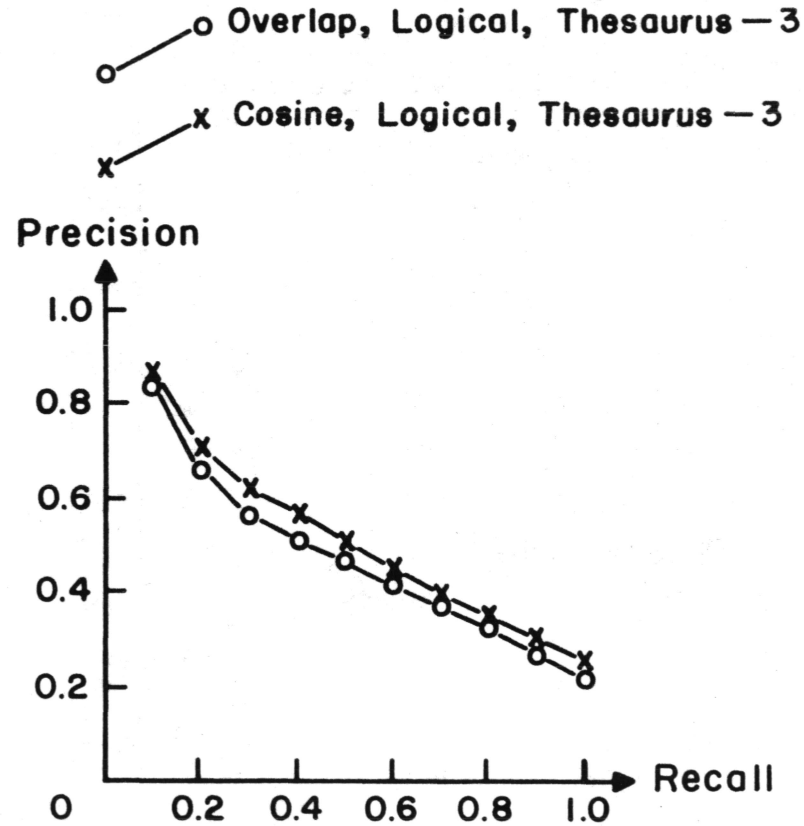
Performance curves comparing cosine and overlap correlation coefficients

using two Dictionaries, IRE-3 Collection

Fig. 4



Cranfield-1, Abstracts, 42 Requests

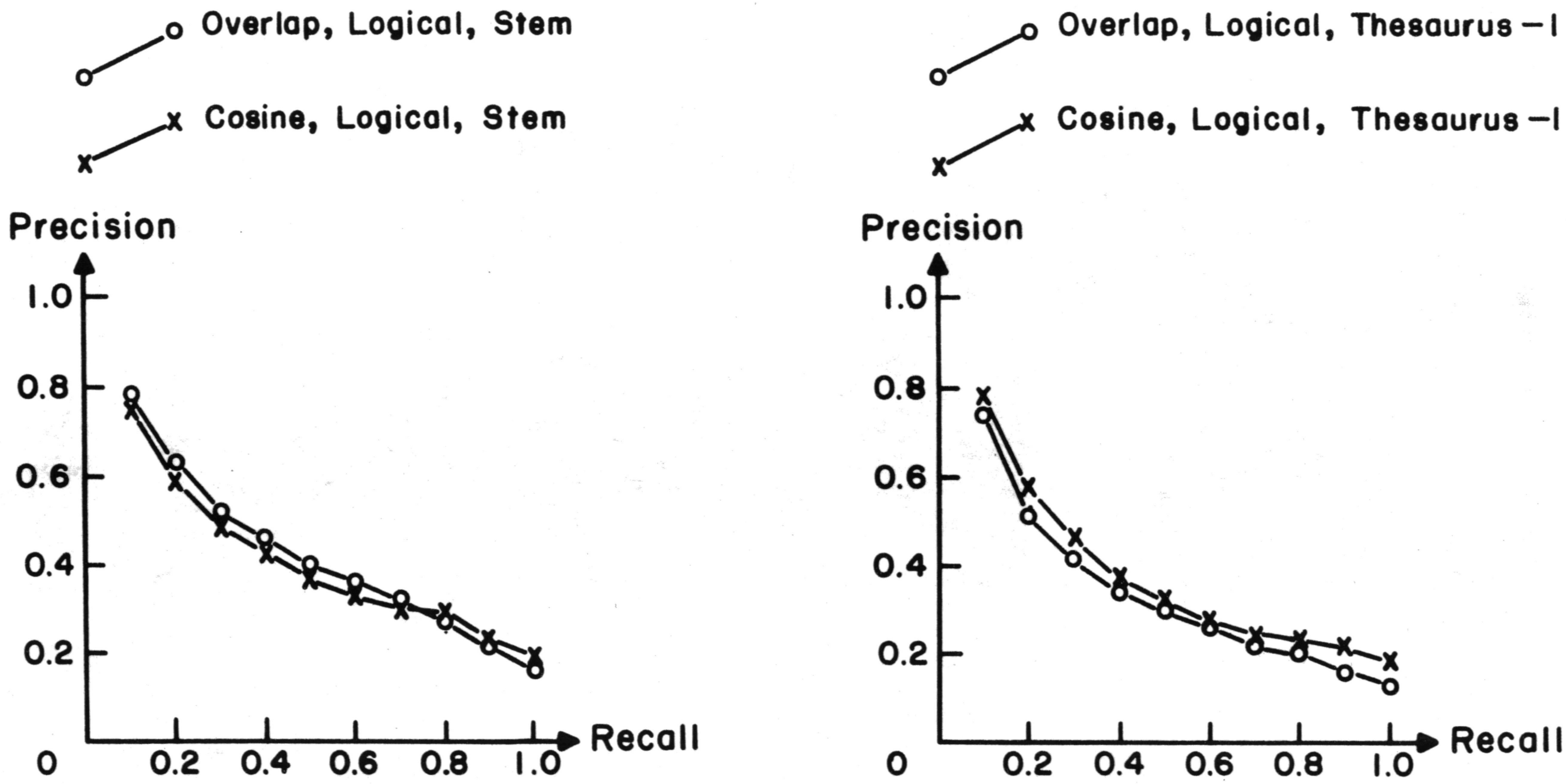


Cranfield-1, Abstracts, 42 Requests

Performance curves comparing cosine and overlap correlation coefficients

using two dictionaries, Cran-1 Collection.

Fig. 5

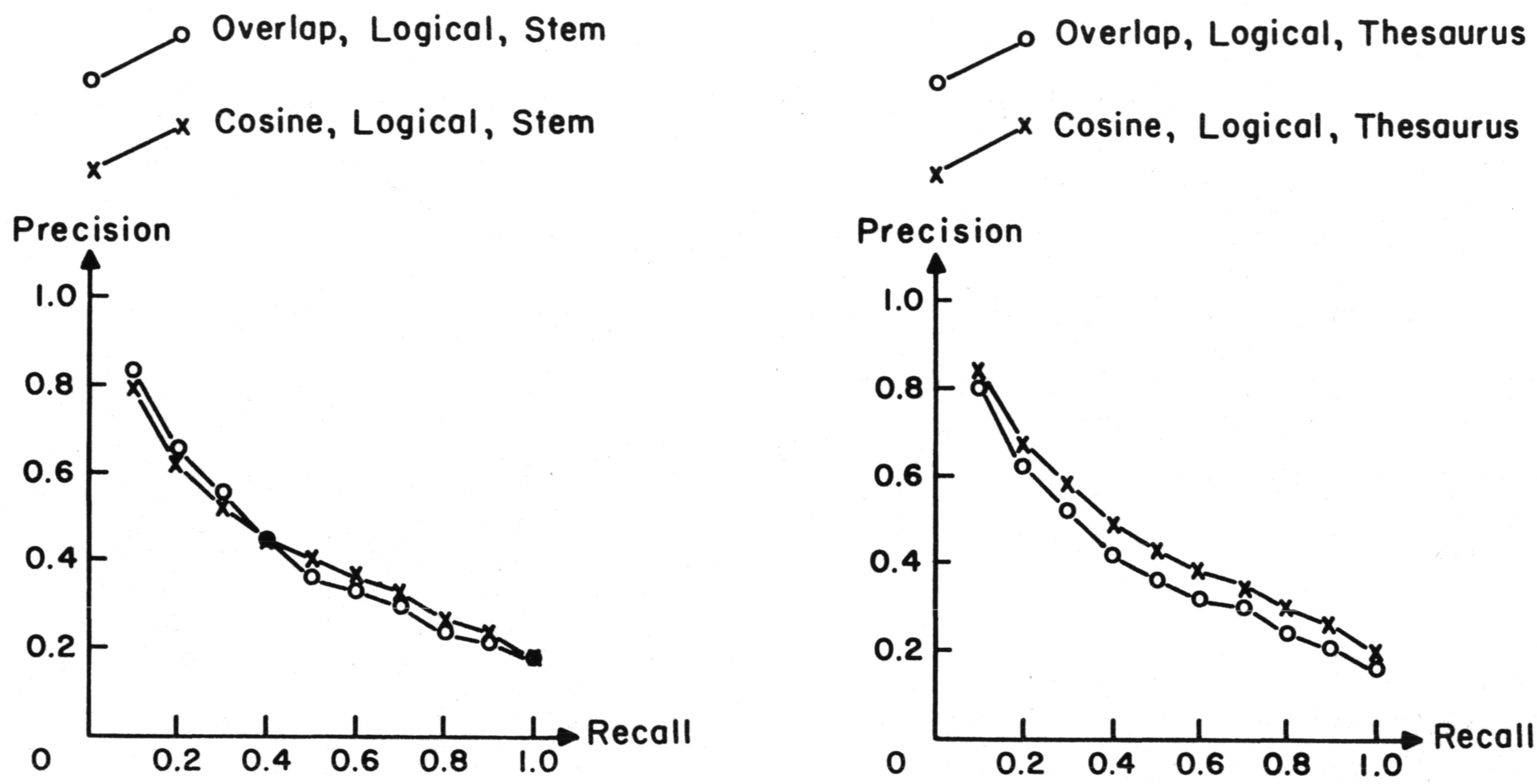


ADI, Text, 35 Requests

ADI, Text, 35 Requests

Performance curves comparing cosine and overlap correlation coefficients using two dictionaries, ADI Text Collection.

Fig. 6



ADI, Abstracts, 35 Requests

ADI, Abstracts, 35 Requests

Performance Curves Comparing Cosine and Overlap Correlation Coefficients Using Two Dictionaries, ADI Abstracts Collection.

Fig. 7.

Comparisons of individual request merit are given in Figure 8, where 58.8% to 76.5% of the requests favor cosine on IRE-3, 70.0% to 85.7% favor cosine on Cran-1, and 50% to 80% favor cosine on ADI. The use of the thesaurus dictionary, which gives a performance superior to stem on all collections, also shows good superiority for cosine.

C) Analysis of Performance

Since cosine consistently performs better than overlap, an adequate explanation must be sought. It should be noted that although the average results show cosine to be superior by only around 5% in precision, individual requests and relevant documents may show large changes in favor of both cosine and overlap. Figures 9 and 10 show results which strongly favor both cosine and overlap, respectively, for two individual requests. The individual relevant documents display large changes in rank with change in correlation coefficient. Using the Cran-1 Stem results, it is found that of 198 documents relevant to all 42 requests, 95 show rank improvements on cosine over overlap, 62 show the reverse improvement, and 41 show no change in rank. Figure 11 shows the amounts of change in rank for the 95 and 62 documents, revealing that the advantage is, as expected, with cosine.

Figure 12 gives a diagrammatic representation of what is happening when the ranking induced by overlap is changed to cosine. Overlap orders documents by match alone; to simplify the diagram five matching strengths only are recorded. At each matching strength both relevant and non-relevant documents may be found intermingled. If the ordering induced by cosine is now imposed on the documents, two types of changes take place. Firstly, some non-relevant documents are decreased in match and rank position, and some relevant documents are increased in match and rank position; naturally, such changes favor cosine as opposed to overlap. Secondly, the changes that

COLLECTION	INPUT AND DICTIONARY	EVALUATION MEASURE USED TO DETERMINE MERIT	NUMBER AND PERCENTAGE* OF INDIVIDUAL REQUESTS				
			COSINE SUPERIOR		OVERLAP SUPERIOR		BOTH EQUAL
IRE-3 34 Requests	Abstract, Stem	Normed Recall	20	58.8%	14	41.2%	0
		Normed Precision	21	61.8%	13	38.2%	0
	Abstract, Thesaurus -3	Normed Recall	26	76.5%	8	23.5%	0
		Normed Precision	22	64.7%	12	35.3%	0
CRAN-1 42 Requests	Abstract, Stem	Normed Recall	28	70.0%	12	30.0%	2
		Normed Precision	34	82.9%	7	17.1%	1
	Abstract, Thesaurus -3	Normed Recall	35	83.3%	7	16.7%	0
		Normed Precision	36	85.7%	6	14.3%	0
ADI 35 Requests	Text, Stem	Normed Recall	24	70.6%	10	29.4%	1
		Normed Precision	23	67.6%	11	22.4%	1
	Text, Thesaurus-1	Normed Recall	26	74.3%	9	25.7%	0
		Normed Precision	28	80.0%	7	20.0%	0
	Abstract, Stem	Normed Recall	17	50.0%	17	50.0%	1
		Normed Precision	18	52.9%	16	47.1%	1
	Abstract, Thesaurus -1	Normed Recall	23	67.6%	11	32.4%	1
		Normed Precision	28	80.0%	7	20.0%	0

* Percentages do not include cases where both options have equal merit.

Comparison of individual request merit giving the numbers of requests forming cosine and overlap with percentages on 8 options from three collections, according to merit assigned by normalized recall and precision.

Figure 8.

Request QA9 ADI Collection, full text, thesaurus dictionary. 2 relevant documents.

Cosine Correlation (Logical Vectors)

Rank	Relevant Documents	Correlation	Length (Thesaurus Concepts)
7	82	0.1867	43
24	50	0.1624	102

Normalized Recall = 0.8250

Normalized Precision = 0.4535

Overlap Correlation (Logical Vectors)

Rank	Relevant Documents	Correlation	Length (Thesaurus Concepts)
36	50	0.6666	102
59	82	0.5000	43

Normalized Recall = 0.4250

Normalized Precision = 0.1406

Example of individual request where cosine is superior to overlap

Figure 9.

Request QA2 ADI Collection, full text, thesaurus dictionary. 2 relevant documents.

Cosine Correlation (Logical Vectors)

Rank	Relevant Documents	Correlation	Length(Thesaurus Concepts)
25	12	0.2141	159
29	71	0.2108	162

Normalized Recall = 0.6813

Normalized Precision = 0.2732

Overlap Correlation (Logical Vectors)

Rank	Relevant Documents	Correlation	Length(Thesaurus Concepts)
14	12	0.8888	159
14	71	0.8888	162

Normalized Recall = 0.8438

Normalized Precision = 0.4345

Example of individual request where overlap is superior to cosine.

Figure 10.

	NUMBERS OF RELEVANT DOCUMENTS WITH RANK CHANGES IN RANGES: 1-5	6-10	11-20	21-30	31-40	41-50	51-75	
COSINE SUPERIOR	46	12	15	12	8	1	1	Total 95
OVERLAP SUPERIOR	31	9	15	4	2	1	0	Total 62

Changes in rank positions between cosine and overlap of 157 of the individual documents relevant to the 42 requests, Cran-1 collection, stem dictionary, logical vectors.

Figure 11.

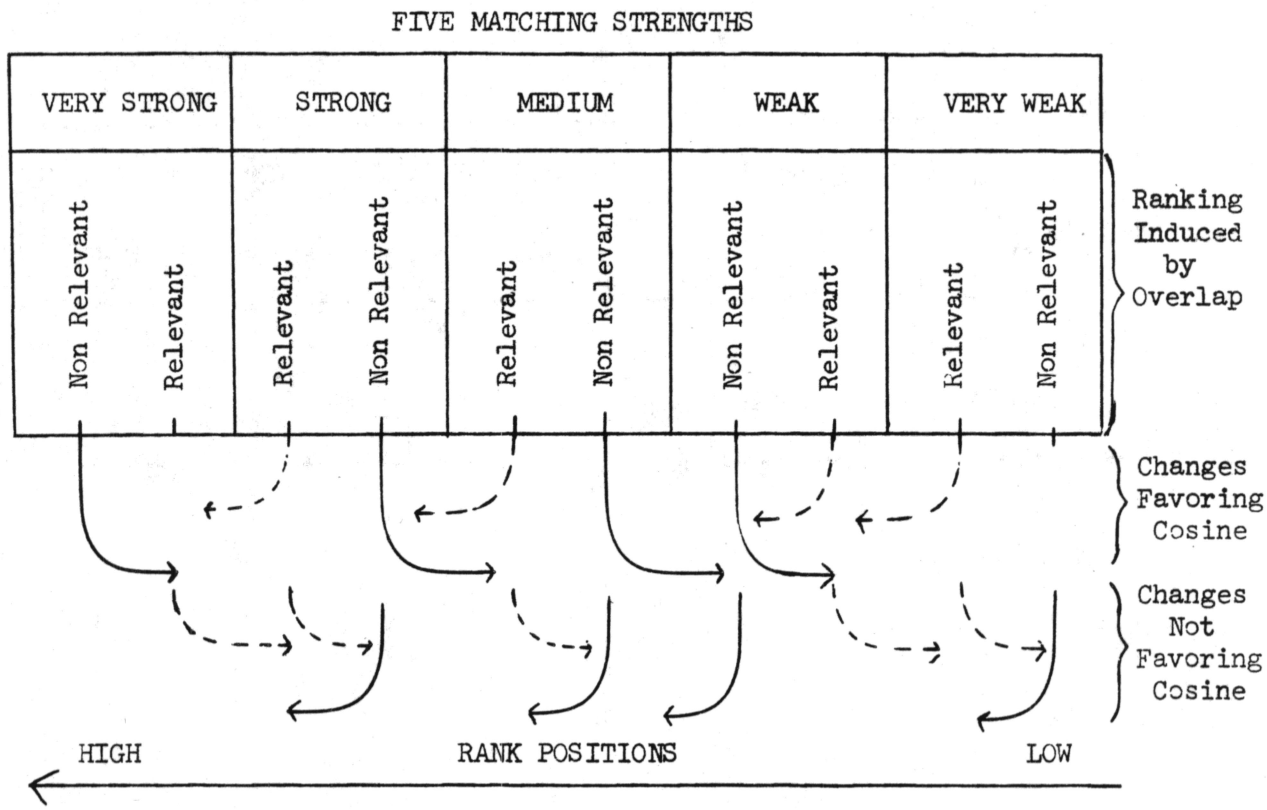


Diagram of changes in rank positions of relevant and non relevant documents, showing which changes favor cosine and which do not, starting with overlap.

Figure 12.

do not favor cosine occur when non-relevant documents are increased in match and relevant documents decreased. Figure 11 shows that both these changes take place on the Cran-1 Stem run at any rate, but that changes favoring cosine as against overlap occur at a ratio of about 3 to 2.

In seeking an explanation for this, it must be assumed that only the factor of document length brought to bear in the cosine correlation can be causing this result. It seems likely therefore that the distribution of documents by length among the strongly matched and weakly matched, and among the relevant and non-relevant is not even in the ordering induced by overlap. The first suggested explanation is simply that relevant documents tend to be short in length, and non-relevant documents are long. This, however, does not turn out to be the case: in the ADI collection 70 of the total collection of 82 are relevant to one or more of the requests, and in the Cran-1 collection 153 out of the 200 are at some time relevant. Figure 13 shows, for the Cran-1 collection, that the average document length varies by trivial amounts comparing all 200 against the 153 which are sometimes relevant, and the 47 which are never relevant.

The remaining explanation is that the distribution of documents by length differs between relevant and non-relevant documents for a given level of match. Specifically, it is hypothesized that highly matched non-relevant documents (i.e. highly ranked on overlap) are longer than average. Analysis is performed to test this hypothesis, by taking pairs of relevant and non-relevant documents, both pairs having an almost identical (and string) match on overlap, and comparing document lengths. Figure 14 gives an individual example, showing how the non-relevant document has 104 stem concepts and the relevant one 39. When more than one relevant and non-relevant document have identical matches, lengths can be averaged over all such documents; the results given in Figure 15 are thus based on over 100 documents. Figure 15 shows

	Average (Mean) Concepts Per Documents
200 Documents	60.6
153 Docts. that are Sometimes Relevant	60.9
47 Docts. that are Always Non-Relevant	59.7

Document length data for Cran-1 Collection, using stem dictionary.

Figure 13.

Q230 Cran-1, stem dictionary.

Overlap Correlation (Logical Vectors)

Rank	Document	Correlation	Length
2	794 (Non-Relevant)	0.5000	104 Stem Concepts
2	10A (Relevant)	0.5000	39 Stem Concepts

Cosine Correlation (Logical Vectors)

Rank	Document	Correlation	Length
1	10A (Relevant)	0.1961	39 Stem Concepts
2	31H (Non-Relevant)	0.1307	39 Stem Concepts

Comparison of document lengths using two pairs of relevant/non-relevant documents that are similarly correlated with the search request, using overlap and cosine.

Figure 14.

	AVERAGE(MEAN) CONCEPTS PER DOCUMENT		
	Highly Ranked Matched Pairs From Overlap, Pairs from 32 Requests	Highly Ranked Matched Pairs From Cosine, Pairs from 36 Requests	Totals in Collection
Relevant	64.3	54.8	60.9
Non-Relevant	97.3	51.9	60.6

Document length data using relevant/non-relevant documents paired on the basis of similar correlation coefficients using highly ranked pairs to contrast overlap and cosine.

Figure 15.

that, on overlap, non-relevant documents are longer than relevant, the length being considerably above the average. Similar parameters are calculated for the cosine ordering in Figures 14 and 15, and very similar lengths are then obtained for relevant and non-relevant documents in this case.

Since the analysis shows that non-relevant documents with strong matches are longer than average, it is now obvious that cosine effectively lowers the ranks of these documents, and thus provides a better retrieval performance than overlap. Although it is certainly the case that non-relevant documents with weak matches must be shorter than average, it seems that their low match (sometimes as low as zero) is never sufficient to increase their rank by any significant amount on overlap; it is the strongly matched non-relevant only that are responsible for the superiority of the cosine correlation.

This phenomenon is probably caused by the fact that not all non-relevant documents have an equal probability of resulting in spurious matches; as seems logical, the probability of spurious matches is greater in larger documents. Spurious matches result from spurious concept combinations, which arise because no judgments of importance are made to discriminate between request concepts; that is, any combination of, say, three concepts (out of six in a request) is assumed to be as important as any other. An example of this is given in Figure 14, where both a non-relevant and a relevant document match in three out of the six concepts; the data of Figure 16 show, however, that the non-relevant match on words such as "during", "report" and "measurement" which turn out to be spurious. Such spurious matches are more likely to occur for long non-relevant documents than for short ones. The logical search formulations used in post-coordinate manual systems would eliminate many such false matches; some success in this direction can be achieved without manual search formulation by use of weighting methods, to be described next.

REQUEST DATA		MATCHES WITH DOCUMENTS	
Stem Concept Number	Word	Relevant 10A	Non-Relevant 794
1229	ABLATION	✓	
1740	TECHNICAL		
2326	DURING		✓
2382	REPORT		✓
2509	FLIGHT	✓	
2513	MEASUREMENT	✓	✓

Matches achieved with a relevant and non-relevant documents in relation to request Q230 ("Technical report on measurement of ablation during flight"), Cran-1 Collection using stem dictionary.

Figure 16.

4. SMART Test Results — Weighting Scheme

A) Description of Weighting Scheme

Weighted document and request vectors, rather than the binary ones presented up to now in considering the overlap and cosine, may be constructed by assigning to each content identifier a 'weight' that reflects the importance or usefulness of that identifier. Since the assignment of weights is ideally done by automatic means, the weighting scheme in use with SMART relies initially on frequency information. When suffix 's' and stem dictionaries are used, concepts are weighted entirely by frequency of occurrence of the concepts in the documents (or requests): thus a concept that occurs three times in a document will receive three times the weight of a concept that appears only once.

With a thesaurus dictionary in use, or any dictionary that permits a word to appear in more than one concept group, an additional adjustment of the weight reflects word ambiguity. Thus, if a word appears in more than one concept group it is assumed to be ambiguous, and the weight assigned to the concept number representing the ambiguous word is decreased according to the number of concept groups in which the word appears. Many other modifications to a weighting procedure of this type can be suggested; for example, where abstracts and titles are used the title words may be given higher weights than the abstract words.

Both the overlap and cosine correlation coefficients may be used with weighted vectors. For example, if a hypothetical request and document are weighted as follows:

Concept	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u
Request Weight	1	2	1	1	1	3	1	2													
Document Weight	–	–	–	1	3	1	4	2	1	3	7	1	1	2	3	5	1	1	1	1	1

using equations (1) and (2) already given;

$$\text{OVERLAP} = \frac{1 + 1 + 1 + 1 + 2}{\text{Min}\ (12\ ,\ 39)} = \frac{6}{12} = 0.50\ ;$$

$$\text{COSINE} =$$

$$\frac{(1 \times 1) + (1 \times 3) + (3 \times 1) + (1 \times 4) + (2 \times 2)}{\sqrt{(1^2+2^2+1^2+1^2+1^2+3^2+1^2+2^2)\times(1^2+3^2+1^2+4^2+2^2+1^2+3^2+7^2+1^2+1^2+2^2+3^2+5^2+1^2+1^2+1^2+1^2+1^2)}}$$

$$= \frac{15}{\sqrt{(22 \times 135)}} = 0.28\ .$$

It should be noted that the overlap numerator uses the lowest weight assigned to a given matching concept in a request and a document; thus, for requests in which all the concepts are assigned a weight of 1 only, none of the weighted concepts in the document can exert any influence on the final correlation. In this situation, which is quite common for the requests tested, the weighted vector result is then identical to the unweighted result for overlap. For this reason, the comparison of overlap and cosine was made using unweighted (logical) vectors; comparisons of the weighted versus unweighted (numeric versus logical) vectors will use the cosine correlation coefficient.

Weighting achieved by manual or semi-manual decisions may also become a part of automatic retrieval systems of the future, under the assumption that such methods do not require large amounts of time and effort, and give useful improvements in performance. An example of this, using selective request weighting to improve vital request notions, is given in section VIII.

B) Retrieval Performance Results

Thirteen comparison runs on three collections are presented in Figure 17, evaluated by normalized recall and normalized precision. All IRE-3 results show numeric to be superior to logical, and most the runs on Cran-1

COLLECTION	INPUT and DICTIONARY	EVALUATION MEASURE	LOGICAL VECTORS	NUMERIC VECTORS
IRE-3 34 Requests	Abstract, Suffix 's'	Normed Recall Normed Precision	.8707 .6134	.8817 .6484
	Abstract, Stem	Normed Recall Normed Precision	.8777 .6167	.8954 .6746
	Abstract, Thesaurus -3	Normed Recall Normed Precision	.9067 .6574	.9268 .7382
CRAN-1 42 Requests	Abstract, Stem	Normed Recall Normed Precision	.8397 .6377	.8644 .6704
	Abstract, Thesaurus -3	Normed Recall Normed Precision	.8729 .6936	.8837 .6952
	Title, Stem	Normed Recall Normed Precision	.8120 .6212	.8112 .6185
ADI 35 Requests	Text, Suffix 's'	Normed Recall Normed Precision	.7768 .5462	.7520 .5308
	Text, Stem	Normed Recall Normed Precision	.7695 .5248	.7779 .5573
	Text, Thesaurus-1	Normed Recall Normed Precision	.7819 .5092	.8206 .6273
	Text, Thesaurus-2 (Hastie)	Normed Recall Normed Precision	.6589 .3602	.7774 .5441
	Abstract, Suffix 's'	Normed Recall Normed Precision	.7296 .5044	.7253 .4997
	Abstract, Stem	Normed Recall Normed Precision	.7423 .4904	.7601 .5326
	Abstract, Thesaurus -1	Normed Recall Normed Precision	.8043 .5823	.8016 .6069

Performance results comparing numeric and logical vectors using cosine correlation for 13 options on three collections, using normalized recall and precision.

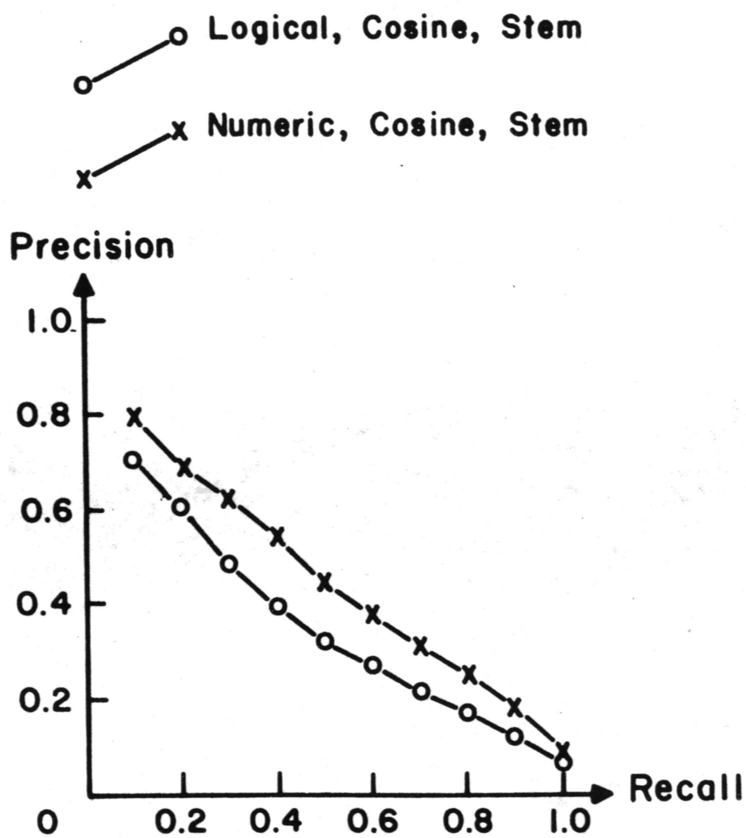
Figure 17.

and ADI give the same result. The exceptions are the stem dictionary on abstracts and titles Cran-1, and suffix 's' dictionary on abstracts and text ADI. Figures 18, 19, 20 and 21 present precision versus recall graphs for the stem and thesaurus dictionaries on the IRE-3 collection (Figure 18), the Cran-1 collection (Figure 19), and the ADI collection using text (Figure 20) and abstracts (Figure 21). General merit strongly favors numeric, the only exceptions being the low recall high precision area on Cran-1 Stem, and the small differences in the curves on ADI abstract stem. Since the normalized measures for both recall and precision show ADI test suffix 's' to prefer logical vectors, a precision versus recall graph of this output together with ASI abstracts suffix 's' is given in Figure 22. The graphs show numeric to be superior on both plots up to 0.8 recall; the difference in merit obtained by the normalized measures compared with the graphs of standard measures is considered in Section II.

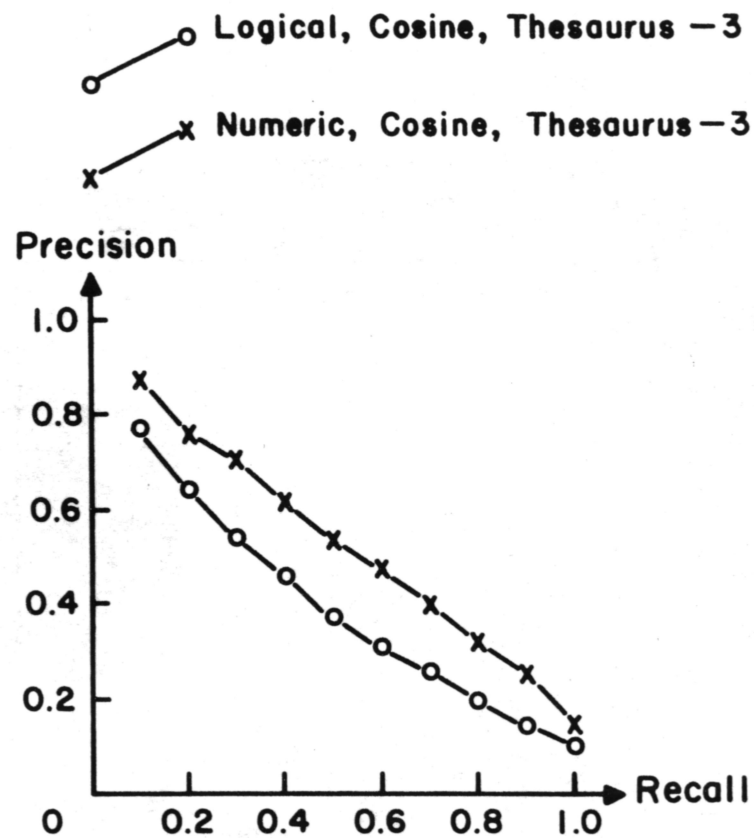
Comparisons of individual request merit are given in Figure 23, where 76.5% to 88.2% of the requests favor numeric on IRE-3, 51.4% to 77.8% favor numeric on Cran-1, and 45.7% to 65.7% favor numeric on ADI.

C) Analysis of Performance

The thesaurus dictionaries show a better improvement for numeric over logical than the stem and suffix 's' dictionaries; a specific reason for this is suggested by the data in Figure 24. Using four ADI dictionaries and the ADI text results, it seems that numeric gives the best increases in performance over logical with dictionaries that contain few concept classes. The dictionary with the smallest number of classes is an exception to this for four of the performance measures used, this dictionary, however, has a performance that is inferior to the stem dictionary thus explaining the discrepancy. The grouping of words achieved by a thesaurus provides a greater



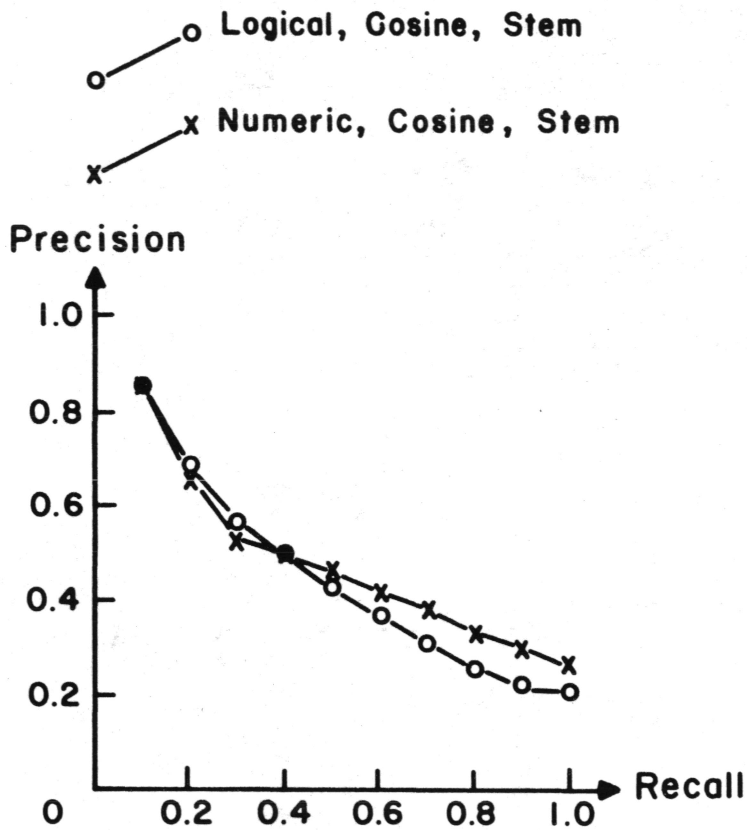
IRE-3, 34 Requests



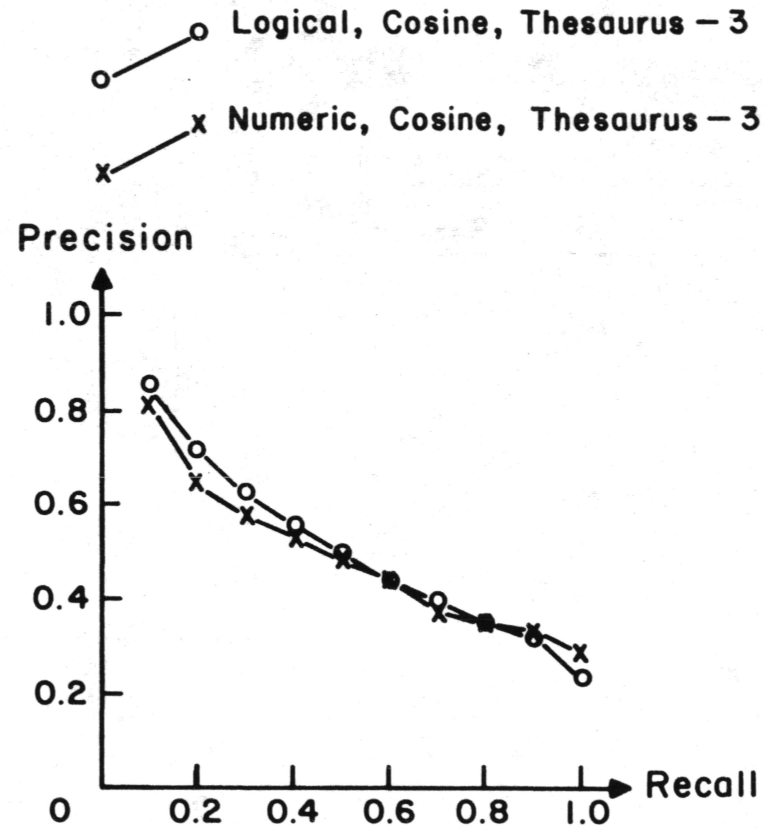
IRE-3, 34 Requests

Performance curves comparing numeric and logical vectors using two dictionaries, IRE-3 Collection.

Fig. 18



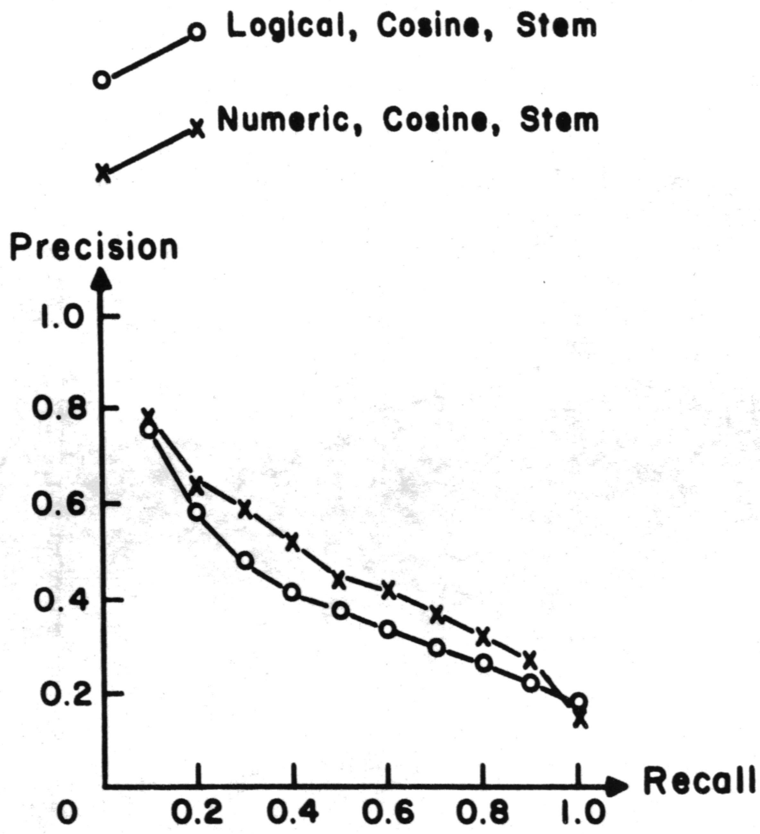
CRAN-1, 42 Requests



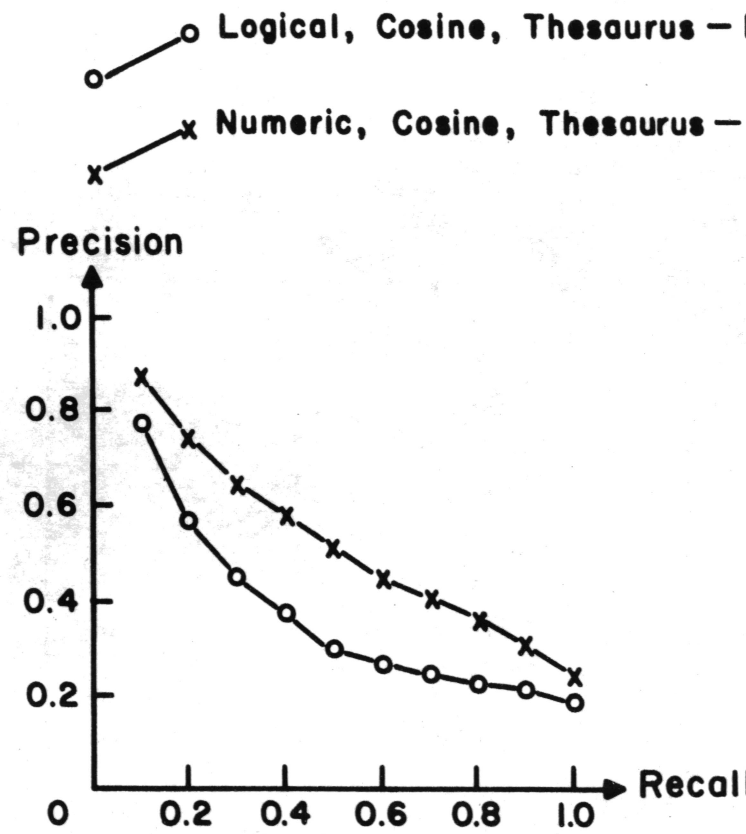
CRAN-1, 42 Requests

Performance curves comparing numeric and logical vectors using two dictionaries, Cran-1 Collection.

Fig. 19.



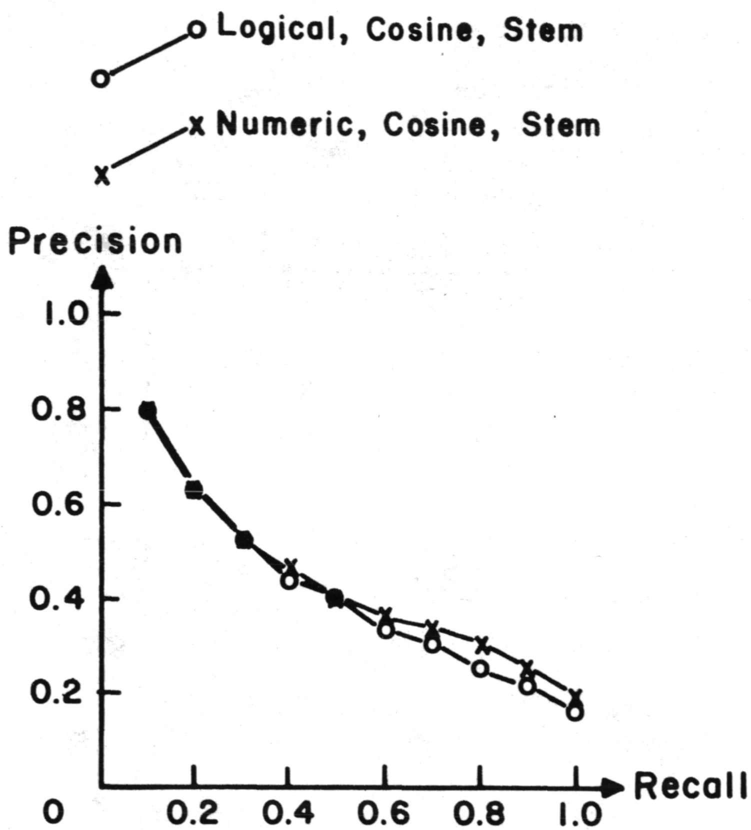
ADI, Text, 35 Requests



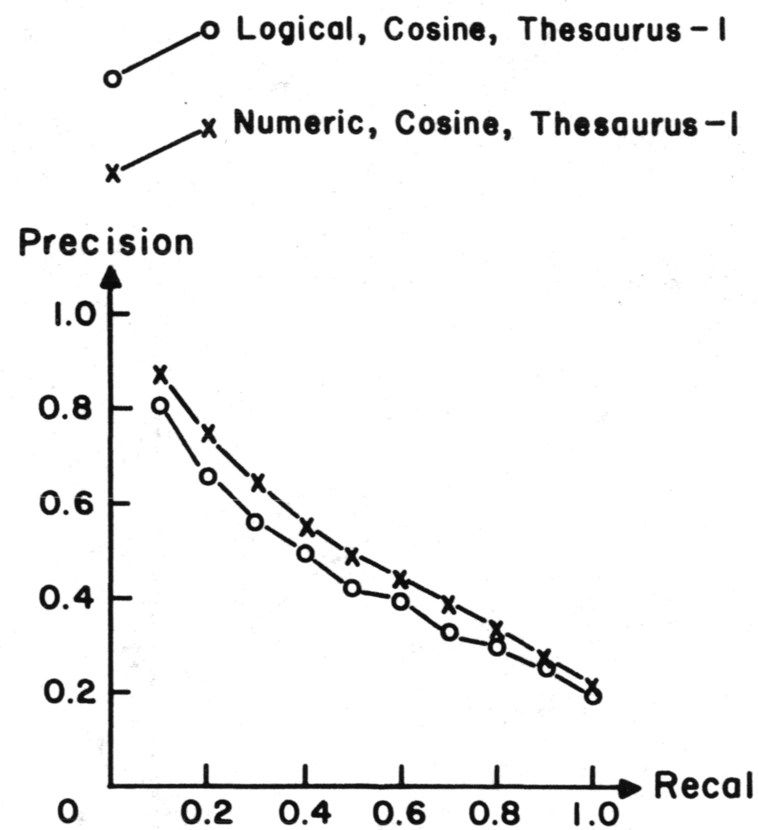
ADI, Text, 35 Requests

Performance curves comparing numeric and logical vectors
two dictionaries, ADI Text Collection.

Fig. 20.



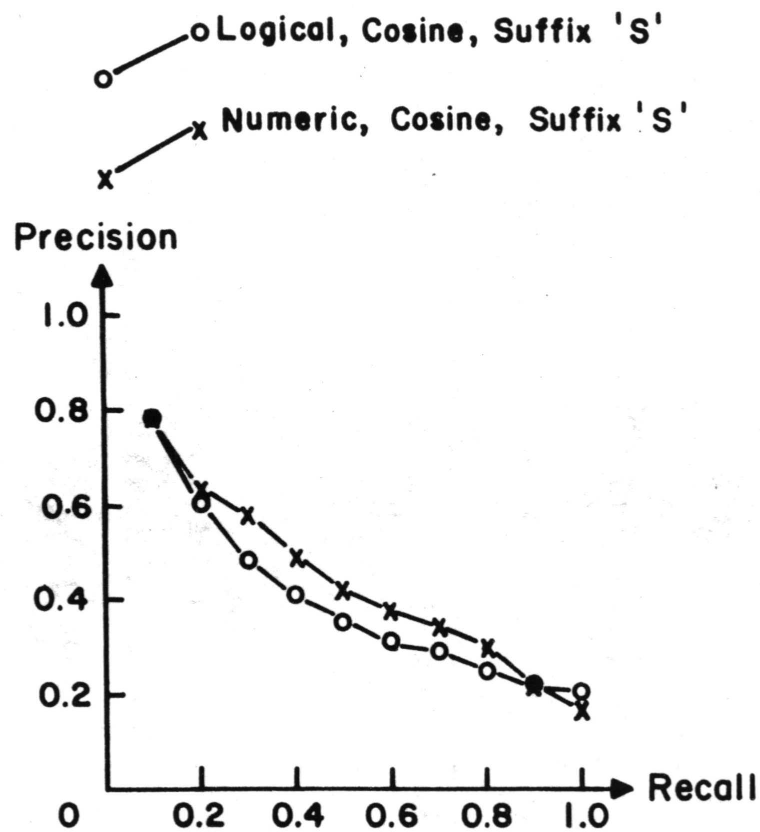
ADI, Abstracts, 35 Requests



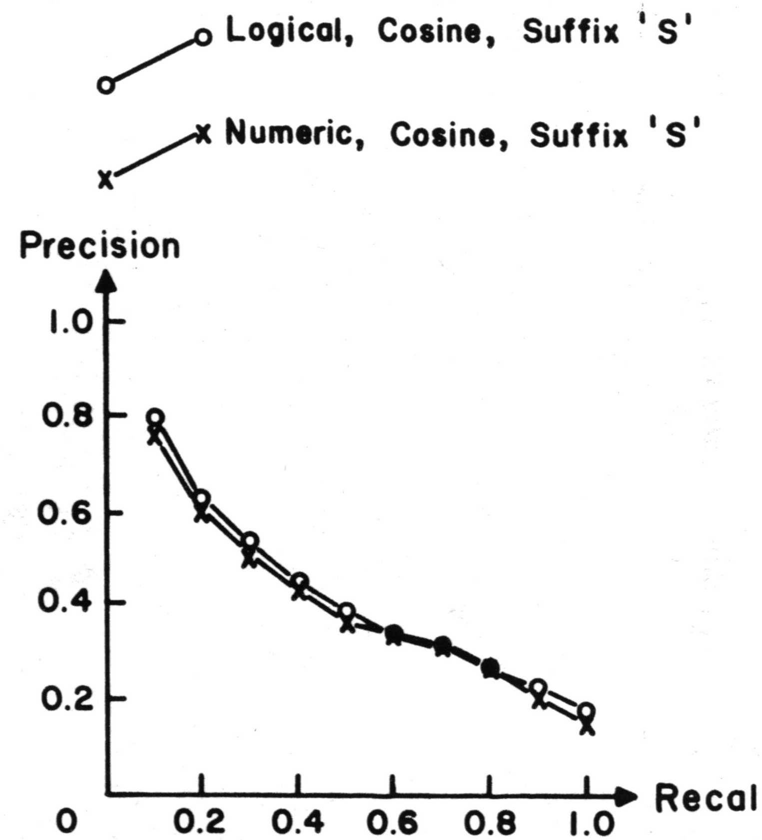
ADI, Abstracts, 35 Requests

Performance curves comparing numeric and logical vectors two dictionaries, ADI Abstracts Collection.

Fig. 21.



ADI, Text, 35 Requests



ADI, Abstracts, 35 Requests

Performance curves comparing numeric and logical vectors using suffix 's' dictionary, ADI Text and Abstracts Collections.

Fig. 22.

COLLECTION	INPUT AND DICTIONARY	EVALUATION MEASURE USED TO DETERMINE MERIT	NUMBER AND PERCENTAGE* OF INDIVIDUAL REQUESTS				
			NUMERIC SUPERIOR		LOGICAL SUPERIOR		BOTH EQUAL
IRE-3 34 Requests	Abstract, Stem	Normed Recall	26	76.5%	8	23.5%	0
		Normed Precision	26	76.5%	8	23.5%	0
	Abstract, Thesaurus -3	Normed Recall	26	78.8%	7	22.2%	1
		Normed Precision	30	88.2%	4	11.8%	0
CRAN-1 42 Requests	Abstract, Stem	Normed Recall	28	77.8%	8	22.2%	6
		Normed Precision	26	70.3%	11	29.7%	5
	Abstract, Thesaurus -3	Normed Recall	19	51.4%	18	48.6%	5
		Normed Precision	21	53.8%	18	46.2%	3
ADI 35 Requests	Text, Suffix 's'	Normed Recall	14	40.0%	21	60.0%	0
		Normed Precision	17	48.6%	18	51.4%	0
	Text, Stem	Normed Recall	18	52.9%	16	47.1%	1
		Normed Precision	16	45.7%	19	54.3%	0
	Text, Thesaurus-1	Normed Recall	22	62.9%	13	37.1%	0
		Normed Precision	23	65.7%	12	34.3%	0
	Abstract, Suffix 's'	Normed Recall	17	51.5%	16	48.5%	2
		Normed Precision	19	57.6%	14	42.4%	2
	Abstract, Stem	Normed Recall	17	51.5%	16	48.5%	2
		Normed Precision	18	51.4%	17	48.6%	0
	Abstract, Thesaurus -1	Normed Recall	16	48.5%	17	51.5%	2
		Normed Precision	18	54.5%	15	45.5%	2

* Percentages do not include cases where both options have equal merit.

Comparisons of individual request merit giving the numbers of requests favoring numeric and logical with percentages on 10 options from three collections, according to merit assigned by normalized recall and precision.

Figure 23.

DICTIONARY (ADI TEXT)	TOTAL CONCEPTS IN DICTIONARY	AMOUNT OF PERFORMANCE INCREASE OF NUMERIC OVER LOGICAL				
		NORMED. RECALL	NORMED. PRECISION	PRECISION AT RECALL: .2	.5	.8
Suffix 's'	7,615	-.0248	-.0154	+.0363	+.0692	+.0414
Stem	5,606	+.0084	+.0325	+.0778	+.0755	+.0571
Thesaurus-1	541	+.0387	+.1181	+.1874	+.2032	+.1003
Thesaurus-2	289	+.0890	+.1109	+.1037	+.0957	+.0867

Comparisons of performance increases from numeric to logical in four ADI dictionaries, using ADI text results with five measures of performance.

Figure 24.

range of weights in the vectors, that is, combinations of words may cause some concepts to be very highly weighted, and rare examples of weights in excess of 100 have been noted.

The superiority of numeric on Cran-1 Stem, which was left in some doubt on the average measures (although not in the individual request figures) is only marginally established by looking at the 198 individual relevant documents separately. Some 41 of the documents have identical ranks on numeric and logical, while 85 have better ranks on numeric, and 72 have better ranks on logical. The 85 that are better on numeric show larger increases in the rank positions involved, as shown in Figure 25.

These large scale changes that work in both directions, some favoring numeric and some logical, are illustrated for one individual request by the data in Figure 26. Six of the ten highest ranked documents on logical receive rank positions below 1.0 in numeric; this large change in document ranks favors numeric in this example. In order to determine how the weighting scheme is used to achieve a more effective discrimination between relevant and non-relevant documents, further data on these 17 documents are given in Figures 27, 28, 29 and 30.

Figure 27 shows the ordering resulting from logical (cosine), giving correlations, matching concepts and total document concepts. Figure 28 gives the numeric ordering, together with data about the matching concepts and document concepts from which the final correlation is derived. For example, the correlation given to document 420 (relevant) of 0.4421 is derived from:

$$\text{Cosine Numeric Correlation} = \frac{\begin{array}{c}\text{Sum of matching concept doct. weights}\\ \text{x request weights}\end{array}}{\sqrt{\left(\begin{array}{c}\text{sum of squares}\\ \text{of req. wts.}\end{array}\right) \times \left(\begin{array}{c}\text{sum of squares}\\ \text{of doct. wts.}\end{array}\right)}}$$

$$= \frac{3312}{\sqrt{1584 \times 35{,}424}} = 0.4421$$

	NUMBERS OF RELEVANT DOCUMENTS WITH RANK CHANGES IN RANGES:								
	1-5	6-10	11-20	21-30	31-40	41-50	51-75	76-100	
NUMERIC SUPERIOR	34	12	10	10	6	6	4	3	Total 85
LOGICAL SUPERIOR	32	10	11	10	5	3	1	0	Total 72

Changes in rank positions between numeric and logical of 157 of the individual documents relevant to the 42 requests, Cran-1 collection, stem dictionary, cosine correlation.

Figure 25.

Request Q 137 8 request concepts, sum of weights 108, sum of squares of weights 1584.

Rank	Document	Cosine Numeric Correlation	MATCHING CONCEPTS Total	Sum of wts. in Doct.	Sum of Doc.Wts. x Req Wts	DOCUMENT CONCEPTS Total	Sum of Weights	Sum of Squares of Weights
1	420 R	.4421	3	180	3312	48	984	35,424
2	921	.4305	4	156	2736	71	1116	25,488
3	779	.4109	3	192	3456	67	1344	44,640
4	699	.4081	5	180	3600	61	1308	49,104
5	797 R	.3679	5	216	3744	109	2136	65,376
6	34A	.3565	5	204	3600	60	1476	64,368
7	916	.3468	6	132	2160	62	1056	24,480
8	794 R	.3444	3	156	3024	104	1824	48,672
9	683	.3355	4	204	3168	77	1764	66,960
10	793 R	.3325	4	132	2160	54	1020	26,640
15	879	.2896	4	60	864	17	276	5,616
17	795 R	.2665	4	48	720	26	336	4,608
27	316	.2147	4	60	864	43	612	10,224
32	33H	.2082	6	120	1728	63	1344	43,488
34	672	.1979	4	48	720	40	552	8,352
40	415	.1655	4	48	720	57	780	11,952
43	796 R	.1492	4	60	1008	79	1296	28,800

R = Relevant Document

Normalized Recall = 0.9459

Normalized Precision = 0.7610

Results of a single request using numeric vectors and cosine correlation

Figure 28

		6 Relevant Docts.	11 Non-Relevent Docts.
Logical Cosine	Average No. of Matching Concepts	3.8	4.5
Numeric Cosine	Average Sum of Weights of Concepts that Match in Document	204	167
	Average Sum of Products of Request + Document Matching Weights	2,328	2,147

Average values for factors that go into the numerator of the cosine correlation comparing relevant and non-relevant documents for logical and numeric vectors, from the results of Figures 27 and 28.

Figure 29.

		6 Relevant Documents	11 Non-Relevant Documents
Logical (Cosine)	Average Number of Concepts in Documents.	70.0	56.2
	Relevant Length as Percentage of Non-Relevant	124.6%	
Numeric (Cosine)	Average Sum of Weights of Concepts in Document.	1,266	1,057
	Relevant as Percentage of Non-Relevant.	119.8%	
	Average Sum of Squares of Weights of Concepts in Document.	34,920	32,243
	Relevant as Percentage of Non-Relevant.	108.3%	

Average values for factors that go into the denominator of the cosine correlation comparing relevant and non-relevant documents for logical and numeric vectors, from the results of Figures 27 and 28.

Figure 30.

Examining Figures 26, 27, and 28, the following individual cases may be noted:

1. Relevant documents that are long receive low ranks on logical cosine unless very highly matched. The great length of document 797 is offset on cosine numeric by the very high weights associated to the matching concepts.

2. Relevant documents having few matching concepts that are ranked below certain higher-matching non-relevant documents with cosine logical receive improved ranks on numeric cosine when matching concepts are highly weighted (see documents 420 and 794). When matching concepts of relevant documents are not highly weighted, the numeric measure usually worsens their rank positions (see documents 793, 795 and 796).

From this data two hypotheses emerge: First, if a relevant and a non-relevant document have similar numbers of matching concepts or similar rank positions using logical cosine, the introduction of weights will on average result in higher matches for the relevant than the non-relevant documents. It seems reasonable that low weighted matching concepts should have a higher probability of reflecting a trivial occurrence of those concepts in the document than is the case for concepts that are highly weighted.

The second hypothesis is that weights assigned to the matching concepts provide some measure of discrimination between concepts according to their importance; this discrimination is of value in matching relevant documents. In such cases spurious matches with many concepts are distinguished from correct matches even if obtained with fewer concepts.

Evidence that the first hypothesis holds for request Q137 is given in Figure 29, showing that the change from logical to numeric produces far better cosine correlation values in the numerator for relevant documents compared with the non-relevant documents. In this example, numeric also produces better denominator values as well for the relevant documents (see Figure 30); no explanation for this phenomenon is suggested.

Support for the second hypothesis is obtained in Figure 31, where the Cranfield request on ablation is again used as an example. The two relevant documents have poor matches with the request, but since the matching concept is the most important request word, weights of 4 are derived from the frequency of occurrence in the document; a non-relevant document with more matching concepts, but spurious ones, is ranked below the relevant with the weights in use.

5. Conclusions and Suggested Further Studies

A matching function that consists of the cosine correlation with numeric vectors has been shown to be nearly always superior to the use of either the overlap correlation or logical vectors. A simplified table of results using precision versus recall graphs, for normalized measures, and individual requests is given in Figure 32.

The cosine correlation coefficient works better than the overlap coefficient because the factor of document length included in the cosine coefficient reduces the request/document correlation for a number of the highly matched non-relevant documents, since there is a strong correlation among non-relevant documents between number of matching concepts and the length of the document. The superiority of weighted concepts evidenced by the superiority of numeric as opposed to logical vectors is due to two reasons. The first is that highly weighted matching concepts tend to distinguish between important and trivial occurrences of those concepts in the documents, and thus tend to make better distinctions between relevant and non-relevant documents. The second reason is that if different concepts in a request receive different weights, such weighting does discriminate between vital

Q 230 Cran-1 Collection, Stem Dictionary. 6 Request concepts.

Cosine Logical

Rank	Document	Correlation	Match	Concepts
4	794 (Non-relevant)	.1200	3	DURING, REPORT, MEASUREMENT
18	09I (Relevant)	.0690	1	ABLATION
32	10+ (Relevant)	.0566	1	ABLATION

Cosine Numeric

Rank	Document	Correlation	Concepts and Weights
3	09I (Relevant)	.1837	ABLATION 4
7	10+ (Relevant)	.1395	ABLATION 4
17	794 (Non-relevant)	.0887	DURING 1, REPORT 1, MEASUREMENT 2

Comparison of logical and numeric results for a single request, showing the superiority of numeric.

Figure 31.

COLLECTION	INPUT AND DICTIONARY	COSINE(C) VERSUS OVERLAP(ø) I R-2	I R-5	I R-8	II NR	II NP	III (NR)	III (NP)	NUMERIC(N) VERSUS LOGICAL(L) I R-2	I R-5	I R-8	II NR	II NP	III (NR)	III (NP)
IRE-3 34 Requests	Abstract, Stem	$\boxed{C}$	$\boxed{C}$	C	C	C	C	C	$\boxed{N}$	$\boxed{N}$	$\boxed{N}$	N	$\boxed{N}$	N	N
	Abstract, Thesaurus-3	$\boxed{C}$	$\boxed{C}$	C	C	$\boxed{C}$	C	C	$\boxed{N}$	$\boxed{N}$	$\boxed{N}$	N	$\boxed{N}$	N	N
CRAN-1 42 Requests	Abstract, Stem	$\boxed{C}$	C	ø	C	$\boxed{C}$	C	C	L	N	N	$\boxed{N}$	N	N	N
	Abstract, Thesaurus-3														
ADI 35 Requests	Abstract, Stem	ø	C	C	C	C	C	C	N	N	N	N	N	N	N
	Abstract, Thesaurus-1														
	Text, Stem	ø	ø	C	C	C	C	C	$\boxed{N}$	$\boxed{N}$	$\boxed{N}$	N	N	N	L
	Text, Thesaurus-1	C	C	C	C	$\boxed{C}$	C	C	$\boxed{N}$	$\boxed{N}$	$\boxed{N}$	N	$\boxed{N}$	N	N

I Precision at Recall 0.2 (R-2), 0.5 (R-5) and 0.8 (R-8). $\boxed{}$ = over .05 better.
[No Box] = up to .05 better.

II Normalized Recall (NR) and Normalized Precision (NP). $\boxed{}$ = over .05 better.
[No Box] = up to .05 better.

III Individual Requests based on merit assigned by Normed Recall (NR) and Normed Precision (NP).
○ = over 60% of requests better. [No Circle] = between 50% and 60% better.

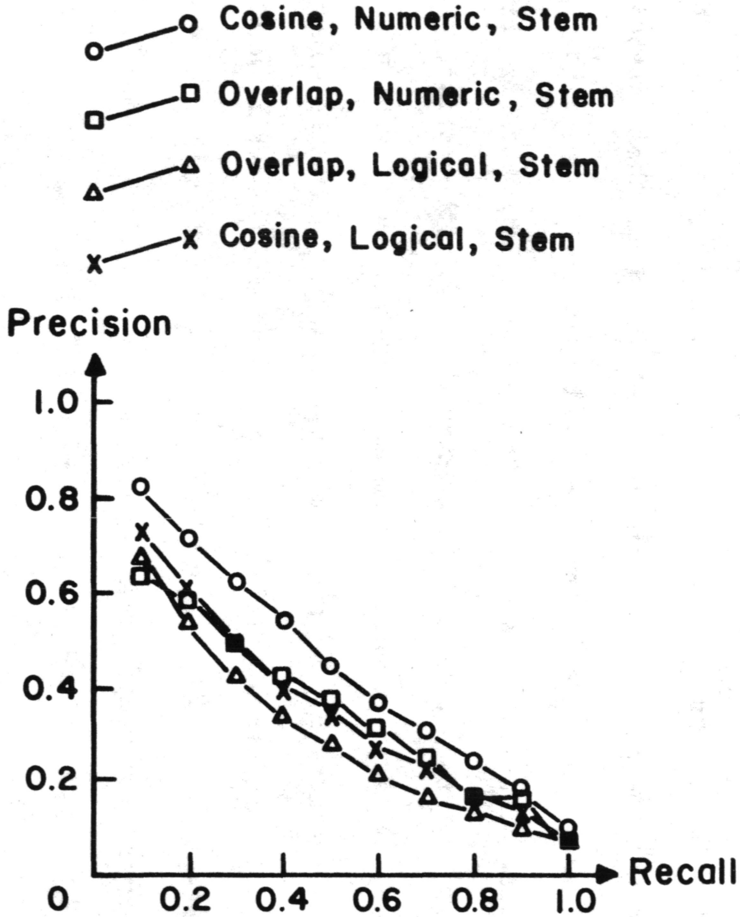
Summary of performance results of matching function comparisons.

Figure 32.

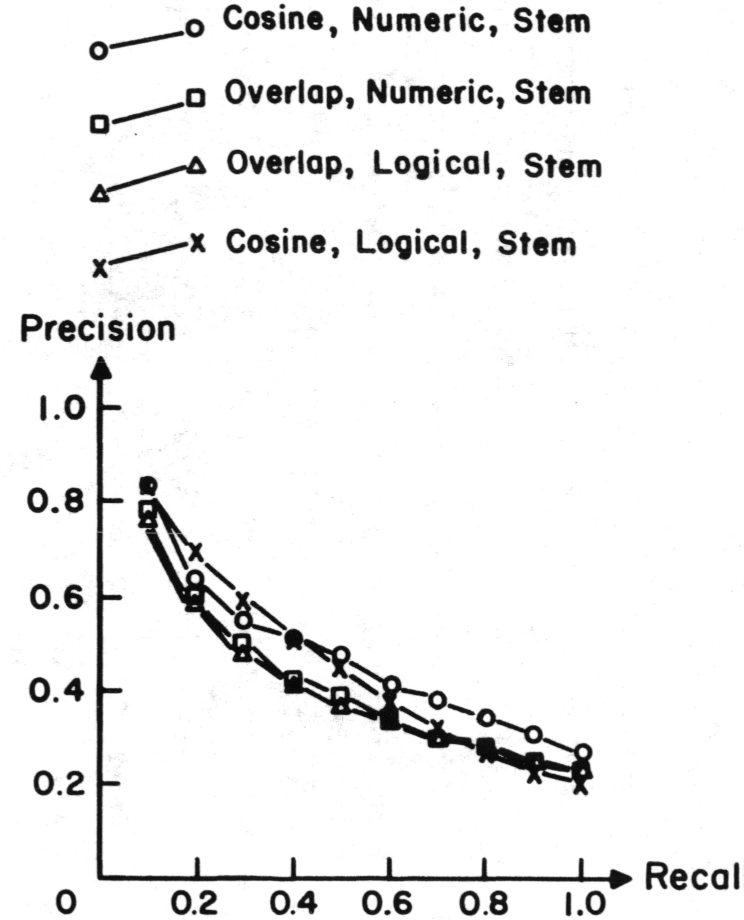
and unimportant request notions so that in some cases otherwise similarly matched relevant and non-relevant documents are correctly separated.

These conclusions are based on the average results observed for three sets of more than 30 requests; since the more detailed results presented show that not all the individual requests perform best with cosine numeric, it must be asked whether cosine numeric should be exclusively used for automatic retrieval systems. It might be possible to develop methods to predict, in advance of the search of a request, the matching function that would perform best, assuming that several optimal functions were provided. Analysis has shown that accurate advance prediction by automatic means is very difficult: the request sets in use have been divided by several criteria, such as length of request measured, number of request concepts, and generality of request, measured by number of documents assessed as relevant in the collection, but no correlation between these criteria and the best matching function has been discovered.

Advance prediction of the most suitable matching function might be obtained from individual users: for example, users with very precisely stated needs who wish to examine only those documents containing all their stated request notions might be best satisfied by an overlap correlation, with or without weighting scheme. Users who supply many possible words to define their general area of interest might be best satisfied with the cosine numeric function. However, two sets of results examined suggest that provision of the best matching function only (cosine numeric) could provide acceptable results for a majority of users. Figure 33 shows precision versus recall graphs for the IRE-3 and Cran-1 collections, comparing the four possible combinations of cosine, overlap, numeric and logical. Figure 34 shows that in the Cran-1 results 66.7% to 78.6% of the requests prefer cosine numeric to any of the other three functions. The IRE-3 collection shows less of a



IRE-3, 34 Requests



CRANFIELD-1, 42 Requests

Performance curves comparing the four matching functions on two collections.

Fig. 33

COLLECTION	INPUT AND DICTIONARY	EVALUATION MEASURE USED TO DETERMINE MERIT	NUMBER AND PERCENTAGE OF INDIVIDUAL REQUESTS COSINE NUMERIC SUPERIOR	COSINE LOGICAL, OVERLAP LOGICAL OR OVERLAP NUMERIC SUPERIOR
IRE-3 34 Requests	Abstract, Stem	Normed Recall	19 (55.9%)	15 (44.1%)
		Normed Precision	23 (67.6%)	11 (32.4%)
CRAN-1 42 Requests	Abstract, Stem	Normed Recall	33 (78.6%	9 (21.4%)
		Normed Precision	28 (66.7%)	14 (33.3%)

Comparisons of individual request merit giving the numbers of requests favoring cosine numeric versus those favoring any of the other three functions, on two collections, according to merit assigned by normalized recall and precision.

Figure 34.

superiority for cosine numeric, but in fact all requests that do better with other functions do so by very small amounts. Even if a perfect advance choice of the best matching function were made for each request, the final result for the 34 requests of IRE-3 and the 42 of Cran-1 would be as given in Figure 35, showing that the final best possible performance is only trivially superior to the use of cosine numeric for all requests.

Studies of other matching functions in the context of the SMART system have been made [2 and Section IV], but have not been subjected to the extensive analysis and evaluation made of those reported here; no correlation coefficient that is superior to cosine has been discovered so far. It is suggested that some studies of a different type are needed. Some quite basic questions about the preferred ordering of documents in a ranked output have not been investigated. For example, using a search request containing five concepts, is it preferable that the matching function places a document with four matching concepts all of low weights in front of one with three matching concepts at high weights? Also, if two documents both match on two equally weighted request concepts, one document having weights of 1 and 3, and the other weights of 2 and 2, should they both be regarded as equally matched with the request (as the numerator of cosine would show), or is the second document perhaps a preferred match?

Questions such as these clearly cannot be answered except in a given retrieval context. A "hand ranking" study is suggested, in which persons would be asked to rank documents in relation to search requests in the order in which they as users would wish to see the documents. The persons doing the ranking would, of course, be given no information as to which documents were actually judged relevant by the requestor, and the experiment could be carried out using several permutations of the variations suggested in Figure 36. The results could be directly evaluated by performance measure-

COLLECTION	INPUT AND DICTIONARY	MATCHING FUNCTION	NORMED. RECALL	NORMED. PRECISION
IRE-3 34 Requests	Abstract, Stem	Cosine Numeric	0.8954	0.6746
	Abstract, Stem	Choice of Best Function for Each Request	0.9014	0.6931
CRAN-1 42 Requests	Abstract, Stem	Cosine Numeric	0.8644	0.6704
	Abstract, Stem	Choice of Best Function for Each Request	0.8720	0.6918

Comparison of Cosine Numeric results with choice of best function for each request done by hindsight, on two collections.

Figure 35.

DATA SUPPLIED		INSTRUCTIONS
REQUEST AND MATCHING CONCEPTS	NON-MATCHING DOCUMENT CONCEPTS	
1. Concept numbers with weights only.	1. No information given about this.	1. Objective is a high precision performance.
2. Concept numbers with weights plus indication of highly weighted request concepts according to importance.	2. Total number of concepts, average weight, and sum of squares of weights given.	2. Objective is a high recall performance.
3. Concepts and weights decoded into actual words.	3. All concepts and weights given and decoded into words.	

Variables for a hand ranking test in which persons would rank documents in relation to search requests in order to develop insight into desirable matching functions.

Figure 36.

ment; in addition, the detailed decisions made by the humans could be examined in order to design new correlation coefficients and weighting schemes.

References

[1] W. T. Brandhorst, Simulation of Boolean Logic Constraints Through the the Use of Term Weights, American Documentation, Vol. 17, July 1966.

[2] Hall and Manning, Student Reports, Computer Science 435, Cornell University, Spring 1966.