AUTOMATIC INDEXING USING TERM DISCRIMINATION
AND TERM PRECISION MEASUREMENTS

G. Salton*, A. Wong*, and C.T. Yu⁺

TR 75-253

July 1975

Computer Science Department
Cornell University
Ithaca, New York 14853

*Department of Computer Science, Cornell University, Ithaca, NY 14853.

⁺Department of Computing Science, University of Alberta, Edmonton, Alberta.

This study was supported in part by the Canadian Research Council and in part by the National Science Foundation under grant GJ-43505.

Automatic Indexing Using Term Discrimination

and Term Precision Measurements

G. Salton*, A. Wong†, and C.T. Yu⁺

Abstract

A variety of abstract automatic indexing models have been developed in recent times in an effort to produce indexing methods that are both effective and usable in practice. Among these are the term discrimination model and the term precision system. These two indexing systems are briefly described and experimental evidence is cited showing that a combination of both theories produces better retrieval performance than either one alone. Appropriate conclusions are reached concerning viable automatic indexing procedures usable in practice.

1. Term Discrimination Model

In information retrieval, the first and probably most crucial operation consists in assigning to the stored documents and to incoming user queries appropriate identifiers capable of representing information content. This operation, known as indexing, is carried out manually by trained indexers, or subject experts in practically all existing, operational retrieval environments.

A great many automatic indexing strategies have been proposed in the literature over the last twenty years, and the available evaluation results indicate that by and large automatic indexing is viable and effective. [1,2] The raw test results have unfortunately proved less than compelling to the managers of operational retrieval systems for several reasons. For one thing,

* Department of Computer Science, Cornell University, Ithaca, New York, 14853.

† Department of Computing Science, University of Alberta, Edmonton, Alberta.

This study was supported in part by the Canadian Research Council and in part by the National Science Foundation under grant GS-43505.

most automatic indexing systems were developed and tested in the laboratory using small sample collections, and no hard evidence could be furnished concerning their applicability in operational user environments. Second, many of the proposed systems were not founded on a recognizable theory, but rather appeared to outside observers to be made up of arbitrary components assembled in an ad-hoc manner to form a complete process.

In recent times, a number of automatic content analysis and indexing theories have, however, been developed based on mathematical or linguistic foundations such as statistical decision theory, information theory, and probabilistic considerations. One such is the term discrimination model which rates a given content indicator, or index term, in accordance with its usefulness as a discriminator among the documents of a collection, and offers in addition a reasonable physical interpretation for the indexing process. [3,4]

Consider, in particular, a collection of documents each identified by a set of content identifiers or index terms. The index term sets, or vectors, for two given documents can be compared and a similarity coefficient can be produced measuring the closeness between the respective documents. The existence of the term sets representing the various documents, and the computation of similarity measures between documents serve to define a document space for the collection. In such a space two documents appear in close proximity when their similarity coefficient is large (that is, when the corresponding term vectors are quite similar); contrariwise, documents exhibiting little similarity are widely separated in the document space.

Since a direct relation exists between a given set of index terms used to characterize a collection of documents and the corresponding space formed by the indexed documents, the problem of choosing an effective term set may be transformed into one of generating a useful document space. In this connection, one may conjecture that a document space which is "bunched up", in the sense that all documents exhibit somewhat similar term sets, is not useful for retrieval since one document cannot then be distinguished from another. A spread out space on the other hand, in which the documents are widely separated from each other may provide an ideal retrieval situation since certain documents — those which appear close to the respective queries — may then be retrieved while others can be rejected.

The discrimination value model leads to a distinction among possible index terms in accordance with their ability to spread out the document space when assigned to the documents of the collection. Specifically, a good discriminator is one which renders the documents less similar to each other when assigned as an index term to the documents of a collection; that is, its assignment decreases the space density. A poor discriminator, on the other hand, increases the space density when assigned as a content identifier.

For each potential index term, a discrimination value (DV) can be computed as the difference in space densities before and after assignment of that term. The greater the difference in space densities, the more the space will have spread after assignment of a given term, and therefore the better that term will function as a discriminator. The space density may be defined as the average pairwise similarity between all documents, or more efficiently as the sum of the similarity coefficients between each document and a dummy "centroid" document defining the center of the space. The space density image is illustrated (in two dimensions) in the example of Fig. 1.

The best discriminators for a given collection will exhibit positive discrimination values because the space spreads when they are assigned; correspondingly the worst discriminators have negative discrimination values. In the middle, one finds a large number of terms whose assignment does not affect the space density; the corresponding discrimination values remain near zero. When the terms are arranged in decreasing order of their discrimination values, the ordering produces a rank for each term, known as the discrimination rank (DR). The highest discrimination value (DV) corresponds to lowest rank (DR = 1).

The term discrimination model can be used as a basis for an indexing policy by relating the discrimination values to the document frequencies of the respective terms. Specifically, it appears that the best discriminators — those with positive discrimination values — exhibit medium document frequencies, that is, they are assigned neither to too many nor to too few documents. Broad terms, assigned to an excessively large number of documents render the documents similar to each other when they are assigned and wind up with

negative discrimination values. Narrow terms, on the other hand, comprising those that appear rarely in the collection produce few changes in the similarity coefficients between the documents. Their assignment therefore does not affect the space density, and the discrimination values remain close to zero.

An appropriate indexing policy then consists in assigning chiefly those indexing units with medium document frequency. This raises the question of what to do with the relatively small number of high-frequency negative discriminators, and with the very large number of low-frequency terms with near-zero discrimination values. The first possibility consists in a simple elimination of these questionable terms — a solution particularly attractive from the point of view of storage efficiency, because the reduced length of the document and query vectors would produce large savings in storage requirements. Unfortunately, the elimination of high-frequency content indicators tends to depress the retrieval performance for high-recall users, that is, for those users expecting to retrieve a high proportion of relevant materials. Similarly, the deletion of the many low-frequency terms may be expected to affect the retrieval precision by making it impossible to reject many extraneous items. If one assumes that the majority of users are not interested in the extremes of the performance ranges (that is, in very high recall combined with low precision, or in very high precision combined with low recall), then the full vocabulary should be used. This may be expected to produce results combining adequate recall with reasonable precision.⁺ [5]

⁺ Two complementary measures are often used for retrieval system evaluation, including the recall defined as the proportion of relevant materials retrieved, and the precision, the proportion of retrieved materials that are relevant. Under normal circumstances, these measures vary inversely, so that high recall generally entails low precision, and vice-versa.

The chosen indexing policy then consists in retaining the full indexing vocabulary while transforming the terms whose frequency characteristics are nonoptimal into indexing units with better assignment frequencies and hence, hopefully, more satisfactory discrimination properties. The following strategy appears then to be appropriate: [6]

- a) Terms with medium document frequency characteristics (with assignment frequencies typically between $n/100$ and $n/10$ for n documents) are used directly for indexing purposes; they comprise the vast majority of the good discriminators.
- b) Terms with excessively high document frequencies comprise the worst discriminators. These terms are too general, or too broad, and their use produces unacceptable precision losses. These terms are transformed into lower frequency units by using them as components of appropriate indexing phrases. When phrases are assigned for purposes of content representation instead of single terms the assignment frequencies are smaller and the discrimination capabilities are enhanced.
- c) Terms with excessively low document frequencies are so rare, or specific that they cannot retrieve an acceptable proportion of the relevant documents; their use depresses the recall performance. Such terms are transformed into higher frequency units by including them in classes of related terms, specified for example by a thesaurus. The assignment frequency of a thesaurus class may be expected to be higher than that of the individual class elements; hence the discrimination values may be more attractive.

Such an indexing policy accounts for three different entities useful for document content representation: single entries used for terms whose discrimination value is already acceptable; term phrases constructed from high-frequency components with negative discrimination values; and term classes formed of low-frequency elements with non-zero discrimination values. A summary of the indexing policy derived from the term discrimination analysis is contained in Fig. 2.

The phrase and thesaurus class assignments have been evaluated under a variety of experimental conditions using document collections in different subject areas such as aerodynamics, medicine, and world affairs. The addition of thesaurus class identifiers, and the replacement of high-frequency terms by phrases was found experimentally to produce improvements in recall and precision ranging from 17 to 39 percent for the various subject fields. [6] Furthermore, under certain well-defined conditions, analytical proofs can be given to show that the phrase and thesaurus transformations must be effective in a retrieval environment (that is, must produce improvement in recall and precision). [7]

2. Term Precision

The previously described term discrimination model is based largely on the occurrence characteristics and frequency distributions of the indexing units across the documents of a collection. Such a model does not, unfortunately, account for the linguistic, or semantic aspects of the texts being processed, and may for this reason be criticized as simplistic and unrealistic. The term precision approach is designed to supply this missing dimension to some extent, though not through the direct use of language processing methods, but rather by utilizing customer opinions concerning the relevance or nonrelevance of certain documents to the queries submitted to the system.

Specifically, the linguistic aspect is introduced by distinguishing among the text words on the basis not only of frequency characteristics, but also of the document type in which the terms occur, that is, whether the terms occur primarily in documents identified as relevant to a given user query, or whether on the contrary they prevail among the nonrelevant. A precision weight attached to each query term is then used as a partial indication of the linguistic characterization of the terms.

As was the case earlier for the discrimination value model, it is possible to show formally that the precision weight system is effective in retrieval, given only certain assumptions concerning the distribution of the vocabulary across the documents of a collection. [8]

Consider, in particular, a given query $Q = (q_1, q_2, \dots, q_m)$. The term precision of query term q_k may be defined as

$$\text{pr}(q_k) = \left(\frac{r_k}{R-r_k} \right) / \left(\frac{h_k}{I-h_k} \right) \quad (1)$$

where r_k and h_k represent the number of documents containing term q_k that are judged to be relevant and nonrelevant, respectively, to query Q ; and R and I represent the number of relevant, and the number of nonrelevant documents for that query. When a term occurs in several queries, its term precision is taken as the average of the precision values obtained for the individual queries. It is clear from expression (1) that high precision implies that the corresponding term is prevalent among the relevant documents in a collection. The reverse obtains for terms that occur mostly among the nonrelevant items.

Given the term precision of the various query terms, a precision weight p_k can now be assigned to each query term q_k , which varies directly with the term precision; that is,

$$p_k > p_e \iff \text{pr}(q_k) \geq \text{pr}(q_e). \quad (2)$$

Correspondingly, the weights originally assigned to the various query terms may be modified in such a way that the original weight w_k of term q_k is multiplied by the precision weight p_k , producing a new weight $w_k * p_k$, $k = 1, \dots, m$.

Various computational procedures suggest themselves for the determination of the precision weights: in some the weight is made to decrease linearly with decreasing term precision; for others the decrease is quite nonlinear. The previously mentioned formal proofs of retrieval effectiveness were constructed for a precision weight system which heavily emphasizes the higher precision terms by giving them relatively large weights and decreasing the weight assignments rapidly to 1 as the term precision diminishes. [8]

The precision weight method emphasizes terms which are concentrated in the relevant, as opposed to the nonrelevant documents. The terms receiving most of the emphasis may then be expected to be the relatively low-frequency narrow terms, since the high-frequency terms tend to occur indiscriminately both in the

relevant and the nonrelevant documents. Like other term weighting systems, the precision weights may be expected to be useful mainly for the enhancement of retrieval precision by making it possible to reject additional extraneous items. However, since the precision weights are based on somewhat subjective criteria, as opposed to the objective measurements which form the basis for the term discrimination model, the two models may be complementary in nature. An effort to confirm this by suitable experimentation is made in the next section.

3. Experimental Results

The precision weighting system is best implemented in a dynamic environment in which a large number of user-system interactions take place. In such a situation it is possible to gather the relevance assessments of documents with respect to queries which form the basis for the determination of term precision values and weights. After precision weights will have been computed for a substantial proportion of the available query terms, the weighting system can be applied routinely to all new incoming queries submitted to the system by the user population.

Since it becomes necessary for present purposes to evaluate the performance of the term precision system, it is obviously not possible to use a single query collection for both the determination of the precision weights and the evaluation of the weighting system proper. Instead, distinct query collections must be used for the two purposes. Two collections of documents representing different subject areas were used for the experiments described in the present study:

- a) 424 documents in aerodynamics, the CRAN collection, available with 155 user queries;
- and b) 425 documents in world affairs originally published in Time magazine, the TIME collection; 83 user queries were available for TIME.

To satisfy the demands of the experimental set-up the query sets were divided somewhat arbitrarily into two pieces: the modifying sets used for precision weight determination consisted of 99 and 46 queries for CRAN and TIME, respectively; the remaining 56 and 37 queries constituting the test sets for the

two collections were used for evaluation purposes. The test sets were so chosen that every test set term occurring in more than one query was also included in the corresponding modifying set. Precision weights were thus computable for all terms included in at least two queries within the test sets.

To render possible the evaluation of an experimental system including both the term discrimination operations and the precision weight assignments, the previously mentioned phrase and thesaurus transformations were applied to the available experimental collections. The following strategy was used for that purpose [6]:

- a) two terms included in a query were incorporated into a phrase if they were separated in the query texts by at most one intervening word; that is, phrases were created from two adjacent terms, or from two terms separated by one intervening word;
- b) thesauruses specifying groupings of low-frequency terms were available for both subject areas (aerodynamics and world affairs):
the thesaurus class identifiers could in each case be used for content representation in addition to (or instead of) the original term identifiers.

Following the phrase creation and the corresponding thesaurus "look-up" operations, phrase and thesaurus class identifiers were added to the query and document vectors whenever the corresponding components were originally present. For phrase assignment, both of the phrase components were expected to be present in the vectors; a thesaurus class was assigned whenever any class element was originally included in the particular vector.

The discrimination value theory indicates that the best content indicators exhibit medium document frequencies. By defining the weight of a phrase as the average of the weights of the respective phrase components, the importance of the many high-frequency phrase components may be appropriately reduced and the corresponding discriminatory power enhanced. For analogous reasons, the weight of a thesaurus class assigned to a vector is taken as the sum of the weights of the class elements present in the original vector. Since many thesaurus class elements exhibit low document frequencies, the importance of these terms is properly raised by the summing of weights implicit in the thesaurus class assignments.

As a last step, those content identifiers which are included in query and document vectors with excessively high document frequencies are eliminated from the respective vectors prior to the query-document comparisons. The cut-off frequency was taken to be 50 for CRAN and 30 for TIME; that is, terms occurring in more than 50 CRAN documents, or more than 30 TIME documents were eliminated. This deletion step agrees with the term discrimination theory in that medium-frequency phrases are effectively introduced as a replacement for the high-frequency vector components which exhibit poor discrimination values.

Appropriate statistical data are shown in Table 1, including the exact number of phrases and thesaurus classes generated by the foregoing procedure for the experimental collections. In view of the relatively small size of the available query sets, it is not surprising that the number of phrases and thesaurus classes assignable by the system is somewhat limited. The bottom line in each portion of Table 1 shows the number of single terms, phrases, and thesaurus classes that are common to query modification and query test sets. These are the only content identifiers that may be subjected to the precision weight treatment, since the weights obtained from the modification sets are later applied to the test sets. Again the figures of Table 1 show that the applicability of the precision weight method is limited when the number of available user queries is small.

In actual practice, the effect of the precision weights is further diminished by the fact that the precision weights decrease rapidly toward 1 as the term precision decreases from its maximum value. For experimental purposes, precision weights are assigned for each query to the ten terms exhibiting the highest term precision for that query. For queries comprising more than ten terms, the extra terms receive a standard weight equal to 1. Table 2 shows the total number of terms affected by the precision weights in the experimental environment. It is obvious from the Table that only relatively few terms can participate in the precision weight experiment. The nonlinear precision weights p_k actually assigned to the respective terms are shown in Table 3. These

values vary with the number of terms present in the query. For seven-term queries, for example, the weights range from 1.6957 for the term exhibiting the highest precision to 1.0000 for the term of lowest precision.⁺

The recall-precision results obtained by processing the test query sets for the two subject areas are shown in Tables 4 and 5 for the CRAN and MED collections, respectively. In each case, retrieval precision values are given at each of ten recall points, averaged over the complete query sets. The top part of each Table uses the term discrimination theory without term precision weighting; precision weights are used for the lower half of the Tables. The percentage figures shown with the average precision values refer to the improvement obtained by the given run over a standard control method consisting of an automatic single-term word stem assignment with term frequency weighting. The control run, shown in the first column of Tables 4(a) and 5(a) respectively, utilizes neither phrases, thesaurus classes, nor precision weighting. For each run, an average percentage improvement figure is also shown at the bottom of the corresponding column.

The following conclusions are immediately apparent from Tables 4 and 5.

- a) The discrimination value model which uses the phrase assignments for high-frequency terms and the thesaurus class transformations for low-frequency terms is highly successful: the runs of Table 4(a) show average performance improvements of 20.3 percent for the phrase assignments and 20.6 percent for the thesaurus transformation using the CRAN collection; the corresponding figures for TIME in Table 5(a) are 13.5 percent and 5.9 percent, respectively. The combined phrases and thesaurus improvement is 27.5 percent for CRAN and 11 percent for TIME.

⁺ Linear precision weight assignments were also tried experimentally. In that case, the weights actually used decreased in even intervals from 1.95 for the term with highest weight to 1.00 for the lowest weighted term. The retrieval results for the linear case were, however, not substantially different from those obtained with the weights of Table 3.

- b) The term precision weight assignments shown in the first column of Tables 4(b) and 5(b) are also successful, producing average improvements of 6 percent for CRAN and 3.2 percent for TIME; these advantages are obtained by assigning term precision weights to less than 300 CRAN terms and only 110 TIME terms (see Table 2).
- c) When the term precision weights are used in conjunction with the phrase and thesaurus assignments additional improvements are apparent, raising the percentage figures to almost 30 percent for CRAN and about 15 percent for TIME. The precision weights do not bring additional advantages for the "phrases only" run; but this may be due to the very small number of precision weights assigned to that case (205 for CRAN and only 27 for TIME).
- d) While the term weight assignments and the phrase transformations tend to operate in the same direction (since they are both designed to improve the retrieval precision), the two processes operate somewhat independently as shown by the combined results in the right-hand column of the Tables.

In view of the relatively restricted test environment, the actual recall and precision values obtained may perhaps be less convincing than one might wish. Still, it is worth noting that for the combined process substantial average precision values are obtainable by fully automatic means, reaching nearly 0.80 for low values of the recall. Overall, the average performance figures are higher for the TIME collection than for CRAN; at the same time, the thesaurus constructed for TIME is not nearly so effective as the CRAN thesaurus. The reason for both phenomena appears to be the same: the TIME collection is more heterogeneous covering the whole world affairs area, with a larger proportion of low-frequency, specific terms. This renders it easier to reject many of the nonrelevant items thus producing the more favorable retrieval precision; at the same time, the heterogeneous vocabulary complicates the grouping of terms into thesaurus classes. The TIME thesaurus might be improved in future tests by using pseudo-classification or other automatic thesaurus construction principles. [9]

The procedures introduced in this study must be more fully tested in operational environments where user relevance assessments obtained over a longer time period may provide reliable data for the term precision measurements. It is hoped that such tests can be carried out soon in an interactive situation under operational conditions.

References

- [1] M.E. Stevens, Automatic Indexing — A State of the Art Report, National Bureau of Standards, NBS Monograph 91, Washington, March 1965.
- [2] G. Salton, Automatic Text Processing, Science, Vol. 168, No. 3929, 17 April 1970, p. 335-343.
- [3] G. Salton, C.S. Yang and C.T. Yu, Contributions to the Theory of Indexing, Information Processing 74, North Holland Publishing Co., Amsterdam, 1974, p. 584-590.
- [4] G. Salton, A Theory of Indexing, Regional Conference Series in Applied Mathematics, No. 18, Society for Industrial and Applied Mathematics, Philadelphia, 1975.
- [5] G. Salton and C.S. Yang, On the Specification of Term Values in Automatic Indexing, Journal of Documentation, Vol. 29, No. 4, December 1973, p. 351-372.
- [6] G. Salton, C.S. Yang and C.T. Yu, A Theory of Term Importance in Automatic Text Analysis, Journal of the ASIS, Vol. 26, No. 1, January-February 1975, p. 33-44.
- [7] C.T. Yu and G. Salton, Effective Information Retrieval Using Term Accuracy, Computer Science Department, Cornell University, Technical Report TR 75-249, Ithaca, New York, June 1975.
- [8] C.T. Yu and G. Salton, Precision Weighting — An Effective Automatic Indexing Method, to be published ACM Journal, Vol. 23, No. 1, January 1976.
- [9] D.M. Jackson, The Construction of Retrieval Environments and Pseudoclassification Based on External Relevance, Information Storage and Retrieval, Vol. 6, 1970, p. 187-219.

- x ———— o distance between original
 document x and centroid o

 x ———— o distance following assignment
 of good discriminator

Document Space Alterations Following
 Assignment of Good Discriminator

Fig. 1

Indexing Strategy Derived From Term Discrimination Model

Fig. 2

		Number of Word Stems	Number of Thesaurus Classes	Number of Phrases
CRAN Collection (424 documents, 155 queries)	Modification Set (99 queries)	372	202	1200
	Test Set (56 queries)	308	181	836
	Common to both sets	210	142	317
TIME Collection (424 documents, 83 queries)	Modification Set (45 queries)	268	198	597
	Test Set (37 queries)	212	146	483
	Common to both sets	105	87	122

Number of Terms, Thesaurus Classes, and
Phrases Generated From CRAN and

TIME Queries

Table 1

	CRAN 424 Collection	TIME 425 Collection
Word Stems and Phrases	205	27
Word Stems and Thesaurus Classes	289	110
Word Stems, Phrases and Thesaurus	289	110

Number of Nondistinct Query Terms Modified
by Precision Weighting
(terms common to modification and test sets
with term precision in top ten)

Table 2

Number of Terms in Query	Rank of Term in Decreasing Term Precision Order									
	1	2	3	4	5	6	7	8	9	10
1	1.0000									
2	1.6667	1.0000								
3	1.6667	1.3333	1.0000							
4	1.7273	1.3636	1.1818	1.0000						
5	1.7000	1.4000	1.2000	1.1000	1.0000					
6	1.7327	1.3784	1.2162	1.1081	1.0541	1.0000				
7	1.6357	1.3714	1.2000	1.1143	1.0571	1.0286	1.0000			
8	1.6356	1.3532	1.1940	1.1045	1.0597	1.0299	1.0149	1.0000		
9	1.6759	1.3538	1.1846	1.1000	1.0538	1.0308	1.0154	1.0077	1.0000	
10	1.6753	1.3453	1.1907	1.0943	1.0511	1.0275	1.0157	1.0079	1.0039	1.0000

Precision Weights Assigned to Query Terms
(nonlinear weight construction method [8])

Table 3

R	Word Stems (term frequency weighting)	Word Stems and Phrases (H terms deleted)	Word Stems and Thesaurus (H terms deleted)	Word Stems, Phrases and Thesaurus (H terms deleted)
.1	.6165	.7081 + 13%	.6569 + 6%	.7304 + 16%
.2	.5291	.6548 + 19%	.5571 + 5%	.6701 + 21%
.3	.4268	.5695 + 25%	.5201 + 18%	.5661 + 25%
.4	.3528	.4683 + 25%	.4578 + 23%	.4745 + 26%
.5	.3118	.4198 + 26%	.3927 + 21%	.4308 + 28%
.6	.2693	.3602 + 25%	.3315 + 19%	.3620 + 28%
.7	.2094	.2610 + 20%	.2777 + 25%	.2820 + 28%
.8	.1556	.1884 + 17%	.2194 + 29%	.2279 + 32%
.9	.1116	.1356 + 18%	.1633 + 32%	.1826 + 39%
1.0	.1097	.1296 + 15%	.1532 + 28%	.1701 + 36%
		+ 20.3%	+ 20.6%	+ 27.5%

a) No Term Precision Weighting

R	Word Stems (term frequency and precision weights)	Word Stems and Phrases (H delete-precision weights added)	Word Stems and Thesaurus (H delete-precision weights added)	Word Stems, Phrases and Thesaurus (precision weights)
.1	.5780 - 7%	.6999 + 12%	.6624 + 7%	.7299 + 16%
.2	.5330 + 1%	.6509 + 19%	.6010 + 12%	.6808 + 22%
.3	.4555 + 6%	.5609 + 24%	.5424 + 21%	.5759 + 26%
.4	.3780 + 7%	.4735 + 26%	.4710 + 25%	.4922 + 28%
.5	.3449 + 10%	.4261 + 27%	.4117 + 24%	.4430 + 30%
.6	.2998 + 10%	.3620 + 26%	.3609 + 25%	.3761 + 28%
.7	.2337 + 10%	.2992 + 19%	.2895 + 28%	.2999 + 30%
.8	.1699 + 8%	.1861 + 14%	.2194 + 29%	.2429 + 36%
.9	.1203 + 7%	.1393 + 19%	.1645 + 32%	.1885 + 41%
1.0	.1186 + 6%	.1315 + 17%	.1542 + 29%	.1761 + 38%
	+ 0.0%	+ 20.3%	+ 23.2%	+ 29.5%

H terms: terms with document frequency exceeding 50

b) With Precision Weighting

Average Precision at Ten Recall Points for
Word Stems, Phrases and Thesaurus
with and without Precision Weights
(CRAN collection, 424 documents 56 test queries)
Table 4

R	Word Stems (term frequency weighting)	Word Stems and Phrases (H terms deleted)	Word Stems and Thesaurus (H terms deleted)	Word Stems, Phrases and Thesaurus (H terms deleted)
.1	.6716	.7518 + 11%	.6938 + 3%	.7240 + 7%
.2	.6557	.7434 + 12%	.6761 + 3%	.7032 + 7%
.3	.6436	.7367 + 13%	.6452 0%	.6871 + 6%
.4	.6253	.7111 + 12%	.6339 + 1%	.6781 + 8%
.5	.6129	.6916 + 12%	.6222 + 3%	.6688 + 8%
.6	.4937	.5790 + 14%	.5649 + 12%	.5910 + 16%
.7	.4570	.5452 + 16%	.5194 + 12%	.5440 + 16%
.8	.4428	.5220 + 14%	.4874 + 8%	.5243 + 14%
.9	.4128	.4856 + 15%	.4486 + 8%	.4811 + 14%
1.0	.4042	.4813 + 16%	.4433 + 9%	.4677 + 14%
		+ 13.5%	+ 5.9%	+ 11.0%

a) No Term Precision Weighting

R	Word Stems (term frequency and precision weights)	Word Stems and Phrases (H delete-precision weights added)	Word Stems and Thesaurus (H delete-precision weights added)	Word Stems, Phrases and Thesaurus (precision weights)
.1	.6922 + 3%	.7439 + 10%	.7081 + 5%	.7847 + 14%
.2	.6721 + 2%	.7356 + 11%	.6797 + 4%	.7605 + 14%
.3	.6342 - 1%	.7229 + 12%	.6535 + 3%	.7217 + 11%
.4	.6148 - 2%	.7137 + 12%	.6535 + 4%	.7121 + 12%
.5	.5935 - 3%	.6931 + 12%	.6496 + 6%	.6880 + 11%
.6	.5213 + 4%	.5842 + 15%	.5708 + 13%	.6161 + 19%
.7	.5005 + 3%	.5527 + 17%	.5293 + 14%	.5589 + 18%
.8	.4801 + 6%	.5273 + 15%	.4933 + 10%	.5365 + 16%
.9	.4436 + 7%	.4928 + 16%	.4579 + 10%	.4916 + 16%
1.0	.4344 + 7%	.4877 + 17%	.4507 + 10%	.4809 + 15%
	+ 3.2%	+ 13.7%	+ 7.9%	+ 14.7%

H terms: terms with document frequency exceeding 30

b) With Term Precision Weighting

Average Precision at Ten Recall Points for
Word Stems Phrases and Thesaurus
with and without Precision Weights
(TIME collection, 425 documents 37 text queries)
Table 5