

Working Paper Series
ISSN 1170-487X

**An Evaluation of Passage-Level
Indexing Strategies for a Technical
Report Archive**

by Michael Williams

Working Paper 98/8
April 1998

© 1998 Michael Williams
Department of Computer Science
The University of Waikato
Private Bag 3105
Hamilton, New Zealand

An Evaluation of Passage-Level Indexing Strategies for a Technical Report Archive

Michael Williams
Department of Computer Science
University of Waikato
Hamilton, New Zealand
mlw@cs.waikato.ac.nz

Abstract

Past research has shown that using evidence from document passages rather than complete documents is an effective way of improving the precision of full-text database searches. However, passage-level indexing has yet to be widely adopted for commercial or online databases.

This paper reports on experiments designed to test the efficacy of passage-level indexing with a particular collection of a full-text online database, the New Zealand Digital Library. Discourse passages and word-window passages are used for the indexing process. Both ranked and Boolean searching are used to test the resulting indexes.

Overlapping window passages are shown to offer the best retrieval performance with both ranked and Boolean queries. Modifications may be necessary to the term weighting methodology in order to ensure optimal ranked query performance.

1. Introduction

As disk storage costs have decreased, it has become increasingly feasible to store complete documents in on-line repositories. Commonly, these on-line document collections are indexed by bibliographic information, or by short abstracts [8,9,18]. When such summary information is not available and cannot readily be extracted from the documents, it may be necessary to index the documents by their complete text [19]. Such a full-text index will provide a very detailed representation of the contents of each document, and so search results will tend to have high recall. However, they may also suffer from low precision.

Full-text retrieval systems tend to favour lengthy and/or discursive documents. Due to the breadth of their vocabulary, these documents are statistically more likely to contain instances of the query terms [16]. However, most current full-text retrieval algorithms cannot determine where in a document the text matches the query, and as a result it is not possible to determine whether a term occurs in the same context in a document as in the query. Consequently, searches may return a high proportion of irrelevant documents which contain scattered, conceptually unrelated instances of the query terms.

Passage-level indexing addresses the problem of long and discursive documents being favoured. The document is broken up into more-or-less equally sized passages during indexing; document passages are then matched against the query, rather than complete documents. Documents may then be returned based on passage-level evidence, or a combination of passage-level and document-level evidence [3].

The choice of passage size and type is somewhat problematic. Documents may be divided into three types of passage: discourse, window, or semantic. Discourse passages are those defined by the author (eg. sentences, paragraphs, and sections). Window passages are passages that contain a specified number of words. As window passage boundaries are arbitrary, it is a common practice to overlap word-windows to reduce the incidence of relevant text being split by unfortunately placed passage boundaries [3]. Semantic passages are bounded by shifts in document content [5]. Whatever choice of passage type is made, a suitable passage size must be chosen. If excessively small passages are used, then overall index performance may suffer, as there will be a high probability of relevant concepts being broken up by passage boundaries. On the other hand, if very large passages are used, then there will be an increased likelihood of irrelevant passages containing sufficient query terms to match the query.

The general efficacy of passage retrieval has been demonstrated in experiments by Salton [12], Callan [3], and others [5, 6, 16], but these experiments do not demonstrate that there is an optimal passage size and type for all collections (in fact, they tend to contradict this idea). The effect of passage-level indexing varies according to the content of the document collection; for instance, discourse passages have been found to work well with highly structured text [16] but may work poorly when document structure is not so explicit [2,3].

There is still considerable uncertainty as to how to choose a good passage size and type for indexing an arbitrary document collection.

This paper reports on experiments designed to test the suitability of passage-level indexing for use with a particular collection: the Computer Science Technical Reports collection of the New Zealand Digital Library (NZDL), an internet-accessible full-text database (<http://www.nzdl.org>). The NZDL's existing indexing system was modified, as opposed to starting from scratch with a new indexing method. The current indexing system is relatively adaptable, and allows for remarkable compression of the index, which were the primary reasons that it was not replaced. However, it also has certain limitations, found with many retrieval systems, which had to be worked around. The current state of the NZDL is described more fully in the following section.

2. Description of the NZDL

The New Zealand Digital Library is an on-line database housed at the University of Waikato, New Zealand. The contents of the database are fully searchable and readable by internet users. The NZDL is comprised of several document collections, of which the Computer Science Technical Reports archive (CSTR) is the largest and most heavily accessed, containing some 40,000 documents at the time of writing, or approximately 2300 MB of ASCII text. The documents in the CSTR vary in size, but generally are long (approximately 25 pages on average)[18]. As the CSTR is the largest collection and sees daily use by the international computer science research community, it was chosen as the testbed for this project.

Document indexing is carried out by a full-text indexing program known as MG. MG is most noteworthy in its highly efficient use of disk space. The full-text inverted file indexes created by MG are stored in compressed form, with the index being only about 5% of the size of the original text collection. This degree of compression is especially impressive when it is considered that stopwords are included in the index [17]. Otherwise, the indexing and searching of the NZDL is unremarkable. Term weighting is accomplished with the *tf.idf* (term frequency.inverse document frequency) function, which assigns high weights to terms that occur frequently within particular documents but rarely within the collection as a whole [10]. The standard cosine measure is used for document ranking [7]. Stemming and case-folding are optional ("on" by default). Phrase searching is accomplished by post-retrieval scanning of query results.

NZDL users may search for documents with either ranked or Boolean queries. Currently, only pure ranked and pure Boolean searching are supported. There is no provision for the use of "Boolean-style" operators in ranked queries, or author/title searching. In preliminary examinations of the query logs, no clear user preference for ranked or Boolean queries is evident; about two-thirds of users employ the default query mode, whether it is ranked or Boolean [4].

At the present time, a form of passage-level indexing is employed by the NZDL. The CSTR collection may be searched at page level, which means that the search terms must appear on the same page. A separate page-level index, in addition to the full-document index, is required. The effectiveness of the NZDL's page-level index in improving search precision has yet to be evaluated.

A major drawback of the NZDL's search engine is that it is difficult to combine multiple sources of evidence in determining whether a document should be retrieved. This problem is not unique to the NZDL search engine, but rather is shared by many operational systems. The documents can only be divided into passages during the initial indexing process, not "on the fly" after the query becomes available. In order to combine information from various levels of a document, it would be necessary to maintain multiple indexes and issue the query to all of them; the costs in disk space would be high, and searching would be relatively slow. Previous experiments have focused on more flexible (but less standard) IR systems such as the inference network-based INQUERY, which are effective at combining evidence from multiple document levels [3,15]. These experiments will investigate whether passage-level indexing can be effective with systems where information from only a single passage level can be used in document retrieval, using MG as a test case.

3. Methodology

There were three major goals to this project:

- To determine whether passage-level indexing could be used to improve the performance of CSTR search results.
- To determine the degree of compromise needed to make passage-level indexing effective for both ranked and Boolean querying; for example, will two different indexes be required?

- To determine whether passage-level indexing is worthwhile when document-level and passage-level evidence cannot be combined in the retrieval process.

A description of the experimental methodology by which these goals were to be accomplished is given in the following sections.

3.1 The Creation of the Sample Database and Set of Test Queries

Resource constraints contraindicated the use of the complete CSTR collection as a test database. Instead, a smaller database was built up by randomly sampling files from the CSTR archive. By this method, it was hoped to create a sample database that would be as representative of the CSTR as possible. The final test database was small, containing only 400 documents and being 7.8 MB in size. It would have been desirable to use a larger database. However, the NPL collection is only 3 MB in size and has been successfully used in experiments of similar type [3]. As the documents were long and generally discursive, a single document could match a number of queries.

A set of test queries was sampled in a pseudo-random manner from the CSTR query logs, with those queries that proved difficult to interpret for meaning being rejected. A total of 58 Boolean and 58 ranked queries were selected. The major drawback of this method of query selection is that it is impossible to determine *a priori* how many relevant documents exist in the sample database for each query. A substantial minority of queries had only one or two relevant documents, which may have produced some degree of instability in the results.

3.2 Passage Sizes

Both discourse passages and window passages were tested, as it was not apparent in advance which type would perform better. Only one discourse passage index was created, as paragraphs are the only discourse passages that occur consistently in CSTR documents. Simple bounding heuristics were used in order to reduce the variation in paragraph size. Short paragraphs (<50 words) were appended onto the start of the following paragraph. Originally, it was also intended to divide large paragraphs. In practice this was found to slightly worsen search results, presumably because the paragraphs were divided arbitrarily, rather than by content. Since very large paragraphs are not common in the CSTR, the heuristic for paragraph splitting was removed.

Page-level indexing is already supported by the NZDL, and is experimentally tested in this paper. Strictly speaking, pages are neither discourse passages nor window passages, as the number of words on a page may vary. Page boundaries are generally poorly placed, occurring even in the middle of sentences. The page-level index created for this experiment places the passage boundaries at the first end-of-sentence after a page break.

Window passages are more flexible than discourse passages, in that they can be of any size, and they can be overlapped in order to reduce the likelihood of relevant text being split by a passage boundary. In fact, if a good window size is chosen, window passages may perform far better than discourse passages [3]. The difficulty lies in choosing a window size that will work well with the collection. Only a few window sizes could be tested in this experiment; the process of manually assessing the results for recall and precision proved too time-consuming to permit a larger number of indexes to be thoroughly tested.

A passage size of 200-250 words has been proposed by Callan [3] as being the most likely to produce good results with an arbitrary collection. However, this result applies only when ranked queries and overlapping windows are in use. It seems intuitively reasonable that a larger window size would be more effective when Boolean queries are in use. Boolean queries are more stringent than ranked queries in that all of the query terms must be present for a document or passage to match the query. As a result, Boolean queries tend to favour precision over recall [1]. If excessively small window passages were used in a Boolean query environment, then a decline in recall may be the result.

Three window sizes were tested: 250 words, 500 words and 1,000 words. It was hoped that the more finely-grained 250 word-window index would produce good results with ranked queries, whereas one of the less finely-grained 500 and 1,000 word-window index would prove to be more suitable for use with Boolean queries.

It was uncertain as to whether the cost of overlapping window passages would be justified by the resultant improvement in recall. In order to create overlapping word-window indexes with the NZDL, it is necessary to divide the documents into overlapping passages at the time of indexing, and then index the

passages as if they were separate entities. The resultant overlapping window index will be nearly twice as large as the corresponding non-overlapping index. To determine whether this increase in index size is justifiable, overlapping 250, 500 and 1,000 word-window indexes were created so that their performance could be compared with that of their non-overlapping counterparts.

To summarise, the indexes used in this series of experiments were:

- Full document (control)
- Paragraph-level
- Page-level
- Window: 250, 500 and 1,000 words
- Overlapping window: 250, 500 and 1,000 words.

3.3 Evaluation

A difficulty for the IR researcher is in knowing how many retrieved documents per query to evaluate. If the results set of a particular query contains hundreds of documents, it is generally not feasible to assess the entire results set for relevance.

The Boolean results sets created in these experiments were, on average, quite small (<20 documents). It was therefore feasible to evaluate the complete results sets returned by Boolean queries. The ranked queries, on the other hand, generally returned a very large number of documents, sometimes virtually every document in the sample collection. For this reason, it was decided to use a fixed rank cutoff for the ranked queries and restrict evaluation to those ranks. Only the twenty most highly ranked documents returned by each ranked query were included in the calculations of recall and precision. With the passage-level indexes, the top twenty most highly ranked documents were identified by their most highly ranked passages.

3.4 Relevance Judgements

Even with the small test set used, it would be tedious to assemble exhaustive relevance judgements for every query. In theory, each of the 400 documents in the test set would have to be read 58 times to determine the complete set of relevant documents just for the Boolean queries. Fortunately, it is not necessary to have an *absolute* measure of recall in order to compare the performance of the various indexes. The difference between the indexes can be indicated by their *relative* recall performance.

With Boolean queries, any documents retrieved by the passage-level indexes in response to a query will also be retrieved by the document-level index; 100% relative recall for a Boolean query may therefore be defined as the set of relevant documents returned by the full-document index. The sole exception to this rule is when the query contains one or more instances of the NOT operator, in which case a document that does not match the query may contain passages that do match the query (in this case, all retrieved documents will have to be evaluated).

Relevance judgements for the ranked queries were more difficult to assemble. A "pooling" method was used, by which the set of known relevant documents is built up by combining the results of several different search techniques [14]. In this case, the set of relevant documents for each query was created by examining the twenty most highly ranked documents returned by each index, and assessing any previously unevaluated documents for relevance. In addition, some previously undiscovered relevant documents were located by issuing Boolean queries containing some of the ranked query terms (the latter approach is feasible only with a very small test database).

4. Results

The results of the experiments are presented in the following sections. Results for the Boolean queries are presented first, followed by the results for the ranked queries.

4.1 Boolean Query Results

Although the standard recall-precision curve may be used for the display of Boolean query results [13,14], it is debatable whether it is an appropriate way of presenting unranked results. A simple approach has been adopted here by which the total number of relevant and irrelevant documents returned by the entire query set has been used as a measure of index performance. The order in which the documents are returned is ignored. The Boolean query results for the non-overlapping indexes are summarised in Table 1.

Table 1: Recall and Precision for Non-Overlapping indexes (Boolean Queries)

(a) Relevant documents returned and recall for each index

DOC	PARA	PAGE	250	500	1000
179 (95.21%)	81 (43.09%)	154 (81.91%)	135 (71.81%)	158(84.04%)	174 (92.55%)

(b) Irrelevant documents returned and precision for each index

DOC	PARA	PAGE	250	500	1000
838 (17.60%)	28 (74.31%)	83 (64.98%)	56 (70.68%)	90 (63.71%)	179 (49.29%)

The expectation was that there would be a positive relationship between passage size and recall and a corresponding negative relationship between passage size and precision [10]. The results clearly bear this out. In particular, the document-level index has a high level of recall, but very low precision, presumably due to the predominance of long documents with extensive vocabularies (as discussed in section 1). Precision improves and recall falls away as passage size decreases. Fortunately, the improvement in precision proceeds faster than the decline in recall.

Paragraphs and pages contain a varying number of words, so it is difficult to quantify the relationship between passage size and recall performance for the paragraph- and page-level indexes. Generally, paragraphs in the CSTR collection are less than 250 words. Averaged over the collection, pages in the CSTR collection contain slightly less than 500 words (although the number of words per page may vary greatly, as the last page of a document is usually short). A comparison of the result for the 500 word-window index and the page-level index shows that the relative heterogeneity of the page-level index's passages does not seem to have affected recall or precision as much as might be expected.

The results for the overlapping window indexes are presented in Table 2. The difference in performance between the overlapping and non-overlapping indexes is also given. The figure in parentheses next to the total number of (ir)relevant documents returned is the number of (ir)relevant documents returned by the overlapping index that were not returned by the corresponding non-overlapping index.

Table 2. Recall and Precision for Overlapping Indexes (Boolean Queries)

a. Relevant documents retrieved and recall for each index

	250	500	1,000
TOTAL:	140 (+5)	173 (+15)	178 (+4)
Recall:	74.47%	92.02%	94.68%
Change:	+3.70%	+9.50%	+2.30%

b. Irrelevant documents retrieved and precision for each index

	250	500	1,000
TOTAL:	65 (+9)	109 (+19)	189 (+10)
Precision:	68.29%	61.35%	48.50%
Change:	-3.38%	-3.70%	-1.60%

Each of the overlapping indexes performs better than its non-overlapping counterpart in terms of recall. Whereas the improvement for the 250 and 1,000 word overlapping window indexes are relatively slight, the improvement for the 500 word overlapping window index is significant. Unfortunately, the overlapping window indexes also have lower precision than their non-overlapping counterparts, as a greater number of additional irrelevant documents were retrieved than additional relevant documents. However, the decline in precision is modest for all of the overlapping window indexes.

Whether these results can be taken as evidence supporting the practice of overlapping passages depends on whether recall and precision are the only considerations, or whether additional costs (disk storage, search time, etc.) are taken into account. If only recall and precision are considered, then these results would tend to support the conclusion that the overlapping of passages is worthwhile, as the gain in recall is higher than the loss in precision for all of the indexes. However, when it is factored in that the overlapping window indexes created for this project were by necessity twice as large as their non-overlapping counterparts, then the practice becomes harder to justify.

It is apparent that when passage-level indexing is used in a Boolean query environment there will have to be some compromise between recall and precision in determining the optimal passage size. The relative importance of recall and performance in determining the quality of search results is hard to define: it may vary from user to user, and even from query to query, and cannot be definitively quantified. A variety of possible interpretations of overall index performance, with index performance being measured as a weighted combination of recall and precision, are given in Table 3.

The 500 word overlapping window index appears to offer the best compromise between recall and precision. Only when precision is heavily favoured over recall (70% precision, 30% recall) is it outperformed by the 250 word non-overlapping window index. Similarly, the 500 word non-overlapping window index offers the best overall performance of the non-overlapping indexes for most of the range covered by the table.

Table 3. Index Performance as a Weighted Combination of Recall and Precision.

INDEX	Percentage weighting (Recall / Precision)				
	30/70	40/60	50/50	60/40	70 / 30
Document	40.88	48.64	56.41	64.17	71.93
Paragraph	64.94	61.82	58.70	55.58	52.46
Page	70.06	71.75	73.45	75.14	76.83
250 word	71.02	71.13	71.25	71.35	71.47
500 word	69.81	71.84	73.88	75.91	77.94
1,000 word	62.26	66.59	70.92	75.25	79.57
250 word overlapping	70.14	70.76	71.38	72.00	72.62
500 word overlapping	70.55	73.62	76.69	79.75	82.82
1,000 word overlapping	62.35	66.97	71.59	76.21	80.82

4.2 Ranked Query Results

Table 4. Non-Overlapping Index Performance on Ranked Queries.

Recall	Precision (% change) - 58 queries					
	Document	Paragraph	Page	250	500	1000
0	55.07	52.60 (-4.49)	53.70 (-2.49)	58.36 (+5.97)	55.74 (+1.22)	50.33 (-8.12)
10	54.50	52.27 (-4.09)	52.78 (-3.16)	56.52 (+3.71)	53.85 (-1.19)	49.34 (-9.47)
20	51.81	47.27 (-8.76)	49.11 (-5.21)	55.93 (+7.95)	52.80 (+1.91)	47.23 (-8.84)
30	46.94	40.95 (-12.76)	45.78 (-2.47)	48.21 (+2.71)	46.30 (-1.36)	42.22 (-10.10)
40	43.08	36.77 (-14.65)	41.79 (-2.99)	45.81 (+6.34)	43.85 (+1.79)	40.06 (-7.01)
50	41.11	31.61 (-23.11)	38.52 (-6.30)	41.94 (+2.02)	41.17 (-0.14)	36.84 (-10.39)
60	31.66	24.11 (-23.85)	31.76 (+0.32)	29.83 (-5.78)	29.37 (-7.23)	27.67 (-12.60)
70	29.48	19.32 (-34.46)	25.10 (-14.86)	26.60 (-9.77)	23.32 (-20.90)	21.86 (-25.85)
80	25.07	17.43 (-30.47)	22.50 (-10.25)	21.62 (-13.76)	19.88 (-20.70)	19.65 (-21.62)
90	23.66	16.26 (-31.28)	19.26 (-18.60)	19.52 (-17.50)	19.88 (-15.98)	18.47 (-21.94)
100	23.66	16.26 (-31.28)	19.26 (-18.60)	19.52 (-17.50)	18.24 (-22.91)	18.47 (-21.94)
avg:	38.73	32.25 (-16.73)	36.32 (-6.22)	38.53 (-0.52)	36.76 (-5.09)	33.83 (-12.65)

The ranked query results for the non-overlapping indexes are presented in Table 4. In contrast to the largely predictable Boolean query results, the ranked query results were unexpected. It is apparent that all of the passage-level indexes have performed poorly on the ranked retrieval task. The paragraph-level index is the poorest overall, suffering from a 16.73% decline in overall average precision, and performing considerably worse than the document-level index at all recall levels. Although the small passage size of the paragraph-level index may have contributed to this result, it is noteworthy that the 1,000 word-window index is the second worst overall, with reduced precision at all recall levels and an overall average precision 12.65% less

than that of the document-level index. Only the 250 word-window index performs reasonably. Although its overall average precision is slightly lower than that of the document-level index, it does perform better than the document-level index at the lower recall levels (50% and below). It is difficult to determine the exact relationship between passage size and index performance from these results.

The poor performance of the non-overlapping indexes might be at least partially due to relevant concepts being split by passage boundaries, in which case the overlapping window indexes might be expected to perform better. The results for the overlapping indexes are given in Table 5. The results for the document-level index have been repeated in this table to give a point of comparison.

Table 5. Overlapping Index Performance on Ranked Queries.

Recall	Precision (% change) - 58 queries			
	Document	250	500	1000
0	55.07	58.36 (+5.97)	58.23 (+5.74)	55.61 (+0.98)
10	54.50	57.21 (+4.97)	56.19 (+3.10)	53.51 (-1.82)
20	51.81	53.21 (+2.70)	53.34 (+2.95)	48.80 (-5.81)
30	46.94	49.87 (+6.24)	49.49 (+5.43)	44.27 (-5.69)
40	43.08	44.14 (+2.46)	43.77 (+1.60)	42.74 (-0.80)
50	41.11	42.41 (+3.16)	42.35 (+3.02)	40.53 (-1.41)
60	31.66	35.03 (+10.64)	35.25 (+11.34)	39.50 (+24.76)
70	29.48	30.69 (+4.10)	27.40 (-7.10)	31.19 (+5.80)
80	25.07	24.23 (-3.35)	22.68 (-9.53)	23.20 (-7.46)
90	23.66	19.81 (-16.27)	20.54 (-13.19)	17.45 (-26.25)
100	23.66	19.81 (-16.27)	19.26 (-18.60)	17.15 (-27.51)
avg:	38.73	39.52 (+2.04)	38.95 (+0.57)	37.63 (-2.84)

As with the Boolean queries, overlapping the window passages leads to improved index performance. The 250, 500 and 1,000 word overlapping window indexes perform 2.5%, 5.9% and 11.23% better than their non-overlapping counterparts, respectively. The 250 and 500 word overlapping window indexes have slightly better overall average precision than the document-level index (+2.04% and +0.57% respectively). However, the 1,000 word overlapping window index performs slightly worse than the document-level index (a 2.84% decline in overall average precision).

The results for the ranked queries are, on first analysis, highly counterintuitive. It has been assumed that passage-level index performance would be poorest when the passages are too small to contain many of the query terms or when the passages are too large to ensure the proximity of the query terms. The optimal passage size for a particular collection should lie somewhere between these two extremes. Callan's results suggest that a graph of index performance on ranked queries vs. passage size should have a single maxima [3]. However, the results for these indexes exhibit several local performance maxima (at a passage size of 250 words and when complete documents are treated as "passages").

A possible contributing factor to the unusual index performance results lies in the term weighting methodology used. It has already been mentioned that the MG system employs the *tf.idf* function to assign term weights. During the indexing process, the term weights were assigned *after* the documents were partitioned into passages; in other words, passages were treated as if they were separate documents in determining the term weights. The effect of this may be to reduce the resolving power of certain terms (the degree to which the terms may distinguish relevant from irrelevant documents [13]). For instance, consider two terms *i* and *j*, of which there are 50 instances of each in the collection. Term *i* appears in only one document, but there are numerous instances of *i* in the document. Term *j* appears in a number of documents, but there are only one or two instances of *j* per document. Term *i* will then have a much higher resolving power than *j*. However, if the collection is broken up into small passages before term weights are applied, it is likely that there will be only one or two instances of term *i* per passage. Since the indexing system treats the passages as separate entities, it will regard terms *i* and *j* as essentially similar in resolving power; both occur once or twice in roughly the same number of passages. As the most highly-weighted terms will be those that occur repeatedly within particular documents, then the effect of passage-level indexing will be to distribute these terms throughout multiple passages, consequently reducing their weight. The effect of this tendency towards homogenisation of the term weights is likely to have been detrimental to the performance of the passage-level indexes. Alternative methods for assigning term weights may produce better results.

5. Conclusions

This paper describes experiments designed to test the performance of various passage-level indexes, with a view to determining their suitability for indexing a collection of relatively large documents: specifically, the Computer Science Technical Reports collection of the NZDL.

It was identified that passage-level indexing is an effective methodology for use in a Boolean query environment and can substantially improve precision, at the cost of a moderate decline in recall. The best passage size for a particular collection will depend on the relative importance placed on the precision and recall of search results. Of the indexes tested, the 500 word overlapping window index seems to offer the best combination of recall and precision for Boolean queries. The practice of overlapping window passages was shown to improve the retrieval performance of the word-window indexes for both ranked and Boolean queries.

The ranked query results are more difficult to interpret than the Boolean query results, as they are influenced by a number of factors (passage size, the weighting and ranking methods used) and it is difficult to isolate the effects of each. It is, however, possible to make inferences based on them, and suggest future work that would clarify the issue.

The ranked query results can be taken as tentative evidence that different passage sizes are required to produce optimal results for ranked and Boolean queries. Ranked queries perform better when relatively small passages (250 words) are used, whereas Boolean queries perform better when larger passages (500 words) are used. A firm determination cannot be given, though, until it is decided how term weights should best be assigned. The results suggest that dividing the documents up into passages before assigning term weights is likely to produce poor results. It would be relatively easy to modify the MG system to partition the documents after term weights had been assigned, or to use a weighting strategy other than the *tf.idf* formula. Further research on this project could focus on determining the best method for assigning term weights during the initial indexing process.

- [1] D.C. Blair and M.E. Maron. An evaluation of retrieval effectiveness for a full-text document retrieval system. *Communications of the ACM*, 28(3), pp. 290-299, 1985.
- [2] C. Buckley, J. Allan and G. Salton. Automatic routing and ad-hoc retrieval using SMART: TREC-2. In *Proceedings of the Second Text REtrieval Conference (TREC-2)*, pp. 45-56, 1994.
- [3] J.P. Callan. Passage-level evidence in document retrieval. In *Proceedings of the 17th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, pp. 302-309, 1994.
- [4] S.J. Cunningham. Personal communication.
- [5] M.A. Hearst and C. Plaunt. Subtopic structuring for full-length document access. In *Proceedings of the 16th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, pp. 59-68, 1993.
- [6] M. Kaszkiel and J. Zobel. 'Passage Retrieval Revisited. In *Proceedings of the Twentieth International ACM-SIGIR Conference on Research and Development in Information Retrieval (SIGIR '97)*, pp 178-185, 1997.
- [7] NZDL On-line Documentation. See <http://www.nzdl.org>, "help" option.
- [8] C.D. Paice. *Information Retrieval and the Computer*. MacDonald and Jane's, London, 1977.
- [9] D. Ridley. *Online Searching: A Scientist's Perspective*. John Wiley and Sons, Chichester, England, 1996.
- [10] J.S. Ro. An evaluation of the applicability of ranking algorithms to improve the effectiveness of full-text retrieval. I. On the effectiveness of full-text retrieval. *Journal of the American Society for Information Science*, 39(2), pp. 73-78, 1988.

- [11] G. Salton. *Automatic Text Processing: the Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley, Reading, MA, 1989.
- [12] G. Salton, J. Allan and C. Buckley. Approaches to passage retrieval in full text information systems. In *Proceedings of the 16th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, pp. 49-58, 1993.
- [13] G. Salton and M.J. McGill. *Introduction to Modern Information Retrieval*, McGraw-Hill, New York, 1983.
- [14] H. Turtle. Natural language vs. Boolean query evaluation: A comparison of retrieval performance. In *Proceedings of the 17th Annual ACM/SIGIR Conference on Research and Development in Information Retrieval*, pp. 212-220, 1994.
- [15] H. Turtle and W.B. Croft. Evaluation of an inference network-based retrieval model. *ACM Transactions in Information Systems*, 9(3), pp. 187-222, 1991.
- [16] R. Wilkinson. Effective retrieval of structured documents. In *Proceedings of the 17th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, pp. 311-317, 1994.
- [17] I.H. Witten, A. Moffat and T.C. Bell. *Managing Gigabytes: Compressing and Indexing Documents and Images*. Van Nostrand Reinhold, New York, 1994.
- [18] I.H. Witten, C.G. Nevill-Manning, and S.J. Cunningham. Digital libraries based on full-text retrieval. *Proc Webnet '96* (invited paper), 1996.
- [19] I.H. Witten, C.G. Nevill-Manning, and S.J. Cunningham. Building a digital library for computer science research: Technical issues. In *Proceedings of the Australasian Computer Science Conference*, pp. 534-542, 1996.