

UNIVERSITY OF MINNESOTA

This is to certify that I have examined this copy of a master's thesis by

Qingyan Chen

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the
final examining committee have been made.

Carolyn J. Crouch
Name of Faculty Adviser(s)

Signature of Faculty Adviser(s)

Date

GRADUATE SCHOOL

Improving the Retrieval Effectiveness of Very Short Queries

A THESIS
SUBMITTED
TO THE FACULTY OF THE GRADUATE SCHEOOL
OF THE UNIVERSITY OF MINNESOTA
BY

QINGYAN CHEN

IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

SEPTEMBER, 1999

© Qingyan Chen 1999

Acknowledgements

This thesis would not have been possible without the support of many people. With much pleasure, I take the opportunity to thank all the people who have helped me through the course of graduate study.

First I would like to thank my advisor, Professor Carolyn Crouch, for her invaluable guidance, advice, and encouragement. Carolyn brought me into the world of Information Retrieval, and she taught me how to do research in the field. Her pleasant-natured style always makes our meetings and research stimulating, enjoyable and fruitful. I can never thank her enough for all I have learned from her.

I would also like to thank Professor Donald Crouch for providing help and guidance in our research. His insightful discussions with us and persistent questioning have been and will always be inspiring to me. His valuable suggestions have made our work straightforward yet successful.

I thank Professor Jonathan Maps for all the good things I have learned from him. He was the advisor for my graduate study in physics during my first two years at UMD, and I owe him for all the good habits I have developed in scientific research and experimentation. His intelligence, assiduousness and carefulness have been and will always be encouraging me to do my best work.

I would like to express my gratitude for my colleague, David Wicklund, for his creative ideas in our discussions. Many thanks to our system administrator, Jim Luttinen, for the good facilities and timely problem solving. Also thanks to Amit Singhal for his quick response on my questions about the Smart system and the TREC experiments.

Thanks to all the people at UMD who have made my graduate study pleasant and memorable.

TABLE OF CONTENTS

- 1 INTRODUCTION..... 1
 - 1.1 INFORMATION RETRIEVAL..... 1
 - 1.2 VECTOR SPACE MODEL AND THE SMART SYSTEM..... 2
 - 1.3 TREC COLLECTION AND EVALUATION..... 5
- 2 MOTIVATION..... 8
- 3 ALGORITHMS..... 11
- 4 EXPERIMENTS AND RESULTS..... 16
 - 4.1 TEST COLLECTIONS AND TEST QUERIES..... 16
 - 4.2 EXPERIMENTAL PARAMETERS..... 18
 - 4.3 EXPERIMENTAL RESULTS..... 20
 - 4.3.1 *WSJ Collection vs. TREC Query 1-50 (title only)*..... 20
 - 4.3.2 *TREC-3 ad-hoc Task (TREC disks 1 & 2 vs. Query 151-200)*..... 23
 - 4.3.3 *TREC-6 ad-hoc Task (TREC disks 4 & 5 vs. Query 301-350)*..... 24
 - 4.4 QUERY DRIFT..... 25
- 5 CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH..... 29
 - 5.1 CONCLUSIONS..... 29
 - 5.2 SUGGESTIONS FOR FUTURE RESEARCH..... 29
- 6 REFERENCES..... 31
- 7 APPENDIX I..... 33
- 8 APPENDIX II..... 41
- 9 APPENDIX III..... 49

LIST OF TABLES

Table 1-1: Most commonly used term weights in Smart 11.0.....	4
Table 1-2: Document statistics for TREC disk 1 & 2. Words are strings of alphanumeric characters. No stopwords were removed and no stemming was performed.....	6
Table 3-1: Term ordering for TREC Query 1.....	13
Table 3-2: Term ordering for TREC Query 50.....	15
Table 4-1: Query length statistics.....	17
Table 7-1: P@20 before and after reranking for different settings (WSJ collection vs. Query 1-50).....	33
Table 7-2: P@30 before and after reranking for different settings (WSJ collection vs. Query 1-50).....	34
Table 7-3: non-interpolated average precision (varying T) (WSJ collection vs. Query 1-50).....	35
Table 7-4: P@20 (varying T) (WSJ collection vs. Query 1-50).....	36
Table 7-5: non-interpolated average precision (varying β) (WSJ collection vs. Query 1-50 baseline pseudo-feedback runs).....	37
Table 7-6: non-interpolated average precision (varying β) (WSJ collection vs. Query 1-50).....	38
Table 7-7: P@20 (varying β) (WSJ collection vs. Query 1-50 baseline pseudo-feedback runs).....	39
Table 7-8: P@20 (varying β) (WSJ collection vs. Query 1-50).....	40
Table 8-1: P@20 before and after reranking for different settings (TREC-3 ad-hoc task).....	41
Table 8-2: P@30 before and after reranking for different settings (TREC-3 ad-hoc task).....	42
Table 8-3: non-interpolated average precision (varying T) (TREC-3 ad-hoc task).....	43
Table 8-4: P@20 (varying T) (TREC-3 ad-hoc task).....	44
Table 8-5: non-interpolated average precision (varying β) (TREC-3 ad-hoc task baseline pseudo-feedback runs).....	45
Table 8-6: non-interpolated average precision (varying β) (TREC-3 ad-hoc task).....	46
Table 8-7: P@20 (varying β) (TREC-3 ad-hoc task baseline pseudo-feedback runs).....	47
Table 8-8: P@20 (varying β) (TREC-3 ad-hoc task).....	48
Table 9-1: P@20 before and after reranking for different settings (TREC-6 ad-hoc task).....	49
Table 9-2: P@30 before and after reranking for different settings (TREC-6 ad-hoc task).....	50
Table 9-3: non-interpolated average precision (varying T) (TREC-6 ad-hoc task).....	51
Table 9-4: P@20 (varying T) (TREC-6 ad-hoc task).....	52
Table 9-5: non-interpolated average precision (varying β) (TREC-6 ad-hoc task baseline pseudo-feedback runs).....	53
Table 9-6: non-interpolated average precision (varying β) (TREC-6 ad-hoc task).....	54
Table 9-7: P@20 (varying β) (TREC-6 ad-hoc task baseline pseudo-feedback runs).....	55
Table 9-8: P@20 (varying β) (TREC-6 ad-hoc task).....	56

LIST OF FIGURES

Figure 4-1: A sample TREC topic statement.....	17
Figure 4-2: Effect of reranking on precision at 20 docs.....	26
Figure 4-3: Query Drift in terms of average precision.	27
Figure 4-4: Query Drift in terms of P@20.....	28

1 Introduction

The emergence of the World Wide Web has necessitated the availability of search capabilities for a user. One needs to find useful information in the enormous amount of information available on the web. Search engines based on information retrieval techniques serve this need and have become popular. It is a continual effort in the field of information retrieval to enhance the search effectiveness of such systems.

The typical users of web search engines formulate very short queries; they are unable to construct long and carefully stated queries [Magennis 1997]. In a study of the user queries submitted to Excite (a major Internet search service), it was found that, on the average, a web query contained only 2.35 terms [Jansen 1998]. Special techniques are needed for such short queries. This work suggests useful methods to improve the retrieval effectiveness of very short queries on large text collections.

1.1 Information Retrieval

Since most information in the modern world is text-based, text retrieval is the basis of information retrieval (IR). The aim of an IR system is to organize and store information and retrieve the useful information when a user poses a query to the system. Modern IR systems usually accept a free-format natural language query from a user. Given a large collection of documents, keywords are matched between a query and the documents to predict the potential usefulness of the documents for fulfilling the user's information need. In a system based on the Vector Space Model, all

documents are ranked in decreasing order of their predicted usefulness. Thus documents that are potentially most useful are presented, in decreasing rank order, to the user.

The effectiveness of an IR system is measured in terms of *recall* and *precision*. *Recall* is a measure of the ability of a system to present all relevant documents. It is defined as the number of relevant documents retrieved divided by the total number of relevant documents in the collection. *Precision* is a measure of the ability of a system to present only relevant documents. It is defined as the number of relevant documents retrieved divided by the total number of documents retrieved. For example, suppose there are 80 documents relevant to *microscope* in the collection. The system returns 60 documents, 40 of which are about *microscope*. Then recall at this point is $40/80 = 50\%$, and precision is $40/60 = 67\%$.

1.2 Vector Space Model and the Smart System

In the Vector Space Model, natural language statements such as documents and queries are converted into vectors. The features of these vectors are usually the words and phrases remaining in the statement after stemming and removing stopwords. The vectors are weighted to give emphasis to terms that exemplify meaning and are most useful in retrieval. In the retrieval process, the query vector is compared to each document vector. Those that correlate highly with the query (based on the similarity measure used) are considered to be similar and are returned to the user. Smart is the most famous example of an IR system based on the Vector Space Model. Smart 11.0 is the system used in all our work.

Smart was developed by Gerard Salton of Cornell University. It performs automatic indexing by removing stopwords, stemming, term weighting, *etc.* These steps are detailed as follows:

- Tokenization: The text is first parsed into individual words and other tokens.
- Stopword Removal: Common function words (like *the, of, an, ...*), also called stopwords, are removed from this list of tokens.
- Stemming: Various morphological variants of a word are normalized to the same stem. Usually simple rules for suffix stripping are used in this process.
- Phrase formulation: Optionally, phrases in the text are recognized and are used in addition to the list of single words to index the text.
- Weighting: The term (word stems and phrases) vector, thus created for a text, is weighted using functions relating term frequency, inverse document frequency, and length normalization considerations.

In this way, both the queries and the documents are transformed into vectors of the form $D_i = (w_{i1}, w_{i2}, \dots, w_{it})$, where D_i represents a document (or query) and w_{ik} is the weight of term T_k in document D_i .

The two basic factors responsible for determining the importance of a term in a document are the *term frequency* factor and the *inverse document frequency* factor. *Term frequency* refers to the number of occurrences of a term in a document, and *document frequency* corresponds to the number of documents in the collection in which a term appears. A term is considered important in a document when it has a high term frequency (*tf*) and low document frequency (*df*). Weights are

assigned to the term in proportion to its term frequency and in inverse proportion to its document frequency. This is the well known $tf \times idf$ weighting system and follows the general term weighting principle — high weights assigned to terms that occur often in particular documents but less often in the collection as a whole.

In the Smart system, term weighting schemes are denoted by triples of letters. The first letter in a triple is a shorthand for the tf factor being used in the term weights, the second letter corresponds to the idf function, and the third letter corresponds to the normalization factor applied to the term weights. Table 1-1 summarizes the most commonly used term weighting function in Smart 11.0.

Table 1-1: Most commonly used term weights in Smart 11.0.

Term Frequency		Inverse Document Frequency		Normalization	
First Letter	$f(tf)$	Second Letter	$f(\frac{1}{df})$	Third Letter	$f(length)$
n (natural)	tf	n (none)	1	n (none)	1
l (logarithmic)	$1 + \ln(tf)$	t (full)	$\ln(\frac{N+1}{df})$	c (cosine)	$\sqrt{w_1^2 + w_2^2 + \dots + w_n^2}$
a (augmented)	$0.5 + 0.5 \times \frac{tf}{max\ tf}$				

$max\ tf$ — the maximum term frequency of any term in the document under consideration

N — the total number of documents in the collection

A retrieval experiment is characterized by a pair of triples — $ddd.qqq$ — where the first triple corresponds to the term weighting used for the documents, and the second triple corresponds

to the query term weights. For example, when *anc.ltn* is used to symbolize a retrieval run, the document term weights used in the run are

$$\frac{0.5 + 0.5 \times \frac{tf}{\max tf}}{\sqrt{\sum_{\text{all terms}} \left(0.5 + 0.5 \times \frac{tf}{\max tf}\right)^2}}$$

and the query term weights are

$$(1 + \ln(tf)) \times \ln\left(\frac{N+1}{df}\right).$$

The inner product similarity measure is used to rank documents, *i.e.*, when document D_i is represented by a vector of the form $(d_{i1}, d_{i2}, \dots, d_{it})$ and query Q_j by the vector $(q_{j1}, q_{j2}, \dots, q_{jt})$, the similarity between the query and the document is computed as

$$Sim(D_i, Q_j) = \sum_{k=1}^t (d_{ik} \times q_{jk}).$$

The documents in the collection are ranked by their decreasing similarity to the query and are presented to the user in this order. Documents presented earlier in a search have a higher degree of vocabulary overlap with the query and are probably more relevant to the user's information need.

1.3 TREC Collection and Evaluation

The Text REtrieval Conference (TREC) is an NIST and DARPA co-sponsored effort that brings together IR researchers from around the world to discuss their work on large test collections. A

common test collection and a common evaluation system are given so that the results produced by various systems can be compared and contrasted on the same data.

The common IR task of ranking documents for a new query is called the *ad-hoc* task in the TREC framework. The TREC data comes on CD-ROMs (the *TREC disks*). The disks are numbered, and a combination of several disks can be used to form a text collection for experimentation. In our experiments, disk 1, 2, 4 and 5 are used. The characteristics of the data on these two disks is listed in Table 1-2.

Table 1-2: Document statistics for TREC disk 1 & 2. Words are strings of alphanumeric characters. No stopwords were removed and no stemming was performed.

	Size (megabytes)	# Docs	Median # Word/Doc	Mean # Word/Doc
Disk 1				
<i>Wall Street Journal</i> , 1987-1989	267	98,732	245	434.0
<i>Associated Press</i> newswire, 1989	254	84,678	446	473.9
<i>Computer Selects</i> articles, Ziff-Davis	242	75,180	200	473.0
<i>Federal Register</i> , 1989	260	25,960	391	1315.9
Abstracts of U.S. DOE publications	184	226,087	111	120.4
Disk 2				
<i>Wall Street Journal</i> , 1990-1992 (WSJ)	242	74,520	301	508.4
<i>Associated Press</i> newswire (1988) (AP)	237	79,919	438	468.7
<i>Computer Selects</i> articles, Ziff-Davis (ZIFF)	175	56,920	182	451.9
<i>Federal Register</i> (1988)	209	19,860	396	1378.1
Disk 4				
the <i>Financial Times</i> , 1991-1994 (FT)	564	210,158	316	412.7
<i>Federal Register</i> , 1994 (FR94)	395	55,630	588	644.7
<i>Congressional Record</i> , 1993 (CR)	235	27922	288	1373.5
Disk 5				
Foreign Broadcast Information Services (FBIS)	470	130,471	322	543.6
the <i>LA Times</i>	475	131,896	351	526.5

The ad-hoc task defines the collection and queries to use. Each year 50 new user queries (or topics) are given to the participants. For example, the ad-hoc task for TREC-3 is running TREC topics 151-200 against all documents from Disk 1 and 2.

Relevance assessments of the documents in the collection are available for each topic and make evaluation possible. One of the important TREC measures is called non-interpolated average precision over all relevant documents. It is a single-valued measure that averages the precision value obtained after each relevant document is retrieved. It reflects performance over all relevant documents. For example, consider a query that has four relevant documents which are retrieved at ranks 1, 2, 4, and 7. The actual precision obtained when each relevant document is retrieved is 1, 0.75, and 0.57, respectively, the mean of which is 0.83. Thus, the average precision over all relevant documents for this query is 0.83.

2 Motivation

Automatic query expansion via relevance feedback is a well-recognized and effective technique which is commonly used to add useful terms to a query. However, web users are unlikely to provide a system with the relevant judgements needed for relevance feedback [Jansen 1998]. In such situations, *pseudo* relevance feedback is commonly used to expand the user-query. In this method, a small set of documents (the *pseudo-relevant set*) is retrieved using the original user-query and these documents are assumed to be relevant. These documents are then used in a relevance feedback process to construct an expanded query, which is then run to retrieve the set of documents actually presented to the user.

When one deals with very short queries (a few words) that are often not very specific in nature, the initial set of retrieved documents can be very poor (*i.e.*, low precision). Because no relevance assessments are made on this list, there can be many non-relevant documents in the pseudo-relevant set. When terms are selected to expand the query, many will be of dubious value. It is therefore surprising that this process works at all. Yet experiments have shown that it works more often than not. Kwok [Kwok 1998] attributes this to the fact that even though the documents used for pseudo-feedback are often not relevant, they fall in the general topic area of the raw query; the retrieval algorithm works to the extent that documents of the same general subject area are biased to rank high. These documents then provide a source of terms that can augment the initial query in topical description; this expanded query can attract documents in the same topical area and improve the chance of producing highly ranked relevant documents in the second retrieval.

Despite of the utility of this approach, it was also observed that poor pseudo-relevant sets could cause a real problem, which is called *query drift* [Mitra 1998]. In query drift, too many non-relevant documents are presented in the pseudo-relevant set, and the focus of a search topic is altered by an improper expansion. To prevent query drift, we want as high as possible a percentage of relevant documents in the pseudo-relevant set used for the pseudo-feedback. Improving the precision of the pseudo-relevant set appears crucial for successful pseudo-feedback retrieval.

A significant reason for low precision in the pseudo-relevant set has been identified by previous work in our IR Research Lab [Wang 1998]. We found that the stemmed terms in the query quite often attract non-relevant documents at top ranks. For example, consider Query 10 (title only) from the TREC collection: *AIDS treatments*. After stemming, it consists of two terms: *aid* and *treat*. This significantly changes the meaning of the original query. As a result, quite a few non-relevant documents are top-ranked for an initial retrieval (*e.g.*, with *atc.atc* weighting scheme in our experiment, 14 non-relevant documents are in the top 20 for an initial retrieval. Some of the articles contain the term *aid* in terms of its normal dictionary meaning.).

Stemming is intended to maximize recall. Consequently, we need to retain stemming for indexing the document and query collections. To get higher precision for the pseudo-relevant set, we use a filtering process to push more relevant documents into the top ranks. This new set of top-ranked documents now becomes our pseudo-relevant set. The approach is as follows:

1. Assume the pseudo-relevant set consists of M documents. Retrieve N documents, where $N > M$.

2. Examine documents in this set and compare the original, unstemmed terms in the query against the original, unstemmed text documents. Assign a new similarity score to each document based on the correlation.
3. The document set is reranked based on the new similarity and the top-ranked documents are chosen to form the pseudo-relevant set.

We call step 2 the *exact match* step. By exact match, we expect to reduce the number of non-relevant documents in the top ranks so that precision in the pseudo-relevant set is increased.

3 Algorithms

Incorporating the precision-enhancing step as discussed in the previous section into the usual pseudo relevance feedback process, we arrive at the following algorithm:

Stage 1 (Reranking):

1. To use M (e.g., 20) documents in the feedback process, retrieve N documents using the original user query, where $N > M$.
2. For each retrieved document, compute a new similarity score Sim_{new} based on exact match.
3. Rerank the N retrieved documents in decreasing Sim_{new} .

Stage 2 (Pseudo-feedback):

1. Select the top M documents in the new ranking list and use them to expand the query.
2. Use the expanded query to retrieve the final list of documents returned to the user.

The reranking in Stage 1 is expected to produce a higher proportion of relevant documents in the pseudo-relevant set. The three reranking algorithms we considered are detailed as below. For each method, an exact match is used as discussed in the previous section, with the exception of singular/plural nouns (all plurals are reduced to singular form).

Method 1 (Simple count):

Count the number of unique query terms occurring in the document. Consider, for example, Query 1 from TREC topics: *Antitrust Cases Pending*. After reducing *cases* to its singular form *case*, the query consists of three words: *antitrust*, *case* and *pending*. Suppose in a document, *antitrust* occurs 3 times, *pending* occurs 2 times, and *case* does not occur. Then the new similarity

score for the document is 2. Simple count ignores the frequency of each query term in the document.

$$Sim_{new}(D) = \text{number of unique query terms in the document}$$

Method 2 (Weighted count):

Sum the frequencies of query terms occurring in the document. The new similarity score for the document in the above example is 5.

$$Sim_{new}(D) = \sum_{t_i \in Q \cap t_i \in D} tf(t_i)$$

where $tf(t_i)$ is the term frequency of query term t_i in the document D .

Method 3 (Importance weights):

Apply the weighting scheme used for the initial retrieval to generate the weights for the original (unstemmed) query terms. Terms are then re-weighted by applying an importance factor derived from the statistics of the top 1000 documents produced by the initial retrieval. Finally the new similarity is calculated as

$$Sim_{new}(D) = \sum_{t_i \in Q \cap t_i \in D} tf(t_i) \times w_{new}(t_i)$$

where $tf(t_i)$ is the term frequency of query term t_i in the document, and $w_{new}(t_i)$ is the adjusted new weight for the query term t_i .

As detailed in [Singhal 1998], the importance factor is used to reflect the importance of a query term (in addition to its original term weight). First, the following function is used to rank the original query terms:

$$\frac{\text{document frequency of the term in first 1,000 documents retrieved (df in 1,000)}}{\text{document frequency of the term within the collection (df)}}$$

After the terms in a query are ranked by the above formula, their weights are modified by multiplying with the following importance factor:

$$1.0 - \sqrt{\frac{\text{rank} - 1}{10}}$$

This factor lowers the weights of the terms ranked poorly in the above ranking, therefore noticeably emphasizing the top few terms.

For example, if we perform an initial retrieval using the *Inc.ltc* weighting scheme, we use *ltc* weights for the unstemmed query terms. Consider Query 1 from the TREC collection: *Antitrust Cases Pending*. The term ordering generated by this scheme is shown in Table 3-1.

Table 3-1: Term ordering for TREC Query 1.

Word	<i>df in 1,000</i>	<i>df</i>	$\frac{\text{df in 1,000}}{\text{df}}$	Original <i>ltc</i> Weight	Importance Factor	Final Weight
antitrust	669	2162	0.3094	0.7267	1.0000	0.7267
pending	426	5020	0.0849	0.5871	0.6838	0.4014
case	628	20156	0.0312	0.3566	0.5528	0.1971

The query terms are listed in decreasing order of their perceived importance in Table 3-1. In this way, the importance of some critical words, like *antitrust* in Query 1, gets emphasized.

To summarize this method, the main idea is to down-weight those query terms which appear more general so that when term frequency is taken into account for calculating the new similarity of a document, documents with many general terms will not be over-emphasized. The *idf* factor is used for this purpose, not only in the original weights (when a *t* factor is used for

query term weighting, such as *atc* and *lrc*), but also in the calculation of the importance factor (*df* in the collection used for ranking the query terms). This requires a *df* value for the unstemmed query terms, which in turn requires that an unstemmed document collection be maintained. In practice, it is not feasible to maintain both stemmed and unstemmed document collections. So a more realistic approach would use the statistical data from the stemmed document collection; *i.e.*, to obtain the *df* value for an unstemmed term from the dictionary of the stemmed document collection. We believe our method is a good approximation of this approach; however, further study is required. Another possible approach is to use statistics based on top 1000 documents retrieved to assign query term weights (to down-weight the general query terms).

Comments on these methods:

Method 1 emphasizes the number of unique query terms in the document text. The disadvantage of this approach is obvious: each word in the query is equally emphasized, and term frequency does not affect the correlation between document and query.

Method 2 takes the term frequency into account. However, each term in the query is still considered of equal importance. It has the disadvantage that if a general term in the query occurs several times in a document, the document will have a high correlation score with the query.

Method 3 takes both into consideration. However, its application may produce weights which are open to question. Consider, for example, Query 50 from the TREC collection: *Potential Military Interest in Virtual Reality Applications*. The ranking of terms generated by *lrc* weighting scheme is shown in Table 3-2. The method ranks *application* ahead of *virtual*, which seems counterintuitive.

Table 3-2: Term ordering for TREC Query 50.

Word	<i>df</i> in 1,000	<i>df</i>	$\frac{df \text{ in } 1,000}{df}$	Original <i>ltc</i> Weight	Importance Factor	Final Weight
Application	382	4045	0.0944	0.4013	1.0000	0.4013
virtual	25	442	0.0566	0.6378	0.5528	0.3526
military	447	7604	0.0588	0.3339	0.6838	0.2283
reality	72	2417	0.0298	0.4563	0.3675	0.1677
potential	352	11356	0.0310	0.2911	0.4523	0.1316
interest	467	36505	0.0128	0.1663	0.2929	0.0487

4 Experiments and Results

We used the WSJ articles (1987-1992) from TREC disks 1 and 2 as the collection and TREC topics 1-50 as the queries for our initial investigations with short queries. In a continued effort, we use them for our new experiments as well. Then we extend our study to the documents/queries specified for the TREC-3 ad-hoc task (using all documents from disks 1 and 2 as the collection and the TREC topics 151-200 as the query set) and the TREC-6 ad-hoc task (using all documents from disks 4 and 5 as the collection and the TREC topics 301-350 as the query set).

4.1 Test Collections and Test Queries

The characteristics of the collections can be found in Table 1-2. The WSJ collection contains 173,252 documents and the average number of terms in a document is 466. The TREC-3 collection contains 741,856 documents and the average number of terms in a document is 415.7. The TREC-6 collection contains 556,077 documents and the average number of terms in a document is 541.9.

An example of the queries (or topics) used for our experiments is shown in Figure 4-1. Only the topic titles are used as queries because they resemble the range of queries that one would expect web users to input. The characteristics of the queries are listed in Table 4-1.

<p><num> Number: 151</p> <p><title> Topic: Coping with overcrowded prisons</p> <p><desc> Description:</p> <p>The document will provide information on jail and prison overcrowding and how inmates are forced to cope with those conditions; or it will reveal plans to relieve the overcrowded condition.</p> <p><narr> Narrative:</p> <p>A relevant document will describe scenes of overcrowding that have become all too common in jails and prisons around the country. The document will identify how inmates are forced to cope with those overcrowded conditions, and/or what the Correctional System is doing, or planning to do, to alleviate the crowded condition.</p>
--

Figure 4-1: A sample TREC topic statement.

	Min	Max	Average
Topics 1-50 (title only) ¹			
including stopwords	1	12	4.4
with stopword removal	1	8	3.2
Topics 151-200 (title only)			
including stopwords	2	20	6.5
with stopword removal	2	11	4.5
Topics 301-350 (title only)			
including stopwords	1	5	2.7
with stopword removal	1	4	2.4

Table 4-1: Query length statistics.

There are 6,094 relevant documents in the WSJ collection for Queries 1-50 (122 documents/query on the average). In TREC-3 ad-hoc task, there are 9,805 relevant documents for Queries 151-200 (196 documents/query on the average). And in TREC-3 ad-hoc task, there are 4,611 relevant documents for Queries 301-350 (92 documents/query on the average).

¹Topic 22 is omitted since it does not get indexed as explained in Section 4.3.1.

4.2 Experimental Parameters

There are many alternatives for each step in the general algorithm stated in Section 3. Many parameter combinations are possible. The detailed settings for our experiments are specified here.

Stage 1 (Reranking):

The initial retrieval is a standard Smart run: after stopword removal and stemming, documents and queries are indexed using single terms; term-weights are computed using one of the standard (*i.e.*, *atc.atc*, *ltc.ltc*, and *inc.ltc*) weighting schemes; and 1000 documents are retrieved for each query using the simple vector inner-product similarity measure. In these experiments, we considered each of the weighting schemes mentioned above. *atc* represents an augmented $tf \times idf$ weighting scheme. *ltc* is another $tf \times idf$ scheme that also seeks to restrain the impact of terms with high *tf* values. *inc* uses the same logarithmic approach applied without the *ltc* factor. TREC-1 experiments run at Cornell University [Buckley 1993] indicate that *inc.ltc* weights produce the best results, at the same time reducing the computation required to calculate *idf* factor in large document collections.

Then **N** top-ranked documents are selected for reranking. Apply the reranking algorithms (Method 1, 2, or 3) to this set of **N** documents and examine the effect of the reranking on the top-ranked **M** documents. The average precision at the top **M** documents is used to measure this effect. In our experiments, **N** ranges from 50 to 1000 in step of 50, and the values of **M** considered are 20 and 30.

Stage 2 (Pseudo-feedback):

Choose the most effective weighting scheme, reranking method, M value, *etc.*, in Stage 1 to carry out the pseudo-feedback. The top M documents are assumed to be relevant and the query is expanded using terms from these documents. We used Rocchio's algorithm for relevance feedback.

In the Rocchio relevance feedback process, all terms occurring in the relevant documents are sorted by the number of relevant documents in which they occur, with ties being broken by considering the highest average weight in the relevant documents. The top T such terms are then re-weighted according to Rocchio's formula and added to the original query to produce the feedback query.

Rocchio's formula to compute the feedback weight for an individual query term Q is:

$$Q_{new} = \alpha \times Q_{old} + \beta \times \text{average weight in relevant documents} \\ - \gamma \times \text{average weight in non-relevant documents}$$

The parameters (α , β , and γ) of Rocchio's algorithm are used to indicate the relative importance of terms from the original query, relevant documents, and non-relevant documents, respectively. In our case, all top M documents are assumed relevant. So γ does not play a role in these experiments, and we can see that only the ratio of α and β is important for the final ranking.

We examine the effect of varying T , the number of terms for expansion, and the parameter β for Rocchio's formula (with α set to 1) in our experiments. Suitable ranges for these parameters are investigated.

4.3 Experimental Results

4.3.1 WSJ Collection vs. TREC Query 1-50 (title only)

For this experiment, only 49 queries remain after indexing. (Query 22, *counternarcotics*, does not get indexed since the word *counternarcotics* does not appear in any of the documents in the collection and thus is not present in the dictionary.)

Stage 1 (Reranking):

We experimented with the *atc.atc*, *ltc.ltc*, and *inc.ltc* weighting schemes. Each of the three reranking methods (simple count, weighted count, and importance weights) was studied. We also studied the effect of varying **N**, the number of top-ranked documents reranked to select the pseudo-relevant set of **M** documents used in feedback. We looked at the average precision at **M** documents in various configurations. The precision at **M** documents before reranking is used as the baseline to compute the percentage improvement.

The experimental results are shown in Tables 7-1 and 7-2 in Appendix I. The baseline figures are listed in parentheses. The percentage improvements over the baselines for various parameter settings are also listed. The figures for the best settings in each weighting scheme are highlighted.

The improvements are generally substantial as we can see for all of the three reranking methods and different weighting schemes when the parameters are properly selected. However, the *inc.ltc* weighting scheme produces the highest baseline precision: 0.2480 at 20 documents (P@20) and 0.2429 at 30 documents (P@30). Generally it also produces the highest precision after reranking. In this scheme, the first reranking method (simple count) seems to produce the

best performance especially when the reranking is done on 150 - 450 documents, although the figures are quite stable on all **N** values. And generally we can see that the proportion of relevant documents in the top 20 is always greater than that at 30.

To summarize, these observations suggest the following settings for Stage 2:

- Weighting scheme [*atc.atc*, *ltc.ltc*, *inc.ltc*]: ***inc.ltc*** .
- Reranking method [simple count, weighted count, importance weights]: **simple count**.
- **N** (number of documents used for reranking) [50, 100, ..., 1000]: **150-450**.
- **M** (number of documents used for pseudo-feedback) [20, 30]: **20**.

The best settings (*i.e.*, those that maximize proportion of relevant documents in the top-ranked set) are used for Stage 2.

Stage 2 (Pseudo-feedback):

The *inc.ltc* weighting scheme and reranking method 1 (simple count) are used for Stage 2. We studied the pseudo-feedback performance by assuming the top 20 documents to be relevant with the document list reranked on various values of **N**. We studied the effect of **T** (number of terms for expansion) and Rocchio parameters (β with α set to 1) as well.

Tables 7-3 and 7-4 in Appendix I show the effect of varying **T** for the expansion. In this study, α and β are arbitrarily set to 1. Reranking is done on the top **N** documents with **N** varying from 50 to 1000 in steps of 50. The non-interpolated average precision and precision at the top 20 documents averaged over all queries are used as the measures of performance. The baseline figures listed in parentheses correspond to an initial retrieval of 1,000 documents without pseudo-feedback. The baseline pseudo-feedback figures are also listed which correspond to pseudo-

feedback runs by assuming the top 20 documents in the initially retrieved document list to be relevant. The percentage improvement over the baseline is computed for each run.

Our pseudo-feedback runs based on the reranked document list improve the performances over the baselines significantly as well as the baseline pseudo-feedback runs. Besides, it is observed that expanding more terms tends to improve both measures. We also observed a correspondence between the precision at the top 20 documents of the reranked document list and the performance of the pseudo-feedback. It is generally true that the higher the precision at the top 20 documents in the reranked list, the better the performance of the pseudo-feedback, although there is some fluctuation.

Tables 7-6 and 7-8 in Appendix I show the effect of varying β with α set to 1 on the pseudo-feedback runs. Reranking is done on the top 400 documents (which gives the best precision at 20 documents after reranking) to get the pseudo-relevant set for pseudo-feedback. The effect of varying T can also be observed in these tables. The baseline pseudo-feedback runs are listed in tables as well (Table 7-5 and 7-7 in Appendix I). Again we observed the improvements of our pseudo-feedback runs based on the reranked list over the baselines and baseline pseudo-feedback runs. And from the tables we can see that a β value of 2 is good for both of the performance measures.

To summarize, the Stage-2 experiments suggest the following:

- A massive query expansion is desired for the pseudo-feedback run; and
- With α set to 1, a β value 2 appears appropriate.

4.3.2 TREC-3 ad-hoc Task (TREC disks 1 & 2 vs. Query 151-200)

Stage 1 (Reranking):

The experiments are carried out in the same way as in the last section. The results are shown in Tables 8-1 and 8-2 in Appendix II. Results indicate that the following approach yields the highest precision:

- Weighting scheme [*atc.atc*, *ltc.ltc*, *inc.ltc*]: ***inc.ltc***.
- Reranking method [simple count, weighted count, importance weights]: **simple count**.
- N (number of documents used for reranking) [50, 100, ..., 1000]: **650-1000**.
- M (number of documents used for pseudo-feedback) [20, 30]: **20**.

The best settings are used for Stage 2.

Stage 2 (Pseudo-feedback):

The performance of the pseudo-feedback is evaluated the same way as in the last section. Tables 8-3 and 8-4 in Appendix II show the effect of varying T for the expansion. Again, in this study, α is set to 1, and β is set to 1. And reranking is done on the top N documents with N varying from 50 to 1000 in steps of 50.

Tables 8-5 to 8-8 in Appendix II show the effect of varying β with α set to 1 (reranking is done on the top 800 documents to get the pseudo-relevant set for pseudo-feedback).

As a result, the following is suggested by Stage 2:

- A massive expansion is desired; and
- With α set to 1, set β to **2-6**.

4.3.3 TREC-6 ad-hoc Task (TREC disks 4 & 5 vs. Query 301-350)

Stage 1 (Reranking):

The experiments are carried out the same as before. The results are shown in Tables 9-1 and 9-2 in Appendix III. Results indicate that the following approach yields the highest precision:

- Weighting scheme [*atc.atc* , *ltc.ltc* , *inc.ltc*]: ***inc.ltc*** .
- Reranking method [simple count, weighted count, importance weights]: **simple count**.
- **N** (number of documents used for reranking) [50, 100, ..., 1000]: **150-500**.
- **M** (number of documents used for pseudo-feedback) [20, 30]: **20**.

The best settings are used for Stage 2.

Stage 2 (Pseudo-feedback):

The performance of the pseudo-feedback is evaluated as before. Tables 9-3 and 9-4 in Appendix III show the effect of varying **T** for the expansion. Again, in this study, **α** is set to 1, and **β** is set to 1. And reranking is done on the top **N** documents with **N** varying from 50 to 1000 in steps of 50.

Tables 9-5 to 9-8 in Appendix III show the effect of varying **β** with **α** set to 1 (reranking is done on the top 300 documents to get the pseudo-relevant set for pseudo-feedback).

As a result, the following is suggested by Stage 2:

- A massive expansion is desired; and
- With **α** set to 1, set **β** to 1.

4.4 Query Drift

Recall that our reranking approach is intended to prevent query drift, caused by term expansion on too many non-relevant documents in the pseudo-relevant set, mainly by improving the precision in the set of documents used for feedback. In this section, we examine the issue of query drift in greater detail.

We concentrate our study on one of our pseudo-feedback runs in the TREC-3 ad-hoc task.

The specification for the run is as below:

1. *Inc.Itc* weighting scheme for initial retrieval;
2. method 1 (simple count) to rerank the top **800** documents;
3. select the top **20** documents from the reranked list to form the pseudo-relevant set;
4. expand the original queries by **300** terms;
5. set α to **1** and β to **6**.

The non-interpolated average precision averaged over all queries for this run is 0.2953, which is a 123.7% improvement over the baseline (31.1% over the baseline pseudo-feedback). The P@20 is 0.5200, which is a 92.6% improvement over the baseline (48.6% over the baseline pseudo-feedback).

We expect query drift to be most severe for queries that initially retrieve few relevant documents in the top ranks. On the other hand, query drift should not occur for queries retrieving a good number of relevant documents at the top ranks. To study the relationship between query drift and initial precision, we partition the 50 queries used in our TREC-3 ad-hoc task into bins corresponding to the number of relevant documents retrieved by the initial query within the top 20

ranks. Thus the first bin contains the queries which retrieve no relevant documents in the top 20, and the last bin contains queries for which all top 20 documents are relevant.

First we examine how our reranking algorithms improve precision in the top ranks for each bin. This is done by computing the average number of relevant documents in the top 20 documents in the original and reranked document list for the set of queries in each bin [Mitra 1998]. We plot histograms with the x-axis representing the query bins and the y-axis the difference in the average number of relevant documents contained in the original and in the reranked set for queries in that bin. A bar above the x-axis indicates the post-reranking $P@20$ (precision at top 20 documents) is better than the initial $P@20$. A bar below the x-axis indicates that reranking hurts $P@20$.

Figure 4-2 shows the histograms corresponding to the reranking technique used in our pseudo-feedback run specified at the beginning of this section. Most of the bars are above the x-axis, showing that reranking does indeed result in an increased concentration of relevant documents in the pseudo-relevant set for these queries.

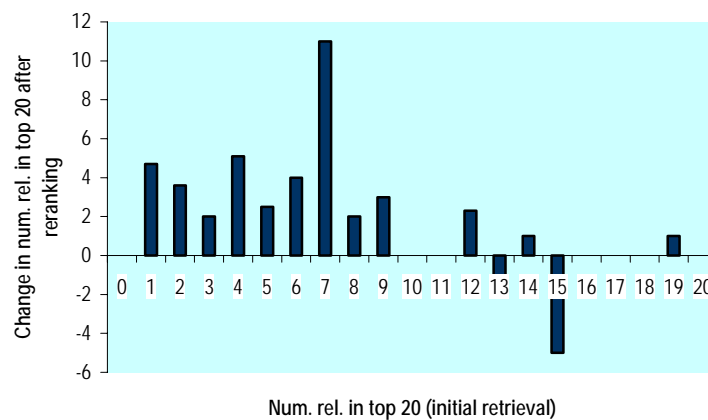


Figure 4-2: Effect of reranking on precision at 20 docs.

Figures 4-3 and 4-4 show how these changes affect query expansion. The histograms shown in Figure 4-3 are generated as follows. For each bin, the average precision is computed for the initial retrieval and the post-feedback run. The percentage difference between these two figures is plotted on the y-axis. A bar above the x-axis shows that there was an improvement in average precision due to expansion. A bar below the x-axis indicates the contrary. The white bars in Figure 4-3 correspond to a baseline pseudo-feedback, where the expansion is based on the initially retrieved document list by assuming the top 20 documents to be relevant and choosing the same parameters as for our pseudo-feedback run. The black bars correspond to our pseudo-feedback run where the expansion is based on the reranked document list. The histograms in Figure 4-4 are generated for P@20 in the same way.

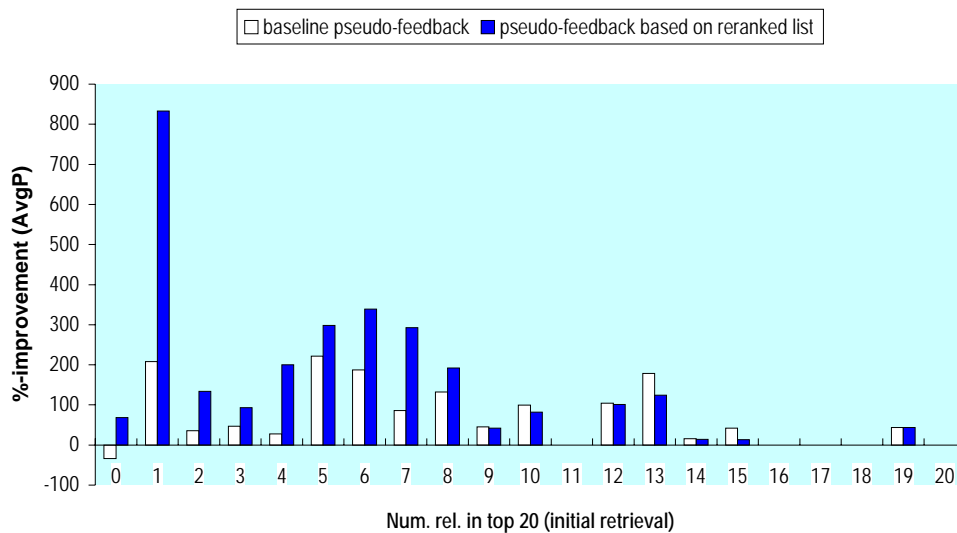


Figure 4-3: Query Drift in terms of average precision.

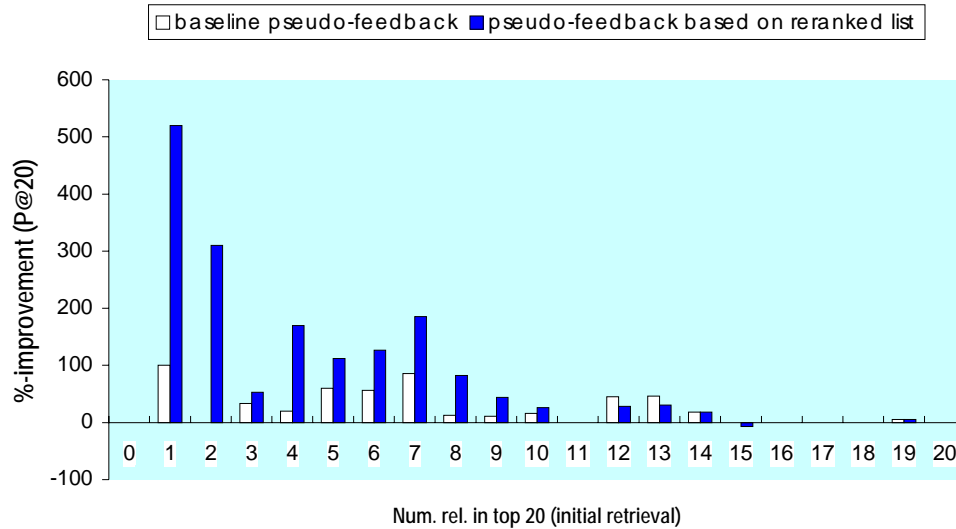


Figure 4-4: Query Drift in terms of P@20.²

As expected, we observe that more significant improvements are achieved for the bins on the left-hand side by expansion based on the reranked list. Although there are drops for the bins on the right-hand side, they are very slight and in most cases there are still substantial improvements over the baselines. Reranking does reduce query drift as we hoped.

² The very first bin is ignored since P@20 is 0 initially so the improvement is infinite.

5 Conclusions and Suggestions for Future Research

5.1 Conclusions

Retrieval via very short queries in an ad-hoc environment presents difficulties for information retrieval. Yet it is an important task, since users of web search engines very commonly formulate a query with a few words. We focus on trying to improve automatic expansion in pseudo-feedback. We attempt approaches to refine the pseudo-relevant set by reranking a larger document set using exact match (simple count, weighted count, importance weights).

We evaluated our techniques on the WSJ collection, the TREC-3 ad-hoc task and the TREC-6 ad-hoc task. Significant improvements have been observed. In a *Inc.Itc* weighting scheme, reranking by simple count appears very successful to improve the precision of the pseudo-relevant set. The pseudo-feedback based on this refined pseudo-relevant set effectively reduces query drift and results in significant improvements in retrieval effectiveness (both non-interpolated average precision and P@20).

5.2 Suggestions for Future Research

1. To verify the general applicability of our techniques, more experiments can be done on various query and collection combinations and analysis carried out on them. If a technique is effective for a large set of queries across multiple collections, this strongly indicates that the technique is generally effective.
2. Better weighting schemes, not available through Smart 11.0, have since been established, such as *Lnu.ltu* [Singhal 1997][Singhal 1998] and *dnb.dtn* [Singhal 1999]. We should test

- our reranking methods applied to these weighting schemes to see if we can further improve retrieval effectiveness. We can also investigate incorporating our reranking techniques with many other techniques (*e.g.*, using phrases, assuming a set of non-relevant documents, breaking documents into overlapping windows to compute the similarity [Singhal 1998], ...) that have been developed to improve the retrieval effectiveness of very short queries.
3. Given the disadvantage of our simple count reranking method (query terms are equally emphasized), we can explore better reranking methods based on exact match to emphasize the most important query terms. A suggested approach might include down-weighting general terms (a) by use of an appropriate function and/or (b) by using statistics derived from the top 1000 documents retrieved (as suggested in Section 3).

6 References

[Buckley 1993] C. Buckley, G. Salton, and J. Allan. *Automatic Retrieval with Locality information Using SMART*. In: Proc. of the First Text REtrieval Conference (TREC-1). NIST Special Publication 500-207, 1993.

[Buckley 1995] C. Buckley, G. Salton, J. Allan, and A. Singhal (1995). *Automatic Query Expansion Using SMART: TREC 3*. In: Proc. of the Third Text REtrieval Conference (TREC-3). NIST Special Publication 500-225, 1995.

[Jansen 1998] B. J. Jansen, A. Spink, J. Bateman, and T. Saracevic. *Real Life Information Retrieval: A Study Of User Queries On The Web*. In: SIGIR Forum, Volume 32 Number 1, 1998.

[Kwok 1998] K.L. Kwok and M. Chan. *Improving Two-Stage Ad-Hoc Retrieval for Short Queries*. In: Proc. of 21st Ann. Intl. ACM-SIGIR Conf. On R&D in IR, 1998.

[Magennis 1997] M. Magennis and C.J. van Rijsbergen. *The potential and actual effectiveness of interactive query expansion*. In: Proc. of 20th Ann. Intl. ACM-SIGIR Conf. On R&D in IR, 1997.

[Mitra 1998] M. Mitra, A. Singhal, and C. Buckley. *Improving Automatic Query Expansion*. In: Proc. of 21st Ann. Intl. ACM-SIGIR Conf. On R&D in IR, 1998.

[Singhal 1997] A. Singhal. *Term Weighting Revisited*. A Dissertation Presented to the Faculty of the Graduate School of Cornell University. January, 1997.

[Singhal 1998] A. Singhal. *AT&T at TREC-6*. In: Proc. of the Sixth Text REtrieval Conference (TREC-6). NIST Special Publication 500-240, 1998.

[Singhal 1999] A. Singhal, J. Choi, D. Hindle, and D.D Lewis. *AT&T at TREC-7*. In: Proc. of the Seventh Text REtrieval Conference (TREC-7). NIST Special Publication 500-242, 1998.

[Wang 1998] X. Wang. *Experiment on generating improved queries for the World Wide Web*. A Thesis Presented to the Faculty of the Graduate School of University of Minnesota, Duluth. 1998.

7 Appendix I

P@20									
	<i>inc.ltc</i> (0.2480)			<i>ltc.ltc</i> (0.2112)			<i>atc.atc</i> (0.1633)		
N↓	1	2	3	1	2	3	1	2	3
50	0.2898	0.2959	0.3020	0.2694	0.2704	0.2643	0.2194	0.2276	0.2347
	16.87%	19.34%	21.81%	27.54%	28.02%	25.12%	34.38%	39.38%	43.75%
100	0.2959	0.2878	0.2959	0.2857	0.2745	0.2765	0.2480	0.2500	0.2480
	19.34%	16.05%	19.34%	35.27%	29.95%	30.92%	51.88%	53.13%	51.88%
150	0.3020	0.2786	0.2929	0.2888	0.2653	0.2714	0.2459	0.2592	0.2602
	21.81%	12.35%	18.11%	36.71%	25.60%	28.50%	50.63%	58.75%	59.38%
200	0.3010	0.2806	0.2918	0.2878	0.2469	0.2561	0.2418	0.2531	0.2561
	21.40%	13.17%	17.70%	36.23%	16.91%	21.26%	48.13%	55.00%	56.88%
250	0.3051	0.2612	0.2827	0.2939	0.2541	0.2633	0.2480	0.2622	0.2653
	23.05%	5.35%	13.99%	39.13%	20.29%	24.64%	51.88%	60.63%	62.50%
300	0.3041	0.2500	0.2776	0.2908	0.2500	0.2694	0.2500	0.2622	0.2663
	22.63%	0.82%	11.93%	37.68%	18.36%	27.54%	53.13%	60.62%	63.12%
350	0.3031	0.2500	0.2786	0.2898	0.2551	0.2714	0.2510	0.2551	0.2592
	22.22%	0.82%	12.35%	37.20%	20.77%	28.50%	53.75%	56.25%	58.75%
400	0.3061	0.2449	0.2765	0.2908	0.2490	0.2673	0.2531	0.2510	0.2582
	23.46%	-1.23%	11.52%	37.68%	17.87%	26.57%	55.00%	53.75%	58.12%
450	0.3051	0.2439	0.2786	0.2918	0.2429	0.2663	0.2582	0.2429	0.2571
	23.05%	-1.65%	12.35%	38.16%	14.98%	26.09%	58.13%	48.75%	57.50%
500	0.2990	0.2429	0.2765	0.2918	0.2408	0.2633	0.2561	0.2469	0.2561
	20.58%	-2.06%	11.52%	38.16%	14.01%	24.64%	56.88%	51.25%	56.88%
550	0.3000	0.2367	0.2724	0.2918	0.2418	0.2633	0.2541	0.2439	0.2510
	20.99%	-4.53%	9.88%	38.16%	14.49%	24.64%	55.63%	49.38%	53.75%
600	0.2980	0.2388	0.2735	0.2918	0.2378	0.2663	0.2541	0.2388	0.2520
	20.16%	-3.70%	10.29%	38.16%	12.56%	26.09%	55.63%	46.25%	54.38%
650	0.2980	0.2316	0.2663	0.2918	0.2388	0.2684	0.2541	0.2367	0.2531
	20.16%	-6.58%	7.41%	38.16%	13.04%	27.05%	55.63%	45.00%	55.00%
700	0.2980	0.2337	0.2663	0.2898	0.2367	0.2663	0.2520	0.2378	0.2551
	20.16%	-5.76%	7.41%	37.20%	12.08%	26.09%	54.38%	45.63%	56.25%
750	0.2980	0.2367	0.2633	0.2898	0.2398	0.2653	0.2520	0.2347	0.2541
	20.16%	-4.53%	6.17%	37.20%	13.53%	25.60%	54.38%	43.75%	55.63%
800	0.2969	0.2316	0.2622	0.2888	0.2398	0.2643	0.2520	0.2388	0.2622
	19.75%	-6.58%	5.76%	36.71%	13.53%	25.12%	54.38%	46.25%	60.63%
850	0.2980	0.2337	0.2622	0.2898	0.2388	0.2643	0.2510	0.2378	0.2633
	20.16%	-5.76%	5.76%	37.20%	13.04%	25.12%	53.75%	45.63%	61.25%
900	0.2949	0.2286	0.2602	0.2888	0.2286	0.2633	0.2520	0.2418	0.2673
	18.93%	-7.82%	4.94%	36.71%	8.21%	24.64%	54.38%	48.13%	63.75%
950	0.2949	0.2286	0.2592	0.2898	0.2265	0.2602	0.2500	0.2378	0.2622
	18.93%	-7.82%	4.53%	37.20%	7.25%	23.19%	53.13%	45.63%	60.63%
1000	0.2949	0.2245	0.2612	0.2898	0.2245	0.2592	0.2490	0.2327	0.2633
	18.93%	-9.47%	5.35%	37.20%	6.28%	22.71%	52.50%	42.50%	61.25%

Table 7-1: P@20 before and after reranking for different settings (WSJ collection vs. Query 1-50).

P@30									
	<i>inc.ltc</i> (0.2429)			<i>ltc.ltc</i> (0.2000)			<i>atc.atc</i> (0.1565)		
N↓	1	2	3	1	2	3	1	2	3
50	0.2653 9.24%	0.2741 12.89%	0.2694 10.92%	0.2401 20.07%	0.2415 20.75%	0.2347 17.35%	0.1925 23.04%	0.1993 27.39%	0.1980 26.52%
100	0.2735 12.61%	0.2660 9.52%	0.2721 12.04%	0.2599 29.93%	0.2517 25.85%	0.2469 23.47%	0.2184 39.57%	0.2333 49.13%	0.2265 44.78%
150	0.2741 12.89%	0.2667 9.80%	0.2741 12.89%	0.2619 30.95%	0.2537 26.87%	0.2544 27.21%	0.2245 43.48%	0.2429 55.22%	0.2367 51.30%
200	0.2721 12.04%	0.2633 8.40%	0.2748 13.17%	0.2633 31.63%	0.2408 20.41%	0.2435 21.77%	0.2224 42.17%	0.2401 53.48%	0.2374 51.74%
250	0.2755 13.45%	0.2551 5.04%	0.2701 11.20%	0.2714 35.71%	0.2415 20.75%	0.2517 25.85%	0.2218 41.74%	0.2395 53.04%	0.2469 57.83%
300	0.2748 13.17%	0.2429 0.00%	0.2660 9.52%	0.2714 35.71%	0.2388 19.39%	0.2463 23.13%	0.2265 44.78%	0.2395 53.04%	0.2503 60.00%
350	0.2741 12.89%	0.2388 -1.68%	0.2639 8.68%	0.2714 35.71%	0.2408 20.41%	0.2537 26.87%	0.2272 45.22%	0.2361 50.87%	0.2449 56.52%
400	0.2762 13.73%	0.2306 -5.04%	0.2626 8.12%	0.2721 36.05%	0.2340 17.01%	0.2497 24.83%	0.2286 46.09%	0.2333 49.13%	0.2442 56.09%
450	0.2748 13.17%	0.2299 -5.32%	0.2605 7.28%	0.2728 36.39%	0.2306 15.31%	0.2524 26.19%	0.2327 48.70%	0.2293 46.52%	0.2422 54.78%
500	0.2735 12.61%	0.2259 -7.00%	0.2626 8.12%	0.2728 36.39%	0.2231 11.56%	0.2531 26.53%	0.2340 49.57%	0.2279 45.65%	0.2435 55.65%
550	0.2735 12.61%	0.2231 -8.12%	0.2585 6.44%	0.2714 35.71%	0.2224 11.22%	0.2537 26.87%	0.2361 50.87%	0.2313 47.83%	0.2456 56.96%
600	0.2728 12.32%	0.2245 -7.56%	0.2578 6.16%	0.2707 35.37%	0.2238 11.90%	0.2524 26.19%	0.2361 50.87%	0.2313 47.83%	0.2476 58.26%
650	0.2728 12.32%	0.2224 -8.40%	0.2524 3.92%	0.2714 35.71%	0.2252 12.59%	0.2510 25.51%	0.2367 51.30%	0.2299 46.96%	0.2476 58.26%
700	0.2721 12.04%	0.2211 -8.96%	0.2531 4.20%	0.2714 35.71%	0.2224 11.22%	0.2476 23.81%	0.2374 51.74%	0.2299 46.96%	0.2497 59.57%
750	0.2735 12.61%	0.2177 -10.36%	0.2517 3.64%	0.2714 35.71%	0.2245 12.24%	0.2469 23.47%	0.2388 52.61%	0.2272 45.22%	0.2442 56.09%
800	0.2735 12.61%	0.2156 -11.20%	0.2524 3.92%	0.2721 36.05%	0.2218 10.88%	0.2456 22.79%	0.2395 53.04%	0.2313 47.83%	0.2497 59.57%
850	0.2735 12.61%	0.2190 -9.80%	0.2524 3.92%	0.2728 36.39%	0.2190 9.52%	0.2442 22.11%	0.2401 53.48%	0.2279 45.65%	0.2469 57.83%
900	0.2735 12.61%	0.2184 -10.08%	0.2490 2.52%	0.2721 36.05%	0.2190 9.52%	0.2449 22.45%	0.2408 53.91%	0.2279 45.65%	0.2483 58.70%
950	0.2748 13.17%	0.2156 -11.20%	0.2463 1.40%	0.2728 36.39%	0.2190 9.52%	0.2429 21.43%	0.2408 53.91%	0.2272 45.22%	0.2476 58.26%
1000	0.2741 12.89%	0.2129 -12.32%	0.2469 1.68%	0.2721 36.05%	0.2190 9.52%	0.2435 21.77%	0.2408 53.91%	0.2245 43.48%	0.2476 58.26%

Table 7-2: P@30 before and after reranking for different settings (WSJ collection vs. Query 1-50).

non-interpolated average precision (baseline: 0.1520)									
T → N ↓	25	50	75	100	200	300	400	500	600
*	0.1756	0.1824	0.1869	0.1903	0.1951	0.1974	0.1982	0.1988	0.1997
	15.53%	20.00%	22.96%	25.20%	28.36%	29.87%	30.39%	30.79%	31.38%
50	0.1816	0.1924	0.1995	0.2023	0.2094	0.2130	0.2136	0.2142	0.2152
	19.47%	26.58%	31.25%	33.09%	37.76%	40.13%	40.53%	40.92%	41.58%
100	0.1865	0.1980	0.2064	0.2098	0.2162	0.2199	0.2218	0.2222	0.2229
	22.70%	30.26%	35.79%	38.03%	42.24%	44.67%	45.92%	46.18%	46.64%
150	0.1867	0.1993	0.2071	0.2112	0.2178	0.2217	0.2238	0.2241	0.2249
	22.83%	31.12%	36.25%	38.95%	43.29%	45.86%	47.24%	47.43%	47.96%
200	0.1868	0.2001	0.2074	0.2119	0.2189	0.2229	0.2249	0.2248	0.2256
	22.89%	31.64%	36.45%	39.41%	44.01%	46.64%	47.96%	47.89%	48.42%
250	0.1888	0.2017	0.2091	0.2130	0.2205	0.2252	0.2266	0.2266	0.2277
	24.21%	32.70%	37.57%	40.13%	45.07%	48.16%	49.08%	49.08%	49.80%
300	0.1874	0.2020	0.2082	0.2131	0.2207	0.2247	0.2265	0.2270	0.2278
	23.29%	32.89%	36.97%	40.20%	45.20%	47.83%	49.01%	49.34%	49.87%
350	0.1862	0.2024	0.2086	0.2123	0.2200	0.2244	0.2266	0.2269	0.2276
	22.50%	33.16%	37.24%	39.67%	44.74%	47.63%	49.08%	49.28%	49.74%
400	0.1860	0.2023	0.2086	0.2122	0.2194	0.2238	0.2258	0.2265	0.2270
	22.37%	33.09%	37.24%	39.61%	44.34%	47.24%	48.55%	49.01%	49.34%
450	0.1867	0.2019	0.2073	0.2118	0.2199	0.2231	0.2253	0.2257	0.2275
	22.83%	32.83%	36.38%	39.34%	44.67%	46.78%	48.22%	48.49%	49.67%
500	0.1859	0.1985	0.2032	0.2070	0.2149	0.2188	0.2208	0.2213	0.2225
	22.30%	30.59%	33.68%	36.18%	41.38%	43.95%	45.26%	45.59%	46.38%
550	0.1856	0.1987	0.2023	0.2066	0.2140	0.2173	0.2195	0.2202	0.2213
	22.11%	30.72%	33.09%	35.92%	40.79%	42.96%	44.41%	44.87%	45.59%
600	0.1860	0.1982	0.2032	0.2072	0.2140	0.2171	0.2194	0.2202	0.2212
	22.37%	30.39%	33.68%	36.32%	40.79%	42.83%	44.34%	44.87%	45.53%
650	0.1851	0.1977	0.2027	0.2067	0.2135	0.2170	0.2195	0.2197	0.2207
	21.78%	30.07%	33.36%	35.99%	40.46%	42.76%	44.41%	44.54%	45.20%
700	0.1851	0.1978	0.2027	0.2067	0.2136	0.2171	0.2195	0.2197	0.2207
	21.78%	30.13%	33.36%	35.99%	40.53%	42.83%	44.41%	44.54%	45.20%
750	0.1851	0.1978	0.2027	0.2067	0.2136	0.2170	0.2194	0.2196	0.2206
	21.78%	30.13%	33.36%	35.99%	40.53%	42.76%	44.34%	44.47%	45.13%
800	0.1846	0.1973	0.2023	0.2063	0.2129	0.2164	0.2185	0.2188	0.2197
	21.45%	29.80%	33.09%	35.72%	40.07%	42.37%	43.75%	43.95%	44.54%
850	0.1844	0.1974	0.2023	0.2066	0.2132	0.2170	0.2192	0.2196	0.2204
	21.32%	29.87%	33.09%	35.92%	40.26%	42.76%	44.21%	44.47%	45.00%
900	0.1841	0.1970	0.2020	0.2063	0.2128	0.2166	0.2187	0.2191	0.2199
	21.12%	29.61%	32.89%	35.72%	40.00%	42.50%	43.88%	44.14%	44.67%
950	0.1836	0.1970	0.2021	0.2064	0.2129	0.2166	0.2187	0.2192	0.2199
	20.79%	29.61%	32.96%	35.79%	40.07%	42.50%	43.88%	44.21%	44.67%
1000	0.1836	0.1970	0.2021	0.2064	0.2129	0.2167	0.2187	0.2192	0.2199
	20.79%	29.61%	32.96%	35.79%	40.07%	42.57%	43.88%	44.21%	44.67%

* baseline pseudo-feedback runs

Table 7-3: non-interpolated average precision (varying T) (WSJ collection vs. Query 1-50).

		P@20 (baseline: 0.2480)							
T → N ↓	25	50	75	100	200	300	400	500	600
*	0.2786	0.2878	0.2929	0.2918	0.2980	0.2980	0.3000	0.2949	0.2969
	12.34%	16.05%	18.10%	17.66%	20.16%	20.16%	20.97%	18.91%	19.72%
50	0.2959	0.3143	0.3255	0.3276	0.3265	0.3265	0.3306	0.3306	0.3327
	19.31%	26.73%	31.25%	32.10%	31.65%	31.65%	33.31%	33.31%	34.15%
100	0.3031	0.3184	0.3398	0.3388	0.3398	0.3418	0.3429	0.3398	0.3408
	22.22%	28.39%	37.02%	36.61%	37.02%	37.82%	38.27%	37.02%	37.42%
150	0.3092	0.3306	0.3418	0.3459	0.3449	0.3429	0.3459	0.3480	0.3500
	24.68%	33.31%	37.82%	39.48%	39.07%	38.27%	39.48%	40.32%	41.13%
200	0.3122	0.3337	0.3439	0.3500	0.3500	0.3510	0.3510	0.3520	0.3541
	25.89%	34.56%	38.67%	41.13%	41.13%	41.53%	41.53%	41.94%	42.78%
250	0.3153	0.3367	0.3469	0.3551	0.3571	0.3592	0.3571	0.3582	0.3612
	27.14%	35.77%	39.88%	43.19%	43.99%	44.84%	43.99%	44.44%	45.65%
300	0.3133	0.3357	0.3480	0.3541	0.3571	0.3571	0.3561	0.3571	0.3612
	26.33%	35.36%	40.32%	42.78%	43.99%	43.99%	43.59%	43.99%	45.65%
350	0.3122	0.3367	0.3490	0.3551	0.3551	0.3551	0.3551	0.3582	0.3622
	25.89%	35.77%	40.73%	43.19%	43.19%	43.19%	43.19%	44.44%	46.05%
400	0.3102	0.3388	0.3490	0.3531	0.3531	0.3551	0.3571	0.3582	0.3612
	25.08%	36.61%	40.73%	42.38%	42.38%	43.19%	43.99%	44.44%	45.65%
450	0.3133	0.3378	0.3500	0.3520	0.3541	0.3541	0.3561	0.3582	0.3622
	26.33%	36.21%	41.13%	41.94%	42.78%	42.78%	43.59%	44.44%	46.05%
500	0.3102	0.3357	0.3469	0.3469	0.3490	0.3510	0.3520	0.3541	0.3582
	25.08%	35.36%	39.88%	39.88%	40.73%	41.53%	41.94%	42.78%	44.44%
550	0.3102	0.3347	0.3439	0.3459	0.3469	0.3469	0.3490	0.3520	0.3561
	25.08%	34.96%	38.67%	39.48%	39.88%	39.88%	40.73%	41.94%	43.59%
600	0.3122	0.3327	0.3408	0.3449	0.3459	0.3469	0.3490	0.3510	0.3551
	25.89%	34.15%	37.42%	39.07%	39.48%	39.88%	40.73%	41.53%	43.19%
650	0.3112	0.3327	0.3408	0.3429	0.3459	0.3459	0.3490	0.3510	0.3551
	25.48%	34.15%	37.42%	38.27%	39.48%	39.48%	40.73%	41.53%	43.19%
700	0.3112	0.3327	0.3408	0.3429	0.3459	0.3469	0.3500	0.3520	0.3561
	25.48%	34.15%	37.42%	38.27%	39.48%	39.88%	41.13%	41.94%	43.59%
750	0.3112	0.3327	0.3408	0.3418	0.3469	0.3480	0.3500	0.3520	0.3561
	25.48%	34.15%	37.42%	37.82%	39.88%	40.32%	41.13%	41.94%	43.59%
800	0.3112	0.3306	0.3398	0.3408	0.3459	0.3459	0.3459	0.3490	0.3531
	25.48%	33.31%	37.02%	37.42%	39.48%	39.48%	39.48%	40.73%	42.38%
850	0.3082	0.3316	0.3388	0.3408	0.3469	0.3469	0.3459	0.3490	0.3531
	24.27%	33.71%	36.61%	37.42%	39.88%	39.88%	39.48%	40.73%	42.38%
900	0.3071	0.3306	0.3388	0.3429	0.3459	0.3459	0.3449	0.3469	0.3510
	23.83%	33.31%	36.61%	38.27%	39.48%	39.48%	39.07%	39.88%	41.53%
950	0.3061	0.3306	0.3398	0.3439	0.3459	0.3459	0.3449	0.3469	0.3510
	23.43%	33.31%	37.02%	38.67%	39.48%	39.48%	39.07%	39.88%	41.53%
1000	0.3061	0.3306	0.3398	0.3439	0.3459	0.3449	0.3449	0.3469	0.3510
	23.43%	33.31%	37.02%	38.67%	39.48%	39.07%	39.07%	39.88%	41.53%

* baseline pseudo-feedback runs

Table 7-4: P@20 (varying T) (WSJ collection vs. Query 1-50).

non-interpolated average precision (baseline: 0.1520)									
T → β ↓	25	50	75	100	200	300	400	500	600
1	0.1756	0.1824	0.1869	0.1903	0.1951	0.1974	0.1982	0.1988	0.1997
	15.53%	20.00%	22.96%	25.20%	28.36%	29.87%	30.39%	30.79%	31.38%
2	0.1750	0.1838	0.1889	0.1928	0.2002	0.2021	0.2035	0.2040	0.2058
	15.13%	20.92%	24.28%	26.84%	31.71%	32.96%	33.88%	34.21%	35.39%
4	0.1693	0.1798	0.1864	0.1906	0.1988	0.2015	0.2027	0.2033	0.2054
	11.38%	18.29%	22.63%	25.39%	30.79%	32.57%	33.36%	33.75%	35.13%
6	0.1657	0.1760	0.1831	0.1881	0.1972	0.1994	0.2003	0.2015	0.2031
	9.01%	15.79%	20.46%	23.75%	29.74%	31.18%	31.78%	32.57%	33.62%
8	0.1635	0.1743	0.1804	0.1861	0.1955	0.1982	0.1992	0.2004	0.2019
	7.57%	14.67%	18.68%	22.43%	28.62%	30.39%	31.05%	31.84%	32.83%
10	0.1615	0.1728	0.1795	0.1848	0.1940	0.1969	0.1983	0.1991	0.2009
	6.25%	13.68%	18.09%	21.58%	27.63%	29.54%	30.46%	30.99%	32.17%
12	0.1598	0.1717	0.1785	0.1840	0.1930	0.1957	0.1974	0.1980	0.1997
	5.13%	12.96%	17.43%	21.05%	26.97%	28.75%	29.87%	30.26%	31.38%
14	0.1509	0.1708	0.1777	0.1832	0.1920	0.1949	0.1967	0.1972	0.1990
	-0.72%	12.37%	16.91%	20.53%	26.32%	28.22%	29.41%	29.74%	30.92%
16	0.1582	0.1701	0.1771	0.1826	0.1917	0.1945	0.1962	0.1965	0.1979
	4.08%	11.91%	16.51%	20.13%	26.12%	27.96%	29.08%	29.28%	30.20%
18	0.1574	0.1695	0.1765	0.1820	0.1912	0.1942	0.1954	0.1961	0.1974
	3.55%	11.51%	16.12%	19.74%	25.79%	27.76%	28.55%	29.01%	29.87%
20	0.1567	0.1690	0.1762	0.1817	0.1907	0.1938	0.1951	0.1957	0.1972
	3.09%	11.18%	15.92%	19.54%	25.46%	27.50%	28.36%	28.75%	29.74%

Table 7-5: non-interpolated average precision (varying β) (WSJ collection vs. Query 1-50 baseline pseudo-feedback runs).

non-interpolated average precision (baseline: 0.1520)									
$T \rightarrow$ $\beta \downarrow$	25	50	75	100	200	300	400	500	600
1	0.1860 22.37%	0.2023 33.09%	0.2086 37.24%	0.2122 39.61%	0.2194 44.34%	0.2238 47.24%	0.2258 48.55%	0.2265 49.01%	0.2270 49.34%
2	0.1876 23.42%	0.2058 35.39%	0.2123 39.67%	0.2159 42.04%	0.2251 48.09%	0.2283 50.20%	0.2300 51.32%	0.2309 51.91%	0.2301 51.38%
4	0.1794 18.03%	0.2009 32.17%	0.2099 38.09%	0.2124 39.74%	0.2213 45.59%	0.2218 45.92%	0.2229 46.64%	0.2234 46.97%	0.2235 47.04%
6	0.1727 13.62%	0.1960 28.95%	0.2055 35.20%	0.2088 37.37%	0.2167 42.57%	0.2190 44.08%	0.2194 44.34%	0.2194 44.34%	0.2187 43.88%
8	0.1685 10.86%	0.1928 26.84%	0.2034 33.82%	0.2060 35.53%	0.2149 41.38%	0.2167 42.57%	0.2165 42.43%	0.2161 42.17%	0.2161 42.17%
10	0.1657 9.01%	0.1901 25.07%	0.2013 32.43%	0.2041 34.28%	0.2133 40.33%	0.2149 41.38%	0.2147 41.25%	0.2144 41.05%	0.2141 40.86%
12	0.1634 7.50%	0.1883 23.88%	0.1995 31.25%	0.2025 33.22%	0.2120 39.47%	0.2131 40.20%	0.2137 40.59%	0.2135 40.46%	0.2127 39.93%
14	0.1613 6.12%	0.1870 23.03%	0.1981 30.33%	0.2022 33.03%	0.2109 38.75%	0.2122 39.61%	0.2127 39.93%	0.2125 39.80%	0.2117 39.28%
16	0.1601 5.33%	0.1863 22.57%	0.1972 29.74%	0.2018 32.76%	0.2102 38.29%	0.2114 39.08%	0.2123 39.67%	0.2115 39.14%	0.2106 38.55%
18	0.1591 4.67%	0.1855 22.04%	0.1958 28.82%	0.2010 32.24%	0.2096 37.89%	0.2106 38.55%	0.2118 39.34%	0.2108 38.68%	0.2100 38.16%
20	0.1583 4.14%	0.1848 21.58%	0.1954 28.55%	0.2006 31.97%	0.2090 37.50%	0.2099 38.09%	0.2112 38.95%	0.2104 38.42%	0.2101 38.22%

Table 7-6: non-interpolated average precision (varying β) (WSJ collection vs. Query 1-50).

P@20 (baseline: 0.2480)									
T → β ↓	25	50	75	100	200	300	400	500	600
1	0.2786	0.2878	0.2929	0.2918	0.2980	0.2980	0.3000	0.2949	0.2969
	12.34%	16.05%	18.10%	17.66%	20.16%	20.16%	20.97%	18.91%	19.72%
2	0.2857	0.2969	0.3000	0.3010	0.3102	0.3082	0.3061	0.3020	0.3061
	15.20%	19.72%	20.97%	21.37%	25.08%	24.27%	23.43%	21.77%	23.43%
4	0.2837	0.2969	0.3061	0.3122	0.3173	0.3112	0.3112	0.3122	0.3153
	14.40%	19.72%	23.43%	25.89%	27.94%	25.48%	25.48%	25.89%	27.14%
6	0.2786	0.3020	0.3071	0.3133	0.3204	0.3163	0.3184	0.3194	0.3184
	12.34%	21.77%	23.83%	26.33%	29.19%	27.54%	28.39%	28.79%	28.39%
8	0.2786	0.3031	0.3082	0.3122	0.3224	0.3173	0.3204	0.3204	0.3194
	12.34%	22.22%	24.27%	25.89%	30.00%	27.94%	29.19%	29.19%	28.79%
10	0.2776	0.3010	0.3061	0.3092	0.3194	0.3173	0.3173	0.3184	0.3194
	11.94%	21.37%	23.43%	24.68%	28.79%	27.94%	27.94%	28.39%	28.79%
12	0.2714	0.2959	0.3051	0.3092	0.3163	0.3184	0.3173	0.3194	0.3214
	9.44%	19.31%	23.02%	24.68%	27.54%	28.39%	27.94%	28.79%	29.60%
14	0.2704	0.2980	0.3061	0.3102	0.3153	0.3184	0.3194	0.3184	0.3204
	9.03%	20.16%	23.43%	25.08%	27.14%	28.39%	28.79%	28.39%	29.19%
16	0.2684	0.2990	0.3061	0.3092	0.3153	0.3184	0.3184	0.3184	0.3224
	8.23%	20.56%	23.43%	24.68%	27.14%	28.39%	28.39%	28.39%	30.00%
18	0.2694	0.2949	0.3061	0.3102	0.3143	0.3194	0.3194	0.3173	0.3235
	8.63%	18.91%	23.43%	25.08%	26.73%	28.79%	28.79%	27.94%	30.44%
20	0.2694	0.2949	0.3082	0.3082	0.3153	0.3184	0.3194	0.3194	0.3245
	8.63%	18.91%	24.27%	24.27%	27.14%	28.39%	28.79%	28.79%	30.85%

Table 7-7: P@20 (varying β) (WSJ collection vs. Query 1-50 baseline pseudo-feedback runs).

P@20 (baseline: 0.2480)									
T → β ↓	25	50	75	100	200	300	400	500	600
1	0.3102 25.08%	0.3388 36.61%	0.3490 40.73%	0.3531 42.38%	0.3531 42.38%	0.3551 43.19%	0.3571 43.99%	0.3582 44.44%	0.3612 45.65%
2	0.3214 29.60%	0.3500 41.13%	0.3673 48.10%	0.3663 47.70%	0.3724 50.16%	0.3765 51.81%	0.3797 53.10%	0.3786 52.66%	0.3796 53.06%
4	0.3204 29.19%	0.3480 40.32%	0.3633 46.49%	0.3663 47.70%	0.3684 48.55%	0.3694 48.95%	0.3796 53.06%	0.3806 53.47%	0.3837 54.72%
6	0.3143 26.73%	0.3449 39.07%	0.3571 43.99%	0.3684 48.55%	0.3714 49.76%	0.3776 52.26%	0.3796 53.06%	0.3786 52.66%	0.3847 55.12%
8	0.3102 25.08%	0.3418 37.82%	0.3571 43.99%	0.3663 47.70%	0.3745 51.01%	0.3796 53.06%	0.3816 53.87%	0.3827 54.31%	0.3827 54.31%
10	0.3041 22.62%	0.3408 37.42%	0.3551 43.19%	0.3694 48.95%	0.3714 49.76%	0.3816 53.87%	0.3776 52.26%	0.3786 52.66%	0.3816 53.87%
12	0.3010 21.37%	0.3378 36.21%	0.3561 43.59%	0.3673 48.10%	0.3704 49.35%	0.3806 53.47%	0.3765 51.81%	0.3776 52.26%	0.3806 53.47%
14	0.3010 21.37%	0.3388 36.61%	0.3551 43.19%	0.3653 47.30%	0.3704 49.35%	0.3796 53.06%	0.3765 51.81%	0.3776 52.26%	0.3806 53.47%
16	0.2980 20.16%	0.3367 35.77%	0.3571 43.99%	0.3633 46.49%	0.3694 48.95%	0.3796 53.06%	0.3776 52.26%	0.3786 52.66%	0.3786 52.66%
18	0.2980 20.16%	0.3347 34.96%	0.3551 43.19%	0.3612 45.65%	0.3684 48.55%	0.3786 52.66%	0.3796 53.06%	0.3796 53.06%	0.3806 53.47%
20	0.2980 20.16%	0.3337 34.56%	0.3551 43.19%	0.3612 45.65%	0.3673 48.10%	0.3786 52.66%	0.3806 53.47%	0.3765 51.81%	0.3806 53.47%

Table 7-8: P@20 (varying β) (WSJ collection vs. Query 1-50).

8 Appendix II

P@20									
	<i>Inc.ltc</i> (0.2700)			<i>ltc.ltc</i> (0.2140)			<i>atc.atc</i> (0.1770)		
N ↓	1	2	3	1	2	3	1	2	3
50	0.3600	0.3630	0.3620	0.3280	0.3240	0.3150	0.2580	0.2600	0.2590
	33.33%	34.44%	34.07%	53.27%	51.40%	47.20%	45.76%	46.89%	46.33%
100	0.3930	0.4030	0.3970	0.3740	0.3670	0.3620	0.2890	0.3050	0.2860
	45.56%	49.26%	47.04%	74.77%	71.50%	69.16%	63.28%	72.32%	61.58%
150	0.4050	0.4160	0.3980	0.3970	0.3940	0.3840	0.3080	0.3170	0.3120
	50.00%	54.07%	47.41%	85.51%	84.11%	79.44%	74.01%	79.10%	76.27%
200	0.4180	0.4120	0.3970	0.4050	0.3950	0.3840	0.3280	0.3440	0.3300
	54.81%	52.59%	47.04%	89.25%	84.58%	79.44%	85.31%	94.35%	86.44%
250	0.4150	0.4060	0.3910	0.4140	0.3930	0.3720	0.3440	0.3440	0.3310
	53.70%	50.37%	44.81%	93.47%	83.64%	73.83%	94.35%	94.35%	87.01%
300	0.4190	0.3980	0.3900	0.4180	0.3870	0.3680	0.3490	0.3510	0.3470
	55.19%	47.41%	44.44%	95.33%	80.84%	71.96%	97.18%	98.31%	96.05%
350	0.4220	0.4000	0.3850	0.4230	0.3890	0.3620	0.3520	0.3590	0.3530
	56.30%	48.15%	42.59%	97.66%	81.78%	69.16%	98.87%	102.82%	99.43%
400	0.4210	0.3900	0.3760	0.4240	0.3760	0.3600	0.3550	0.3670	0.3570
	55.93%	44.44%	39.26%	98.13%	75.70%	68.22%	100.56%	107.34%	101.69%
450	0.4230	0.3880	0.3730	0.4300	0.3790	0.3600	0.3590	0.3780	0.3630
	56.67%	43.70%	38.15%	100.93%	77.10%	68.22%	102.82%	113.56%	105.08%
500	0.4240	0.3860	0.3700	0.4300	0.3630	0.3540	0.3630	0.3770	0.3640
	57.04%	42.96%	37.04%	100.93%	69.63%	65.42%	105.08%	112.99%	105.65%
550	0.4260	0.3800	0.3660	0.4300	0.3640	0.3510	0.3650	0.3810	0.3670
	57.78%	40.74%	35.56%	100.93%	70.09%	64.02%	106.21%	115.25%	107.34%
600	0.4270	0.3710	0.3590	0.4310	0.3620	0.3370	0.3670	0.3860	0.3630
	58.15%	37.41%	32.96%	101.40%	69.16%	57.48%	107.34%	118.08%	105.08%
650	0.4300	0.3610	0.3560	0.4290	0.3540	0.3300	0.3670	0.3850	0.3650
	59.26%	33.70%	31.85%	100.47%	65.42%	54.20%	107.34%	117.51%	106.21%
700	0.4280	0.3580	0.3550	0.4280	0.3540	0.3320	0.3660	0.3800	0.3630
	58.52%	32.59%	31.48%	100.00%	65.42%	55.14%	106.78%	114.69%	105.08%
750	0.4300	0.3620	0.3520	0.4270	0.3550	0.3330	0.3690	0.3760	0.3600
	59.26%	34.07%	30.37%	99.53%	65.89%	55.61%	108.47%	112.43%	103.39%
800	0.4300	0.3530	0.3520	0.4270	0.3490	0.3320	0.3690	0.3770	0.3660
	59.26%	30.74%	30.37%	99.53%	63.08%	55.14%	108.47%	112.99%	106.78%
850	0.4320	0.3480	0.3520	0.4290	0.3540	0.3330	0.3690	0.3820	0.3630
	60.00%	28.89%	30.37%	100.47%	65.42%	55.61%	108.47%	115.82%	105.08%
900	0.4310	0.3410	0.3440	0.4270	0.3500	0.3290	0.3710	0.3750	0.3590
	59.63%	26.30%	27.41%	99.53%	63.55%	53.74%	109.60%	111.86%	102.82%
950	0.4300	0.3380	0.3410	0.4280	0.3460	0.3240	0.3710	0.3690	0.3560
	59.26%	25.19%	26.30%	100.00%	61.68%	51.40%	109.60%	108.47%	101.13%
1000	0.4290	0.3350	0.3370	0.4280	0.3400	0.3230	0.3710	0.3710	0.3540
	58.89%	24.07%	24.81%	100.00%	58.88%	50.93%	109.60%	109.60%	100.00%

Table 8-1: P@20 before and after reranking for different settings (TREC-3 ad-hoc task).

P@30									
	<i>Inc.ltc</i> (0.2640)			<i>ltc.ltc</i> (0.2127)			<i>atc.atc</i> (0.1707)		
N ↓	1	2	3	1	2	3	1	2	3
50	0.3227 22.22%	0.3253 23.23%	0.3207 21.46%	0.2767 30.09%	0.2800 31.66%	0.2760 29.78%	0.2193 28.52%	0.2227 30.47%	0.2133 25.00%
100	0.3607 36.62%	0.3587 35.86%	0.3567 35.10%	0.3380 58.93%	0.3340 57.05%	0.3233 52.04%	0.2660 55.86%	0.2653 55.47%	0.2580 51.17%
150	0.3847 45.71%	0.3780 43.18%	0.3727 47.16%	0.3580 68.34%	0.3540 66.46%	0.3500 64.58%	0.2800 64.06%	0.2933 71.88%	0.2833 66.02%
200	0.3960 50.00%	0.3807 44.19%	0.3653 38.38%	0.3667 72.41%	0.3610 69.91%	0.3507 64.89%	0.2973 74.22%	0.3140 83.98%	0.3013 76.56%
250	0.3947 49.49%	0.3740 41.67%	0.3573 35.35%	0.3747 76.18%	0.3660 72.41%	0.3533 66.14%	0.3093 81.25%	0.3213 88.28%	0.3100 81.64%
300	0.3953 49.75%	0.3687 39.65%	0.3547 34.34%	0.3807 79.00%	0.3627 70.53%	0.3473 63.32%	0.3120 82.81%	0.3300 93.36%	0.3127 83.20%
350	0.3993 51.26%	0.3753 42.17%	0.3627 37.37%	0.3907 83.70%	0.3680 73.04%	0.3467 63.01%	0.3207 87.89%	0.3300 93.36%	0.3233 89.45%
400	0.4020 52.27%	0.3700 40.15%	0.3553 34.60%	0.3947 85.58%	0.3607 69.59%	0.3440 61.76%	0.3253 90.63%	0.3367 97.27%	0.3260 91.02%
450	0.4000 51.52%	0.3653 38.38%	0.3520 33.33%	0.3987 87.46%	0.3567 67.71%	0.3433 61.44%	0.3320 94.53%	0.3407 99.61%	0.3360 96.87%
500	0.4013 52.02%	0.3620 37.12%	0.3473 31.57%	0.4000 88.09%	0.3447 62.07%	0.3340 57.05%	0.3360 96.88%	0.3440 101.56%	0.3327 94.92%
550	0.4033 52.78%	0.3547 34.34%	0.3467 31.31%	0.4000 88.09%	0.3433 61.44%	0.3307 55.49%	0.3387 98.44%	0.3480 103.91%	0.3373 97.66%
600	0.4053 53.54%	0.3507 32.83%	0.3427 29.80%	0.4033 89.66%	0.3380 53.98%	0.3213 51.10%	0.3407 99.61%	0.3547 107.81%	0.3413 100.00%
650	0.4060 53.79%	0.3453 30.81%	0.3373 27.78%	0.4020 89.03%	0.3300 55.17%	0.3167 48.90%	0.3420 100.39%	0.3540 107.42%	0.3400 99.22%
700	0.4060 53.79%	0.3393 28.54%	0.3353 27.02%	0.4013 88.71%	0.3307 55.49%	0.3147 47.96%	0.3440 101.56%	0.3513 105.86%	0.3360 96.88%
750	0.4073 54.29%	0.3407 29.04%	0.3360 27.27%	0.4020 89.03%	0.3333 56.74%	0.3153 48.28%	0.3420 100.39%	0.3547 107.81%	0.3313 94.14%
800	0.4067 54.04%	0.3353 27.02%	0.3347 26.77%	0.4007 88.40%	0.3320 56.11%	0.3160 48.59%	0.3420 100.39%	0.3553 108.20%	0.3327 94.92%
850	0.4067 54.04%	0.3347 26.77%	0.3347 26.77%	0.4027 89.34%	0.3320 56.11%	0.3153 48.28%	0.3413 100.00%	0.3533 107.03%	0.3367 97.27%
900	0.4053 53.54%	0.3347 26.77%	0.3333 26.26%	0.4007 88.40%	0.3293 54.86%	0.3113 46.39%	0.3427 100.78%	0.3480 103.91%	0.3373 97.66%
950	0.4047 53.28%	0.3280 24.24%	0.3360 27.27%	0.4007 88.40%	0.3247 52.66%	0.3080 44.83%	0.3433 101.17%	0.3487 104.30%	0.3387 98.44%
1000	0.4047 53.28%	0.3220 21.97%	0.3320 25.76%	0.3987 87.46%	0.3220 51.41%	0.3087 45.14%	0.3453 102.34%	0.3467 103.12%	0.3400 99.22%

Table 8-2: P@30 before and after reranking for different settings (TREC-3 ad-hoc task).

non-interpolated average precision (baseline: 0.1320)									
T → N ↓	25	50	75	100	200	300	400	500	600
*	0.0509	0.1859	0.1914	0.1937	0.1997	0.2026	0.2035	0.2043	0.2055
	-61.44%	40.83%	45.00%	46.74%	51.29%	53.48%	54.17%	54.77%	55.68%
50	0.2064	0.2215	0.2295	0.2352	0.2437	0.2466	0.2485	0.2500	0.2504
	56.36%	67.80%	73.86%	78.18%	84.62%	86.82%	88.26%	89.39%	89.70%
100	0.2149	0.2328	0.2404	0.2475	0.2566	0.2617	0.2640	0.2653	0.2660
	62.80%	76.36%	82.12%	87.50%	94.39%	98.26%	100.00%	100.98%	101.52%
150	0.2169	0.2357	0.2444	0.2508	0.2611	0.2663	0.2685	0.2699	0.2705
	64.32%	78.56%	85.15%	90.00%	97.80%	101.74%	103.41%	104.47%	104.92%
200	0.2188	0.2360	0.2461	0.2534	0.2644	0.2695	0.2723	0.2738	0.2745
	65.76%	78.79%	86.44%	91.97%	100.30%	104.17%	106.29%	107.42%	107.95%
250	0.2189	0.2372	0.2492	0.2555	0.2657	0.2706	0.2732	0.2746	0.2752
	65.83%	79.70%	88.79%	93.56%	101.29%	105.00%	106.97%	108.03%	108.48%
300	0.2189	0.2371	0.2499	0.2562	0.2676	0.2727	0.2748	0.2762	0.2762
	65.83%	79.62%	89.32%	94.09%	102.73%	106.59%	108.18%	109.24%	109.24%
350	0.2193	0.2379	0.2483	0.2565	0.2672	0.2716	0.2743	0.2759	0.2767
	66.14%	80.23%	88.11%	94.32%	102.42%	105.76%	107.80%	109.02%	109.62%
400	0.2186	0.2367	0.2482	0.2547	0.2662	0.2706	0.2732	0.2747	0.2754
	65.61%	79.32%	88.03%	92.95%	101.67%	105.00%	106.97%	108.11%	108.64%
450	0.2192	0.2373	0.2485	0.2551	0.2666	0.2712	0.2736	0.2753	0.2760
	66.06%	79.77%	88.26%	93.26%	101.97%	105.45%	107.27%	108.56%	109.09%
500	0.2188	0.2370	0.2478	0.2552	0.2670	0.2713	0.2735	0.2752	0.2758
	65.76%	79.55%	87.73%	93.33%	102.27%	105.53%	107.20%	108.48%	108.94%
550	0.2187	0.2380	0.2496	0.2567	0.2691	0.2734	0.2752	0.2774	0.2781
	65.68%	80.30%	89.09%	94.47%	103.86%	107.12%	108.48%	110.15%	110.68%
600	0.2185	0.2378	0.2494	0.2566	0.2689	0.2734	0.2751	0.2771	0.2779
	65.53%	80.15%	88.94%	94.39%	103.71%	107.12%	108.41%	109.92%	110.53%
650	0.2187	0.2381	0.2501	0.2572	0.2694	0.2741	0.2757	0.2779	0.2790
	65.68%	80.38%	89.47%	94.85%	104.09%	107.65%	108.86%	110.53%	111.36%
700	0.2191	0.2380	0.2505	0.2568	0.2690	0.2740	0.2757	0.2775	0.2787
	65.98%	80.30%	89.77%	94.55%	103.79%	107.58%	108.86%	110.23%	111.14%
750	0.2193	0.2383	0.2504	0.2570	0.2698	0.2744	0.2764	0.2781	0.2792
	66.14%	80.53%	89.70%	94.70%	104.39%	107.88%	109.39%	110.68%	111.52%
800	0.2197	0.2393	0.2510	0.2582	0.2713	0.2760	0.2780	0.2797	0.2809
	66.44%	81.29%	90.15%	95.61%	105.53%	109.09%	110.61%	111.89%	112.80%
850	0.2196	0.2398	0.2516	0.2592	0.2715	0.2760	0.2789	0.2802	0.2812
	66.36%	81.67%	90.61%	96.36%	105.68%	109.09%	111.29%	112.27%	113.03%
900	0.2195	0.2396	0.2515	0.2588	0.2716	0.2759	0.2787	0.2799	0.2807
	66.29%	81.52%	90.53%	96.06%	105.76%	109.02%	111.14%	112.05%	112.65%
950	0.2195	0.2394	0.2510	0.2588	0.2712	0.2757	0.2784	0.2797	0.2806
	66.29%	81.36%	90.15%	96.06%	105.45%	108.86%	110.91%	111.89%	112.58%
1000	0.2197	0.2393	0.2508	0.2588	0.2708	0.2755	0.2783	0.2797	0.2804
	66.44%	81.29%	90.00%	96.06%	105.15%	108.71%	110.83%	111.89%	112.42%

* baseline pseudo-feedback runs

Table 8-3: non-interpolated average precision (varying T) (TREC-3 ad-hoc task).

P@20 (baseline: 0.2700)									
T → N ↓	25	50	75	100	200	300	400	500	600
*	0.2730 1.11%	0.3160 17.04%	0.3220 19.26%	0.3170 17.41%	0.3230 19.63%	0.3210 18.89%	0.3180 17.78%	0.3180 17.78%	0.3190 18.15%
50	0.3690 36.67%	0.3860 42.96%	0.3940 45.93%	0.3950 46.30%	0.4060 50.37%	0.4060 50.37%	0.4080 51.11%	0.4110 52.22%	0.4070 50.74%
100	0.3880 43.70%	0.4140 53.33%	0.4110 52.22%	0.4210 55.93%	0.4300 59.26%	0.4350 61.11%	0.4360 61.48%	0.4370 61.85%	0.4350 61.11%
150	0.3910 44.81%	0.4150 53.70%	0.4250 57.41%	0.4320 60.00%	0.4430 64.07%	0.4490 66.30%	0.4490 66.30%	0.4480 65.93%	0.4490 66.30%
200	0.4000 48.15%	0.4170 54.44%	0.4290 58.89%	0.4370 61.85%	0.4470 65.56%	0.4520 67.41%	0.4540 68.15%	0.4530 67.78%	0.4560 68.89%
250	0.3980 47.41%	0.4180 54.81%	0.4290 58.89%	0.4360 61.48%	0.4470 65.56%	0.4480 65.93%	0.4510 67.04%	0.4490 66.30%	0.4530 67.78%
300	0.3960 46.67%	0.4220 56.30%	0.4330 60.37%	0.4400 62.96%	0.4500 66.67%	0.4530 67.78%	0.4570 69.26%	0.4570 69.26%	0.4630 71.48%
350	0.3980 47.41%	0.4270 58.15%	0.4350 61.11%	0.4460 65.19%	0.4500 66.67%	0.4580 69.63%	0.4610 70.74%	0.4640 71.85%	0.4680 73.33%
400	0.4000 48.15%	0.4280 58.52%	0.4410 63.33%	0.4500 66.67%	0.4560 68.89%	0.4630 71.48%	0.4670 72.96%	0.4690 73.70%	0.4700 74.07%
450	0.4030 49.26%	0.4290 58.89%	0.4410 63.33%	0.4520 67.41%	0.4590 70.00%	0.4640 71.85%	0.4670 72.96%	0.4680 73.33%	0.4720 74.81%
500	0.3990 47.78%	0.4280 58.52%	0.4370 61.85%	0.4480 65.93%	0.4610 70.74%	0.4670 72.96%	0.4660 72.59%	0.4710 74.44%	0.4750 75.93%
550	0.3960 46.67%	0.4300 59.26%	0.4400 62.96%	0.4480 65.93%	0.4680 73.33%	0.4740 75.56%	0.4710 74.44%	0.4750 75.93%	0.4780 77.04%
600	0.3950 46.30%	0.4290 58.89%	0.4380 62.22%	0.4480 65.93%	0.4700 74.07%	0.4770 76.67%	0.4720 74.81%	0.4750 75.93%	0.4790 77.41%
650	0.3960 46.67%	0.4310 59.63%	0.4380 62.22%	0.4460 65.19%	0.4660 72.59%	0.4770 76.67%	0.4750 75.93%	0.4780 77.04%	0.4810 78.15%
700	0.4000 48.15%	0.4320 60.00%	0.4400 62.96%	0.4480 65.93%	0.4670 72.96%	0.4760 76.30%	0.4780 77.04%	0.4790 77.41%	0.4830 78.89%
750	0.4010 48.52%	0.4290 58.89%	0.4380 62.22%	0.4480 65.93%	0.4670 72.96%	0.4740 75.56%	0.4750 75.93%	0.4760 76.30%	0.4820 78.52%
800	0.4020 48.89%	0.4330 60.37%	0.4440 64.44%	0.4530 67.78%	0.4700 74.07%	0.4780 77.04%	0.4780 77.04%	0.4780 77.04%	0.4850 79.63%
850	0.4020 48.89%	0.4330 60.37%	0.4450 64.81%	0.4560 68.89%	0.4700 74.07%	0.4770 76.67%	0.4820 78.52%	0.4840 79.26%	0.4890 81.11%
900	0.4010 48.52%	0.4350 61.11%	0.4470 65.56%	0.4560 68.89%	0.4730 75.19%	0.4780 77.04%	0.4830 78.89%	0.4860 80.00%	0.4900 81.48%
950	0.4040 49.63%	0.4350 61.11%	0.4470 65.56%	0.4560 68.89%	0.4740 75.56%	0.4750 75.93%	0.4810 78.15%	0.4850 79.63%	0.4900 81.48%
1000	0.4050 50.00%	0.4310 59.63%	0.4450 64.81%	0.4560 68.89%	0.4730 75.19%	0.4740 75.56%	0.4790 77.41%	0.4830 78.89%	0.4880 80.74%

* baseline pseudo-feedback runs

Table 8-4: P@20 (varying T) (TREC-3 ad-hoc task).

non-interpolated average precision (baseline: 0.1320)									
T → β ↓	25	50	75	100	200	300	400	500	600
1	0.0509	0.1859	0.1914	0.1937	0.1997	0.2026	0.2035	0.2043	0.2055
2	-61.44%	40.83%	45.00%	46.74%	51.29%	53.48%	54.17%	54.77%	55.68%
4	0.1882	0.1986	0.2043	0.2076	0.2141	0.2175	0.2184	0.2195	0.2207
6	42.58%	50.45%	54.77%	57.27%	62.20%	64.77%	65.45%	66.29%	67.20%
8	0.1935	0.2042	0.2104	0.2134	0.2204	0.2243	0.2252	0.2260	0.2272
10	46.59%	54.70%	59.39%	61.67%	66.97%	69.92%	70.61%	71.21%	72.12%
12	0.1940	0.2052	0.2116	0.2145	0.2212	0.2252	0.2262	0.2271	0.2282
14	46.97%	55.45%	60.30%	62.50%	67.58%	70.61%	71.36%	72.05%	72.88%
16	0.1941	0.2051	0.2118	0.2143	0.2211	0.2252	0.2261	0.2271	0.2284
18	47.05%	55.38%	60.45%	62.35%	67.50%	70.61%	71.29%	72.05%	73.03%
20	0.1938	0.2047	0.2116	0.2142	0.2208	0.2249	0.2259	0.2269	0.2281
	46.82%	55.08%	60.30%	62.27%	67.27%	70.38%	71.14%	71.89%	72.80%
	0.1933	0.2045	0.2113	0.2140	0.2205	0.2246	0.2256	0.2266	0.2277
	46.44%	54.92%	60.08%	62.12%	67.05%	70.15%	70.91%	71.67%	72.50%
	0.1929	0.2043	0.2109	0.2138	0.2203	0.2243	0.2254	0.2263	0.2274
	46.14%	54.77%	59.77%	61.97%	66.89%	69.92%	70.76%	71.44%	72.27%
	0.1927	0.2041	0.2108	0.2136	0.2202	0.2242	0.2252	0.2260	0.2270
	45.98%	54.62%	59.70%	61.82%	66.82%	69.85%	70.61%	71.21%	71.97%
	0.1925	0.2037	0.2106	0.2134	0.2200	0.2241	0.2249	0.2258	0.2268
	45.83%	54.32%	59.55%	61.67%	66.67%	69.77%	70.38%	71.06%	71.82%
	0.1924	0.2035	0.2102	0.2132	0.2198	0.2239	0.2246	0.2255	0.2267
	45.76%	54.17%	59.24%	61.52%	66.52%	69.62%	70.15%	70.83%	71.74%

Table 8-5: non-interpolated average precision (varying β) (TREC-3 ad-hoc task baseline pseudo-feedback runs).

non-interpolated average precision (baseline: 0.1320)									
T → β ↓	25	50	75	100	200	300	400	500	600
1	0.2197	0.2393	0.2510	0.2582	0.2713	0.2760	0.2780	0.2797	0.2809
	66.44%	81.29%	90.15%	95.61%	105.53%	109.09%	110.61%	111.89%	112.80%
2	0.2425	0.2643	0.2764	0.2836	0.2950	0.2994	0.2999	0.3012	0.3012
	83.71%	100.23%	109.39%	114.85%	123.48%	126.82%	127.20%	128.18%	128.18%
4	0.2512	0.2718	0.2836	0.2889	0.2976	0.2993	0.2989	0.2984	0.2985
	90.30%	105.91%	114.85%	118.86%	125.45%	126.74%	126.44%	126.06%	126.14%
6	0.2508	0.2696	0.2812	0.2864	0.2940	0.2953	0.2936	0.2928	0.2927
	90.00%	104.24%	113.03%	116.97%	122.73%	123.71%	122.42%	121.82%	121.74%
8	0.2492	0.2675	0.2788	0.2839	0.2911	0.2914	0.2902	0.2890	0.2883
	88.79%	102.65%	111.21%	115.08%	120.53%	120.76%	119.85%	118.94%	118.41%
10	0.2476	0.2656	0.2762	0.2821	0.2888	0.2887	0.2874	0.2860	0.2853
	87.58%	101.21%	109.24%	113.71%	118.79%	118.71%	117.73%	116.67%	116.14%
12	0.2461	0.2639	0.2747	0.2804	0.2871	0.2867	0.2852	0.2836	0.2832
	86.44%	99.92%	108.11%	112.42%	117.50%	117.20%	116.06%	114.85%	114.55%
14	0.2447	0.2623	0.2731	0.2788	0.2857	0.2851	0.2836	0.2822	0.2814
	85.38%	98.71%	106.89%	111.21%	116.44%	115.98%	114.85%	113.79%	113.18%
16	0.2435	0.2611	0.2721	0.2778	0.2846	0.2839	0.2823	0.2809	0.2800
	84.47%	97.80%	106.14%	110.45%	115.61%	115.08%	113.86%	112.80%	112.12%
18	0.2426	0.2601	0.2712	0.2770	0.2834	0.2830	0.2814	0.2798	0.2790
	83.79%	97.05%	105.45%	109.85%	114.70%	114.39%	113.18%	111.97%	111.36%
20	0.2419	0.2594	0.2706	0.2764	0.2825	0.2822	0.2805	0.2791	0.2782
	83.26%	96.52%	105.00%	109.39%	114.02%	113.79%	112.50%	111.44%	110.76%

Table 8-6: non-interpolated average precision (varying β) (TREC-3 ad-hoc task).

P@20 (baseline: 0.2700)									
T → β ↓	25	50	75	100	200	300	400	500	600
1	0.2730	0.3180	0.3220	0.3170	0.3230	0.3210	0.3180	0.3180	0.3190
	1.11%	17.78%	19.26%	17.41%	19.63%	18.89%	17.78%	17.78%	18.15%
2	0.3130	0.3180	0.3280	0.3290	0.3300	0.3300	0.3310	0.3330	0.3330
	15.93%	17.78%	21.48%	21.85%	22.22%	22.22%	22.59%	23.33%	23.33%
4	0.3210	0.3270	0.3370	0.3290	0.3380	0.3450	0.3450	0.3430	0.3480
	18.89%	21.11%	24.81%	21.85%	25.19%	27.78%	27.78%	27.04%	28.89%
6	0.3240	0.3280	0.3350	0.3330	0.3380	0.3500	0.3490	0.3470	0.3480
	20.00%	21.48%	24.07%	23.33%	25.19%	29.63%	29.26%	28.52%	28.89%
8	0.3270	0.3310	0.3380	0.3350	0.3410	0.3500	0.3450	0.3440	0.3490
	21.11%	22.59%	25.19%	24.07%	26.30%	29.63%	27.78%	27.41%	29.26%
10	0.3310	0.3300	0.3410	0.3360	0.3430	0.3520	0.3470	0.3440	0.3530
	22.59%	22.22%	26.30%	24.44%	27.04%	30.37%	28.52%	27.41%	30.74%
12	0.3300	0.3280	0.3410	0.3380	0.3420	0.3490	0.3460	0.3460	0.3520
	22.22%	21.48%	26.30%	25.19%	26.67%	29.26%	28.15%	28.15%	30.37%
14	0.3310	0.3280	0.3430	0.3390	0.3410	0.3510	0.3480	0.3480	0.3530
	22.59%	21.48%	27.04%	25.56%	26.30%	30.00%	28.89%	28.89%	30.74%
16	0.3330	0.3290	0.3430	0.3390	0.3400	0.3530	0.3480	0.3480	0.3540
	23.33%	21.85%	27.04%	25.56%	25.93%	30.74%	28.89%	28.89%	31.11%
18	0.3340	0.3290	0.3440	0.3390	0.3390	0.3530	0.3510	0.3490	0.3540
	23.70%	21.85%	27.41%	25.56%	25.56%	30.74%	30.00%	29.26%	31.11%
20	0.3350	0.3280	0.3440	0.3380	0.3400	0.3550	0.3500	0.3480	0.3530
	24.07%	21.48%	27.41%	25.19%	25.93%	31.48%	29.63%	28.89%	30.74%

Table 8-7: P@20 (varying β) (TREC-3 ad-hoc task baseline pseudo-feedback runs).

P@20 (baseline: 0.2700)									
T → β ↓	25	50	75	100	200	300	400	500	600
1	0.4020	0.4330	0.4440	0.4530	0.4573	0.4780	0.4780	0.4780	0.4850
	48.89%	60.37%	64.44%	67.78%	69.37%	77.04%	77.04%	77.04%	79.63%
2	0.4400	0.4720	0.4850	0.4960	0.5100	0.5070	0.5080	0.5080	0.5070
	62.96%	74.81%	79.63%	83.70%	88.89%	87.78%	88.15%	88.15%	87.78%
4	0.4650	0.4860	0.5020	0.5080	0.5220	0.5200	0.5190	0.5180	0.5220
	72.22%	80.00%	85.93%	88.15%	93.33%	92.59%	92.22%	91.85%	93.33%
6	0.4720	0.4820	0.5010	0.5150	0.5190	0.5200	0.5230	0.5240	0.5220
	74.81%	78.52%	85.56%	90.74%	92.22%	92.59%	93.70%	94.07%	93.33%
8	0.4750	0.4770	0.5000	0.5150	0.5210	0.5250	0.5240	0.5210	0.5230
	75.93%	76.67%	85.19%	90.74%	92.96%	94.44%	94.07%	92.96%	93.70%
10	0.4740	0.4780	0.4980	0.5130	0.5170	0.5220	0.5250	0.5230	0.5220
	75.56%	77.04%	84.44%	90.00%	91.48%	93.33%	94.44%	93.70%	93.33%
12	0.4730	0.4740	0.4980	0.5100	0.5170	0.5190	0.5200	0.5200	0.5200
	75.19%	75.56%	84.44%	88.89%	91.48%	92.22%	92.59%	92.59%	92.59%
14	0.4710	0.4720	0.4970	0.5070	0.5150	0.5220	0.5190	0.5210	0.5180
	74.44%	74.81%	84.07%	87.78%	90.74%	93.33%	92.22%	92.96%	91.85%
16	0.4710	0.4740	0.4970	0.5070	0.5140	0.5210	0.5190	0.5210	0.5160
	74.44%	75.56%	84.07%	87.78%	90.37%	92.96%	92.22%	92.96%	91.11%
18	0.4700	0.4730	0.4962	0.5060	0.5170	0.5200	0.5200	0.5180	0.5160
	74.07%	75.19%	83.78%	87.41%	91.48%	92.59%	92.59%	91.85%	91.11%
20	0.4680	0.4710	0.4980	0.5040	0.5170	0.5200	0.5190	0.5180	0.5150
	73.33%	74.44%	84.44%	86.67%	91.48%	92.59%	92.22%	91.85%	90.74%

Table 8-8: P@20 (varying β) (TREC-3 ad-hoc task).

9 Appendix III

P@20									
	<i>Inc.ltc</i> (0.2500)			<i>ltc.ltc</i> (0.2280)			<i>atc.atc</i> (0.2070)		
N↓	1	2	3	1	2	3	1	2	3
50	0.3030	0.2480	0.2430	0.2740	0.2520	0.2350	0.2460	0.2310	0.2230
	21.20%	-0.80%	-2.80%	20.18%	10.53%	3.07%	18.84%	11.59%	7.73%
100	0.3200	0.2270	0.2250	0.2880	0.2109	0.2150	0.2670	0.2220	0.2200
	28.00%	-9.20%	-10.00%	26.32%	-3.95%	-5.07%	28.99%	7.25%	6.28%
150	0.3250	0.2270	0.2160	0.2940	0.2120	0.2070	0.2730	0.2140	0.2170
	30.00%	-9.20%	-13.60%	28.95%	-7.02%	-9.21%	31.88%	3.38%	4.83%
200	0.3260	0.2260	0.2130	0.2970	0.2130	0.2040	0.2740	0.2060	0.2020
	30.40%	-9.60%	-14.80%	30.26%	-6.58%	-10.53%	32.37%	-0.48%	-2.42%
250	0.3350	0.2190	0.2160	0.2990	0.2140	0.2040	0.2740	0.2040	0.1930
	34.00%	-12.40%	-13.60%	31.14%	-6.14%	-10.53%	32.37%	-1.45%	-6.76%
300	0.3390	0.2150	0.2130	0.3010	0.2110	0.2020	0.2740	0.2020	0.1910
	35.60%	-14.00%	-14.80%	32.02%	-7.46%	-11.40%	32.37%	-2.42%	-7.73%
350	0.3340	0.2080	0.2130	0.3020	0.2090	0.2020	0.2730	0.2020	0.1890
	33.60%	-16.80%	-14.80%	32.46%	-8.33%	-11.40%	31.88%	-2.42%	-8.70%
400	0.3330	0.2070	0.2070	0.3000	0.2010	0.2020	0.2740	0.2010	0.1930
	33.20%	-17.20%	-17.20%	31.58%	-11.84%	-11.40%	32.37%	-2.90%	-6.76%
450	0.3310	0.2050	0.2030	0.3050	0.2050	0.2060	0.2800	0.2130	0.1990
	32.40%	-18.00%	-18.80%	33.77%	-10.09%	-9.65%	35.27%	2.90%	-3.86%
500	0.3300	0.2040	0.2020	0.3040	0.1980	0.1960	0.2830	0.2080	0.1920
	32.00%	-18.40%	-19.20%	33.33%	-13.16%	-14.04%	36.71%	0.48%	-7.25%
550	0.3280	0.2030	0.2050	0.3040	0.1970	0.1930	0.2810	0.2020	0.1890
	31.20%	-18.80%	-18.00%	33.33%	-13.60%	-15.35%	35.75%	-2.42%	-8.70%
600	0.3280	0.2060	0.2020	0.3050	0.1950	0.1900	0.2840	0.2000	0.1930
	31.20%	-17.60%	-19.20%	33.77%	-14.47%	-16.67%	37.20%	-3.38%	-6.76%
650	0.3260	0.2010	0.1980	0.3050	0.1930	0.1880	0.2830	0.2010	0.1970
	30.40%	-19.60%	-20.80%	33.77%	-15.35%	-17.54%	36.71%	-2.90%	-4.83%
700	0.3250	0.2010	0.1990	0.3060	0.1910	0.1880	0.2830	0.2000	0.2020
	30.00%	-19.60%	-20.40%	34.21%	-16.23%	-17.54%	36.71%	-3.38%	-2.42%
750	0.3250	0.1980	0.2000	0.3060	0.1890	0.1870	0.2830	0.1980	0.1980
	30.00%	-20.80%	-20.00%	34.21%	-17.11%	-17.98%	36.71%	-4.35%	-4.35%
800	0.3260	0.1960	0.1970	0.3040	0.1870	0.1870	0.2830	0.2000	0.1970
	30.40%	-21.60%	-21.20%	33.33%	-17.98%	-17.98%	36.71%	-3.38%	-4.83%
850	0.3240	0.1940	0.1940	0.3040	0.1900	0.1840	0.2830	0.2000	0.2000
	29.60%	-22.40%	-22.40%	33.33%	-16.67%	-19.30%	36.71%	-3.38%	-3.38%
900	0.3240	0.1920	0.1970	0.3010	0.1870	0.1820	0.2830	0.1990	0.1980
	29.60%	-23.20%	-21.20%	32.02%	-17.98%	-20.18%	36.71%	-3.86%	-4.35%
950	0.3230	0.1900	0.1980	0.3010	0.1850	0.1820	0.2820	0.1960	0.1960
	29.20%	-24.00%	-20.80%	32.02%	-18.86%	-20.18%	36.23%	-5.31%	-5.31%
1000	0.3220	0.1860	0.1950	0.3010	0.1800	0.1780	0.2830	0.1950	0.2000
	28.80%	-25.60%	-22.00%	32.02%	-21.05%	-21.93%	36.71%	-5.80%	-3.38%

Table 9-1: P@20 before and after reranking for different settings (TREC-6 ad-hoc task).

P@30									
	<i>Inc.ltc</i> (0.2260)			<i>ltc.ltc</i> (0.2053)			<i>atc.atc</i> (0.1800)		
N↓	1	2	3	1	2	3	1	2	3
50	0.2460	0.2240	0.2173	0.2247	0.2187	0.2087	0.2067	0.1973	0.2230
	8.85%	-0.88%	-3.83%	9.42%	6.49%	1.62%	14.81%	9.63%	7.73%
100	0.2687	0.2047	0.2007	0.2473	0.2047	0.1953	0.2340	0.2047	0.2200
	18.88%	-9.44%	-11.21%	20.45%	-0.32%	-4.87%	30.00%	13.70%	6.28%
150	0.2800	0.1980	0.1920	0.2607	0.1907	0.1793	0.2413	0.1960	0.2170
	23.89%	-12.39%	-15.04%	26.95%	-7.14%	-12.66%	34.07%	8.89%	4.83%
200	0.2860	0.1973	0.1880	0.2647	0.1873	0.1847	0.2453	0.1853	0.2020
	26.55%	-12.68%	-16.81%	28.90%	-8.77%	-10.06%	36.30%	2.96%	-2.42%
250	0.2940	0.1980	0.1873	0.2673	0.1860	0.1873	0.2473	0.1813	0.1930
	30.09%	-12.39%	-17.11%	30.19%	-9.42%	-8.77%	37.41%	0.74%	-6.76%
300	0.2947	0.1940	0.1827	0.2713	0.1860	0.1767	0.2473	0.1733	0.1910
	30.38%	-14.16%	-19.17%	32.14%	-9.42%	-13.96%	37.41%	-3.70%	-7.73%
350	0.2933	0.1880	0.1820	0.2713	0.1820	0.1753	0.2467	0.1753	0.1890
	29.79%	-16.81%	-19.47%	32.14%	-11.36%	-14.61%	37.04%	-2.59%	-8.70%
400	0.2927	0.1827	0.1767	0.2720	0.1760	0.1693	0.2467	0.1740	0.1930
	29.50%	-19.17%	-21.83%	32.47%	-14.29%	-17.53%	37.04%	-3.33%	-6.76%
450	0.2940	0.1787	0.1747	0.2740	0.1753	0.1693	0.2507	0.1780	0.1990
	30.09%	-20.94%	-22.71%	33.44%	-14.61%	-17.53%	39.26%	-1.11%	-3.86%
500	0.2940	0.1780	0.1733	0.2740	0.1700	0.1687	0.2540	0.1727	0.1920
	30.09%	-21.24%	-23.30%	33.44%	-17.21%	-17.86%	41.11%	-4.07%	-7.25%
550	0.2940	0.1780	0.1727	0.2727	0.1700	0.1667	0.2540	0.1747	0.1890
	30.09%	-21.24%	-23.60%	32.79%	-17.21%	-18.83%	41.11%	-2.96%	-8.70%
600	0.2947	0.1780	0.1720	0.2747	0.1707	0.1673	0.2553	0.1833	0.1930
	30.38%	-21.24%	-23.89%	33.77%	-16.88%	-18.51%	41.85%	1.85%	-6.76%
650	0.2933	0.1767	0.1680	0.2760	0.1700	0.1633	0.2553	0.1847	0.1970
	29.79%	-21.83%	-25.66%	34.42%	-17.21%	-20.45%	41.85%	2.59%	-4.83%
700	0.2927	0.1727	0.1673	0.2773	0.1720	0.1613	0.2553	0.1840	0.2020
	29.50%	-23.60%	-25.96%	35.06%	-16.23%	-21.43%	41.85%	2.22%	-2.42%
750	0.2933	0.1700	0.1667	0.2773	0.1673	0.1613	0.2547	0.1773	0.1980
	29.79%	-24.78%	-26.25%	35.06%	-18.51%	-21.43%	41.48%	-1.48%	-4.35%
800	0.2933	0.1700	0.1660	0.2780	0.1673	0.1607	0.2560	0.1740	0.1970
	29.79%	-24.78%	-26.55%	35.39%	-18.51%	-21.75%	42.22%	-3.33%	-4.83%
850	0.2927	0.1687	0.1640	0.2780	0.1673	0.1613	0.2567	0.1733	0.2000
	29.50%	-25.37%	-27.43%	35.39%	-18.51%	-21.43%	42.59%	-3.70%	-3.38%
900	0.2927	0.1673	0.1640	0.2773	0.1647	0.1600	0.2573	0.1693	0.1980
	29.50%	-25.96%	-27.43%	35.06%	-19.81%	-22.08%	42.96%	-5.93%	-4.35%
950	0.2920	0.1673	0.1653	0.2773	0.1653	0.1580	0.2567	0.1680	0.1960
	29.20%	-25.96%	-26.84%	35.06%	-19.48%	-23.05%	42.59%	-6.67%	-5.31%
1000	0.2907	0.1673	0.1660	0.2773	0.1647	0.1553	0.2567	0.1700	0.2000
	28.61%	-25.96%	-26.55%	35.06%	-19.81%	-24.35%	42.59%	-5.56%	-3.38%

Table 9-2: P@30 before and after reranking for different settings (TREC-6 ad-hoc task).

non-interpolated average precision (baseline: 0.1471)									
T → N ↓	25	50	75	100	200	300	400	500	600
*	0.1113	0.1152	0.1197	0.1230	0.1320	0.1350	0.1364	0.1376	0.1383
	-24.34%	-21.69%	-18.63%	-16.38%	-10.27%	-8.23%	-7.27%	-6.46%	-5.98%
50	0.1168	0.1247	0.1322	0.1380	0.1493	0.1555	0.1596	0.1590	0.1604
	-20.60%	-15.23%	-10.13%	-6.19%	1.50%	5.71%	8.50%	8.09%	9.04%
100	0.1182	0.1320	0.1422	0.1473	0.1593	0.1658	0.1688	0.1696	0.1709
	-19.65%	-10.27%	-3.33%	0.14%	8.29%	12.71%	14.75%	15.30%	16.18%
150	0.1198	0.1343	0.1437	0.1482	0.1620	0.1687	0.1721	0.1731	0.1744
	-18.56%	-8.70%	-2.31%	0.75%	10.13%	14.68%	17.00%	17.68%	18.56%
200	0.1209	0.1354	0.1449	0.1495	0.1634	0.1706	0.1740	0.1750	0.1763
	-17.81%	-7.95%	-1.50%	1.63%	11.08%	15.98%	18.29%	18.97%	19.85%
250	0.1239	0.1390	0.1497	0.1554	0.1694	0.1780	0.1822	0.1831	0.1845
	-15.77%	-5.51%	1.77%	5.64%	15.16%	21.01%	23.86%	24.47%	25.42%
300	0.1237	0.1385	0.1497	0.1553	0.1695	0.1781	0.1824	0.1832	0.1846
	-15.91%	-5.85%	1.77%	5.57%	15.23%	21.07%	24.00%	24.54%	25.49%
350	0.1258	0.1405	0.1515	0.1571	0.1716	0.1812	0.1851	0.1862	0.1878
	-14.48%	-4.49%	2.99%	6.80%	16.66%	23.18%	25.83%	26.58%	27.67%
400	0.1259	0.1409	0.1518	0.1582	0.1720	0.1816	0.1855	0.1869	0.1883
	-14.41%	-4.21%	3.20%	7.55%	16.93%	23.45%	26.10%	27.06%	28.01%
450	0.1260	0.1405	0.1514	0.1581	0.1723	0.1816	0.1853	0.1868	0.1882
	-14.34%	-4.49%	2.92%	7.48%	17.13%	23.45%	25.97%	26.99%	27.94%
500	0.1259	0.1403	0.1512	0.1578	0.1723	0.1820	0.1858	0.1869	0.1882
	-14.41%	-4.62%	2.79%	7.27%	17.13%	23.73%	26.31%	27.06%	27.94%
550	0.1260	0.1402	0.1510	0.1578	0.1721	0.1819	0.1858	0.1872	0.1885
	-14.34%	-4.69%	2.65%	7.27%	17.00%	23.66%	26.31%	27.26%	28.14%
600	0.1262	0.1399	0.1515	0.1579	0.1726	0.1828	0.1864	0.1876	0.1892
	-14.21%	-4.89%	2.99%	7.34%	17.34%	24.27%	26.72%	27.53%	28.62%
650	0.1262	0.1397	0.1510	0.1573	0.1716	0.1819	0.1851	0.1866	0.1877
	-14.21%	-5.03%	2.65%	6.93%	16.66%	23.66%	25.83%	26.85%	27.60%
700	0.1262	0.1397	0.1510	0.1573	0.1717	0.1818	0.1850	0.1863	0.1876
	-14.21%	-5.03%	2.65%	6.93%	16.72%	23.59%	25.76%	26.65%	27.53%
750	0.1261	0.1397	0.1511	0.1577	0.1719	0.1819	0.1851	0.1865	0.1877
	-14.28%	-5.03%	2.72%	7.21%	16.86%	23.66%	25.83%	26.78%	27.60%
800	0.1262	0.1398	0.1512	0.1578	0.1721	0.1821	0.1852	0.1865	0.1879
	-14.21%	-4.96%	2.79%	7.27%	17.00%	23.79%	25.90%	26.78%	27.74%
850	0.1261	0.1397	0.1506	0.1575	0.1717	0.1818	0.1849	0.1861	0.1876
	-14.28%	-5.03%	2.38%	7.07%	16.72%	23.59%	25.70%	26.51%	27.53%
900	0.1261	0.1397	0.1506	0.1575	0.1717	0.1818	0.1850	0.1862	0.1876
	-14.28%	-5.03%	2.38%	7.07%	16.72%	23.59%	25.76%	26.58%	27.53%
950	0.1261	0.1396	0.1506	0.1575	0.1718	0.1817	0.1848	0.1862	0.1877
	-14.28%	-5.10%	2.38%	7.07%	16.79%	23.52%	25.63%	26.58%	27.60%
1000	0.1266	0.1401	0.1510	0.1582	0.1720	0.1820	0.1852	0.1862	0.1876
	-13.94%	-4.76%	2.65%	7.55%	16.93%	23.73%	25.90%	26.58%	27.53%

* baseline pseudo-feedback runs

Table 9-3: non-interpolated average precision (varying T) (TREC-6 ad-hoc task).

P@20 (baseline: 0.2500)									
T → N ↓	25	50	75	100	200	300	400	500	600
*	0.2220	0.2240	0.2260	0.2270	0.2420	0.2430	0.2450	0.2450	0.2440
	-11.20%	-10.40%	-9.60%	-9.20%	-3.20%	-2.80%	-2.00%	-2.00%	-2.40%
50	0.2370	0.2500	0.2570	0.2620	0.2680	0.2820	0.2830	0.2850	0.2820
	-5.20%	0.00%	2.80%	4.80%	7.20%	12.80%	13.20%	14.00%	12.80%
100	0.2440	0.2670	0.2740	0.2790	0.2870	0.2980	0.2980	0.2990	0.2970
	-2.40%	6.80%	9.60%	11.60%	14.80%	19.20%	19.20%	19.60%	18.80%
150	0.2490	0.2680	0.2770	0.2840	0.2910	0.3010	0.3030	0.3060	0.3020
	-0.40%	7.20%	10.80%	13.60%	16.40%	20.40%	21.20%	22.40%	20.80%
200	0.2460	0.2690	0.2760	0.2850	0.2920	0.3040	0.3050	0.3070	0.3050
	-1.60%	7.60%	10.40%	14.00%	16.80%	21.60%	22.00%	22.80%	22.00%
250	0.2540	0.2760	0.2860	0.2940	0.3030	0.3170	0.3220	0.3240	0.3230
	1.60%	10.40%	14.40%	17.60%	21.20%	26.80%	28.80%	29.60%	29.20%
300	0.2550	0.2770	0.2870	0.2980	0.3040	0.3180	0.3230	0.3230	0.3240
	2.00%	10.80%	14.80%	19.20%	21.60%	27.20%	29.20%	29.20%	29.60%
350	0.2570	0.2790	0.2880	0.2960	0.3040	0.3210	0.3230	0.3260	0.3250
	2.80%	11.60%	15.20%	18.40%	21.60%	28.40%	29.20%	30.40%	30.00%
400	0.2580	0.2800	0.2880	0.2970	0.3060	0.3200	0.3220	0.3250	0.3240
	3.20%	12.00%	15.20%	18.80%	22.40%	28.00%	28.80%	30.00%	29.60%
450	0.2560	0.2790	0.2880	0.2970	0.3060	0.3190	0.3210	0.3240	0.3210
	2.40%	11.60%	15.20%	18.80%	22.40%	27.60%	28.40%	29.60%	28.40%
500	0.2570	0.2760	0.2890	0.2990	0.3050	0.3190	0.3210	0.3240	0.3210
	2.80%	10.40%	15.60%	19.60%	22.00%	27.60%	28.40%	29.60%	28.40%
550	0.2550	0.2780	0.2860	0.2950	0.3050	0.3180	0.3190	0.3230	0.3210
	2.00%	11.20%	14.40%	18.00%	22.00%	27.20%	27.60%	29.20%	28.40%
600	0.2550	0.2780	0.2860	0.2950	0.3060	0.3200	0.3210	0.3240	0.3210
	2.00%	11.20%	14.40%	18.00%	22.40%	28.00%	28.40%	29.60%	28.40%
650	0.2560	0.2790	0.2850	0.2950	0.3070	0.3200	0.3210	0.3240	0.3210
	2.40%	11.60%	14.00%	18.00%	22.80%	28.00%	28.40%	29.60%	28.40%
700	0.2550	0.2780	0.2850	0.2950	0.3070	0.3190	0.3210	0.3230	0.3200
	2.00%	11.20%	14.00%	18.00%	22.80%	27.60%	28.40%	29.20%	28.00%
750	0.2540	0.2760	0.2840	0.2930	0.3060	0.3170	0.3200	0.3220	0.3190
	1.60%	10.40%	13.60%	17.20%	22.40%	26.80%	28.00%	28.80%	27.60%
800	0.2580	0.2750	0.2830	0.2930	0.3070	0.3170	0.3200	0.3220	0.3190
	3.20%	10.00%	13.20%	17.20%	22.80%	26.80%	28.00%	28.80%	27.60%
850	0.2560	0.2750	0.2800	0.2930	0.3070	0.3170	0.3200	0.3200	0.3180
	2.40%	10.00%	12.00%	17.20%	22.80%	26.80%	28.00%	28.00%	27.20%
900	0.2560	0.2750	0.2800	0.2930	0.3070	0.3170	0.3200	0.3200	0.3180
	2.40%	10.00%	12.00%	17.20%	22.80%	26.80%	28.00%	28.00%	27.20%
950	0.2570	0.2740	0.2810	0.2930	0.3070	0.3170	0.3180	0.3190	0.3170
	2.80%	9.60%	12.40%	17.20%	22.80%	26.80%	27.20%	27.60%	26.80%
1000	0.2610	0.2760	0.2820	0.2930	0.3040	0.3160	0.3180	0.3190	0.3170
	4.40%	10.40%	12.80%	17.20%	21.60%	26.40%	27.20%	27.60%	26.80%

* baseline pseudo-feedback runs

Table 9-4: P@20 (varying T) (TREC-6 ad-hoc task).

non-interpolated average precision (baseline: 0.1471)									
T → β ↓	25	50	75	100	200	300	400	500	600
1	0.1113	0.1152	0.1197	0.1230	0.1320	0.1350	0.1364	0.1376	0.1383
2	-24.34%	-21.69%	-18.63%	-16.38%	-10.27%	-8.23%	-7.27%	-6.46%	-5.98%
4	0.0841	0.0952	0.1031	0.1058	0.1141	0.1188	0.1198	0.1205	0.1212
6	-42.83%	-35.28%	-29.91%	-28.08%	-22.43%	-19.24%	-18.56%	-18.08%	-17.61%
8	0.0685	0.0738	0.0875	0.0906	0.1012	0.1046	0.1040	0.1050	0.1049
10	-53.43%	-49.83%	-40.52%	-38.41%	-31.20%	-28.89%	-29.30%	-28.62%	-28.69%
12	0.0600	0.0638	0.0801	0.0820	0.0906	0.0959	0.0976	0.0984	0.0982
14	-59.21%	-56.63%	-45.55%	-44.26%	-38.41%	-34.81%	-33.65%	-33.11%	-33.24%
16	0.0558	0.0596	0.0749	0.0780	0.0873	0.0918	0.0932	0.0943	0.0945
18	-62.07%	-59.48%	-49.08%	-46.97%	-40.65%	-37.59%	-36.64%	-35.89%	-35.76%
20	0.0523	0.0569	0.0717	0.0756	0.0854	0.0885	0.0906	0.0916	0.0919
	-64.45%	-61.32%	-51.26%	-48.61%	-41.94%	-39.84%	-38.41%	-37.73%	-37.53%
	0.0493	0.0553	0.0701	0.0741	0.0840	0.0871	0.0895	0.0901	0.0903
	-66.49%	-62.41%	-52.35%	-49.63%	-42.90%	-40.79%	-39.16%	-38.75%	-38.61%
	0.0474	0.0541	0.0689	0.0728	0.0826	0.0862	0.0883	0.0885	0.0892
	-67.78%	-63.22%	-53.16%	-50.51%	-43.85%	-41.40%	-39.97%	-39.84%	-39.36%
	0.0462	0.0530	0.0677	0.0719	0.0814	0.0851	0.0875	0.0875	0.0886
	-68.59%	-63.97%	-53.98%	-51.12%	-44.66%	-42.15%	-40.52%	-40.52%	-39.77%
	0.0453	0.0523	0.0671	0.0712	0.0806	0.0844	0.0866	0.0871	0.0876
	-69.20%	-64.45%	-54.38%	-51.60%	-45.21%	-42.62%	-41.13%	-40.79%	-40.45%
	0.0446	0.0520	0.0664	0.0704	0.0799	0.0838	0.0861	0.0864	0.0869
	-69.68%	-64.65%	-54.86%	-52.14%	-45.68%	-43.03%	-41.47%	-41.26%	-40.92%

Table 9-5: non-interpolated average precision (varying β) (TREC-6 ad-hoc task baseline pseudo-feedback runs).

non-interpolated average precision (baseline: 0.1471)									
T → β ↓	25	50	75	100	200	300	400	500	600
1	0.1237	0.1385	0.1497	0.1553	0.1695	0.1781	0.1824	0.1832	0.1846
	-15.91%	-5.85%	1.77%	5.57%	15.23%	21.07%	24.00%	24.54%	25.49%
2	0.1013	0.1192	0.1348	0.1408	0.1561	0.1671	0.1702	0.1720	0.1732
	-31.14%	-18.97%	-8.36%	-4.28%	6.12%	13.60%	15.70%	16.93%	17.74%
4	0.0851	0.0959	0.1162	0.1248	0.1459	0.1535	0.1555	0.1575	0.1595
	-42.15%	-34.81%	-21.01%	-15.16%	-0.82%	4.35%	5.71%	7.07%	8.43%
6	0.0769	0.0848	0.1065	0.1163	0.1375	0.1446	0.1484	0.1513	0.1534
	-47.72%	-42.35%	-27.60%	-20.94%	-6.53%	-1.70%	0.88%	2.86%	4.28%
8	0.0714	0.0799	0.0994	0.1119	0.1333	0.1398	0.1447	0.1480	0.1488
	-51.46%	-45.68%	-32.43%	-23.93%	-9.38%	-4.96%	-1.63%	0.61%	1.16%
10	0.0681	0.0770	0.0965	0.1096	0.1303	0.1366	0.1417	0.1456	0.1466
	-53.70%	-47.65%	-34.40%	-25.49%	-11.42%	-7.14%	-3.67%	-1.02%	-0.34%
12	0.0651	0.0749	0.0940	0.1075	0.1282	0.1348	0.1397	0.1433	0.1450
	-55.74%	-49.08%	-36.10%	-26.92%	-12.85%	-8.36%	-5.03%	-2.58%	-1.43%
14	0.0638	0.0729	0.0928	0.1061	0.1281	0.1332	0.1367	0.1421	0.1435
	-56.63%	-50.44%	-36.91%	-27.87%	-12.92%	-9.45%	-7.07%	-3.40%	-2.45%
16	0.0628	0.0720	0.0913	0.1052	0.1271	0.1321	0.1362	0.1408	0.1423
	-57.31%	-51.05%	-37.93%	-28.48%	-13.60%	-10.20%	-7.41%	-4.28%	-3.26%
18	0.0622	0.0711	0.0906	0.1046	0.1263	0.1313	0.1354	0.1401	0.1415
	-57.72%	-51.67%	-38.41%	-28.89%	-14.14%	-10.74%	-7.95%	-4.76%	-3.81%
20	0.0608	0.0707	0.0901	0.1038	0.1256	0.1308	0.1348	0.1394	0.1406
	-58.67%	-51.94%	-38.75%	-29.44%	-14.62%	-11.08%	-8.36%	-5.23%	-4.42%

Table 9-6: non-interpolated average precision (varying β) (TREC-6 ad-hoc task).

P@20 (baseline: 0.2500)									
T → β ↓	25	50	75	100	200	300	400	500	600
1	0.2220	0.2240	0.2260	0.2270	0.2420	0.2430	0.2450	0.2450	0.2440
	-11.20%	-10.40%	-9.60%	-9.20%	-3.20%	-2.80%	-2.00%	-2.00%	-2.40%
2	0.1980	0.2080	0.2170	0.2210	0.2290	0.2360	0.2380	0.2390	0.2370
	-20.80%	-16.80%	-13.20%	-11.60%	-8.40%	-5.60%	-4.80%	-4.40%	-5.20%
4	0.1820	0.1920	0.2110	0.2170	0.2270	0.2300	0.2320	0.2310	0.2300
	-27.20%	-23.20%	-15.60%	-13.20%	-9.20%	-8.00%	-7.20%	-7.60%	-8.00%
6	0.1740	0.1840	0.2050	0.2070	0.2170	0.2250	0.2270	0.2290	0.2270
	-30.40%	-26.40%	-18.00%	-17.20%	-13.20%	-10.00%	-9.20%	-8.40%	-9.20%
8	0.1660	0.1770	0.1960	0.2020	0.2140	0.2230	0.2250	0.2250	0.2260
	-33.60%	-29.20%	-21.60%	-19.20%	-14.40%	-10.80%	-10.00%	-10.00%	-9.60%
10	0.1590	0.1690	0.1950	0.1970	0.2120	0.2200	0.2220	0.2210	0.2220
	-36.40%	-32.40%	-22.00%	-21.20%	-15.20%	-12.00%	-11.20%	-11.60%	-11.20%
12	0.1570	0.1680	0.1950	0.1960	0.2140	0.2170	0.2200	0.2190	0.2200
	-37.20%	-32.80%	-22.00%	-21.60%	-14.40%	-13.20%	-12.00%	-12.40%	-12.00%
14	0.1560	0.1660	0.1910	0.1930	0.2120	0.2150	0.2170	0.2190	0.2210
	-37.60%	-33.60%	-23.60%	-22.80%	-15.20%	-14.00%	-13.20%	-12.40%	-11.60%
16	0.1530	0.1630	0.1890	0.1900	0.2110	0.2140	0.2170	0.2180	0.2200
	-38.80%	-34.80%	-24.40%	-24.00%	-15.60%	-14.40%	-13.20%	-12.80%	-12.00%
18	0.1520	0.1630	0.1870	0.1890	0.2080	0.2130	0.2170	0.2170	0.2200
	-39.20%	-34.80%	-25.20%	-24.40%	-16.80%	-14.80%	-13.20%	-13.20%	-12.00%
20	0.1510	0.1630	0.1860	0.1890	0.2080	0.2130	0.2150	0.2170	0.2200
	-39.60%	-34.80%	-25.60%	-24.40%	-16.80%	-14.80%	-14.00%	-13.20%	-12.00%

Table 9-7: P@20 (varying β) (TREC-6 ad-hoc task baseline pseudo-feedback runs).

P@20 (baseline: 0.2500)									
T → β ↓	25	50	75	100	200	300	400	500	600
1	0.2550 2.00%	0.2770 10.80%	0.2870 14.80%	0.2980 19.20%	0.3040 21.60%	0.3180 27.20%	0.3230 29.20%	0.3230 29.20%	0.3240 29.60%
2	0.2180 -12.80%	0.2600 4.00%	0.2800 12.00%	0.2990 19.60%	0.3160 26.40%	0.3240 29.60%	0.3320 32.80%	0.3340 33.60%	0.3320 32.80%
4	0.1880 -24.80%	0.2280 -8.80%	0.2740 9.60%	0.2840 13.60%	0.3190 27.60%	0.3240 29.60%	0.3220 28.80%	0.3210 28.40%	0.3250 30.00%
6	0.1760 -29.60%	0.2140 -14.40%	0.2630 5.20%	0.2730 9.20%	0.3040 21.60%	0.3150 26.00%	0.3140 25.60%	0.3120 24.80%	0.3180 27.20%
8	0.1700 -32.00%	0.2090 -16.40%	0.2500 0.00%	0.2650 6.00%	0.2930 17.20%	0.3100 24.00%	0.3110 24.40%	0.3090 23.60%	0.3130 25.20%
10	0.1670 -33.20%	0.2020 -19.20%	0.2470 -1.20%	0.2640 5.60%	0.2940 17.60%	0.3020 20.80%	0.3070 22.80%	0.3090 23.60%	0.3120 24.80%
12	0.1640 -34.40%	0.2010 -19.60%	0.2440 -2.40%	0.2640 5.60%	0.2940 17.60%	0.2980 19.20%	0.3070 22.80%	0.3050 22.00%	0.3080 23.20%
14	0.1620 -35.20%	0.1960 -21.60%	0.2400 -4.00%	0.2630 5.20%	0.2910 16.40%	0.2940 17.60%	0.3040 21.60%	0.3070 22.80%	0.3070 22.80%
16	0.1590 -36.40%	0.1930 -22.80%	0.2380 -4.80%	0.2610 4.40%	0.2900 16.00%	0.2930 17.20%	0.3030 21.20%	0.3030 21.20%	0.3040 21.60%
18	0.1610 -35.60%	0.1900 -24.00%	0.2350 -6.00%	0.2590 3.60%	0.2860 14.40%	0.2930 17.20%	0.3010 20.40%	0.3020 20.80%	0.3040 21.60%
20	0.1580 -36.80%	0.1900 -24.00%	0.2330 -6.80%	0.2560 2.40%	0.2830 13.20%	0.2920 16.80%	0.3000 20.00%	0.3000 20.00%	0.3040 21.60%

Table 9-8: P@20 (varying β) (TREC-6 ad-hoc task).