Experiments in Query Processing at LEXIS-NEXIS for TREC-7

Ashwin G. Rao, Timothy Humphrey, Afsar Parhizgar, Christi Wilson, Daniel Pliske
Applied Research
LEXIS-NEXIS
(ashwin.g.rao,timothy.humphrey,afsar.parhizgar,christi.wilson,daniel.pliske)@LEXIS-NEXIS.com

Introduction

The purpose of this report is to provide an overview of LEXIS-NEXIS' entries to the TREC-7 competition. The report will describe the experiments we conducted, the results we obtained, and our future research directions. The report is divided into three sections. The first section describes the experimental setup and gives a brief account of some of the research activities that led to the TREC-7 entries. The second section explains how the techniques developed during our research culminated into the three entries that were submitted. Our experiences with these new techniques gave us insight into new research directions for improving query processing. In the third section, we conclude by sharing these ideas with the reader.

1 - TREC-7 Research at LEXIS-NEXIS

In the past [1] [2] [3], we concentrated on the evaluation of various query enhancement techniques to improve the quality of the final retrieval. This year, we decided to focus our attention on one specific technique. While preparing for our TREC submission last year, we came to the conclusion that retaining the focus of the original queries was critical during query expansion, and we could adopt co-occurrence analysis techniques to help us meet this need. Therefore, this year, we decided to spend a major portion of our time fine-tuning our co-occurrence metrics. Adding related terms to the query normally helps retrieve more documents, at the expense of precision (i.e. reducing the number of relevant documents at the top of the ranked list.) We believed that if we could improve our co-occurrence analysis process, we could add as many terms as possible to the queries and let the co-occurrence analysis process eliminate superfluous terms to bring documents that are more relevant into the ranked list. Our goal was to improve the precision of the final retrieval by 10% over last year's best official result. The best official result for TREC-6 was from the City University of London. The best unofficial result was from UMASS. Table 1 contains the precision values from the two universities and our final training results.

TITLE+DESC (TRAINING)				***Precision***		
Entry	***Avg. Precision***	***Exact Precision***	***Rel_Ret***	***At 5 docs***	***At 20 docs***	***At .20***
City University	0.2327	0.2595	2422	0.4360	0.3320	0.3892
UMASS (unofficial)	0.2730	0.3021	N/A	N/A	0.4200	N/A
LN (training)	0.2749	0.3080	2685	0.5080	0.3850	0.4771
(Improvement over City Univ.)	+18.1%	+18.7%	+10.9%	+16.5%	+16.6%	+22.6%
(Improvement over UMASS)	+0.7%	+1.95%	N/A	N/A	-8.33	N/A

Table 1. Comparison of results from the best performers last year vs. results after training.

In order to utilize our computing resources effectively, we re-engineered our processes so that we could distribute them and run them in parallel over a large set of hardware resources (around 200 Solaris workstations and servers).
Our initial plan was to participate in both Ad hoc and Filtering tracks, but we later decided to limit our participation to the Ad hoc task due to time constraints. We have submitted three entries, LNaTitleDesc7, LNaTitle7, and LNmanual7. As the names denote, the LNaTitleDesc7 entry used terms from both the title and the description fields of the TREC-7 topics to automatically retrieve documents. The LNaTitle7 entry used just the terms from the title field of TREC-7 topics

for automatic retrieval. Our manual entry, LNmanual7, used any information available to the human analysts for manual retrieval. All three entries will be described in detail in this report.

In any information retrieval system, text data goes through various stages before it can be used by the search engine. We used TREC-7 topics to make queries that are then used by the search engine to retrieve relevant documents from the TREC-7 corpus. Apart from the original topic terms, the queries can contain additional terms from external sources like dictionaries, thesauri and the LEXIS-NEXIS' REL[1] feature. The inverted files, corpus statistics, and queries become inputs into the search engine that creates a ranked list of documents. During the training process, we score the performance of our retrieval system by comparing the ranked list of documents output by the search engine, with a list of relevant documents provided by NIST. We use this scoring process to learn how to improve our retrieval processes.

This year we used the TREC-6 topics to train the system. We opted to use the unofficial results from UMASS as the benchmark for our training.

1.1 - Search Engine

To choose a search engine for this year's TREC competition, we revisited our best implementation of the well-known algorithms that were used at previous TREC conferences. We chose to implement Ocelot (based on City University's BM25 [4]), Inquery (based on UMASS' Inquery [6] [8]), and Panther (based on Cornell's Lnu.ltu [7]). Please refer to the original papers for more information about the algorithms.

During our initial training runs using the title and description fields of the TREC-6 queries, we highlighted nouns by incrementing their frequencies and we identified and added phrases to queries. We used these new queries to help us pick one or two of the best implementations of the algorithms described above. Ocelot turned out to be a clear winner as can be seen from Table 2. We therefore adopted it for the remainder of our experiments.

TITLE + DESC (TRAINING)				***Precision***	
Method	***Rel_Ret***	***Avg. Precision***	***Exact Precision***	***At 0.20***	***At 20 docs***
Ocelot	2542	.2560	.2885	.4747	.3640
Panther	2330	.2125	.2377	.3961	.2920
Inquery	2431	.2555	.2835	.3901	.3340

Table 2. Comparison of results of the three search algorithms.

We modified the Ocelot algorithm by incorporating query term coverage and query term dependency techniques. Both techniques are related to query processing and they will be described in the next section.

1.2 - Query Creation and Enhancement

This section will briefly describe various steps taken to enhance the queries. The first sub-section describes how we pre-processed our topics. Section 1.2.2 describes our attempts in adding more relevant terms to the topics to effectively improve the recall while keeping the precision as high as possible.

Section 1.2.3 describes two techniques that we have found to be effective for very short queries (consisting of 2-3 terms). The last sub-section explains the relevance feedback approach.

1.2.1 - Query Pre-processing

In our experiments, we have found that title terms carry much more information than the terms in the description, validating the observations by Voorhees [12]. In order to ensure the dominance of the title terms, we multiplied the within-query term frequencies by a factor. We arrived at the multiplier after some ad hoc experimentation.

[1] The LEXIS-NEXIS' REL (RELated concepts) feature of the LEXIS-NEXIS commercial system provides the user with a list of related terms that can be automatically incorporated into his search request. For this exercise, we logged on to the CURNWS file within the NEWS library of the online system, and transmitted the REL command along with a term selected from the query's title or the description field. The system returned a list of several dozen related terms, including multi-word terms.

For the automatic entries, we decided to repeat the query processing steps that gave us the most significant boost in TREC-6. During our preparation for TREC-6 [1], we experimented with highlighting query terms that were classified by WordNet [9] as nouns. We then added potential phrases and synonyms based on WordNet. This year we just performed the first step. The addition of synonyms consistently gave us poor results, so we skipped this step for our final processing. For the first query-processing step, we highlighted nouns that were identified by a table look-up of nouns found in WordNet by incrementing their frequencies by one. In addition, we also detected phrases in the query that were also present in the TREC-7 corpus.

The improvements due to phrase-detection were not as large as we had expected. Quite a few important phrases in the corpus were missed because of a bug in the phrase-detection program. Table 3 shows the improvement we gained in the ranking after applying the above technique.

TITLE + DESC (TRAINING)				***Precision***	
Run	***Rel_Ret***	***Avg. Precision***	***Exact Precision***	***At 0.20***	***At 20 docs***
Baseline	2377	.2560	.2741	.4749	.3620
Nouns+Phrases	2542	.2560	.2885	.4747	.3640

Table 3. Improvements gained by adding nouns and phrases.

1.2.2 - Adding More Terms Using LEXIS-NEXIS' REL Feature

Last year we found that adding related terms from the LEXIS-NEXIS online system helped us retrieve a larger number of relevant documents in the LNaShort (title only) entry [1]. This work was done after Voorhees [10] had found that related terms help by enhancing recall in the retrieval process. We decided to use this technique for the two automatic entries in TREC-7. We wanted to use the new co-occurrence analysis process to filter out noisy query terms added by the LEXIS-NEXIS REL feature. During training, we realized that precision and recall statistics were drastically effected by the inclusion of the related terms, but they seemed to bring in relevant documents not present in the original Nouns+Phrases run. We therefore retained this run to serve as input into the data-fusion step. Table 4 shows incremental improvements due to REL+Co-occurrence and data-fusion steps.

TITLE + DESC (TRAINING)				***Precision***	
Run	***Rel_Ret***	***Avg. Precision***	***Exact Precision***	***At 0.20***	***At 20 docs***
Nouns+Phrases	2542	.2560	.2885	.4747	.3640
Nouns+Phrases+REL+Co	2243	.2240	.2582	.3976	.3060
Data-Fusion	2526	.2620	.2898	.4493	.3810

Table 4. Incremental improvements due to REL + Co-occurrence and Data-Fusion.

1.2.3 - Query Term Dependency and Query Term Coverage Factors

For the title only entry (LNaTitle7), we also added term dependency and term coverage factors to improve the precision in the initial (pre-relevance feedback) run. The term dependency technique attempts to capture multiple concepts within the title. We give more weight to terms that are physically separated from each other than to those that are close together. The assumption behind this approach is that normally within short queries (2-3 terms), two terms located adjacent to each other are about the same concept. The probability of two terms not adjacent to each other being of different concepts increases as the distance between them increases. The query term coverage factor ranks higher the documents with a larger number of query terms. The assumption here is that the title fields of TREC topics are very specific. Hence, the larger the number of query terms in the document, the more on-point it is. These techniques can only be used in queries with 2-3 terms because longer queries become less focused and we can't assume that all query terms are equally important. To compound this problem, longer queries result in larger inter-term distance factors that confuse the original probabilities of relevance of various documents. The benefit of adding the term dependency factor and query term coverage processing can be seen in Table 5.

TITLE ONLY (TRAINING)				***Precision***	
Run	***Rel_Ret***	***Avg. Precision***	***Exact Precision***	***At 0.20***	***At 20 docs***
Nouns+Phrases	2231	.2250	.2576	.3988	.3300
Nouns+Phrases+trmDep	2265	.2386	.2791	.4291	.3450
Nouns+Phrases+trmDep+Cov	2235	.2382	.2792	.4299	.3450

Table 5. Incremental improvements due to query term dependency and coverage.

1.2.4 - Relevance Feedback

This year we were planning to experiment with both LDA [6] and Rocchio [11] relevance ranking approaches, but due to resource constraints, we were left with just enough time to submit the Rocchio runs. We performed Rocchio relevance ranking on the two automatic entries. We found that we got consistently better results after the inclusion of a factor representing non-relevant documents to our last year's Rocchio formula. Our Rocchio re-weighting formula was:

*8 * original query vector + 4 * average relevant vector - 4 * average non-relevant vector*

where *average relevant vector* consists of terms from the top ranked documents, and *average non-relevant vector* consists of terms from the documents ranked at the bottom of the ranked list. The results of our Rocchio processing and subsequent fusion with original ranking can be found in Table 6.

TITLE + DESC (TRAINING)				***Precision***	
Run	***Rel_Ret***	***Avg. Precision***	***Exact Precision***	***At 0.20***	***At 20 docs***
Nouns+Phrases	2542	.2560	.2885	.4747	.3640
Rocchio	2616	.2657	.3054	.4495	.3690
Data-Fusion	2685	.2749	.3080	.4771	.3850

Table 6. Incremental improvements due to Rocchio and Data-Fusion processes.

1.3 - Data Fusion

It is a well-known fact that different ranking algorithms and different query processing techniques retrieve different sets of documents. Belkin [16] used probability theory to arrive at a technique that merges results from different ranking techniques. If the fusion is done right, merging different sources of evidence (rankings) will improve the retrieval effectiveness, because the merged results will contain the best documents from all the sources. During preparation for the TREC-6 conference, we had experimented with various data-fusion techniques. We settled on a technique that gave us consistently better results. We used this technique for the TREC-7 conference too. We have found that data-fusion improves both precision and recall. Results in Table 6 show the improvement gained over the baseline algorithms as a result of the application of the data-fusion step.

1.4 - Co-occurrence Metrics

Most IR techniques are based on standard statistical measures that use the term frequency information within the text to determine whether a document is relevant to a query. Unfortunately, query terms can be used in multiple contexts with distinct meanings, so merely looking at term frequencies is not enough. When we try to add terms to the queries to improve recall, the problem is magnified. The reason is that we tend to add terms that expand the alternate meanings of the query terms along with the terms that belong to the same concept as the query. Several techniques exist that counter this effect. One of the most popular techniques has been the use of the mutual-information metric [17]. We investigated this and other techniques and arrived at a hybrid approach to evaluate the importance of various terms with respect to the original query terms.

Co-occurrence and Its Importance in Query Enhancement

To determine the best terms to add to a query, we began with the terms in the title of each TREC topic. These terms are known to be relevant to the respective query because of the way that they were created [12]. We use these terms as anchors to add new terms. We make an assumption that terms that frequently co-occur with all the title terms have a good probability of being relevant to the query, and are therefore added to the query. Terms that don't co-occur with all the title terms are eliminated.

There are many methods for measuring co-occurrence (e.g. Dice's coefficient, Jaccards's coefficient, cosine coefficient, the overlap measure, and many others [14]). Some of these measures work better than others in different situations. We experimented with several of these methods and derived one of our own that worked well on TREC data.

2 - Description of Entries

2.1 - LNaTitleDesc7

LNaTitleDesc7 entry used the title and the description fields of the TREC-7 topics. Two initial ranking runs were performed as a precursor to the Rocchio process. For the first ranking run, we enhanced the topics as described in section 1.2. We chose a multiplication factor of two to ensure that title terms were more significant for retrieval than the description terms.

For the second run, we added REL terms from the online system. To maintain the focus of the topics, weights of the title terms were tripled, the weights of the description terms were doubled, and both sets were added to the related terms from the online system. The new set of topics was pre-processed by the new co-occurrence metric before being ranked.

The output of the two initial rankings was combined by the data-fusion process. The combined ranking was then processed by a Rocchio relevance ranking process as described in section 1.2.4. We again processed the newly added terms through our co-occurrence analysis process to weed out terms that were deemed superfluous, or not on-point, before re-ranking the results. The results were later merged with the original Title+Description ranking to arrive at the final LNaTitleDesc7 ranking. Table 7 and Figure 1 show the results that were obtained after running the TREC-7 evaluation program after each intermediate step of the LNaTitleDesc7 entry.

LNaTitleDesc7					***Precision***	
Step	***Run***	***Rel_Ret***	***Avg. Precision***	***Exact Precision***	***At 0.20***	***At 20 docs***
1	Baseline	2442	.2030	.2559	.3712	.3950
2	1 + Nouns + Phrases	2581	.2127	.2628	.3672	.4050
3	2 + REL	2586	.2310	.2700	.3908	.3860
4	Co-occurrence on 3	2870	.2405	.2807	.3969	.3980
5	Data Fusion of (2+4)	2899	.2410	.2867	.4105	.4120
6	Rocchio on 5	3112	.2418	.2714	.3870	.3860
7	Co-occurrence on 6	3106	.2427	.2734	.3970	.3860
8	Data Fusion of (2+7)	3020	.2394	.2783	.4073	.4050

Table 7. The results of incremental steps in the LNaTitleDesc7 entry.

By re-weighting nouns and detecting phrases, we were able to get a 4.8% improvement in retrieval performance. Unlike the training run (Table 4), addition of REL terms from the LEXIS-NEXIS online system didn't hurt the precision as much while bringing more relevant documents to the top 1000. The co-occurrence analysis step (Step 4) helped us eliminate the noisy REL terms to improve the retrieval by 13% over Step 2. The data-fusion step (Step 5) improved the precision further especially among the top 20 documents. We then applied the Rocchio relevance feedback approach to Step 5 to get a new set of queries. The retrieval performance of Rocchio seems to be consistent with our earlier observations. We were able to bring in more documents to the ranked-list at the expense of precision at the top 20 documents. The co-occurrence analysis step improved the precision further although some of the relevant documents were lost because some good terms that did not co-occur with title-terms were also eliminated. The final data-fusion step was undertaken as a conservative measure to ensure that the focus of the original queries was not lost. This step undid some of the improvements we had gained in our previous steps. As an after-thought, we could have done without this step.

The co-occurrence analysis steps in the LNaTitleDesc7 entry proved to be quite effective in improving the precision. Co-occurrence analysis provides an automatic approach for choosing terms that are more important than others for query expansion.

2.2 - LNaTitle7

The LNaTitle7 entry used the title field of the TREC-7 topics for retrieval. Figure 4 illustrates the processing steps that were taken to create the LNaTitle7 entry. Two initial ranking runs were performed before the Rocchio process. The first ranking run had as input the title terms with nouns being enhanced and phrases being identified as described in section 1.2. The ranking took into consideration term dependency and term coverage of the queries.

The second run used the REL terms from the online system and the title terms whose weights had been altered to maintain the focus of the original topics. The new topics were pre-processed by the co-occurrence analysis process to remove extraneous terms. The documents that were retrieved using these new topics were ranked based on the basic formula of the Ocelot algorithm (i.e. without the term dependency and term coverage processing).

The output of the two initial rankings were combined by the data-fusion processes, and the combined ranking was then processed by a Rocchio relevance ranking process as described above. We again processed the newly added terms through the co-occurrence analysis process before ranking the documents for the final time. The final ranking was then fused with the initial 'title' rank, and the output was named LNaTitle7. Table 8 and Figure 2 contain actual values obtained after each intermediate step in the LNaTitle7 entry.

LNaTitle7					***Precision***	
Step	***Run***	***Rel_Ret***	***Avg. Precision***	***Exact Precision***	***At 0.20***	***At 20 docs***
1	Baseline	2178	.1807	.2320	.3308	.3440
2	1 + Nouns + Phrases	2197	.1877	.2408	.3323	.3550
3	2 + REL	1740	.1577	.1981	.2946	.2960
4	Co-occurrence on 3	2618	.2302	.2667	.3860	.3800
5	2 + coverage + termDep	2137	.1892	.2420	.3368	.3810
6	2 + termDep	2241	.1955	.2452	.3413	.3820
7	Data Fusion of (6+4)	2699	.2310	.2728	.3915	.4040
8	Rocchio on 7	2923	.2444	.2700	.4043	.3790
9	Co-occurrence on 8	2916	.2410	.2702	.4096	.3790
10	Data Fusion of (6+9)	2791	.2338	.2691	.3887	.3910

Table 8. The results of the incremental steps in the LNaTitle7

Step 1 shows the ranked results of unmodified queries. After we detected phrases and re-weighted noun terms, we got a boost of just 3.8%. Addition of REL terms from the LEXIS-NEXIS online system reduced both the precision and the recall figures. However, application of the co-occurrence analysis process on the new REL-modified queries improved the average precision by 22.6% to .2302.

Steps 5 and 6 show the results of using term coverage and term dependency factors. Using the term dependency factor improved both the recall and the precision. We got better results by omitting the term coverage factor as can be seen when we compare Step 6 and Step 5.

The data-fusion step, (Step 7), resulted in a marginal improvement of both the precision and recall. The new Rocchio approach helped improve the precision and recall figures by 5.8%. The co-occurrence step after Rocchio and the subsequent data-fusion step didn't add any value to the ranking.

The final data-fusion step had a major negative impact on our results. We believe that the Rocchio relevance feedback run must have retrieved many on-point documents that are ranked higher up in the ranked list. So mixing relevance ranking with a noisier ranking (output of Step 6) automatically added noise to the final results.

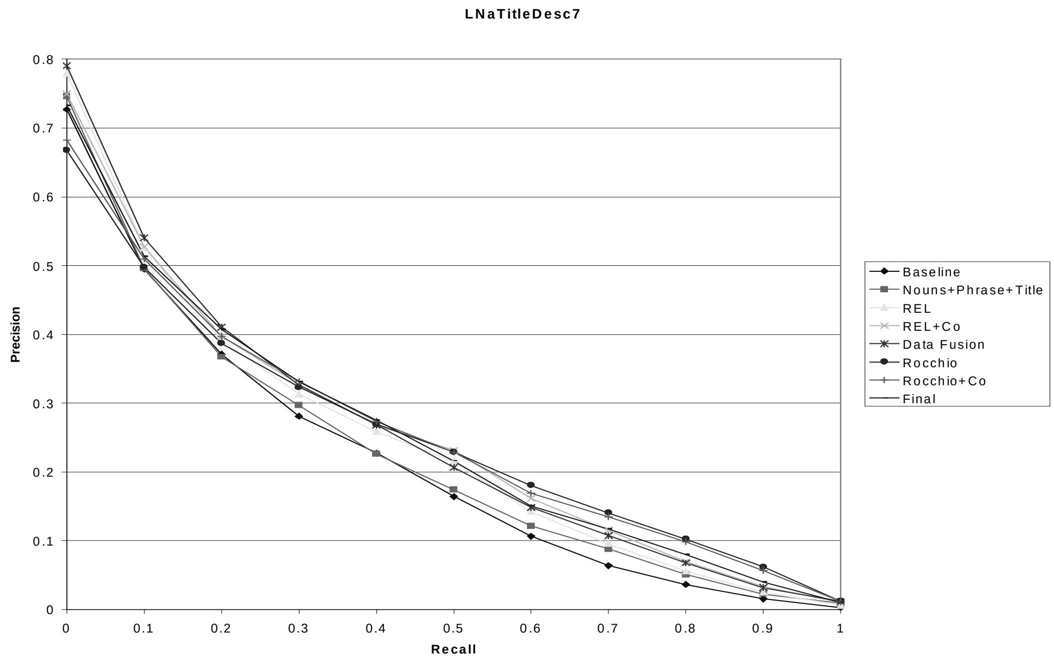

Figure 1. Recall/Precision graph of the processing steps for the LNaTitleDesc7 entry.

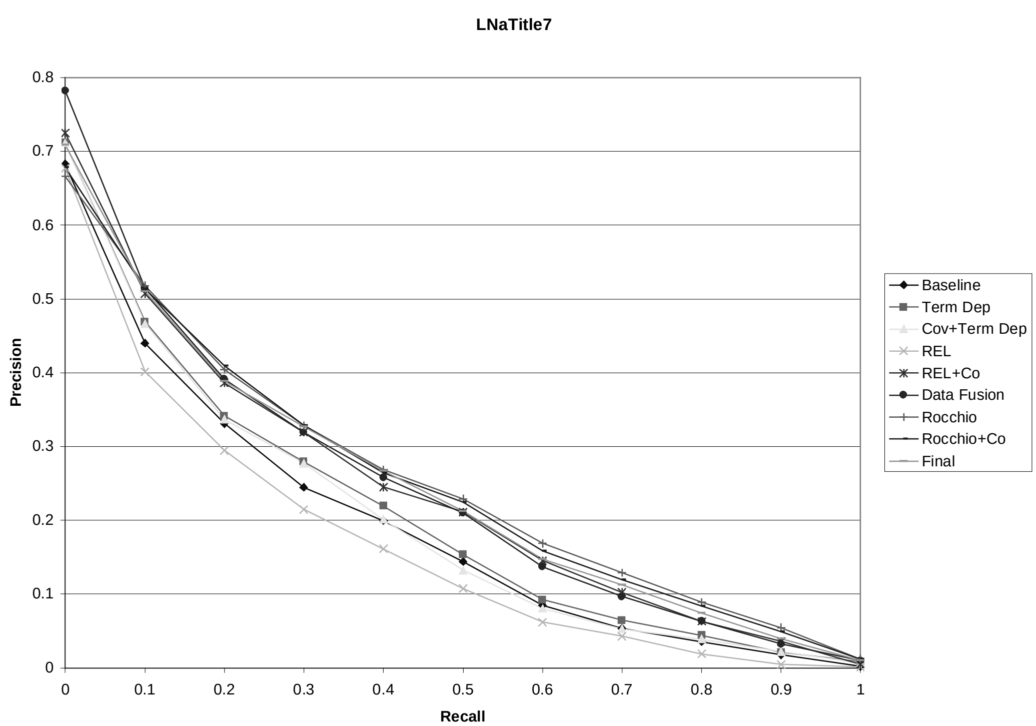

Figure 2. Recall/Precision graph of the processing steps for the LNaTitle7 entry.

2.3 - LNmanual7

We used the Ocelot algorithm to create a sample of 20 documents for each query that human analysts could use to refine the query terms for the manual entry. The algorithm used terms from the title and the description fields of TREC-7 topics. We doubled the weight of the terms in the title, incremented nouns by one and identified and added phrases. No terms from the narrative field were involved in this process. We evaluated these terms and either replaced, re-weighted, or supplemented them using terms from the top 20 documents, the narrative field, and the LEXIS-NEXIS' REL.

Table 9 and Figure 3 present the results of our manual entry. We gained a 28% improvement on the average precision by manually enhancing the queries. The largest gain is at 0.20 precision (i.e. 32%) indicating that the manual intervention improved the ranking of some on-point documents.

LNmanual7					Precision	
Step	**Run**	**Rel_Ret**	**Avg. Precision**	**Exact Precision**	**At 0.20**	**At 20 docs**
1	Baseline	2442	.2030	.2559	.3712	.3950
2	1 + Nouns + Phrases	2581	.2127	.2628	.3672	.4050
3	Manual on 2 (Final)	3005	.2722	.3190	.4848	.4490
4[2]	Co-occurrence on 3	2922	.2606	.3089	.4770	.4510

Table 9. The results of the incremental steps in theTREC-7 manual entry.

The co-occurrence analysis process (Step 4) which has improved the performance of the algorithm in the automatic entries did not have the same contribution as the human judgment. One explanation is that the human analysts made a judgment about each word based on their world knowledge and experience. The co-occurrence analysis process increases precision by eliminating terms that do not co-occur with the title terms. Some of the terms added by the analysts are relevant to the topic even though they don't co-occur with the title terms. These are the terms that are eliminated by the co-occurrence process. Analysis at the query level indicated that the average precision has increased when the co-occurrence analysis was applied to the manual run (Step 3), but because important terms were eliminated, the actual number of relevant documents retrieved was lower. This resulted in higher recall values in Step 3 versus Step 4 as indicated in Figure 3.

Because time was short, we had only one chance to select terms and re-run the document selection. It was not possible to fine-tune the term selection or have multiple iterations to see what kinds of terms were useful.

We could only add phrases to the query that have been recognized by the indexing process. In many occasions, the existence of one unambiguous phrase in a document could indicate relevance but we could not add it to the query because the indexing program hadn't recognized it as a phrase. For example, the phrase *human cargo* is a unique combination of terms for a query about *human smuggling* (TREC-7 topic #362). Unlike the phrase *human cargo*, the two individual terms *human* and *cargo* are very common and they are repeated in many documents.

The absence of proximity indicators was another limiting factor. For example, "technology transfer" and "illegal" may occur within a document but if they are in close proximity, the meaning and relevance change.

We couldn't use numbers although they were significant relevance indicators in some queries. For example, documents about international waters disputes often mentioned the "12 mile" limit and relevant documents about blood alcohol fatalities required a report on the blood alcohol level of the driver.

One of the decisions that the human reader had to make for each added query term was the weight that had to be assigned to it. Terms from the title proved to be more relevant than either the description or the narrative [12]. Therefore, they were usually assigned a much higher weight.

---

[2] Not submitted

Figure 4 compares our three entries in TREC-7. The contribution from terms in the description field was minimal this year because the LnaTitle7 and LnaTitleDesc7 have almost identical results. A subjective survey of TREC-7's 50 topics showed that in at least 9 topics, the description section did not contribute any non-noise words to the query. The median number of useful words in all 50 topics was just 3.

The performance of our manual entry exceeded the automatic entries. This is due to the differences in the term expansion processes. The expansion process in the automatic entries is blind to terms in the expansion set. These terms are pulled in simply by the virtue of occurring in the top ranked documents. On the other hand, the query expansion terms in the manual entry were highly filtered through human experience.

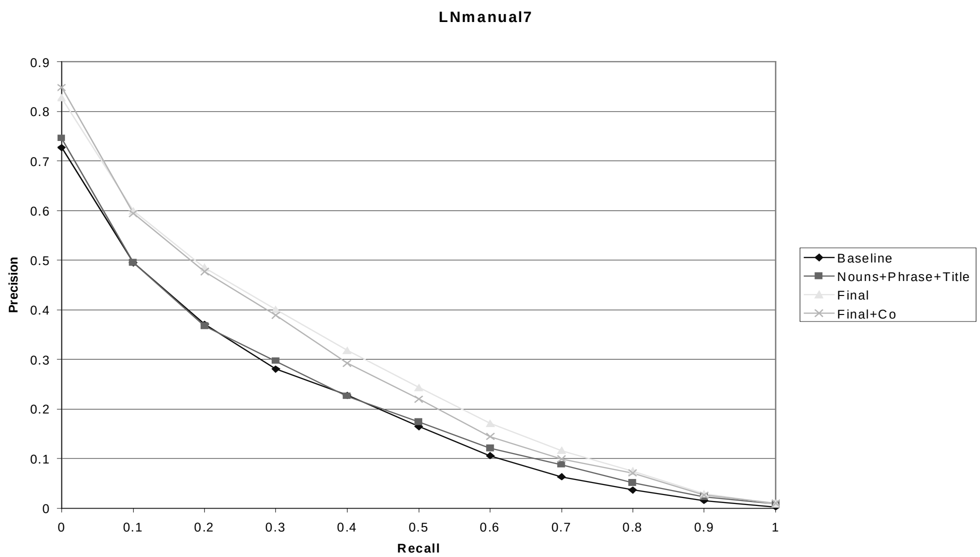

Figure 3. Recall/Precision graph for processing steps for the LNmanual7 entry.

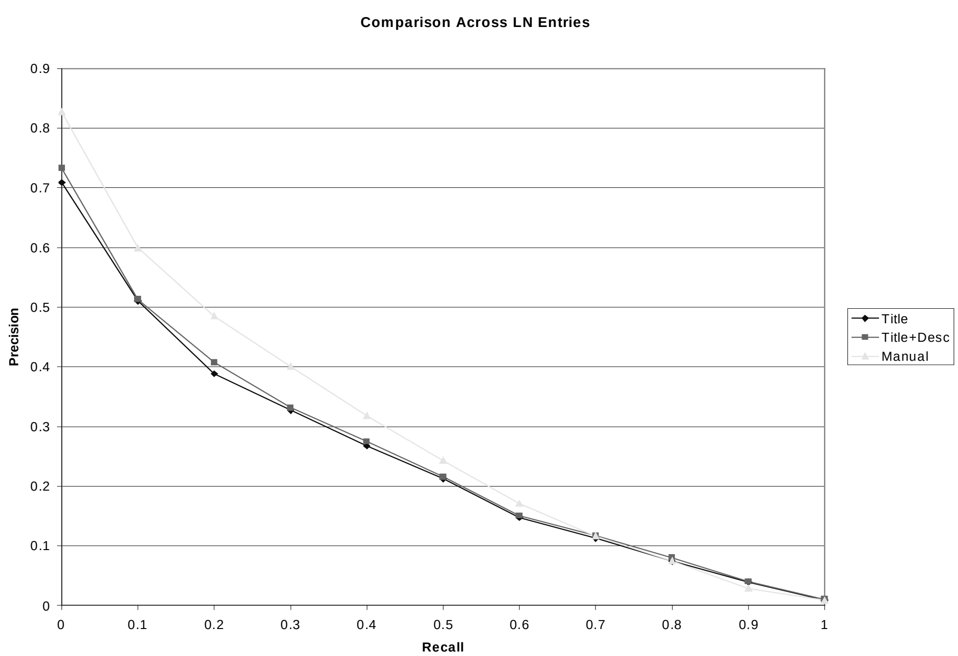

Figure 4. Comparison across all LEXIS-NEXIS entries.

3 - Conclusion

This report described our research effort for TREC-7. We performed a comparative analysis of several well-known search algorithms, and we improved our query enhancement techniques. While we were researching these issues, we developed two new tools to help us create more robust and scaleable search solutions. The first tool helps us fine-tune the co-occurrence metrics. This technique has many applications in the commercial world such as improving text filtering, identification of related concepts and text classification. The second tool helps us distribute the TREC processing over multiple systems to increase the processing speed.

In summary, we found that the new co-occurrence analysis step improved our precision without significantly degrading the recall. We believe that the experience gained by participating in TREC-7 was instrumental in answering some of our research questions but there are still other research issues that we need to explore. We experimented with variations of query term coverage factors, but we didn't come up with a consistent way of improving ranked-results. Another research issue that was raised during this study was the need for finding better ways to identify on-point documents for the relevance feedback process. We also need to find out ways to adjust the performance of our system based on the specific genre of the corpus.

References

[1] Lu, A., Meier, E., Rao, A., Miller, D., and Pliske, D. in "Query Processing in TREC-6", The Sixth Text Retrieval Conference, TREC-6 Notebook, 1997.

[2] Lu, X. A., Ayoub, M., and Dong, J. in "Ad Hoc Experiments using EUREKA" The Fifth Text Retrieval Conference, NIST Special Publication 500-238, pp. 229-240, edited by Voorhees E.M. and Harman D.K.

[3] Lu, X.A., and Keefer, R.B. in "Query expansion/reduction and its impact on retrieval effectiveness", The Third Text Retrieval Conference, NIST Special Publication 500-225, pp. 231-239, edited by Harman D., 1995.

[4] Robertson, S.E., Walker, S., Jones, S., Hancock-Beaulieu, M.M., and Gatford, M. in "Okapi at TREC-3", The Third Text Retrieval Conference, NIST Special Publication 500-225, pp. 109-126, edited by Harman D, 1995.

[5] Broglio, J., Callan, J.P., Croft, W.B., and Nachbar, D.W. in "Document Retrieval and Routing Using the INQUERY System", The Third Text Retrieval Conference, NIST Special Publication 500-225,, pp. 29-38, edited by Harman D, 1995.

[6] Allan, J., Ballesteros, L., Callan, J.P., Croft, W.B., and Lu, Z. in "Recent Experiments with INQUERY", The Fourth Text Retrieval Conference, NIST Special Publication, pp. 49-64 edited by Harman D. 1996.

[7] Singhal, A., Salton, G., Mitra, M., and Buckley, C. in "Document length normalization," Information Processing & Management, Vol. 32, pp.619-633, 1996.

[8] Allan, J., Callan, J., and Croft, W.B. in "INQUERY does battle with TREC-6", The Sixth Text Retrieval Conference, TREC-6 Notebook, 1997.

[9] Miller, G. in "WordNet: An online lexical database", International Journal of Lexicography, Vol. 3(4), 1990

[10] Voorhees, E.M. in "On expanding query vectors with lexically related words," The Second Text Retrieval Conference, NIST Special Publication 500-215, pp. 223-231, 1994.

[11] Rocchio, J.J. in "Relevance Feedback in Information Retrieval", The SMART Retrieval System, pp. 313-323, edited by Salton G.

[12] Voorhees, E.M., and Harman, D. in "Overview of the Sixth Text REtrieval Conference (TREC-6)", The Sixth Text Retrieval Conference, TREC-6 Notebook, 1997.

[13] Robertson, S.E., and Sparck, J.K. in "Relevance weighting of search terms", Journal of the American Society for Information Science 27: p129-146, 1976.

[14] Boyce, B.R, Meadow, C.T., and Kraft, D.H., Measurement in Information Science, Academic Press, pp. 86-87, 1994.

[15] Turtle, H.R., and Croft, W.B in "Inference Networks for document retrieval", Proceedings of the 13th International Conference on Research and Development in Information Retrieval, pp. 1-24, New York: Association for Computing Machinery, 1990.

[16] Belkin, N.J., Cool, C., Croft, W.B., and Callan, J.P. in "The effect of multiple query representations on information retrieval performance", Proceedings of the 16 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 339-346, 1993.

[17] Church, K., and Hanks, P. in "Word association norms, mutual information, and lexicography", Proceedings of the ACL Meeting, pp. 76-83, 1989.