The Impact of Document Collection Characteristics on Information Access in Digital Libraries†

David Kirk Evans
Dept. of Computer Science
Columbia University
NY, NY 10027, USA
devans@cs.columbia.edu

Judith L. Klavans
Center for Research on Information Access
Columbia University
NY, NY 10027, USA
klavans@cs.columbia.edu

Nina Wacholder
Center for Research on Information Access
Columbia University
NY, NY 10027, USA
nina@cs.columbia.edu

ABSTRACT
In this paper, we report on an experiment in which we measure the contribution of words of different grammatical categories—nouns, noun phrases, adjectives, adverbs and verbs—to the performance of an information retrieval system. We compare the performance of the reduced versions of documents across four different text corpora from the Text Retrieval Conference (TREC) totaling 330 MBs in size. We find that in general, reduced versions of documents in which only nouns or only noun phrases are indexed perform as well the original full-text versions of the documents. This is definitive quantitative confirmation of the fact that nominals contribute strongly to the 'aboutness' of documents. Equally importantly, our results show surprising differences among corpora: although the contribution of noun phrases to average precision is high for three of the corpora, noun phrases make much less of a contribution to fourth corpus (the Federal Register). These results suggest that information access for specific collections in large digital libraries can be improved by advanced characterization of a collection using shallow and efficient natural language processing techniques.

KEYWORDS: information retrieval, part of speech tagging, natural language processing, information access

OVERVIEW
One of the over-arching goals of our research is to devise techniques for identifying significant topics in unrestricted domains using a combination of linguistic and statistical techniques. The definition of what information is "significant" is highly dependent upon the task; however, one task that we can obtain quantifiable, though indirect, measures for is IR. As a step in identifying significant topics, we have devised a set of experiments to measure the effect of running an IR system over corpora with different representations of original full text documents. Our reasoning is that document representation is one of the central issues in determining relevance. That is, if an IR system is able to effectively represent potential significant topics in a document collection, then query matching technology will be more effective at actually retrieving documents that are relevant.

In the experiment we report on here, representations are built based on grammatical categories, such as nouns, adjectives and full noun phrases, to allow us to measure the independent contribution of each category across each corpus. We can measure the effect of part of speech identification by the difference in precision and recall from two versions of the collection, one that contains the original document and the other that contains just a reduced version. Figure 1 illustrates an excerpt of original full text and Figure 2 shows a reduced version of that text consisting only of nouns, the NN version: [1]

> Two men were dead; valuable manuscripts had changed hands; false names were involved. The mystery was solved Tuesday as Columbia University announced that it had been given the papers of Ellery Queen, the dashing, upper-crust detective whose adventures helped elevate American detective fiction to an art form. Ellery Queen was created by Brooklyn-born cousins Frederic Dannay and Manfred B. Lee, who entered a 1928 mystery contest that required a pseudonym.

Figure 1: Example full text document

† This work has been supported under NSF Grant IRI-97-12069, "Automatic Identification of Significant Topics in Domain Independent Full Text Analysis", PI's: Judith L. Klavans and Nina Wacholder and NSF Grant CDA-97-53054 "Computationally Tractable Methods for Document Analysis", PI: Nina Wacholder.

[1] From the Associated Press section of the TREC corpus.

men dead manuscripts hands names mystery Tuesday Columbia University papers Ellery Queen detective adventures fiction art form Ellery Queen cousins Frederic Dannay Manfred B. Lee 1928 contest pseudonym

Figure 2: Example noun (NN) version

This paper builds on our previous work (Wacholder et al., under review [14]) which showed that different grammatical categories contribute differently to the performance of an IR system than would be expected from simple frequency proportions. In this paper, we present new results from running a larger set of documents from four different genres of text within the TREC collection. Our results demonstrate two points: first, the grammatical properties of document collections have a measurable effect on information retrieval, and second, that the effect of specific grammatical categories is different depending upon the corpus. For example, our analysis reveals that the noun phrase (NP) and noun (NN) versions generally performed as well as the full text version with respect to recall and precision. That the performance was roughly equivalent is quite surprising, given that corpus size was reduced by an average of 45% for the NP versions and 58.1% for the NN versions.

Our results suggest that a method that uses grammatical information could improve performance of IR by identifying document properties, and thus help alleviate information overload. More specifically, these results suggest that for distributed collections in digital libraries, analysis of the different collections may have an impact upon the tools selected to organize and access information contained in the document sets.

RELATED WORK

Historically, the most effective information retrieval systems have been statistically based, forgoing as much natural language processing as possible due to its computationally intensive nature and dubious contribution to increased performance. Indeed, the impact of natural language processing techniques on information retrieval has been questionable and questioned. For example, Smeaton 1992 [9] reviews the fact that domain dependent ambiguity is poorly handled in keyword based information retrieval. Smeaton contrasts conceptual-based retrieval, where full-fledged NLP is used, with traditional retrieval. The former, although more sophisticated, requires far more processing time and is still unproven as more effective. Salton et al. 1990 [8] identify head-modifier relations, and from this generate index terms. Gay and Croft 1990 [4] extracted nominal compounds to examine their impact on IR.

At the same time, widely available statistical systems, such as the Cornell SMART system (Lesk 1964 [5], Salton 1964 [6], Salton 1971 [7]) have achieved remarkable success in terms of speed, coverage, and efficiency. Recently, techniques using powerful statistical clustering (Chen et al. 1997 [2]) have added other dimensions to successful representation of document sets to deal with information overload. Our approach is to examine the careful integration of shallow natural language techniques into statistical systems. Recent advances in fast statistical part-of-speech (POS) taggers (Tzoukermann and Radev 1998 [12]) and techniques for noun phrase parsing (Zhai 1997 [15], Evans 1998 [3]) make it worthwhile to explore the role that part of speech information can play across multiple genres of text, even faced with the large collection sizes seen today.

Strzalkowksi, Lin, and Carballo 1997 [10] have outlined an approach to NLP using what they call the “stream-based information retrieval model.” This model was conceived to enable a careful evaluation of representation methods, from simple quantitative to complex linguistic methods. Alternative methods of document indexing, called “streams” are performed in parallel, and then weighted for maximum performance. The research that we have undertaken is similar to the head-modifier pairs stream, although our pairs retain strict structure whereas Strzalkowski et al. permit variants. Furthermore, we reduce all noun phrases to the minimal NP unit, called a simplex NP (Wacholder 1998 [13]) whereas Strzalkowski et al. consider largest scope NPs reduced to pairs. The approaches are similar, however, in representing the document in terms of the linguistically based index term, although the details of identifying these terms differ. Our results are consistent in observing that linguistic data improves precision to a modest degree.

THE EXPERIMENT

The goal of the research described in this paper is to measure the differences among text corpora and to assess the effects of these differences on the performance of a statistical IR system. In order to achieve comparable results, we took four corpora (330 megabytes of text) from the Text Retrieval Conference (TREC) CD-ROMs (http://trec.nist.gov), tagged the text with grammatical part of speech (e.g., noun, verb, adjective) and used this tagged text to create different versions of the document consisting only of those grammatical categories. We then ran the SMART retrieval system[3] over these corpora, using topic descriptions from the TREC competition as our queries. We took the manually prepared TREC relevance judgements as our gold standard. By holding the search engine, queries and relevance judgements constant, and by manipulating the document versions, we were able to measure the relative contributions of different grammatical categories to the results. In the remainder of this section, we describe the four aspects of the experiment – search engine, corpus, document proxies, and results in greater detail.

Our test corpus consisted of 330MBs of text from Disks 1 and 2 of the Tipster CD-ROMs prepared for the TREC conferences (TREC 1994 [11]). We used documents from four collections—Associated Press (AP), Federal Register (FR), Wall Street Journal 1990 and 1992 (WSJ) and Ziff Davis (ZD). Table 1 shows the size of the full text corpora

[3] We used SMART (version 11) since it is widely used and publicly available. We indexed all versions using the lnc.ltc weighting scheme.

in MBs, as well as the number of documents in each corpus. The FR corpus is considerably larger than the other three in part because it consists of longer documents, so that a larger corpus is needed to achieve a roughly comparable number of documents.

	AP	FR	WSJ	ZF	ALL
Size	60	120	78	76	334
# Docs	20184	12766	24171	25900	83021

Table 1: Size of original full-text (FT) corpora in MBs of text.

In addition to comparing the performance of SMART over full-text documents, we designed the experiment so as to obtain a quantitative measure of the contribution of different grammatical categories such as nouns and verbs to SMART's performance. These categories are identified by the Alembic system belonging to the Mitre Corporation (Aberdeen et al. 1995 [1]) which we have found to be robust and reliable. However, tagging 330MBs of text took approximately two days, by far the slowest part of the experiment.

For example, consider again the text of Figure 1:

> Two men were dead; valuable manuscripts had changed hands; false names were involved. The mystery was solved Tuesday as Columbia University announced that it had been given the papers of Ellery Queen, the dashing, upper-crust detective whose adventures helped elevate American detective fiction to an art form. Ellery Queen was created by Brooklyn-born cousins Frederic Dannay and Manfred B. Lee, who entered a 1928 mystery contest that required a pseudonym.

this would be reduced to the following for the adjective (JJ) version:

> valuable false dashing upper-crust American detective

Figure 3: Example adjective version

for the noun phrase (NP) version:

> two men valuable manuscripts hands false names mystery Tuesday Columbia University papers Ellery Queen dashing, upper-crust detective adventures Amreican detective fiction art form Ellery Queen Brooklyn-born cousins Frederic Dannay Manfred B. Lee 1928 mystery contest pseudonym

Figure 4: Example noun phrase (NP) version

for the verb (VB) version:

> dead changed involved solved announced given helped elevate created entered required

Figure 5: Example verb (VB) version

Notice the difference between the NN version given in Figure 2, which includes all nouns and proper nouns, whereas the NP version is a superset, and includes all words in simplex noun phrases, e.g. primarily adjectives and some adverbs.

In this experiment, we analyze the contribution of grammatical categories by creating different versions of documents in the corpora. Each version is grammatically motivated: we removed all words except those belonging to a particular grammatical category. The categories we used are nouns (NN), adjectives (JJ), adverbs (RB), verbs (VB) and one phrasal category, noun phrases (NPs) which consists of text from noun phrases, e.g. nouns and adjectives. Table 2 shows the size of each of these reduced corpora and the percentage of reduction in size of the corpora as compared to the full text version.

	AP	FR	WSJ	ZF	ALL
FT	60	120	78	76	334
NP	36 (40%)	70 (42%)	47 (40%)	30 (60%)	183 (45%)
NN	26 (56%)	58 (51%)	34 (57%)	22 (71%)	140 (58.1%)
JJ	5.9 (90%)	11 (91%)	7.4 (92%)	4.6 (94%)	29 (91%)
RB	2.3 (96%)	3 (97%)	3.2 (96%)	1.9 (97%)	--
VB	10.5 (94%)	17 (86%)	10.9 (86%)	6.6 (91%)	--

Table 2: Corpus sizes by version in MBs of text and percentage of reduction in size of the reduced corpora relative to full text.

It is apparent from Table 2 that there are some important differences in size reduction by grammatical category; for example, the NP version of Ziff-Davis (ZF) is reduced by 60%, while the NP version of the other collections is reduced by only about 40%. The NN version of ZF is reduced by 76%, while the NN versions of the other collections are reduced by about 60%.

Our queries consisted of the first 150 topics used in the TREC competitions unmodified except that the words "relevant" and "query" were removed, as were sentences which explicitly referred to topics excluded from the search goal. By using identical queries over all of these different versions of the original full-text corpus, we were able to determine the effects of reducing a document to words of a certain grammatical category, and indirectly the contribution of words of different parts of speech.

	# Rel Qs	Avg # Rel Docs	Med # Rel Docs
AP	74	36.78	12
FR	57	7.67	3
WSJ	132	19.66	14
ZF	74	36.78	12
All	141	52.41	37

Table 3 : Number of relevant queries and average and median numbers of relevant documents per query

For evaluating our results, we used TREC relevance judgements as our gold standard. Table 3 shows the number of queries that contain documents judged relevant for each corpus, the average number of relevant documents per query, and the median number of relevant documents per query. SMART returns a ranked list of documents for each retrieval run, however, the TREC relevance judgments are binary, either relevant or irrelevant. We therefore ignore the ranking information when looking at performance across the versions of the corpus: therefore, in our tests a perfect retrieval run would retrieve any document judged to be relevant before any document judged as irrelevant, but any ordering of the retrieved documents would be acceptable.

We compare results of the different runs using the standard IR precision (P) and recall (R) metrics, where precision measures accuracy, and recall measures coverage.

$$P = \frac{\text{\# Relevant Docs Retrieved}}{\text{Total \#Docs Retrieved}} \qquad R = \frac{\text{\# Relevant Docs Retrieved}}{\text{Total \#Relevant Docs}}$$

We retrieve the 105 documents most highly ranked documents. We compute average precision at ten different target recall points, 0.1, 0.2, …, 0.9, and 1.0. Since we retrieve only the top 105 documents, even using all 105 documents it may not be possible to get an arbitrary recall value. For example, if only 30 relevant documents out of a possible 100 relevant documents are retrieved in the top 105 documents, we are limited to a recall of 30% as a maximum. If we retrieved all of the documents in the collection, the average recall would always be 50%, however, the actual values that we computed reflect the recall using a fixed number, in this case 105, of documents. For each of the ten target recall points, the actual recall that the retrieval run performed at could be less than, greater than, or equal to the target recall. Table 4 reflects the average of the actual recall rates at each of the ten points.

RESULTS

Highlights

Our results clearly show that differences among full and reduced corpora produce different retrieval performance in some unexpected ways. Among the most interesting results of our experiment are the following:

- The full noun phrase and noun versions of the corpora generally performed as well as the full text version with respect to recall and precision. That the performance was roughly equivalent is quite surprising, given that corpus size was reduced by an average of 45% for the NP versions and 58.1% for the NN versions (as shown in Table 4). This is definitive quantitative confirmation of the fact that nominals contribute strongly to the 'aboutness' of documents.

- The verb (VB) and adverb (RB) versions do not perform very well for all versions of all corpora. This is not unexpected, given the general assumption about the contribution of nominals and given the large size reduction effected by removing the other categories. What is interesting, though, is that adverbs and verbs do not consistently perform as poorly as one might have expected. While adverbs has 0% precision and 0% recall for the ZF corpus, it has 29.8% recall and 3.6% precision for the Federal Register (FR) corpus. VB has poor precision in all cases, but does achieve 43.6% recall at 1.9% precision for the FR corpus.

- Although the adjective (JJ) versions of the documents were reduced relative to the full text version by an average of 91%, the JJ versions of the document still produce fairly good recall for WSJ and FR. In terms of precision, the JJ versions do not do well on an absolute measure; however, the JJ versions do far better than VB and RB which are the clear losers in all cases.

- The performance of the Federal Register corpus does not conform to the general patterns of the other corpora. For example, the precision for the NP version of all of the corpora is 46%; however, the performance for the NP version of the FR corpus is only 15.9%.

Discussion of Results

Table 4 shows the average precision and average recall percentages per corpus for each of the different versions. On the x axis of Table 4 are the different versions: FT for full-text, NP for noun phrase, NN for nouns, JJ for adjectives, RB for adverbs, and VB for verbs. On the y axis are the different corpora: the Associated Press (AP) newswire, the Federal Register (FR), the Ziff-Davis news service (ZF), the Wall Street Journal (WSJ), and all corpora together (All).

		FT	NP	NN	JJ	RB	VB
AP	R:	53.4	51.3	53.1	33.2	0.69	10.1
	P:	**47.4**	46.3	44.9	15.3	0.13	2.5
FR	R:	74.5	42.0	73.9	61.2	29.8	43.6
	P:	40.6	15.9	**54.9**	11.2	3.6	1.9
ZF	R:	50.4	48.2	48.7	29.0	0	2.3
	P:	53.3	**53.9**	**53.9**	2.6	0	0.27
WSJ	R:	50.2	49.7	51.5	24.4	4.3	5.6
	P:	46.5	**47.2**	46.9	14.8	0.71	1.6
All	R:	42.2	40.1	42.7	20.8	—	—
	P:	44.9	46.1	**47.0**	14.6		

Table 4: Average actual recall and average precision over 10 target recall points. The highest precision for each corpus is in bold.

Average precision is computed over ten target recall levels

(0.1, 0.2, 0.9, 1.0). In all corpora except the Federal Register, the full text, noun phrase and noun versions of the documents perform approximately equally well in terms of precision. full-text had the highest average precision only for the AP corpus; NP performs best for WSJ and ties with NN for FR.

In general, the precision figures in Table 4 are more meaningful than the recall figures; this is because it is easy to achieve perfect recall by retrieving all documents in the corpus. However, the recall figures are more meaningful when viewed in combination with the precision figures because there is generally a trade-off between precision and recall; the ideal system would achieve perfect precision and perfect recall. The tradeoff is reflected in Figure 6, where precision is plotted at ten levels of target recall. In this figure, the corpus, i.e. FR, AP, WEJ, All, ZF in its reduced version is represented as FRNP for the NP version of FR, FRFT for the full text version of FR, and so on. In this way, the x axis represents a plot of precision vs. recall according to the linguistically determined version of each corpus. Note that we show the three important versions of each corpus, namely the noun phrase (NP), the noun (NN), and of course full-text (FT).

While all of the precision and recall data series follow the typical inverse relationship of precision to recall in Figure 6, the AP and WSJ corpora are most representative, as seen by the group of AP bars (columns 4-6) and WSJ bars (columns 7-9.) This is shown by the downward slope of precision as target recall increases. However, the FR and ZF corpora in columns 1-3 and 13-15 respectively) have different characteristics. First of all, the results for the NP version of the FR corpus are clearly much worse than results for all other versions of all other collections, especially at the higher target recall levels. Observe in column 2 of Figure 6 that, at a target recall level of .1, the full text version of the FR corpus, FRFT, has precision near 50% and the noun version in column 3, FRNN, has a precision of 85%. This means that for FRFT and FRNN, there is a high number of relevant documents in the first 10% of documents retrieved. The precision stays the same until there is a drop in precision at a target recall of about 50%, after which the precision plateaus again; however, the FRNP plateaus at a much lower level than for any other.

Compare all three versions of the Ziff-Davis ZF corpus which produce quite similar results; at target recall of .1, the ZF corpus starts off near 60% precision at .1 recall, and then plateaus at 50%. ZF overall has a higher precision than FR, much less of a drop, and in its best performing versions, is almost flat, as can be clearly observed in the nearly flat columns 13-15 in the graph. This behavior suggests that the ZF corpus has many queries that do well at all precision levels, bring up the average score. A relatively small number of queries do well at the lowest recall levels, bringing the score up slightly, but perform poorly at all other levels, and reduce the precision a uniform level. ZF full-text, ZFFT, and ZF NP perform exactly the same, while ZFNN actually outperforms or ties with ZF full-text at every point.

Results by collection

The second interesting point to observe is the difference that corpus makes on results. Our data clearly shows that the contribution of words based on POS is dependent upon the corpus. NN performs more consistently, and in the cases that it does not do better than full-text, it does not do much worse than it. This result could be explained by the NP version of the corpus containing more text than the NN version, since it contains modifiers other than nouns. This provides more keywords being added to the index of the NP version, which might weight other non-relevant documents higher than relevant documents, pushing relevant documents out of the top 105 retrieved set, thus lowering precision. It seems that the NP version is sensitive to differences in corpus, as it can perform the best in some instances, yet fall to mediocre performance in others.

Results by grammatical category

The results of this experiment show that in general nouns and noun phrases have approximately the same discriminatory power among documents for the SMART

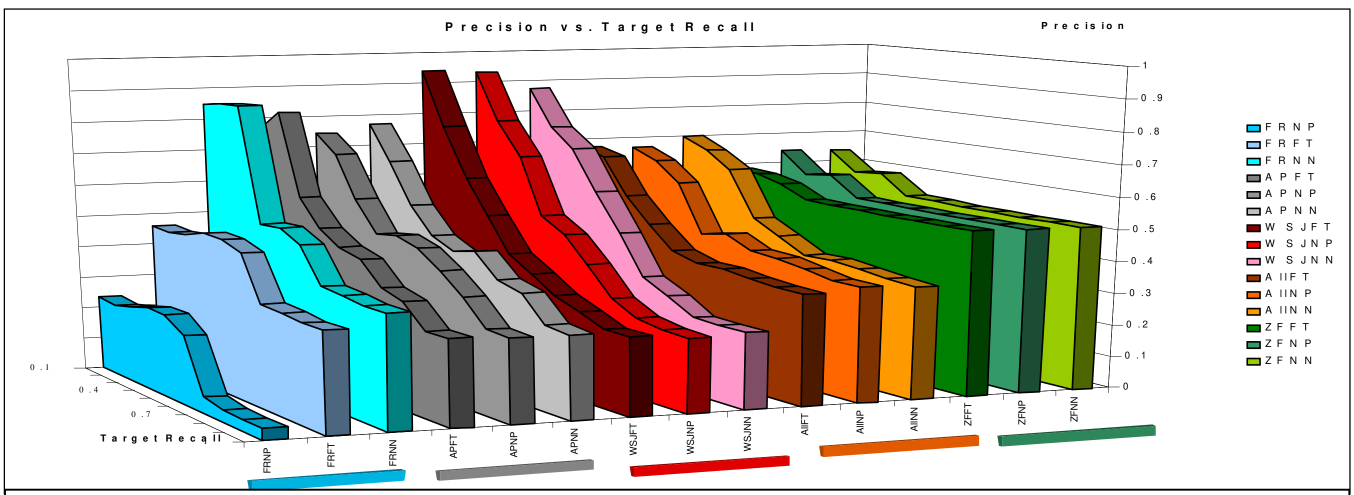

Figure 6: Precision vs. Target Recall

information retrieval system, and that adjectives are much more discriminatory than either verbs or nouns. As a practical matter, the inclusion of the full-text version of the document is a sensible choice, since reducing the documents to NP and NN requires signficant additional processing without providing commensurate improvment in precision or recall.

We plan to add to the NP and NN tests by utilizing some of the data that we gather from the LinkIT system (Evans 1998). This system was built to identify linguistically meaningful relationships between phrases. We have not yet used this information in our version, but we envision adding higher weight to frequent linked nouns, under the hypothesis that these nouns represent cohesion and salience within documents.

However, in general IR results of the system are quite far below the ideal results, which in our experiment would have all relevant documents retrieved before all other documents. For example, the average recall for the full 330MB corpus for full-text documents is 42.2% and the average precision is 44.9%, as shown in Figure 4. This leaves considerable room for improvement. It is clear that one of the limiting factors is the query itself, although the queries we used in our experiment were considerably longer than the typical two or three word query used in most actual searches. However, it is also logical that if signficant information in the document is not included in the document representation, then even the best query will not be sufficiently similar to the impoverished document representation.

One of the surprising results of this experiment is the fact that documents reduced to adjectives performed as well as they did. One explanation for this rests on the discriminatory power of adjectives; an alternative explanation is that in reducing the document, we reduce the total number of keywords in the corpus overall; this reduces the number of keywords in the query and in the document which overlap. If the overlap between document and query occurs for keywords that are not associated with a signficant concept in the document, non-relevant documents will be retrieved. Therefore, reduction of a document to signficant topics in the document might improve retrieval. There are many keywords that are simply not central to the user's information need. Through reduction, we intend not only to highlight important words, but to filter out unimportant or misleading topics and thereby improve results.

CONCLUSIONS AND FUTURE WORK

We have shown that different grammatical categories from different corpora affect the performance of information access. The results have important implications for understanding the match between effective search and retrieval, and the linguistic properties of document collections. The primary focus of our future work will be to utilize additional linguistic and statistical information from natural language processing that we have not yet included in our experiments to further test and analyze the impact of NLP on document analysis. The impact of our research on Digital Libraries will be to enable more targeted access to the large variety of text available to users.

REFERENCES

1. Aberdeen, John, John Burger, David Day, Lynette Hirschman, Patricia Robinson and Marc Vilain. MITRE: Description of the Alembic System as Used for MUC-6, in Proceedings of the Sixth Message Understanding Conference (MUC-6), November 1995.

2. Chen, Hsinchun, Bruce R. Schatz, Andrea L. Houston, Robin R. Sewell, Tobun Dorbin Ng, and Chienting Lin. Internet Browsing and Searching: User Evaluations of Category Map and Concept Space Techniques, in Proceedings of ACM-DL 1997 (Philadelphia, PA, July 23rd – 26th 1997) ACM Press, p. 257.

3. Evans, David Kirk. A Technical Description of the LinkIT System. Unpublished report. Columbia University, Department of Computer Science. 1998.

4. Gay, L. S., and W. B. Croft, Interpreting nominal compounds for information retrieval. *Information Processing and Management* **26** (1), 21-38 (1990).

5. Lesk, Michael E. 1964. The SMART automatic text processing and document retrieval system. Report ISR-8, sec II. Harvard Computation Laboratory, Cambridge, Massachusetts.

6. Salton, Gerard. 1964. A flexible automatic system for the organization, storage, and retrieval of language data (SMART). Report ISR-5, sec. I. Harvard Computation Laboratory, Cambridge, Massachusetts.

7. Salton, Gerard, ed. 1971. *The SMART retrieval system – experiments in automatic document processing.* Englewood Cliffs, New Jersey: Prentice-Hall.

8. Salton, Gerard, et al., On the application of syntactic methodologies in automatic text analysis in *Information Processing and Management* **26** (1), 73-92 (1990).

9. Smeaton, Alan F. Progress in the Application of Natural Language Processing to Information Retrieval Tasks, in The Computer Journal Vol 35. No. 3, 1992.

10. Strzalkowski, Tomek, Fang Lin, and Jose Perez-Carballo. Natural Language Information Retrieval TREC-6 Report in ed. Ellen M. Voorhees and Donna K. Harmon, NIST Special Publication 500-240: The Sixth Text REtrieval Conference (TREC-6), Dept. of Commerce, 1997.

11. Text Research Collection (TREC) Volume 2 (revised March 1994), available from the Linguistic Data Consortium.

12. Tzoukermann, Evelyne and Dragomir R. Radev. Use of Weighted Finite State Transducers in Part of Speech Tagging, in Andras Kornai, editor, *Extended Finite State Models of Language*. Cambridge University Press,

1998.

13. Wacholder, Nina (1998) "Simplex NPs clustered by head: a method for identifying significant topics in a document", Proceedings of Workshop on the Computational Treatment of Nominals, edited by Federica Busa, Inderjeet Mani and Patrick Saint-Dizier, pp.70-79. COLING-ACL, October 16, 1998, Montreal.

14. Wacholder, Nina, David Kirk Evans, and Judith L. Klavans, 1999. An Analysis of the Role of Grammatical Categories in a Statistical Information Retrieval System, under review.

15. Zhai, Chengziang. Fast statistical parsing of noun phrases for document indexing, in *Proceedings of the Fifth Conference on Applied Natural Language Processing*, Association for Computational Linguistics, (Washington D.C., March 31-April 1, 1997)