Producing Biographical Summaries: Combining Linguistic Knowledge with Corpus Statistics[1]

Barry Schiffman
Columbia University
1214 Amsterdam Avenue
New York, NY 10027, USA
Bschiff@cs.columbia.edu

Inderjeet Mani[2]
The MITRE Corporation
11493 Sunset Hills Road
Reston, VA 20190, USA
imani@mitre.org

Kristian J. Concepcion
The MITRE Corporation
11493 Sunset Hills Road
Reston, VA 20190, USA
kjc9@mitre.org

Abstract

We describe a biographical multi-document summarizer that summarizes information about people described in the news. The summarizer uses corpus statistics along with linguistic knowledge to select and merge descriptions of people from a document collection, removing redundant descriptions. The summarization components have been extensively evaluated for coherence, accuracy, and non-redundancy of the descriptions produced.

1 Introduction

The explosion of the World Wide Web has brought with it a vast hoard of information, most of it relatively unstructured. This has created a demand for new ways of managing this often unwieldy body of dynamically changing information. The goal of automatic text summarization is *to take a partially-structured source text, extract information content from it, and present the most important content in a condensed form in a manner sensitive to the needs of the user and task* (Mani and Maybury 1999). Summaries can be 'generic', i.e., aimed at a broad audience, or topic-focused, i.e., tailored to the requirements of a particular user or group of users. Multi-Document Summarization (MDS) is, by definition, the extension of single-document summarization to collections of related documents. MDS can potentially help the user to see at a glance what a collection is about, or to examine similarities and differences in the information content in the collection.

Specialized multi-document summarization systems can be constructed for various applications; here we discuss a biographical summarizer. Biographies can, of course, be long, as in book-length biographies, or short, as in an author's description on a book jacket. The nature of descriptions in the biography can vary, from physical characteristics (e.g., for criminal suspects) to scientific or other achievements (e.g., a speaker's biography). The crucial point here is that facts about a person's life are selected, organized, and presented so as to meet the compression and task requirements.

While book-quality biographies are out of reach of computers, many other kinds can be synthesized by sifting through large quantities of on-line information, a task that is tedious for humans to carry out. We report here on the development of a biographical MDS summarizer that summarizes information about people described in the news. Such a summarizer is of interest, for example, to analysts who want to automatically construct a dossier about a person over time.

Rather than determining in advance what sort of information should go into a

[1] This work has been funded by DARPA's Translingual Information Detection, Extraction, and Summarization (TIDES) research program, under contract number DAA-B07-99-C-C201 and ARPA Order H049.
[2] Also at the Department of Linguistics, Georgetown University, Washington, D. C. 20037.

biography, our approach is more data-driven, relying on discovering how people are actually described in news reports in a collection. We use corpus statistics from a background corpus along with linguistic knowledge to select and merge descriptions from a document collection, removing redundant descriptions. The focus here is on synthesizing *succinct* descriptions. The problem of assembling these descriptions into a coherent narrative is not a focus of our paper; the system currently uses canned text methods to produce output text containing these descriptions. Obviously, the merging of descriptions should take temporal information into account; this very challenging issue is also not addressed here.

To give a clearer idea of the system's output, here are some examples of biographies produced by our system (the descriptions themselves are underlined, the rest is canned text). The biographies contain descriptions of the salient attributes and activities of people in the corpus, along with lists of their associates. These short summaries illustrate the extent of compression provided. The first two summaries are of a collection of 1300 wire service news documents on the Clinton impeachment proceedings (707,000 words in all, called the 'Clinton' corpus). In this corpus, there are 607 sentences mentioning Vernon Jordan by name, from which the system extracted 82 descriptions expressed as appositives (78) and relative clauses (4), along with 65 descriptions consisting of sentences whose deep subject is Jordan. The 4 relative clauses are duplicates of one another: "who helped Lewinsky find a job". The 78 appositives fall into just 2 groups: "friend" (or equivalent descriptions, such as "confidant"), "adviser" (or equivalent such as "lawyer"). The sentential descriptions are filtered in part based on the presence of verbs like "testify, "plead", or "greet" that are strongly associated with the head noun of the appositive, namely "friend". The target length can be varied to produce longer summaries.

Vernon Jordan *is a presidential friend and a Clinton adviser. He is 63 years old. He helped Ms. Lewinsky find a job. He testified that Ms. Monica Lewinsky said that she had conversations with the president, that she talked to the president. He has numerous acquaintances, including Susan Collins, Betty Currie, Pete Domenici, Bob Graham, James Jeffords and Linda Tripp.*

1,300 docs, 707,000 words (**Clinton** corpus) 607 Jordan sentences, 78 extracted appositives, 2 groups: friend, adviser.

Henry Hyde *is a Republican chairman of House Judiciary Committee and a prosecutor in Senate impeachment trial. He will lead the Judiciary Committee's impeachment review. Hyde urged his colleagues to heed their consciences , "the voice that whispers in our ear , 'duty, duty, duty.'"*

Clinton corpus, 503 Hyde sentences, 108 extracted appositives, 2 groups: chairman, impeachment prosecutor.

Victor Polay *is the Tupac Amaru rebels' top leader, founder and the organization's commander-and-chief. He was arrested again in 1992 and is serving a life sentence. His associates include Alberto Fujimori, Tupac Amaru Revolutionary, and Nestor Cerpa.*

73 docs, 38,000 words, 24 Polay sentences, 10 extracted appositives, 3 groups: leader, founder and commander-in-chief.

2 Producing biographical descriptions

2.1 Preprocessing

Each document in the collection to be summarized is processed by a sentence tokenizer, the Alembic part-of-speech tagger (Aberdeen et al. 1995), the Nametag named entity tagger (Krupka 1995) restricted to people names, and the CASS parser (Abney 1996). The tagged sentences are further analyzed by a cascade of finite state machines leveraging patterns with lexical and syntactic information, to identify constructions such as pre- and post-modifying appositive phrases, e.g., "Presidential candidate George Bush", "Bush, the presidential candidate", and relative clauses, e.g., "Senator ..., who is running for re-election this Fall,". These appositive phrases and relative clauses capture descriptive information which can correspond variously to a person's age, occupation, or some role a person played in an incident. In addition, we also extract sentential

descriptions in the form of sentences whose (deep) subjects are person names.

2.2 Cross-document coreference

The classes of person names identified within each document are then merged across documents in the collection using a cross-document coreference program from the Automatic Content Extraction (ACE) research program (ACE 2000), which compares names across documents based on similarity of a window of words surrounding each name, as well as specific rules having to do with different ways of abbreviating a person's name (Mani and MacMillan 1995). The end result of this process is that for each distinct person, the set of descriptions found for that person in the collection are grouped together.

2.3 Appositives

2.3.1 Introduction

The appositive phrases usually provide descriptions of attributes of a person. However, the preprocessing component described in Section 2.1 does produce errors in appositive extraction, which are filtered out by syntactic and semantic tests. The system also filters out redundant descriptions, both duplicate descriptions as well as similar ones. These filtering methods are discussed next.

2.3.2 Pruning Erroneous and Duplicate Appositives

The appositive descriptions are first pruned to record only one instance of an appositive phrase which has multiple repetitions, and descriptions whose head does not appear to refer to a person. The latter test relies on a person typing program which uses semantic information from WordNet 1.6 (Miller 1995) to test whether the head of the description is a person. A given string is judged as a person if a threshold percentage θ_1 (set to 35% in our work) of senses of the string are descended from the synset for Person in WordNet. For example, this picks out "counsel" as a person, but "accessory" as a non-person.

2.3.3 Merging Similar Appositives

The pruning of erroneous and duplicate descriptions still leaves a large number of redundant appositive descriptions across documents. The system compares each pair of appositive descriptions of a person, merging them based on corpus frequencies of the description head stem, syntactic information, and semantic information based on the relationship between the heads in WordNet. The descriptions are merged if they have the same head stem, or if both heads have a common parent below Person in WordNet (in the latter case the head which is more frequent in the corpus is chosen as the merged head), or if one head subsumes the other under Person in WordNet (in which case the more general head is chosen).

When the heads of descriptions are merged, the most frequent modifying phrase that appears in the corpus with the selected head is used. When a person ends up with more than one description, the modifiers are checked for duplication, with distinct modifiers being conjoined together, so that "Wisconsin lawmaker" and "Wisconsin democrat" yields "Wisconsin lawmaker and Democrat". Prepositional phrase variants of descriptions are also merged here, so that "chairman of the Budget Committee" and "Budget Committee Chairman" are merged. Modifiers are dropped but their original order is preserved for the sake of fluency.

2.3.4 Appositive Description Weighting

The system then weights the appositives for inclusion in a summary. A person's appositives are grouped into equivalence classes, with a single head noun being chosen for each equivalence class, with a weight for that class based on the corpus frequency of the head noun. The system then picks descriptions in decreasing order of class weight until either the compression rate is achieved or the head noun is no longer in the top θ_2 % most frequent descriptions (θ_2 is set to 90% in our work). Note that the summarizer refrains from choosing a subsuming term from WordNet that is not present in the descriptions, preferring to not risk inventing new descriptions, instead confining

itself to cutting and pasting of actual words used in the document.

2.4 Relative Clause Weighting

Once the relative clauses have been pruned for duplicates, the system weights the appositive clauses for inclusion in a summary. The weighting is based on how often the relative clause's main verb is strongly associated with a (deep) subject in a large corpus, compared to its total number of appearances in the corpus. The idea here is to weed out 'promiscuous' verbs that are weakly associated with lots of subjects. The corpus statistics are derived from the Reuters portion of the North American News Text Corpus (called '**Reuters**' in this paper) -- nearly three years of wire service news reports containing 105.5 million words.

Examples of verbs in the Reuters corpus which show up as promiscuous include "get", "like", "give", "intend", "add", "want", "be", "do", "hope", "think", "make", "dream", "have", "say", "see", "tell", "try". In a test, detailed below in Section 4.2, this feature fired 40 times in 184 trials.

To compute strong associations, we proceed as follows. First, all subject-verb pairs are extracted from the Reuters corpus with a specially developed finite state grammar and the CASS parser. The head nouns and main verbs are reduced to their base forms by changing plural endings and tense markers for the verbs. Also included are 'gapped' subjects, such as the subject of "run" in "the student promised to run the experiment"; in this example, both pairs 'student-promise' and 'student-run' are recorded. Passive constructions are also recognized and the object of the by-PP following the verb is taken as the deep subject.
Strength of association between subject *i* and verb *j* is measured using mutual information (Church and Hanks 1990):

$$MI(i,j)=\ln\left(\frac{N \cdot tf_{ij}}{tf_i \cdot tf_j}\right).$$

Here tf_{ij} is the maximum frequency of subject-verb pair *ij* in the Reuters corpus, tf_i is the frequency of subject head noun *i* in the corpus, tf_j is the frequency of verb *j* in the corpus, and *N* is the number of terms in the corpus. The associations are only scored for *tf* counts greater than 4, and a threshold θ_3 (set to log score > -21 in our work) is used for a strong association.

The relative clauses are thus filtered initially (Filter 1) by excluding those whose main verbs are highly promiscuous. Next, they are filtered (Filter 2) based on various syntactic features, as well as the number of proper names and pronouns. Finally, the relative clauses are scored conventionally (Filter 3) by summing the within-document relative term frequency of content terms in the clause (i.e., relative to the number of terms in the document), with an adjustment for sentence length (achieved by dividing by the total number of content terms in the clause).

3 Sentential Descriptions

These descriptions are the relatively large set of sentences which have a person name as a (deep) subject. We filter them based on whether their main verb is strongly associated with *either of the head nouns of the appositive descriptions found for that person name* (Filter 4). The intuition here is that particular occupational roles will be strongly associated with particular verbs. For example, politicians vote and elect, executives resign and appoint, police arrest and shoot; so, a summary of information about a policeman may include an arresting and shooting event he was involved with. (The verb-occupation association isn't manifest in relative clauses because the latter are too few in number).

A portion of the results of doing this is shown in Table 1. The results for "executive" are somewhat loose, whereas for "politician" and "police", the associations seem tighter, with the associated verbs meeting our intuitions.

All sentences which survive Filter 4 are extracted and then scored, just as relative clauses are, using Filter 1 and Filter 3. Filter 4 alone provides a high degree of compression; for example, it reduces a total of 16,000 words in the combined sentences that include Vernon Jordan' s name in the Clinton corpus to 578 words in 12 sentences; sentences up to the target length can be selected from these based on scores from Filter 1 and then Filter 3.

However, there are several difficulties with these sentences. First, we are missing a lot of them due to the fact that we do not as yet handle

pronominal subjects which are coreferential with the proper name. Second, these sentences contain lots of dangling anaphors, which will need to be resolved. Third, there may be redundancy between the sentential descriptions, on one hand, and the appositive and relative clause descriptions, on the other. Finally, the entire sentence is extracted, including any subordinate clauses, although we are working on refinements involving sentence compaction. As a result, we believe that more work is required before the sentential descriptions can be fully integrated into the biographies.

executive	police	politician
reprimand 16.36	shoot 17.37	clamor 16.94
conceal 17.46	raid 17.65	jockey 17.53
bank 18.27	arrest 17.96	wrangle 17.59
foresee 18.85	detain 18.04	woo 18.92
conspire 18.91	disperse 18.14	exploit 19.57
convene 19.69	interrogate 18.36	brand 19.65
plead 19.83	swoop 18.44	behave 19.72
sue 19.85	evict 18.46	dare 19.73
answer 20.02	bundle 18.50	sway 19.77
commit 20.04	manhandle 18.59	criticize 19.78
worry 20.04	search 18.60	flank 19.87
accompany 20.11	confiscate 18.63	proclaim 19.91
own 20.22	apprehend 18.71	annul 19.91
witness 20.28	round 18.78	favor 19.92
testify 20.40	corner 18.80	denounce 20.09
shift 20.42	pounce 18.81	condemn 20.10
target 20.56	hustle 18.83	prefer 20.14
lie 20.58	nab 18.83	wonder 20.18
expand 20.65	storm 18.90	dispute 20.18
learn 20.73	tear 19.00	interfere 20.37
shut 20.80	overpower 19.09	voice 20.38

Table 1. Verbs strongly associated with particular classes of people in the Reuters corpus (negative log scores).

4 Evaluation

4.1 Overview

Methods for evaluating text summarization can be broadly classified into two categories (Sparck-Jones and Galliers 1996). The first, an extrinsic evaluation, tests the summarization based on how it affects the completion of some other task, such as comprehension, e.g., (Morris et al. 1992), or relevance assessment (Brandow et al. 1995) (Jing et al. 1998) (Tombros and Sanderson 1998) (Mani et al. 1998). An intrinsic evaluation, on the other hand, can involve assessing the **coherence** of the summary (Brandow et al. 1995) (Saggion and Lapalme 2000).

Another intrinsic approach involves assessing the **informativeness** of the summary, based on to what extent key information from the source is preserved in the system summary at different levels of compression (Paice and Jones 1993), (Brandow et al. 1995). Informativeness can also be assessed in terms of how much information in an ideal (or 'reference') summary is preserved in the system summary, where the summaries being compared are at similar levels of compression (Edmundson 1969).

We have carried out a number of intrinsic evaluations of the accuracy of components involved in the summarization process, as well as the succinctness, coherence and informativeness of the descriptions. As this is a MDS system, we also evaluate the non-redundancy of the descriptions, since similar information may be repeated across documents.

4.2 Person Typing Evaluation

The component evaluation tests how accurately the tagger can identify whether a head noun in a description is appropriate as a person description The evaluation uses the WordNet 1.6 SEMCOR semantic concordance, which has files from the **Brown** corpus whose words have semantic tags (created by WordNet' s creators) indicating WordNet sense numbers. Evaluation on 6,000 sentences with almost 42,000 nouns compares people tags generated by the program with SEMCOR tags, and provided the following results: right = 41,555, wrong = 1,298, missing = 0, yielding Precision, Recall, and F-Measure of 0.97.

4.3 Relative Clause Extraction Evaluation

This component evaluation tests the well-formedness of the extracted relative clauses. For this evaluation, we used the Clinton corpus. The

relative clause is judged correct if it has the right extent, and the correct coreference index indicating which person the relative clause description pertains to. The judgments are based on 36 instances of relative clauses from 22 documents. The results show 28 correct relative clauses found, plus 4 spurious finds, yielding Precision of 0.87, Recall of 0.78, and F-measure of .82. Although the sample is small, the results are very promising.

4.4 Appositive Merging Evaluation

This component evaluation tests the system's ability to accurately merge appositive descriptions. The score is based on an automatic comparison of the system's merge of system-generated appositive descriptions against a human merge of them. We took all the names that were identified in the Clinton corpus and ran the system on each document in the corpus. We took the raw descriptions that the system produced before merging, and wrote a brief description by hand for each person who had two or more raw descriptions. The hand-written descriptions were not done with any reference to the automatically merged descriptions nor with any reference to the underlying source material. The hand-written descriptions were then compared with the final output of the system (i.e., the result after merging). The comparison was automatic, measuring similarity among vectors of content words (i.e., stop words such as articles and prepositions were removed).

Here is an example to further clarify the strict standard of the automatic evaluation (words scored correct are underlined):

System: E. Lawrence Barcella is a *Washington lawyer, Washington white-collar defense lawyer, former federal prosecutor*
System Merge: *Washington white-collar defense lawyer*
Human Merge: *a Washington lawyer and former federal prosecutor*
Automatic Score: Correct=2; Extra-Words=2; Missed-Words=3

Thus, although 'lawyer' and 'prosecutor' are synonymous in WordNet, the automatic scorer doesn't know that, and so 'prosecutor' is penalized as an extra word.

The evaluation was carried out over the entire Clinton corpus, with descriptions compared for 226 people who had more than one description. 65 out of the 226 descriptions were Correct (28%), with a further 32 cases being semantically correct 'obviously similar' substitutions which the automatic scorer missed (giving an adjusted accuracy of 42%). As a baseline, a merging program which performed just a string match scored 21% accuracy. The major problem areas were errors in coreference (e.g., Clinton family members being put in the same coreference class), lack of good descriptions for famous people (news articles tend not to introduce such people), and parsing limitations (e.g., "Senator Clinton" being parsed erroneously as an NP in "The Senator Clinton disappointed…"). Ultimately, of course, domain-independent systems like ours are limited semantically in merging by the lack of world knowledge, e.g., knowing that Starr' s chief lieutenant can be a prosecutor.

4.5 Description Coherence and Informativeness Evaluation

To assess the coherence and informativeness of the relative clause descriptions[3], we asked 4 subjects who were unaware of our research to judge descriptions generated by our system from the **Clinton** corpus. For each relative clause description, the subject was given the description, a person name to whom that description pertained, and a capsule description consisting of merged appositives created by the system. The subject was asked to assess (a) the **coherence** of the relative clause description in terms of its *succinctness* (was it a good length?) and its *comprehensibility* (was it and understandable by itself or in conjunction with the capsule?), and (b) its **informativeness** in terms of whether it was an *accurate* description (does it conflict with the capsule or with what you know?) and whether it was *non-redundant* (is it distinct or does it repeat what is in the capsule?).

The subjects marked 87% of the descriptions as accurate, 96% as non-redundant, and 65% as coherent. A separate 3-subject inter-

[3] Appositives are not assessed in this way as few errors of coherence or informativeness were noticed in the appositive extraction.

annotator agreement study, where all subjects judged the same 46 decisions, showed that all three subjects agreed on 82% of the accuracy decisions, 85% of the non-redundancy decisions and 82% of the coherence decisions.

5 Learning to Produce Coherent Descriptions

5.1 Overview

To learn rules for coherence for extracting sentential descriptions, we used the examples and judgments we obtained for coherence in the evaluation of relative clause descriptions in Section 4.5. Our focus was on features that might relate to content and specificity: low verb promiscuity scores, presence of proper names, pronouns, definite and indefinite clauses. The entire list is as follows:

badend:	boolean. is there an impossible end, indicating a bad extraction (... Mr.)?
bestverb:	continuous. use the verb promiscuity threshhold θ_3 to find the score of the most non-promiscuous verb in the clause
classes (label):	boolean. accept the clause, reject the clause
count pronouns:	continuous. number of personal pronouns
count proper:	continuous. number of nouns tagged as NP
hasobject:	continuous. how many np's follow the verb?
haspeople:	continuous. how many "name" constituents are found?
has possessive:	continuous. how many possessive pronouns are there?
hasquote:	boolean. is there a quotation?
hassubc:	boolean. is there a subordinate clause?
isdefinite:	continuous. how many definite NP's are there?
repeater:	boolean. is the subject's name repeated, or is there no subject?
timeref:	boolean. is there a time reference?
withquit:	is there a "quit" or "resign" verb?
withsay:	boolean. is there a "say" verb in the clause?

5.2 Accuracy of Learnt Descriptions

Table 2 provides information on different learning methods. The results are for a ten-fold cross-validation on 165 training vectors and 19 test vectors, measured in terms of Predictive Accuracy (percentage test vectors correctly classified).

Tool	Accuracy
Barry's Rules	.69
MC4 Decision Tree	.69
C4.5Rules	.67
Ripper	.62
Naive Bayes	.62
Majority Class (coherent)	.60

Table 2. Accuracy of Different Description Learners on Clinton corpus

The best learning methods are comparable with rules created by hand by one of the authors (Barry's rules). In the learners, the *bestverb* feature is used heavily in tests for the negative class, whereas in Barry's Rules it occurs in tests for the positive class.

6 Related Work

Our work on measuring subject-verb associations has a different focus from the previous work. (Lee and Pereira 1999), for example, examined verb-object pairs. Their focus was on a method that would improve techniques for gathering statistics where there are a multitude of sparse examples. We are focusing on the use of the verbs for the specific purpose of finding associations that we have previously observed to be strong, with a view towards selecting a clause or sentence, rather than just to measure similarity. We also try to strengthen the numbers by dealing with 'gapped' constructions.

While there has been plenty of work on extracting named entities and relations between them, e.g., (MUC-7 1998), the main previous body of work on biographical summarization is that of (Radev and McKeown 1998). The fundamental differences in our work are as follows: (1) We extract not only appositive phrases, but also clauses at large based on

corpus statistics; (2) We make heavy use of coreference, whereas they don't use coreference at all; (3) We focus on generating succinct descriptions by removing redundancy and merging, whereas they categorize descriptions using WordNet, without a focus on succinctness.

7 Conclusion

This research has described and evaluated techniques for producing a novel kind of summary called biographical summaries. The techniques use syntactic analysis and semantic type-checking (from WordNet), in combination with a variety of corpus statistics. Future directions could include improved sentential descriptions as well as further intrinsic and extrinsic evaluations of the summarizer as a whole (i.e., including canned text).

References

J. Aberdeen, J. Burger, D. Day, L. Hirschman, P. Robinson, and M. Vilain. 1995. "MITRE: Description of the Alembic System system as used for MUC-6". In Proceedings of the Sixth Message Understanding Conference (MUC-6), Columbia, Maryland.

S. Abney. 1996. "Partial parsing Via Finite-State Cascades". *Proceedings of the ESSLLI '96 Robust Parsing Workshop*.

Automatic Context Extraction Program. http://www.nist.gov/speech/tests/ace/index.htm

R. Brandow, K. Mitze, and L. Rau. 1995. "Automatic condensation of electronic publications by sentence selection." *Information Processing and Management* 31(5): 675-685. Reprinted in *Advances in Automatic Text Summarization*, I. Mani and M.T. Maybury (eds.), 293-303. Cambridge, Massachusetts: MIT Press.

K. W. Church and P. Hanks. 1990. "Word association norms, mutual information, and lexicography". *Computational Linguistics* 16(1): 22-29.

H. P. Edmundson. 1969. "New methods in automatic abstracting". *Journal of the Association for Computing Machinery* 16 (2): 264-285. Reprinted in *Advances in Automatic Text Summarization*, I. Mani and M.T. Maybury (eds.), 21-42. Cambridge, Massachusetts: MIT Press.

G. Krupka. 1995. "SRA: Description of the SRA system as used for MUC-6". In Proceedings of the Sixth Message Understanding Conference (MUC-6), Columbia, Maryland.

L. Lee and F. Pereira. 1999. "Distributional Similarity Models: Clustering vs. Nearest Neighbors". In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, 33-40.

I. Mani and T. MacMillan. 1995. "Identifying Unknown Proper Names in Newswire Text". In *Corpus Processing for Lexical Acquisition*, B. Boguraev and J. Pustejovsky (eds.), 41-73. Cambridge, Massachusetts: MIT Press.

I. Mani and M. T. Maybury. (eds.). 1999. *Advances in Automatic Text Summarization*. Cambridge, Massachusetts: MIT Press.

G. Miller. 1995. "WordNet: A Lexical Database for English". *Communications of the Association For Computing Machinery (CACM)* 38(11): 39-41.

A. Morris, G. Kasper, and D. Adams. 1992. "The Effects and Limitations of Automatic Text Condensing on Reading Comprehension Performance". *Information Systems Research* 3(1): 17-35. Reprinted in *Advances in Automatic Text Summarization*, I. Mani and M.T. Maybury (eds.), 305-323. Cambridge, Massachusetts: MIT Press.

MUC-7. 1998. Proceedings of the Seventh Message Understanding Conference, DARPA.

C. D. Paice and P. A. Jones. 1993. "The Identification of Important Concepts in Highly Structured Technical Papers." In *Proceedings of the 16th International Conference on Research and Development in Information Retrieval (SIGIR'93)*, 69-78.

D. R. Radev and K. McKeown. 1998. "Generating Natural Language Summaries from Multiple On-Line Sources". *Computational Linguistics* 24(3): 469-500.

H. Saggion and G. Lapalme. 2000. "Concept Identification and Presentation in the Context of Technical Text Summarization". In *Proceedings of the Workshop on Automatic Summarization*, 1-10.

K. Sparck-Jones and J. Galliers. 1996. *Evaluating Natural Language Processing Systems: An Analysis and Review*. Lecture Notes in Artificial Intelligence 1083. Berlin: Springer.

A. Tombros and M. Sanderson. 1998."Advantages of query biased summaries in information retrieval". In *Proceedings of the 21st International Conference on Research and Development in Information Retrieval (SIGIR'98)*, 2-10.