A Scalable Summarization System Using Robust NLP

Chinatsu Aone†, Mary Ellen Okurowski‡, James Gorlinsky†, Bjornar Larsen†

†SRA International
4300 Fair Lakes Court
Fairfax, VA 22033
{aonec, gorlinsk, larsenb}@sra com

‡Department of Defense
9800 Savage Road
Fort George G Meade, MD 20755-6000
meokuro@afterlife ncsc mil

Abstract

We describe a scalable summarization system which takes advantage of robust NLP technology such as corpus-based statistical NLP techniques, information extraction and readily available on-line resources The system attempts to compensate for the bottlenecks of traditional frequency-based, knowledge-based or discourse-based summarization approaches by utilizing features derived by these robust techniques Preliminary evaluation results are reported, and the multi-dimensional summary viewer is described

1 Introduction

Summarization research and system development can be broadly characterized as frequency-based, knowledge-based or discourse-based These categories correspond to a continuum of increasing understanding of a text and increasing complexity in text processing

Earliest attempts at summarization (Luhn, 1958, Edmundson, 1969, Rush, Salvador, and Zamora, 1971) essentially relied on lexical and locational information within the text, i e , frequency of words or key terms, their proximity, and location within the text More recent adaptations of this approach have employed an automated method to combine these types of feature sets through classification techniques (Kupiec, Pedersen, and Chen, 1995) or have drawn upon traditional information retrieval indexing methods to incorporate knowledge of a text corpus (Brandow, Mitze, and Rau, 1995) To a large extent, these types of shallow approaches are ignorant of domain knowledge and the text macrostructure They create summaries by extracting sentences from the original document

Knowledge-based approaches generally depend on rich domain knowledge sources to interpret the conceptual structure of the text Systems like TOPIC (Reimer and Hahn, 1988), SUSY (Fum, Guida, and Tasso, 1985) or SCISORS (Rau, Jacobs, and Zernik, 1989) parse domain specific texts and create conceptual representations for the generation of text summaries These types of knowledge-based systems apply knowledge of the domain to characterize specific conceptual knowledge of a text Paice (Paice and Jones, 1993) provides a good example of the role of this conceptual information and Riloff (Riloff, 1995) gives a method for automatically identifying relevant concepts highly correlated with a category of interest Because these systems create a rich conceptual representation, there are multiple ways in which a text summary may be created For example, SUMMONS (McKeown and Radev, 1995) generates a text summary from such a template representation, while (Maybury, 1995) describes multiple methods for selecting events and presenting event summaries Knowledge-based approaches are usually very knowledge-intensive and domain-specific

Discourse-based approaches are grounded in theories of text cohesion and coherence and vary considerably in how much they push the limits of text understanding and the complexity as well as automation of that processing Spearheaded by the lack of cohesion and coherence in extracts produced by frequency-based approaches, much of the work typifying discourse-based approaches focuses on linguistic processing of the text to identify the best cohesive sentence candidates (Paice, 1990, Johnson et al , 1993) or the best sentence candidates for representing the rhetorical structure of the text (Mike et al , 1994) Both approaches involve parsing the text and analyzing discourse relations to select sentences for extraction

Frequency-based approaches (Brandow, Mitze, and Rau, 1995) may incorporate heuristics to handle readability related issues and knowledge-based approaches systematically perform discourse processing in analyzing and condensing the text, but in a broad classification schema it is the discourse-based approaches that tend to focus on the text macrostructure and surface clues to that structure At the far end of the continuum lies work by Sparck Jones (Jones, 1993, Jones, 1995) in describing a

manual method for source text representation based on linguistic, domain and communicative information From the NLP technology point of view, discourse theory is the least understood among subfields of linguistics

Our work addresses challenges encountered in these previous approaches by applying robust and proven NLP techniques such as corpus-based statistical NLP, robust information extraction, and readily-available on-line NLP resources These techniques and resources allow us to create a richer indexed source of linguistic and domain knowledge than other frequency approaches Our approach attempts to approximate text discourse structure through these multiple layers of information, obtained from automated methods in contrast to labor-intensive, discourse-based approaches Moreover, our planned training methodology will also allow us to exploit this productive infrastructure in ways which model human performance while avoiding hand-crafting domain-dependent rules of the knowledge-based approaches Our ultimate goal is to make our summarization system scalable and portable by learning summarization rules from easily extractable text features

2 System Description

Our summarization system DimSum consists of the Summarization Server and the Summarization Client The Server extracts features (the Feature Extractor) from a document using various robust NLP techniques, described in Section 2 1, and combines these features (the Feature Combiner) to baseline multiple combinations of features, as described in Section 2 2 Our work in progress to automatically train the Feature Combiner based upon user and application needs is presented in Section 2 2 2 The Java-based Client, which will be discussed in Section 4, provides a graphical user interface (GUI) for the end user to customize the summarization preferences and see multiple views of generated summaries

2.1 Extracting Summarization Features

In this section, we describe how we apply robust NLP technology to extract summarization features Our goal is to add more intelligence to frequency-based approaches, to acquire domain knowledge in a more automated fashion, and to approximate text structure by recognizing sources of discourse cohesion and coherence

2.1.1 Going Beyond a Word

Frequency-based summarization systems typically use a single word string as a unit for counting frequencies While such a method is very robust, it ignores the semantic content of words and their potential membership in multi-word phrases For example, it does not distinguish between "bill" in "Bill Clinton" and "bill" in "reform bill" This may introduce noise in frequency counting as the same strings are treated uniformly no matter how the context may have disambiguated the sense or regardless of membership in multi-word phrases For DimSum, we use term frequency based on tf*idf (Salton and McGill, 1983, Brandow, Mitze, and Rau, 1995) to derive *signature words* as one of the summarization features If single words were the sole basis of counting for our summarization application, noise would be introduced both in term frequency and inverse document frequency

Table 1 Collocations with "chips"

{potato tortilla corn chocolate bagle} chips
{computer pentium intel microprocessor memory} chips
{wood oak plastic} chips
bargaining chips
blue chips
mr chips

However, recent advances in statistical NLP and information extraction make it possible to utilize features which go beyond the single word level Our approach is to extract multi-word phrases automatically with high accuracy and use them as the basic unit in the summarization process, including frequency calculation

First, just as word association methods have proven effective in lexical analysis, e g (Church and Hanks, 1990), we are exploring whether frequently occurring collocational information can improve on simple word-based approaches We have preprocessed about 800 MB of LA times/Washington Post newspaper articles using a POS tagger (Brill, 1993) and derived two-word noun collocations using mutual information The result included, for example, various "chips" phrases as shown in Table 1 The word "chips" occurred 1143 times in this corpus, and the table shows that this word is semantically very ambiguous In word associations, it can refer to food, computer components, abstract concepts, etc By incorporating these collocations, we can disambiguate different meanings of "chips "

Secondly, as the recent Message Understanding Conference (MUC-6) showed (Adv, 1995), the accuracy and robustness of name extraction has reached a mature level, equaling the level of human performance in accuracy (mid-90%) and exceeding human speed by many thousands of times We employed SRA's NameTag^{TM} (Krupka, 1995) to tag the aforementioned corpus with names of people, entities, and places, and derived a baseline database for tf*idf calculation In the database, we not only treated multi-word names (e g , "Bill Clinton") as single tokens but also disambiguated the semantic types of names so that, for instance, the company "Ford"

is treated separately from President "Ford" Our approach is thus different from (Kupiec, Pedersen, and Chen, 1995) where only capitalization information was used to identify and group various types of proper names

2.1.2 Acquiring Knowledge of the Domain

In knowledge-based summarization approaches, the biggest challenge is to acquire enough domain knowledge to create conceptual representations for a text Though summarization from conceptual representation has many advantages (as discussed in Section 1), extracting such representations constrains a system to domain dependency and is too knowledge-intensive for our approach

Instead, we took an automatic and robust approach where we acquire *some* domain knowledge from a large corpus and incorporate that knowledge as summarization features in the system We incorporated corpus knowledge in three ways, that is, by using a large corpus baseline to calculate idf values for selecting signature words, by deriving collocations statistically from a large corpus, and by creating a word association index derived from a large corpus (Jing and Croft, 1994) With this method, the system can automatically adapt to each distinct domain, like newspapers vs legal documents without manually developing domain knowledge Domain knowledge is embraced in *signature words*, which indicate key concepts of a given document, in *collocation phrases*, which provide richer key concepts than single-word key concepts (e g "appropriations bill," "omnibus bill," "brady bill," "reconciliation bill," "crime bill," "stopgap bill,", etc), and in their *associated words*, which are clusters of domain related terms (e g , "Bayer" and "aspirin," "Columbia River" and "gorge," "Dead Sea" and "scrolls")

2.1.3 Recognizing Sources of Discourse Cohesion and Coherence

Past research (Paice, 1990) has described the negative impact on abstract quality of failing to perform some type of discourse processing Since discourse knowledge (e g , effective anaphora resolution and text segmentation) cannot currently be automatically acquired easily with high accuracy and robustness, heuristic techniques are often employed in summarization systems to suppress sentences with interdependent cohesive markers

However, there are several shallower but robust methods we can employ now to acquire some discourse knowledge Namely, we exploit the discourse features of lexical items within a text by using name aliases, synonyms, and morphological variants

Within a document, subsequent references to full names are often aliases Thus, linking name aliases provides some indication as to which sentences are interrelated, as shown below

> *The* **Institutional Revolutionary Party**, *or* **PRI**, *capped its landmark assembly to reform itself with a flourish of pomp and promises Among the measures coming out of the assembly's fiercest public debate, in which party members rose up against their leadership Saturday night, are new requirements for future* **PRI** *presidential candidates, qualifications that neither Zedillo nor any of Mexico's previous four presidents would have met*

The NameTag name extraction tool discussed in the previous section performs linking of name aliases within a document such as "Albright" to "Madeleine Albright," "U S " to "United States," and "IBM" for "International Business Machine" We used this tool to link full names and their aliases so that term frequency can be more accurately reflected, i e , "IBM" and "International Business Machine" are counted as two occurrences of the same term

Another overt clue for discourse cohesion and coherence is synonymous words When a theme of an article is developed throughout the text, synonymous words often appear as variants in the text In the example below, for instance, "pictures" and "images" are used interchangeably

> *A new medical* **imaging** *technique may someday be able to detect lung cancer and diseases of the brain earlier than conventional methods, according to doctors at the State University of New York, Stony Brook, and Princeton University If doctors want to take* **pictures** *of the lungs, he noted, they have to use X-ray machines, exposing their patients to doses of radiation in the process The new technique uses an anesthetic, xenon gas, instead of water to create* **images** *of the body*

Although synonym sets have not proven effective in information retrieval for query expansion (Vorhees, 1994), we are using WordNet (Miller et al , 1990) to link synonymous words in an article In the IR task, a query term is expanded with its synonyms without disambiguating the senses of the term Thus, semantically irrelevant query terms are added, and the system typically retrieves more irrelevant documents, decreasing the precision Our summarization approach, in contrast, attempts to exploit WordNet synonym sets of only signature terms in a *single* document Our hypothesis is that if a synonym of a signature term exists in the article, the term has been disambiguated by the context of the article and the "correct" synonym, not a synonym of the term in a different sense, is likely to co-occur in the same document

In addition, morphological analysis allows us to link morphological variants of the same word within a document Morphological variants are often used to refer to the same concept throughout a document,

providing discourse clues In the above example, "imaging" and "images" are morphologically linked

Like synonyms, morphology or stemming has not proven to be useful for improving information retrieval (Salton and Lesk, 1968, Harman, 1991) However, the recent work by (Church, 1995) showed that effectiveness of morphology, or correlations among morphological variants within a document, varies from word to word A word like "hostage" has a large correlation with its variant "hostages" while a word like "await" does not According to his experiments, good keywords like "hostage" and its variants are likely to be repeated more than chance within a document and highly correlate with variant forms This implies that important signature words we use for summarization are likely to appear in a single document multiple times using their variant forms

2.2 Combining Summarization Features

The DimSum summarizer exploits our flexible definition of a signature word and sources of domain and discourse knowledge in the texts through

- the creation of multiple baseline databases corresponding to multiple definitions of signature words
- the application of the discourse features in multiple-term frequency calculation methods

Different baseline databases can affect the inverse document frequency (idf) values We have created multiple baseline databases based upon multiple definitions of the signature words Signature words are flexibly defined as collections of features Presently, we derive databases consisting of (a) terms alone, (b) terms plus multi-word names, (c) stemmed terms plus muti-words names, and (d) terms plus multi-word names and collocations The discourse features, i e , synonyms, morphological variants or name aliases, for signature words, on the other hand, can affect the term frequency (tf) values Using these discourse features boosts the term frequency score within a text when they are treated as variants of signature words Having multiple baseline databases available makes it easy to test the contribution of each feature or combination of features

2.2.1 The Feature Combiner: Current

In order to select sentences for a summary, each sentence in the document is scored using different combinations of signature word features and discourse features Currently, every token in a document is assigned a score based on its tf*idf value The token score is used, in turn, to calculate the score of each sentence in the document More specifically, the score of a sentence is calculated as the average of the scores of the tokens contained in that sentence with the exception that certain types of tokens can be eliminated from the sentence as discussed That is, the DimSum system can ignore any combination of name types (i e , person, place, entity) from a given document for scoring (cf Section 3 for more details)

After every sentence is assigned a score, the top n highest scoring sentences are chosen as a summary of the content of the document Currently, the DimSum system chooses the number of sentences equal to a power k (between zero and one) of the total number of sentences Thus, the system can vary the length of a summary according to k For instance, if 0 5 is chosen as the power, and the document consists of 100 sentences, the output summary would contain 10 sentences This scheme has an advantage over choosing a given percentage of document size as it yields more information for longer documents while keeping summary size manageable We use the results of this method as the baseline summary performance (i e , without any training), and report them in Section 3

2.2.2 The Feature Combiner. Future

As our goal is to make our summarization system trainable to different user and application needs, we are currently working on learning the best feature combination method from a training corpus automatically For training and evaluating our summarization system, we had a user create extract summaries by selecting relevant sentences in articles In order to compare with the results of a trainable summarizer reported by (Kupiec, Pedersen, and Chen, 1995), we first use Bayes' rules to learn the best scoring method Then, we will use an inductive learning algorithm such as the decision tree algorithm (Quinlan, 1993) to learn summarization rules which can deal with feature dependencies across sentences

3 Evaluation

Much research has been devoted to assessing correspondence between human and machine abstracts because of the complexity of analyzing "aboutness" as illustrated in (Hahn, 1990) As a result, most of the preliminary evaluations of summarization systems have been developer-based A common approach is to compare correspondence between automatic performance and human performance (Rath, Resnick, and Savage, 1961, Edmundson, 1969, Kupiec, Pedersen, and Chen, 1995) or summary acceptability (Brandow, Mitze, and Rau, 1995) Others have been task-based, comparing abstract and full text originals in terms of the browsing and search time (Miike et al , 1994, Sumita, Ono, and Miike, 1993) or recall and precision in document retrieval (Brandow, Mitze, and Rau, 1995)

Our evaluation methodology is two-pronged First, we evaluate the system by scoring for correspondence with human generated extracts (Sec-

tion 3 1) Second, in our future work we are collaborating with the University of Massachusetts to evaluate retrieval effectiveness for system-generated and human-generated summaries (Section 3 2)

3.1 Developer-based Evaluation

The DimSum development environment software incorporates automatic scoring software to calculate system recall and precision for any user's training or test data This allows us to evaluate system performance for any user and for variations in summary preferences

We performed an informal experiment in which 6 users created summary extract versions of the same set of 15 texts These versions varied considerably among users, which supports our view that a summarization system should be trained for user preference Then, we ran the DimSum system over these 15 texts using multiple feature combinations (i e , combinations among names, synonyms, and morphological variants), and scored against the six versions of summary extracts Though correspondence between the DimSum summaries and user summaries was low (ranging between 14% and 31% F-measures), clearly some feature sets were more effective for some users than for others For example, the best feature combination for the best-case correspondence between the user and DimSum (i e , 31% case) was the combination of name, synonym and morphological information On the other hand, the best combination for the worst-case correspondence between the user and DimSum (i e , 14% case) was the combination of name and synonym information Some summary extracts, however, were not affected by different combinations of features

The second step was to obtain a "bottom-line" score for a single user We ran the DimSum system over a set of 86 texts using multiple feature combinations The features were combined by taking an average of tf*idf, tf or idf scores of each token in a sentence No training was performed We varied the length of summaries (by changing k from 0 5 to 1 0), use of different types of names (i e , person, place, and entity), use of aliases, and use of synonyms for different parts of speech (i e , adjective, adverb, noun, and verb)

Table 2 shows the top three F-measure scores (1–3), and the score for using the simplest baseline (4) For the best summary (1), place names were used while person and entity names were recognized but removed for sentence scoring Synonyms were also used The k value was set to 0 65 (about 20-25% of a document as a summary) Use of aliases and synonyms didn't make much difference in the scores (2–3) However, they all scored slightly higher than the summary which didn't use any of these features, i e , a summary which didn't use names, synonyms, or aliases (4)

It is interesting that using name tagging in a reverse way, i e , recognizing and then deleting person names from sentence scoring, made a significantly positive effect on summarization The best summary score with the person name used in sentence scoring was 38 6% (5) The reason why person names made negative contributions to the summary seems to be because personal names were often mentioned as passing references (e g , names of spokespeople) in the corpus, but they had high idf values

Finally, in every feature combination, taking tf*idf scores of each word outperformed the idf-based calculation, and the latter in turn outperformed the tf-based score calculation

These results further motivate us to apply automated learning to combine summarization features The fact that humans vary in summarization suggests that recall/precision evaluation is not meaningful unless a summarization system is trainable to a particular summary style Our current work is to identify through training what feature combinations produce an optimal summary for a given user We anticipate that the summary performance will improve with training as DimSum learns automatically how or whether these different signature word definitions are contributing to the summary The current design does not incorporate paragraph/sentence location information or genre-specific indicator phrases We are exploring if these features can be indirectly subsumed by the derived features we have already identified

Also, the cursory look at the summaries of DimSum shows that the system-generated summary may be providing the same information as the summary provided by the user, but the sentences were chosen differently This happens because the same information can be conveyed by different sentences within the same document This motivates us to conduct a more task-oriented summarization evaluation, which is discussed below

3.2 Task-based Evaluation

As a more task-oriented evaluation, we are collaborating with the University of Massachusetts to evaluate retrieval effectiveness for DimSum system-generated and human-generated extracts for topics from TREC-5 (Text REtrieval Conference-5) We have selected 30 topics, five assessed as difficult, five assessed as easy (Harman, 1996), and the remaining 15 randomly The top 50 documents judged relevant by the INQUERY system in TREC-5 for each topic have been identified For each document, two extract versions are being manually created One extract is based on the topic description, while the second is generated independent of the topic description In addition, the DimSum system will automatically generate two versions (query dependent and generic) for each of the texts With the TREC-5 full text results as a baseline, multiple iterations of the INQUERY system will test retrieval performance on

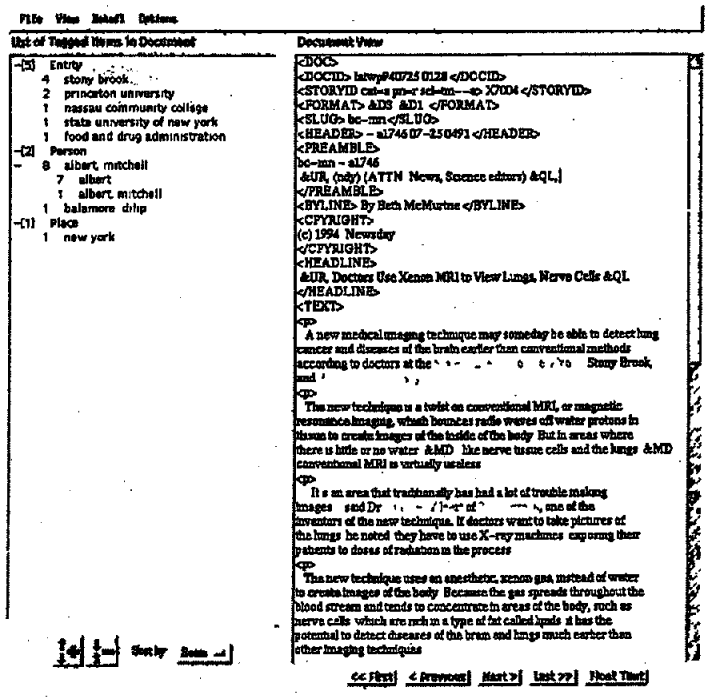

Figure 1 Name Mode Summary

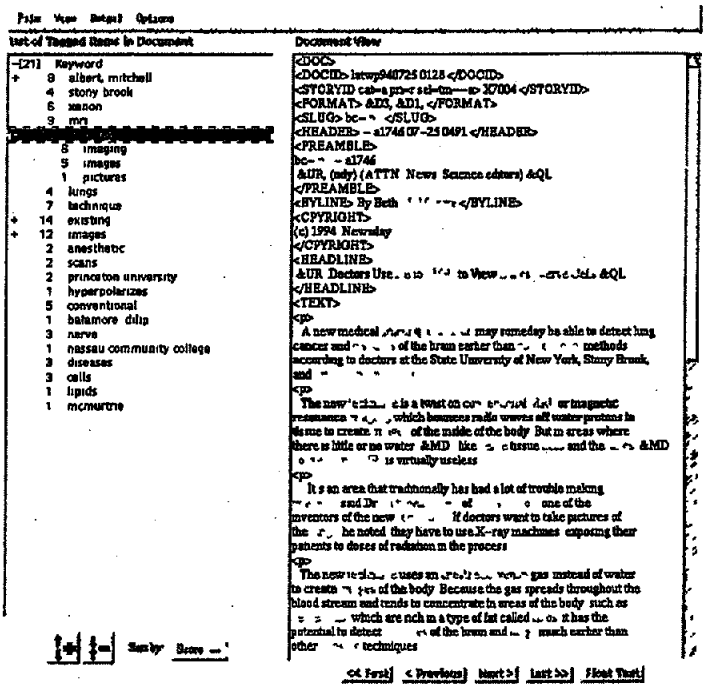

Figure 2 Keyword Mode Summary

Table 2 Summary scores for different feature combinations

Feature Combination	tf*idf	idf	tf	
term+place+synonym	41 5	32 3	20 9	(1)
term+place+entity	41 2	33 9	21 2	(2)
term+place+alias+synonym	40 9	32 4	21 0	(3)
term	39 9	32 4	21 0	(4)
term+person+place+entity+alias+synonym	38 6	32 1	22 5	(5)

the human and machine generated extracts to compare retrieval effectiveness

4 Multi-dimensional Summary Views

The DimSum Summarization Client provides a summary of a document in multiple dimensions through a graphical user interface (GUI) to suit different users' needs In contrast to a static view of a document, the system brings the contributing linguistic and other resources to the desktop and the user chooses the view he wants As shown in Figure 1, the GUI is divided into the List Box on the left and the Text Viewer on the right

When a user asks for a summary of a text, extracted summary sentences are highlighted in the Text Viewer The user can dynamically control a percentage of sentences to highlight for a summary In addition, the Client can automatically color-code top keywords in different colors for different types (i e , person, entity, place and other) for quick and easy browsing

In the List Box, the user can explore two different summary views of a text First, the user can choose the "Name Mode," and all the names of people, entities, and places which were recognized by the name extraction tool are sorted and displayed in the List Box (cf Figure 1) The user can also select a subset of name types (e g , only person and entity, but not place) to display Aliases of a name are indented and listed under their full names

In the "Keyword Mode," the top keywords, or signature words, (including names) are displayed in the List Box Analogous to the name aliases, for each keyword its synonyms and morphological variants, if exist, are indented and listed below it (cf Figure 2) The user can choose the score threshold or percentage to vary the number of keywords for display

In both modes, the names and signature words in the List Box can be sorted alphabetically, by frequency, or by the tf*idf score Clicking on a term in the List Box also causes the first occurrence of the term to be highlighted in the Text Viewer From there, the user can use the FIRST, PREVIOUS, NEXT, or LAST button at the bottom of the GUI to track the other occurrences of the term, including its aliases, synonyms, and morphological variants This provides the user with a way to track themes of the text interactively

5 Summary

The DimSum summarization system leverages off of the works of (Kupiec, Pedersen, and Chen, 1995) and (Brandow, Mitze, and Rau, 1995), and advances summarization technology by applying corpus-based statistical NLP techniques, robust information extraction, and readily available on-line resources Our preliminary experiments with combining different summarization features have been reported, and our current effort to learn to combine these features to produce the best summaries has been described The features derived by these robust NLP techniques were also utilized in presenting multiple summary views to the user in a novel way

References

Advanced Research Projects Agency 1995 *Proceedings of Sixth Message Understanding Conference (MUC-6)* Morgan Kaufmann Publishers

Brandow, Ron, Karl Mitze, and Lisa Rau 1995 Automatic condensation of electronic publications by sentence selection *Information Processing and Management*, 31, forthcoming

Brill, Eric 1993 *A Corups-based Approach to Language Learning* Ph D thesis, University of Pennsylvania

Church, Kenneth and Patrick Hanks 1990 Word Association Norms, Mutual Information, and Lexicography *Computational Linguistics*, 16(1)

Church, Kenneth W 1995 One term or two? In *Proceedings of the 17th Annual International SIGIR Conference on Research and Development in Information Retrieval*, pages 310–318

Edmundson, H P 1969 New methods in automatic abstracting *Journal of the ACM*, 16(2) 264–228

Fum, Danilo, Giovanni Guida, and Carlo Tasso 1985 Evalutating importance A step towards text summarization In *IJCAI85*, pages 840–844 IJCAI, AAAI

Hahn, Udo 1990 Topic parsing Accounting for text macro structures in full-text analysis *Infor-*

mation Processing and Management, 26(1) 135–170

Harman, Donna 1991 How effective is suffixing? *Journal of the American Society for Information Science*, 42(1) 7–15

Harman, Donna 1996 Overview of the fifth text retrieval conference (trec-5) In *TREC-5 Conference Proceedings*

Jing, Y and B Croft 1994 *An Association Thesaurus for Information Retrieval* Umass Technical Report 94-17 Center for Intelligent Information Retrieval, University of Massachusetts

Johnson, F C , C D Paice, W J Black, and A P Neal 1993 The application of linguistic processing to automatic abstract generation *Journal of Documentation and Text Management*, 1(3) 215–241

Jones, Karen Sparck 1993 What might be in a summary? In Knorz, Krause, and Womser-Hacker, editors, *Information Retrieval '93*, pages 9–26

Jones, Karen Sparck 1995 Discourse modeling for automatic summaries In E Hajicova, M Cervenka, O Leska, and P Sgall, editors, *Prague Linguistic Circle Papers*, volume 1, pages 201–227

Krupka, George 1995 SRA Description of the SRA System as Used for MUC-6 In *Proceedings of Sixth Message Understanding Conference (MUC-6)*

Kupiec, Julian, Jan Pedersen, and Francine Chen 1995 A trainable document summarizer In *Proceedings of the 18th Annual International SIGIR Conference on Research and Development in Information Retrieval*, pages 68–73

Luhn, H P 1958 The automatic creation of literature abstracts In *IBM J Research Development*, volume 2, pages 159–165

Maybury, Mark T 1995 Automated even summarization techniques In B Endres-Niggemeyer, J Hobbs, and Karen Sparck Jones, editors, *Summarizing Text for Intelligent Communication*, pages 101–149

McKeown, Kathleen and Dragomir Radev 1995 Generating summaries of multiple news articles In *Proceedings of the 18th Annual International SIGIR Conference on Research and Development in Information*, pages 74–78

Miike, Seiji, Etsuo Itho, Kenji Ono, and Kazuo Sumita 1994 A full text retrieval system with a dynamic abstract generation function In *Proceedings of 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 152–161

Miller, George, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller 1990 Five papers on WordNet Technical Report CSL Report 43, Cognitive Science Laboratory, Princeton University

Paice, C 1990 Constructing literature abstracts by computer Techniques and prospects *Information Processing and Management*, 26(1) 171–186

Paice, C and P A Jones 1993 The identification of important concepts in highly structured technical papers In *Proceedings of the 16th Annual International ACM SIGIR Conference of Research and Development in Information Retrieval*, pages 69–78

Quinlan, J Ross 1993 *C4 5 Programs for Machine Learning* Morgan Kaufmann Publishers

Rath, G J , A Resnick, and T R Savage 1961 The formation of abstracts by the selection of sentences *American Documentation*, 12(2) 139–143

Rau, Lisa F , Paul S Jacobs, and Uri Zernik 1989 Information extraction and text summarization using linguistic knowledge acquisition *Information Processing and Management*, 25(4) 419–428

Reimer, Ulrich and Udo Hahn 1988 Text condensation as knowledge base abstraction In *IEEE Conference on AI Applications*, pages 338–344

Riloff, Ellen 1995 A corpus-based approach to domain-specific text summarization In B Endres-Niggemeyer, J Hobbs, and K Sparck Jones, editors, *Summarizing Text for Intelligent Communication*, pages 69–84

Rush, J E , R Salvador, and A Zamora 1971 Automatic abstracting and indexing Production of indicative abstracts by application of contextual inference and syntactic criteria *Journal of the American Society for Information Science*, 22(4) 260–274

Salton, G and M McGill, editors 1983 *An Introduction to Modern Information Retrieval* McGraw-Hill

Salton, Gerald and Mark Lesk 1968 Computer evaluation of indexing and text processing *Journal of the ACM*, 15(1) 8–36

Sumita, Kazuo, Kenji Ono, and Seiji Miike 1993 Document structure extraction for interactive document retrieval systems In *Proceedings of SIGDOC'93*, pages 1301–310

Vorhees, E 1994 Query expansion using lexical-semantic relations In *Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and Development of Information Retrieval*, pages 61–69