

Feasibility of Using Citations as Document Summaries

A Thesis

Submitted to the Faculty

of

Drexel University

by

Jeff Hand

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

December, 2003

© Copyright 2003
Jeff Hand. All Rights Reserved.

Dedications

This dissertation is dedicated with love to my wife, Dr. Mira Lalovic, without whose support and gentle persuasion this research would not have happened.

Acknowledgments

I would like to express my sincere thanks to the members of my dissertation committee:

Drs. Carl Drott, Jim Bass, Xia Lin, Henry Small and Chaomei Chen. In particular, I would like to express my gratitude to Carl Drott for his support and guidance, and uncanny ability to be a “tough” or “soft” advisor at just the right time.

Table of Contents

| | |
|---------------------------------------------------------------------------|------------|
| List of Tables | v |
| List of Figures..... | vii |
| Abstract..... | x |
| 1. Introduction..... | 1 |
| 1.1 Aim of Research..... | 1 |
| 1.2 Context of the Research | 2 |
| 1.3 Significance of the Research | 7 |
| 1.4 Specific Goals of this Research..... | 9 |
| 1.5 Definition of Terms | 10 |
| 2. Review of the Literature..... | 20 |
| 2.1 Introduction | 20 |
| 2.2 Advances in Computer Science Text Summarization..... | 23 |
| 2.2.1 Classic Techniques | 23 |
| 2.2.2 Extraction..... | 24 |
| 2.2.3 Natural Language..... | 28 |
| 2.2.4 Cue Phrase and Location | 29 |
| 2.2.5 Entity Summarization | 31 |
| 2.2.6 Discourse Summarization..... | 34 |
| 2.2.7 Information Retrieval — Hypertext..... | 39 |
| 2.2.8 Information Extraction..... | 40 |
| 2.2.9 Information Retrieval..... | 43 |
| 2.2.10 Multiple Documents | 45 |
| 2.2.11 Evaluation Methods | 46 |
| 2.2.12 Summary..... | 49 |
| 2.3 Advances in Citation Analysis in Library and Information Science..... | 50 |
| 2.3.1 Citation Analysis | 52 |
| 2.3.2 Citation Motivation..... | 55 |
| 2.3.3 Citation Context..... | 56 |
| 2.4 Automating Citation Analysis for Summarization..... | 64 |
| 2.5 Conclusion..... | 71 |
| 3. Research Methods..... | 73 |
| 3.1 Review of the Problem | 73 |
| 3.1.1 Rater Data | 74 |
| 3.1.2 Document Data | 74 |
| 3.1.3 Data Analysis..... | 75 |
| 3.2 Subject Survey Data | 75 |
| 3.2.1 Document and Citation Sample | 75 |

| | |
|---------------------------------------------------------------------------------------------------------|------------|
| 3.2.2 Raters | 78 |
| 3.2.3 Survey Instrument..... | 79 |
| 3.3 Document Data..... | 80 |
| 3.3.1 Similarity Metric | 80 |
| 3.3.2 Citation Categories | 84 |
| 3.3.3 Citation Location | 86 |
| 3.4 Variables and Research Questions | 86 |
| 3.4.1 Research Questions..... | 88 |
| 4. Results | 90 |
| 4.1 Introduction | 90 |
| 4.2 Descriptive Statistics | 91 |
| 4.3 Comparisons of Expert and Non-expert Means | 95 |
| 4.4 Document Metrics Descriptive Statistics | 100 |
| 4.4.1 Citation Location | 101 |
| 4.4.2 Citation Category | 102 |
| 4.5 Partial Coefficient of Determination Analyses | 104 |
| 4.6 Regression Analyses | 110 |
| 4.7 Conclusion..... | 116 |
| 5. Discussion..... | 118 |
| 5.1 Summary.. | 118 |
| 5.2 Discussion | 119 |
| 5.3 Conclusion..... | 127 |
| List of References..... | 129 |
| Appendix A. Survey Instrument..... | 148 |
| Appendix B. Document Abstracts | 154 |
| Appendix C. Citation Context Extracts | 168 |
| Appendix D. Citation Contexts Per Document | 204 |
| Appendix E. Citation Categorization Data for Raters 1 and 2..... | 205 |
| Appendix F. Data Collection, Storage, and Calculations | 216 |
| Appendix G. Ranking of Citation Context for Expert, Non-expert, and the Combined Groups..... | 219 |
| Appendix H. Stepwise Regression Tables..... | 227 |
| Curriculum Vitae..... | 230 |

List of Tables

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| 1-1. Small's Categories of Citations | 13 |
| 3-1. Small's Modified Categories of Citations | 85 |
| 3-2. Variables in the Study | 87 |
| 4-1. Means, Median, Mode, and Standard Deviation of Expert Ratings | 92 |
| 4-2. Means, Median, Mode, and Standard Deviation of Non-expert Ratings | 93 |
| 4-3. Means, Median, Mode, and Standard Deviation of All Participants' Ratings ... | 94 |
| 4-4. T-Tests Between Non-experts and Experts on Seven Variables..... | 95 |
| 4-5. Seven Rank Ordering of Scores | 100 |
| 4-6. Summary of Citation Rank Order for Experts, Non-experts, and Combined Group..... | 99 |
| 4-7. Correlation between Non-experts and Experts on Seven Variables | 100 |
| 4-8. Citations Location Table..... | 102 |
| 4-9. Category Agreement Between Raters A and B..... | 102 |
| 4-10. Identifying Phrase Agreement Between Raters A and B..... | 103 |
| 4-11. Citation Categories | 103 |
| 4-12. Data Example..... | 103 |
| 4-13. Partial Coefficient of Determinants (Partial R² Values) Among Frequency Counts/Cosine Scores and Expert Variables | 107 |
| 4-14. Partial Coefficient of Determinants Among Cosine/Frequency Count Scores and Non-expert Variables..... | 108 |
| 4-15. Partial Correlations Among Frequency Counts/ Cosine Similarity Metrics Combined Variables of Experts and Non-Experts..... | 109 |
| 4-16. Partial Correlations Among Cosine /Binary Scores and Nominal Variables..... | 110 |

| | |
|---------------------------------------------------------------------------------------------------------------------|------------|
| 4-17. Regression Analysis for Predicting Cosine Unweighted Using the Expert Ratings | 110 |
| 4-18. Variance Explained R^2 by Variables..... | 112 |
| 4-19. Regression Analysis for Predicting Cosine Unweighted Using the Non-expert Ratings | 113 |
| 4-20. Regression Analysis for Predicting Cosine Unweighted Combining the Expert and Non-expert Ratings | 113 |
| 4-21. Regression Analysis for Frequency Counts Using Expert Ratings..... | 114 |
| 4-22. Regression Analysis for Frequency Counts Using Non-expert Ratings | 115 |
| 4-23. Regression Analysis for Frequency Counts Combining the Expert and Non-expert Ratings | 116 |

List of Figures

| | |
|-----------------------------------------------------------------------------------------------------------|------------|
| 1-1. Co-Citation Map of Leukemia Viruses, 1978–1981; Step 1 of Specification Narrative | 4 |
| 1-2. Narrative Sequence — Spanning Tree and Search Method; Step 2 of Specification Narrative | 4 |
| 3-1. Distribution of Citations per Abstract..... | 77 |
| 4 1. Citation Distribution per Abstract..... | 101 |
| F-1. Database Diagram..... | 218 |

Abstract

Feasibility of Using Citations as Document Summaries

Jeff Hand

Carl Drott, Ph.D.

The purpose of this research is to establish whether it is feasible to use citations as document summaries. People are good at creating and selecting summaries and are generally the standard for evaluating computer generated summaries. Citations can be characterized as concept symbols or short summaries of the document they are citing. Similarity metrics have been used in retrieval and text summarization to determine how alike two documents are. Similarity metrics have never been compared to what human subjects think are similar between two documents. If similarity metrics reflect human judgment, then we can mechanize the selection of citations that act as short summaries of the document they are citing.

The research approach was to gather rater data comparing document abstracts to citations about the same document and then to statistically compare those results to several document metrics; frequency count, similarity metric, citation location and type of citation. There were two groups of raters, subject experts and non-experts. Both groups of raters were asked to evaluate seven parameters between abstract and citations: purpose, subject matter, methods, conclusions, findings, implications, readability, and understandability. The rater was to identify how strongly the citation represented the content of the abstract, on a five point likert scale. Document metrics were collected for

frequency count, cosine, and similarity metric between abstracts and associated citations. In addition, data was collected on the location of the citations and the type of citation. Location was identified and dummy coded for introduction, method, discussion, review of the literature and conclusion. Citations were categorized and dummy coded for whether they refuted, noted, supported, reviewed, or applied information about the cited document. The results show there is a relationship between some similarity metrics and human judgment of similarity.

1. Introduction

1.1 Aim of Research

After a four hour fire and brimstone Sunday sermon, by a well known preacher of the day. Calvin Coolidge was asked by a reporter, “What did the preacher talk about for such a long period of time?” Mr. Coolidge replied, “Sin.” Having to fill up an entire column the reporter pressed for more information by asking, “What did the preacher say about sin?” Mr. Coolidge replied, “He was against it.”

People are good summarizers because they are capable of comparing, generalizing, and distinguishing the content and context among pieces of text. These abilities allow them to select a citation to serve as the summary of a source document. In contrast, computers do not summarize documents well, but they are far superior to people at counting, sorting, and matching. The study aims to determine the feasibility of using word metrics to select the citation context, to be used as a document summary. Citation contexts are the citation text and the text before and after the citation. The citation context represents a summarization of the cited document by an author. This study examines whether there is a relationship between citation context selection by human subjects and similarity word metrics.

The aim of this dissertation is to evaluate the feasibility of using word metrics as a tool to select the best citation context that can be used as a document summary. It examines the relationship between citation context selection by human subjects and similarity word metrics, and the possibility of using machines for text summarization.

Document metrics will be collected for abstracts and their citation contexts for word frequency, un-weighted cosine, and Salton's similarity metrics. In addition citation contexts will be evaluated for citation location, and category of citation. Subject responses of similarity between citation context and abstracts will be based on parameters, identified in the literature (Edmundson, 1969). Data will be collected on the following seven parameters: subject matter, purpose, methods, conclusion or findings, implications, understandability, and readability.

The two sets of data — document metrics and subject responses — will be compared to determine whether any relationship between them exists, and if so, the strength of the relationship and how it may be applied to a system for automated citation context selection. If such a relationship between document metrics and subject responses can be shown, the task of automated citation context selection can be executed with a degree of certainty and known error.

1.2 Context of the Research

The impetus for this research was Henry Small's 1986 paper *The Synthesis of Specialty Narrative from Co-citation Clusters*. The specialty narrative is a technique to generate a text summary of multiple scientific papers from the Institute for Scientific Information (ISI) database. The specialty narrative process has four steps. For the purpose of this research, step three, citing passage analysis and consensus passage selection, is the primary step. A brief explanation of the entire process will show the research in the context of the entire method.

Step one is creating a co-citation map of a scientific discipline. Figure 1-1 is an example of a co-citation map. This particular map was generated by counting the number

of co-citations between published papers for leukemia virus research from 1978 to 1981 (Small, 1986) and then by graphing the co-citation relationships. Figure 1-1 reflects the key documents and shows their co-relationship pattern.

Step two; create the spanning tree and search method. This is a way to “linearize” the graph established in step one. Figure 1-2 gives a sequence to the ideas presented in the co-citation map. This is equivalent to a sequential outline of a document.

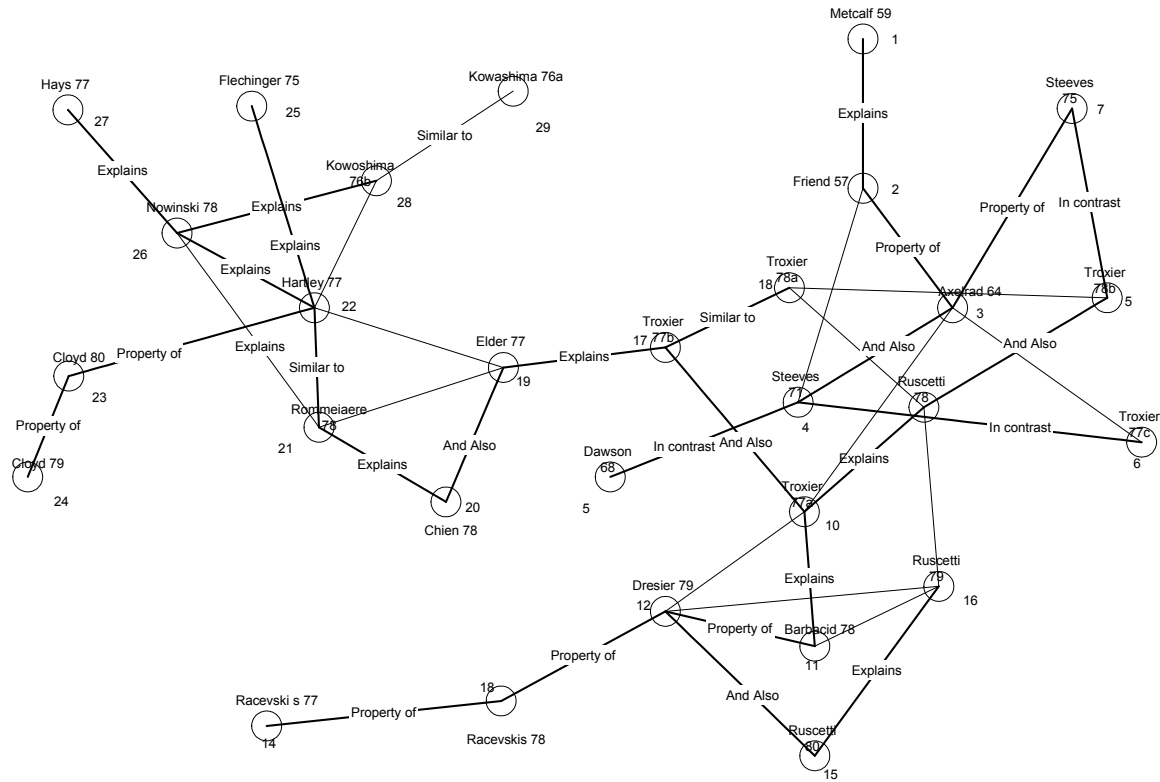


Figure 1-1. Co-Citation Map of Leukemia Viruses, 1978–1981; Step 1 of Specification Narrative

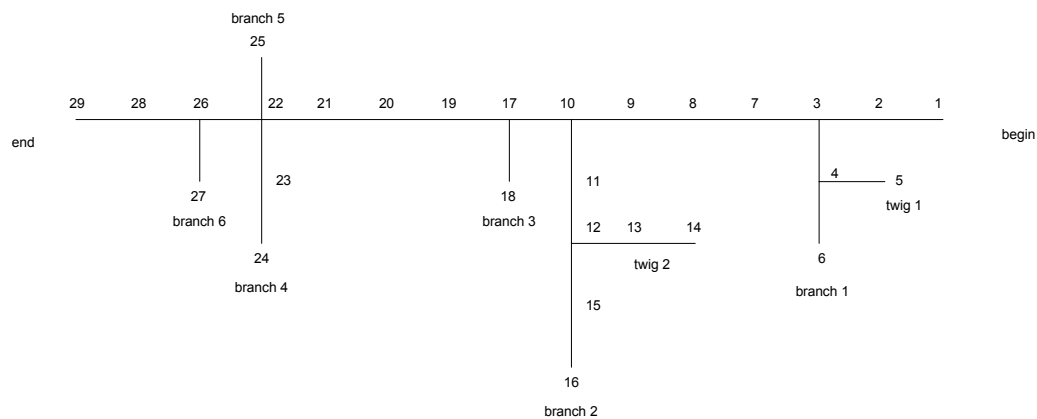


Figure 1-2. Narrative Sequence — Spanning Tree and Search Method; Step 2 of Specification Narrative

Step three is citing passage analysis and consensus passage selection. As stated by Small (1986), a consensus passage is an “... expression of ideas in words, which entails a selection of wording which is representative of existing formulations by others.” In this step, we find the citation contexts written by authors about a document. These citation contexts serve as surrogates for the documents identified by the co-citation analysis from step one. Small accomplished this step in the original research manually. The focus of the present research is to find a way to automate the selection of a citation context that can stand as a document summary.

Step four, the creation of transitional sentences. This step identifies the relationship between documents and subsequent labeling with associative and transitional phrases. Figure 1-2 shows transition phrases between nodes (documents).

In addition to Small’s paper on the specialty narrative, several other research papers are fundamental to a solid grounding in the research presented here. The present study is organized around two academic research domains, information science and computer science, with the same aim: text summarization. Chapter 2, the Review of the Literature, expands on the relevance and contributions to the present study from these academic domains, which are briefly described below.

Three influential computer science papers that provide a foundation for this research are by Luhn, Edmundson, and Salton and McGill. Luhn’s (1958) research describes a statistical approach to text summarization, based on word frequency counts to score sentences; those with the highest scores are extracted to create the abstract. Edmundson’s (1969) paper extends Luhn’s work, user questionnaires were compared to word frequency counts, word locations, and cue phrases to determine which sentences are

to be extracted for an abstract. Edmundson's work provides the conceptual framework for the present research, which substituted citation contexts for sentences and added the additional parameter of the similarity metric. Salton and McGill's (1983) similarity metric measures the similarity of two documents, based on word frequency counts and weighting factors of the common words between documents. The similarity metric is integral to this study and is used as a comparison against human subjects' selection of abstracts and citation contexts that are written about the same material.

Several important information science papers are fundamental to this study as well: citations as concept symbols (Small, 1978), citation classification (Small, 1982), the ISI Atlas of Science project, and automated text summarization (Nanba, Kando, and Okumura, 2000). Small (1978) explored the idea of using citation context as concept symbols to create a compressed representation of the full text of a cited document. In the same time frame, the Institute for Scientific Information developed the Atlas of Science, an endeavor to map the literature of a scientific discipline and generate reviews of the discipline's literature. The technique was limited by the inability to automate the process and its dependence on human writers to interpret the complex relationships in the maps. In the 1980s, Small (1982) started to examine citation text analysis, which we will refer to here as citation classification. Building on others' work, he devised a simplified, rational classification system for citation context. Nanba, Kando, and Okumura (2000) developed a system called PRESRI that combines the software technology of NEC's ResearchIndex with citation categorization and citations as concept symbols in an attempt to automate a review of the literature in a specific subject domain. The work is promising, and although the system incorporates the idea of citations as a concept

symbol, it could be considerably improved with the incorporation of co-citation analysis and Small's work on citation analysis.

The present research builds on the research cited above to develop an automated citation context selection system that mediates the speedy, consistent technique of word metrics with the quality of human selection. The possibility of using similarity metrics to select citation context as a document summary will be examined in relation to variables including citation location, citation category, and expertise of the human subjects who evaluate the similarities between the citation context and document abstracts.

1.3 Significance of the Research

Given the proliferation of new and existing sources of information in electronic form, there is no question that the scientific community is experiencing information overload. Information overload is a problem scholars have been wrestling with since the Middle Ages, and the pace of overload has been increasing exponentially ever since. One reaction to overload in the Middle Ages was the creation of text summaries by hand; today these efforts are known as encyclopedias and indexes (Hutchins, 1993). Text summaries written by people are of high quality but also expensive. Our current research communities struggle to effectively summarize seemingly inexhaustible amounts of information.

Today so much information exists, from such a variety of sources, that researchers cannot be sure whether they have adequately accessed the available information appropriate for their research needs. No single Internet search engine or meta-search engine can adequately access all the available Web-based information. Individuals rarely have the time to process all the information they find. This overwhelming amount of

information requires a method, some sort of summarization/extraction tool. Information “predigested” to a manageable size would be extremely useful to individual users.

Although computer systems can handle large amounts of data relatively inexpensively, the text summaries generated by computers are of questionable quality. Most automated text extraction and summarization systems perform poorly in the areas of understandability and readability. Sentence extraction is indicative only of content, and may or may not present a summary that is coherent. Current techniques that aim to create summaries based on understanding sentences or documents are limited but evolving.

One benefit of an effective text summarization tool is rooted in the findings of cognitive science: People understand concepts more quickly and retain the concepts for longer periods of time when the information is presented as a condensation (Morris, 1999). Another benefit of an effective text summarization tool is the saving of time in information retrieval and analysis. Traditional human information searching and retrieval techniques are time-and labor-intensive, and results are entirely contingent upon the researcher's familiarity and skill with both the information system being used and the unique language of the discipline being researched. Therefore, researchers can benefit from any techniques that reduce the time and cognitive effort devoted to information retrieval and analysis.

Yet another benefit is that the costs of devoting extensive time to uncovering documents could be reduced, which would increase researcher productivity. For example, using this technique, a novice could become acquainted with the literature of a discipline in less time. More time could be spent analyzing the conceptual framework of a discipline instead of collecting information.

Some prototype computer systems, such as CORA and ResearchIndex by NEC, have automated the extraction and linking of citations. This is a precursor to a fully automated specialty narrative. The significance of this research is its contribution to the automated construction of specialty narratives, specifically step three in Small's construction process, the citation selection acting as a document summary.

1.4 Specific Goals of this Research

Previous studies (Small, 1986; Nanba, Kando, and Okumura, 2000) have demonstrated that citations can be used as document summaries. The question for this study is whether it is feasible to use computer-calculated similarity metrics as predictors of human selection of citations that can serve as document summaries. The research will compare quantitative measures of human similarity judgments with word cosine similarity metrics calculated between an abstract and its citation contexts. The similarity judgments of subjects will be measured based on a questionnaire developed by Edmundson (1969). The computer measures will be based on word frequency and cosine similarity metrics (Salton and McGill, 1983), type of citation, and location of citation.

Specifically, this study will ask the following:

- Is there any relationship between human-selected citations and similarity measures, word frequency counts, cosine similarity metrics, and weighted similarity metrics?
- Is there a difference between experts and non-experts in selecting the similarity between an abstract and its citations?
- Does the location of the citation context within the document have an influence on the citation's selection as a document summary?

- Does the type or category of a citation have any influence on its selection as a document summary?

1.5 Definition of Terms

Following is a list of terms used in this dissertation. Other researchers might use these terms differently; this section establishes how these terms are to be understood in the context of this dissertation.

Abstract: “An abstract may be defined as an abbreviated, accurate representation of a document. An abstract may be said to consist of one or more portions of a document selected to represent the whole.” (Well, V.H. “Standards for Writing Abstracts,” JASIS, 4:22, 1970, 351-7).

An abstract might be an *indicative* abstract or an *informative* abstract and may be written by the author or by an expert. For the purposes of this research, an abstract is whatever is published as an abstract.

Atlas of Science: Published by ISI, this was an attempt to map the literature of science. The indexing terms assigned to documents in ISI’s SciSearch database (Dialog File 4,434) were used to create “research fronts” (see “Research front”). The *Atlas of Science* used these research fronts to track, map, and describe temporal changes and developments in the particular research areas represented. Hence the title. The *Atlas of Science* is no longer published. The research fronts still exist, but updates ceased in 1996 and accessibility via the Dialog SciSearch database is limited.

Automatic Text Summarization: Automatic text summarization “has been taken to include extraction and abstracting. The greater portion of techniques developed has been within either text extraction or fact extraction” (Jones, 1999). Automatic text summarization has three phases (Mani and Maybury 1999):

- Analyzing the input text
- Transforming it into a summary representation
- Synthesizing an appropriate output form.

The text that is *input* into the text summarization system could be a single document or multiple documents. The *transformation* phase involves text selection. *Synthesis* involves producing the final summary via text compaction, smoothing the text, and adding fluency and coherence (i.e., adding connecting phrases between extracted segments of text).

Citation (noun): “Look from the *cited* document to the *citing* document.” (Small, 1978). Hence, look from the inside to the outside: A small set of core cited documents is utilized by a larger world of documents that do the citing. *Citation* (noun) will be understood in this dissertation to mean “cited reference.” *Citation* will also sometimes be used to mean “cited passage,” or the phrase (quotation) that is drawn from the cited document.

Citation analysis: Citation analysis is used in different ways. One aspect of citation analysis centers on the study and ultimate classification of “cited references” (see “Citation”). Classification schemes usually center on (1) the function or role of the citation, or (2) the author’s motivations for citing.

Another type of citation analysis is the straight *citation count*, which is more easily quantifiable. Citation counts operate on measuring the relative worth and disciplinary value of a scholarly document based on the number of times the document is cited by other documents. Citation counts are powerful grading tools to evaluate scientific work, based on the assumption that there is a correlation between citation counts and other performance measures (e.g., department and lab performance).

Some problems with citation analysis include *non-citation*: ignoring the work of others for reasons including because the fact that the material has become “common knowledge” and is no longer explicitly cited. Kuhn (1970) uses the term “exemplars” to describe these types of documents. Also problematic are citations whose functions are merely adulatory and not directly significant, along with acts of self-citation. These and other issues can skew the results of citation analyses. See also “Citation content analysis” and “Citation context analysis.”

Citation category: Several models exist for the categorization of citations. Most focus on the function the citations perform in the citing document. Other schemes include underlying author motivations as part of the categories. For instance, Small (1982) devised 13 function-based citation categories (see Table 1-1).

Although much citation activity in a mature field or area of research is commonly used to lend support and thus demonstrate the author’s affiliation with a particular viewpoint or school of thought, the scheme presented here demonstrates the various other possibilities and motivations for citation.

Table 1-1. Small's Categories of Citations

| | |
|-------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|
| Negation | Disagrees with opinion or findings of original document; critical |
| Partial Negation | Disapproves or questions; mixed opinion |
| Replaced | Offers new interpretation or explanation; negational |
| Confirmation | Approves, verifies, or substantiates a claim |
| Support | Legitimizes or substantiates statement or assumption |
| Background Info | Forms part of relevant literature; historical; further reading |
| Bibliographic Info | |
| Review or Compared | Adds information; data used for comparative purposes |
| Distinguished | Acts in a juxtapositional manner |
| Applied | Acts in an operational manner; uses concepts, definitions, interpretations, data, material, methods, and/or equations or methodology |
| Improved/Modified | Extends or modifies a theory; offers a specific point of departure; acknowledges pioneering work. |
| Changed the Precision | |
| Future Research Implications | |

Citation Context Analysis: This is a type of citation analysis that analyzes the text within which citations (cited references) are embedded in the citing document, either to discern the function of those citations at that point in the citing work or to “characterize some feature of the cited or citing work” (Small, 1978). See Table 1-1 for additional information. Under scrutiny is the overall manner in which the citations (previous works or previous research) are being used by the citing document, as well as the functions of the citations and the author’s reasons for citing. Citation context analysis, unlike citation content analysis, focuses on establishing and analyzing the relationships between the cited and citing documents (Small, 1978; McCain and Turner, 1989).

Citation location: This is the location of the cited passage in the overall text of the citing document. The significance of citation location lies in the pattern(s) that emerge from scrutinizing the same cited document as it is described in several citing documents. For

example, a citation may appear repeatedly in the same section of several scientific papers. If a citation appears repeatedly in the literature review portion of a scientific paper, that citation might become classified as “having historical significance” or “paying homage to pioneers.” Similarly, another citation might occur primarily in the methodology section of a scientific paper and become associated with the idea of “scholarly bricklaying” (Merton) for the extension of ideas presented in the citing document. See also “Citation category.” The combination of citation category and citation location can contribute to the perception/classification of the cited document as a whole.

Cited document: This is the cited reference, the document that is cited by a later document.

Citing document: This is the document that is doing the citing. A few citing documents draw on many cited documents.

Co-citation: This is the citation of multiple documents by another document. The cited documents are connected to each other by the source document that cites them. When citations (or references) for two or more authors, documents, or publications appear in a bibliography, this is called “co-citation.”

“Frequently cited papers represent the key concepts, methods, or experiments in a field, then co-citation patterns can be used in great detail to map out the relationships between these key ideas” (Small, 1973). Furthermore, “[c]o-citation is a rough measure of association between concepts symbolized by the highly cited papers” (Small, 1977).

Co-citation analysis: Co-citation analysis is made possible by citation indexing. Authors — usually first authors — mentioned in the citing document are indexed in a searchable field. It is therefore possible, using a cited reference database such as SciSearch, to search on the “cited reference” field of the records. Combinations of cited authors referenced by the same paper can reveal patterns, changes, and shifts in thought that affect a certain field of study. (See also “Co-citation” and “Concept symbol.”)

Co-citation analysis suggests that the ideas expressed in the separate documents inhabit a similar proximity in intellectual space (White, 1990; White, 1992). If co-citation patterns for different authors are the same, there could be a significant relationship between the ideas being expressed by these authors. White has said that “it is the piling up of co-citations — the fact that their counts over time exceed a certain threshold — that indicates a relationship” (1990).

Concept symbol: In “Cited Documents as Concept Symbols” (1978) and “The Lives of a Scientific Paper” (1984, 83-97), Henry Small explains how, over time, an entire document can be perceived as a “concept symbol” within the area of research to which it belongs, its “specialty community,” and how this concept symbol can change over time as new research replaces old. Small (1984) discusses “what happens to a scientific paper in its intellectual life, that is, the process of conceptual and symbolic transformation a paper can undergo.” In addition, a single paper can have multiple conceptual identities and be a different concept symbol to different research communities (Cozzens, 1982).

Using the ISI Science Citation Index, Small tracked the citation patterns of specific documents over a number of years (how many times the documents were cited, and by whom). In addition to straight citation counts, Small employed co-citation analysis to

track the changes in citation patterns for a paper about basement membrane collagen (Kefalides, 1974). Over time, the paper under scrutiny shifted from one set of co-citation clusters to another, indicating a shift in thinking and perception, not just about the content of the paper itself but also about the research area to which the paper belonged, as that area began to expand and branch off into smaller specialty areas of research. With the passage of some years, the paper began to be used in a different fashion: An initially benign citation pattern (citation function or type) changed with the burgeoning area of basement membrane collagen research. The Kefalides paper was then invoked in the context of controversial viewpoints within the research arena. Thus, the paper changed over time as a concept symbol in its field.

Cue words: As explained by Mani and Maybury (1999) and Endres-Niggemeyer (1998), cue words are words and phrases that can be compiled and “classified as bonus words (positive value), stigma words (negative weight), or null words (irrelevant to sentence selection). “Cue phrases” are based on the same principle as cue words. Some cue words can be used as triggers in the process of extracting relevant text from the citing document. For example, Nanba and Okumura (1999) describe types of cue words that help establish which portions of a document are relevant to text extraction, for the purpose of constructing a coherent document summary.

Fact extraction: The searcher is looking for specific information for example corporate purchases and take-overs, The fact extraction system is only looking to find the specific information everything else is ignored. “[I]ndividual manifestations of specified important notions” are being sought in the source text (Sparck-Jones, 1999). Therefore,

fact extraction is a “closed” approach, “in that the source text does no more than provide source instantiation for previously established generic content requirements” (Sparck-Jones, 1999).

Research front: A research front is “a cluster of co-cited core papers as well as the group of current source papers that cite one or more of these core papers” (Garfield, 1994). ISI has delineated thousands of research fronts and applied them to the SciSearch database in a searchable field.

Research fronts are identified via a system that consists of the last two digits of the year the research front was created, followed by a hyphen, a four-digit number, and a string of descriptors that “name” the research front (e.g., “88-0152; neural nets, associative memory”). The research front’s name is generated by combining the “most frequently occurring words and phrases used in the titles of the citing (source) papers” (Garfield 1994). A paper can belong to more than one research front in a particular year. Research fronts were assigned to records in the SciSearch database from the 1970s to 1997.

Review of the Literature: This is a summary of the previous research presented by a group of scientific documents that are addressing a similar hypothesis. The review of the literature attempts to present the state of knowledge within a given discipline (Cooper, 1989). The review is intended to “replace those earlier papers that have been lost from sight behind the research front” (Price, 1965). The review of the literature is a standard section of formally written scientific papers and often follows the abstract or the introduction. Alternately, an entire paper might be a review of the literature.

Similarity Metric: A similarity word metric is a number between 0 and 1 that is used to compare how alike documents are. A 0 means the documents have nothing in common; a 1 says they are duplicate documents. Essentially, it counts the number of words in common and then normalizes the result, so that large documents do not skew the resulting metric (Salton and McGill, 1983).

Specialty narrative: This is a condensation of the literature in a field of research. Using co-citation analysis, Small (1986) described a way to generate “reviews or synopses of scientific fields called specialty narratives” (see “Co-citation analysis”). The purpose of the specialty narrative is to provide a coherent review of the literature in a particular area of study, in order to bring the researcher “up to speed” in that area, with which he or she may or may not be familiar. Co-citation analysis identifies the key papers in a field of study. “Consensus passages” — those that have been identified as being most representative of the cited document (see “Citation content analysis”) — are extracted from the citing documents. Consensus passages are linked using a unique methodology, resulting in the specialty narrative.

Text extraction: According to Sparck-Jones (1999), “the source text is taken as its own representation, without any interpretation [*italics mine*], and this representation is then subject to a transformation stage which is simply extractive.” Sparck-Jones terms text extraction an “open” approach to summarizing, as there is no prior presumption about what information in the document is important.

Text summarization: This is the process of producing a condensed version of a source text, including only the most “important” parts of the original work. What is deemed important hinges on what the user needs and what the summary is going to be used for, what task it is to be applied to, or what question it will be used to answer. “Text summarization is the process of distilling the most important information from a source (or sources) to produce an abridged version for a particular user (or users) and task (or tasks)” (Mani and Maybury, 1999). “A summary is a reductive transformation of source text to summary text through content reduction by selection or generalization on what is important in the source.” (Sparck-Jones, 1999).

2. Review of the Literature

2.1 Introduction

Until recently, research publication was primarily paper-oriented, with secondary publishers often producing monthly journals containing abstracts of newly published research material so that researchers could stay informed about publications in their area of study (Mani & Maybury, 1998; Teufel & Moens, 1998). Short and indicative, human-created abstracts served as a decision tool, so that readers could decide whether “the source text is worth a visit to the library” (Teufel & Moens, 1998, p. 155).

Because of the labor involved in human abstracting, the concept of automated abstracting emerged as early as the 1950s, but pioneers in the field had limited options in storage, processing speed, and character sets (Endres-Niggemeyer, 1998). In recent years, the picture for abstracting has changed dramatically. Today, research articles are increasingly being made available online, and more and more information services are now serving researchers online (Mani & Maybury, 1998). Search engines and other information retrieval systems can now search for full-text articles (Myaeng & Jang, 1998). The amount of information available online has grown many fold and increases exponentially every year.

Because researchers lack the time to read all the available texts and, worse, have no possibility of ever reading all available texts, text summarization technology has been developed to read texts by machine. Faced with this conundrum—researchers having to

make critical decisions about available information, yet not having the time to read it all—text summarization has become the focus of considerable interest and investment in both the research and commercial sectors (Mani & Maybury, 1998).

Text summarization tries to help users digest the information content of an article by “taking a partially structured source text, extracting information content from it, and presenting the most important content to the user in a manner sensitive to the user’s needs” (Mani & Bloedorn, 1998, p. 358). Faced with such an enormous amount of material to summarize, automatic text summarization systems have altered the nature of the abstraction process, generally reducing it to the task of extraction. That is, text summarization most often uses “heuristics based upon a detailed statistical analysis of word occurrence to identify the text pieces that are likely to be most important and concatenates the selected pieces together to form a final extract” (Mitra, Singhal, & Buckley, 1997, p. 1). The changing nature of summarization continues to be an important area of research.

The emergence of the World Wide Web and the concomitant growth in text summarization systems have provided an “important opportunity for broad application of text summarization systems” (Firmin & Chrzanowski, 1998, p. 325). The goals of automatic summarization presuppose that the full article from which an abstract or extract is created is available in machine-readable form. As a result, abstracts can now have both additional and quite different functions compared with what human-created abstracts used to have (Teufel & Moens, 1998).

Many new techniques of text summarization have been proposed and developed. Domain-independent techniques for automatic paragraph-level summarization (as

opposed to the more common sentence-level) extraction have been proposed (Mitra, Singhal & Buckley, 1997). These new techniques have varied and often have more severe mandates than human-created abstracts. With the information overload on the World Wide Web, it is important that users feel that a text summary has provided information that helps them easily understand the contents of the article without feeling a need to read the full text (Myaeng & Jang, 1998).

Automated summaries therefore must be more accurate and more functional than their human-generated counterparts. Retrieval problems in an overloaded information environment have also necessitated the development of new text summarization techniques. When a user employs a conventional information retrieval system to call up documents, the number of documents retrieved often means that not all documents are of interest or that their interest level is unclear. By matching document summaries with related summaries, a system may give users more information about the relevance of a retrieved document. A matched document summary can be either a generic summary, which gives an overall sense of the document's contents, or a user-query summary, in which the system presents "the content that is most closely related to the initial search query" (Goldstein & Kantrowitz et al., 1999, p. 1). Finally, with the exponential growth of information available on the Web, good hypertext linking between different information items can help users in their searches by "successfully navigating through the colossal amount of information on the Web" (Salton & Singhal et al., 1997, p. 193). Thus, tools for automatically discovering hypertext links have been developed, and methods for automatic text linking have been proposed (Salton & Singhal et al., 1997).

2.2 Advances in Computer Science Text Summarization

2.2.1 Classic Techniques

The history of automated text summarization is characterized by a gradual evolution, proceeding from approaches that focused on surface-level features of the text to entity-level and finally discourse structure (Mani & Maybury, 1998). In the 1950s, Luhn developed a surface-level information retrieval approach that focused on term frequency to extract important sentences that, when put together in a short text or abstract, represented the whole text (Endres-Niggemeyer, 1998). This subset of representative sentences was selected from a text based on a formula that weighted words based on frequency and distribution (Rath & Resnick et al., 1961). Keywords consisted of mentions by the author of important concepts within each sentence, and values assigned to the words were based on an index of keywords (Paice, 1990). The justification for measuring word significance by frequency of use “is based on the fact that a writer normally repeats certain words as he advances or varies his argument and as he elaborates on an aspect of a subject” (Mani and Maybury, 1999, p. 16). Frequency keyword heuristics were based on the assumption that “important sentences contain content words that are present in the title and headings” (Kupiec & Pedersen, 1998, p. 56). The Luhn process entailed weighting words in a complete text by score and deleting all common function words by means of a stoplist (Paice, 1990). Luhn’s method “defined the criteria and direction for subsequent research” (Endres-Niggemeyer, 1998, p. 302).

Such surface-level abstracting methods coincided with information retrieval environments focusing on the use of keywords, or terms, to conduct searches for articles, as well as statistical techniques. “Statistical techniques take advantage of large document collections to identify words that are useful indexing terms automatically” (Riloff &

Lehnert, 1994, p. 297). Statistical techniques, popular because they can be fully automated and can sift through large volumes of documents with relative ease, nonetheless struggle with limitations caused by synonymy, polysemy, phrases, local context, and even global context (Riloff & Lehnert, 1994). Research has progressed from Luhn's surface-level methods as a result of the growing conclusion that "while it is quite feasible to produce a short indicative abstract from a text, it is still very hard to produce a sufficiently coherent and readable one" (Zechner, 1997, p. 2).

2.2.2 Extraction

Increased machine power, exponential increases in data amounts, and a renewed interest in the efficacy of statistical techniques are among external factors that have brought a change of focus in text summarization methods (Zechner, 1997). The sheer size of the World Wide Web has placed enormous pressure on mechanized systems to meet user demands.

As a result, research and development in automatic summarizing in the 1990s witnessed a renewed interest in sentence extraction, which is the "only domain-independent summarization technique that works in application environments" (Endres-Niggemeyer, 1998, p. 331). Sentence extraction selects a few representative sentences from the source documents and represents them in a summary that is indicative only of contents. "The task is well understood with the current technology, but the result may look unnatural because there is no guarantee that a list of representative sentences forms a coherent summary" (Myaeng & Jang, 1998, p. 2).

Sentence extraction techniques are also derived from Luhn's methods. Expanding the idea of weighting keywords, Luhn assigned an overall weighting score to whole

sentences in a document according to statistical criteria, “a simple function of the number of high-frequency words occurring in the sentence” (Edmundson, 1998, p. 24). Variations of stochastic methods of measuring words or phrases coding sentences according to other measures of importance have been developed and have selected the best-rated sentences for extraction in a similar manner (Teufel & Moens, 1997). Since then, many robust summarization systems have opted for statistical sentence extraction, even though the “result of this process is an extract, i.e. a collection of sentences selected verbatim from the text” (Teufel & Moens, 1998, p. 156), and not, strictly speaking, an abstract. Moreover, many sentence-based extracts are “just a collection of sentences, possibly difficult to interpret because of phenomena like unrelated anaphora [common words referring to other areas of the text] and unexpected topic shifts” (Teufel & Moens, 1998, p. 156).

A serious limitation of sentence extraction techniques is that they “do not work very well for high-compression summarization” (Teufel & Moens, 1998, p. 156). With a compression ratio of about 10% to 15% of the original, sentence extraction works fine with short newspaper articles, but with growing interest and emphasis on summarizing larger, longer journal articles, sentence extraction’s compression ratio, which would produce a two-page summary of a 20-page article, produces “a document surrogate which is not adequate as an abstract” (Teufel & Moens, p. 156). “Sentence selection technique produces extracts which can be incoherent and which, because of the generality of the methodology, can give under-informative results” (Teufel & Moens, 1999, p. 1). Fact extraction techniques, moreover, are tailored to particular domains and have not been scaled up from restricted to unrestricted domains (Teufel & Moens, 1999).

In time, researchers came up against several limitations of the keyword- and sentence-extraction shallow summarizing techniques. It is difficult to form coherent abstracts simply by extracting sentences without in turn ranking sentences according to their salience (Goldstein & Kantrowicz et al., 1999). Improvements were made in sentence selection by introducing sharper measurement factors, including textual cohesion, balance, and coverage. Textual cohesion was attained by the resolution of anaphora, while textual balance was determined by the “prejudicial selection of sentences based upon the location of the sentences in the original document” (Brandow & Mitze et al., 1995, p. 676). Even with these improvements, problems persisted in the resulting summary’s acceptability or usefulness (Brandow & Mitze et al., 1995). The attempt to use natural language and other text condensation approaches has also been shown to result in limited improvement, and those approaches “generally require the selection of a narrow domain and the availability of domain knowledge” (Brandow & Mitze et al., 1995, p. 676). As a result of these shortcomings, “natural language approaches to text condensation currently [remain] infeasible for generic text condensation tasks” (Brandow & Mitze et al., p. 676).

Early techniques designed to enhance surface-level keyword and sentence extraction developed, in the 1970s, into the emergence of more extensive entity-level approaches to summarization, “as well as the first discourse-based approaches based on story grammars” (Mani & Maybury, 1998). Methods based on measuring importance through the location of a word, phrase, or sentence came later (Mani & Maybury, 1998). These newer demands resulted from the observation that what really counts in estimating the importance of a concept denoted by a keyword in a text is not the occurrence of the

keyword per se but the referents the word points to (Endres-Niggemeyer, 1998).

Researchers began to sense that the outputs of extraction systems were not summaries, but “simply portions of the text, produced verbatim” (Hovy & Lin, 1998, p. 91). A true summary, should involve “the fusion of various concepts of the text into a smaller number of concepts, to form an abstract” (Hovy & Lin, 1998, p. 91). An abstract is more a “novel phrasing describing the content of the original” (Hovy & Lin, p. 91). Extraction systems such as ADAM address the issue of lack of cohesion and lack of balance by attaching an added feature to sentence extraction for the purpose of reducing anaphora: ADAM simply adds preceding sentences to accompany any sentence whose first clause contains an anaphor. By doing so, ADAM became “the first computer program to produce abstracts rather than extracts” (Paice, 1990, p. 17).

The problem of anaphora was ameliorated by adding adjacent sentences. Such techniques created “tidy” passages, defined as passages having “no anaphor or other referential device that is not resolved within the same sentence or passage” (Paice, 1990, p. 179). The computer program GARP scans a text and applies appropriate contextual rules “whenever it encounters a potentially anaphoric word” (Paice, 1990, p. 176). The program decides not only whether a word is anaphoric but also “whether the antecedent lies within the current sentence or elsewhere” (Paice, p. 182). Attempts to deal with all the anaphors in text—third-person pronouns, nominal demonstratives, indefinites, quantifiers, and normal substitutes—forced researchers to explore the logical and rhetorical structures of a whole text, and to factor into summarization “cohesive features which indicate the nature of the relationship between a sentence or clause and its predecessor or successor” (Paice, 1990, p. 117).

2.2.3 Natural Language

Sentence extraction is but one end of the spectrum of summarization techniques now being explored. At the other end of the spectrum are attempts to create summaries based on understanding sentences or documents. These methods produce a more natural-looking summary, but the task is difficult and unreliable, “primarily because of the limited state of natural language processing techniques” (Myaeng & Jang, 1998, p. 62).

Natural language processing is an area of renewed interest but with a change: Whereas natural language was formerly explored as a method for communicating with computers, now it is studied as a “method of obtaining information from computer-readable text” (Jacobs & Rau, 1993, p. 4). Those who support natural language processing stress its urgency, not only because of the need for automatic methods to deal with the proliferation of information on the Web, but also because it seems to be the technology with the most potential for extracting accurate information (Jacobs & Rau, 1993). At one end of the spectrum, then, is a shifting paradigm, moving away from research motivated by linguistic theory and toward applied technology (Zechner, 1997), while at the other end the development of more storage and more processing speed has given researchers the freedom to “constructively think of interactivity, visualization, and multimedia solutions” (Endres-Niggemeyer, 1998, p. 331).

In this new environment, the abstract or summary has changed dramatically—a change that itself has motivated still more change. The old type of abstract was a “fixed, long-lived stand-alone text, targeted at one particular type of user” (Teufel & Moens, 1998, p. 156). The new type of abstract is more dynamic and user-responsive, is

generated automatically when needed, and is admittedly of lower quality than human-crafted abstracts, but it is more useful in many situations.

One area of research that is seen as critical in helping users wade through the overload of information is the use of an abstract as a navigational tool. Because abstracts are generated as needed, they can now show how certain articles are related, can contain pointers to certain select passages in the full articles, can be embedded in the source text itself (highlighting the most relevant sentences, thus playing an important role in the non-linear reading of text), and can assist users who jump from titles to section heads to captions and on to citations and other sources (Teufel & Moens, 1998). A summarization that takes all navigational possibilities of the abstract into consideration can, it is believed, give users an advantage against information overload (Teufel & Moens, 1998).

2.2.4 Cue Phrase and Location

With the focus in summarization turning to content factors, several researchers (as early as Edmundson in 1969) developed abstract generation systems that use the cue, title, and location methods, each based on a weighting heuristic, to determine word importance by means of overall corpus features (Zechner, 1997). The cue method is based on the observation that “certain of the words or phrases occurring in a sentence, though not in themselves keywords, nonetheless provide an indication of whether or not the sentence deals with important material” (Paice, 1990). Using the cue method, non-keyword cue words are coded as either bonus words, which increase scores, or stigma words, which decrease scores (Paice, 1990).

The indicator-phrase method is based on the observation that most authors use common expressions to introduce or accompany explicit statements about the topic of a

text. Again, although these phrases—such as “the main aim of our paper is” and “the purpose of this article is”—are not themselves keywords, important material is likely to be located where they occur (Paice, 1990, p. 174).

The location or position method is one of the most popular entity-level extraction techniques. It “springs from the recognition that texts in a genre generally observe a predictable discourse structure, and that sentences of greater topic centrality tend to occur in certain specifiable locations” (Lin & Hovy, 1995, p. 1) The location method examines “certain general characteristics of the corpus provided in the skeleton of the document” (Edmundson, 1998, p. 31), such as titles and headings, to locate important concepts. The method is reinforced by the observation that sentences occurring directly under headings are usually positively relevant, and that other important topic sentences often occur very early or very late in the document . In the location method, sentences receive various scores according to whether they occurred at the beginning or end of a paragraph (with first sentences, usually central to a theme, scoring high), at the beginning or end of the document, or below a heading (Paice, 1990). The location method has been found to outperform methods based on word counting, which are robust but generally less accurate (Lin & Hovy, 1995).

The location method has produced some surprising successes, such as the finding that summaries produced by simply extracting the first sentences of a story or article are adequate (Brandow & Mitze et al., 1995). But the fact that the location method is based on a resolution power focused on sentences, whereas output usually occurs at the word or phrase level, means that results of a search can “contain too much spurious material” (Lin & Hovy, 1995, p. 1). Other experiments have failed to find a correspondence between the

position of a word in a sentence or paragraph and its relation to significance (Lin & Hovy, 1995). Still, Lin & Hovy did empirically determine a high degree of relevance in sentences extracted by the position method and a “high degree of coverage” of relevance compared with human abstracts (p. 2). They concluded that “the results gained from what is, after all, a fairly simple technique are rather astounding” (Lin & Hovy, 1995, p. 8).

2.2.5 Entity Summarization

Today, summarization can be characterized as “approaching the problem at the surface, entity, and discourse levels” (Mani & Maybury, 1998, p. X). Surface-level approaches represent information with shallow features selectively combined to yield a salience factor that extracts information. Such features can include thematic features, location, background, and cue words and their relationships, and are related to “classic” information retrieval systems. Entity-level summarization builds an internal representation of the text, modeling all corpus entities and their relationships—including similarity; proximity; co-occurrence; thesaural relationships among words; co-reference; and logical relations such as agreement, contradiction, entailment, and consistency—to determine salience (Mani & Maybury, 1998). Finally, discourse structure approaches model the text and mark its rhetorical or narrative structures for hypertext links. The discourse structure in the paper is thus used to create coherent summaries, including extracting paragraphs instead of sentences, or training extraction systems on the model of human abstractors to make relevance judgments about individual documents in the context of a discourse as a whole (Endres-Niggemeyer, 1998).

One entity-level extraction method focuses on syntactic criteria, “based on the hypothesis that the extract-worthiness of sentences might be correlated with their syntactic structure” (Paice, 1990, p. 173). Some recent studies of this method have determined that at present it has limited usefulness (Paice, 1990).

A more complicated entity-level method has developed around the idea of studying the relationships between elements in a text as a whole by creating a relationship map of the text, in which “every text or text excerpt is represented in vector form as a set of weighted terms” (Salton & Singhal et al., 1995, p. 1). According to this method, theme and segment text decomposition is used to distinguish between simple and complex text structures (Salton & Singhal et al., 1995). The resulting text classification system is used “as a basis for the generation of text retrieval and text traversal operations” (Salton & Singhal et al., 1995).

Maps of texts have been most useful in extraction systems that extract paragraphs, not sentences. Paragraph extraction has grown because a paragraph gives more context than a sentence and is usually self-contained, so extraction at the paragraph level avoids some of the “evident coherence problems such as dangling references” (Endres-Niggemeyer, 1998, p. 333) in sentence extraction. To use a paragraph extraction system, the whole text must be represented in vector form, computed by pair-wise similarity coefficients, with the paragraphs represented by nodes and joined by links “based on the numerical similarity computed for each pair of paragraphs” (Endres-Niggemeyer, 1998, p.333). The resulting text relationship map makes clear how the paragraphs of the text are related to each other (Mitra, Singhal & Buckley, 1999, p. 2). As in a sentence extraction semantic structure map, the idea is that the most important paragraphs in the document

will show the largest number of relationships on the map (Paice, 1990). This quality, termed the “bushiness” of the node, displays links connecting one paragraph to another paragraph’s node; the more nodes that are connected, the bushier and more significant is the node (Mitra, Singhal & Buckley, 1999, p. 3). The bushiness of the node in turn represents the overlapping vocabulary between one paragraph and another, leading the researcher to determine the most relevant and salient ideas in the text as a whole (Salton & Singhal et al., 1997). The presence of bushiness, or intradocument text linking, allows searchers to isolate the “functionally homogeneous” passages, which are called text segments, defined as “a contiguous piece of text that is well linked internally, but largely disconnected from the adjacent text” (Salton & Singhal et al., 1997, p. 196). Once all relevant text segments are determined, a program traverses “the selected node in the text in order to construct an extract” (Salton & Singhal et al., p. 198).

Text-level structures of research papers have been applied to areas such as “indexing, automatic abstraction, reference interviews, text retrieval, information extraction, and the design of user interfaces and electronic journals or digital library systems” (Kando, 1999, p. 3). Their usefulness derives from the fact that lexical cues allow one to analyze the functional structure of a paper, a technique “useful for grasping the outlines of the studies” (Nanba & Okumura, 1999, p. 2). Systems such as BREVIDOC improve on keyword-based systems by retrieving full texts using a document structure analyzer, a syntactic analyzer, and a text structure analyzer, all to determine the outer structure of a document (Zechner, 1997). Along with these techniques, the 1980s witnessed an explosion of other entity-level approaches, often

based on artificial intelligence, such as the “use of scripts, logic and production rules, semantic networks, as well as hybrid approaches” (Mani & Maybury, 1998, p. 339).

2.2.6 Discourse Summarization

The basis of a “text understanding” approach to text summarization is that every sentence contains relevant information—if not explicitly in keywords or by location or position, then in some unseen relation between the sentence and other sentences “through some relation of coherence or rhetorical structure” (Appelt, 1999, p. 161). Thus far, however, work has suffered “from a lack of appropriately annotated corpora that can be used for building, training, and evaluating summarization systems” (Teufels & Moens, 1999, p. 1). Somewhat like entity-level analysis of position or location, discourse-level analysis of text for summarization focuses on the discourse macro structure or rhetorical structure of the document (Strzalkowski & Stein, 1998). The discourse structure is obtained by asking “what is a summary?”—automating the way in which humans intuitively review the full linguistic structure of a text—and does so by using a corpus of typical or ideal abstracts that analyze the summaries and then creates predefined summary templates that are filled by extracting information from the text being summarized (Strzalkowski & Stein, 1998).

Summarization working at the discourse level is therefore concerned with training a system by having it learn the properties of sentences in abstracts and using that knowledge to extract similar abstract-worthy sentences from unseen texts (Teufel & Moens, Discourse-level argumentation). This is done by a statistically based model that “determines the degree to which individual sentences belong to a summary, by taking

into account lexical and statistical information obtained from a training corpus” (Myaeng & Jang, 1998, p. 61).

Discourse macro structure analysis was inspired by the observation that certain types of texts, from news articles to technical reports, conform to a set of style or organizational constraints “which help the author to achieve a desired communicative effect” (Strzalkowski & Stein, 1998, p. 138). Indeed, others claim that the “predictable patterns of methodology” in research have resulted in “a rigid, highly structured building plan for research articles, in which rhetorical division—introduction, purpose, experimental design, results, discussion, conclusions—are clearly marked in section headers” (Teufel & Moens, 1998). As a result of this predictability, the main discourse structure of a paper has come to be seen as a “fixed rhetorically annotated tree structure” (Teufel & Moens, 1998, p. 170).

Discourse structure will be useful, some researchers claim, because it models strategies used by human abstractors. Teufel & Moens (1998) envisaged a full rhetorical structure as an argumentative template, “where the slots represent certain argumentative or rhetorical roles, such as goal, achievement, background, and method” (p. 156). Once these roles are marked, “abstracting means simply analyzing the argumentative structure of the source text and identifying textual extracts which constitute appropriate filters for the template” (Teufel & Moens, 1998, p. 156).

Discourse structure approaches involve several complicated steps in the process of summarization. First, rules must be generated and the program “trained” by analyzing a “training corpus” of typical or ideal texts. Second, a number of preprocessing programs must be run, to mark up the whole text in various ways. Finally, these mark-up processes

can annotate any aspect of the document for reference to other similar documents in the discourse as a whole, by hypertext linking. A number of programs using a combination of these methods are being developed.

Fundamentally, discourse methods of summarization select sentences from a full document based not on keyword counts or location scoring but on a “statistically based model that determines the degree to which individual sentences belong to the summary, by taking into account lexical and statistical information obtained from a training corpus” (Myaeng & Jang, 1998, p. 61). Some of the statistical information acquired from the training corpus involves the identification of text components, the ranking of sentences, the elimination of similar sentences, and the production of a summary (Myaeng & Jang, 1998). In searching for sentences based on their argumentative roles, the search is undertaken with the support of training materials, “a collection of texts where each sentence is annotated with information about the argumentative roles that the sentence plays in the paper” (Teufel & Moens, 1999, p. 4). Teufel & Moens’ (1999) annotation scheme includes seven rhetorical moves, from explicit statements of research goal to good characterizations of the paper as a whole, to provide a good overall summary of the article (p. 4). Argumentative zoning of a text consists of “the task of breaking a text containing scientific argument into linear zones of the same argumentative status, or zones of the same intellectual attribution” (Teufel & Moens, 1998, p. 7). This type of segmentation tries to capture the argumentative state of every part of the text with respect to the overall argument of the text. Other argumentative goals could be the degree to which a sentence expresses the main goal of the source text or calls into question

someone else's work, in which case larger discourse relevance is brought into the annotation (Teufel & Moens, NO DATE, Discourse level, p. 2).

Along with using training corpora to model the search process, discourse-level summarization also uses a number of natural language preprocessing programs (Appelt, 1999; Jacobs & Rau, 1993; Zechner, 1997). Using algorithms of various kinds, preprocessing “seems to have particular promise for the quality and efficiency of later processing” (Jacobs & Rau, 1993, p. 166). Preprocessing modes or techniques include tagging, template activation (including topical analysis), and segmentation (or bracketing). All of these methods—comparable to the human capacity to skim documents—assign additional information to each word in a text to assist in interpreting the word in a summarization process (Jacobs & Rau, 1993). Statistics based on the relative frequency of words occurring in the document, the query, and the corpus as a whole (Appelt, 1999); tokenization—which, like tagging, attaches additional features onto the input texts, including part of speech, root forms, role in phrases, and other features; and still other techniques such as text-tiling, which “identifies coherent passages in a given text, to find topic boundaries, which is done by computing similarity of adjacent blocks of text” (Zechner, 1997, p. 22), as well as hidden Markov models and other clustering methods are all part of the broad yet customized area of preprocessing.

Not only do the most advanced forms of text summarization move from a corpus entity level to a discourse structure analysis, but they exist at a junction of various paradigms of text retrieval or summarization, including information extraction, information retrieval, cognitive science, natural language processing, and linguistics (Endres-Niggemeyer, 1998; Mani & Maybury, 1998). Text summarization now proceeds

along its own theoretical path, and it is increasingly exploited in the commercial and academic fields, with summarization tools appearing in the telecommunication industry, in data mining of text databases, in filters for Web-based information retrieval, and in word processing tools (Mani & Maybury, 1998). As a result, an explosion of approaches that exploit the global discourse structure in various ways has developed. New summarization systems do not simply extract information on a practical level, as in earlier computerized summarization systems, but, inspired by cognitive science, seek to incorporate a model of the intellectual process itself during the summarization process (Endres-Niggemeyer, 1998). Also, theories about human cognitive organization have created base system architectures for newer summarization systems, which use semantic representations of text meaning derived from a pre-search study of the input (Endres-Niggemeyer, 1998). Discourse-level approaches to text summarization therefore are primarily involved in modeling the global structure of the text and its relationship to its overall communicative goals in the literature. Determining this structure has included such techniques as a hypertext markup of the format elements of the document, revealing threads of topics that are embedded in the text, and outlining and making explicit the rhetorical structure of the text, such as the argumentative structure and the narrative structure (Mani & Maybury, 1998).

Generally speaking, these approaches also move from text extraction to fact extraction, because the use of a corpus has resulted in algorithms that have already decided what to look for in the source documents, and the search simply provides “an instantiation for previously established generic content requirements” (Jones, 1998, p. 2). Finally, such systems have become more deeply involved than ever in comparing the

quality of human versus machine-made abstracts, and have determined that human abstracts are characterized by an intuitive sense of abstract-worthiness in semantic content and appropriateness of a segment for representing the contents of a document that must be replicated in a machine-generated extract by finding “an operational approximation to the subjective notion of abstract-worthiness” (Teufel & Moens, 1997, p. 2). Some researchers have called the criteria of what constitute abstract-worthiness “the gold standard” and have set out to create summaries entirely out of such sentences: “A gold standard sentence is a sentence in the source text that is matched with a summary sentence on the basis of semantic and syntactic similarity” (Teufel & Moens, 1997, p. 2).

2.2.7 Information Retrieval — Hypertext

A final and important level in the process of discourse-level summarization is the evolution of hypertext linking and other techniques, which have derived from the field of information retrieval. Related to summarization, information retrieval is more directly aimed at retrieving useful documents in response to a user query (Salton & Singhal et al., 1997). In this task, it is important to link articles by pair-wise links of similarity (an idea related to co-citation analysis in information science). Salton & Allan (1993) proposed a vector-processing model as an alternative to Boolean operators, because vectors can be compared with each other and obtained by vector similarity coefficients between texts (Salton & Allan, 1993). By traversing full texts from relevant text segment to related text segments elsewhere, one not only develops a way of “reading” text that extends beyond sequential beginning-to-end reading, but also can traverse large collections of documents moving from text to text (Salton & Allan, 1993). A text structure linking various and similar texts at many levels of detail enhances searching as well as summarization (Salton

& Allan, 1993). Such linked structures, often called hypertext, “make it possible for the reader to start with particular text passages and use the linked structure to find related text elements” (Salton & Allan et al., 1994, p. 1,421).

Because up to now there have been no viable methods for automatically building large hypertext structures and for “using such structures in a sophisticated way” (Salton & Allan et al., 1994, p. 1,421), the expansion of discourse-level summarization to hypertext represents an important development, as well as convergence with information retrieval, in text summarization. An awareness not only of the rhetorical structure of a paper itself but also of its role in a larger body of literature opens the possibility of measuring still finer context factors—ranging from input, purpose, and output factors to consideration of audience and intended use of the paper—in improving summarization. In this manner, queries can be tailored in a detailed way to meet the context requirements of the searchers, and text summaries can be created that meet users’ queries more accurately (Jones, 1998).

2.2.8 Information Extraction

Numerous working summarization systems that function on discourse-level extraction are in operation. The TOPIC/TWRM-TOPOGRAPHIC system processes relatively short articles in the field of information science (Endres-Niggemeyer, 1998), while SCISOR (System for Conceptual Information Summarization, Organization, and Retrieval) summarizes newspaper stories “using a conceptual representation of knowledge about possible events” (Endres-Niggemeyer, 1998, p. 319). Like many of these systems, the memory organization of SCISOR is inspired by “current ideas about human memories and cognition” (Endres-Niggemeyer, 1998, p. 323). SCISOR performs text analysis and question answering in the domain of financial news and selects stories

for abstracting using a combination of artificial intelligence and more shallow methods (Zechner, 1997). FASTUS extracts information from natural language texts using a “multiple-stage cascaded non-deterministic finite-state transducer” and is noted for its fast runtime and its quick adaptation time when moving to a new domain (Zechner, 1997). Typical of discourse-level systems, the University of Massachusetts system uses both decision trees and other machine-learning techniques, such as string specialists, a part-of-speech sentence analyzer, and a fully automated dictionary construction system, that allow for the automatic training of the system (Zechner, 1997).

In the category of systems that make relevance judgments based on rhetorical structure trees, SIMPR “draws upon surface linguistic constraints developed in corpus linguistics and upon the knowledge of professional indexers” (Endres-Niggemeyer, 1998, p. 344). The STREAK system uses conceptual summarization to produce short reports on professional basketball games, while SUMMONS demonstrates how “summaries of a series of news articles about the same event can be generated ... from sets of templates produced by information extraction systems” (Endres-Niggemeyer, 1998, p. 356). In SUSY, the summarizing strategies follow the human approach, using internal text representation, word expert parsing, and the syntax specialist techniques (Endres-Niggemeyer, 1998). While SIMPR skillfully integrates methods from different backgrounds, STREAK and SUMMONS “focus on summary text generation from structured knowledge” (Endres-Niggemeyer, 1998, p. 333).

Like most information extraction systems, “defined as the identification of instances of a particular class or events of relationships in a natural language text, and the extraction of the relevant arguments of the event or relationship” (Grishman, 1995, p. 1),

these systems both extract facts from documents through local text analysis and integrate the facts to produce larger facts that are in turn integrated to create “the required output format” (Grishman, 1995, p. 3). Although it is clearly discourse level in its approach, information extraction is “situated somewhere between information retrieval and text understanding on the spectrum” (Appelt, 1999, p. 162). Information extraction differs from information retrieval by analyzing “unrestricted text in order to extract specific types of information”, where relevance is “determined by predefined domain guidelines which must specify exactly what types of information are expected to be found” (Lehnert, 1996, p. 1). Information extraction is also characterized by its interest in extracting relevant facts and representing them in a useful form, as well as identifying passages that may contain relevant information (Appelt, 1999). Because of their direct and pragmatic nature, information extraction systems have been designed for use in health-care delivery, monitoring of technical and scientific literature, intelligence gathering, and competitive intelligence (Lehnert, 1996). Because these systems operate at many levels, “from word recognition to sentence analysis, and from understanding at the sentence level on up to discourse analysis at the level of the full text document” (Lehnert, 1996, p. 1), and because the systems must build dictionaries in the face of all the jargon and proper names included in any literature, “it is not easy to build an operational information extraction system” (Lehnert, 1996, p. 1).

Several hybrid forms of summarization or extraction exist, based on the specific goals and needs of particular systems. Riloff & Lehnert (1994) presented “an alternative to full-blown natural language processing” with an information extraction system, eschewing complete analysis of a document in favor of fulfilling a well-defined task,

which simply “extracts certain types of information from a document” (p. 299), allowing everything else to be dropped. “Information extraction is less computationally expensive than natural language processing because many phrases and even sentences can be ignored if they are not relevant to the domain” (Riloff & Lehnert, 1994, p. 299). A variation of this technique is the selective concept extraction method, “essentially a form of text skimming” (Riloff & Lehnert, 1994, p. 300), which selectively processes texts using a conceptual sentence analyzer called CIRCUS, built of a “domain-specific dictionary of concept nodes, or structures that extract relevant information from a sentence” (Rilof & Lehnert, 1994, p. 300).

Elsewhere, SUMMARIST employs information technology techniques “as far as they can take us” (Hovy & Lin, 1998, p. 83) and then augments them with symbolic-semantic and statistical methods. SUMMARIST counts concepts as well as words, operating on both surface and deeper levels (Hovy & Lin, 1998).

2.2.9 Information Retrieval

The fact that much text summarization today takes place in the context of online search engines has caused a convergence of automatic text summarization and information retrieval (Ruthven & Tombros et al., no date). The short-term role of the query process has placed pressure on summarization systems to deliver results. Researchers are not only looking more closely at the query formulation process as a possible shortcut to better summarization; they are also investigating the use of summaries for interactive searching. Several techniques using highly matching sentences now seek to produce summaries “tailored to the user’s query in a detailed manner“(Ruthven & Tombros et al., no date, p. 1). Moreover, query formulation,

recognized as “one of the demanding activities in information seeking” (Ruthven & Tombros et al., no date, p. 1), is being studied, to improve searches from the outset. The use of “relevance feedback methods ... designed to overcome the difficulties in selecting query terms by detecting which terms in a collection are good at retrieving relevant documents” (Ruthven & Tombros et al., no date, p. 1) will greatly relieve some pressure from summarization itself. Relevance feedback weights all terms in a document, in a way that “reflects how well the term discriminates relevant from non-relevant documents” (Ruthven & Tombros et al., p. 4).

One area where text summarization and information retrieval have converged is the renewed interest in knowledge-based information retrieval systems: “A great deal of work has recently been done on knowledge-based information retrieval systems” (Riloff & Lehnert, 1994, p. 298). Like discourse structure text summarization systems, knowledge-based information retrieval systems rely on an explicit knowledge base, “such as a rule base, semantic network, patterns, or case frames” (Riloff & Lehnert, 1994, p. 298). Although many of these systems have had good success, that success is in limited domains. Also, they require “an extensive manual knowledge-engineering effort to create the knowledge base” (Riloff & Lehnert, 1994, p. 298). Manual knowledge engineering is “a time-consuming and tedious process that may require several years of effort by experts who are highly experienced with the domain and the task” (Riloff & Lehnert, p. 298).

Another new front of research is text interpretation systems, which can extract key information while tolerating a high degree of error. With the surprising advances in recent text understanding research, along with the demands of online text retrieval, “text interpretation systems have scaled up from sketchy processing of a few stories to fairly

accurate categorization and data extraction operating on significant volumes of input text” (Jacobs & Raus, 1993, p. 144).

2.2.10 Multiple Documents

The scale-up in online information has also been the “forcing function” behind new natural language research—statistical pre-processing, corpus analysis, and a variety of parsing architectures have been mentioned—and the explosion of new natural language techniques. An additional problem caused by the explosion of online information is the existence of multiple documents covering similar information, “as in the case of multiple news stories about an event or a sequence of events” (Mani & Bloedorn, 1998).

Multidocument summarization has been developed to summarize the similarities between such documents and relieve the searcher from reading all of them. Several approaches have been developed to extract content from multiple documents and summarize them. Natural language message understanding systems are able to extract relationships of similarity between documents and identify areas of agreement (Mani & Bloedorn, 1998). Mani & Bloedorn (1998) also developed a tool that can be used to detect and align similar regions of text among members of a collection and “detect relevant differences among members” (p. 359). In multidocument summarization, the number of documents to be summarized “can range from large gigabyte-sized collections to just pairs of documents” (Mani & Maybury, 1998, p. 338). Various ways of characterizing relationships between documents have been developed, including “part-whole relationships, differences of perspectives, identifying algorithms which scale up to large-sized collections, eliminating redundancy across documents, exploiting orderings among

documents in intelligent ways, and making use of effective presentation and visualization strategies to represent relationships” (Mani & Maybury, 1998, p. 338).

2.2.11 Evaluation Methods

A final area of research that is helping to improve text summarization is the increased interest in evaluation methods of text summarization systems (Hovy & Lin, 1998; Zechner, 1997). Summarization evaluation begins with a comparison of machine-made and human-made abstracts. Hovy & Lin (1998) argue that “compared to the complex processing people perform when summarizing, automated summarization techniques are likely to remain mere approximations for a long time yet” (p. 82). Noting that the introduction of semantics-based artificial-intelligence techniques in the 1970s promised to bring summarization up to human level, researchers praise renewed interest in this area of research (Hovy & Lin, 1998). On the other hand, Zechner (1997) noted that in a review of both human and machine-made abstracts, although subjective judgments indicate the superiority of human abstracts in readability, “there was no marked difference in terms of conveying the essential information between extracts and abstracts” (Firmin & Chrzanowski, 1998, p. 325). Zechner’s (1998) conclusion is that although less fun to read, machine extracts “still fulfill the same purpose as abstracts do and do not even produce noticeable time delay in digesting the information” (p. 25). One issue is decided in a definitive manner when comparing human and machine summaries: Humans will never compete with machines in throughput. At present, humans require between 15 and 60 minutes to process a single document, while a slow information extraction system takes less than 3 minutes and the fastest take only 30 seconds (Lehnert, 1996).

An issue that complicates an evaluation of summaries is that humans themselves show little predictable reliability in their selection of representative sentences during the abstracting process. In a recent study of how professional human abstractors summarized documents, “each subject on the average selected the same sentences only 55% of the time” (Resnick, 1998, p. 291). This lack of reliability suggests both that a single set of representative sentences for a text may not exist and that many equally representative sets of sentences exist for any article (Resnick, 1998).

Other evaluation systems attempt to match a system summary against an ideal summary, but researchers find that “the ideal summary is hard to establish” (Mani & Maybury, 1998, p. 283). In information extraction systems, two important metrics are used to assess the performance of the summarizing system: recall, or “how much of the information that should have been extracted was correctly extracted” (Lehnert, 1996, p. 2), and precision, which “refers to the reliability of the information extracted” (Lehnert, p. 2). Goldstein & Kantrowitz et al. (1999) developed a normalized version of precision-recall curves to evaluate a summarization system, finding that an evaluation must also take into account both the compression ratios and the characteristics of the document set being used (p. 8). That is, simply measuring precision and recall does not indicate “whether the improvement of one summarizer over another is significant or not” (Goldstein & Kantrowitz et al., 1999, p. 6).

Additional summarization systems use linguistic knowledge and a statistical control. A recent study reviewed a summarizer that uses, in its evaluation phase, a “cosine distance metric (of the SMART search engine) to score sentences with respect to query” (Goldstein & Kantrowitz et al., 1999, p. 8). Query expansion methods are added

to summarization systems when a query has been found to be wanting. These expansions have been shown to improve performance, and they may also improve the original summarization process as well. Goldstein & Kantrowitz et al. (1999) evaluated “the relative benefits of various forms of query expansions for summarization by forming a new query, adding the top-ranked sentence of the document, the title, and the document’s first sentence” (p. 8), and saw improvements in the summaries.

In general, there are two types of summary evaluation: intrinsic, which measures a system’s quality, and extrinsic, which measures a system’s performance in a specific task (Goldstein & Kantrowitz et al., 1999). Intrinsic evaluations strive for a “gold star summary” (Firmin & Chrzanowski, 1998, p. 325)—that is, qualities believed to constitute the best summary possible of a given text. An example of how evaluation has been added to summarization systems is DARPA’s expansion of TIPSTER to include a full-text summary evaluation post-processing stage. The evaluation reviews the “output of various system approaches with respect to specific summarization tasks and provides feedback to their developers” (Firmin & Chrzanowski, 1998, p. 325). The qualities reviewed include intent, or the potential use of the summary; focus, or the scope of the summary; and coverage, which “refers to whether the summary is based on a single document or multiple documents relating to the same subject matter” (Firmin & Chrzanowski, 1998, p. 325). In intrinsic evaluations, sentence selection in the abstract is compared with a target abstract, often a template of manually generated key concepts, and also uses a “set of statistics to determine if the summary effectively captured the focal concept, the nonfocal concepts, and conclusions of the full text” (Firmin & Chrzanowski, 1998, p. 326). At present, despite advances, “serious questions remain” about which type of evaluation is

best (Mani & Bloedorn, 1998, p. 373), and there is a lack of consensus about what is the best basis for the comparisons on which evaluations are built. “The problem of evaluating text summarization is a very deep one” (Mani & Bloedorn, 1998, p. 373).

2.2.12 Summary

All in all, text summarization research is going through a highly robust period, with developments in surface-level, entity-level, and discourse-level summarization, and the concomitant development of cognitive science and the renewal of interest in natural language processing to help summarizing machines better understand the texts they are processing, much in the manner of human abstractors of old. Mani & Maybury (1998) go so far as to call the current period in summarization research a “renaissance in the field” (p. xi). While all three approaches are being explored aggressively, current research is characterized by a focus on extracts, not abstracts, although more natural-language-generation work is expected to focus on summarization in the coming years (Mani & Maybury, 1998). And multidocument summarization, multilingual summarization, and even multimedia summarization are developing fields.

For all of these advances in discourse-level abstracting, however, research in the corpus-based approach still faces challenges, including the difficulty in creating and making available suitable text corpora, ensuring that a suitable number of summaries is available, and extending these approaches to the production of coherent extracts (Mani & Maybury, 1998). There is also a larger, intractable problem: “Compared to the complex processing people perform when summarizing, automated summarization techniques are likely to remain mere approximations for a long time yet” (Hovy & Lin, 1998, p. 82).

2.3 Advances in Citation Analysis in Library and Information Science

Parallel to the progress in summarization methods, from surface level to entity level to discourse structure level, citation analysis—a field of library and information science—has undergone a similar evolution and now appears to converge with computer science to propose still another means of automated text summarization: analyzing citations (Sengupta, 1992; Garfield, 1994; Small, 1995). Citation count in library and information science, are used to determine an author's output, the corporate source of the article, or the life or use of the paper as information retrieval strategies (Small, 1999). Here too, these “classic” information science techniques have faltered in the face of information overload, and research has proceeded to engage in citation content analysis and citation context analysis as a comparable means of transforming a citation count into a navigational tool (Liu, 1993). By generating accurate summaries defined by a consideration of the semantic content of a citation, as well as the author's citing behavior, and by how citation context can—through techniques such as co-citation analysis and bibliometric coupling—pair up in clusters with other documents, lists of co-citations, which define a field, are developed. The task of creating a further summary of a field called a specialty narrative represents a convergence with text summarization (Small, 1974). As advances in citation analysis in the field of library and information science proceed and merge with the field of automated text summarization in computer science, questions about whether or not co-citation analysis can be effectively automated have become a pressing area of research (Small & Greenlee, 1980).

Citation analysis has progressed from the 1950s, when citation counting techniques were developed as a way to measure the importance of a document in a bibliography, to citation content and citation context analysis—and extensions of such analysis, such as

co-citation clustering—to develop specialty narrative summaries of fields of research as a whole (Garfield, 1994).

When, in the 1950s, statistical and mathematical techniques began to be used to study the nature of bibliographical organizations and services, the science of bibliometrics was born (Sengupta, 1992). The term *bibliometrics* was coined in 1969 by Pritchard, using specific statistical techniques to examine the interconnectedness of written documents (Sengupta, 1992). (When statistical techniques were subsequently used to study library organization, information systems, and scientific disciplines, respectively, librametrics, informetrics, and scientometrics were also developed [Sengupta, 1992]). The premise of bibliometrics is that “communication can be quantified ... [by] the study and measurement of the publication patterns of all forms of written communication and their authors” (Sengupta, 1992, p. 79). A major thrust of bibliometrics is citation analysis, “which is based on a hypothesis that any act of citing the author of an earlier paper is always meaningful” (Sengupta, 1992, p. 81). The citation count is the basic method of citation analysis. Later, Garfield (1955, 1967, 1970) developed a new direction to citation analysis by introducing the Scientific Citation Index (SCI), which has triggered movement in the direction of bibliographic coupling, co-citation analysis, and the clustering of scientific papers (Sengupta, 1992). Garfield’s research has suggested the “possibility of creating citation indexes for science and the possibility of using citations to search the literature across disciplinary boundaries” (Small, 1995, p. 118).

Citation-based indexes were created to facilitate information retrieval and dissemination using source-reference connections (Garfield, 1997). Originating as

“derivative subject indexing,” in which the titles of papers cited in reviews were used to provide descriptive terms to automatically index papers, citation-index–based searching now includes such strategies as simple cited-reference searchers, reference cycling, co-citation, and bibliographic coupling (Garfield, 1997, p. 5). Subsequent research determined that the sentences in cited or citing texts included “detailed, descriptive indexing statements about papers or books cited” (Garfield, 1994, p. 1). Through “context analysis,” these statements were added, as descriptive words, to indexes, to enhance index descriptors and thus retrievability (Garfield, 1994, p. 3). Further research has explored the idea of drawing citation network maps and, by traversing the “narrative” of such maps, creating specialty narratives summarizing fields as a whole (Small, 1995; Small, 1974). By mapping out and utilizing the reference information that flows between papers, the goal of automated generation of summaries for an entire field of science has emerged as well (Nanba & Okumura, 1999).

2.3.1 Citation Analysis

Citation analysis is based on a few simple assumptions. First, citation analysis argues that “the number of citations a paper receives is a guide to its quality, and that authors producing high-quality papers tend to be more visible to the scientific community and to receive more recognition” (Gilbert, 1977, p. 118). Second, citation analysis argues that papers that jointly cite other papers are related to each other (Nanba & Okumura, 1999). The instrument of citation analysis is the citation count, which continues to be used to “study and draw conclusions about nearly every aspect of scientific work today” (Shadish & Tolliver et al., 1995, p. 477). Citation counts are used to study everything including national scientific policies, disciplinary development, department performance,

research laboratory performance, the careers of scientists, journals, and individual scientific workers (Shadish & Tolliver et al., 1995). Citation counts are powerful grading tools: “Partly on the basis of citation counts, some departments are ranked higher than others in national surveys, and some scientists are deemed more eminent and some scientific works deemed more important” (Shadish & Tolliver et al., 1995, p. 477). The use of citation counts to evaluate performance in this manner is based on the assumption, supported by empirical findings and by Merton’s theory of citations, that there is a correlation between citation counts and other performance measures (Shadish & Tolliver et al., 1995).

Citation analysis has grown more popular thanks to the advent of the Social Science Citation Index and other citation indexes (Kaplan, 1965). The index has made citation counting much easier and has also made the use of citations in studying the structure of knowledge less intrusive (Line, 1981). With citation indexes, citation studies have also merged with scientometrics, a new discipline that, given the expansion of fields of expert study, is deeply concerned with the structure of science (Line, 1981). The presence of citation indexes is expected to play “an important role in effecting the citation behavior of scientists in the future” (Kaplan, 1965, p. 179) and in fact has triggered greater study in the norms and behavior surrounding the practice of citing in scientific discourse. The citation index facilitates the analysis of citation patterns and the various reasons why citations are given (Kaplan, 1965). Generally, citation indexes were designed to “facilitate information retrieval and dissemination using source-reference connections” (Garfield, 1997, p. 2). Citation-based databases now allow researchers to “navigate the literature in unique ways” (Garfield, p. 2) and to locate papers “independent

of language, nomenclature, title words or author key words” (Garfield, p. 2). Some of the strategies being used in citation indexing are simple cited-reference searches, reference cycling, co-citation clustering, and bibliographic coupling, which is “the linking of related papers through shared references” (Garfield, p. 2). As with the development of text summarization from surface-level to entity- and discourse-level approaches, the development from simple citation counting to citation indexing—and the new discourse-searching techniques accompanying it—has created a convergence in searching, in which cited reference strategies can be combined with a variety of keyword, author, and institutional searches to achieve improved results in queried searches (Garfield, 1997).

Although the evolution of citation counting to citation indexing has itself pushed the development of new citation-based search techniques, questions about simple citation analysis have also forced the literature forward. “Strong objections have been raised against the use of citation counts” (Osareh, 1996, p. 221). Although citation analysis has been “widely accepted as a useful method for studying a wide range of topics in bibliometrics and the sociology of science” (Osareh, p. 222), a review of the literature shows that citation analysis has some problems and suffers from abnormalities (Osareh, 1996). Among the criticisms leveled at the practice of citation analysis are that authors have begun to gratuitously cite their own work in order to raise their citation counts and that the norms that rule citation counts in different disciplines vary so widely—based on citing behavior in each discipline—that conclusions drawn from studying a citation count in one field may not apply to another (Snyder & Bonzi, 1998). “With decisions such as journal retention in libraries or the research productivity of scholars made through the

technique of citation analysis, these are not minor concerns” (Snyder & Bonzi, 1998, p. 431).

2.3.2 Citation Motivation

More important criticism has been leveled against other assumptions of citation counts (Line, 1981; Liu, 1993). Several writers have questioned the validity of such counts, noting that not all cited works were read in connection with the research paper present, that not all papers read in the research for the paper were in the end cited, and that many citations were included for negative purposes (Line, 1981). Overall, citation counts “ignored the underlying purposes of why an author cited them” (Liu, 1993, p. 375). That is, the purpose of many citations is more to serve scientific, political, and personal goals than to describe the intellectual ancestry of a paper (Liu, 1993). In sum, researchers have found that the criterion of quality drawn from citation counts has been found to be always “imposed on the study by a preconceived hypothesis about what citations might measure” (Shadish & Tolliver et al., 1995, p. 478). What citation counts never did was ask scholars to construct the meaning and intent behind various kinds of citations in the work (Shadish & Tolliver et al., 1995). In a maturation of understanding directly parallel to the development from surface-level to entity- and discourse-level extraction in the field of text summarization, moving from citation count analysis to citation content and context analysis has also shifted the focus of citation analysis from the simple grading of a paper to exploring the relationship of the paper to the larger world of the discourse and literature. This alteration echoes changes in science as well: “The citation has functioned as the specific historic codifier of scientific communications in a period when the sciences have evolved from a transcendental project into a multifaceted

project that has become embedded in a myriad of contexts” (Leydesdorff & Wouters, 1999, p. 178).

Why does an author provide references to other work? That remains a central question in bibliometric research, and its answer remains difficult to obtain. Thus far, a grand theory of citation has proved to be elusive, because of the “constitutionally complex nature of modern citation behavior” (Cronin, 1998, p. 45). Would-be theory builders from sociologists to information scientists have been stymied by the complexity of the citation process (Cronin, 1998). More and more studies are examining the roles and motivations of citing (Wang & White, 1995). Unlike the assumptions behind citation counting, new research sees document use and citation as “cognitive behavior that is dynamic and situational” (Wang & White, 1995, p. 184). The dynamic quality of citing and searching involves the fact that “users may apply similar or different rules and criteria at different stages for selecting, reading, and citing the documents, and that the users’ personal knowledge of the topic, authors, journals, and information needs changes over time (Wang & White, 1995). By situational is meant that the users’ “tasks and goals affect the decisions on reading or citing a document, and these tasks or goals are modified over the course of a research project” (Wang & White, 1995, p. 184). This level of understanding has encouraged new research that seeks to clarify the practices or motivation of the citation process (Whit & Wang, 1997).

2.3.3 Citation Context

Current research in this area deduces the behavior of researchers by analyzing either the context or the content of the citation. This line of research assumes that a researcher’s behavior is revealed only “in the content or the context of the citations”

(White & Wang, 1997, p. 123). Studies in citation classification and other studies using citation motivation surveys have also revealed different theoretical beliefs about the nature and complexities of the citing process. Two major schools of thought have developed: the normative school and the micro-sociological school (Liu, 1993). The normative theory believes that people give credit to colleagues whose work they make use of by citing that work and that the citations therefore represent an explicit influence on their work. Based on the normative theory, simple citation analysis therefore believes that it evaluates the impact of scientists and the work, establishes cognitive pedigrees, and maps scientific networks and specialties (Liu, 1993). The normative interpretation also highlights rules, tacit or codified, that “govern the dispensing of credit within the scholarly communication system (Cronin, 1998, p. 47). The normative view, originally expounded by Merton, “provided a sociological interpretation of citation analysis” (Luukkonen, 1997, p. 27), insofar as recognition of the previous work of other scientists was seen as an institutional form of rewarding effort in the discipline.

Within the micro-sociological theory, however, ambiguities and vagueness surround the process of citing other documents (Liu, 1993). Under this paradigm, three basic approaches have been explored: classification of citation context, citation content analysis, and citer motivation survey. “Citation context studies have tried to devise a classification or taxonomy based on a text analysis in order to find out the interdocument relationship in the presence of reference citations” (Liu, 1993, p. 378). There are two principal approaches to the analysis of citation context: “classification of the types of functions of references in scholarly texts, and uses of the semantic context of the citing passage to characterize the cited work” (Small, 1982, p. 288). Citation context analysis

seeks to determine how “writers choose certain works to cite in specific academic disciplines, and how the writer integrates the literature into the construction of new knowledge claims” (Dong, 1996). Sociologists and psychologists argue that “people may have complex citation motives that have not yet been clearly understood” (Liu, 1993, p. 370). It is argued, in citation context analysis, that citing practice is associated with underlying effects and complex motivation (Liu, 1993). Elements studied in citation behavior include several questions: What reasons do the participants acknowledge in making citing decisions? What are the reasons for acknowledging some users and not others? Do uncited documents differ from cited documents in characteristics such as age or reputation of journal (White & Wang, 1997)? What is the purpose of contribution references, to persuade the scientific community, for example, that the work presented is valuable because it’s based on classic scholarship, to demonstrate the validity and significance of the work reports in the paper, to provide justification for the positions adopted in the current paper, to demonstrate the novelty of one’s results, or to indicate how the presented findings illuminate or solve problems that arise from the cited work? (Gilbert, 1977). Others felt that citing a standard or classic source “allowed them to avoid developing an explanation of a concept, but still provided a reference point for any reader who needs an explanation” (White & Wang, 1997, p. 145).

It has also been found that some citing occurs simply for sociological reasons linked to norms in the scientific community. Sometimes citations of work not directly related to the presented paper are included “because the author hopes that the referenced papers will be regarded as authoritative by the intended audience” (Gilbert, 1977, p. 115) and thus reflect well on the current work. References also constitute a means of

protecting individual property rights to ideas. But papers also include negational citations (those the author wishes to challenge) and perfunctory references, “which cite, almost as an aside, work which is not apparently strictly relevant to the author’s immediate concerns” (Gilbert, 1977, p. 114). Reasons behind these more complicated motivations for citing could be paying homage to pioneers, giving credit to related work, identifying aspects of the methodology in the paper, providing background reading, correcting one’s own work or that of another, substantiating claims made in the paper, alerting researchers to forthcoming work, or identifying the original publications in which a concept was discussed (Liu, 1993). All of these factors—biased citing, informational influences not cited, self-citing, and different types of citing—have forced citation analysis to proceed to citation context analysis in order to truly measure the relevance or salience of a citation in a paper (Liu, 1993).

One specialty area of citation behavior that is receiving more attention of late is the phenomenon of self-citation (Snyder & Bonzi, 1998). Because self-citation was one of the stigmatized behaviors that caused some researchers to question the validity of simple citation counts vis-à-vis their accuracy in indicating the importance of a document, reappraisals of this particular behavior from a sociological perspective are important. Early research indicated that there “are no significant differences in the motivations of citing between self-citations and citations to other works” (Snyder & Bonzi, 1998, p. 431). It has also been found that self-citing behavior is fairly similar across disciplines. Citation analysis has explored self-citation in various studies, some confirming that the “exposure given to self-citations does not differ significantly from that given to citations to others” (Snyder & Bonzi, 1998, p. 431). In the matter of journal self-citation, however,

Rousseau (1999) found that “high self-citing rate is ... an indicator of the isolation of the field covered by the journal” and that “the self-cited rate is relatively low for leading journals and high for peripheral ones” (p. 521). One finding determined that 50% of articles contain at least one self-citation and that full-length theoretical research articles in information science are more likely than any other kind of paper to contain self-citation (Dimitroff, 1995).

Further complicating the study of citation behavior and context is the fact that many young researchers do not know, and are not formally trained to know, when to cite (Garfield, 1996): “Students are not given good explicit guidelines for when to cite materials” (Garfield, 1996, p. 450). Moreover, a recent literature review “failed to turn up very many explicit normative guides for citation practices” (Garfield, p. 450). Many authors and inventors “are not aware of their intellectual debts, while others make naïve assumptions about what is common knowledge” (Garfield, p. 452). The fact that citation behavior is not formally taught is somewhat scandalous, given the recent “explosion of attention to fraud, plagiarism, and misconduct” (Garfield, p. 454). Finally, the requirements for citations “vary considerably from journal to journal and according to the types of materials involved” (Garfield, p. 456). Some citations occurred because “researchers were aware of referees and editors as filters for publication and modified their citing according to their perception of the gatekeeper expectations” (White & Wang, 1997, p. 145).

A subcategory of citation context analysis is the bibliometric study of the aging of various literatures, evidenced by changes in citation counts from journal to journal over time (McCain & Turner, 1989). From all of this contextual information gleaned from an

analysis of a citation, a classification and taxonomy is developed that assists users in better understanding the interdocument relationship of citations (Liu, 1993).

In sum, citation context analysis emphasizes the importance of “examining the contexts in which citations are made in order to understand the form of appraisal intended by the citing author or the specific concepts associated with the cited work” (Small, 1986, p. 85). Elements of the citation context have been categorized according to Moravcsik’s four areas: conceptual/operational, evolutionary/juxtapositional, organic/perfunctory, and confirmative/negational (McCain & Turner, 1989). It is not only the citation history of a paper that indicates usefulness, but also its context within a multiplicity of references and, finally, as studied in citation content analysis below, “the concept symbol that the key paper represented to later researchers” (McCain & Turner, 1989, p. 149).

By contrast, citation content analysis tries to “characterize the cited works by analyzing the semantic content of the citing papers” (Liu, 1993, p. 378). Once again, a classification system or taxonomy based on the semantic content of a cited document assists researchers in obtaining “a better understanding of relationships between citing and cited works” (Liu, 1993, p. 384). A new sociology of scientific knowledge in particular, “which paid attention to the technical content of science” (Luukkonen, 1997, p. 28), developed in the 1980s, displacing Mertonian normative views of science and citation in science. A break with the Mertonian tradition is a theory by Latour that elaborates the rhetorical function of citations. Emphasizing that the boundaries between fields are blurry, Latour argues that scientists transform earlier literature when they cite or even misquote earlier texts or cite them for reasons completely different from their authors’ intentions (Luukkonen, 1997). As a result, a statement undergoes a

transformation into a “black box,” wherein the concept is stylized or recorded (according to Latour) or obliterated by incorporation (according to Garfield). After time, a scientist’s work becomes “so generic to the field, so integrated into its body of knowledge that people neglect to cite it explicitly” (Luukkonen, 1997, p. 30). Such a theory “makes understandable many empirical findings in citation content and context studies” (Luukkonen, 1997, p. 31), because the idea that citations are rhetorical devices set in varying discourse practices explains many different motivations for citing (Luukkonen, 1997).

Although it is noted that Latourian views have largely been ignored by the bibliometric community because of differences between their approach and that of social constructivism or other new analytical approaches, the theory corresponds to other studies that see referencing as a “labeling process” and contend that citations should be interpreted semantically as “concept symbols” and classified accordingly (Small, 1978, p. 327). A theory of citation practice “must take account of the symbolic act of authors’ association of particular ideas with particular documents”; that is, analysis shows that “when a scientist cites, he or she is creating a link between a concept, procedure, or kind of data, and a document” (Small, 1978, p. 337). Thus, citations symbolize the conceptual association of scientific ideas as recognized by published research authors; and those authors, by citing papers, “make explicit linkages between their current research and prior work in the archive of scientific literature” (Garfield, 1994, p. 1). To analyze the conceptual or semantic content of a citation, relationships between content-related variables are examined. One study has determined the shape of a cluster pattern that outlined “the use that citing authors make of each others’ papers and whether the citation

history of a paper can be associated with the type of use or changes in use made by citing authors early or late in the paper's life span" (McCain & Turner, 1989, p. 129). Citation content analysis further explores a phenomenon about which little is known, how a scholarly work becomes an exemplar—that is, a highly cited article that becomes a classic in a field (Shadish & Tolliver et al., 1995). The fact that an author is a recognized authority in his field, that the work is thought to be a classic, that the citation is to an early work that represented a whole genre of studies, and that it generated novel research—all may affect the elaboration of that citation into a concept symbol (Shadish & Tolliver, et. al, 1995).

Both citation context analysis and citation content analysis are at the vanguard of a changing view of the citation process. The first, citation context analysis, "involves the recontextualizing of citation practices to accommodate the interplay of the political and the personal in the production and exploitation of symbolic capital" (Cronin, 1998, p. 50). The second, citation content analysis, "proposes a structurally informed analysis of the citation process, designed to bridge the existing interpretative divide and to articulate the relationship between private acts and public worlds" (Cronin, 1998, p. 50). Small (1986) believes that "perhaps bibliometric methods have not worked dramatically up to now because we have been content with statistical constructs and have not taken the crucial step of returning to the texts themselves to extract the specific concepts that are embodied in our bibliometric results" (p. 95). Both citation context analysis and citation content analysis seek to achieve this end, with a more practical importance as well, in that both means of analysis derive from texts' "detailed, descriptive indexing statements about papers or books cited" (Garfield, 1994, p. 1), which are increasingly being used in such

search engines as ISI Keyword Plus to enhance descriptors of a citation index and thus retrievability (Garfield, 1994). On the basis of these descriptors, citation-based search strategies have expanded to include not only simple cited-reference searches but also strategies such as co-citation and bibliographic coupling (Garfield, 1997).

2.4 Automating Citation Analysis for Summarization

Citation indexing has “made possible a further transformation of information retrieval” (Liu, 1993, p. 175). Starting, as noted, with simple citation counts, proceeding through adding new descriptors to keywords by means of citation classification and by deriving still other author-supplied keywords taken from citing or descriptive indexing statements made by authors in the texts from which citations are taken (through citation context and content analysis), such citation-based retrieval systems as ISI Keyword Plus have expanded the possibilities of retrieval through citation. A culmination of these areas of study is the development of abstract-like summaries of citation-based descriptions of papers based on the exploration of the structure of the overall citation network of a work and its citations. This emphasis is in keeping with contemporary views that now see citations as “a structurally embedded component of the primary communication process” (Cronin, 1998, p. 49), which deserve to be included in any epistemological critiques of science, for example. The fact that more and more research is being undertaken by networking, and that co-authorship is on the rise, indicates that collaboration and cooperation in broad “webs of citations and acknowledgements” (Cronin, 1998, p. 50) have become the frontier of research.

Some navigational aids, based on citation analysis, have been developed to help researchers find their way around the increasingly vast citation networks (Small, 1995).

Bibliographic coupling highlights the “relationship between two offspring who share a common parent,” that is, two articles cited in the same paper. Developed by Kessler, bibliographic coupling indicates the “degree of similarity of contents of the citing papers” (Liu, 1993, p. 3) based on the number of references the two documents share. The logic behind bibliographic coupling is that “papers will have a relationships to each other ... dependent upon the number of references they have in common” (Voos & Dagaev, 1976, p. 20). Co-citation analysis examines the “relationship between two parents who share a common offspring” (Small, 1995, p. 118). This concept was introduced by Small in 1973 to generate clusters of related papers: “The number of times two papers are cited together in subsequent literature determines the co-citation strength of the two papers” (Liu, 1993, p. 3). Again, the strength of a co-citation is “defined as the number of times two documents have been cited together” (Small & Griffith, 1974, p. 19). There is also longitudinal coupling, which is the “relationship between a grandparent and a grandchild” (Small, 1995, p. 118).

To navigate a citation network, maps identifying nodes of relationships (not significantly different from the text structure maps developed in corpus-based text summarization methods in computer science) are being created (Small, 1995). As in computer science corpus maps, the nodes that develop bushy linearity of relationships with other nodes are “strongly linked” (Small, 1995, p. 118) and point to positive directions in a citation network search. Citation indexes such as the Science Citation Index have greatly facilitated the study of citation networks and provide a tool for another possibility once citation networks are created: studying the lives of scientific papers (Small, 1984). Examining how a work is cited, and how it expands in the literature, goes

beyond mere citation counting to get at the true meaning of the paper in the literature (Small, 1984).

Another tool developed for citation network analysis using co-citation analysis and bibliographic coupling is ISI's Integration Citation File, which explores "the structure of the full citation network for scholarly literature" (Small, 1995, p. 121). These explorations show that documents that cite, and are in turn cited, do indeed play a key role in the structure of the citation network, verifying the notion "that scientists build on each other's work, and that ... citations represent an important communication link through which knowledge passes from one generation of researchers to the next" (Small, 1995, p. 121). Also, the fact that linkages are different for every discipline indicates that the structures of fields or disciplines can be revealed through the webs of citations and acknowledgments (Cronin, 1998).

Finally, citation networks appear to model accurately the network of information that researchers search, especially online. In reality, most scientists and researchers start searches "without developing a formalized search query" (Garfield, 1997, p. 10). Nonetheless, even piecemeal searches often produce useful results, which in turn help them find their way "into the relevant pathway" (Garfield, 1997, p. 10). As such, the literature should be viewed "as one gigantic topological network, with each published paper as a node" (Garfield, p. 10), and a researcher is "simply traversing the network" (Garfield, p. 10). Thus from the observation that citations are entities in a social system of research, citation indexing, co-citation analysis, and bibliographic coupling "have added new dimensions to the practice of citation" (Mitra, 1970, p. 118).

One application of co-citation analysis is to map out the structure of science by clustering highly co-cited documents (Osareh, 1996). Co-citation analysis studying similarity indicators between papers generates lists of new documents that are highly co-cited, and it can provide a list of core or more important papers from earlier materials for a specific field, “which may be a profile of the field and the basis of a selective dissemination of information” (Osareh, 1996, p. 217). Small & Griffith (1974) developed a computer-based technique to “identify clusters of highly interactive documents” (p. 17) and argue that those clusters represent highly active current structures of scientific specialties. Their technique, they believe, “opens the way to a systematic exploration of the entire specialty structure of science” (Small & Griffith, 1974, p. 17). Expanding citation context analysis to the study of co-citation contexts, these techniques can also provide an analysis of the “logical structure of shared knowledge in scientific specialties” (Small & Greenlee, 1980, p. 278). As these techniques generate lists of citations, co-citation analysis begins to output products that in effect summarize specific fields.

Co-citation analysis of fields is also related to other quantitative techniques recently introduced—such as cluster analysis, multidimensional scaling, factor analysis, and block modeling—to study the structure of disciplines. Many of these studies have used “large computerized files of citation data as input to the statistical techniques” (Small & Greenlee, 1980, p. 277). This new combination of large files, automation, and sophisticated manipulative procedures “has made it possible for the analyst to pose new questions” (Small & Greenlee, 1980, p. 278) about the structure of knowledge in fields.

Indeed, the large files generated by co-citation analysis and the output of a typical text summarization system merge in a method that uses co-citation clusters to generate

reviews or synopses of scientific fields, called “specialty narratives” (Small, 1986, p. 97). Small believes that the generation of summaries or synopses of specific scientific areas may be particularly amenable to computer-assisted intelligence (p. 97). The process begins with a co-citation cluster, a “bibliometrically defined network structure” (Small, 1986, p. 97), built on the hypothesis that it defines the structure of both a field of knowledge and the invisible college surrounding it. The resulting co-citation network of the specialty can indeed be used to generate a synthetic review statement (Small, 1986).

A combination of citation context analysis and co-citation clustering can also reveal a pattern of common use and an estimation of the degree of commonality—called the “percent uniformity”—which allows researchers to “identify the concept identified with each highly cited document” (Small & Greenlee, 1980, p. 299) and the relationships between them.

The model for thought processes involved in reviewing a field is like a walk through a co-citation network (Small, 1986). First, the researcher forms the co-citation cluster by selecting key ideas and documents and recognizing their pattern of co-relation. Next, the ideas are ordered to establish a “sequence of presentation consistent with their co-relation (a spanning tree)”. Third, the ideas must be expressed in words, “which entails a selection of a wording which is representative of existing formulations by others” (processes called citing passage analysis and consensus passage selection), and finally the process must create “transitional sentences” (Small, 1986, p. 100). To represent this information as a narrative, all nodes of a map must be traversed in an orderly fashion (Small, 1986). A search method called depth-first search is therefore applied to the tree structure representing the co-citation network. This method “traverses

each branch and twig of the tree as deeply as possible before proceeding to the next branch of the twig” (Small, 1986, p. 100). The mechanism ensures a continuity of ideas, creating a narrative that is a “combination of statements by several individuals from several sources, melded together by common usage” (Small, 1986, p. 100). Because the resulting narrative, constructed to walk through a co-citation network in order to understand the structure of a field, is an expression of the specialty field as a whole, the resulting expression is called the “specialty narrative” (Small, 1986, p. 98). The specialty narrative is a text that its creators believe summarizes an entire specialty area.

In short, citation analysis in its advanced forms converges with text summarization because of the observation that all computer-readable documents have embedded in them citing statements by authors, which, when analyzed in terms of context and content, can be summarized and used to provide computer-readable information about other documents (O’Connor, 1983). The use of citing statements has become economically feasible only with context and content analysis, both identifying the citing statement in the text. Experiments undertaken to use citing statements to augment retrieval in a full-text search, in effect as summaries, have shown that the use of citing statements can in fact improve retrieval because “they might use somewhat different terminology than that of the full text of the cited papers, and that terminology might better match some search formulations” (O’Connor, 1983, p. 365).

Little research on practical automated technology that exploits the more advanced levels of citation analysis has been conducted as yet. One such system is PRESRI, a prototype that relies on citation relationships to generate review articles automatically. PRESRI first identifies citing areas in a paper and the type of citing relationships in the

paper. It then uses this information for a citation-based topical clustering of papers, pointing as well to the citing areas they contain (Nanba & Kando et al., 2000). PRESRI citing areas are “defined as a succession of sentences that have a connection with the sentence that includes the citation in the paragraph” (Nanba & Kando et al., 2000, p. 4). Such connections are indicated by cue phrases, which are used for citing area extraction. The system then classifies research methods in a database automatically, using citation links and citation types (these heuristics are based on Weinstock’s notable categorization of the reasons for citations, also used in citation context analysis) (Nanba & Kando et al., 2000). Preliminary results of the PRESRI system indicate that a method based on bibliographic coupling is more effective than other search methods (Nanba & Kando et al., 2000).

Another advantage of automated citation indexing is that, unlike human abstraction, which has a delay between publication time and appearance in the database, “citation indexing computer-enhanced algorithmic systems permits the citation index to be virtually concurrent with the literature” (Garfield, 1997, p. 5). Citation indexing thus enables researchers to search both cited and citing papers (Leydesdorff & Wouters, 1999). Citation indexing is the result of new possibilities spawned by the advances in citation analysis, and it makes possible a further transformation of information retrieval (Leydesdorff & Wouters, 1999).

With advances in citation analysis and expansion of the field with citation content and context analysis, on through co-citation analysis to citation indexing and an understanding of citation networks, library and information science’s once-simple tool of citation counting has expanded into a bibliometric science that will greatly assist

researchers not only in searching vast literatures but also in developing summaries for navigating those literatures. A full theory of citation, “continuously contingent on the development of scientific communication” (Leydesdorff & Wouters, 1999, p. 178), is as yet unresolved, but the new awareness of the quality of a citation is itself an indicator of a new reflexivity in which analysts, aware of the increasing complexity of communication, must be both more focused and more reflexive about the status of their arguments in the citation network of the literature (Leydesdorff & Wouters, 1999).

2.5 Conclusion

Citation analysis has expanded from simple citation counts to approaches such as citation context analysis and citation content analysis, which see citation behavior as a varied sociologically motivated communicative channel as well as a process that transforms cited material into concept symbols that relate to their original semantic content in various ways (Garfield, 1997; Gilbert, 1977; Line, 1981; Liu, 1993; Sengupta, 1992; Shadish & Tolliver et al., 1995; Small, 1995; White & Wang, 1997). Further, studies of citation networks, fueled by techniques such as co-citation analysis and bibliographic coupling and made possible by citation indexing, have opened citation analysis to studies of the structure and interrelatedness of whole fields of research disciplines (Small, 1995). As such, automated co-citation analysis has the prospect of strongly assisting in literature searches and the creation of summaries—called specialty narratives—and abstracts for those systems.

Advances in citation analysis in library and information science parallel, and have borrowed much from, longer-term developments in the computer science field of text summarization and the ways in which it intersects with information retrieval and other

commercial purposes (Mani & Maybury, 1998). This science, too, has progressed from simple sentence extraction based on keywords to much more sophisticated methods of extraction based on entity-level (or corpus) and discourse-structure-level analysis. This evolution also culminates in efforts to examine whole fields through interrelationships of users and researchers. Finally, both evolutions—in library and information science and in computer science—have been forced by the exponential growth in online information and the understanding in various areas of the literature that too much literature now exists for an individual researcher to ever read without assistance from text summarization or citation analysis systems. Whereas these fields used to be characterized by comparisons between human and machine operations, automation is now essential for making sense—via maps of structures of literatures—of the vast amount of information available today.

3. Research Methods

3.1 Review of the Problem

The problem under study is to determine the feasibility of using similarity word metrics, such as frequency of words in common or cosine similarity metric that can be easily calculated by computer for selecting a citation context to serve as a document summary. When writing a scientific paper about the work of other researchers, many authors are in effect creating an abstract, or surrogate, of the original source document they cite.

Ideally, people are better text summarizers than computers, but given the enormous amount of information available on any research topic, human text summarization on a wide, systematic level is impractical, if not impossible. If we could measure and replicate the selection process that people use, citation selection could be optimized into a “best citation selection” method for use in a computer program. Such a computer program would automate the selection of an appropriate citation context to fill in as a document summary. This research will take three steps to investigate the feasibility of using word metrics to select a citation context that would act as a document surrogate: 1) gather information from subjects on which documents and citations they rank as similar, 2) gather the data about each of the documents and citations (word metrics, citation location, and type of citation context), 3) perform a statistical analysis of the data from steps 1 and 2.

3.1.1 Rater Data

Subjects read a document abstract and were then given several citation contexts to read. Next, they were asked to evaluate the citations based on a scale developed by Edmundson (1969) to identify and measure the conformity of the citation contexts to important parameters of the original document abstract. These parameters identify how well the citation context reflects the original document's subject matter, purpose, methods, findings, and implications. The survey instrument appears in Appendix A. Two questions beyond those used by Edmundson, on readability and understandability, were added to the survey. These survey parameters were added because there was concern by members of the committee that the higher the degree of understandability and/or readability, the more likely a subject would rate a citation context as an acceptable summary.

3.1.2 Document Data

Several types of document data were collected: frequency counts of the words in common between abstract and citation context; similarity metrics, using several different weighting schemes, between abstracts and citations; category of citation; and location of the citation in the citing document. Salton's cosine similarity metric, with several different weighting methods, (Salton, 1975 and 1983) was used to compare citation contexts with the original document's abstract. This measure indicates how alike the abstract and the citation context are. The category of the citation (Small, 1978) and citation location within the document were also collected from the citing documents. This additional data was used to determine whether those factors mediate or improve the

potential of the similarity metric to predict which citations are the most representative of their source documents.

3.1.3 Data Analysis

Regression analysis was used to determine whether a relationship exists between the dependent variable (the user data) and the independent variable (the document data). Regression techniques were used to determine whether the category of citation or the citation location within the citing document influences the selection of citation.

3.2 Subject Survey Data

3.2.1 Document and Citation Sample

The sample of documents was originally selected for an in-class, nongraded exercise that was voluntarily handed in by students for an ISYS 300 class in Information Retrieval Systems in June 2002. The objective of the assignment was to underscore the difficulty in determining the nature or “aboutness” of a document and the implications for successful information retrieval. The electronic source documents and corresponding citation context extracts were drawn from the CiteSeer (<http://CiteSeer.nj.nec.com/cs>) scientific digital library database. CiteSeer is a digital library built from scientific papers in computer science that are posted on the World Wide Web. The database features autonomous citation indexing, citation statistics, powerful searches, and query-sensitive summaries, among many other useful features. CiteSeer provides valuable information about articles and authors with tools such as citation graph analysis and citation context. CiteSeer algorithmically extracts the citation sentence from a document in addition to the context, usually half of the sentence before and after the citation sentence. Tools on the system also provide information on the common usage and on the authority of a given

source. These features allow this study to quickly identify source documents, citations with context, and the documents that contain the citations.

The search terms used for document selection were *text summarization*, *text analysis*, *natural language processing*, and *Web [AND] text analysis*. Documents and citations were not selected randomly. CiteSeer can list documents in rank order based on the number of citation contexts in the database. For a document to be considered for selection, there had to be at least five citations. Documents with the highest number of citations were selected first. A document was rejected for any or all of the following reasons: There was no abstract in ASCII format; the material was a book, dissertation, or section of a book; or the material was heavily math based. Because of these restrictions, the originally selected 60 documents were reduced to 45 that met the criteria.

The citation contexts were also selected from the CiteSeer database. Originally, 410 citations were selected, ranging from 5 to 14 citations per document. The average was nine citation contexts per document. Citation contexts were also scrutinized to make sure they met the same criteria as the documents above, but with the added restriction of readability. Character recognition software has a difficult time with some characters and symbols. In some instances, the citation contexts were confusing and unreadable and were therefore rejected on inspection. After the elimination process, 249 citation context extracts were left, an average of 5.5 per source technical document.

Variations in the selected sample were due to the restrictions listed above. Appendix D shows the distribution of citations per paper. Figure 3-1, which shows the distribution of citations per document, indicates that 79% of documents had between four and six citations each. An examination of Figure 3-1 shows a range of 1 to 11 citations

per abstract. Although some of the citations fell below the initial target of five and some were much higher than five, they were kept to see whether the results were affected in any way. Document titles and abstracts are listed in Appendix B. Citation extracts are listed in Appendix C.

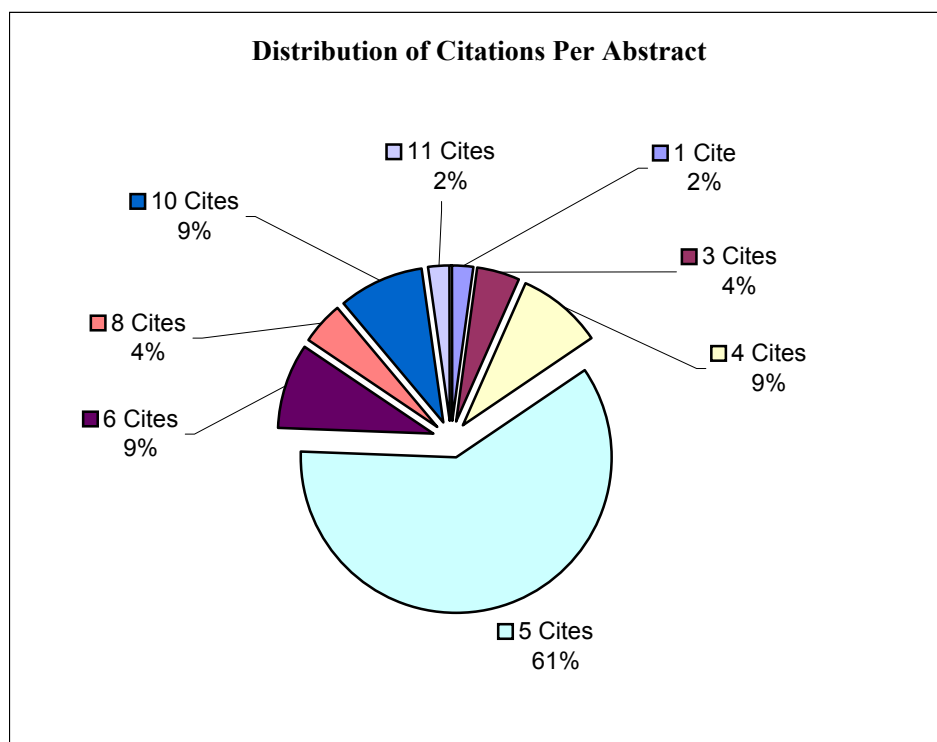


Figure 3-1. Distribution of Citations per Abstract

3.2.2 Raters

The performance of citation contexts as document summaries was judged by two groups of raters. The first group comprised upper-class undergraduate juniors and seniors with only general knowledge of information systems and cursory knowledge in the language processing subject area.

The student or non-expert group had 71 raters, out of a possible 86, who were upper-class undergraduate and master's-level students from two classes of ISYS 300, Text Information Retrieval Systems. This study uses an in-class nongraded exercise that students handed in voluntarily. No demographic or identifying data was requested of the survey raters. The exercise was a study in the difficulties of text retrieval and text summarization. This study utilizes the in-class exercise for the user data portion of the study. By necessity, a class exercise had to be short, so students were given two abstracts and approximately 10 citations each. That distribution allowed for the entire data set to be rated approximately four times.

The second group, referred to as the expert group, comprised four Ph.D. candidates, two in information science and two in natural language processing. This group countered some of the possible errors and misunderstanding of the non-expert group. The group offered an opportunity to determine how robust the similarity metric is compared with the judgment of experts familiar with the literature. Participants were selected based on a working knowledge of the literature in text summarization and natural language processing.

Both groups evaluated the same abstracts and citations by answering seven questions on a five-point Likert scale. To identify the “best” citation surrogates for a given document, raters were shown a sample citation extract with attached survey

questions and were instructed in the assignment procedures. The raters were permitted to ask questions about survey procedures and purpose but were not given information about the expectations of the study.

3.2.3 Survey Instrument

Raters were given the abstract of a source document and the referring citation extracts, each with an attached survey questionnaire. An example of a typical survey questionnaire can be found in Appendix A. The raters evaluated the extracted citation text according to how well it represented the meaning of the abstract. Raters indicated the performance of each citation extract as it compared to the original abstract, in each of the following categories:

- Subject matter: How well does the citation represent the general subject area of the abstract?
- Purpose: Does the citation represent or identify the abstract's main purpose?
Possible purposes include presenting original research findings, discussing theory, and surveying the work of others.
- Methods: Does the citation indicate the methods used in the research?
- Conclusions or findings: Does the citation indicate any research findings or conclusions?
- Generalization or implications: Does the citation indicate the significance of the research and its influence on the larger technical problem or theory?
- Readability: How readable is the citation?
- Understandability: How understandable is the citation?

The citation extract's performance in each category was evaluated according to a 5-point scale used in previous studies of automatic extraction techniques (Edmundson, 1969). The scale points are as follows:

- Not at all (0%)
- Slightly (25%)
- Adequately (50%)
- Very Well (75%)
- Completely (100%)

To determine which of the citation extractions were most representative of the source document, the student ratings for the seven categories of response (subject matter, purpose, methods, conclusions, generalizations, readability, and understandability) were averaged across subjects and across variables. That procedure resulted in one averaged number for each citation.

3.3 Document Data

3.3.1 Similarity Metric

To rank citation contexts for comparison with those chosen by human raters, similarity metrics were used. Each citation context was paired with the abstract and similarity scores were calculated. For this study word frequency counts and a cosine similarity word metric with four different weighting factors were used. Documents or abstracts, are used interchangeably, and can be represented by term vectors of the form:

$$D = (d_i, d_j, \dots, d_n) \quad (1)$$

where each d_i represents the frequency count of a content word in document D . The citation context would be represented in the vector form and formulated as follows:

$$C = (c_a, c_i, \dots, c_m) \quad (2)$$

where c_i represents the frequency of a word i contained within citation C . Equations 1 and 2 are both vectors in an n -dimensional vector space.

The length of D or C is given by extension of Pythagoras' theorem:

$$|D|^2 = d_i^2 + d_j^2 + \dots d_n^2 \quad (3)$$

$$|C|^2 = c_i^2 + c_j^2 + \dots c_n^2 \quad (4)$$

Given that D_i and C_i are vectors, the inner product or dot product is given by the following:

$$D \bullet C = d_i c_i + d_i c_j + d_j c_i + d_j c_j + \dots + d_n c_n \quad (5)$$

The cosine of the angle between the vectors D_i and C_i can be represented as follows:

$$\text{Similarity}(D, C) = \cos(\theta) = \frac{D \bullet C}{|D| |C|} \quad (6)$$

A more formal expression of the term vectors 1 and 2 is rendered by including weight assignments to provide finer points to the terms. If w_{dk} and w_{ck} are the weights of term d_k in document D and citation C , can be written as shown here:

$$D = (d_0w_{d0}; d_1w_{d1}; \dots; d_kw_{dk}) \text{ and } C = (c_0w_{c0}; c_1w_{c1}; \dots; c_kw_{ck}) \quad (7)$$

The values of w_{dt} and w_{ct} are assumed to be equal to 0 when a term is not assigned to either the document D or the citation C . In this research, as is frequent practice all the words have been stemmed. The Porter algorithm was used for this study.

The following Equation 8 shows the vector product for the similarity metric. When the weights are restricted to 0 or 1 the equation measures the terms that are in common to both document D and citation C :

$$Similarity(D, C) = \sum_{i=1}^n w_{di} * w_{ci} \quad (8)$$

where n is the total number of unique words in the citation and document; in this research *document* and *abstract* are used interchangeably.

The similarity (SIM(D,C)) metric counts the number of words in common and then normalizes the result, so that large documents do not skew the resulting metric (Salton & McGill, 1983). The equation is as follows:

$$SIM(D, C) = \frac{\sum_{i=1}^n w_{di} * w_{ci}}{\sqrt{\sum_{i=1}^n (w_{di})^2 + \sum_{i=1}^n (w_{ci})^2}} \quad (9)$$

where w_{id} is the weight of term i in document D and n is the number of unique terms, also known as types, in the two documents D and C . If the two documents have no words in common, $SIM(D,C) = 0$. Comparing a document with itself would result in a $SIM(D,C) = 1$. Salton and Buckley (1987) have done work to determine the optimal weighting factors. Their research has indicated that the best weighting factor for documents is shown in Equation 10 and that the best weighting factor for citations is shown in Equation 11.

$$\text{Document Term Weights (} w_{id} \text{)} = \frac{tf_{id} * \log_{10}(N / n_i)}{\sqrt{\sum |tf_{id} * \log_{10}(N / n_i)|^2}} \quad (10)$$

tf_{id} is called the Term Frequency Component. It is a raw term frequency count – the number of times a term i occurs in a document d .

$\log_{10}(N/n_i)$ is called the Collection Frequency component and is multiplied with the Term Frequency Component. N is the total number of documents in a collection and n_i is the number of documents in which the term i occurs. The denominator is the Normalization component, represented by the Euclidian vector length.

$$\text{Citation Term Weights}(c_{ic}) = |0.5 + (0.5 \, tf_{ic} / \max \, tf_i)| * \log_{10}(N/n_i) \quad (11)$$

This is an augmented normalized term frequency for the citation context, where the tf_{ic} is term frequency for a given term i in a citation context c and $\max \, tf_i$ is the maximum term frequency in the vector. This is further normalized to lie between .5 and 1.0. This is multiplied by the Collection Frequency Component $\log_{10}(N/n_i)$.

3.3.2 Citation Categories

Henry Small (1982) examined eight major citation classification schemes for scholarly text to develop a matrix of simpler, more useful parameters with which to categorize the use and meaning of citation text. Table 3-1 shows how Small reconciled several categorization schemes, collapsing the overlapping features and clarifying the unique dimensions among them into 14 categories, which are listed on the right side of the table. The left side of the table is a simplified or compressed version of the categories. The left side categories (Refute, Support, Noted, Review, and Applied) were created to simplify the categorization of citations. Five categories were much easier for the raters to identify and select the cue phrase. In addition, fewer categories will make it easier to see a statistical pattern more easily given the relatively small sample size of citations for this study.

Table 3-1. Modified Categories of Small's Categories of Citations

| Generalized Category | Small's Categories of Citations |
|----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| Refute | Negation: Disagrees with opinion or findings of original document; critical |
| Support | Partial Negation: Disapproves or questions; mixed opinion |
| | Replaced: Offers new interpretation or explanation; negational |
| Noted | Confirmation: Approves, verifies, or substantiates a claim |
| | Support: Legitimizes or substantiates statement or assumption |
| Review | Background Info: Forms part of relevant literature; historical; further reading |
| Applied | Bibliographic Info |
| | Review or Compared: Adds affirmative information; data used for comparative purposes |
| Applied | Distinguished: Acts in a juxtapositional manner |
| | Applied: Acts in an operational; uses concepts, definitions, interpretations, data, material, methods, and/or equations or methodology |
| | Improved/Modified: Extends or modifies a theory; offers a specific point of departure; acknowledges pioneering work |
| | Changed the Precision |
| | Future Research Implications |

This study uses the five categories in the left column for two reasons: to simplify the categorization of citations and to reduce the number of categories so as to simplify the statistical analysis. Two experts — librarians with expertise in cataloging—classified each citation extract into one of the five categories. In addition, they extracted the cue words that identify each citation category. The categorical information was used in the statistical analysis part of the study. The cue phrases are what allow the automatic classification of the citation in a computerized system. The cue words and phrases will be collated into a table and may be used to automate a computer system in the future if categorical information is useful in the selection of an appropriate citation.

3.3.3 Citation Location

A citation may contribute more or less to overall performance depending on its relative location in the referring document. The researcher determined and recorded the position of the citations. All of the documents that contained the citations were located. Each citation was found in the document, and the location of the citation recorded. Each location was recorded as a dummy variable, for example 1 = introduction, 2 = review of the literature, 3 = discussion, 4 = methodology, and 5 = conclusion and so on.

Refer to Appendix F for explanation of how the survey and document data was stored and manipulated. Appendix F also contains information on the database construction.

3.4 Variables and Research Questions

Table 3-2 summarizes the variables that will be used in this study. The ratings variables (Subject Matter, Purpose, Methods, Conclusion/Findings, Generalizations, Understandability, and Readability) are from the subject surveys. These variables were considered the independent variables. The next variable in the table is citation categories (Refute, Noted, Support, Reviewed, and Applied). This variable is a dependent variable. The next variable is citation location and is also a dependent variable. Both Citation Category and Location are “enhancement” variables, meaning that they are not expected to stand alone as predictors of user responses on similarity of documents, but rather to add to or enhance the predictability of the similarity metrics. The last variable in the table is the similarity metric. This is a dependent variable and will have a value between 0 and

1. This variable will be compared to user responses on the similarity of an abstract and citation.

Table 3-2. Variables in the Study

| Variables in the Study | Data Type | Response | Purpose, Contribution and/or Expected Direction of Influence | IV or DV |
|----------------------------------------------------------------------------------------------------------------------------------|------------------|----------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|
| Ratings: Subject Matter, Purpose, Methods, Conclusions/Findings, Generalizations or Implications, Understandability, Readability | Numeric | Average or Sum 1-5 | To establish a measurement/rationale for subject selection of a citation. The higher the user scores, the better the citation can act as a document summary. The higher the subject rating, the higher the similarity metric should be. | IV |
| Citation Categories | Nominal | 1 Refute 2 Noted 3 Support 4 Reviewed 5 Applied | This variable is intended to refine the similarity metric. It is expected that some citation categories will not make good document summaries. It is expected that “2 Support” category will most likely improve citation selection, and “1 Refute” is expected to not produce any viable citations. | DV |
| Citation Location: Introduction, Review of the Literature, Methodology, Discussion, Conclusion | Nominal | 1. Intro 2. R of L 3. Method 4. Discussion 5. Conclusion | This variable is intended to refine the selection of an appropriate citation that is an adequate document summary. It is expected that taking citations from different areas of the document may have a positive or negative affect on the selection of a citation as a document summary. | DV |
| Similarity Metric | Numeric | 0–1 | Metric to identify similarity between citation and the original document or abstract. The higher the similarity rating, the more likely it is that the citation can stand in as a summary for the original document. | DV |

3.4.1 Research Questions

All of the variable data has been collected to answer the primary question: Do similarity metrics identify the same citations and abstracts that people would identify as similar? The data collected can be grouped in two major categories: 1) subject survey scores on document similarity and document measurements, and 2) similarity metrics, citation location, and citation category. The similarity metrics are evaluated alone and then in conjunction with the citation location variable and citation classification variable to determine whether a workable automated citation extraction model can be derived using a combination of those measures that compares in performance to those citations chosen by human raters.

To evaluate abstracts and citations to determine the reasonable surrogate for a document, the following research questions were investigated:

1. Are document surrogates selected by word metrics the same as document surrogates chosen by subject raters? (Can word metrics be used to automate the extraction process of citations to serve as reasonable surrogates for original source documents?)
 - 1.1. Is there a difference between subject matter experts and non-experts in a) how they responded on the survey b) how their responses correlated with similarity metrics?
 - 1.2. Do different similarity metrics correlate differently with human subjects' responses to document similarity? The similarity metrics used for this study, in order of complexity are:
 - frequency counts of words in common
 - simple cosine (Equation 6 in Section 3.3).

$$\begin{aligned}
& \bullet \sum w_{id} w_{ic} \text{ where } w_{id} = \frac{tf_{id} * \log_{10}(N / n_i)}{\sqrt{\sum |tf_{id} * \log_{10}(N / n_i)|^2}} \\
& \text{and } w_{ic} = |0.5 + (0.5 \, tf_{ic} / \max \, tf_i)| * \log_{10}(N/n_i) \\
& \bullet \sum w_{id} w_{ic} \text{ where } w_{id} = \frac{tf_{id} * \log_{10}(N - n_i / n_i)}{\sqrt{\sum |tf_{id} * \log_{10}(N - n_i / n_i)|^2}} \\
& \text{and } w_{ic} = |0.5 + (0.5 \, tf_{ic} / \max \, tf_i)| * \log_{10}(N-n_i/n_i)
\end{aligned}$$

2. Are there any easily computer-calculated document metrics that can increase the explanatory power of the similarity metric?
 - 2.1. Can the location of citation add additional explanatory power to the similarity metric, so that the combination explains more variance?
 - 2.2. Can the category of citation add additional explanatory power to the similarity metric, so that the combination explains more variance?
 - 2.3. Can the citation category and location combined add additional explanatory power to the similarity metric, so that the three variables explain more variance.

4. Results

4.1 Introduction

The purpose of this study is to examine the feasibility of using a computer-calculated word metric to select citation context as a document summary compared to human selection. The research questions tested in the study are:

- Is there agreement between the similarity metric and human subjects about how alike an abstract and a citation context extraction are?
- Does the similarity metric correlate better with survey takers who are content experts or those who are non-experts?
- Are any of the various similarity metrics more efficacious than the others in terms of agreeing with the selection of similar documents compared to the choices of human subjects?
- Does the location of a citation reduce the variance between the similarity metric and subjects' ratings?
- Does the type of citation reduce the variance between the similarity metric and subjects' ratings?
- Do citation categories and citation locations together reduce the variance between the similarity metric and subject's ratings?

This chapter presents the results from the data analysis. The chapter is divided into seven sections. Section 4.2 presents the descriptive statistics and related discussion. Section 4.3 presents the document metrics data, including descriptions of document and

citation context, citation location, and citation categories. Section 4.4 presents results from the correlation analyses. Section 4.5 presents results from comparisons of means. Section 4.6 discusses regression analyses, and a brief summary concludes the chapter.

4.2 Descriptive Statistics

This analysis will indicate whether there is anything interesting in the data that might indicate any anomalies that might require additional statistical analysis. There was concern that the differences of data collection between the expert and non-expert groups could introduce data discrepancies. All the variables were tested for normality and all were normal, with a few minor and insignificant variations that are indicated in the following explanations of the data.

Descriptive statistics were computed for all variables and for the three groups — experts, non-experts, and the combined group — of respondents. First, descriptive statistics for the experts were computed. These appear in Table 4-1. The most common score for the first variable, subject matter, was a 3.0, with a mean of 3.42 and a median of 3.50. The standard deviation of 0.65 indicates that there is little variation in the responses.

Table 4-1. Means, Median, Mode, and Standard Deviation of Expert Ratings

| Variable | Mean | Median | Mode | Standard Deviation |
|-------------------|-------------|---------------|-------------|---------------------------|
| Subject Matter | 3.42 | 3.50 | 3.00 | 0.65 |
| Purpose | 3.32 | 3.25 | 3.00 | 0.82 |
| Method | 2.27 | 2.33 | 2.00 | 0.72 |
| Findings | 2.72 | 2.75 | 3.00 | 0.66 |
| Implications | 2.78 | 2.75 | 3.00 | 0.47 |
| Readability | 4.17 | 4.25 | 4.00 | 0.40 |
| Understandability | 3.88 | 4.00 | 4.00 | 0.61 |

N=249

The second variable, purpose, also had a median of 3.0, indicating that this was the most common response. The average score for purpose was 3.32, and the standard deviation of 0.82 indicates that there is slightly more variation in responses to this variable than the first. Two other variables, findings and implications, also had modes of 3.0 for each and averages of 2.72 and 2.78 respectively.

The readability and understandability variables both have higher average scores than averages for the other variables: 4.17 and 3.88, respectively. The most common score for both was 4.0, and both variables also have a narrow spectrum of variation, showing that the subjects were in close agreement with these variables.

Expert means are higher for the variables subject matter, purpose, readability, and understanding and slightly lower for method, findings, and implications. The mode shows that subject matter, purpose, findings, and implications are all at the middle value of 3; readability and understandability have a value of 4, and the variable method has the lowest value, at 2. All of the variables are normal except for a strong negative skew for readability and understandability.

Descriptive statistics were then computed for all variables for the non-expert raters. These are depicted in Table 4-2. Starting with the variable subject matter, the most common score for the non-experts was 2.0, compared to a mode of 3.0 among the experts. The average score for subject matter for non-experts was 2.87, which was also lower than the average for the experts, 3.42. For the second variable, purpose, the mode for non-experts was 3.0, as was that for the experts. Again the average for the non-experts, 2.70, was lower than for the expert's average of 3.32. The average score for the variable method was higher among non-experts, 2.54, than it was for experts, 2.27. For the variable findings, the average non-expert score was 2.50, which was lower than that of the experts, 2.72, and the most common score for non-experts was 2.0, which is also lower than for the experts, 3.0.

Table 4-2. Means, Median, Mode, and Standard Deviation of Non-expert Ratings

| Variable | Mean | Median | Mode | Standard Deviation |
|-------------------|-------------|---------------|-------------|-------------------------------|
| Subject Matter | 2.87 | 2.83 | 2.00 | 0.67 |
| Purpose | 2.70 | 2.67 | 3.00 | 0.65 |
| Method | 2.54 | 2.50 | 3.00 | 0.63 |
| Findings | 2.50 | 2.57 | 2.00 | 0.64 |
| Implications | 2.46 | 2.43 | 2.00 | 0.52 |
| Readability | 3.22 | 3.29 | 3.00 | 0.60 |
| Understandability | 3.14 | 3.20 | 3.00 | 0.61 |

N=249

The fifth variable, implications, had an average score of 2.46 among non-experts, which was slightly lower than that for experts, 2.78. The last two variables, readability (M = 3.22) and understandability (M = 3.14), both had lower averages than those recorded for the experts, 4.17 for readability and 3.88 for understandability. Standard

deviations are similar between the two groups. All the variables are approximately normal.

The third group of descriptive statistics computed was for the entire group of respondents, both non-experts and experts together. These descriptives are depicted in Table 4-3. The mean of subject matter was 3.09, with a standard deviation of 0.57, indicating little variation. The average score for the variable purpose was 2.95 (SD = 0.56). The mean of method was 2.44 (SD = 0.45), and the mean of findings was 2.59 (SD = 0.49). The mean of implications is 2.60, with the standard deviation indicating little variation in scores (SD = 0.38). The mean of readability is 3.59 (0.45), and the mean of understandability is 3.45 (SD = 0.52). All variables are normal, except a that there is a slight negative skew for understandability and readability.

Table 4-3. Means, Median, Mode, and Standard Deviation of All Participants' Ratings

| Variable | Mean | Median | Mode | Standard Deviation |
|-------------------|-------------|---------------|-------------|---------------------------|
| Subject Matter | 3.09 | 3.13 | 3.00 | 0.57 |
| Purpose | 2.95 | 2.91 | 3.00 | 0.56 |
| Method | 2.44 | 2.40 | 2.00 | 0.54 |
| Findings | 2.59 | 2.56 | 2.50 | 0.49 |
| Implications | 2.60 | 2.60 | 2.33 | 0.38 |
| Readability | 3.59 | 3.67 | 4.00 | 0.45 |
| Understandability | 3.45 | 3.44 | 3.00 | 0.52 |

N=249

Comparing experts and non-experts, it seems that there is a 0.5 to 1.0 difference in the means for most variables. There are two possible reasons for this: First, they may be interpreting the variables differently when comparing abstract and citation context, or second, the two groups may be interpreting the Likert scale differently. The next section

will test the means to see if our observation of the differences of the means has any significance.

4.3 Comparisons of Expert and Non-expert Means

Matched-pair t-tests were conducted to ascertain whether there were significant differences between average scores of experts and non-experts on all the study variables. Significant differences were found between all the variables at the .005 level. Results appear in Table 4-4.

Table 4-4. T-Tests Between Non-experts and Experts on Seven Variables

| Dependent Variable | DF | Mean | T-value* | P-value |
|---------------------------|-----------|--------------------|-----------------|----------------|
| | | Difference* | | |
| Subject Matter | 44 | 0.55 | 8.36 | <0.0001 |
| Purpose | 44 | 0.63 | 8.20 | <0.0001 |
| Method | 44 | -0.24 | -3.41 | 0.0014 |
| Findings | 44 | 0.24 | 3.38 | 0.0015 |
| Implications | 44 | 0.36 | 5.91 | <0.0001 |
| Readability | 44 | 0.96 | 17.19 | <0.0001 |
| Understandability | 44 | 0.74 | 11.64 | <0.0001 |

* Test statistic vs. the null hypothesis that the true mean difference is 0.

Starting with the variable subject matter, the mean of experts ($M = 3.42$) was significantly higher than the mean recorded for non-experts ($M = 2.87$; $t = 8.36$). This may indicate that experts and the non-experts had differences in the interpretation of the rating scale. The average score for purpose was also significantly higher among experts than non-experts ($M = 3.32$ vs. 2.70 ; $t = 8.2$).

For the variable method, non-experts scored significantly higher on average ($M = 2.54$) than the experts ($M = 2.27$; $t = -3.41$). This indicates that the non-experts thought

the citation more accurately reflected the methods used in the research than the experts did.

Average scores for the variable findings were significantly lower for non-experts ($M = 2.50$) than they were for experts ($M = 2.72$; $t = 3.38$). This indicates that the non-experts did not think the citation accurately reflected the findings of the study as well as did the experts.

Implication scores were also significantly lower among non-experts ($M = 2.46$) than among experts ($M = 2.78$; $t = 5.91$). This indicates that non-experts were less inclined to think that the citation accurately reflected the significance of the study.

Readability scores were also significantly higher among experts ($M = 4.17$) than among non-experts ($M = 3.22$; $t = 17.19$). This difference was the largest difference found in the analysis and indicates that experts thought the citations were far more readable than the non-experts indicated.

Finally, the difference between the means for understandability indicate that experts scored significantly higher than did non-experts ($M = 3.88$ vs. 3.14 ; $t = 11.64$). This indicates that experts thought the citation was more understandable than the non-experts did.

The results show that the two groups answered the same questions substantially differently. Looking back at the descriptive data in Tables 4-1 and 4-2 and the Mean Difference column in Table 4-4, we can see that the non-experts scored all questions lower by approximately half a point to one point, out of a five-point Likert scale. This could possibly indicate that on average, experts thought that the citation contexts

represented the abstract better than the non-experts thought it did. On the other hand, it could mean that experts and non-experts interpreted or valued the Likert scale differently.

Another perspective, on the two groups, is to present rank order of the averaged raw scores raters gave to citation contexts associated with each abstract. Table 4.5 identifies the abstracts in the first column and the abstract in the second column. So we can identify that the first abstract (1) has five citations associated with it, one to five. The second abstract has citations six to ten associated with it and so one. The table only identified abstracts one to four. Refer to Appendix G for a complete listing of all citation contexts ratings listed by each abstract. Each pairing of abstract and citation context shows the frequency count and the Cosine in column three and four respectively. Frequency Count column identifies the number of terms in common between the abstract and the citation context. Each term is only counted once even though it may occur multiple times between the abstract and citation context. Looking at the table the citations are not in sequential order, they are in ranked order by Frequency Count (and Cosine) in descending order, for example (1,1), (2,6), (3,14) and (4,19) represent the abstract and citation context pairings that have the highest Frequency Count and Cosine rankings. The highest rankings are highlighted in grey. Columns five, six and seven are the averaged ratings of Experts, Non-Experts and the Combined results of both groups of raters. The highest rankings are shown here in grey also. Looking at the first abstract, we can see that citation context 1 had the highest Frequency Count and Cosine. The Expert group selected citation context number four as the most similar to the abstract 1. The Non-experts picked citation context number three and the combined data also designated citation context number three as the most similar to abstract one. This shows that the

Frequency Count, Experts and Non-experts all selected different citations contexts as similar to abstract number 1. Abstract two shows the Frequency Count and Cosine metric selected citation context number six, but Experts and Non-experts selected citation context number 7 as the most similar to abstract number 2. Abstract three shows that Frequency Count, Cosine, Experts and the Combined results all selected citations context number 14 as most similar. The Non-experts selected citation context number fifteen as the most similar to abstract two.

Table 4-5. Rank Ordering of Scores

| Abstract | Citation | Freq Count | Cosine | Combine | Experts | Non-Experts |
|-----------------|-----------------|-------------------|---------------|----------------|----------------|--------------------|
| 1 | 1 | 16 | 0.383131 | 2.5 | 2.68 | 2.38 |
| 1 | 3 | 13 | 0.345349 | 3.49 | 3.29 | 3.62 |
| 1 | 2 | 10 | 0.302891 | 2.89 | 3.32 | 2.6 |
| 1 | 4 | 10 | 0.302891 | 3.46 | 3.43 | 3.48 |
| 1 | 5 | 9 | 0.287348 | 2.73 | 3.14 | 2.45 |
| 2 | 6 | 10 | 0.418854 | 2.47 | 3.14 | 2.02 |
| 2 | 9 | 10 | 0.418854 | 2.17 | 2.61 | 1.88 |
| 2 | 8 | 8 | 0.374634 | 3.14 | 3.61 | 2.83 |
| 2 | 7 | 7 | 0.350438 | 3.29 | 3.68 | 3.02 |
| 2 | 10 | 5 | 0.296174 | 2.69 | 2.71 | 2.67 |
| 3 | 14 | 9 | 0.323498 | 3.1 | 3.61 | 2.82 |
| 3 | 11 | 8 | 0.304997 | 2.55 | 3.36 | 2.08 |
| 3 | 12 | 4 | 0.215666 | 2.51 | 3.11 | 2.16 |
| 3 | 15 | 4 | 0.215666 | 3.09 | 3.25 | 3 |
| 3 | 13 | 3 | 0.186772 | 2.39 | 3.07 | 2 |
| 4 | 19 | 10 | 0.5 | 2.88 | 3.29 | 2.65 |
| 4 | 18 | 7 | 0.41833 | 3.43 | 3.54 | 3.37 |
| 4 | 16 | 5 | 0.353553 | 2.52 | 3.14 | 2.16 |
| 4 | 20 | 4 | 0.316228 | 2.52 | 2.79 | 2.37 |
| 4 | 17 | 2 | 0.223607 | 2.39 | 2.61 | 2.27 |

Table 4-6 is a summation of the rank orders of citation contexts. All averaged scores were ordered by highest frequency count to lowest (cosine numbers matched

frequency counts) for each citation context. A rank order of 1 means that the frequency count, cosine, and the rater's selection all matched. A ranking of 2 to 5 means the group selection was that many citations removed from the highest value. Values that tied were counted as the same rank. For example, there may have been a frequency count of 12 for two citation contexts with a given abstract. If they were both the highest number of words in common, both citations would have received a ranking of 1, because the similarity metric or frequency could have no way of making a distinction between the two citation contexts.

Looking at Table 4-6 we can see that the combined groups have a definite trend; the more words in common between two documents the more likely they are to be selected by the raters, and the fewer words in common the less likely the citation contexts are to be selected. Selecting the citation context with the highest number of words in common yields, on average, a 42% agreement with what the raters selected. Rank 1 and 2 accounted for 75% of all the selections made by the survey raters.

Table 4-6. Summary of Citation Rank Order of Averaged Raw Scores for Experts, Non-experts, and Combined Group

| | Rank Order | | | | |
|-------------------|-------------------|----------|----------|----------|----------|
| | 1 | 2 | 3 | 4 | 5 |
| Combined | | | | | |
| Count | 19 | 15 | 7 | 3 | 1 |
| % | 42% | 33% | 16% | 7% | 2% |
| Non-Expert | | | | | |
| Count | 7 | 18 | 8 | 1 | 1 |
| % | 38% | 40% | 18% | 2% | 2% |
| Experts | | | | | |
| Count | 20 | 9 | 10 | 4 | 2 |
| % | 45% | 20% | 22% | 9% | 4% |

Given the difference in the means between the two groups as demonstrated by the t-test in Table 4-4, the next question to ask is, do the two sets of variable responses correlate? Table 4-7 shows the correlations between the expert and non-expert subject responses for each variable in the questionnaire. Subject matter shows a .3009 correlation, which is significant at the .05 level. The variables method, readability, and understandability are highly correlated between the two groups with $p < .01$. Purpose and Findings variables did not correlate between the two groups, and the variable Implications was slightly negatively correlated. This shows that although the t-test identified the two groups as different, there are still similarities in their pattern of answers for subject matter, method, readability, and understandability. This seems to demonstrate that although the two group means are different for all the variables, there is still an underlying agreement on some of the variables.

Table 4-7. Correlation between Non-experts and Experts on Seven Variables

| Dependent Variable | Correlation Experts by Non-Experts |
|---------------------------|-----------------------------------------------|
| Subject Matter | .3009* |
| Purpose | .2393 |
| Method | .3890** |
| Findings | .1448 |
| Implications | -.1127 |
| Readability | .4025** |
| Understandability | .4650** |

* $p < .05$, ** $p < .01$.

4.4 Document Metrics Descriptive Statistics

Both abstracts and citations were evaluated for the number words, to derive some simple metrics. The average number of citations per abstract was 5.5 (see Appendix E)..

The average number of words per abstract was 70 The average number of words per citation was 46.

Figure 4-1 shows the number of citations per abstract. For example, there are 27 abstracts that have five citations associated with them and four abstracts that have six citations associated with them.

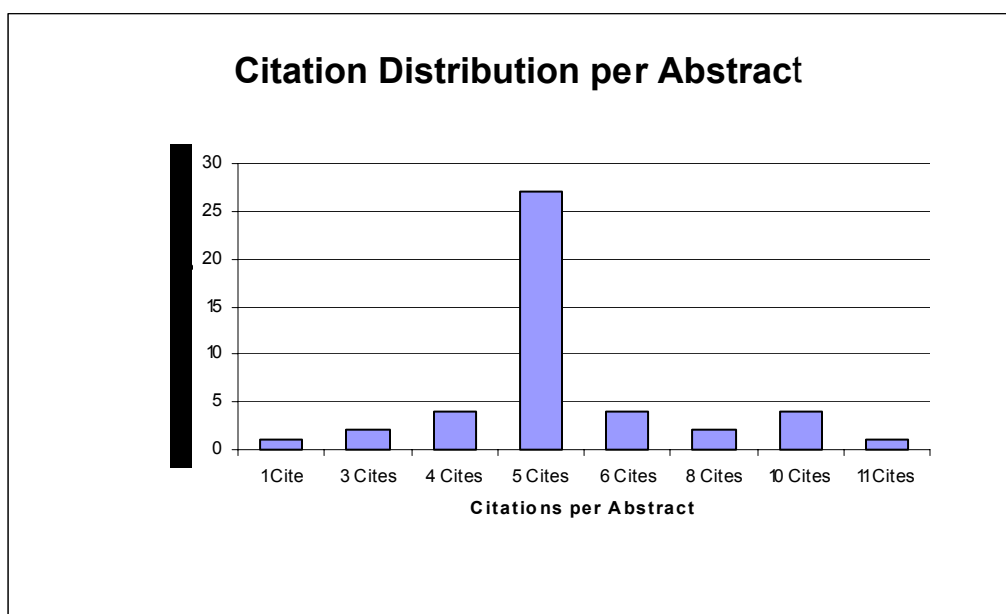


Figure 4 1. Citation Distribution per Abstract

4.4.1 Citation Location

Table 4-8 shows the areas of the document from which a given citation context was extracted. Literature Review had the highest number of citation extractions, at 32.1%, with Introduction following close behind with 31.7%. The Uncertain category is used for citations for which the author was unable to determine original document location.

Table 4-8. Citations Location Table

| Area of Document | Frequency Count | Percentage |
|-------------------------|------------------------|-------------------|
| Literature Review | 80 | 32.1% |
| Introduction | 79 | 31.7% |
| Method | 35 | 14.1% |
| Background | 18 | 7.2% |
| Discussion | 11 | 4.4% |
| Uncertain | 10 | 4.0% |
| Conclusion | 9 | 3.6% |
| Results | 3 | 1.2% |
| Future Work | 2 | 0.8% |
| Summary | 1 | 0.4% |
| Abstract | 1 | 0.4% |
| Total | 249 | |

4.4.2 Citation Category

Two raters looked through 249 of the citation contexts and determined which of the five categories (Refute, Support, Noted, Review, and Applied) the citation extracts belonged to. Table 4-9 shows how the two raters agreed and disagreed on placing the citations into one of the five given categories. They agreed 32% of the time and disagreed 68% of the time.

Table 4-9. Category Agreement Between Raters A and B

| | Category Agree | Category Disagree |
|-----------------|-----------------------|--------------------------|
| Frequency Count | 79 | 169 |
| Percentage | 32% | 68% |

In order for a computer system to categorize a citation that has been extracted, it matches trigger words or phrases with a category. The raters for the citations were asked to identify the trigger words that identified the category selection. Surprisingly, there was

fairly good agreement on the trigger words at 52%, as shown in Table 4-10. The raters agreed more often on the triggers than on the category the trigger represented. This is surprising, considering that there are probably thousands of possible trigger words/phrases and only five categories. Reinich's (1998) research showed that subjects on average selected the same sentences 55% of the time. This is close to the results we show here.

Table 4-10. Identifying Phrase Agreement Between Raters A and B

| | <i>Phrase Agree</i> | <i>Phrase Disagree</i> |
|------------------|---------------------|------------------------|
| Frequency Counts | 130 | 119 |
| Percentage | 52% | 48% |

Table 4-11 presents data on the categorization of citations by the raters. The data shows that rater A favored the Noted category with 46% of the responses. Rater B, on the other hand, is spread between Noted at 40% and Review at 30%.

Table 4-11. Citation Categories

| Cite Categories | Rater A | | Rater B | | Rater Agree | |
|------------------------|----------------|----------------|----------------|----------------|--------------------|----------------|
| | Count | Percent | Count | Percent | Count | Percent |
| Applied | 30 | 12% | 32 | 13% | 8 | 11% |
| Review | 14 | 6% | 74 | 30% | 4 | 5% |
| Noted | 115 | 46% | 100 | 40% | 50 | 67% |
| Support | 42 | 17% | 17 | 7% | 2 | 3% |
| Refute | 31 | 12% | 13 | 5% | 11 | 15% |
| Total | 249 | | 249 | | 75 | 30% |

The overall results from the document metrics descriptive statistics show two things: First, citation locations are very skewed and occur most frequently in the Introduction and Literature Review sections of documents. Second, two raters often agreed on the trigger words for a citation category, but often disagreed on the

classification of the category. The results give us a number of trigger words and phrases that we could use to begin to automate the process of categorizing citations, but a lot more work and data is required to be exhaustive.

4.5 Partial Coefficient of Determination Analyses

The next step in the analyses was to correlate the different similarity metrics. Correlation turned out not to be the appropriate technique, because we do not have independence among the 249 citations, because there are multiple citations contexts associated with each abstract. The 45 abstracts are independent of each other. The data was analyzed for all 249 citations using correlation, and then the citations associated with each of the 45 abstracts were analyzed using the coefficient of determination, R^2 . The results showed that grouping the citations by abstracts explained about 50% of the variance compared to considering citations independently. Coefficient of determination (R^2) is a “measure of the proportion of the variance of the dependent variable about its mean that is explained by the independent, or predictor, variables. The coefficient can vary between 0 and 1. If the regression model is properly applied and estimated, the analyst can assume that the higher the value of R^2 , the greater the explanatory power of the regression equation, and therefore the better the prediction of the criterion variable” (Hair et al, 1995).

Therefore, partial coefficient of determination (partial R^2) is the more appropriate analysis technique, allowing us to see the true effect of variables.

Partial correlations between each metric for each of the variables were calculated. Again, these were run for experts and non-experts separately and then for the entire group together.

Equations for each of the metrics used can be found in Chapter 3. A word contained in the citation context and the abstract was counted once, even if it occurred multiple times in each document, it was only counted once. In the table below the first column identifies the words that occur in the Abstract and the Citation Context. The second column is the number of times the term occurs in the Abstract. The Citation Context also shows the same words that occur between Citation Context and Abstract. The fourth column identifies the number of times that the word occurs in the citation context. This Frequency Count between Abstract and Citation Context would be 3 in this example.

Table 4-12. Data Example

| Abstract | Word Count | Citation Context | Word Count |
|-----------------|-------------------|-------------------------|-------------------|
| acquire | 1 | acquire | 1 |
| approach | 3 | approach | 1 |
| base | 2 | base | 2 |

The Un-weighted cosine is equation 6 in Chapter Three,

$$\text{Similarity}(D, C) = \cos(\theta) = \frac{D \bullet C}{|D| |C|}. \quad D \bullet C \text{ is the inner product, and is defined as the}$$

sum of the products of the vector components and can be written as follows:

$$D = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix} \quad C = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

This part of the equation was calculated by multiplying the number of times each word in common occurs. $D \bullet C$ can be calculated as:

$$D \bullet C = (1 * 1) + (3 * 1) + (2 * 2) = 7$$

This equation can also be written as $\sum_{i=1}^n DiCi$ as shown in Chapter 3. The $|D| |C|$ part of the equation is the vector length of the document and the citation context. The calculation of the vector length is a special case of the inner product of a vector, the calculation is performed on itself. Given

$$D = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix} \quad C = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

then $D \bullet D$ can be calculated as $1^2 + 3^2 + 2^2 = 14$ and $C \bullet C = 1^2 + 1^2 + 2^2 = 6$. $|D| |C|$ can be calculated : $\sqrt{14} * \sqrt{6}$. Now the entire equation can be calculated .

$$\text{Similarity}(D, C) = \cos(\theta) = \frac{D \bullet C}{|D| |C|} = \frac{7}{\sqrt{14} * \sqrt{6}} = .76$$

The similarity metric identified in the tables as $\log_{10}(N/n)$ is equations 10 and 11 in Chapter 3. These equations represent the best weighting technique according to Salton and Buckley (1987). The $\log_{10}(N-n/n)$ similarity metric is a slight modification of equations 10 and 11. $(N - n)$ is the probabilistic inverse collection frequency factor. This similarity metric's small modification put it in agreement with the probabilistic model of information retrieval. This was the rationale for selecting this metric in addition to the traditional similarity metrics. The calculation of $\log_{10}(N/n)$ and $\log_{10}(N-n/n)$ were fairly

straight forward, following the equations and easily calculated from data shown in the above examples.

Table 4-13. Partial Coefficient of Determinants (Partial R² Values) Among Frequency Counts/Cosine Scores and Expert Variable Ratings

| Variable | Frequency Count | Cosine Unweighted | Cosine Log10(N/n) | Cosine Log10 (N-n/n) |
|-------------------|------------------------|--------------------------|--------------------------|-----------------------------|
| Subject Matter | 0.191*** | 0.184*** | 0.018 | 0.006 |
| Purpose | 0.109*** | 0.106*** | 0.000 | 0.002 |
| Method | 0.087*** | 0.076*** | 0.001 | 0.003 |
| Findings | 0.018* | 0.018* | 0.002 | 0.000 |
| Implications | 0.038** | 0.037** | 0.005 | 0.000 |
| Readability | 0.006 | 0.010 | 0.002 | 0.001 |
| Understandability | 0.008 | 0.006 | 0.009 | 0.013 |

*p < .05, ** p < .01, ***p < .0001

Table 4-13 depicts the partial correlations among the metrics for each variable for the experts. Results indicate that there are significant correlations between subject matter and frequency count and between subject matter and un-weighted cosine. Both correlations are positive and strong with a p less than .0001.

Significant correlations were also found between purpose and frequency count and purpose and un-weighted cosine. Again, both associations were strong with significance less than .0001. The partial correlations between method and frequency count, method and un-weighted cosine, findings and frequency count, and findings and un-weighted cosine, are also significant, but all are also strong with p < .001. Lastly, the correlations between implications and frequency count and implications and un-weighted cosine were also significant. There were no significant correlations among any of the cosine log measures.

Partial correlations were then run for the metric similarities for the non-experts.

Results appear in Table 4-14.

Table 4-14. Partial Coefficient of Determinants Among Cosine/Frequency Count Scores and Non-expert Variable Ratings

| Variable | Frequency Count | Cosine Unweighted | Cosine Log10 (N/n) | Cosine Log10 (N-n/n) |
|-------------------|------------------------|--------------------------|---------------------------|-----------------------------|
| Subject Matter | 0.241** | 0.022** | 0.008 | 0.001 |
| Purpose | 0.245** | 0.214** | 0.015 | 0.005 |
| Method | 0.185** | 0.165** | 0.005 | 0.003 |
| Findings | 0.044* | 0.041* | 0.000 | 0.005 |
| Implications | 0.091** | 0.097** | 0.000 | 0.001 |
| Readability | 0.077** | 0.060* | 0.000 | 0.001 |
| Understandability | 0.060* | 0.051* | 0.002 | 0.001 |

* $p < .001$, ** $p < .0001$.

Results indicate that there are significant correlations between all the variables and frequency count and between all the variables and the un-weighted cosine measures. Again, there are no significant correlations between any of the variables and the weighted cosine measures, cosine $\log_{10}(N/n)$ or $\log_{10}(N-n/n)$.

Examining the significant correlations in Table 4-14, there are significant ($p < .0001$) correlations between Subject Matter, Purpose, Method, Implications and Readability with frequency count. Correlations with Cosine are the same except Readability is less significant ($p < .001$). Overall the correlations are strong for frequency count and slightly weaker for Cosine. Non-experts show slightly stronger associations than found with the experts.

Last, the correlations between readability and frequency count, readability and un-weighted cosine, understandability and frequency count, and understandability and un-

weighted cosine are also significant. Note that these relationships were not significant for the experts.

The third set of correlations was computed for the experts and the non-experts together. Significant correlations were found for all the variables with the frequency count scores and with the cosine scores. Like previous analyses discussed above, there were no significant correlations found between the study variables and the cosine log 10 measures. Results appear below in Table 4-15.

It is interesting to note that the combined data of experts and non-experts has higher partial correlation values with the variables than the separate data sets do.

Table 4-15. Partial Correlations Among Frequency Counts/ Cosine Similarity Metrics Combined Variables of Experts and Non-Experts

| Variable | Frequency Count | Cosine Unweighted | Cosine Log10 (N/n) | Cosine Log10 (N-n/n) |
|-------------------|------------------------|--------------------------|---------------------------|-----------------------------|
| Subject Matter | 0.298*** | 0.281*** | 0.013 | 0.002 |
| Purpose | 0.289*** | 0.263*** | 0.008 | 0.001 |
| Method | 0.210*** | 0.183*** | 0.001 | 0.003 |
| Findings | 0.057** | 0.055** | 0.001 | 0.000 |
| Implications | 0.114*** | 0.118*** | 0.005 | 0.001 |
| Readability | 0.069*** | 0.060** | 0.001 | 0.001 |
| Understandability | 0.055** | 0.046* | 0.006 | 0.005 |

*p < .01, **p < .001, ***p < .0001.

Partial correlations were also run for the location of the citations and the type or category of citation, with the four similarity measures. Citation Category 1 and Citation Category 2 are the classifications of each of the two experts. Results are depicted in Table 4-16. Only two significant correlations were found. First, a significant, positive, but weak correlation was found between citation category 2 and un-weighted cosine ($p < 0.05$). Second, there was also a significant, positive, but also very weak correlation

between cosine log and citation category 2. This means that the selections of one of the raters were in agreement with the un-weighted cosine and the cosine $\log_{10}(N/n)$.

Table 4-16. Partial Correlations Among Cosine/Binary Scores and Nominal Variables

| Variable | Frequency Count | Cosine Unweighted | Cosine Log10 (N/n) | Cosine Log10 (N-n/n) |
|--------------------|------------------------|--------------------------|---------------------------|-----------------------------|
| Citation Location | 0.098 | 0.093 | 0.065 | 0.045 |
| Citation Category1 | 0.040 | 0.041 | 0.024 | 0.03 |
| Citation Category2 | 0.031 | 0.054* | 0.048* | 0.046 |

*p < .05

4.6 Regression Analyses

In the last part of the analysis regression models were constructed for cosine un-weighted and frequency counts for the three groups — non-experts, experts, and the combined group — for a total of six runs. In addition, a stepwise regression was completed for all the groups and combination of groups. Results from the analysis using just the experts are presented in Table 4-1.

Table 4-17. Regression Analysis for Predicting Cosine Un-weighted Using the Expert Ratings

| Dependent Variable | DF | Slope | SE | F-Value | P-Value | VIF* |
|---------------------------|-----------|--------------|-----------|----------------|----------------|-------------|
| Abstract | 44 | | | 3.3810 | <.0001 | 4.9587237 |
| Area | 10 | | | 1.3870 | 0.1890 | 3.0839627 |
| Subject Matter | 1 | 0.0391165 | 0.010278 | 14.4858 | 0.0002 | 3.1352385 |
| Purpose | 1 | 0.0067579 | 0.008029 | 0.7085 | 0.4010 | 3.0374485 |
| Method | 1 | 0.0136574 | 0.008512 | 2.5743 | 0.1103 | 2.6088998 |
| Findings | 1 | -0.010871 | 0.009544 | 1.2974 | 0.2561 | 2.733435 |
| Implications | 1 | 0.0076571 | 0.013022 | 0.3458 | 0.5572 | 2.6150985 |
| Readability | 1 | 0.001463 | 0.019005 | 0.0059 | 0.9387 | 4.0648467 |
| Understandability | 1 | -0.017414 | 0.015848 | 1.2074 | 0.2733 | 6.4703664 |

* For the categorical variables, the greatest VIF from their corresponding dummy variables is shown.

The results show that abstract is highly significant, with a $p < .0001$. This shows that matching the citation context to the abstract they are referencing accounts for 44% of the total variance (see Table 4-18). VIF or collinearity is at 4.96, which is below the threshold of 10. So we can ignore collinearity as a problem for the data. There are no slope and standard error (SE) and abstract and area because they are discrete variables. On the balance of the variables, subject matter and area are the only other variables with significant explanatory power, accounting for 7.2% and 6.9% of the variance of the model (see Table 4-18). In total the expert model accounted for 58% of the total variance. A stepwise regression was also executed. The variables added to the model that accounted for greater than 0.1 of the variance are listed in the order added: Abstract, Subject Matter, Method.

The second regression analysis was conducted with non-experts. Results appear below in Table 4-18. It is interesting to note that the non-experts accounted for more variance in the model than the experts: R^2 for the entire model is 65% (Table 4-18). In addition, variables accounting for the most variance are Abstract (.485), Area (.065), Subject Matter (.052), and Methods (.033). Readability and Understandability variables have VIF values of 10.16 and 10.41 respectively, which indicates that there is collinearity between the variables. This can be explained by students not making a clear distinction for which citation contexts were readable or understandable. They tended to answer these two questions with the same number for all citation contexts. A stepwise regression was also executed. The variables added to the model that accounted for greater than 0.1 of the variance are listed in the order added: Abstract, Subject Matter, Method, and Purpose.

Table 4-20 shows the regression analysis results for cosine un-weighted and the combined data from experts and non-experts. The combined data accounts for slightly more total variance, .657 (refer to Table 4-17), compared to either experts (.652) or non-experts (.580). The variables that have the most variance explained are Abstract (.488), Area (.064), Subject Matter (.049), and Method (.039). A stepwise regression was also executed. The variables added to the model that accounted for greater than 0.1 of the variance are listed in the order added: Abstract, Subject Matter, Method, Understandability, Purpose. Combining the data appears to make a slightly more robust model than either data set by itself.

Table 4-18. Variance Explained R2 by Variables

| Variables | Cosine Unwt'd Expert Ratings | Cosine Unwt'd Non- expert Ratings | Cosine Unwt'd Comb. Expert and Non- expert Ratings | Freq. Counts Expert Ratings | Freq. Counts Non- expert Ratings | Freq. Counts Comb. Expert and Non- expert Ratings |
|-------------------|-------------------------------------------------|----------------------------------------------------------|-----------------------------------------------------------------------------------|------------------------------------------------|---------------------------------------------------------|----------------------------------------------------------------------------------|
| Abstract | 0.44439 | 0.48535 | 0.48767 | 0.54958 | 0.56243 | 0.58732 |
| Area | 0.06940 | 0.06498 | 0.06374 | 0.07530 | 0.06920 | 0.07070 |
| Subject Matter | 0.07225 | 0.05204 | 0.04945 | 0.06476 | 0.01648 | 0.03682 |
| Purpose | 0.00379 | 0.01507 | 0.01636 | 0.00322 | 0.01773 | 0.02360 |
| Method | 0.01365 | 0.03279 | 0.03859 | 0.02625 | 0.05670 | 0.05469 |
| Findings | 0.00693 | 0.00072 | 0.00075 | 0.00827 | 0.00002 | 0.00077 |
| Implications | 0.00186 | 0.00361 | 0.00329 | 0.00181 | 0.00000 | 0.00038 |
| Readability | 0.00003 | 0.00285 | 0.00308 | 0.00346 | 0.00018 | 0.00291 |
| Understandability | 0.00645 | 0.00618 | 0.00511 | 0.00039 | 0.01035 | 0.00308 |
| Total | 0.580018 | 0.651869 | 0.65697 | 0.628213 | 0.67427 | 0.69786 |

Table 4-19. Regression Analysis for Predicting Cosine Un-weighted Using the Non-expert Ratings

| Dependent Variable | DF | Slope | SE | F-Value | P-Value | VIF* |
|---------------------------|-----------|--------------|-----------|----------------|----------------|-------------|
| Abstract | 44 | | | 3.9867 | <.0001 | 5.2786283 |
| Area | 10 | | | 1.2925 | 0.2373 | 3.0352305 |
| Subject Matter | 1 | 0.0461157 | 0.014431 | 10.2116 | 0.0016 | 5.6130434 |
| Purpose | 1 | 0.0241056 | 0.014292 | 2.8450 | 0.0933 | 5.3705068 |
| Method | 1 | 0.0265452 | 0.010571 | 6.3058 | 0.0129 | 2.6291472 |
| Findings | 1 | -0.004032 | 0.011024 | 0.1338 | 0.7150 | 2.390346 |
| Implications | 1 | 0.0129085 | 0.015721 | 0.6742 | 0.4127 | 3.0047902 |
| Readability | 1 | -0.017679 | 0.024241 | 0.5319 | 0.4667 | 10.161612 |
| Understandability | 1 | -0.023129 | 0.021507 | 1.1565 | 0.2836 | 10.40845 |

* For the categorical variables, the greatest VIF from their corresponding dummy variables is shown.

Table 4-20. Regression Analysis for Predicting Cosine Un-weighted Combining the Expert and Non-expert Ratings

| Dependent Variable | DF | Slope | SE | F-Value | P-Value | VIF* |
|---------------------------|-----------|--------------|-----------|----------------|----------------|-------------|
| Abstract | 44 | | | 6.0485 | <.0001 | 5.2734482 |
| Area | 10 | | | 1.4227 | 0.1729 | 3.0405411 |
| Subject Matter | 1 | 1.4655271 | 0.548093 | 7.1496 | 0.0082 | 5.6239325 |
| Purpose | 1 | 1.1543206 | 0.542917 | 4.5205 | 0.0348 | 5.3950885 |
| Method | 1 | 1.3057607 | 0.397002 | 10.8179 | 0.0012 | 2.6229415 |
| Findings | 1 | -0.159506 | 0.419619 | 0.1445 | 0.7043 | 2.4194593 |
| Implications | 1 | 0.1586937 | 0.598174 | 0.0704 | 0.7911 | 3.0030602 |
| Readability | 1 | -0.68167 | 0.922408 | 0.5461 | 0.4608 | 10.156319 |
| Understandability | 1 | -0.620133 | 0.815928 | 0.5777 | 0.4482 | 10.355316 |

*For the categorical variables the greatest VIF from their corresponding dummy variables is shown.

The next three regression tables (4-22, 4-22, and 4-23) use the frequency counts of words in common between the citation context and the abstract. This similarity metric is by far the easiest to determine, and as the data will show is also a better fit for the data.

Table 4-21 shows the regression analysis results for frequency count of words in common between abstracts and the citation contexts for the expert group. The variables that have the most variance explained are Abstract (.550), Area (.075), Subject Matter

(.065), and Method (.026); refer to Table 4-18. The frequency count and expert model accounted for a total variance of .628. Collinearity is not significant for any of the variables. A stepwise regression was also executed for this model. The variables added to the model that accounted for greater than 0.1 of the variance are listed in the order added: Abstract, Subject Matter.

Table 4-21. Regression Analysis for Frequency Counts Using Expert Ratings

| Dependent Variable | DF | Slope | SE | F-Value | P-Value | VIF* |
|---------------------------|-----------|--------------|-----------|----------------|----------------|-------------|
| Abstract | 44 | | | 5.1856 | <.0001 | 4.9574575 |
| Area | 10 | | | 1.5227 | 0.1340 | 3.0880715 |
| Subject Matter | 1 | 1.4213022 | 0.394978 | 12.9487 | 0.0004 | 3.1385362 |
| Purpose | 1 | 0.2400178 | 0.308601 | 0.6049 | 0.4377 | 3.0440869 |
| Method | 1 | 0.7303658 | 0.325299 | 5.0410 | 0.0259 | 2.6107771 |
| Findings | 1 | -0.456443 | 0.365609 | 1.5586 | 0.2134 | 2.718267 |
| Implications | 1 | 0.2903451 | 0.499131 | 0.3384 | 0.5615 | 2.6095051 |
| Readability | 1 | -0.588664 | 0.730495 | 0.6494 | 0.4214 | 4.0659493 |
| Understandability | 1 | -0.163776 | 0.608904 | 0.0723 | 0.7883 | 6.463984 |

* For the categorical variables the greatest VIF from their corresponding dummy variables is shown.

Table 4-22 shows the regression analysis of word in common frequency counts conducted with non-experts. It is interesting to note that the non-experts accounted for more variance in the model than the experts: R^2 for this regression analysis is .674 (see Table 4-18). In addition, variables accounting for the most variance are Abstract (.562), Area (.069), Methods (.057), Purpose (.018), and Subject Matter (.016). Readability and Understandability variables have VIF values of 10.61 and 9.05 respectively, which indicates there is collinearity between the variables. A stepwise regression was executed for this analysis. The variables added to the model that accounted for greater than 0.1 of the variance are listed in the order added: Abstract, Purpose, Method, Understandability, Subject Matter.

**Table 4-22. Regression Analysis for Frequency Counts
Using Non-expert Ratings**

| Dependent Variable | DF | Slope | SE | F-Value | P-Value | VIF* |
|---------------------------|-----------|--------------|-----------|----------------|----------------|-------------|
| Abstract | 44 | | | 5.4627 | <.0001 | 5.4887928 |
| Area | 10 | | | 1.3903 | 0.1874 | 2.9321314 |
| Subject Matter | 1 | 1.029935 | 0.58189 | 3.1328 | 0.0784 | 8.2677331 |
| Purpose | 1 | 1.0448515 | 0.568778 | 3.3746 | 0.0678 | 7.3748609 |
| Method | 1 | 1.1627461 | 0.346807 | 11.2408 | 0.0010 | 2.5338548 |
| Findings | 1 | -0.021587 | 0.331961 | 0.0042 | 0.9482 | 2.3968819 |
| Implications | 1 | 0.0039739 | 0.476059 | 0.0001 | 0.9933 | 3.2915069 |
| Readability | 1 | -0.137267 | 0.740816 | 0.0343 | 0.8532 | 10.6061 |
| Understandability | 1 | -0.934684 | 0.668463 | 1.9551 | 0.1637 | 9.0513105 |

* For the categorical variables the greatest VIF from their corresponding dummy variables is shown.

The results in Table 4-23 show that Abstract is again highly significant with a $p > .0001$. Abstract accounts for .587 (refer to Table 4-16) of the total variance, followed by Area (.071), Method (.055), Subject Matter (.037), and Purpose (.024).

**Table 4-23. Regression Analysis for Frequency Counts
Combining the Expert and Non-expert Ratings**

| Dependent Variable | DF | Slope | SE | F-Value | P-Value | VIF* |
|---------------------------|-----------|--------------|-----------|----------------|----------------|-------------|
| Abstract | 44 | | | 6.0485 | <.0001 | 5.2734482 |
| Area | 10 | | | 1.4227 | 0.1729 | 3.0405411 |
| Subject Matter | 1 | 1.4655271 | 0.548093 | 7.1496 | 0.0082 | 5.6239325 |
| Purpose | 1 | 1.1543206 | 0.542917 | 4.5205 | 0.0348 | 5.3950885 |
| Method | 1 | 1.3057607 | 0.397002 | 10.8179 | 0.0012 | 2.6229415 |
| Findings | 1 | -0.159506 | 0.419619 | 0.1445 | 0.7043 | 2.4194593 |
| Implications | 1 | 0.1586937 | 0.598174 | 0.0704 | 0.7911 | 3.0030602 |
| Readability | 1 | -0.68167 | 0.922408 | 0.5461 | 0.4608 | 10.156319 |
| Understandability | 1 | -0.620133 | 0.815928 | 0.5777 | 0.4482 | 10.355316 |

* For the categorical variables the greatest VIF from their corresponding dummy variables is shown.

A stepwise regression was executed for this analysis. The variables added to the model that accounted for greater than 0.1 of the variance are listed in the order added: Abstract, Subject Matter, Method.

Total variance explained by each regression model is listed in Table 4-18. The expert group explained .628 of the total variance. The non-expert group accounted for .674 of the total variance and the combined groups explained .698 of the variance.

4.7 Conclusion

The purpose of this study was to examine the feasibility of an easily computer-calculated word metric to select citation context as a document summary compared to human selection. In this chapter, data analyses were presented.

Correlation analyses indicated that there were significant association between frequency count of words in common between the two documents and the un-weighted cosine measures for non-experts, experts, and the combined sample. Comparisons of means indicated that there were a number of differences in average scores of the predictor

variables. Experts assigned higher on average on subject matter, purpose, findings, implications, and understandability than non-experts did. Non-experts assigned significantly higher than experts only on the variable Method.

Regression analysis revealed that different predictor variables significantly predicted frequency count accuracy depending on the group of respondents. In the combined sample, subject matter and subject method were significant.

The following chapter will provide an overview of the results, discuss shortcomings of the study, provide recommendations for future research, and apply results to literature.

5. Discussion

5.1 Summary

The overall findings of this study suggest that it is feasible to use simple similarity metrics to select citation contexts as document surrogates. The similarity metric can predict up to 69.8% of the variance, giving results close to what human subjects would select. This study compared several types of similarity metrics and how well the metrics predicted choices of citation contexts that humans make, considering as variables the level of familiarity the rater has with the topic area, the type of similarity metric used, the location of the citation, and the category of the citation.

The results show that simple metrics, word frequency counts, and cosine measurements are better matches to rater selections than the more complex weighted cosine measures, with small variations between expert and non-expert groups. The expert and non-expert raters showed different patterns of responses. There are several interesting points on the two groups; although the experts and non-experts showed significant differences on the t-test, there was a significant amount of correlation between the groups on several of the variables. The groups answered the survey questions differently, resulting in different means, but there was also a pattern of agreement between the groups on several of the variables. It's not clear from this study whether the differences were due to different insights into the subject matter or because the two groups viewed the Likert scale differently. The similarity metric varied slightly between the two groups as a

predictor of user responses, but when the group data were combined the explained variance went up, indicating that the combination smoothed group scoring differences, creating a better fit with the similarity metric.

The category and location of the citation context were tested in this research to determine if these citation measures could enhance the explanatory power of the similarity metric. The citation category did not enhance the similarity metric's explanatory power. The location of the citation had the effect, explaining additional variance by 6 to 7 %. This finding was fairly consistent across groups and similarity metrics.

The present research used ten variables in the study (Abstract, (Citation) Area, (Citation) Category, Subject Matter, Method, Purpose, Findings, Implications, Readability, and Understandability). Regression analysis and stepwise regression analysis show that number of variables could be reduced to just those that made a significant contribution to explanation of the variance. For the combined group data compared to the frequency count of words in common, the variables could be simplified to just four variables: Abstract, Subject Matter, Method, and Purpose.

5.2 Discussion

The evaluation of text summarization systems by comparing machine-made and human-made abstracts is a growing area of research. Hovy and Lin (1998) argue that automated summarization techniques present only approximations of complex human processing. This study indicates that using a similarity metric to select a citation context gives a good approximation, within a range of certainty, that it will match a human selection. The advantage of using citation contexts counters this argument, because

citation contexts are written by knowledgeable authors and are not subject to the same limitations as machine-generated summarizations. Another argument is that humans do not show predictable reliability in choices of representative sentences during the process of summarizing documents. This indicates that there may be many reasonably representative sets of sentences existing in any article (Resnick, 1998). The general results in this study seem to partially agree with Resnick's argument, in that given more than one equally reasonable alternative the similarity metric will select a reasonable citation context that more often than not will agree with selection by a person.

Although the results of this study are promising there are a few points that could limit the usefulness of citation contexts as document summaries. Citation contexts are not readily available, in fact, electronically extracted citations are very difficult to find. The data used for this study used an automated system, but the data is only for the discipline of computer science. So if you wanted extractions in any other discipline they would have to be extracted, and unfortunately it would probably be manually. Another limiting factor is that the citation contexts are extracted, very often in the middle of a sentence, and therefore have grammatical references that do not appear in the extraction. Making the text confusing, which unfortunately is the same argument used against machine generated abstracts.

These research results provide information on the methods by which matches, with a given amount of error, can be made between document abstracts and citation contexts, that match the human selection process. Each objective of this study is listed below, with supporting evidence for the conclusions drawn. The results answer several important questions:

1. Does the similarity metric between a citation context and an abstract of the document correspond with the citation contexts chosen by subject raters for document surrogates?

This was answered by getting survey data from raters evaluating the similarity of citation contexts compared to the abstract of the document they are representing.

Similarity metrics were then calculated for all abstracts and citation contexts. The survey data was statistically compared to the similarity metrics to determine any relationships.

The raters were given the abstract of a source document and the referring citation extracts. They evaluated extracted citation text on seven dimensions, rating how well it represented the meaning of the abstract. The variables included subject matter, purpose, methods, findings, implications, readability, and understandability. Each variable was evaluated, by the subjects, on a five-point Likert scale (*not at all*, *slightly*, *adequately*, *very well*, and *completely*). The higher the score, the better the citation acted as a document summary, with respect to the indicated variable.

For comparison, similarity metrics were calculated for each pairing of abstract and citation context. The four metrics calculated were 1) simple frequency counts of words in common between the abstract and the citation context, 2) un-weighted cosine, 3) weighted cosine, and 4) weighted cosine adjusted for sample size. These were then compared to the survey data by using correlation and regression analysis.

The results of the research indicate that there are grounds for the statement that similarity metrics do reflect human choices in selecting a citation context as a document

surrogate. Using the calculated similarity metrics as the dependent variable and the subject-assigned scores, as the independent variables, the total variance explained ranged from 58% to 69.8%. The range of scores depended on whether the subjects were considered as separate groups of experts and novice raters or as a single group. This result was far higher than the expectation of the researcher and by text summarization standards, a very good result.

2. Is there a difference between subject matter experts and non-experts in

2.1. How they responded on the survey?

2.2. How their responses correlated with similarity metrics?

To determine the robustness of the similarity metric, two groups were established, one expert and one non-expert. The expert group had familiarity with the literature of information science and natural language processing. The non-expert group was composed of undergraduate students. Expert group similarity ratings were compared to the ratings of non-experts. Both groups performed the same task of rating citation surrogates in terms of their relationships to a given abstract.

Comparing the two groups' responses show a difference in the response pattern. The t-test showed a significant difference between experts and non-experts for every variable. On the other hand, correlation tests on the two groups' survey data show positive correlation on four of the variables (Subject Matter, Method, Readability, and Understandability). This means that although the groups differed overall in the scores

they gave when rating the citation contexts, there was a correlation between the groups for some of the variables, indicating similarity of responses between the two groups.

The matched paired t-tests between rating groups found significant differences for all the variables. The experts' ratings averaged higher scores, in the variables subject matter, findings, and purpose than the scores of the students. The students averaged higher scores on methods used in the research. The experts were clearly using the scale differently than the students were. For every variable except one (methods) the experts gave higher marks. We could speculate that the higher scores of the experts reflect a greater confidence in their judgments.

3. Are different similarity metrics correlated differently with human subjects' responses to document similarity?

The similarity metrics used for this study, in order of complexity are:

- 1) frequency counts of words in common

- 2) simple cosine, $Similarity(D, C) = \cos(\theta) = \frac{D \bullet C}{\|D\| \|C\|}$.

- 3) $\sum w_{id} w_{ic}$ where $w_{id} = \frac{tf_{id} * \log_{10}(N / n_i)}{\sqrt{\sum |tf_{id} * \log_{10}(N / n_i)|^2}}$

$$\text{and } w_{ic} = |0.5 + (0.5 \text{ } tf_{ic} / \max \text{ } tf_i)| * \log_{10}(N / n_i)$$

$$4) \sum w_{id} w_{ic} \text{ where } w_{id} = \frac{tf_{id} * \log_{10}(N - n_i / n_i)}{\sqrt{\sum |tf_{id} * \log_{10}(N - n_i / n_i)|^2}}$$

$$\text{and } w_{ic} = |0.5 + (0.5 \text{ } tf_{ic} / \max \text{ } tf_i)| * \log_{10}(N - n_i / n_i)$$

The choice of similarity metric employed made a difference. Frequency count and the un-weighted cosine measure are better for the selection of similar citation contexts when compared to the choices of human subjects. To assess which of the similarity metrics gave the best match with the rater data, the correlations among simple cosine (2), similarity metrics (3 & 4) and frequency count (1) scores and the group's scores were evaluated. For the experts, correlations of the five rating variables; subject matter, purpose, method, findings, and implications with the similarity measures frequency count and cosine were positive. There were no significant correlations among either of the cosine log 10 measures (3 & 4).

The simplest measure, the frequency count of words in common, did the best job of fitting user responses. Frequency count and cosine measures are both efficacious in terms of agreeing with the selection of similar citation contexts similar to the choices of human subjects. The log 10 similarity measures show no such power.

For the data on the student raters, there were significant correlations between all seven rating variables and the frequency count (1) and simple cosine (2) measures. There were no significant correlations between the rating variables and the cosine log 10 similarity (3 & 4) measures. The significant correlations between the rating variables; subject matter and purpose, with frequency count and cosine were slightly stronger than those for the experts. Student data showed significant but weak correlations between

readability and understandability with frequency count and simple cosine. This was not the case in the experts' ratings.

The third set of correlations was computed for the combined group of experts and students together. Significant correlations were found for all the rating variables with the frequency count and cosine scores. There were no significant correlations for the study variables with the cosine log 10 similarity measures.

If we think of the similarity measures on a continuum of simple calculations (frequency count and cosine) to more complex calculations (cosines with weighting factors), it seems that simpler similarity measures are a better match to what the raters thought was similar between citation contexts and the abstracts. A possible answer is that the weighting factors, which add complexity to the log measures, were developed for larger documents. When used in this study, they may be too sensitive for the small abstracts and citation contexts.

This result also has the added benefit of being easier to calculate on a computer system. The application of the similarity metric is greatly simplified and demands fewer computing resources. This bodes well for the development of an automated system to select citation contexts to fill in as document surrogates.

4. Are there any easily computer-calculated document metrics that can increase the explanatory power of the similarity metric?

4.1. Does the location of citation add additional explanatory power to the similarity metric?

The location method, an extraction technique in text summarization, recognizes that sentences with a higher degree of centrality to the topic tend to be found in specifiable locations (Lin & Hovy, 1995). This analysis is to test if there is any credence to location in the selection of a citation context. The citation location was determined by examination of the documents. Each location was recorded as a dummy variable as follows: 1 = introduction, 2 = literature review, 3 = discussion, 4 = methodology, 5 = conclusion.

Correlations among similarity metric scores and location were measured. No correlations were found between similarity metrics and location. Regression analysis was also run on the location data in relation to the rater variables. The results show that citation location consistently explained between 6 and 7 percent of the total variance.

4.2. Can the type of citation add additional explanatory power to the similarity metric?

Citation extracts were classified into one of five categories: refute, support, noted, review, and applied. The citation contexts were classified by two Ph.D.-level librarians. The raters disagreed, with each other, 68% of the time when classifying the citations.

Raters were also asked to identify the “trigger words” which they used to classify citation extracts. Trigger words are identifying markers that can be programmed into a computer to classify a citation. When identifying the trigger words, raters agreed on 52% of the trigger words. An interesting note Resnick (1998) found similar results with subjects selecting sentences, for text summarization of a larger document, an average

55% agreement in subject selection. Results with only two raters is inadequate to draw any general conclusions, but it is an interesting side note to this study, that further research might yield interesting results.

Correlation and regression analysis were then performed on the two raters' scores and the four similarity metrics. One rater had no correlation with any of the similarity metrics. The second rater had a significant correlation ($p < .05$) with the un-weighted cosine and the weighted cosine log 10 (N/n). The regression analysis showed that citation category had almost no explanatory power. The data was too "dirty" for further analysis and it was removed from further regression analysis.

4.3. Can the type of citation and location combined add additional explanatory power to the similarity metric, so that the three variables explain more variance?

The data for citation type was eliminated so there is no definitive way to assess this question without further research.

5.3 Conclusion

The primary findings of the current study show that it is feasible to use citation contexts as document surrogates. Either a simple cosine similarity metric or frequency count could be used to make the citation context selection. Using the similarity metric does not give exactly the same results as human selections, but it does give a reasonable chance of selecting the right citation contexts with a bounded error. The secondary findings of this study are that location of a citation context has a minor influence,

increasing the likelihood of selecting the same citation context selected by human raters. Citation category show no value based on the data gathered for this research.

Small's (1987) model was the conceptual basis for this research. That paper made a compelling case for using citation contexts as document summaries. The relationship between citation context selection by human subjects and similarity word metrics, as researched here, holds promise for the automatic creation of a specialty narrative, despite some of the limitations listed in the discussion.

List of References

- Proceedings of the Ninth International Conference on Computational Linguistics(1982, July 5-10). Paper presented at the Ninth International Conference on Computational Linguistics (COLING), Prague, Czechoslovakia.
- MUC-4(1992, June 16-18). Paper presented at the Fourth Message Understanding Conference (MUC-4), McLean Virginia.
- Alterman, R. (1991). Understanding and Summarization. Artificial Intelligence Review, 5, 239-254.
- Alterman, R., & Bookman, A. L. (1990). Some Computational Experiments in Summarization. Discourse Processes, 13, 143-174.
- Appelt, D. E. (1999). Introduction to Information Extraction. AI Communications, 12(3), 161-172.
- Atkins, H. (1999). The ISI Web of Science - Links and Electronic Journals - How links work today in the Web of Science, and the challenges posed by electronic journals. D-Lib Magazine, 5(9).
- Baldi, S. (1998). Normative Versus Social Constructivist Processes in the Allocation of Citations: A Network-Analytic Model. American Sociological Review, 63, 829-846.
- Baldwin, B., & Morton, T. S. (1998). Dynamic Coreference-Based Summarization, Proceedings of the Third Conference on Empirical Methods in Natural Language Processing (EMNLP-3). Granada, Spain.
- Bara, B. G., & Guida, G. (Eds.). (1984). Computational models of natural language processing. New York: North-Holland.
- Barker, K., Chali, Y., Copeck, T., Matwin, S., & Szpakowicz, S. (1998). The Design of a Configurable Text Summarization System (TR 98-04). Ottawa: School of Information Technology and Engineering, University of Ottawa.
- Barker, T., Szpakawicz, K., Copeck, T., Delannoy, J. F., & Matwin, S. (1996). Preliminary Validation of a Text Summarization Algorithm (TR-96-04): University of Ottawa, Department of Computer Science.
- Barzilay, R. (1997). Lexical Chains for Summarization. Unpublished M.Sc., Ben-Gurion University of the Negev.

- Barzilay, R., McKeown, K. R., & Elhadad, M. (1999). Information Fusion in the Context of Multi-Document Summarization. Paper presented at the Proceedings of the 37th Association for Computational Linguistics, Maryland.
- Bassi, A. A. (1999). An Associative Semantic Model for Text Processing. Electronic Edition, IEEE Computer Society DL, 31-37.
- Bates, M. J. (1998). Information Science: The Invisible Substrate. Los Angeles: University of California.
- Bates, M. J. (1999). The Invisible Substrate in Information Science. Journal of the American Society for Information Science, 50(12), 1043-1050.
- Berger, A. L., & Mittal, V. O. (2000). OCELOT: A System for Summarizing Web Pages. In N. J. Belkin & P. Ingwersen & M.-K. Leong (Eds.), 23rd Annual Conference on Research and Development in Information Retrieval (ACM SIGIR) (pp. 144--151). Athens, Greece: ACM.
- Black, W. J. (1993). Parsing, Linguistic Resources and Semantic Analysis for Abstracting and Categorisation.
- Black, W. J., & Johnson, F. C. (1988). A Practical Evaluation fo Two Rule-based Automatic Abstracting Techniques. Expert Systems for Information Management, 1(3), 159-177.
- Bollacker, K. D., Lawrence, S., & Giles, C. L. (1999). A System for Automatic Personalized Tracking of Scientific Literature on the Web. Paper presented at the Fourth ACM Conference on Digital Libraries, Berkeley, CA.
- Braam, R. R., Moed, H. F., & Van Raan, A. F. J. (1991). Mapping of Science by Combined Co-Citation and Word Analysis I. Structural Aspects. Journal of the American Society for Information Science, 42(4), 233-251.
- Braam, R. R., Moed, H. F., & Van Raan, A. F. J. (1991). Mapping of Science by Combined Co-Citation and Word Analysis. II. Dynamical Aspects. Journal of the American Society for Information Science, 42(4), 252-266.
- Brandow, R., K. Mitze, & Rau, L. F. (1995). Automatic Condensation of Electronic Publications by Sentence Selection. Information Processing and Management, 31(5), 675-685.
- Broek, P. v. d., & Trabasso, T. (1986). Causal Networks versus Goal Hierarchies in Summarizing Text. Discourse Processes, 9, 1-15.
- Buckley, C., Mitra, M., Walz, J., & Cardie, C. (1997, November 19-21). Using Clustering and SuperConcepts Within SMART. Paper presented at the The Sixth Text REtrieval Conference (TREC 6), Gaithersburg, Maryland.

Burghardt, W., & Holker, K. (Eds.). (1979). Text Processing (Vol. 3). New York: Wlaler de Gruyter.

Bush, V. (1945). As We May Think. Atlantic Monthly, 176(1), 101-108.

Callan, J. P., Croft, W. B., & Broglio, J. (1995). TREC and TIPSTER Experiments With INQUERY. Information Processing & Management, 31(3), 327-343.

Cameron, R. D. (1997). A Universal Citation Database as a Catalyst for Reform in Scholarly Communication.: Simon Fraser University.

Caplan, P., & Arms, W. Y. (1999). Reference Linking for Journal Articles. D-Lib Magazine, 5(7/8, July/August), 18 pages.

Carbonell, J., & Goldstein, J. (1998). The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. Paper presented at the SIGIR 1998, Melbourne, Australia.

Childs, L. C. (1997). Information Extraction Using the NLToolset. Paper presented at the IEEE Military Communications Conference (MILCOM '97) Unclassified Technical Paper Session, Monterey, CA.

Clifton, C., & Steinheiser, R. (1998). Data Mining on Text., 630-635.

Climenson, W. D., Hardwick, M. H., & Jacobson, S. N. (1961). Automatic Syntax Analysis in Machine Indexing and Abstracting. American Documentation, 12(3), 178-183.

Cohen, W. W., & Singer, Y. (1996). Context-Sensitive Learning Methods for Text Categorization. Paper presented at the SIGIR '96 (Special Interest Group Information Retrieval), Zurich, Switzerland.

Cooper, H. M. (1984). Integrating Research, A Guide for Literature Reviews Second Edition (Vol. 2). Newbury Park, CA: Sage Publications.

Corston-Oliver, S. H. (1998). Beyond String Matching and Cue Phrases: Improving Efficiency and Coverage in Discourse Analysis (Technical Report MSR-TR-98-66). Redmond, WA: Microsoft.

Coulter, N., Monarch, I., Konda, S., & Carr, M. (1996). An Evolutionary Perspective of software engineering research through co-word analysis (CMU/SEI-96-TR-019). Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University.

Craven, T. C. (1993). A Computer Aided Abstracting Tool Kit. Canadian Journal of Information and Library Science, 18(2), 19-31.

Cronin, B. (1998). Metatheorizing Citation. Scientometrics, 43(1), 48-55.

- Crookes, G. (1986). Towards a Validated Analysis of Scientific Text Structure. Applied Linguistics, 7(1), 57-70.
- Damashek, M. (1995). Gauging Similarity with n-Grams; Language-Independent Categorization of Text. Science, February 10(267), 843-848.
- Danes, F. (Ed.). (1974). papers in functional sentence perspective. Prague: Publishing House of the Czechoslovak Academy of Sciences.
- Daniels, J. J. (1997). Retrieval of Passages for Information Reduction. Unpublished PhD, University of Massachusetts, Amherst.
- Davis, M. W., & Foltz, P. W. (1998, Aug 1-4, 1998). Learning Via Compact Data Representation. Paper presented at the Twentieth Annual Meeting of the Cognitive Science Society (CogSci 98), University of Wisconsin-Madison.
- Dean, J., & Henzinger, M. R. (1999, May 1999). Finding Related Pages in the World Wide Web. Paper presented at the 8th International World Wide Web Conference, Toronto, Canada.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by Latent Semantic Analysis. Journal of the American Society for Information Science, 41(6), 391-- 407.
- Dijk, T. A. v. (1980). Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition. Hillsdale, NJ: Lawrence Erlbaum.
- Dimitroff, A. (1995). Self-Citations in the Library and Information Science Literature. Journal of Documentation, 51(1), 44-56.
- Dolin, R., Agrawal, D., Abbadi, A. E., & Pearlman, J. (1998). Using Automated Classification for Summarizing and Selecting Heterogeneous Information Sources. D-Lib Magazine.
- Dong, Y. R. (1996). Learning How to Use Citations for Knowledge Transformation: Non-Native Doctoral Students' Dissertation Writing in Science. Research in the Teaching of English, 30(4), 428-457.
- Earl, L. L. (1970). Experiments in Automatic Extracting and Indexing. Information Storage and Retrieval, 6, 313-334.
- Edmundson, H. P. (1999). New Methods in Automatic Extracting. In I. Mani & M. Maybury (Eds.), Advances in Automatic Text Summarization (pp. 22-42). Cambridge, MA: MIT Press.
- Embley, D. W., Campbell, D. M., Jiang, Y. S., Liddle, S. W., Lonsdale, D. W., Ng, Y. K., & Smith, R. D. (1999). Conceptual-model-based data extraction from multiple-record Web pages. Data & Knowledge Engineering, 31, 227-251.

Endres-Niggemeyer, B. (1993, December 13-17, 1993). A Naturalistic Model of Abstracting. Paper presented at the Preprints of Summarizing Text for Intelligent Communication. Dagstuhl Seminar Report 79, Schloss Dagstuhl, Germany.

Endres-Niggemeyer, B. (1998). Summarizing Information. New York: Springer-Verlag.

Endres-Niggemeyer, B. (2000). SimSum: An Empirically Founded Simulation of Summarizing. Information Processing and Management, 36(4), 659-682.

Endres-Niggemeyer, B., & Neugebauer, E. (1995, May 2-6, 1995). Professional Summarising: No Cognitive Simulation Without Observation. Paper presented at the International Conference in Cognitive Science 1995, San Sebastian.

Endress-Niggemeyer, B. (1993, Dec 13-17, 1993). A Research Platform for Intelligent Summarizing. Paper presented at the Dagstuhl Seminar on Summarizing Text for Intelligent Communication, Dagstuhl, Germany.

Firmin, T., & Chrzanowski, M. J. (1999). An Evaluation of Automatic Text Summarization Systems. In I. Mani & M. Maybury (Eds.), Advances in Automatic Text Summarization (pp. 325-336). Cambridge, MA: MIT Press.

Foltz, P. W. (1990, April 1990). Using Latent Semantic Indexing for Information Filtering. Paper presented at the Proceedings of the ACM Conference on Office Information Systems, ACM/SIGOIS, New York.

Foltz, P. W. (1996). Latent Semantic Analysis for text-based research., (28(2)), 197-202.

Foltz, P. W., Landauer, T. K., & Kintsch, W. (1998). The Measurement of Textual Coherence with Latent Semantic Analysis., 25(2&3), 285-307.

Fukuhara, T., Takeda, H., & Nishida, T. (1999). Multiple-text Summarization for Collective Knowledge Formation (WWW): Graduate School of Information Science, Nara Institute of Science and Technology

8916-5 Takayama, Ikoma City, Nara, 630-0101, Japan.

Futrelle, R. P., & Gauch, S. (1993). The Role of Automated Word Classification in the Summarization of the Contents of Sets of Documents. Paper presented at the Second International Conference on Information and Knowledge Management (CIKM '93), Washington, D.C.

Garfield, E. (1955). Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. Science, 122(3159), 108-111.

Garfield, E. (1967). Primordial Concepts, Citation Indexing, and Historio-Bibliography. The Journal of Library History, 2(3), 235-249.

Garfield, E. (1970). Citation Indexing for Studying Science. Essays of an Information Scientist, 1, 132.

Garfield, E. (1983). Citation Indexing: Its Theory and Application in Science, Technology, and Humanities (2nd Edition ed.). Philadelphia: ISI Press.

Garfield, E. (1996). When to Cite. Library Quarterly, 66(4), 449-458.

Garfield, E. (1997). Concept of Citation Indexing: a Unique and Innovative Tool for Navigating the Research Literature. Paper presented at the Presentation (Unpublished), Far Eastern State University, Vladivostok, Russia.

Garfield, E., & Sher, I. H. (1963). New Factors in the Evaluation of Scientific Literature Through Citation Indexing. American Documentation, 14(8).

Garfield, E., & Small, H. (1989). Michael J. Moravcsik: Multidimensional Scholar and Hero of Third World Science. Scientometrics, 20(1), 19-24.

Garner, R. (1982). Efficient Text Summarization: Costs and Benefits. Journal of Educational Research, 75, 275-279.

Gibb, F. (1993). Knowledge-based Indexing SIMPR: Integration of Natural Language Processing and Principles of Subject Analysis in an Automated Indexing System. Journal Of Document and Text Mangement, 1(2), 131-153.

Gilbert, G. N. (1977). Referencing As Persuasion. Social Studies of Science, 7, 113-122.

Giles, C. L., Bollacker, K. D., & Lawrence, S. (1998). CitedSeer: An Automatic Citation Indexing System. Paper presented at the Digital Libraries 98 - Third ACM Conference on Digital Libraries, New York.

Goldstein, J. (1999). Automatic Text Summarization of Multiple Documents. Unpublished PhD, Carnegie Mellon, Pittsburgh.

Goldstein, J., Kantrowitz, M., Mittal, V., & Carbonell, J. (1999). Summarizing Text Documents: Sentence Selection and Evaluation Metric. Paper presented at the ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-99), Berkeley, CA.

Griffith, B. C., Small, H. G., Stonehil, J. A., & Dey, S. (1974). The Structure of Scientific Literatures II: Toward a Macro- and Microstructure for Science. Science Studies, 4, 339-365.

Grishman. Information Extraction: Techniques and Challenges.

Grobelnik, M., & Mladenic, D. (1998). Efficient Text Categorization. Paper presented at the European Conference on Machine Learning Workshop on Text Mining, Chemnitz, Germany.

Grosz, B. J., & Sidner, C. L. (1986). Attention, intention, and the structure of discourse. Computational Linguistics, 12(3).

Hahn, U. (1990). TOPIC Parsing: Accounting for Text macro structures in full-text analysis. Information Processing and Management, 26(1), 135-170.

Hahn, U. (1993, Dec 13-17 1993). Concept-Oriented Summarizing in the Text Condensation System TOPIC. Paper presented at the Dagstuhl Seminar on Summarizing Text for Intelligent Communication, Dagstuhl, Germany.

Hare, V. C., & Borchardt, K. (1984). Direct Instruction of Summarization Skills. Reading Research Quarterly, 20, 62-78.

Hargens, L. L. (2000). Using the Literature: Reference networks, reference contexts, and the social structure of scholarship. American Sociological Review, 65(6), 846-865.

Harter, S. P., & Kim, H. J. (1996, May 20 - 22). Electronic Journals and Scholarly Communication: A Citation and Reference Study. Paper presented at the American Society for Information Science Midyear Meeting, San Diego.

Hasher, L., & Zacks, R. T. (1984). Automatic Processing of Fundamental Information: The Case of Frequency of Occurrence. American Psychologist, December, 1372-1388.

Hatzvassilioglou, V., Klavans, J. L., HOLcombe, M. L., Barzilay, R., Kan, M.-Y., & McKeown, K. R. SIMFINDER: A Flexible Clustering Tool Summarization.

Hazewinkel, M. (1998). Topologies and Metrics on Information Spaces. CWI Quarterly, 12(2).

Hearst, M., Kopec, G., Brotsky, D., & PARC, S. (1996). Research in Support of Digital Libraries at Xerox PARC. Part II: Paper and Digital Documents. D-Lib Magazine, issn 1082-9873, June, 17 pages.

Hearst, M. A., Kopec, G., & Brotsky, D. (1996). Research in Support of Digital Libraries at Xerox PARC II: Paper and Digital Documents. D-Lib Magazine.

Hearst, M. A., & Plaunt, C. (1993). Subtopic Structuring for Full Length Document Access. Paper presented at the ACM-SIGIR'93, New York.

Helfman, J. I. (1994). Similarity Patterns in Language. Paper presented at the IEEE Symposium on Visual Languages '94: Proceedings of the 10th Institute of Electrical and Electronics Engineers (IEEE) Computer Society Technical Committee on Multimedia Computing Symposium on Visual Languages, St. Louis, MO.

Hellman, E. S. (1999, May 10-12, 1999). Scholarly Link Specification Framework (S-Link S). Paper presented at the The Third ICC/IFIP Conference on Electronic Publishing, Ronneby, Sweden.

- Herring, C. (1968). Distill or Drown: The Need for Reviews. Physics Today, 21(9), 27-33.
- Hitchcock, S., Car, L., Hall, W., Harris, S., Proberts, S., Evans, D., & Brailsford, D. (1998). Linking Electronic Journals: Lessons from the Open Journal Project. D-Lib Magazine, December, 17 pages.
- Hjorland, B., & Albrechtsen, H. (1995). Toward a New Horizon in Information Science: Domain-Analysis. Journal of the American Society for Information Science, 46(6), 400-425.
- Hobbs, J. R. (1993, December 13--17 1993). Summaries from Structure. Paper presented at the In Working Notes of the Dagstuhl Seminar on Summarizing Text for Intelligent Communication, Dagstuhl, Germany.
- Hoch, R. (1995, August 14-15, 1995). READLEX: A Lexicon for the Recognition and the Analysis of Structured Documents. Paper presented at the Proceedings of the Third International Conference on Document Analysis and Recognition (Vol. 2), Montréal, Canada.
- Hoffman, T. Learning and Representing Topic: A Hierarchical Mixture Model for Word Occurrences in Document Databases. Available: Available from CiteSeer; <http://citeseer.nj.nec.com/cachedpage/119738/1>.
- Hovy, E. (1988). Generating Natural Language Under Pragmatic Constraints. Hillsdale, NJ: Lawrence Erlbaum & Associates.
- Hovy, E. (1993). Automated discourse generation using discourse structure relations. Artificial Intelligence, 63, 341-385.
- Hovy, E., & Lin, C.-Y. (1999). Automated Text Summarization in SUMMARIST. In I. Mani & M. Maybury (Eds.), Advances in Automatic Text Summarization (pp. 82-94). Cambridge, MA: MIT Press.
- Hutchins, J. (1993). Introduction to "Text Summarization" Workshop. In B. Endres-Niggemeyer, Hobbs, J., and Sparck Jones, K. (Ed.), Workshop on Summarizing Text for Intelligent Communication. Dagstuhl: AAAI Press.
- Jacobs, P. S. (1992). TRUMP: a transportable language understanding program. international journal intelligent systems, 7, 245-276.
- Jacobs, P. S. (1993). Innovations in text interpretation. Artificial Intelligence, 63, 143-191.
- Jacobs, P. S., & Rau, L. F. (1990). SCISOR: Extracting Information from On-Line News. Communications of the ACM, 33(11), 88-97.

Jing, H., Barzilay, R., McKeown, K., & Elhadad, M. (1998). Summarization Evaluation Methods: Experiments and Analysis. In AAAI Spring Symposium on Intelligent Text Summarization, Technical Report SS-98-06 of the AAAI. Stanford, CA: Stanford University.

Jing, H., & McKeown, K. R. Cut and Paste Based Text Summarization, [PDF]. Available: <http://citeseer.nj.nec.com/cachedpage/294438/12002>].

Jones, K. S. (1993, Nov 1999). Summarising: Analytic Framework, Key Component, Experimental Method; SERC Grant GR/F 2647. Available: <http://www.ik.fh-hannover.de/ik/projekte/Dagstuhl/Abstract/Abstracts/Sparck-Jones/Sparck-Jones.html> [1999, 12/15/99].

Jones, K. S. (1999). Automatic Summarising: Factors and Directions. In I. Mani & M. Maybury (Eds.), Advances in Automatic Text Summarization (pp. 1-12). Cambridge, MA: MIT Press.

Justeson, J. S., & Katz, S. M. (1995). Technical Terminology: Some Linguistic Properties and an Algorithm for Identification in Text. Natural Language Processing, 1(1), 9-27.

Kando, N. (1997). Text-level Structure of Research Papers: Implications for Text-Based Information Processing Systems. Paper presented at the Proceedings of the British Computer Society Annual Colloquium of Information Retrieval Research (BCS-IRSG Colloquium 1997).

Kando, N. (1999, Nov 11-12 1999). Text Structure Analysis as a Tool to Make Retrieved Documents Useable. Paper presented at the IRAL '99, Taipei, Taiwan.

Kaplan, N. (1965). The Norms of Citation Behavior: Prolegomena to the Footnote. American Documentation, 16(3), 179-184.

Kintsch, E., Steinhart, D., Stahl, G., & Group, L. R. (2000). Developing Summarization Skills through the Use of LSA-Based Feedback. Interactive Learning Environments, 8(2), 87-109.

Kintsch, W. (2000). Metaphor Comprehension: A Computational Theory. Psychonomic Bulletin and Review, 7, 257-266.

Kintsch, W. (2001). Predication. Cognitive Science, 25, 173-202.

Kintsch, W., & Dijk, T. A. v. (1978). Toward a Model of Text Comprehension. Psychological Review, 85, 363-394.

Klavans, J. L., McKeown, K. R., Kan, M.-Y., & Lee, S. Resources for Evaluation of Summarization Techniques.

Knauf, R., & Gonzalez, A. J. A Turing Test Approach to Intelligent System Validation.

- Kupiec, J., Pedersen, J., & Chen, F. (1995). A Trainable Document Summarizer. In P. I. Edward A. Fox, and Raya Fidel (Ed.), Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 68-73). Seattle, WA: ACM.
- Laham, D. (1997). Latent Semantic Analysis Approaches to Categorization. In M. G. Shafto & P. Langley (Eds.), Proceedings of the 19th Annual Meeting of the Cognitive Science Society (CogSci) (pp. 979): Lawrence Erlbaum Associates.
- Landauer, T. K., & Dumais, S. T. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge. Psychological Review, 104(2), 211-240.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An Introduction to Latent Semantic Analysis. In R. F. Lorch & E. J. O'Brien (Eds.), Discourse Processes (pp. 259-284).
- Landauer, T. K., Laham, D., & Foltz, P. (1998). Learning Human-like Knowledge by Singular Value Decomposition: A Progress Report. In M. I. Jordan (Ed.), Advances in Neural Information Processing Systems 10 (pp. 45-51). Cambridge: MIT Press.
- Landauer, T. K., Laham, D., Rehder, B., & Schreiner, M. E. (1997). How Well Can Passage Meaning be Derived without Using Word Order? A Comparison of Latent Semantic Analysis and Humans, Proceedings of the 19th Annual Conference of the Cognitive Science Society (CogSci) (pp. 412--417): Lawrence Erlbaum Associates.
- Lawrence, S., Bollacker, K., & Giles, C. L. (1999). Indexing and Retrieval of Scientific Literature. Paper presented at the Eighth International Conference on Information and Knowledge Management, Kansas City, Missouri.
- Lawrence, S., Giles, C. L., & Bollacker, K. (1999). Digital Libraries and Autonomous Citation Indexing. IEEE Computer, 32(6), 67-71.
- Lehman, A. (1999). Text structuration leading to an automatic summary system: RAFI. Information Processing and Management, 35(2), 181-191.
- Lehnert, W., Cardie, C., Fisher, D., McCarthy, J., Riloff, E., & Soderland, S. (1994). Evaluation of Information Extraction Systems. Integrated Computer-Aided Engineering, 1(6), 453-472.
- Lehnert, W. G. (1981). Plot Units and Narrative Summarization. Cognitive Science, 4, 293-331.
- Lehnert, W. G., & Ringle, M. H. (1982). Strategies for Natural Language Processing. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Lehnert, W. G., & Sundheim, B. A. (1991). A Performance Evaluation of Text Analysis Technologies. AI Magazine, 12(3), 81-94.

- Lewis, D. D., & Croft, W. B. (1990). Term Clustering of Syntactic Phrases. Paper presented at the ACM SIGIR Conference on Research and Development in Information Retrieval, Brussels,.
- Leydesdorff, L., & P. Wouters. (1999). Between Texts and Contexts: Advances in Theories of Citation? (A Rejoinder). Scientometrics, 44(2), 169-182.
- Liddy, E. D. (1991). Discourse-Level Structure of Empirical Abstracts: An Exploratory Study. Information Processing and Management, 27(1), 55-81.
- Liddy, E. D. (1993, December 13--17 1993). Development and Implementation of a Discourse Model for Newspaper Texts. Paper presented at the In Working Notes of the Dagstuhl Seminar on Summarizing Text for Intelligent Communication., Dagstuhl, Germany.
- Lin, C.-Y. (1998, March 1998). Assembly of Topic Extraction Modules in SUMMARIST. Paper presented at the In Symposium on Intelligent Text Summarization, Stanford, California.
- Lin, C.-Y., & Hovy, E. Identifying Topics by Position.
- Line, M. B. (1981). The Structure of Social Science Literature as Shown by Large-Scale Citation Analysis. Social Science Information Studies, 1, 67-87.
- Liu, M. (1993). Progress in Documentation The Complexities of Citation Practice: A Review of Citation Studies. Journal of Documentation, 49(4), 370-408.
- Luhn, H. P. (1999). The Automatic Creation of Literature Abstracts. In I. Mani & M. Maybury (Eds.), Advances in Automatic Text Summarization (pp. 15-21). Cambridge, MA: MIT Press.
- Luukkonen, T. (1997). Why has Latour's Theory of Citations been Ignored by the Bibliometric Community? Discussion of Sociological Interpretations of Citation Analysis. Scientometrics, 38(1), 27-37.
- Mani, I., & Bloedorn, E. (1997). Multi-document Summarization by Graph Search and Matching. In Proceedings of the Fourteenth National Conference on Artificial Intelligence (pp. 622--628). Providence, Rhode Island.
- Mani, I., & Bloedorn, E. (1999). Summarizing Similarities and Differences Among Related Documents. In I. Mani & M. Maaybury (Eds.), Advances in Automatic Text Summarization (pp. 357-379). Cambridge, MA: MIT Press.
- Mani, I., Firmin, T., & Sundheim, B. (1999). The TIPSTER SUMMAC Text Summarization Evaluation, Ninth Conference of the European Chapter of the ACL (pp. 77--85). Bergen, Norway: Association for Computational Linguistics.

Mani, I., & Maybury, M. T. (Eds.). (1999). Advances in Automatic Text Summarization. Cambridge, MA: The MIT Press.

Marcu, D. (1997). From Discourse Structures to Text Summaries, ACL'97/EACL'97 Workshop on Intelligent Scalable Text Summarization (pp. 82--88). Madrid, Spain: Association for Computational Linguistics and the European Chapter of the Association for Computational Linguistics.

Marcu, D. (1997). The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts. Unpublished PhD, University of Toronto, Toronto.

Marcu, D. (1997). The Rhetorical Parsing of Natural Language Texts, ACL'97/EACL'97 Workshop on Intelligent Scalable Text Summarization (pp. 96-103). Madrid, Spain: Association for Computational Linguistics and the European Chapter of the Association for Computational Linguistics.

Marcu, D. (1998, March 1998). To Build Text Summaries of High Quality, Nuclearity Is Not Sufficient. Paper presented at the In Working Notes of the the AAAI-98 Spring Symposium on Intelligent Text Summarization, Stanford, CA.

Marcu, D. (1999). The Automatic Construction of Large-Scale Corpora for Summarization Research, SIGIR'99. University of Berkeley, CA.

Marcu, D. (1999). Discourse Trees are Good Indicators of Importance in Text. In I. Mani & M. Maybury (Eds.), Advances in Automatic Text Summarization (pp. 123-136): The MIT Press.

Maybury, M. (1999). Scalable Text Summarization: Wrap-up, [http://www-i.mitre.org/resources/centers/advanced_info/mark.html] [1999].

Maybury, M. T. (1995). Generating summaries from event data. Information Processing and Management, 31(5), 735-751.

McCain, K. W. (1989). Citation Context Analysis and Aging Patterns of Journal Articles in Molecular Genetics. Scientometrics, 17(1-2), 127-163.

McCain, K. W. (1994). Islands in the Stream: Mapping the Fisheries & Aquatic Sciences Literatures. Fisheries, 19(10), 20-27.

McCain, K. W. (2000). Sharing Digitized Research-Related Information on the World Wide Web. Journal of the American Society for Information Science, 51(14), 1321-1327.

McHale, M. L. (1998, August, 1998). A Comparison of WordNet and Roget's Taxonomy for Measuring Semantic Similarity. Paper presented at the COLING/ACL Workshop on Usage of WordNet in Natural Language Processing Systems, Montreal, Canada.

McKeown, K. (1993, December 13-17, 1993). Generating the Complex Sentences of Summaries Using Syntactic and Lexical Constraints: Two Applications. Paper presented at the Summarizing Text for Intelligent Communication. Dagstuhl Seminar Report 79, Schloss Dagstuhl, Germany.

McKeown, K., & Radev, D. R. (1995). Generating Summaries of Multiple News Articles. Paper presented at the Proceedings of ACM-SIGIR 1995, Seattle, Washington.

McKeown, K., Robin, J., & Radev, D. R. (1995). Generating concise natural language summaries. Information Processing and Management, 31(5), 703-733.

McKeown, K. R., Jordan, D. A., & Hatzivassiloglou, V. (1998). Generating Patient-Specific Summaries of Online Literature. Paper presented at the AAAI 1998 Spring Symposium on Intelligent Text summarization.

Mitra, M., Singhal, A., & Buckley, C. (1997, July 11, 1997). Automatic Text Summarization by Paragraph Extraction. Paper presented at the ACL'97/EACL'97 Intelligent Scalable Text Summarization Workshop, Madrid, Spain.

Mittal, V. O., Kantrowitz, M., Goldstein, J., & Carbonell, J. (1999). Selecting Text Spans for Document Summaries: Heuristics and Metrics.

Moens, M. F., Uyttendaele, C., & Dumortier, J. (1999). Information Extraction from Legal Texts: The Potential Discourse Analysis. International Journal of Human-Computer Studies, 51(6), 1155-1171.

Moravcsik, M. J., & Murugesan, P. (1975). Some Results on the Function and Quality of Citations. Social Studies of Science, 5, 86-92.

Moris, A. H., Kasper, G. M., & Adams, D. A. (1992). The Effects and Limitations of Automated Text Condensing on Reading Comprehension Performance. Information Systems Research, 3(1), 17-35.

Mushakoji, S. (1993). Constructing 'Identity' and 'Differences' in Original Scientific Texts and Their Summaries: Its Problems and Solutions. Dagstuhl, Germany.

Myaeng, S. H., & Jang, D. (1999). Development and Evaluation of a Statistically-Based Document Summarization System. In I. Mani & M. Maybury (Eds.), Advances in Automatic Text Summarization (pp. 61-70). Cambridge, MA: MIT Press.

Nanba, H., Kando, N., & Okumura, M. (2000, Nov.12, 2000). Classification of Research Papers Using Citation Links and Citation Types: Towards Automatic Review Article Generation. Paper presented at the 12th American Society for Information Science SIG/CR Workshop, Chicago, IL.

Nanba, H., & Okumura, M. (1999). Towards Multi-paper Summarization Using Reference Information. Paper presented at the Proceedings of the 16th International Joint Conferences on Artificial Intelligence(IJCAI-99).

Nanba, H., & Okumura, M. (2000). Producing More Readable Extracts by Revising Them. Paper presented at the 18th International Conference on Computational Linguistics (COLING 2000), Saarbrücken, Germany.

O'Connor, J. (1983). Biomedical Citing Statements: Computer Recognition and Use to Aid Full-Text Retrieval. Information Processing & Management, 19(6), 361-368.

Ono, K., Sumita, K., & Miike, S. (1994). Abstract Generation Based on Rhetorical Structure Extraction. Proceedings of the 16th International Conference on Computational Linguistics, 344-348.

Osareh, F. (1996). Bibliometrics, Citation Analysis and Co-Citation Analysis: A Review of Literature II. Libri, 46, 217-225.

Paice, C. D. (1990). Constructing Literature Abstracts by Computer: Techniques and Prospects. Information Processing and Management, 26(1), 171-186.

Paik, W. Chronological Information Extraction System (CIES). Available: citeseer.nj.nec.com/371086.html.

Pao, M. L. (1989). Concepts of Information Retrieval. Englewood, CO: Libraries Unlimited.

Peritz, B. C. (1992). On the Objectives of Citation Analysis: Problems of Theory and Method. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE, 449-451.

Perrin, A. J. (2000, August 16, 2000). CodeRead: A Multiplatform coding engine for text-based data. Paper presented at the American Sociological Association annual meeting, Washington, D.C.

Polanyi, L. (1993). Linguistic Dimensions of Text Summarization. In B. Endres-Niggemeyer & J. Hobbs & K. S. Jones (Eds.), Workshop on Summarising Text for Intelligent Communication - Dagstuhl Seminar Report (9350). Dagstuhl, Germany.

Radev, D., & McKeown, K. (1998). Generating Natural Language Summaries from Multiple On-line Sources. Computational Linguistics, 24(3), 469-500.

Rath, G. J., Resnick, A., & Savage, T. R. (1961). The Formation of Abstracts by the Selection of Sentences: Part 1, Sentence Selection by Men and Machines. American Documentation, 12(2), 139-141.

Rau, L. F. (1987). Knowledge organization and access in a conceptual information system. Information Processing and Management, 23(4).

Rau, L. F., Jacobs, P. S., & Zernik, U. (1989). information extraction and text summarization using linguistic knowledge acquisition. information processing and management, 24(4), 419-428.

- Reder, L. M., & Anderson, J. R. (1980). A Comparison of Texts and Their Summaries: Memorable Consequences. Journal of Verbal Learning and Verbal Behavior, 19, 121-134.
- Reimer, U., & Hahn, U. (1988, March 1988). Text Condensation as Knowledge Base Abstraction. Paper presented at the Fourth Conference on Artificial Intelligence Applications.
- Rijsbergen, C. J. v. Information Retrieval.
- RikPietersa, HansBaumgartnerb, JeroenVermuntc, & TammoBijmolta. (1999). Importance and similarity in the evolving citation network of the International Journal of Research in Marketing. Intern. J. of Research in Marketing, 16, 113-127.
- Riloff, E., & Lehnert, W. (1994). Information Extraction as a Basis for High-Precision Text Classification. ACM Transactions on Information Systems, 12(3), 296-333.
- Rino, L. H. M., & Scott, D. (1994, July 1994). Content Selection in Summary Generation. Paper presented at the Third International Conference on the Cognitive Science of Natural Language Processing, Dublin City University, Ireland.
- Robin, J. (1994). Revision-Based Generation of Natural Language Summaries Providing Historical background
- Corpus-Based Analysis, Design, Implementation and Evaluation (CUCS-03-94). New York: Columbia University.
- Rosemarie, G. (1999). Summarizing Texts as Genres of Academic Writing [1999, 1/15/1999].
- Rousseau, R. (1999). Temporal Differences in Self-Citation Rates of Scientific Journals. Scientometrics, 44(3), 521-531.
- Ruthven, I., Tombros, A., & Jose, J. M. A Study on the Use of Summaries and Summary-Based Query Expansion for Question-Answering Task. ECIR 01.
- Salton, G. (Ed.). (1971). The SMART Retrieval System, Experiments in Automatic Document Processing. Englewood Cliffs, NJ: Prentice-Hall.
- Salton, G. (1989). automatic text processing. The transformation analysis and retrieval of information by computer. New York: Addison-Wesley.
- Salton, G., & Allan, J. (1993, November 1993). Selective Text Utilization and Text Traversal. Paper presented at the Hypertext '93.
- Salton, G., Allan, J., Buckley, C., & Singhal, A. (1994). Automatic Analysis, Theme Generation, and Summarization of Machine Readable Texts. Science, 264(June), 1421-1426.

- Salton, G., Allan, J., & Singhal, A. (1996). Automatic Text Decomposition and Structuring. Information Processing and Management, 32(2), 127-138.
- Salton, G., & Araya, J. (1989). On the Use of Clustered File Organization in Information Search and Retrieval (TR89-989). Ithaca: Cornell University.
- Salton, G., Bergmark, D., & Wong, A. Generation and Search of Clustered Files. Ithaca, NY: Cornell University, Department of Computer Science.
- Salton, G., & Buckley, C. (1987). Term Weighting Approaches in Automatic Text Retrieval (87-881). Ithaca: Cornell.
- Salton, G., & McGill, M. (1983). Introduction to Modern Information Retrieval. New York: McGraw-Hill.
- Salton, G., & Singhal, A. (1994). Automatic Text Theme Generation and the Analysis of Text Structure (TR94-1438). Ithaca, NY: Cornell University, Department of Computer Science.
- Salton, G., Singhal, A., Buckley, C., & Mitra, M. (1995). Automatic Text Decomposition Using Text Segments and Text Themes (TR 95-1555): Cornell University Technical Report.
- Salton, G., Singhal, A., Mitra, M., & Buckley, C. (1997). Automatic Text Structuring and Summarization. Information Processing and Management, 33(2), 193-208.
- Salton, G., Yang, C. S., & Yu, C. T. (1975). A Theory of Term Importance in Automatic Text Analysis. Journal of the American Society fo Information Science, 18(11), 613-620.
- Salton, G. a. B., C. (1988). Term weighting approaches in automatic text retrieval. Information Processing & Management, 24(5), 513-523.
- Savoy, J. Citation Schemes in Hypertext Information Retrieval.
- Schutzer, D. (1996). A Need For A Common Infrastructure. D-Lib Magazine.
- Sengupta, I. N. (1992). Bibliometrics, Informetrics, Scientometrics and Librametrics: An Overview. Libri, 42(2), 75-98.
- Shadish, W. R., Tolliver, D., Gray, m., & Gupta, S. K. S. (1995). Author Judgments about Works They Cite: Three Studies from Psychology Journals. Social Studies of Science, 25, 477-498.
- Singhal, A., Buckley, C., and Mitra, M. (1996). Pivoted Document Lenght Normalization. Paper presented at the Proceedings of the 19th SIGIR Conference on Research and Development in Information Retrieval (SIGIR '96).

Skorohodko, E. F. (1971). Adaptive Method of Automatic Abstracting and Indexing. Paper presented at the Proceeding IFIP Conference, Ljubljana.

Small, H. (1980). Co-Citation Context Analysis and the Structure of Paradigms. Journal of Documentation, 36(3), 183-196.

Small, H. (1982). Citation Context Analysis. In B. J. Dervin & M. J. Voight (Eds.), Progress in Communication Sciences (Vol. 3, pp. 287-310). Norwood, NJ: Ablex.

Small, H. (1984). The Lives of a Scientific Paper. In K. Warren (Ed.), Selectivity in Information Systems (pp. 83-97). New York: Praeger.

Small, H. (1986). The Synthesis of Specialty Narrative from Co-citation Clusters. Journal of the American Society for Information Science, 37(3), 97-110.

Small, H. (1995, October 9-12, 1995). Navigating the Citation Network. Paper presented at the The 58th ASIS Annual Meeting (ASIS '95), Chicago, IL.

Small, H. (1999). Visualizing Science by Citation Mapping. Journal of the American Society for Information Science.

Small, H., & Greenlee, E. (1980). Citation Context Analysis of a Co-Citation Cluster: Recombinant-DNA. Scientometrics, 2(4), 277-301.

Small, H., & Greenlee, E. (1989). A Co-Citation Study of AIDS Research., 16(5), 642-666.

Small, H., & Griffith, B. G. (1974). The Structure of Scientific Literatures I: Identifying and Graphing Specialities. Science Studies, 4, 17-40.

Small, H., & Sweeney, E. (1985). Clustering the science citation index using co-citation 1. A comparison of methods. Scientometrics, 7, 391-409.

Small, H., Sweeney, E., & Greenlee, E. (1985). Clustering the Science Citation Index Using Co-Citations. II. Mapping Science. Scientometrics, 8(5-6), 321-340.

Small, H. G. (1978). Cited Documents as Concept Symbols. Social Studies of Science, 8, 327-340.

Smith, P. D. (1990). An Introduction to Text Processing. Cambridge, MA: MIT Press.

Snyder, H., & Bonzi, S. (1998). Patterns of Self-Citation Across Disciplines (1980-1989). Journal of Information Science, 24(6), 431-435.

Stein, G., Bagga, A., & Wise, G. B. (2000, Oct 16-18). Multi-Document Summarization: Methodologies and Evaluations. Paper presented at the Conference TALN 2000, Lusanne, Switzerland.

- Strzalkowski, T., Stein, G., Wang, J., & Wise, B. (1999). A Robust Practical Text Summarizer. In I. Mani & M. Maybury (Eds.), Advances in Automatic Text Summarization (pp. 137-154). Cambridge, MA: MIT Press.
- Strzalkowski, T., Stein, G., Wise, G. B., Perez-Carballo, J., Tapanainen, P., Jarvinen, T., Voutilainen, A., & Karlgren, J. (1998, November 9-11 1998). Natural Language Information Retrieval. Paper presented at the Text Retrieval Conference (TREC-7), Gaithersburg, Maryland.
- Teufel, S., Carletta, J., & Moens, M. (1999). An Annotation Scheme for Discourse Level Argumentation in Research Articles. Paper presented at the Ninth Conference of the European Chapter of the Association for Computational Linguistics (EACL-99).
- Teufel, S., & Moens, M. (1997, July 11, 1997). Sentence Extraction as a Classification Task. Paper presented at the ACL/EACL'97
- Teufel, S., & Moens, M. (1999). Argumentative Classification of Extracted Sentences as a First Step Towards Flexible Abstracting. In I. Mani & M. Maybury (Eds.), Advances in Automatic Text Summarization (pp. 155-171). Cambridge, MA: MIT Press.
- Titscher, S., Meyer, M., Wodak, R., & Vetter, E. (2000). Methods of Text and Discourse Analysis. Thousand Oaks, CA: Sage Publications.
- Tsujimoto, S., & Asada, H. (1992). Major Components of a Complete Text Reading System. Proceedings of the IEEE, 80(7), 1133-1149.
- Turney, P. D. Learning Algorithms for Keyphrase Extraction. submitted to Information Retrieval.
- Turney, P. D. (1997). Extraction of Keyphrases from Text: Evaluation of Four Algorithms (ERB-1051): National Research Council of Canada.
- Turney, P. D. (1999). Learning to Extract Keyphrases from Text (ERB-1057): National Research Council of Canada.
- Voorhees, E. M. (1998). Variations in relevance Judgements and the measurement of retrieval effectiveness. Paper presented at the 21st ACM-SIGIR international conference on research and development in information retrieval, Melbourne, Australia.
- Voos, H., & Dagaev, K. S. (1976). Are All Citations Equal? Or, Did We Op. Cit. Your Idem? The Journal of Academic Librarianship, 1(6), 19-21.
- Wang, P., & White, M. D. (1995, October 9-12, 1995). Document Use During a Research Project: A Longitudinal Study. Paper presented at the The 58th ASIS Annual Meeting (ASIS '95), Chicago, IL.
- Wang, P., & White, M. D. (1996). A Qualitative Study of Scholars' Citation Behavior. Paper presented at the The 59th ASIS Annual Meeting (ASIS '96), Baltimore, MD.

Ward, S. A., & Reed, L. J. (Eds.). (1983). Knowledge Structure and Use: Implications for Synthesis and Interpretation. Philadelphia: Temple University Press.

Wee, L. K., Tong, L. C., & Tan, C. L. (1998). Knowledge Representation Issues in Information Extraction. Kent Ridge Digital Labs, 15(31), 448-458.

White, M. D., & Wang, P. (1997). A Qualitative Study of Citing Behavior: Contributions, Criteria, and Metalevel Documentation Concerns. Library Quarterly, 67(2), 122-154.

Winograd, P. (1984). Strategic Difficulties in Summarizing Texts. Reading Research Quarterly, 19(4), 404-425.

Wissmann, C. (1993). Techniques of Data Retrieval for Scientometric Research in the ISI Citation Indexes. Journal of Information Science, 19(5), 363-376.

Wolfe, M. B., Schreiner, M. E., Rehder, B., Laham, D., Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). Learning from text: Matching readers and texts by Latent Semantic Analysis. (25), 309-336.

Zechner, K. (1997). Literature Survey of Information Extraction and Text Summarization.

Appendix A

Survey Instrument

INSTRUCTIONS: Read the abstract below, then read each of the citations listed below and answer the associated questions. The objective is to determine which of the extracted citation text best represents the meaning of the abstract.

Generating Summaries of Multiple News Articles

Kathleen McKeown and Dragomir R. Radev

ABSTRACT

We present a natural language system which summarizes a series of news articles on the same event. It uses summarization operators, identified through empirical analysis of a corpus of news summaries, to group together templates from the output of the systems developed for ARPA's Message Understanding Conferences. Depending on the available resources (e.g., space), summaries of different length can be produced. Our research also provides a methodological framework for future work on the summarization task and on the evaluation of news summarization systems.

CITATION:**138. Document Fusion for Comprehensive Event Description — Christof Monz Institute**

... comprehensive document, containing the information of all original documents without repeating information which is conveyed by two or more documents. **The work described in this paper is closely related to the area of multi document summarization (Barzilay et al. 1999; Mani and Bloedorn, 1999; McKeown and Radev, 1995; Radev, 2000) where related documents are analyzed to use frequently occurring segments for identifying relevant information that has to be included in the summary.** Our work differs from the work on multi document summarization as we focus on document fusion disregarding summarization. **On the ...**

Subject Matter: How well does the above citation represent the general subject area of the abstract?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Purpose: Does the above citation represent or identify the abstract's main purpose? Here are some examples of possible purposes: present original research findings, discuss theory, or survey the work of others.

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Methods: Does the citation above indicate the methods used in the research?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Conclusions or Findings: Does the citation indicate any research findings or conclusions?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Generalization or Implications: Does the citation indicate the significance of the research and its influence on the larger technical problem or theory?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Readability: How readable is the citation?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Understandability: How understandable is the citation?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

CITATION:**139. Towards an ANN-based Approach to Automatic Sentence Extraction — Li Mu Gao**

... Automatic Summarization Automatic Abstracting is receiving more and more attention of NLP researchers along with the IE (Information Extraction) IR (Information Retrieval) and IF (Information Filtering) technique recently. **Many automatic abstracting systems have been proposed. For example, SUMMONS [McKeown et al., 1995; Radev et al., 1998] SUMMARIST [Hovy et al., 1997; Lin, 1998] COSYMATS [Aretoulaki, 1997] SUMMAC [Sanderson, 1998] SJTUCAA [Wang et al., 1996] FDASCT [Wu et al., 1996] and so on.** Tombros (1997) presented a general automatic text abstracting model which generates the abstract of the text in two. ...

Subject Matter: How well does the above citation represent the general subject area of the abstract?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Purpose: Does the above citation represent or identify the abstract's main purpose? Here are some examples of possible purposes: present original research findings, discuss theory, or survey the work of others.

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Methods: Does the citation above indicate the methods used in the research?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Conclusions or Findings: Does the citation indicate any research findings or conclusions?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Generalization or Implications: Does the citation indicate the significance of the research and its influence on the larger technical problem or theory?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Readability: How readable is the citation?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Understandability: How understandable is the citation?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

CITATION:140. Thinksheet: A Tool for Information Navigation — Piatko (1998)

... to produce a summary [KR96] 8 These systems and Thinksheet share the common goal of providing only the relevant information to the reader, but summary generators operate on a different class of documents. **Generally, they have been used to generate summaries of newspaper or magazine articles** [AL97, BE97, **MR95**]. These papers may have complicated subject matter; i.e., they may be technical articles [TM97] but they generally do not meet our criteria for complex documents because the structure of the articles is mostly simple and linear. These systems also do not provide the individual tailoring capability. ...

Subject Matter: How well does the above citation represent the general subject area of the abstract?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Purpose: Does the above citation represent or identify the abstract's main purpose? Here are some examples of possible purposes: present original research findings, discuss theory, or survey the work of others.

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Methods: Does the citation above indicate the methods used in the research?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Conclusions or Findings: Does the citation indicate any research findings or conclusions?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Generalization or Implications: Does the citation indicate the significance of the research and its influence on the larger technical problem or theory?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Readability: How readable is the citation?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Understandability: How understandable is the citation?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

CITATION:

141. Salience-Based Content Characterisation of Text Documents — Boguraev, Kennedy (1997)

... certain core entities and facts in a document, which are packaged together in a template. **There are shared intuitions among researchers that generation of smooth prose from this template would yield a summary of the document's core content; recent work, most notably by McKeown and colleagues, cf. (McKeown Radev 1995), focuses on making these intuitions more concrete.** While providing a rich context for research in generation, this framework requires an analysis front end capable of instantiating a template to a suitable level of detail. **Given** the current state of the art in text analysis in general, and of ...

Subject Matter: How well does the above citation represent the general subject area of the abstract?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Purpose: Does the above citation represent or identify the abstract's main purpose? Here are some examples of possible purposes: present original research findings, discuss theory, or survey the work of others.

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Methods: Does the citation above indicate the methods used in the research?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Conclusions or Findings: Does the citation indicate any research findings or conclusions?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Generalization or Implications: Does the citation indicate the significance of the research and its influence on the larger technical problem or theory?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Readability: How readable is the citation?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Understandability: How understandable is the citation?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

CITATION:142. Using Lexical Chains for Text Summarization — Barzilay, Elhadad (1997)

... representation in order to create a summary. **There** are three types of source text information: linguistic, domain and communicative. **Each** of these text aspects can be chosen as a basis for source representation. **Summaries** can be built on a deep semantic analysis of the source text. **For example, in (McKeown and Radev, 1995), McKeown and Radev investigate ways to produce a coherent summary of several texts describing the same event, when a detailed semantic representation of the source texts is available (in their case, they use MUC style systems to interpret the source texts) Alternatively, early summarization ...**

Subject Matter: How well does the above citation represent the general subject area of the abstract?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Purpose: Does the above citation represent or identify the abstract's main purpose? Here are some examples of possible purposes: present original research findings, discuss theory, or survey the work of others.

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Methods: Does the citation above indicate the methods used in the research?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Conclusions or Findings: Does the citation indicate any research findings or conclusions?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Generalization or Implications: Does the citation indicate the significance of the research and its influence on the larger technical problem or theory?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Readability: How readable is the citation?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Understandability: How understandable is the citation?

| | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Not at all | Slightly | Adequately | Very Well | Completely |
| 0% | 25% | 50% | 75% | 100% |

Appendix B

Document Abstracts

| Abs # | Title | Abstract |
|-------|--------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | A Case-Based Approach to Knowledge Acquisition for Domain-Specific Sentence Analysis | This paper describes a case-based approach to knowledge acquisition for natural language systems that simultaneously learns part of speech, word sense, and concept activation knowledge for all open class words in a corpus. The parser begins with a lexicon of function words and creates a case base of context-sensitive word definitions during a human-supervised training phase. Then, given an unknown word and the context in which it occurs, the parser retrieves definitions from the case base to infer the word's syntactic and semantic features. By encoding context as part of a definition, the meaning of a word can change dynamically in response to surrounding phrases without the need for explicit lexical disambiguation heuristics. Moreover, the approach acquires all three classes of knowledge using the same case representation and requires relatively little training and no hand-coded knowledge acquisition heuristics. We evaluate it in experiments that explore two of many practical applications of the technique and conclude that the case-based method provides a promising approach to automated dictionary construction and knowledge acquisition for sentence analysis in limited domains. In addition, we present a novel case retrieval algorithm that uses decision trees to improve the performance of a k-nearest neighbor similarity metric. |
| 2 | Database Models for Infinite and Indefinite Temporal Information | Representation and querying of temporal information can benefit from the integration of techniques from constraint databases, database models for indefinite information and reasoning about temporal constraints. With this perspective in mind, we present a hierarchy of temporal data models: temporal relations, generalized temporal relations and temporal tables. We study the semantics of these models and develop algebraic and calculus query languages for them. The proposed models can be useful to several novel applications include planning, scheduling, project management, medical information systems, geographical information systems and natural language processing systems. |

| Abs # | Title | Abstract |
|-------|--------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 3 | Deterministic Part-of-Speech Tagging with Finite State Transducers | Stochastic approaches to natural language processing have often been preferred to rule-based approaches because of their robustness and their automatic training capabilities. This was the case for part-of-speech tagging until Brill showed how state-of-the-art part-of-speech tagging can be achieved with a rule-based tagger by inferring rules from a training corpus. However, current implementations of the rule-based tagger run more slowly than previous approaches. In this paper, we present a finite-state tagger inspired by the rule-based tagger which operates in optimal time in the sense that the time to assign tags to a sentence corresponds to the time required to follow a single path in a deterministic finite-state machine. This result is achieved by encoding the application of the rules found in the tagger as a non-deterministic finite-state transducer and then turning it into a deterministic transducer. The resulting deterministic transducer yields a part-of-speech tagger whose speed is dominated by the access time of mass storage devices. We then generalize the techniques to the class of transformation-based systems. |
| 4 | Translation by Quasi Logical Form Transfer | The paper describes work on applying a general purpose natural language processing system to transfer-based interactive translation. Transfer takes place at the level of Quasi Logical Form (QLF), a contextually sensitive logical form representation which is deep enough for dealing with cross-linguistic differences. Theoretical arguments and experimental results are presented to support the claim that this framework has good properties in terms of modularity, compositionality, reversibility and monotonicity. |
| 5 | Natural Language Interfaces to Databases: An Introduction | This paper is an introduction to natural language interfaces to databases (Nlids). A brief overview of the history of Nlids is first given. Some advantages and disadvantages of Nlids are then discussed, comparing Nlids to formal query languages, form-based interfaces, and graphical interfaces. An introduction to some of the linguistic problems Nlids have to confront follows, for the benefit of readers less familiar with computational linguistics. The discussion then moves on to Nlid architectures, portability issues, restricted natural language input systems (including menu-based Nlids), and Nlids with reasoning capabilities. Some less explored areas of Nlid research are then presented, namely database updates, meta-knowledge questions, temporal questions, and multi-modal Nlids. The paper ends with reflections on the current state of the art. |

| Abs # | Title | Abstract |
|-------|---------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 6 | Using Lexical Chains for Text Summarization | <p>We investigate one technique to produce a summary of an original text without requiring its full semantic interpretation, but instead relying on a model of the topic progression in the text derived from lexical chains. We present a new algorithm to compute lexical chains in a text, merging several robust knowledge sources: the WordNet thesaurus, a part-of-speech tagger and shallow parser for the identification of nominal groups, and a segmentation algorithm derived from [8]. Summarization proceeds in three steps: the original text is first segmented, lexical chains are constructed, strong chains are identified and significant sentences are extracted from the text. We present in this paper empirical results on the identification of strong chains and of significant sentences. Preliminary results indicate that quality indicative summaries are produced and are extensively documented in http://www.cs.bgu.ac.il/summarization-test. Pending problems are identified: the need for anaphora resolution, a model for reconstructing a coherent summary out of the selected sentences, a method to handle long sentences and a method to control the degree of condensation of the original text. Plans to address these short-comings are briefly presented.</p> |
| 7 | Shopbots and Pricebots | <p>Shopbots are agents that automatically search the Internet to obtain information about prices and other attributes of goods and services. They herald a future in which autonomous agents profoundly influence electronic markets. In this study, a simple economic model is proposed and analyzed, which is intended to quantify some of the likely impacts of a proliferation of shopbots and other economically-motivated software agents. In addition, this paper reports on simulations of pricebots [adaptive, price-setting agents which firms may well implement to combat, or even take advantage of, the growing community of shopbots. This study forms part of a larger research program that aims to provide insights into the impact of agent technology on the nascent information economy.</p> |
| 8 | Embedding Knowledge in Web Documents | <p>This paper argues for the use of general and intuitive knowledge representation languages (and simpler notational variants, e.g. subsets of natural languages) for indexing the content of Web documents and representing knowledge within them. We believe that these languages have advantages over metadata languages based on the Extensible Markup Language (XML). Indeed, the retrieval of precise information is better supported by languages designed to represent semantic content and support logical inference, and the readability of such a language eases its exploitation, presentation and direct insertion within a document (thus also avoiding information duplication). We advocate the use of Conceptual Graphs and simpler notational variants that enhance knowledge readability. To further ease the representation process, we propose techniques allowing users to leave some knowledge terms undeclared. We also show how lexical, structural and knowledge-based techniques may be combined to retrieve or generate knowledge or Web documents. To support and guide the knowledge modeling approach, we present a top-level ontology of 400 concept and relation types. We have implemented these features in a Web-accessible tool named WebKB (http://meganesia.int.gu.edu.au/~phmartin/WebKB/), and show examples to illustrate them</p> |

| Abs # | Title | Abstract |
|-------|---------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 9 | Results and Challenges in Web Search Evaluation. | A frozen 18.5 million page snapshot of part of the Web has been created to enable and encourage meaningful and reproducible evaluation of Web search systems and techniques. This collection is being used in an evaluation framework within the Text Retrieval Conference (TREC) and will hopefully provide convincing answers to questions such as, "Can link information result in better rankings?", "Do longer queries result in better answers?", and, "Do TREC systems work well on Web data?" The snapshot and associated evaluation methods are described and an invitation is extended to participate. Preliminary results are presented for an effectiveness comparison of six TREC systems working on the snapshot collection against five well-known Web search systems working over the current Web. These suggest that the standard of document rankings produced by public Web search engines is by no means state-of-the-art. |
| 10 | A Machine Learning Architecture for Optimizing Web Search Engines | Indexing systems for the World Wide Web, such as Lycos and Alta Vista, play an essential role in making the Web useful and usable. These systems are based on Information Retrieval methods for indexing plain text documents, but also include heuristics for adjusting their document rankings based on the special HTML structure of Web documents. In this paper, we describe a wide range of such heuristics--including a novel one inspired by reinforcement learning techniques for propagating rewards through a graph ---which can be used to affect a search engine's rankings. We then demonstrate a system which learns to combine these heuristics automatically, based on feedback collected unintrusively from users, resulting in much improved rankings. |
| 11 | Examining the Role of Statistical and Linguistic Knowledge Sources in a General-Knowledge Question-Answering System | We describe and evaluate an implemented system for general-knowledge question answering. The system combines techniques for standard ad-hoc information retrieval (IR), query-dependent text summarization, and shallow syntactic and semantic sentence analysis. In a series of experiments we examine the role of each statistical and linguistic knowledge source in the question-answering system. In contrast to previous results, we find first that statistical knowledge of word co-occurrences as computed by IR vector space methods can be used to quickly and accurately locate the relevant documents for each question. The use of query-dependent text summarization techniques, however, provides only small increases in performance and severely limits recall levels when inaccurate. Nevertheless, it is the text summarization component that allows subsequent linguistic filters to focus on relevant passages. We find that even very weak linguistic knowledge can offer substantial improvements over purely IR- based techniques for question answering, especially when smoothly integrated with statistical preferences computed by the IR subsystems. |

| Abs # | Title | Abstract |
|-------|-------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 12 | Unsupervised Models for Named Entity Classification | <p>This paper discusses the use of unlabeled examples for the problem of named entity classification. A large number of rules is needed for coverage of the domain, suggesting that a fairly large number of labeled examples should be required to train a classifier. However, we show that the use of unlabeled data can reduce the requirements for supervision to just 7 simple "seed" rules. The approach gains leverage from natural redundancy in the data: for many named-entity instances both the spelling of the name and the context in which it appears are sufficient to determine its type. We present two algorithms. The first method uses a similar algorithm to that of (Yarowsky 95), with modifications motivated by (Blum and Mitchell 98). The second algorithm extends ideas from boosting algorithms, designed for supervised learning tasks, to the framework suggested by (Blum and Mitchell 98).</p> |
| 13 | Multilingual lexical representation | <p>We describe an approach to the representation of lexical translation equivalence which allows cross-linguistic generalisations to be expressed within a typed unification-based formalism, and illustrate how a large scale multilingual lexical knowledge base that uses this representation may be constructed semi-automatically from machine readable dictionaries</p> |
| 14 | Forgetting Exceptions is Harmful in Language Learning | <p>We show that in language learning, contrary to received wisdom, keeping exceptional training instances in memory can be beneficial for generalization accuracy. We investigate this phenomenon empirically on a selection of benchmark natural language processing tasks: graphemeto-phoneme conversion, part-of-speech tagging, prepositional-phrase attachment, and base noun phrase chunking. In a first series of experiments we combine memory-based learning with training set editing techniques, in which instances are edited based on their typicality and class prediction strength. Results show that editing exceptional instances (with low typicality or low class prediction strength) tends to harm generalization accuracy. In a second series of experiments we compare memory-based learning and decision-tree learning methods on the same selection of tasks, and find that decision-tree learning often performs worse than memory-based learning. Moreover, the decrease in performance can be linked to the degree of abstraction from exceptions (i.e., pruning or eagerness). We provide explanations for both results in terms of the properties of the natural language processing tasks and the learning algorithms. Keywords: memory-based learning, natural language learning, edited nearest neighbor classifier, decision-tree learning</p> |

| Abs # | Title | Abstract |
|-------|----------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 15 | Word sense disambiguation using a second language monolingual corpus | <p>This paper presents a new approach for resolving lexical ambiguities in one language using statistical data from a monolingual corpus of another language. This approach exploits the differences between mappings of words to senses in different languages. The paper concentrates on the problem of target word selection in machine translation, for which the approach is directly applicable. The presented algorithm identifies syntactic relationships between words, using a source language parser, and maps the alternative interpretations of these relationships to the target language, using a bilingual lexicon. The preferred senses are then selected according to statistics on lexical relations in the target language. The selection is based on a statistical model and on a constraint propagation algorithm, which handles simultaneously all ambiguities in the sentence. The method was evaluated using three sets of Hebrew and German examples and was found to be very useful for disambiguation. The paper includes a detailed comparative analysis of statistical sense disambiguation methods.</p> |
| 16 | A Scaleable Comparison Shopping Agent for the World-Wide Web | <p>The Web is less agent-friendly than we might hope. Most information on the Web is presented in loosely structured natural language text with no agent-readable semantics. HTML annotations structure the display of Web pages, but provide virtually no insight into their content. Thus, the designers of intelligent Web agents need to address the following questions: (1) To what extent can an agent understand information published at Web sites? (2) Is the agent's understanding sufficient to provide genuinely useful assistance to users? (3) Is site-specific hand-coding necessary, or can the agent automatically extract information from unfamiliar Web sites? (4) What aspects of the Web facilitate this competence?</p> <p>In this paper we investigate these issues with a case study using the ShopBot. ShopBot is a fully implemented, domain-independent comparison-shopping agent. Given the home pages of Several on-line stores, ShopBot autonomously learns how to shop at those vendors. After its learning is complete, ShopBot is able to speedily visit over a dozen software stores and CD vendors, extract product information, such as availability and price, and summarize the results for the user. Preliminary studies show that ShopBot enables users to both find superior prices and substantially reduce Web shopping time. Remarkably, ShopBot achieves this performance without sophisticated natural language processing, and requires only minimal knowledge about different product domains. Instead, ShopBot relies on a combination of heuristic search, pattern matching, and inductive learning techniques.</p> |

| Abs # | Title | Abstract |
|-------|-----------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 17 | From manual to semi-automatic semantic annotation: About ontology-based text annotation tools | Semantic Annotation is a basic technology for intelligent content and is beneficial in a wide range of content-oriented intelligent applications. In this paper we present our work in ontology-based semantic annotation, which is embedded in scenario of a knowledge portal application. Starting with seemingly good and bad manual semantic annotation, we describe our experiences made within the KA2 initiative. The experiences gave us the starting point for developing an ergonomic and knowledge base-supported annotation tool. Furthermore, the annotation tool described are currently extended with mechanisms for semi-automatic information-extraction based annotation. Supporting the evolving nature of semantic content we additionally describe our idea of evolving ontologies supporting semantic annotation. |
| 18 | Inheritance and Complementation: A Case Study of Easy Adjectives and Related Nouns | Mechanisms for representing lexically the bulk of syntactic and semantic information for a language have been under active development, as is evident in the recent studies contained in this volume. Our study serves to highlight some of the most useful tools available for structured lexical representation, in particular, (multiple) inheritance, default specification, and lexical rules. It then illustrates the value of these mechanisms in illuminating one corner of the lexicon involving an unusual kind of complementation among a group of adjectives exemplified by easy. The virtues of the structured lexicon are its succinctness and its tendency to highlight significant clusters of linguistic properties. From its succinctness follow two practical advantages, namely its ease of maintenance and modification. In order to suggest how important these may be practically, we extend the analysis of adjectival complementation in several directions. These further illustrate how the use of inheritance in lexical representation permits exact and explicit characterizations of phenomena in the language under study. We demonstrate how the use of the mechanisms employed in the analysis of easy enable us to give a unified account of related phenomena featuring nouns like pleasure, and even the adverbs (adjectival specifiers) too and enough. Along the way we motivate some elaborations of the HPSG (Head-Driven Phrase Structure Grammar) framework in which we couch our analysis, and offer several avenues for further study of this part of the English lexicon. |
| 19 | Evaluating Natural Language Processing Systems | This report presents a detailed analysis and review of NLP valuation, in principle and in practice. Part 1 examines evaluation concepts and establishes a framework for NLP system evaluation. This makes use of experience in the related area of information retrieval and the analysis also refers to evaluation in speech processing. Part 2 surveys significant evaluation work done so far, for instance in machine translation, and discusses the particular problems of generic system evaluation. The conclusion is that evaluation strategies and techniques for NLP need much more development, in particular to take proper account of the influence of system tasks and settings. Part 3 develops a general approach to NLP evaluation, aimed at methodologically-sound strategies for test and evaluation motivated by comprehensive performance factor identification. The analysis throughout the report is supported by extensive illustrative examples |

| Abs # | Title | Abstract |
|-------|---------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 20 | TextTiling: A Quantitative Approach to Discourse Segmentation | This paper presents TextTiling, a method for partitioning full-length text documents into coherent multiparagraph units. The layout of text tiles is meant to reflect the pattern of subtopics contained in an expository text. The approach uses lexical analyses based on tf.idf, an information retrieval measurement, to determine the extent of the tiles, incorporating thesaural information via a statistical disambiguation algorithm. The tiles have been found to correspond well to human judgments of the major subtopic boundaries of science magazine articles |
| 21 | Specifying Filler-Gap Dependency Parsers in a Linear-Logic Programming Language | An aspect of the Generalized Phrase Structure Grammar formalism proposed by Gazdar, et al. is the introduction of the notion of "slashed categories" to handle the parsing of structures, such as relative clauses, which involve unbounded dependencies. This has been implemented in Definite Clause Grammars through the technique of gap threading, in which a difference list of extracted noun phrases (gaps) is maintained. However, this technique is cumbersome, and can result in subtle soundness problems in the implemented grammars. Miller and Pareschi have proposed a method of implementing gap threading at the logical level in intuitionistic logic. Unfortunately that implementation itself suffered from serious problems, which the authors recognized. This paper builds on work first presented with Miller in which we developed a filler-gap dependency parser in Girard's linear logic. This implementation suffers from none of the pitfalls of either the traditional implementation, or the intuitionistic one. It serves as further demonstration of the usefulness of sub-structural logic in natural language applications. |
| 22 | Learning information extraction patterns from examples | A growing population of users want to extract a growing variety of information from on-line texts. Unfortunately, current information extraction systems typically require experts to hand-build dictionaries of extraction patterns for each new type of information to be extracted. This paper presents a system that can learn dictionaries of extraction patterns directly from user-provided examples of texts and events to be extracted from them. The system, called LIEP, learns patterns that recognize relationships between key constituents based on local syntax. Sets of patterns learned by LIEP for a sample extraction task perform nearly at the level of a hand-built dictionary of patterns |
| 23 | Summarization Evaluation Methods: Experiments and Analysis | Two methods are used for evaluation of summarization systems: an evaluation of generated summaries against an "ideal" summary and evaluation of how well summaries help a person perform in a task such as information retrieval. We carried out two large experiments to study the two evaluation methods. Our results show that different parameters of an experiment can dramatically affect how well a system scores. For example, summary length was found to affect both types of evaluations. For the "ideal" summary based evaluation, accuracy decreases as summary length increases, while for task based evaluations summary length and accuracy on an information retrieval task appear to correlate randomly. In this paper, we show how this parameter and others can affect evaluation results and describe how parameters can be controlled to produce a sound evaluation |

| Abs # | Title | Abstract |
|-------|-------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 24 | Linear Segmentation and Segment Significance | We present a new method for discovering a segmental discourse structure of a document while categorizing each segment's function and importance. Segments are determined by a zero-sum weighting scheme, used on occurrences of noun phrases and pronominal forms retrieved from the document. Segment roles are then calculated from the distribution of the terms in the segment. Finally, we present results of evaluation in terms of precision and recall which surpass earlier approaches. |
| 25 | Natural language processing for information retrieval | The paper summarizes the essential properties of document retrieval and reviews both conventional practice and research findings, the latter suggesting that simple statistical techniques can be effective. It then considers the new opportunities and challenges presented by the user's ability to search full text directly (rather than e.g. titles and abstracts), and suggests appropriate approaches to doing this, with a focus on the potential role of natural language processing. The paper also comments on possible connections with data and knowledge retrieval, and concludes by emphasizing the importance of rigorous performance testing. |
| 26 | Identifying Topics by Position | This paper addresses the problem of identifying likely topics of texts by their position in the text. It describes the automated training and evaluation of an Optimal Position Policy, a method of locating the likely positions of topic-bearing sentences based on genre-specific regularities of discourse structure. This method can be used in applications such as information retrieval, routing, and text summarization. |
| 27 | Assembly of topic extraction modules in summarist | Over the past two years we have been developing the text summarization system SUMMARIST. In this paper, we describe the current status of SUMMARIST and its use in TIPSTER Phase III text summarization research |
| 28 | From discourse structures to text summaries | We describe experiments that show that the concepts of rhetorical analysis and nuclearity can be used effectively for determining the most important units in a text. We show how these concepts can be implemented and we discuss results that we obtained with a discourse-based summarization program. |
| 29 | Generating Summaries of Multiple News Articles | We present a natural language system which summarizes a series of news articles on the same event. It uses summarization operators, identified through empirical analysis of a corpus of news summaries, to group together templates from the output of the systems developed for ARPA's Message Understanding Conferences. Depending on the available resources (e.g., space), summaries of different length can be produced. Our research also provides a methodological framework for future work on the summarization task and on the evaluation of news summarization systems. |

| Abs # | Title | Abstract |
|-------|---------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 30 | An Automatic Method for Generating Sense Tagged Corpora | <p>The unavailability of very large corpora with semantically disambiguated words is a major limitation in text processing research. For example, statistical methods for word sense disambiguation of free text are known to achieve high accuracy results when large corpora are available to develop context rules, to train and test them. This paper presents a novel approach to automatically generate arbitrarily large corpora for word senses. The method is based on (1) the information provided in WordNet, used to formulate queries consisting of synonyms or definitions of word senses, and (2) the information gathered from Internet using existing search engines. The method was tested on 120 word senses and a precision of 91% was observed</p> |
| 31 | Text Generation in a Dynamic Hypertext Environment | <p>This paper describes PEBA-II, a working natural language generation system which interactively describes animals in a taxonomic knowledge base via the production of World Wide Web pages. Our aim is to construct a natural language document generation system with real practical applicability: to this end, the system reconstructs and combines a number of existing ideas in the literature in a novel way, and proposes a solution to the problem of breadth of coverage that is based on a pragmatic approach to knowledge representation and linguistic realisation. The system embodies the following features:</p> <ul style="list-style-type: none"> · a reconstruction of some of the core ideas in schema--based text generation [McKeown 1985], applied to the generation of hypertext documents; · the principled use of a phrasal lexicon to ease surface generation, in concert with a knowledge base whose elements may correspond to pre--compiled collections of atomic units; · a user model and discourse model that permit interesting variations in the texts produced. <p>We describe each of the above aspects of the existing system in some detail, and point to a number of interesting research directions it opens up.</p> |
| 32 | Two-level Description of Turkish Morphology | <p>This paper describes a full two-level morphological description [5,9] of Turkish word structures. The description has been implemented using the PC-KIMMO environment [2] and is based on a root word lexicon of about 23,000 roots words. The phonetic rules of contemporary Turkish (spoken in Turkey) have been encoded using 22 two-level rules while the morphotactics of the agglutinative word structures have been encoded as nite-state machines for verbal, nominal paradigms and other categories. Almost all the special cases of, and exceptions to phonological and morphological rules have been taken into account. In this paper, we describe the rules and the nite state machines along with examples and a discussion of how various special cases were handled. We also describe some known limitations and problems with this description. We then briefly describe various natural language processing applications that use this description as the morphological analysis component.</p> |

| Abs # | Title | Abstract |
|-------|--------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 33 | INTENTION-BASED SEGMENTATION: HUMAN RELIABILITY AND CORRELATION WITH LINGUISTIC CUES | <p>Certain spans of utterances in a discourse, referred to here as segments, are widely assumed to form coherent units. Further, the segmental structure of discourse has been claimed to constrain and be constrained by many phenomena. However, there is weak consensus on the nature of segments and the criteria for recognizing or generating them. We present quantitative results of a two part study using a corpus of spontaneous, narrative monologues. The first part evaluates the statistical reliability of human segmentation of our corpus, where speaker intention is the segmentation criterion. We then use the subjects' segmentations to evaluate the correlation of discourse segmentation with three linguistic cues (referential noun phrases, cue words, and pauses), using information retrieval metrics.</p> |
| 34 | INDUCING FEATURES OF RANDOM FIELDS | <p>We present a technique for constructing random fields from a set of training samples. The learning paradigm builds increasingly complex fields by allowing potential functions, or features, that are supported by increasingly large subgraphs. Each feature has a weight that is trained by minimizing the Kullback-Leibler divergence between the model and the empirical distribution of the training data. A greedy algorithm determines how features are incrementally added to the field and an iterative scaling algorithm is used to estimate the optimal values of the weights. The random field models and techniques introduced in this paper differ from those common to much of the computer vision literature in that the underlying random fields are non-Markovian and have a large number of parameters that must be estimated. Relations to other learning approaches including decision trees and Boltzmann machines are given. As a demonstration of the method, we describe its application to the problem of automatic word classification in natural language processing.</p> |
| 35 | On the Complexity of Qualitative Spatial Reasoning: A Maximal Tractable Fragment of the Region Connection Calculus | <p>The computational properties of qualitative spatial reasoning have been investigated to some degree. However, the question for the boundary between polynomial and NP-hard reasoning problems has not been addressed yet. In this paper we explore this boundary in the "Region Connection Calculus" RCC-8. We extend Bennett's encoding of RCC-8 in modal logic. Based on this encoding, we prove that reasoning is NP-complete in general and identify a maximal tractable subset of the relations in RCC-8 that contains all base relations. Further, we show that for this subset path-consistency is sufficient for deciding consistency.</p> |
| 36 | A Maximum Entropy Approach to Identifying Sentence Boundaries. | <p>We present a trainable model for identifying sentence boundaries in raw text. Given a corpus annotated with sentence boundaries, our model learns to classify each occurrence of ., ?, and ! as either a valid or invalid sentence boundary. The training procedure requires no hand-crafted rules, lexica, part-of-speech tags, or domain-specific information. The model can therefore be trained easily on any genre of English, and should be trainable on any other Roman alphabet language. Performance is comparable to or better than the performance of similar systems, but we emphasize the simplicity of retraining for new domains</p> |

| Abs # | Title | Abstract |
|-------|------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 37 | Selective Text Utilization and Text Traversal | <p>Many large collections of full-text documents are currently stored in machine-readable form and processed automatically in various ways. These collections may include different types of documents, such as messages, research articles, and books, and the subject matter may vary widely. To process such collections, robust text analysis methods must be used, capable of handling materials in arbitrary subject areas, and flexible access must be provided to texts and text excerpts of varying size. In this study, global text comparison methods are used to identify similarities between text elements, followed by local context-checking operations that resolve ambiguities and distinguish superficially similar texts from texts that actually cover identical topics. A linked text structure is then created that relates similar texts at various levels of detail. In particular, text links are available for full texts, as well as text sections, paragraphs, and sentence groups. The linked structures are usable to identify important text passages, to traverse texts selectively both within particular documents and between documents, and to provide flexible text access to large text collections in response to various kinds of user needs. An automated 29-volume encyclopedia is used as an example to illustrate the text accessing and traversal operations.</p> |
| 38 | Automatic text decomposition using text segments and text themes | <p>With the widespread use of full-text information retrieval, passage-retrieval techniques are becoming increasingly popular. Larger texts can then be replaced by important text excerpts, thereby simplifying the retrieval task and improving retrieval effectiveness. Passage level evidence about the use of words in local contexts is also useful for resolving language ambiguities and improving retrieval output. Two main text decomposition strategies are introduced in this study, including a chronological decomposition into text segments, and semantic decomposition into text themes. The interaction between text segments and text themes is then used to characterize text structure, and to formulate specifications for information retrieval, text traversal, and text summarization.</p> |
| 39 | Tailoring the Interaction with Users in Web Stores | <p>We describe the user modeling and personalization techniques adopted in SETA, a prototype toolkit for the construction of adaptive Web stores which customize the interaction with users. The Web stores created using SETA suggest the items best fitting the customers' needs and adapt the layout and the description of the store catalog to their preferences and expertise. SETA uses stereotypical information to handle the user models and applies personalization rules to dynamically generate the hypertextual pages presenting products. The system adapts the graphical aspect, length and terminology used in the descriptions to parameters like the user's receptivity, expertise and interests. Moreover, it maintains a model associated with each person the goods are selected for; in this way, multiple criteria can be applied for tailoring the selection of items to the preferences of their beneficiaries.</p> |

| Abs # | Title | Abstract |
|-------|--------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 40 | PARADISE: A Framework for Evaluating Spoken Dialogue Agents | This paper presents PARADISE (PARAdigm for Dialogue System Evaluation), a general framework for evaluating spoken dialogue agents. The framework decouples task requirements from an agent's dialogue behaviors, supports comparisons among dialogue strategies, enables the calculation of performance over subdialogues and whole dialogues, specifies the relative contribution of various factors to performance, and makes it possible to compare agents performing different tasks by normalizing for task complexity. |
| 41 | Japanese Discourse and the Process of Centering | This paper has three aims: (1) to generalize a computational account of the discourse process called CENTERING, (2) to apply this account to discourse processing in Japanese so that it can be used in computational systems for machine translation or language understanding, and (3) to provide some insights on the effect of syntactic factors in Japanese on discourse interpretation. We argue that while discourse interpretation is an inferential process, syntactic cues constrain this process, and demonstrate this argument with respect to the interpretation of ZEROS, unexpressed arguments of the verb, in Japanese. The syntactic cues in Japanese discourse that we investigate are the morphological markers for grammatical TOPIC, the postposition wa, as well as those for grammatical functions such as SUBJECT, ga, OBJECT, o and OBJECT2, ni. In addition, we investigate the role of speaker's EMPATHY, which is the viewpoint from which an event is described. This is syntactically indicated through the use of verbal compounding, i.e. the auxiliary use of verbs such as kureta, kita. Our results are based on a survey of native speakers of their interpretation of short discourses, consisting of minimal pairs, varied by one of the above factors. We demonstrate that these syntactic cues do indeed affect the interpretation of ZEROS, but that having previously been the TOPIC and being realized as a ZERO also contributes to the salience of a discourse entity. We propose a discourse rule of ZERO TOPIC ASSIGNMENT, and show that CENTERING provides constraints on when a ZERO can be interpreted as the ZERO TOPIC. |
| 42 | Using Inductive Logic Programming to Automate the Construction of Natural Language Parsers | Designing computer systems to understand natural language input is a difficult task. In recent years there has been considerable interest in corpus-based methods for constructing natural language parsers. These empirical approaches replace hand-crafted grammars with linguistic models acquired through automated training over language corpora. A common thread among such methods to date is the use of propositional or probabilistic representations for the learned knowledge. This dissertation presents an alternative approach based on techniques from a subfield of machine learning known as inductive logic programming (ILP). ILP, which investigates the learning of relational (first-order) rules, provides an empirical method for acquiring knowledge within traditional symbolic parsing frameworks. This dissertation details the architecture, implementation and evaluation of Chill, a computer system for acquiring natural language parsers by training over corpora of parsed text. Chill treats language acquisition as the learning of search-control rules within a logic program that implements a shift-reduce parser. Control rules are induced using a novel ILP algorithm which handles difficult issues arising |

| Abs # | Title | Abstract |
|-------|-------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 43 | INTELLIGENT AGENTS ON THE INTERNET: Fact, Fiction, and Forecast | Computer technology has dramatically enhanced our ability to generate, deliver, and store information. Unfortunately, our tools for locating, filtering, and analyzing information have not kept pace. A popular solution is intelligent agents. But what are they? We provide a survey showing the myriad ways in which the evocative term "agent" is interpreted by researchers in the field. Following the "Information Superhighway" metaphor, an intelligent agent may be: A backseat driver who makes suggestions at every turn, or A taxi driver who drives you to your destination, or even A concierge whose knowledge and skills eliminate the need to personally approach the Superhighway at all. We briefly describe a number of prototype Internet agents and elaborate on the Internet Softbot, a concierge under development at the University of Washington. |
| 44 | Shopping Models: A Flexible Architecture for Information Commerce | In a digital library, there are many different interaction models between customers and information providers or merchants. Subscriptions, sessions, pay-per-view, shareware, and pre-paid vouchers are different models that each have different properties. A single merchant may use several of them. Yet if a merchant wants to support multiple models, there is a substantial amount of work to implement each one. In this paper, we formalize the shopping models which represent these different modes of consumer to merchant interaction. In addition to developing the overall architecture, we define the application program interfaces (API) to interact with the models. We show how a small number of primitives can be used to construct a wide range of shopping models that a digital library can support, and provide examples of the shopping models in operation, demonstrating their flexibility |
| 45 | Human Performance on Clustering Web Pages: A Preliminary Study | With the increase in information on the World Wide Web it has become difficult to quickly find desired information without using multiple queries or using a topic-specific search engine. One way to help in the search is by grouping HTML pages together that appear in some way to be related. In order to better understand this task, we performed an initial study of human clustering of web pages, in the hope that it would provide some insight into the difficulty of automating this task. Our results show that subjects did not cluster identically; in fact, on average, any two subjects had little similarity in their web-page clusters. We also found that subjects generally created rather small clusters, and those with access only to URLs created fewer clusters than those with access to the full text of each web page. Generally the overlap of documents between clusters for any given subject increased when given the full text, as did the percentage of documents clustered. When analyzing individual subjects, we found that each had different behavior across queries, both in terms of overlap, size of clusters, and number of clusters. These results provide a sobering note on any quest for a single clearly correct clustering method for web pages. |

Appendix C

Citation Context Extracts

| ABS # | Citation # | Citation Context Text |
|----------|---------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | 1 | ... or example driven learning applied to text classification and document processing (14] More recent approaches span from very specific language processing tasks (e.g. induction of PP disambiguation rules [8] Machine Learning driven subcategorization frames acquisition [2] or case based discourse modeling [9]) to integration of symbolic induction and quantitative techniques for the construction of extensive components of lexical knowledge bases (e.g. 17, 12, 3] 2 Adapting Information Extraction systems to domains and users In Information Extraction the precise interaction between syntax and semantics provides from one side a more ... |
| 1 | 2 | ... Wafer Scale Integration (Kitano, 1993) In Natural Language Processing, lazy learning techniques are currently being applied by various Japanese groups to parsing and machine translation under the names exemplar based translation or memory based translation and parsing (Kitano, 1993) In work by Cardie (Cardie, 1993) and by the present authors (Daelemans, 1995; Daelemans et al. 1994) variants of lazy learning are applied to disambiguation tasks at different levels of linguistic representation (from phonology to semantics) One lazy learning variant, Analogical Modeling (Skousen, 1989) was explicitly ... |
| 1 | 3 | ... components. Another is that most IE systems are capable of extracting only limited structures from the input text. Creating dictionaries of concept patterns for information extraction consumes a great amount of time and is required for each new domain. Research efforts have begun to address this: Cardie (1993) , Riloff (1993) Soderland et al. 1995) Huffman (1996) These corpus based methods all employ a large training corpus annotated with examples for each concept. From these examples, machine learning algorithms induce conceptual patterns for extraction. However, these methods have not ... |
| 1 | 4 | ... plus a semantic hierarchy and associated lexicon. LIEP (Huffman 1996) is another system that learns extraction patterns but relies on predefined keywords, object recognizers (e.g. to identify people and companies) and human interaction to annotate each relevant sentence with an event type. Cardie (Cardie 1993) and Hastings (Hastings Lytinen 1994) also developed lexical acquisition systems for information extraction, but their systems learned individual word meanings rather than extraction patterns. Both systems used a semantic hierarchy and sentence contexts to learn the meanings of unknown words. |
| 1 | 5 | ... knowledge, and some research has concentrated in areas as specific as the problem solving process [Tor] The term is sometimes blurred to refer to the whole process of acquiring any form of knowledge from textual input, which would encompass the areas of DM, IE and IR. A paper by Claire Cardie, Car93] provides some literature to a case based approach in KA. Knowledge acquisition is not necessarily restricted to processing unfiltered text input some KA systems are tools for structuring and organizing information supplied by human experts especially for the system. KA may be a useful method ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2 | 6 | ... basis (represented by constraints) offers a spatial relation based language, and, integrates geometric types into the data model. Our work is related also to temporal databases where the issues of granularity, scale and partially specified information have received considerable attention (e.g. [Kou94a], DS93] WJS93] in the last few years. To illustrate the expressiveness of the proposed spatial data model, consider the following query examples. 1. Querying Spatial Relations with Granularity. Assume the query: Find if the CS Dept. is inside the scope of V 1. This query may be initially ... |
| 2 | 7 | ... values that are unknown but are not to have occurred, values that are known if they have occurred, and values that are unknown even if they occurred. Gadia [25] shows that his model is sound. However, he makes no use of probabilistic information. An important body of work is that of Koubarakis [33, 34] who proposes the use of constraints for representing temporal data. In this sense, our work is directly related and builds upon Koubarakis work. Like our work on TP Databases, Koubarakis uses constraints to represent when an event occurs. Koubarakis framework allows stating the facts that event ... |
| 2 | 8 | ... for realizing calendars, by using their languages to express constraints on the top time unit, which is isomorphic to the integers in our system. In a similar vein, Koubarakis has published an elegant series of papers in which he reasons about definite and indefinite temporal specifications [16, 17] in particular, he shows that constraints may be used to capture indefinite temporal information (points and interval) and manipulated these constraints to implement operations that extend those in the relational algebra of databases. However, he does not use symbolic time that refers to an ... |
| 2 | 9 | ... in secondary storage N non crossing but possibly touching plane segments (for brevity, called NCT segments) Segment databases are the basis for data representation in several large scale applications, including spatial databases and geographical information systems (GIS) [18] temporal databases [15] and constraint databases [13] Among all possible applications, GIS certainly represent the main target of segment databases. Indeed, GIS databases often store data as layers of maps, where each map is typically stored as a collection of NCT segments. Some relevant query types for segment ... |
| 2 | 10 | ... negative and non negative) Temporal constraints can express addition constraints. Hence the recognition problem for Datalog with temporal constraints is undecidable. However, an evaluation of relational calculus queries with temporal constraints is possible and is considered by Koubarakis in [19, 20]. Efficient tests for temporal constraint satisfaction are described in [9] and for monotone two variable constraints in [11] Chomicki and Imielinski [7] consider the language Datalog 1S which is like Datalog extended with an increment operator which may occur only in the first argument of ... |
| 3 | 11 | ... (1997) The main advantage of transforming an HMM is that the resulting transducer can be handled by finite state calculus. Among others, it can be composed with transducers that encode: ffl correction rules for the most frequent tagging errors which are automatically generated (Brill, 1992; Roche and Schabes, 1995) or manually written (Chanod and Tapanainen, 1995) in order to significantly improve tagging accuracy 2 . These rules may include long distance dependencies not handled by HMM taggers, and can conveniently be expressed by the replace operator (Kaplan and Kay, 1994; Karttunen, 1995; Kempe and ... |
| 3 | 12 | ... be replaced by another one in a certain context or contexts. Phonological rewrite rules (Kaplan and Kay, 1994) two level rules (Koskenniemi 1983) syntactic disambiguation rules (Karlsson et al. 1994, Koskenniemi, Tapanainen, and Voutilainen 1992) and part of speech assignment rules (Brill 1992, Roche and Schabes 1995) are examples of replacement in context of finite state grammars. Kaplan and Kay (1994) describe a general method representing a replacement procedure as finite state transduction. Karttunen (1995) takes a somewhat simpler approach by introducing to the calculus of regular expression a ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 3 | 13 | ... and Jurafsky, 1996) Indeed, many interesting transducers are of the type change all occurrences of in some specific context into , and pass on the rest of the input unaltered . The various replacement and local extension operators all produce transducers of this kind (Karttunen, 1995; Roche and Schabes, 1995; Karttunen, 1996; Kempe and Karttunen, 1996; Gerdemann and van Noord, 1999) 1.3 Smaller Automata Another motivation for the introduction of predicates is the observation that the resulting automata are smaller. The size of automata is an important problem in practice (Daciuk, 1998; Kiraz, ... |
| 3 | 14 | ... rarely used as verbs (e.g. recht right, to rake (3rd person, sg) Apart from manually constructed rules mentioned above, we also use rules determined by Brill s tagger (Brill, 1993) 2 All rules are compiled into a single finite state transducer according to the approach described in (Roche and Schabes, 1995). Named entity finder Named entities such as organizations, persons, locations and time expressions are identified using finite state grammars. Since some named entities (e.g. company names) may appear in the text either with or without a designator, we use a dynamic lexicon to store recognized ... |
| 3 | 15 | ... the FSA Utilities toolbox has been the rapidly growing interest in finite state techniques for computational linguistics. In particular, finite state techniques are being used in computational phonology and morphology (Kaplan and Kay 1994) efficient dictionary lookup and part of speech tagging (Roche and Schabes 1995), natural language parsing (Voutilainen and Tapanainen 1993) techniques for parsing ill formed input (Lang 1989, van Noord 1995) and speech recognition (Oerder and Ney 1993, Pereira and Riley 1996) etc. The FSA Utilities toolbox has been developed to experiment with the techniques presented in ... |
| 4 | 16 | ... same as in the monolingual case, and so are not considered further here. For example the ELU system (Estival et al. 1990) uses lexical transfer rules which could be derived from tlinks on the assumption that transfer variables are equivalent to reentrancy between tlink output FSs. The BCI system (Alshawhi et al. 1991) uses transfer between quasi logical forms; the appropriate lexical transfer rules between predicates can be derived from tlinks under the same assumptions. Since at its most general the tlink mechanism states correspondences between FSs its use is not specific to transfer based MT. In our current ... |
| 4 | 17 | ... underspecified. Thus, both possible scope interpretations of the modal operator are captured by this kind of representation. 2. 2 Ambiguity Preservation In order to avoid expensive resolution procedures, it is most desirable to preserve ambiguities that hold within a language pair ([Alshawhi et al. 1991] and [Kay et al. 1994] Considering the language pair German English, these are among others: ffl Scope ambiguities ffl Modifier attachment ambiguities ffl Polysemy ffl Interpretation of possessive relations Ambiguity preservation is primarily a representational problem. An underspecified ... |
| 4 | 18 | ... of SL and TL lexical signs. This has the advantage of modularity since no information of the monolingual grammars is involved, but also the disadvantage that complex equivalences, which involve more than one lexical item, are difficult to express. The Quasi logical form (QLF) transfer, cf. [Alshawhi et al. 1991], provides another solution to this problem. Here, the transfer produces only semantic representations that are syntactically and semantically equivalent to a QLF accepted by the generator. Based on the QLF transfer and the Shake and Bake translation approach, Copestake et al. 1995] propose a ... |
| 4 | 19 | ... it is natural to think of transfer in terms of signs, and to concentrate on lexical transfer. Thus we can classify MT systems in terms of such a model according to the sort of translation constraints they assume, and how they control the process of translation. For example in SRI's BCI system (Alshawhi et al. 1991) the source language string is parsed and transfer is carried out on the (quasi)logical form representation. This produces a quasi logical form appropriate for the TL, which can be used to drive a head driven generator. In contrast, the Shake and Bake approach (Whitelock, 1992) relies on lexical ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 4 | 20 | ... ordering for the application of transfer rules according to their level of specificity (see Section 4.3) Our transfer rules do not, in themselves, trigger calls to the recursive application of subsequent transfer rules. This is a main distinction between our approach and the one presented in Alshawi et al. 1991). Because the recursive rule application is not part of the rules, our approach solves problems with discontinuous translation equivalences which the former approach cannot handle well (Dorna, 2000b) Finally, we have extended the transfer rules with conditions describing local structures or ... |
| 5 | 21 | ...Section 4 describes details from the implementation of VIENA Classroom and Section 5 summarizes the results. 2. Related work In spite of the large amount of work on natural language interfaces, they are today still far away from widespread practical use (for good recent surveys see [Copestake90, Androutsopoulos94] The reason for this are the many limitations which still exist and which are caused by two main factors: missing customization, resulting in unexpected restrictions, and missing integration, responsible for insufficient performance and wrong interpretation (see [McFetridge90, Sparck Jones94] ... |
| 5 | 22 | ...of natural language interfaces has a long tradition and many researchers spent a vast amount of efforts to deal with this complex issue. Nevertheless, we have to face the situation today that natural language interfaces are still far away from widespread practicable use (for a recent survey see [2]) The reasons for this are the many limitations that still exist and that are due to two main factors: missing customisation, resulting in unexpected Information Systems and Technologies for Network Society, Fukuoka, Japan, September 1997 restrictions, and missing integration, which is ... |
| 5 | 23 | ... Natural language is especially appealing as an interface for database queries because the user is able to express her or his information request naturally without the need to learn a formal query language such as for example SQL [4] For a nice and concise overview of the eld consult for instance [1, 3]. Natural language technology is also a potential key for the success of applications in ecommerce. In particular, the provision of multilingual access to information resources is crucial, even stressed in such a multilingual environment as Europe. We have developed an interface prototype called ... |
| 5 | 24 | ...as much on general software quality factors like functionality, usability, reliability, performance, maintainability and portability, as on their linguistic coverage. The special linguistic resources make natural language interfaces relatively complex, unstable and difficult to maintain or adapt [Androutsopoulos et al. 1995] Because of their interactive nature and because of the promise of natural language as an easy to use interface, usability is the most important quality factor for a dialogue system: can the system actually be learnt and can it be used in an effective, efficient and satisfactory way An iterative ... |
| 5 | 25 | ...front end. We discuss some of the directions required in order to turn the system from a research prototype to a working tool. 2 Related work There is voluminous literature on the design of NL interfaces to general nontemporal databases (see Perrault and Grosz, 1988; Copestake and Jones, 1990; Androutsopoulos et al. 1995) for surveys and by now their main advantages and disadvantages are well understood. Much less work has been devoted to the design of NL interfaces to TDBs (Clifford, 1990; Hinrichs, 1988) or other computer systems involving a temporal dimension (Crouch and Pulman, 1993). Of particular relevance ... |
| 6 | 26 | ...in the document. The locations of these high activity areas are likely to be good candidates for selection by a summarizer. The role of FreeNet in this algorithm would simply be to provide answers on which pairs of concepts are related, and how strongly. Other approaches involving lexical chains (Barzilay and Elhadad, 1997) could benefit from the FreeNet engine. An intriguing question to ask about 9 a discourse unit is "What is the maximum length chain for subsequence, but not necessarily contiguous # of lexically related words appearing in this region?" An approximate answer to such a longest path question, ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 6 | 27 | ... how salient, are far from coherent prose; this is why the issue of appropriate presentation metaphors for such topical phrases is an important one, and is discussed at length in (Boguraev, Bellamy, and Kennedy, 1999) Uniquely, this kind of intuition is explicitly addressed in recent work by Barzilay and Elhadad (1999). They introduce the notion of lexical chains , derived by grouping together items related by repetition and certain lexical relations calculated by reference to the WORDNET lexical database (Fellbaum, 1999) A sequence of items in a chain highlights a discussion focused on a topic related to ... |
| 6 | 28 | ...(57.0 22.5) ProperNames = FALSE: false (64.0 30.5) Apparently, high cohesion sentences are evidence that the sentence should not be included in the summary. This is surprising, since there are many techniques that select for the summary sentences with a high degree of cohesion [Mitra 97] Barzilay 97] However, these results are not entirely conclusive, because the summaries of our training base were not produced by humans. In addition, it is possible that these results reflect a tendency of the document base used in the experiments. 4.2 Final |
| 6 | 29 | ...in previous implementations of lexical chains. Because all possible senses of the word are not taken into account, except at the time of insertion, potentially pertinent context information that appears after the word is lost. The problem that results is referred to as greedy disambiguation [1]. Barzilay and Elhadad presented a less greedy algorithm that constructs all possible interpretations of the source text using lexical chains. Their algorithm then selects the interpretation with the strongest cohesion. They then use these strong chains to generate a summary of the original ... |
| 6 | 30 | ...phrases that corefer with expressions in the query. The resulting extract is used to support relevancy judgments with respect to the query. The use of chains of related expressions in documents to select sentences for inclusion in a generic (i.e. non user focused) summary is also not novel. Barzilay and Elhadad (1997) describe a technique for text summarization based on lexical chains. Their technique, which builds on work of Morris and Hirst (1994) and ultimately Halliday and Hasan (1976) who stressed the role of lexical cohesion in text coherence, is to form chains of lexical items across a text based on ... |
| 7 | 31 | ...The effects of shopbots and similar software agents in electronic markets have been studied using several approaches. In microeconomic theory, simple models of the relationship between information of price, utility and consumer choice are applied. Based upon these models, Greenwald and Kephart [5] have made predictions of increased efficiency in electronic markets as a result of shopbot use. Their predictions correspond well to the analytic predictions of Bakos [2] Another approach is the application of software agents with a cognitive architecture consisting of beliefs, preferences and ... |
| 7 | 32 | ... sell the good it is interested in, picks out ones that offer prices less than its valuation, and then selects a seller at random from this lot (this is the same as picking a seller at random and then transacting if the price offered by seller is less than the price buyer's valuation, as suggested in [8] (since both the events are unrelated) 2. Bargain Hunter: buyer checks the price of all sellers having prices lower than its valuation, determines the seller with the lowest price, and then purchases the good (This type of buyer agents corresponds to people who take advantage of shopbots) The ... |
| 7 | 33 | ...price in the market at time t , and, $t = p t (BRS S) b s$ if the seller is not charging the minimum price in the market. The expression in parentheses in the above equations represent the expected number of times a seller is selected by the buyers during the time interval t . In [5, 6], Greenwald and Kephart view the price setting problem as a one shot game, and provide a detailed game theoretic analysis of the shopbot economy showing that although there is no pure strategy Nash equilibrium, there exists a symmetric mixed strategy Nash equilibrium. Greenwald and Kephart also ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 7 | 34 | ... good in multiple auctions (Ito, Fukuta, Shintani, Sycara, 2000; Anthony, Hall, Dang, Jennings, Preist, Bartolini, Phillips, 2001) Outside of, but related to, the auction scenario, automatic shopping and pricing agents for internet commerce have been studied within a simplified model (Greenwald Kephart, 1999). Twenty two agents from 6 countries entered TAC, 12 of which qualified to compete in the semi finals and finals in Boston. The designs of these agents were motivated by a wide variety of research interests including machine learning, artificial life, experimental economics, real time systems, ... |
| 7 | 35 | ...are likely to reduce the size of the marketplace and to introduce bias, as it is difficult to obtain a sufficient number of ratings for every existing vendor, and to control the reliability of the sources. Finally, a fourth general approach is to further automate and generalize the search process [11, 7]. As early as 1995, shopping agents (also referred to as comparison shopping agents) were proposed as a solution to find a product under the best terms (where price was the most important feature early on) among different e-commerce sites. A shopping agent queries multiple sites on behalf of a ... |
| 8 | 36 | ...level that is provided in our approach by the domain ontology and the associated inference engine. Knowledge representation for the web: The Ontobroker project (Decker et al. 1999) lays the technological foundations for the KA2 portal. Similar to Ontobroker are SHOE (Luke et al. 1997) and WebKB (Martin Eklund 1999). All three systems aim at providing intelligent access to Web documents (though, with different means) However, they all lack an environment of methods and tools that are needed to build a community portal on their top and, thus, to make an application out of a core technology. From our point ... |
| 9 | 37 | ... link information in some form [1, 25] How much more effective are link based methods in the web environment as compared to a state of the art keyword-based method developed for the TREC ad hoc task This question has been studied in a limited number of studies, especially under TREC s web track [5, 6, 7]. The results from these studies indicate that for web search, link based methods do not hold any advantage over the state of the art keyword based methods developed for TREC ad hoc search. These results are quite counter intuitive given the general wisdom in the web search community that some kind ... |
| 9 | 38 | ...keyword based methods developed for TREC ad hoc search. These results are quite counter intuitive given the general wisdom in the web search community that some kind of linkage analysis does improve web page site ranking. Our work is motivated by this discrepancy between the results presented in [5, 6, 7], and the general belief in the web search community. Different web search engines make competing claims regarding their coverage and search effectiveness. In this study, we don't concentrate on comparing the search effectiveness of different web search engines. There have been several studies that ... |
| 9 | 39 | ...column holds the proportion of unrelated documents. We averaged these individual proportions to get a precision indicator of 23 for documents relevant to intended concept. While this may seem low, it is understandable in view of the fact that Internet search engines have a precision of 23 to 38 (Hawking, et al. 1999). Moreover, some documents, although irrelevant for the target concept, may still be useful for some other concepts in the structure. Note that the simple strategy of formulating queries by extracting terms from the concept descriptor is generally effective; there are notable exceptions, however. ... |
| 9 | 40 | ...judgement procedure was used by Hawking et al. 8] in a comparison between some TREC IR systems and some well known Web search engines. Using the described procedure, the P 20 values were estimated to 0.33 for the official version of the engine, and to 0.36 for the new version. Hawking et al. [8] found a precision range of 0.23 to 0.38 for the Web search engines they tested. Is an observed precision difference of 0.03 significant A standard paired data two tailed test with the null hypothesis that the difference is zero, is not rejected at the 0.05 level of significance ($z=1.79$ against z ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 9 | 41 | ...search engines, such as: single keyword search; plural search capability; phrase search; Boolean search (with proper noun) and complex Boolean. In the next section, we discuss some of the differences and similarities in classical and Internet based search, access and retrieval of information. [Hawking et al. 1999] discusses evaluation studies of six TREC 34 search engines. In particular, they examine answers to questions, such as: Can link information result in better rankings and Do longer queries result in better answers . 34 U.S. National Institute of Standards and Technology (NIST) Text ... |
| 10 | 42 | ...WebWatcher uses a combination of supervised and reinforcement learning to learn the value of each word on a hyperlink. Our work is not user-centric and strives to find a method for learning an optimal decision policy for locating relevant documents when hyperlink selection is unlimited. Laser [Boyan et al. 1996] is a search engine that automatically optimizes a number of parameters to achieve improved retrieval performance. The CMU CS Web is used as the testbed and evaluation is based on the user s selection of links presented by Laser. The work finds that incorporating HTML markup into the TFIDF ... |
| 10 | 43 | ...has to attempt to reach. That is, the equation reaches a xed point when $r + \gamma V_t(s_t)$ equals to $V_t(s_t)$ i.e. the sum of the reward and the discounted expected reward of the next state becomes the same as the value of the current state. It should be mentioned that WebWatcher (Boyan, Freitag, Joachims, 1996; Joachims, Freitag, Mitchell, 1997) learns the user interests using reinforcement learning like in WAIR. In Web Watcher, it is assumed that the information space is linked with hyperlinks. While the retrieval agent seeks the relevant documents, it is directed by the value of reinforcement ... |
| 10 | 44 | ...different and clear views of the templates discovered. Past work on web log mining has been done. However, most has focused on mining to change the web structure for easier browsing [Craven, et al., 1999; Sundaresan and Yi, 2000] predicting browsing behaviors for prefetching [Zaine, et al., 1998; Boyan, 1996] or predicting user preference for active advertising [Pei, et al., 2000; Perkowitz, 1997] Some work has been done on mining user logs for improving search engine s performance. The most notable example is the Google search engine [Google] in which data mining is used to gather statistical ... |
| 10 | 45 | ...retrieval systems [57] they create an index of words within documents, and return a ranked list of documents in response to user queries. Web search engines are good at returning long lists of relevant documents for many user queries, and new methods are improving the ranking of search results [8, 10, 21, 36, 41]. However, few of the results returned by a search engine may be valuable to a user [6, 50] Which documents are valuable depends on the context of the query for example, the education, interests, and previous experience of a user, along with information about the current request. Is the user ... |
| 10 | 46 | ...Spidering CS Departments RL Immediate RL Future Breadth First Figure 2: Average performance of two reinforcement learning spiders versus traditional breadth first search. Recommend relevant hyperlinks to the user. Laser uses reinforcement learning to tune the search parameters of a search engine [Boyan et al. 1996] 2.1 Experimental Results In August 1998 we fully mapped the CS department web sites at Brown University, Cornell University, University of Pittsburgh and University of Texas. They include 53,012 documents and 592,216 hyperlinks. We perform four test train splits, where the data from three ... |
| 11 | 47 | ...art of the field. 1 Introduction The goal of information retrieval is to find all documents relevant for a user query in a collection of documents. Decades of research in information retrieval were successful in developing and refining techniques that are solely word based (see e.g. [2]) With the advent of the web new sources of information became available, one of them being the hyperlinks between documents and records of user behavior. To be precise, hypertexts (i.e. collections of documents connected by hyperlinks) have existed and have been studied for a long time. What ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 11 | 48 | ...of the passages in which the candidate appears, the likelihood that the candidate matches the assigned answer category, and other special case information provided by the selection rules. This general approach of question analysis followed by IR followed by IE is nearly ubiquitous in QA systems [1, 4, 8, 10,13,14,16,20,23]. Using relatively simple question parsing and answer selection components our QA system provides good performance. For the TREC 9 QA task [21] the system placed in the top three for both 50 and 250 byte runs. Our experience with the TREC QA task indicated that three specific features make the ... |
| 11 | 49 | ...entities of the correct type for a given question are ranked using a set of heuristics. Moldovan et al. and Aliod et al. [13, 2] present systems that re rank and postprocess the results of regular information retrieval systems with the goal of returning the best passages. Cardie et al. [7] describe a system that combines statistical and linguistic knowledge for question answering and employs sophisticated linguistic filters to postprocess the retrieved documents and extract the most promising passages to answer a question. The systems above use the general approach of retrieving ... |
| 12 | 50 | ...the co training algorithm that iteratively selects an unlabeled example, gives it a label, and relearns. Nigam and Ghani (2000) argue that the co training algorithm and its variants succeed in part because they are more robust to the assumptions of their underlying classifier representations. Collins and Singer (1999) present a boosting based algorithm, coBoost, for learning in the co training setting; it tries to minimize the disagreement on the unlabeled data between classifiers that use different views of the data. Goldman and Zhou (2000) show that co training approaches can succeed on datasets without ... |
| 12 | 51 | ...to increase agreement between a pair of statistical models by exploiting mutual constraints between their output. Co Training has been used before in applications like word sense disambiguation (Yarowsky, 1995) web page classification (Blum and Mitchell, 1998) and named entity identification (Collins and Singer, 1999). In all of these cases, using unlabeled data has resulted in performance that rivals training solely from labeled data. However, these previous approaches were on tasks that involved identifying the right label from a small set of labels (typically 2-3) and in a relatively small parameter ... |
| 12 | 52 | ...Romanian, Greek, Turkish and Hindi) and achieved 75.4 F measure on the Romanian text. Although this performance is less than many of the systems reported in MUC7 it is achieved with a small set of seed names (annotated names) using a method which was applied to a diverse set of languages. Collins and Singer (1999) proposed another technique using unsupervised machine learning and extremely small amounts of seed data. This approach performed well when evaluated but has not been incorporated in a full NE identification system and so it is difficult to compare their results with others. It is possible that ... |
| 12 | 53 | ...rule classifier over the neighboring words of the token. Yarowsky [18] performs word sense disambiguation by building a sense classifier using the local context of the word and a classifier based on the senses of other occurrences of that word in the same document. Finally, Collins and Singer [4] introduce the CoBoost algorithm to perform named entity classification which boosts classifiers that use either the spelling of the named entity or the context in which that entity occurs. Datasets whose features naturally partition into two sets, and algorithms that use this division, fall into ... |
| 12 | 54 | ...last name, and place. The baseline for this task for Romanian is 98.67 precision and 34.01 recall, yielding an F measure of 50.58. Final system performance is 76.95 precision and 64.99 recall (F measure 70.47) 8 The only precision error is on Zweden which can either mean Sweden or Swedes. (Collins and Singer, 1999) s approach to named entity classification is applied to English only. Instead of using seed lists, they have 7 hand written seed rules (e.g. that any name that contains Mr. belongs to class person) Several algorithms are tested, using name internal and (restricted) contextual clues. Three ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 12 | 55 | ...that rely on hand crafted rules and or supervised learning techniques have limitations in terms of their portability into new domains as well as in the robustness over time. For the purpose of overcoming those limitations, the minimally supervised approach to named entity recognition proposed by (Collins and Singer, 1999; Cucerzan and Yarowsky, 1999) is more promising. The central idea of the minimally supervised approach relates to bootstrapping utilizing redundancy in unlabeled data, with the help of a minimal number of labeled data as initial seeds. The idea of utilizing redundancy in the unlabeled data for ... |
| 13 | 56 | ...for multilingual inheritance based lexical representation which allows sharing of information across (related) languages at all levels of linguistic description. Most work on multilingual lexicons up to now has assumed monolingual lexicons linked only at the level of semantics (MULTILEX 1993; Copestake et al. 1992). Cahill and Gazdar (1999) show that this approach might be appropriate for unrelated languages, as for example English and Japanese, but that it makes it impossible to capture useful generalisations about related languages such as English and German. Related languages share many linguistic |
| 13 | 57 | ...construct the lexicon. The syntactic frames for the verb classes are represented by a Lexicalized Tree Adjoining Grammar augmented with semantic predicates, which allows a compositional interpretation. Introduction Despite many different approaches to lexicon development (Pustejovsky 1991) (Copestake Sanfilippo 1993), Lowe, Baker, Fillmore 1997) Dorr 1997) the field of Natural Language Processing (NLP) has yet to develop a clear consensus on guidelines for computational verb lexicons, which has severely limited their utility in NLP applications. Many approaches make no attempt to associate the semantics ... |
| 13 | 58 | ...and lexical processes. It is possible to write grammars, parse sentences and carry out a limited amount of translation with the LKB. Furthermore, a number of substantial lexicons have been developed semi automatically from machine readable dictionaries (Sanfilippo and Poznanski 1992; Vossen and Copestake 1993), and proposals have been made for representing translation mismatches in the LKB (Sanfilippo et al. 1992) We begin with a description of lexical entries. A lexical entry in the LKB is called a psort. This is simply a TFS which has been given an identifier (cf. sense identifiers in other ... |
| 13 | 59 | ...Linking Rules Tree Families Predicate Argument Lexical Conceptual Elementary Trees Structure Structure Selectional Restrictions Added to LCS as Features in Trees Constraints Semantic Components Predicates of LCS Features for Class Membership 6.2. ACQUILEX DISCUSSION The ACQUILEX system (Copestake and Sanfilippo, 1993; Briscoe et al. 1994) provides a typed feature structure framework for doing MT, based on a HPSG categorical grammar formalism for the source and target languages. Like an LTAG based MT approach, the ACQUILEX MT framework uses a set of bi directional transfer rules, called tlinks, to pair up ... |
| 13 | 60 | ... et al. 1986, Kegl, 1989, Dorr, 1993) the generative lexicon approach of Pustejovsky and his coauthors (e.g. Pustejovsky, 1991, Pustejovsky et al. 1993) and the work of group most recently associated with the ACQUILEX project in Europe (e.g. Boguraev and Briscoe, 1989, Briscoe et al. 1990, Copestake et al. 1992 and selected chapters from Briscoe et al. eds. 1993) A most interesting feature of all these, otherwise distinct, approaches is that they all Page 7 avoid the concept of language neutral ontology in their theoretical framework, while, in reality introducing elements of metalinguistic ... |
| 14 | 61 | ...is with the memory based learning (MBL) approach to machine learning. In MBL all instances are retained and a new instance is classified according to the familiar instances which it most resembles. The approach has recently been shown to be well suited to a range of natural language leaning tasks [Daelemans, van der Bosch, and Zavrelto appear] In MBL, where numbers of instances are similar, they will contribute to future classifications jointly, so do not appear to have roles as individual recollections in memory. Exceptional instances, by contrast, play an explicit role in classification when a new instance matches. Correspondingly, ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 14 | 62 | ...Eager learning can help to obtain theories with coverage far beyond the one guaranteed by lazy learning. For instance, Explanation Based Learning (EBL) is potentially able to learn a theory from a single example. On the other hand, there are tasks where lazy learning guarantees the best results [5, 27]. However, if recall times are an issue, eager learning can be used to compress training examples (previous observations) into a more concise theory in the hope that this will also result in faster recall. Different ways of combining learning with recall can be summarised in the following ... |
| 14 | 63 | ...regularities that are present in the patterns of usage in memory, in combination with the use of an appropriate similarity metric. No abstractions such as grammatical rules or stochastic rules are extracted from the examples; in fact, it is claimed that editing the data in any way is harmful [Daelemans et al. 1999] Overall, the use of MBL in NLP tasks (mainly various disambiguation tasks) has been very successful, but a principled understanding of what makes these methods successful, and what are the relations to other methods is still lacking 1 . 2.1 Technical description Memory based methods are ... |
| 14 | 64 | ...between noise and genuine class exceptions. Recent work suggests that natural language domains, such as word pronunciation, are problematic in the context of instance deletion as the class definitions are not composed of large homogeneous regions but rather many small regions or exceptions (Daelemans et al. 1999). Deleting an instance in this kind of situation is a real problem, and reinforces the point we make in Section 2. we need a knowledge of the problem to effectively deploy a deletion scheme. 3.2 COMPETENCE PRESERVATION Hart s Condensed Nearest Neighbour rule (CNN) was an early attempt at ... |
| 14 | 65 | ...to handle new text genres would likely yield unpredictable results with the transformation based bracketer. In Empire, individual rules can easily be removed from or added to the grammar. In conclusion, we presented a new approach to partial parsing of natural language texts that combines error5 Daelemans et al. 1999) present a learning algorithm for base NP chunking that is similar to MBSL in its memory based approach. Their results, however, are not comparable to those reported here: They measure performance in terms of the number of words correctly bracketed rather than measuring the number of completely ... |
| 15 | 66 | ...inference mechanism and proven to be useful in a task oriented evaluation procedure. To illustrate these features, let us have a look at some recent approaches to statistical disambiguation. One of the most influential works on target word selection in machine translation is described in Dagan and Itai (1994). The SD of Dagan and Itai (1994) uses statistics 118 AIMS VOL. 4 NO. 3 1998 exclusively on monolingual data. The information gathered is statistics on all grammatical relations in which an ambiguous lexical item participates. In an experiment on translating Hebrew to English, their statistical ... |
| 15 | 67 | ...Yarowsky and Schutze minimize the amount of supervision, it is still tremendous in the face of thousands of ambiguous lexicon entries. Both report results only on very few examples (less than 20) The idea of using a second language monolingual corpus for word sense disambiguation is exploited by Dagan and Itai [1994]. They use a target language model to find the correct word level translation. We expand on this notion and achieve better results, as reported below. Research in statistical machine translation [Brown et al. 1993] demonstrates that word level translation models can be learned from large ... |
| 15 | 68 | ...of corpus based approaches to the problem of disambiguation within the same part of speech. However, most of them use additional language specific data, such as thesauri (Yarowsky, 1992) bilingual corpora (Brown et al. 1991# Gale et al. 1993) monolingual (Lesk, 1986) and bilingual dictionaries (Dagan and Itai, 1994). In order to apply these method to some language, the required information for that language has to be obtained before. The method we present is unsupervised, and the only information required is a sufficiently large corpus texts of the language. It can be used for both morphological ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 15 | 69 | ...measures of overlap with dictionary definitions, they have not realized the potential for combining the relatively limited seed information in such definitions with the nearly unlimited co occurrence information extractable from text corpora. Other unsupervised methods have shown great promise. Dagan and Itai (1994) have proposed a method using co occurrence statistics in independent monolingual corpora of two languages to guide lexical choice in machine translation. Translation of a Hebrew verb object pair such as lah . tom (sign or seal) and h . oze (contract or treaty) is determined using the most ... |
| 15 | 70 | ...with the category T_j for which $P(T_j C)$ is highest. Yarowsky found that the algorithm disambiguated about 90 of occurrences correctly for a sample of 12 ambiguous words. 1.3. 3 Disambiguation Based on Translations in a Second Language Corpus The third dictionary based algorithm, proposed by Dagan and Itai (1994), makes use of word correspondences in a bilingual dictionary. Let us call the language of application for which we want to do disambiguation the first language and the target language in the bilingual dictionary the second language. The basic idea of Dagan and Itai s algorithm is best explained ... |
| 16 | 71 | ...Multi Agent Frameworks The term scalability is not always used to refer to architecture, services and performance of systems. In some cases it is used to refer to scalable functionality. For example, the SAIRE approach [19] claims to be scalable because it supports heterogeneous agents. Shopbot [6] claims to be scalable because its agents can adapt to understand new websites. In both cases, the term extensible functionality would seem to be more appropriate. Researchers and developers of multi agent frameworks are beginning to realise that scalability in the sense of architecture, ... |
| 16 | 72 | ...in Latex documents and how to strip position information from postscript files. Harvest neither discovers new documents nor learns new models of document structure. Similarly, FAQ Finder [13] extracts answers to frequently asked questions (FAQ s) from FAQ files available on the web. ShopBot [17] and ILA (Internet Learning Agent) [18] attempt to interact with and learn the structure of unfamiliar information sources. ShopBot retrieves product information from a variety of vendor sites using only general information about the product domain. ILA, on the other hand, learns models of various ... |
| 16 | 73 | ...the concept of collaborative filtering has become widely used, including in simplified ways by large commercial vendors such as Amazon. The ShopBot was an agent that could learn how to submit queries to e commerce sites and interpret the resulting hits to identify lowest priced items [4]. ShopBot automated the process of building wrappers to parse semistructured (HTML) documents and extract features such as product descriptions and prices. Our goals are similar but we focus on learning the user preferences (with respect to many features) and we use a different approach for ... |
| 16 | 74 | ...size=3 (font Figure 4: Simplified example of a vendor module. The module has logic to submit queries to a vendor site and to interpret the results. wrappers employed by shopping bots and the growing complexity of HTML interfaces. Early shopping agents such as the ShopBot [4] demonstrated the interesting learning challenges stemming from this competition. However the IntelliShopper described here does not focus on this goal, therefore we followed a different route in our implementation. Rather than trying to build automatic wrappers, we simplified the task of ... |
| 16 | 75 | ...BargainFinder [8] was able to scan product listings and prices from a set of on line web stores and place them into a unified ordered table. However, it was not extensible, as it was based entirely on hand coded wrappers which needed to be tailored specifically to each source site. ShopBot [7] went a step further by defining a set of heuristics which could be used to automatically parse pages from new sites and extract prices, although the heuristics used were quite specific to parsing on line store pages. In our system, much greater emphasis is placed on the post processing and ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 16 | 76 | ...here. Several researchers have explored the problem of text extraction from the Web and other Internet sources. One example is ILA [50] a system designed to learn the semantics of the human readable output of online databases by comparing it with information whose is already known. Shopbot [18], a bargain hunting agent, is designed to learn patterns from HTML to support the extraction of pricing information from online commercial catalogs. Shopbot is one solution to the general problem of wrapper induction [33,47,26] learning extraction patterns for highly regular sources. At the ... |
| 16 | 77 | ...information from newly discovered Web resources, such that the need for hand coded wrappers to access the resource and parse its response can be eliminated or reduced. Generalization discovers regular patterns at (or inductively learns) individual Web sites and across multiple sites. ShopBot [2], a Web mining agent specialized on electronic catalogs, uses descriptions of domains and vendors as prior knowledge to compare vendors by an attribute (say, price) for given a characterization of the desired product. The domain description includes information about product attributes useful for ... |
| 16 | 78 | ...training (entering and rating CDs) returning a complete set of Miles Davis albums after we entered a reference to John Coltrane (both jazz artists) This system is a fine collection browser, even though it is still limited to one retailer and does not have any way of comparison shopping. ShopBot [4] is a domain independent comparison shopping agent that explores home pages of several vendors on the World Wide Web and learns how to shop. All these systems are interesting shopping experiences, but they do not go all the way in providing a virtual free market metaphor. The system that most |
| 16 | 79 | ... which focus on applying that technology to the Web, such as the Information Manifold [Levy et al. 1996] Occam [Kwok and Weld 1996] Infomaster [Genesereth et al. 1997] and InfoSleuth [Bayardo Jr. et al. 1997] as well as work done specifically about information extraction [Hammer et al. 1997; Doorenbos et al. 1997; Kushmerick 1997] But what is noticeably absent from the literature is a study on what it takes to put together an entire application using the various integration technologies. To that end, we describe the details of how TheaterLoc works and what our plans are for extending the application. ... |
| 16 | 80 | ...that agent technology may have a profound effect upon the way goods are bought and sold. For example, shopping agents, such as Bargain Finder 2 and Jango 3 (a commercial product based on 1 Source: Forrester Research, Inc. 2 http://bf.cstar.ac.com 3 http://www.jango.com the ShopBot [3]) can make online comparison shopping dramatically more efficient, potentially shifting the competitive balance between consumers and retailers. Firefly [12] expands a consumer s awareness by suggesting products in this case, music CDs based upon the reported preferences of others with ... |
| 17 | 81 | ...population of the ontology was to be carried out by a number of project officers within a distance learning group at The Open University who did not have a computing background. It was important that all of the observatory team understood and had ownership of the ontology. Also as outlined in [5] in their analysis of the KA 2 initiative, and in [12] in their description of a SHOE case study, ontology development and representing specific resources are intertwined activities. The conceptual design of the ontology was developed in a series of weekly meetings involving the whole ... |
| 17 | 82 | ...instance edit form. Figure 10. A screen snapshot of the help given when selecting the other involved parties button of the form shown in figure 9. Many errors in semantic annotation occur because of errors in naming existing entities and in selecting the class of new instances [5]. The forms in WebOnto seek to alleviate this by prompting users with the names of relevant knowledge items. An example of an automatically generated form for editing an instance of a learning community is shown in figure 9. Each slot is displayed as a row. The slot name is a button which ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 17 | 83 | ...searching and browsing or to enable information integration # The work of this author is funded by the German Research Society (DFG grant no. SP 572 4 1) with related data sources. Unfortunately, most users are not willing to manually create metadata due to the efforts and costs involved [7]. Thus, text mining techniques are required that (semi) automatically create semantic markup and tag documents accordingly. In this paper, we present the KDD approach pursued in the research project DIAsDEM whose German acronym stands for Data Integration for Legacy Systems and SemiStructured ... |
| 17 | 84 | ...in supporting other portions of the semantic web lifecycle, there has been little progress in the markup of manually composed documents. The prevalent approach is to create specialized tools that specifically support the association of semantic markups with the content of existing documents [1], 2] These tools provide a GUI that permits an author to browse ontologies, find appropriate terms, generate syntactically correct markups, and associate them with (portions of) the document s content. This activity remains an extra effort that does not directly reward to the person performing ... |
| 17 | 85 | ...is able to generate DAML descriptions from the talks contained in its database. In this sense, ITTALKS is an example of a tool that generates descriptions from highly structured data. Closer to the scope of the BA are the Annotation Tool of the KA 2 initiative (under the Ontobroker project) [1] and the Knowledge Annotator of the Shoe project [2] These tools offer a GUI for authoring and attaching semantic annotations to web documents. They make available context sensitive instances and ontology browsers that facilitate the authoring of semantic descriptions. A second incarnation of the ... |
| 18 | 86 | ...systems. 1.4.2 Literature review (Chapter 4, 5) In Chapter 4, the organization of lexical semantic knowledge is discussed. Most of the hierarchical organization of lexical knowledge is used for syntactic information processing, similarly as Kilgarrieff and Flickinger's work (Kilgarrieff, 1992) (Flickinger and Nerbonne, 1992). Verb semantic representation has a long history. Chapter 5 chronicles the historically important theories in this issue. From Fillmore, Jackendoff, Dowty and Levin, we attempt to draw a line of semantic theory development within the generative framework. What we attempt to point out is that the ... |
| 18 | 87 | ...to some features, there is no way to get the subnodes of a word. When word semantics is considered, this style of structured lexicon that maintains only the inheritance relation will not be enough for defining a good lexicon. 4. 3 Flickinger and Nerbonne: Easy adjectives Flickinger and Nerbonne (Flickinger and Nerbonne, 1992) illustrated that structured lexicon has its practical advantages. By analyzing a concrete example the linguistic properties of the easy adjective, they show that the structured lexicon is easy to be maintained, modified and extended. The following example sentences are used to illustrate the ... |
| 18 | 88 | ... a thorough going account of the interaction of lexical probabilities with probabilities associated with specific sentential interpretations, or 28 Modulo the probabilistic interpretation, this manner of encoding the (non)application of a lexical rule has been deployed in many theories; e.g. Flickinger and Nerbonne (1992) and Sanfilippo (1993) in recent accounts of verbal diathesis alternations. 29 It is plausible to imagine that language users are able to memorise some estimate of the relative frequency with which a word form and sense occur, though it is unlikely that this process is accurate enough to derive ... |
| 18 | 89 | ...least because, given appropriate tools, both general and idiosyncratic properties of language can be captured within a uniform framework. Among the tools normally employed one finds lexical rules (Dowty 1978; Flickinger 1987; Pollard and Sag 1994) and inheritance mechanisms (Briscoe et al. 1993; Flickinger and Nerbonne 1992). Lexical rules may be thought of as establishing a relationship between lexical items such that given the presence of one lexical item in the lexicon the existence of a further item may be inferred. The regularities captured by lexical rules might include changes in the subcategorization and ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 18 | 90 | ...for cases such as (85) where the phrase easy to play on has a nonempty slash value. 85) Which violins are these sonata easy to play on 4. 5 Subbinding The analysis just presented has the further advantage that it provides a potential analysis for more complicated cases, such as those noted in Flickinger and Nerbonne (1992): 86) an easy man to please Flickinger and Nerbonne argue for the following constituent structure: 87) N N ADJ easy N man VP to please The adjective easy is not the head of this structure, nor is it the head of the substructure easy man. Under a configurational approach, such as that presented |
| 19 | 91 | may seem obvious that language processing systems should be subject to empirical criteria of effectiveness. The big problem, of course, is determining precisely what the criteria of success should be. There is now a substantial recent literature on this question [Crouch et al. 95, Sparck Jones 94, Sparck Jones Galliers 96, Gaizauskas 97, Gaizauskas et al. 98] and practical solutions to the evaluation problem have emerged in a number of areas. Participants in the MUC (Message Understanding Conference, an Information Extraction competition) and TREC (Text REtrieval Conference, a competition for Information ... |
| 19 | 92 | ... all problems, a final issue, both in generation and in dialogue: how to evaluate the text produced In particular, how can we say that a text is coherent Some researchers have given hints on what coherence is not 3 , and evaluation heuristics have been identified for some categories of systems [9]. However, researchers in NLP still have to agree on a standard approach, and this is probably not surprising (how can we say that a text is good) Can Argumentation Help As most, if not all, of the Argumentation process is based on the use language it is surprising that Argumentation theory ... |
| 19 | 93 | ...part addresses whether Rhetorica output is also persuasive, by tackling the key problem of system evaluation. VII. PERORATION 149 VII Peroration 7. 1 Appraisal The evaluation of natural language processing systems presents a range of problems and has become a field of research in its own right (Galliers and Sparck Jones, 1993). Evaluation of the Rhetorica system is confounded by several further problems. In the first instance, it is not currently applied in any specific domain: Galliers and Sparck Jones make repeated use of case studies often in information retrieval to examine evaluation methods. In such ... |
| 19 | 94 | ..adequate for system design purposes) and a value of a good coverage at one point versus bad coverage at another is not in itself indicative of the system s fitness to a user s purpose. Evaluation is also likely to be affected by the laboratory situation in which it takes place: as pointed out by [8] what is being evaluated is a setup, a system embedded in a context of use. HCI literature stresses that the usability of a system can be properly assessed only in real situations. The evaluation goals may also be interdependent: a system may gain accuracy at the expense of real time interaction. ... |
| 20 | 95 | ...process than centering would be required to explain the relationship between utterances (32) and (33) simply because these utterances span a discourse segment boundary. The second problem is that listeners perceive segment boundaries at various levels of granularity [Pasonneau and Litman, 1993; Hearst, 1994; Flammia and Zue, 1995; Hirschberg and Nakatani, 1996] and some segment boundaries are fuzzy [Pasonneau and Litman, 1996] For example in discourse A above, 5 out of 7 subjects placed a segment boundary between utterances 29 and 30, while 4 out of 7 subjects placed a segment boundary ... |
| 20 | 96 | ...into a larger number of smaller segments might be possible and be necessary for the given texts. And so we will have to consider the evaluation method that the agreement with human subjects is tested in future. However, since human subjects do not always agree with each other on segmentation[5, 3, 14], our evaluation method using the texts in the questions with model answers is considered to be a good simplification. Several other methods to text segmentation have been proposed. Kozima[7] and Youmans[17] proposed statistical measures(they are named LCP and VMP respectively) which indicate the ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 20 | 97 | ... on where segment boundaries are, either because they construct different mental representations of the segmentation of a discourse, or because segments are naturally defined at varying levels of granularity (Passonneau and Litman, 1993; Grosz and Hirschberg, 1992; Passonneau and Litman, 1994; Hearst, 1994). To illustrate the problem, consider the continuation in 14 of the discourse excerpt in 9 from (Walker and Prince, In Press) 14) it was an emergency for her j to pick up the phone right away. Her i sister] j not being home, she i hung up. Her i sister] j came home a short time later, ... |
| 20 | 98 | ...on the Internet. The FAQ Finder project has shown that when there is an existing collection of questions and answers, as found in the FAQ files, question answering can be reduced 12 Non frivolous examples of such files also exist in the RTFM archive. 13 The TextTiling technique proposed in (Hearst, 1993) seems particularly suitable here. SOLAR STAR SYSTEMS Acamar TNG The Vengeance Factor Alpha Centauri TOS Metamorphosis . Minos Korva (11 lightyears from McAlister C5 ... |
| 20 | 99 | ... conclusion:premise Phi Phi H H Also D core:evidence conclusion:premise Phi Phi H H and E Figure 2: The RDA analysis of (1) These rates of agreement are similar to those found in studies of (nonembedded) segmentation agreement (Grosz Hirschberg 1992; Passonneau Litman 1993; Hearst 1993). However, our assessment of RDA reliability differs from this work in several key ways. For one thing, our subjects coders are not naive about their task and the data is not spoken. Further, the task is more complex than identifying locations of segment boundaries. Example hypotheses and initial ... |
| 21 | 100 | ...language, LO [4] The work of Miller and Hodas [20] describes a linear logic programming language that refines the logics behind both Prolog and Prolog. A detailed analysis of this linear logic programming language is given in [20] including a semantics based on Kripke models. Hodas [19] develops this linear logic programming language further, and explains the natural language parsing issue of gap threading parsers within this logic programming setting. As part of this work, an implementation of this linear logic programming language was built in Standard ML. This system, called ... |
| 21 | 101 | ...an upgrading of the grammar than by the use of some extra grammatical table of operators or table of types. Neither Prolog nor HHG handle at the rule level such constraint as that some rule must be used exactly once. There are proposals for using linear logic programming for solving this problem [13, 6]. We believe that our method of handling logical connectives at the grammar rule level applies to the linear connectives too. Note that when, as usual, the strategy of interpreters for Horn clause programs is recursive descent, the resulting DCG analyser is a poor one. However, complete bottom up ... |
| 21 | 102 | ...(A] The restrictions on the weakening rule in linear logic require every (linear) assumption to be eventually used. Often, when assumptions range over the current continuation, this requirement seems too strong, except for the well known situation of handling relatives through the use of gaps [Hod92]. Therefore, BinProlog s linear implication will succeed even if not all the assumptions are consumed (weakening is allowed) while in systems like Lolli their consumption is a strong requirement, i.e. it is enforced for success. We found our choice practical and not unreasonably restrictive, ... |
| 21 | 103 | ...[16] has clear connections with linear logic. Abrusci and De Paiva [2, 7] independently proved the Lambek calculus to be a fragment of intuitionistic linear logic, and Hodas and Miller [13] have used linear logic to extend Definite Clause Grammars via a linear logic programming language. Hodas [12] has also discussed parsing with gaps based on a similar linear logic programming language. Within computational linguistics, Hepple [11] and Morrill [20] have shown that substructural logic systems can be used directly to characterize constructions in natural language. 1.3. The Approach of This ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 21 | 104 | ... features such as those in Algol and ML [Chirimar 1995, Miller 1994] Chirimar has also specified in Forum the operational semantics of a pipe lined, RISC processor [Chirimar 1995] Natural language parsing Lolli has provided a declarative approach to gap threading within English relative clauses [Hodas 1992]. Object logic proof systems Lolli has been used to refine the usual, intuitionistic specifications of object level natural deduction systems [Hodas Miller 1994] and Forum has been used to provide specifications of object-level sequent systems [Miller 1994] 5 Research in Sequent Calculus Proof ... |
| 22 | 105 | ... While AutoSlog s patterns perform best when semantic class information is available, the learning algorithm and the resulting concept nodes can still operate effectively when no semantic class information can be obtained. There have been a few additional attempts to learn extraction patterns. Huffman s LIEP system [1996] learns patterns that recognize semantic relationships between two target noun phrases, i.e. between two slot fillers of an information extraction output template. The patterns describe the syntactic context that falls between the target noun phrases as well as the semantic class of the heads of ... |
| 22 | 106 | ...however, there have been several efforts to automate the acquisition of extraction patterns. Two of the earliest systems to generate extraction patterns automatically were AutoSlog [Riloff, 1993] and PALKA [Kim and Moldovan, 1993] More recently, CRYSTAL [Soderland et al. 1995] and LIEP [Huffman, 1996] have been developed. All of these systems use some form of manually tagged training data or user input. For example, AutoSlog requires text with specially tagged noun phrases. CRYSTAL requires text with specially tagged noun phrases as well as a semantic hierarchy and associated lexicon. PALKA ... |
| 22 | 107 | ...of information showed good results. This indicates that the ILP system RHB has a high potential in IE tasks. 7 Related Work Previous researches on generating IE rules from texts with templates include AutoSlogTS (Riloff,1996) CRYSTAL (Soderland et al. 1995) PALKA (Kim et al. 1995) LIEP (Huffman, 1996) and RAPIER (Califf and Mooney, 1997) In our approach, we use the type-oriented ILP system RHB , which is independent of natural language analysis. This point differentiates our approach from the others. Learning semantic level IE rules using an ILP system from semantic representations is also ... |
| 22 | 108 | ...IE systems is the high cost involved in manually adapting them to new domains and text styles. In recent years, a variety of Machine Learning (ML) techniques has been used to improve the portability of IE systems to new domains, as in SRV (Freitag, 1998) RAPIER (Califf and Mooney, 1997) LIEP (Huffman, 1996), CRYSTAL (Soderland et al. 1995) and WHISK (Soderland, 1999) However, some drawbacks remain in the portability of these systems: a) existing systems generally depend on the supported text style and learn IE rules either for structured texts, semi structured texts or free text , b) IE systems ... |
| 22 | 109 | ...extracting only limited structures from the input text. Creating dictionaries of concept patterns for information extraction consumes a great amount of time and is required for each new domain. Research efforts have begun to address this: Cardie (1993) Riloff (1993) Soderland et al. 1995) Huffman (1996) . These corpus based methods all employ a large training corpus annotated with examples for each concept. From these examples, machine learning algorithms induce conceptual patterns for extraction. However, these methods have not eliminated the cost of building information extraction systems, but ... |
| 23 | 110 | ...textual cohesion, balance and coverage, it is possible to produce domain independent summaries that are indicative [12] 3. Summary Evaluation In order to compare the quality of summaries produced by different ATS systems it is important to have some form of standard evaluation. Hongyan et al. [5] suggests that one of the main failings in the field of automatic summarisation is the lack of just such a methodology. Many developers adopt non standard techniques that are only suitable for their particular implementation making direct comparison across systems impossible. However, this does ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 23 | 111 | ...has been to measure the similarity between summaries that are produced automatically and by hand. However, this evaluation method has been criticized because it assumes that there is only one correct summary. A task-based evaluation scheme has been recently adopted as new way of evaluating summaries(Jing et al. 1998; Mani et al. 1998; Tombros and Sanderson, 1998) It evaluates the performance of a summarization system in a given task, such as information retrieval and text categorization. This paper compares ten different summarization methods based on information retrieval tasks. To evaluate the system ... |
| 23 | 112 | ... keep the percentage of documents relevant to the query higher than 50 because a smaller number of relevant documents would make the results of the 1 We admit that we should make more thorough experiments with multiple summary lengths, since different summary lengths will yield different results(Jing et al. 1998; Mittal et al. 1999) 2 BMIR J2 was constructed by the SIG Database Systems of the Information Processing Society of Japan, in collaboration with the Real World Computing Partnership. experiments less reliable. The average length of the queries is 3.2 words, and the average length of the ... |
| 23 | 113 | ...for generic summaries (an indicative summary) 3) establishing whether summaries can answer a specified set of questions (an informative summary) by comparison to an ideal summary. In each task, the summaries were rated in terms of confidence in decision, intelligibility and length. Jing et al. [10] performed a pilot experiment (40 sentences) in which they examined the precision recall performance of three summarization systems. They found that different systems achieved their best performance at different lengths (compression ratios) They also found the same results for determining ... |
| 23 | 114 | ... research is notorious for its lack of adequate corpora, a situation that prevents rapid progress in the field: today, there exist only a few small collections of texts whose units have been manually annotated for textual importance [Edmundson, 1968, Kupiec et al. 1995, Teufel and Moens, 1997, Jing et al. 1998, Marcu, 1999] Given the cost and tediousness of the annotation process, it is very unlikely that we will ever manually annotate for textual importance sufficiently large corpora. To circumvent this problem, we have developed an algorithm that constructs such corpora automatically. 1.2 Towards ... |
| 24 | 115 | ...of a new subject of discussion, a new set of topics. This creates a burst of new lexical chains starting at this boundary, and possibly another group of chains that end there. Thus, lexical chains are well accepted predictors of discourse boundaries (Morris and Hirst, 1991; Hearst, 1994; Kan, Klavans, and McKeown, 1998). What is unique in Texplore is that the lexical chains represent more salient topics of the text, as discovered by n gram analysis and by anaphora resolution. This has the potential of providing a better prediction for boundary analysis, and, as we will see later, for facilitating hierarchical ... |
| 24 | 116 | ...carefully because of the difference in text genre and in subject performance. In comparison to the data published by Hearst, the figures here are much lower. The reason for this is the lower density of boundary assignment in our test corpus (0.22 vs. Hearst 0. 39) The performance data reported in (Kan, Klavans, and McKeown, 1998) shows even lower performance figures, both for Hearst s method and for the author s. However, preferring one method over the other is not very reliable unless the algorithms are applied on the same texts. 3.4.2 Evaluating hierarchy reconstruction To score the hierarchy reconstruction, we use ... |
| 24 | 117 | ... under the covers , service function used by the summarizer, or might the results of discourse segmentation be of any interest, and use, to the end user We discuss, in the following section, strategies for incorporating segmentation results in the summary generation process. However, unlike (Kan, Klavans, and McKeown, 1998) whose work also seeks to leverage linear segmentation for the explicit purposes of document summarization, we further take the view that with an appropriate interface metaphor where the user has an overview of the relationships between a summary sentence, the key salient phrases within it, and ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 24 | 118 | ...such as what is the source of information, and whether information is being presented as fact or opinion. These questions are particularly important in news reporting, in which segments presenting opinions and verbal reactions are mixed with segments presenting objective fact (van Dijk, 1988; Kan et al. 1998). The definitions of the categories in our coding manual are intention based: If the primary intention of a sentence is objective presentation of material that is factual to the reporter, the sentence is objective. Otherwise, the sentence is subjective. 1 We focus on sentences about ... |
| 24 | 119 | ...raises two related questions. The first concerns the relationship between segmentation and summarization: is segmentation a strictly under the covers , service, function used by the summarizer, or might the results of discourse segmentation be of any interest, and use, to the end user Unlike [17] (whose work also seeks to leverage linear segmentation for the explicit purposes of document summarization) we take the view that with an appropriate interface metaphor, where the user has an overview of the relationships between a summary sentence, the key salient phrases within it, and its ... |
| 25 | 120 | ... language, the corpus based learning approach is used to recognise unknown words and typographical errors [5] Most systems make heavy use of linguistic and natural language processing (NLP) techniques in order to resolve the ambiguity of the language structure, and to improve the IR result [17]. Grammatical and semantic information are becoming more essential as this information helps the analyser to understand more about the role of words in the sentences. In [13, 14, 30] the n gram probabilistic model is used to improve the efficiency of the information retrieval system. Kawtrakul [13] ... |
| 25 | 121 | ... This index, document s content representation, would then help the user in retrieving this document. Thus, retrieval depends on indexing , that is on some means of indicating what documents are about. Indexing is the basis for retrieving documents that are relevant to the user s need. [31]. The main aim of indexing is to increase precision. One of the main problems is to obtain an accurate representation of that document, which will be stored by our proposed CINDI system. A document representation could for example be a list of extracted words considered to be significant, if not ... |
| 25 | 122 | ... Retrieval (IR) domain, usually to improve precision and recall [12, 13, 14] Natural language processing has been used to automatically generate concept thesauri, generate document summaries, handle natural language queries, and reduce the feature space for vector space models, as discussed in [15]. Term clustering has also been used for automatic thesaurus generation, as well as for document clustering [16] However, these techniques have rarely been used to understand a collection, as opposed to individual documents. Perhaps the best work in understanding collections is the Topic ... |
| 25 | 123 | ...not quite as high as the best reported results of [Joachims, 1998] and [Lam and Ho, 1998] The fact that this technique improved the micro averaged breakeven point validates the prediction that the various representations were making uncorrelated errors. 8. Conclusions and Future Directions Lewis and Sparck Jones [1996] wrote that statistical techniques for classification and retrieval have picked some of the low hanging fruit off the tree. They believe that significant advances must be made before Natural Language Processing techniques can be used to improve text classification. The results of our work ... |
| 25 | 124 | ...to create more accurate indexing terms (Woods W. A. 1997; Strzalkowski T. et al 1998) However, to the best of our knowledge, no DR system produces full syntactic parses of either documents or queries. In fact, the common belief holds that it does not pay off to use deep linguistic analysis in DR (Lewis D. D. Sparck Jones K. 1996). Information Extraction (IE) techniques are similar to DR techniques in that they, too, are suitable for processing text collections of basically unlimited size (normally, messages in a stream) covering a potentially wide range of topics. However, IE systems differ from DR systems in that they ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 26 | 125 | ...or italic characters, etc. This technique might be more useful for summarising Web documents since it can make use of small portions of information, although it still can not handle graphic content. The third and last trend is the discourse model based extraction and summarisation ([15] 7] 19] [16] [17] 25] This technique is using natural language cues in the text such as lexical choice in the text, order of words, proper names identification, reiterations, synonymy, anaphora, lists of predefined cue phrases, connectives, etc. Again, this technique assembles sentences from the text based ... |
| 26 | 126 | ...resources for the scientific and technical domain. We addressed this by constructing our own evaluation resources with technical articles published on electronic journals on the Web. We use as Gold Standard for evaluation the abstracts published with the source documents (as was the case in (Lin and Hovy, 1997) but for different purposes) and we compared the terms appearing in the automatic abstracts with the terms appearing in the journal provided abstracts. We do not compare sentences with sentences because the abstracts published together with source documents usually contain sentences difficult to ... |
| 26 | 127 | ... 1968) ffl important sentences are located at the beginning or end of paragraphs (Baxendale 1958) ffl important sentences are located at positions in a text that are genre dependent, and these positions can be determined automatically, through training techniques (Kupiec, Pedersen, Chen 1995; Lin Hovy 1997; Teufel Moens 1997) ffl important sentences use bonus words such as greatest and significant or indicator phrases such as the main aim of this paper and the purpose of this article , while unimportant sentences use stigma words such as hardly and impossible (Edmundson 1968; Rush, ... |
| 27 | 128 | ...attention of NLP researchers along with the IE(Information Extraction) IR(Information Retrieval) and IF (Information Filtering) technique recently. Many automatic abstracting systems have been proposed. For example, SUMMONS [McKeown et al., 1995; Radev et al., 1998] SUMMARIST [Hovy et al., 1997; Lin, 1998], COSYMATS [Aretoulaki, 1997] SUMMAC [Sanderson, 1998] SJTUCAA [Wang et al., 1996] FDASCT [Wu et al., 1996] and so on. Tombros(1997) presented a general automatic text abstracting model which generates the abstract of the text in two steps: the source text interpretation and the target text ... |
| 27 | 129 | ...are useful in most cases and most hypnoses that are important to abstracting systems (i.e. the title, keywords, cue words and the position of the sentence or the paragraph in the document) are based on the statistical information. Some hybrid approaches, for instance, SUMMARIST[Hovy et al., 1997; Lin, 1998], COSY MATS[Aretoulaki, 1997]#FDASCT[Wu et al., 1996] have been tried by a few researchers. In this paper, we propose an ANN based automatic sentence extraction approach. We have implemented an automatic Chinese text abstracting system based on it. In order to illustrate how to use an ANN ... |
| 27 | 130 | ...Typically, statistical approaches, augmented with key word or phrase matching, are used to identify which full sentences in the article can serve as a summary. Many schemes to rate sentences and methods for combining ratings exist [Paice, 1990, Kupiec et al. 1995, Mani and Bloedorn, 1997, Lin, 1998] Most of the work in this category produces a summary for a single article, although there are a few exceptions. The second two categories correspond to the two stages of processing that have to be carried 131 out if sentence extraction is not used: analysis of the input document to process and ... |
| 27 | 131 | ...characters, etc. This technique might be more useful for summarising Web documents since it can make use of small portions of information, although it still can not handle graphic content. The third and last trend is the discourse model based extraction and summarisation ([15] 7] 19] 16] [17] [25] This technique is using natural language cues in the text such as lexical choice in the text, order of words, proper names identification, reiterations, synonymy, anaphora, lists of predefined cue phrases, connectives, etc. Again, this technique assembles sentences from the text based on ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 28 | 132 | ...Luhn's work at IBM in the fifties [12] Most of the work in sentence extraction applied statistical techniques (frequency analysis, variance analysis, etc. to linguistic units such as tokens, names, anaphora, etc. e.g. 27, 19, 9, 18, 2] Other approaches include the utility of discourse structure [14], the combination of information extraction and language generation [11, 17, 24, 21, 16] and using machine learning to find patterns in text [28, 4, 26] Several researchers have extended various aspects of the single document approaches to look at multi document summarization [13, 21, 3, 7, ... |
| 28 | 133 | ...operational environment, even very simple heuristics such as, for instance, take the first sentence from each segment have remarkably noticeable impact. In essence, this paper argues that a lexical repetition based model of linear segmentation offers highly 2 As opposed to hierarchical; see (Marcu, 1997). plausible schemes for deriving sentence based summaries with certain discourse properties, as a result improving upon an already respectable system. What follows is organized in four main sections. Our summarizer benefits from a number of linguistic analysis filters; these, as well as some ... |
| 28 | 134 | ...approach (Kupiec et al. 1995) uses a corpus of articles with summaries for training to identify the features of sentences that are typically included in abstracts. Other approaches use lexical chains (Barzilay and Elhadad 1997) sentence position (Lin and Hovy 1997) discourse structure (Marcu 1997; Marcu 1998) and user features from the query (Strzalkowski et al. 1998) to find key sentences. While most of the work to date focuses on summarization of single articles, early work is beginning to emerge on summarization across multiple documents. Radev and McKeown 1998) use a symbolic ... |
| 28 | 135 | ... more sophisticated techniques are being deployed in attempts to improve the quality of sentence-based summaries, by seeking to mediate the passage selection process with, for instance, strong notions of topicality (Hovy Lin 1997) lexical chains (Barzilay Elhadad 1997) and discourse structure (Marcu 1997), Reimer Hahn 1997) 1 Also at: http://www.nytimes.com/library/cyber/digicom/012797digicom.html . 1.3 Capsule overviews The approach we take in this work, while addressing a slightly different problem to that of strict summarisation, can be construed as striving for the best of both worlds. ... |
| 28 | 136 | ...how various discourse processing techniques (e.g. rhetorical structure relations) can be used to both identify important information and form the actual summary. While promising, this work does not involve an implementation as of yet, but provides a framework and strategies for future work. [Marcu, 1997] uses a rhetorical parser to build rhetorical structure trees for arbitrary texts and produces a summary by extracting sentences that span the major rhetorical nodes of the tree. 138 In addition to domain specific information extraction systems, there has also been a large body of work on ... |
| 28 | 137 | ...defines the macro-level semantic structure of a connected discourse, while cohesion creates connectedness in a non structural manner. Coherence is represented in terms of coherence relations between text segments, such as elaboration, cause and explanation. Some researchers, e.g. 23] and [19], use discourse structure (encoded using RST [18] as a source representation for summarization) Discourse representation can be used to prune a hierarchical tree of discourse segments and keep only the nucleus of the discourse. In contrast to lexical cohesion, however, coherence is difficult to ... |
| 29 | 138 | ...comprehensive document, containing the information of all original documents without repeating information which is conveyed by two or more documents. The work described in this paper is closely related to the area of multi document summarization (Barzilay et al. 1999; Mani and Bloedorn, 1999; McKeown and Radev, 1995; Radev, 2000) where related documents are analyzed to use frequently occurring segments for identifying relevant information that has to be included in the summary. Our work differs from the work on multi document summarization as we focus on document fusion disregarding summarization. On the |

| ABS # | Citation # | Citation Context Text |
|-------|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 29 | 139 | ...Automatic Summarization Automatic Abstracting is receiving more and more attention of NLP researchers along with the IE(Information Extraction) IR(Information Retrieval) and IF (Information Filtering) technique recently. Many automatic abstracting systems have been proposed. For example, SUMMONS [McKeown et al., 1995; Radev et al., 1998] SUMMARIST [Hovy et al., 1997; Lin, 1998] COSYMATS [Aretoulaki, 1997] SUMMAC [Sanderson, 1998] SJTUCAA [Wang et al., 1996] FDASCT [Wu et al., 1996] and so on. Tombros(1997) presented a general automatic text abstracting model which generates the abstract of the text in two ... |
| 29 | 140 | ...to produce a summary [KR96] 8 These systems and Thinksheet share the common goal of providing only the relevant information to the reader, but summary generators operate on a different class of documents. Generally, they have been used to generate summaries of newspaper or magazine articles [AL97, BE97, MR95]. These papers may have complicated subject matter i.e. they may be technical articles [TM97] but they generally do not meet our criteria for complex documents because the structure of the articles is mostly simple and linear. These systems also do not provide the individual tailoring capability ... |
| 29 | 141 | ...certain core entities and facts in a document, which are packaged together in a template. There are shared intuitions among researchers that generation of smooth prose from this template would yield a summary of the document s core content; recent work, most notably by McKeown and colleagues, cf. (McKeown Radev 1995), focuses on making these intuitions more concrete. While providing a rich context for research in generation, this framework requires an analysis front end capable of instantiating a template to a suitable level of detail. Given the current state of the art in text analysis in general, and of ... |
| 29 | 142 | ...representation in order to create a summary. There are three types of source text information: linguistic, domain and communicative. Each of these text aspects can be chosen as a basis for source representation. Summaries can be built on a deep semantic analysis of the source text. For example, in (McKeown and Radev, 1995), McKeown and Radev investigate ways to produce a coherent summary of several texts describing the same event, when a detailed semantic representation of the source texts is available (in their case, they use MUC style systems to interpret the source texts) Alternatively, early summarization |
| 30 | 143 | ...if a costly tuning procedure is required before applying any existing system to each new domain. Due to this fact, recent works have focused on reducing the acquisition cost as well as the need for supervision in corpus based methods. It is our belief that the research by (Leacock et al. 1998; Mihalcea and Moldovan, 1999) 2 provide enough evidence towards the opening of the bottleneck in the near future. For that reason, it is worth further investigating the robustness and portability of existing supervised ML methods to better resolve the WSD problem. It is important to note that the focus of this work will ... |
| 30 | 144 | ...LazyBoosting on the WSD task. This would include taking into account additional alternative attributes and testing the algorithm in other corpora specially on sense tagged corpora automatically obtained from Internet or large text collections using unsupervised methods (Leacock et al. 1998; Mihalcea and Moldovan, 1999). Since most of the knowledge learned from a domain is not useful when changing to a new domain, further investigation is needed on tuning strategies, specially on those using non supervised algorithms. It is known that mislabelled examples resulting from annotation errors tend to be hard ... |
| 30 | 145 | ... art systems. We test how far can we go with existing hand tagged corpora like SemCor (Miller et al. 1993) and the DSO corpus (Ng and Lee, 1996) which have been tagged with word senses from WordNet. Besides we test an algorithm that automatically acquires training examples from the Web (Mihalcea Moldovan, 1999). In this paper we focus on one of the most successful algorithms to date (Yarowsky 1994) as attested in the Senseval competition (Kilgariff Palmer, 2000) We will evaluate it on both SemCor and DSO corpora, and will try to test how far could we go with such big corpora. Besides, the ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 30 | 146 | ...word sense disambiguation algorithm with very good results, but no provision was made to enrich WordNet with them. The main shortcoming of this strategy is that limiting the search to monosemous relatives, only 65 of the concepts under study could get training examples. Mihalcea and Mondovan [23] present a similar work which tries to improve the previous method. When a monosemous synonym for a given concept is not found, additional information from the definition of the concept is used, in the form of defining phrases constructed after parsing and processing the definition. The whole ... |
| 30 | 147 | ...to the concept. This allows us to be less constraining; the more documents the better, because that allows to found more distinctively co occurring terms. That is why we chose to use all close relatives for a given concept, in contrast to [22] which only focuses on monosemous relatives, and [23], which uses synonyms and a different strategy to process the gloss. Another difference is that our method forbids the cue words of the rest of the senses. We have found that searching the web is the weakest point of our method. The quality and performance of the topic signatures and clusters ... |
| 31 | 148 | ...and length of exhibit descriptions, and records showing which exhibits the visitor has seen and what the system has told the visitor about them. These records allow the system to avoid repeating information that has already been conveyed, and to compare the current exhibit to previous ones [Milosavljevic Dale 1997; Milosavljevic Oberlander 1998; Milosavljevic 1999] In Figure 3, for example, the description reminds the user that both the current and the previous exhibit were created in the archaic period, helping the visitor build a more coherent view of the collection. Following ILEX, the user model ... |
| 31 | 149 | ...and characteristics. Since users expect real time interaction, efficient and robust applied NLG techniques are typically used for hypertext generation. For instance, ILEX [Knott et al. 1996] uses a combination of canned stories and templates; EXEMPLARS [White, 1998] is rule based; and PEBA [Milosavljevic et al. 1996] uses text schemas [McKeown, 1986] and a phrasal lexicon. Also for efficiency reasons, dynamic hypertext generation systems have pipeline architectures where modules are executed sequentially and no module later in the architecture can request information from an earlier module. For example, 1 In ... |
| 31 | 150 | ...as much traffic per user relative to than the normal Web page (Basse, 1999) Adaptive systems attempt to anticipate the needs and desires of the particular user. To do this effectively, the application design must be based on a model of the user (Knott, Mellish, Oberlander, O Donnell, 1996; Milosavljevic, Tulloch, Dale, 1996). Some systems develop this model utilizing the user s previous actions. Other An adaptive system may also gather information by monitoring what the user is doing, or the system may ask questions of the user. For example, some intelligent tutoring systems build a user model based on what reading ... |
| 31 | 151 | ...of the link level, i.e. the navigation structure, and the personalization of the content level, i.e. the information to be presented [7] Some researchers, like [8] have focused on the dynamic adaptation of the hypertextual structure to users with different backgrounds. Others, like [16], [17] [12] 9] and [10] have focused on the dynamic generation of text tailored to the user. Some recent applications are also focused on the generation of personalized presentations exploiting life like characters [1] Although, as mentioned in section 1, e commerce has strong adaptivity ... |
| 31 | 152 | ...also recognized very early the importance of the user model [16] focusing mainly on the user s level of expertise. More recent work takes into account a more dynamic model of the user s expertise by tracing her browsing behaviour and adapting object descriptions to earlier visited hypertext nodes [17]. Our approach aims at capturing the extra linguistic context of the user, including its most dynamic aspects, i.e. her minute to minute evolving intentions. The mechanism of competition for attention constitutes an innovative way of treating the user s context. It can be compared with ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 31 | 153 | ...hyperbook is an information repository which integrates and personalizes a set of distributed information using explicit conceptual models. Research on the development of hyperbooks has focused on the educational sector, where hypertext technologies are used to implement learning environments [81, 68, 10, 19, 25, 110, 51, 54] and intelligent tutors [7, 9] These systems provide the contents which have been covered so far in normal text books, and integrate them into a hypertext system which guides the users during their learning processes. 4.2 Conceptual Modeling for Adaptive Hyperbooks Structuring concepts in domain |
| 31 | 154 | ...of decomposition in the lexicon, and the system can more efficiently generate texts expressing particular concepts. It is better to explicitly represent the realisation of certain complex concepts which are repeatedly realised the same way, to avoid rebuilding the surface form for each occurrence (Milosavljevic, Tulloch and Dale 1996). The use of a phrasal lexicon clearly has benefits for monolingual generation systems for which the level of granularity of elements in the knowledge representation can directly correspond to the level of granularity in the lexicon. We are currently exploring the extension of these techniques to ... |
| 31 | 155 | ...a certain amount of adaptation to the context, such as production of effective referring expressions. The system keeps track of the discourse history and does not repeat the description of a museum item: a pointer to the previously generated description is given instead. PEBA II [Dale and Milosavljevic 1996, Milosavljevic et al. 1996] dynamically generates hypertext descriptions of a zoological database through an online interface. Discourse history is used to obtain context sensitive text. The authors argue that hypertext significantly eases the user modeling task since part of the content selection is made by the user. ... |
| 32 | 156 | ...of corpus based supervised statistical methods to Turkish text processing. 1 Although there are earlier studies on information retrieval, such as [Solak and Can, 1994] and language processing (such as machine translation, parsing, morphological analysis and disambiguation) for Turkish [Oflazer, 1993; Hakkani and Oflazer, 1998; Oflazer and Tur, 1997; Hakkani et al. 1998, among others] neither of these systems employ statistical techniques. The use of statistical techniques in processing Turkish text had a number of problems in general: ffl Currently there is no Turkish corpus large enough ... |
| 32 | 157 | ...sense is defined as a concept instance. Since morphological analysis is a well known topic with satisfactory computational models, how morphological analysis of Turkish is achieved is not explained here. Morphological analyses of Turkish words are directly taken from an engine developed by Oflazer [9]. In the rest of this section, semantic analysis which is the core of this paper is explained in detail, and TMR construction methodology is presented with some examples. 4.1 Semantic Analysis In order to simplify the presentation, first the core idea of the methodology is presented through ... |
| 32 | 158 | ...of feeding the parser with sentences and printing the output of the parser in a suitable format. We will explain in some detail the ATN parser, network definitions, and arguments of verbs later in this chapter. For the morphological analysis and Turkish lexicon one can refer to Oflazer [13] if more information is needed. The current version of grammar includes an S network which includes frequently used simple and complex sentence structures of Turkish. The network makes use of two other networks: NP and ADVP. The NP network is the most commonly used one and is called recursively by ... |
| 32 | 159 | ...our grammar on the Generalized LR Parser Compiler which is the syntactic part of the Universal Parser used in the CMU Machine Translation project. No attempt has been made to include morphological rules since it would be a duplication of the contributions of Hankamer [3] Solak [15] and Oflazer [10]. The parser compiler lets us incorporate our own morphological analyzer, and we use a full two level specification of Turkish morphology based on a lexicon of about 24,000 root words, for morphological analysis of words [1, 10] A Turkish sentence is given as input to the program, and the program ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 32 | 160 | ...proposals made in Chapters 3 and 4. Finally, Chapter 8 summarizes the results of the dissertation. 1.3 Background 1.3. 1 Turkish Morphology Turkish is an agglutinative language where word structures are formed by productive affixations of derivational and inflectional suffixes to root words (Oflazer (1994)) In this section, I present an overview of certain aspects of Turkish morphology (and certain consequences for the syntax) to help the reader follow the rest of the dissertation. In Turkish, noun stems can be marked for plurality, possessiveness, case, etc. The plural suffix is lEr . 1 ... |
| 33 | 161 | ...of previously occurring objects will be treated as known. Related Work There has been recent growth in empirically oriented work in discourse processing. Several researchers have addressed evaluation, investigating the degree to which human subjects agree with one another on discourse tasks (Passonneau Litman 1993, Hirschberg Grosz 1992, Hearst 1993) Others have used frequency information to evaluate algorithms. Passonneau and Litman (1993) tagged a corpus with classes and features and tested algorithms hypothesized from the literature. They consider just one feature at a time, so do not address ... |
| 33 | 162 | ...to be continuous if its discontinuity is close to 0. Sum. ratio in word number indicates the average length of summaries when it is calculated in the number of words. We tested the statistical significance of the agreement among subjects. Using the same methodology as in (Jing et al. 1998; Passonneau and Litman, 1993), we performed Cochran s Q test (Degroot et al. 1981) on the data from the subjects. For our task, Cochran s Q test evaluates the null hypothesis that the total number of human subjects judging the same document as relevant is randomly distributed. The results show that this hypothesis is false... |
| 33 | 163 | ...such as the how the current utterance relates to prior discourse. We have identified four types of discourse cues. The first type is perceptible silence, or pauses, observed at the end of an utterance, which has been found to correlate with discourse boundaries (Grosz and Hirschberg, 1992;Passonneau and Litman, 1993; Swerts, 1997) We believe that in the context of initiative modeling, silence at the end of an utterance may suggest that the speaker has nothing more to say in the current turn and intends to give up his task dialogue initiative. For instance, in the following dialogue segment, the silence at ... |
| 33 | 164 | ...and Beer Sheva University in Israel. Agreement Among Human Subjects We measured agreement among human subjects using percent agreement, a metric defined by (Gale, Church, Yarowsky 1992) for the sense disambiguation task, but also used in other applications such as discourse segmentation (Passonneau Litman 1993; Hearst 1994) Percent agreement is the ratio of observed agreements with the majority opinion to possible agreements with the majority opinion. For our experiments, agreement among 3 or more subjects is a majority opinion. The total possible agreements with the majority opinion is the number of ... |
| 33 | 165 | ...If we discard unfilled hesitations, the judgment of whether or not a repair occurs between two particular words or during a particular dialogue move appears to be sufficiently reliable to use. Reliability is frequently cited as some form of percent agreement (e.g. Passonneau and Litman [14], TOBI [17] Kowtko et al. 8] such as the percentage of judgments on which two coders agreed. in this context, this is misleading because repairs are coded during only approximately 5 of moves and after only approximately 3 of words. For instance, the distribution of agreement rates for two... |
| 33 | 166 | ...units 1 and 3 in text (4) are assigned the score 3. Agreement among judges Overall agreement among judges. I measured the agreement of the judges with one another, by means of the notion of percent agreement that was defined by Gale (1992) and used extensively in discourse segmentation studies (Passonneau Litman 1993; Hearst 1997) Percent agreement reflects the ratio of observed to possible agreements with the majority opinion. The percent agreements computed for each of the five texts and each level of importance are given in table 3. The agreements among judges for my experiment seem to follow the same ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 33 | 167 | ...very important, 63 for those considered less important, and 77 for those considered unimportant. The overall percent agreement in this case is 75 . Statistical significance. It has often been emphasized that agreement figures of the kinds computed above could be misleading (Krippendorff 1980; Passonneau Litman 1993). Since the true set of important textual units cannot be independently known, we cannot compute how valid the importance assignments of the judges were. Moreover, although the agreement figures that would occur by chance offer a strong indication that our data are reliable, they do not provide ... |
| 33 | 168 | ...space of anaphoric antecedents to those discourse entities actually referred to in the discourse, while the cache model allows unrestricted retrieval in the main or long term memory. Many studies on discourse segmentation highlight the role of cue words for signaling segment boundaries (cf. e.g. Passonneau Litman (1993)) or the use of overspecified referential expressions to indicate a thematic shift (Vonk et al. 1992; Walker, 1996b) However useful these strategies might be, we see the danger that such a surface level description may actually hide structural regularities at deeper levels of investigation ... |
| 33 | 169 | ... Each of the 20 articles in the corpus was segmented by at least four human judges, and the majority opinion of segment boundaries was computed as the evaluation standard (Klavans et al. 1998) Human judges achieved on average only 62.4 agreement with the majority opinion, as seen in Table 2. Passonneau and Litman (1993) show that this surprisingly low agreement is often the result of evaluators being divided between those who regard segments as more localized and those who prefer to split only on large boundaries. We then verified that the task was well defined by testing for a strong correlation between the ... |
| 33 | 170 | ...the concurring judges. The next three major boundaries occur after paragraphs 5, 9, 12, and 13. There is some contention in the later paragraphs; three readers marked both 16 and 18, two marked 18 alone, and two marked 17 alone. The outline in Section 1 gives an idea of what each segment is about. (Passonneau Litman 1993) discuss at length the considerations that must go into the evaluation of segmentation algorithms according to reader judgement information. As Figure 4 shows, agreement among judges is not perfect, but trends can be discerned. In our evaluation we follow the suggestions of (Passonneau Litman ... |
| 34 | 171 | ... For example, an active feature with a large weight might indicate that some parse had a high probability. Each weight i is associated with a feature f_i . Weights are real valued numbers and are automatically determined by an estimation process (for example using Improved Iterative Scaling (Lafferty et al. 1997)) One of the nice properties of RFMs is that the likelihood function of a RFM is strictly concave. This means that there are no local minima, and so we can be sure that scaling will result in estimation of a RFM that is globally optimal. The (unnormalised) total weight of a parse $x, x)$ is ... |
| 34 | 172 | ... from formalisms such as the theory of Markov random fields [4, 14, 28, 29, 33] they have formed some of the core analytical frameworks in areas including image processing [5, 15, 20, 23] and biometric analysis [4] and they have found applications in many other fields, including language modeling [18] and the categorization of hypertext documents [12] To motivate the formulation, we consider, as an illustrative example, the well studied problem of restoring an image that has been degraded by noise [5, 23] We are given a large grid of pixels; each pixel has a true intensity that we are ... |
| 34 | 173 | ... and Roukos [110, 187] have proposed a new approach for combining statistical evidence from different sources, that is based on the Maximum Entropy Principle (ME) This work was originated within the speech recognition field [187] but it has also been successfully applied to word morphology [168], PoS tagging [101, 177] PP attachment disambiguation [180] identification of clause boundaries [181] partial and general parsing [211, 178] text categorization [161] and machine translation [10] See [179] for a broad introduction to ME methods and a survey of existing applications. ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 34 | 174 | ... It is well known that this model has an exponential form $P(w_i jw_i \Gamma_1; w_i \Gamma_2) = \frac{1}{Z} e^{(w_i \Gamma_1 + w_i \Gamma_2)}$ where the parameters are estimated by a variant of the Generalized Iterative Scaling algorithm described in [4]. In addition to the N gram constraints, if the model is required to satisfy constraints of the form $X w_i \Gamma_2; w_i \Gamma_1; n_i \Gamma_2; n_i \Gamma_1 P(w_i \Gamma_1; w_i \Gamma_2; n_i \Gamma_2; n_i \Gamma_1; w_i jh_i \Gamma_2; h_i \Gamma_1) \# [h_i \Gamma_2; h_i \Gamma_1; w_i] \# [h_i \Gamma_2; \dots$ |
| 34 | 175 | ...to consider the above two issues in a uniform way. First, we introduce a model of generating a collocation of a verb and argument adjunct nouns (section 2) and then view the model as a probability model (section 3) As a model learning method, we adopt the maximum entropy model learning method (Della Pietra, Della Pietra, and Lafferty, 1997; Berger, Della Pietra, and Della Pietra, 1996) and apply it to the task of model learning of subcategorization preference. Case dependencies and noun class generalization are represented as features in the maximum entropy approach. In the maximum entropy approach, features are allowed to have ... |
| 35 | 176 | ...Having identified a tractable fragment of intuitionistic logic, several interesting questions arise. Nebel [27] uses his tractable class in order to obtain a tractable class for spatial reasoning, in the so called RCC 8 spatial algebra [28, 29] This class was later extended by Renz and Nebel [30] to a maximal tractable subclass of that spatial algebra. Thus, since this class is incomparable to Nebel's, is it possible to use it to obtain other tractable subclasses of RCC 8, by reducing their satisfiability problem to that of intuitionistic logic For the case of the RCC 5 spatial algebra, ... |
| 35 | 177 | ... class is incomparable to Nebel's, is it possible to use it to obtain other tractable subclasses of RCC 8, by reducing their satisfiability problem to that of intuitionistic logic For the case of the RCC 5 spatial algebra, all possible cases of tractable subclasses have already been characterised [30, 24]. Another relevant question is whether we can use our tractable class of set constraints for tractable inference in other logical systems. For instance, Renz and Nebel [30] use classes of modal logics in order to prove classes of the RCC 5 and RCC 8 spatial algebras tractable, so there is ... |
| 35 | 178 | ...a logical framework with a precise semantics within which a variety of more practical representation languages might be embedded. For instance, a decision procedure for a significant set of topological relations was presented by Bennett (1994, 1996) and tractable subsets of these identified by Renz and Nebel (1997, 1999). Cristani et al. 2000) investigate the complexity of reasoning with a combination of mereological and morphological relations and proves tractability of a significant constraint language, which is a fragment of our formalism. 16 Recursive axioms such as that for MoveWithin, which is clearly of ... |
| 35 | 179 | ...in GIS. RCC adopts a region topology in which regions are primary objects and the connection relation is the primary relation. Other relations between regions are defined upon the connection relation with a set of axioms and Boolean functions using first order logic. RCC research (Bennett 1994; Renz Nebel 1999) studies the composition rules of different spatial relations and uses these rules to uncover unknown relations from known ones. The 9 intersection model adopts a point set topology in which points are primary objects and regions are defined as sets of points. A topological relation between two ... |
| 35 | 180 | ... others) has shown that some simple combinations of modalities together with very simple interaction axioms yield undecidable systems; on the positive side, there are a number of examples of quite expressive fragments of multi modal languages, whose decision procedures are polynomial, for example (Renz and Nebel, 1997). Thus, the viability of reasoning with combined modal logics depends very much on the particular combination of modalities and interaction axioms. An obvious way of reducing the complexity of a logical language is to restrict its syntax. We consider such an approach, called layering (Finger and ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 36 | 181 | ...gives the lowest segmentation error rate possible given the word boundary mismatches due to recognition errors. Even with the punctuation marks, this task is not easy. We will not give the details of these systems, as with the punctuation marks, it is possible to achieve more than 99 accuracy. Reynar and Ratnaparkhi [1997] have presented a maximum entropy approach for this task. Palmer and Hearst [1997] have integrated neural networks with decision trees in order to detect sentence boundaries. Other researchers have used regular grammars, or some simple rules CHAPTER 5. SENTENCE SEGMENTATION 46 to detect boundaries ... |
| 36 | 182 | ...framework with respect to many linguistic problems and demonstrates how tools using it can achieve high accuracy and domain independence. At present, we have built a tokenizer and sentence detector which use maximum entropy models with features based on those presented in Ratnaparkhi (1998) and Reynar and Ratnaparkhi (1997). Other tasks such as paragraph detection or dealing with figures and tables have not been implemented as yet, 108 but the existing components have been designed to work with XML documents in a manner which permits additions such as these with minimal effort. With tokenized and sentence detected ... |
| 36 | 183 | ...Gates. Burns , an. and . Golden added . Another four unclassified proper names were capitalized words which followed the U.S. abbreviation e.g. U.S. Supreme Court . This is a difficult case even for sentence boundary disambiguation systems (Mikheev, 1998) Palmer Hearst, 1997) and (Reynar Ratnaparkhi, 1997)) which are built for exactly that purpose, i.e. to decide whether a capitalized word which follows an abbreviation is attached to it or whether there is a sentence boundary between them. The U.S. abbreviation is one of the most difficult ones because it can be as often seen at the end of a ... |
| 36 | 184 | ...developed for period disambiguation in the past. Palmer and Hearst [Palmer and Hearst, 1994] obtained 98.5 accuracy on Wall Street Journal data with a neural network. Michael Riley [Riley, 1989] reported 99.8 accuracy on the Brown corpus for his decision tree approach. Reynar and Ratnaparkhi [Reynar and Ratnaparkhi, 1997] achieved 98.8 accuracy on Wall Street Journal data and 97.9 on the Brown corpus with maximum entropy modelling. A similar approach with slightly better results was presented in [Mikheev, 1998] In contrast to these methods, Grefenstette and Tapanainen s semi automatically trained ... |
| 36 | 185 | ...detection program suming a zero word error rate. This result is in agreement with the results from the human annotation experiments described in Section 3. However, there is a far greater difference between the automatic system s performance on standard and ASR text than the human annotators. Reynar and Ratnaparkhi (1997) (Section 2) argued that a context of one word either side is sufficient for the punctuation disambiguation problem. However, the results of our system suggest that this may be insufficient for the sentence boundary detection problem even assuming reliable part of speech tags (cf note 5) These ... |
| 36 | 186 | ...2.4 Computational Work on Punctuation Computational linguists have worked on the recognition of sentence boundaries for part-of-speech tagging and sentence alignment in bilingual corpora. Palmer and Hearst [1994] use a neural network with part of speech probabilities to label sentence boundaries. Reynar and Ratnaparkhi [1997] use a maximum entropy model (for training) that requires little prior information to detect valid boundaries. Garside and his colleagues [1987] describe a research programme undertaken between 1976 1986. Their aim was to base NLP on the probabilistic analysis of a large corpus. In describing the |
| 37 | 187 | ...the project. Instead the project concentrated on the construction of content bearing links between text fragments. Text structure identification still has many open problems, such as passage retrieval, theme extraction, and text structuring. To date some local solutions have been already proposed (Salton and Allan, 1993; Callan, 1994; Salton et al. 1996) but no general techniques are available. Therefore, in order to maintain the structure and the features of the original paper version, the original book pages were preserved. The original pages were translated into HTML and presented to the user through a ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 37 | 188 | ...be used in various information retrieval tasks. Examples include automatic discovery of the topic structure of text [7] use of rhetorical boundaries (e.g. chapters, paragraphs, and sentences) to improve retrieval performance [8] and the automatic creation of hypertext links within documents [16]. However, their effective use outside the laboratory may still be some way off. Identifiable Duplication of Contents Selecting libraries would be easier if there were not duplication in libraries collections of documents. Documents held in one library are often held in other libraries also. ... |
| 37 | 189 | ...(e.g. [3]) Most systems that use queries as navigational aids use them to identify relevant neighborhoods in the hypertext, and then rely on manually created links to support further navigation. IR techniques have been used to segment long articles into shorter, more focused nodes (e.g. [34], [22]) Similarity among passages has been used to create links between specific nodes. This work, however, has focused on text segmentation techniques rather than on the hypertext interface. Although it is clear that such approaches are promising, little evidence has been published to date ... |
| 37 | 190 | ...documents on the basis of their content and to categorize the texts. DR LINK is an ambitious project to have a computer identify texts and analyse their structure. The successful combination of those approaches could lead to many excellent hypertext links being forged by computer alone. Allan [SA93, All95] found links using techniques from IR. Links were further refined by applying varying cutoffs to subdocument units (paragraphs for example) The method was not rigorously evaluated at that time. This A variation of Allan's techniques is employed (see Chapter 4) to generate links and a ... |
| 37 | 191 | ...product, where the vector terms are the term frequency inverse document frequency (TFIDF) of those words in the node whose term frequency is closest to the median term frequency. In this way, HieNet is able to create only links that are likely to be of interest to the reader. Allan and Salton [2, 44] describe another approach to automated document linking based on the vector space model. They also use TFIDF to cluster documents. They then divide each document into small pieces (typically paragraphs) and perform clustering analysis at this more local level; doing so helps to resolve ... |
| 38 | 192 | ...or more nouns as an effective identification of concepts found within a document. A related area is the automatic determination of text themes, or topics that are emphasized in the text and represented by selected text excerpts. More complex methodologies such as this appear in more recent work by (Salton et al. 1996; 1994) that uses adjacent words to describe simple themes, and non adjacent text possibly spanning multiple paragraphs to define more complex themes. We too use words and word phrases to identify topics or trends in text databases. Methodology The methodology we describe is a general approach ... |
| 38 | 193 | ...case, queries for related documents in the intranet may be a big help for the processes of finding and the integration of documents. For effective retrieval and maintenance in these steadily growing intranets, considerable approaches have been developed (Agosti, Crestani, Melucci 1994, Allan 1995, Salton, Singhal, Buckley, Mitra 1996, Shin, Nam 1997) but further supporting tools are still needed. We developed a new tool: Weaving Intranet Relations WIR which basically gives an innovative retrieval function. Additionally, it is able to support the organization and maintenance of the web content by suggesting new structures ... |
| 38 | 194 | ... with strongly pragmatic constraints, for instance: what kinds of documents are optimally suited for being abstracted in such a way (e.g. Preston Williams 1994; Brandow, Mitze, Rau 1995) how to derive more representative scoring functions, e.g. for complex documents, such as multi topic ones (Salton et al. 1996), or where training from professionally prepared abstracts is possible (Kupiec, Pedersen, Chen 1995) what heuristics might be developed for improving readability and coherence of narratives made up of discontinuous source document chunks (Paice 1990) or with optimal presentations of such ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 38 | 195 | ...least one direct link to that multimedia node. In their extraction operation Dunlop et al. [5] took into consideration all the information kept in the neighbouring nodes and not simply the contextual information in the node that could be related to the file being described. However, Salton et al. [17, 18] suggested a different approach to retrieve information from text nodes through decomposition of the text in text segments and text themes. It seems reasonable that a text segment, or theme, could be much more relevant to the construction of the non text node descriptor than the whole text ... |
| 38 | 196 | ...[6] For our task, paragraph and section information is not available and topic segments consist of a very small number of sentences, sometimes only one, so choosing an arbitrary block size as is often done for passage retrieval is not appropriate. Salton and Singhal [8] and Salton et al. [7], discuss the decomposition of text into segments and themes where a segment is contiguous block of text discussing a single subtopic and a theme is a chain of such segments possibly interleaved with other themes. The segmenting process begins at paragraph level. Then, paragraphs are compared by ... |
| 39 | 197 | ...cannot tell that she is not interested in blue telephones and red cars. Or, for that matter, the system cannot tell that the user is interested in large texts and small pictures, but not in small texts and large pictures. This problem gets even worse when more attributes and objects are involved. [1] suggest a similar approach: with probabilities it is difficult to specify a property as simply irrelevant, and it is difficult to deal with a lack of data: if we don't know yet about the user's preferences wrt. some feature then what? Furthermore the technique of a Bayesian classifier is hard to ... |
| 39 | 198 | ...by the terminal User Agent (browser) The Application Layer is the core of the system: it collects the user behaviour and characteristics and implements the adaptation process. It comprises two main modules: the Adaptive Hypermedia Application Server (AHAS) and the User Modelling Component (UMC) [4]; they run together with a Web Server. The UMC maintains the most recent actions of the user and executes the algorithm for the evaluation of the user's profile. After a user has selected the next page and the system has determined his/her user's position in the Adaptation Space, the AHAS ... |
| 39 | 199 | ...presented to the user depending on his or her expertise (Sales Assistant, Popp and Ldel, 1996) presenting expertise dependent explanations and technical details (Metadoc, Boyle and Encarnacion, 1994; KN AHS, Kobsa et al. 1994) and . generating expertise dependent product descriptions (SETA, Ardissono and Goy 1999, 2000b; Ardissono et al. 1999) Adaptation to a user's knowledge of domain concepts, of rules and of other items is also a typical feature of intelligent tutoring systems. In many such systems, user knowledge is taken into account when guiding the user through the learning material. Examples are the ... |
| 39 | 200 | ...based on concept relationships represented as domain knowledge. Figure 5 shows a graphical notation of a concept hierarchy (i.e. an ontology) from the animal kingdom, as it might be used for representing user knowledge about this domain (cf. Akoulchina and Ganascia, 1997; Milosavljevic, 1997; Ardissono and Goy, 1999, 2000b; Ardissono et al. 1999) 10 A fourth type, abductive reasoning (from the consequences to the premises) is also sometimes employed (e.g. in plan recognition, see Section 2.1.5) but will not be discussed here. 25 thing mammal fish shark whale orca dolphin Figure 5: A concept ... |
| 9 | 201 | ...not very satisfactory: they are often too many, disturbed, not very precise with a lot of noise. Preliminary results obtained with a test collection of the TREC Conference Web Track showed the poor results quality of 5 well known search engines, compared with those of 6 systems taking part in TREC [17]. Berghel considers the current search engines as primitive versions of the future means of information access on the Web [5] mainly because of their incapacity to distinguish the good from the bad in this huge amount of information. According to him, this evolution cannot be done by simple ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 9 | 202 | ...terms encountered may not be present in the training collection. We use IDF (Inverse Document Frequency, which is high for rare terms, and low for common terms) weights derived from a much larger sample (one million web pages, obtained from the collection of pages used in the TREC Web Track [9]) The last, fall back case is necessary in order to handle phrases not present in the training data. Intuitively, it assigns the weight of phrase # inversely proportional to the probability that all the terms in # appear together, scaled to weight occurrences of multi word phrases higher. This ... |
| 9 | 203 | ...for the web, which contains billions of documents. Researchers have addressed this problem by developing standard document collections with queries and associated relevance judgments, and by limiting the domain of documents that are judged [21] Recently, TREC has incorporated a Web Track [9] that employs a collection of web documents (small relative to the size of the web) This collection of documents is a valuable resource to evaluate information retrieval algorithms over web data. However, the collection is not well suited to evaluate a system like Tritus, where we aim to ... |
| 9 | 204 | ...then would like to examine other similar documents. Search engines are still in their infancy. Although the exact nature of many of the algorithms they use for finding appropriate pages is proprietary [24] there is some research that suggests that the algorithms they use are not very accurate [15]. Existing search engines and their users are typically at cross purposes. While these systems normally retrieve documents based on low level features, users usually have a more abstract notion of what will satisfy them when conducting a query for certain information. For instance, most search ... |
| 9 | 205 | ...the MultiText [7] structured text database system that has been used to support large digital library and WWW search applications. Using a specialized index and adaptive algorithms, simple queries can be executed in under a second over 100 GB of data. As a result of using a large capacity database [20], Jupiter avoids any database related scalability issues [6] While Jupiter is usable on its own as a program comprehension tool it can also function as a backend for other tools. For example, MARS [13] the predecessor to Jupiter, was used to extend the Software Bookshelf [18] with search ... |
| 40 | 206 | ...element which contributes to the overall quality of the human machine communication. Unfortunately, there is no fixed relationship between recognition performance on the one hand, and human machine communication quality on the other. Approaches to set up such a relation have been proposed by Walker et al. 1997) with the PARADISE framework, but they are not universal and have to be determined for each application anew. For the planner of transmission systems, it is important that good speech quality as well as good recognition performance are provided by the system, because speech transmission channels ... |
| 40 | 207 | ...rate and error recovery rate are good indicators of usability. A system response will be classified as an error, when it is incoherent with the user s actual utterance, the task, the information in the database or with the dialogue structure. Task success is difficult to measure quantitatively. Walker et al. [1997] work with confusion matrices that represent the deviations from some fixed attribute value matrix that gives the ideal values for a task. By calculating agreement over the confusion matrices for a large corpus of dialogues, they derive a general measure of task success. This method is not ... |
| 40 | 208 | ...instance, corrections appear to be more prosodically marked than other utterances (higher, longer, louder, slower, which is in agreement with our current results. 6.2. On online evaluation In many evaluation schemes the frequency of errors is one of the ingredients (e.g. Nielsen 1993; Walker et al. 1997). Arguably, the most useful kind of evaluation is online evaluation, since this gives the option of automatically adapting to the current situation. The analyses of this paper suggest that the presence of cues such as a prolonged delay before answering or a high pitched narrow focus accent are ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 40 | 209 | ...simple form filling strategy. Rather, the dialogue goal specification is an encapsulation of a method invocation which, when triggered, causes the back end application to do what the user intended the system to do. The assumptions made here are similar to those in the general Paradise framework [Walker et al., 1997] for dialogue evaluation where the task model for dialogue managers is equally described in attribute value matrices. Example (continued) We continue the fast food service example. We concentrate on the dialogue goals relevant to the pizza and pasta objects, as we assume that we have recourse ... |
| 40 | 210 | ...we have begun to design a Task Based Evaluation for JANUS (Thomas, 1999) which measures goal completion. This paper describes what we have learned by comparing the two types of evaluation. 2. Design Criteria Most previous work on TBE has been conducted on human machine dialogue (for example (Walker et al. 1997)) For machine translation, we need a TBE that is suitable for two humans each expressing communicative goals, but mediated by a machine. Our coding scheme for communicative goals is described below. In particular, we have to separate human clumsiness and error from machine error, because we are |
| 40 | 211 | ...and inexperienced users accepted some bad translations as long as they can be understood in context. For example, in the context of the question How much does it cost , users will accept the answer 128 hours. The percent of task success, however, does not provide a measure of user frustration (Walker et al. 1997). This is why we formulated the TBE scoring function to take into account success failure of goals as well as the number of attempts at each goal. In future work, we will give some thought to making the TBE score (on a minus one to one scale) more comparable to the ABE score (expressed as a ... |
| 41 | 212 | ...of 5.51 would be with a definite NP, e.g. Ja, das Hotel hab ich. However, null topic utterances containing such ambiguous main verbs are not taken care of by the extended Compatibility Rule. In the Centering ranking the Discourse Unit is ranked lower than the overt centres (Dimitriadis 1996; Walker et al. 1994) (cf Chapter 2, Section 2.2.3) implying that discourse deictic reference is marked and less CHAPTER 5. DISCOURSE DEIXIS AND NP FORM 191 likely to be established in topic position than reference to concrete entities. What this ranking also implies is the following: if there is only one pronoun ... |
| 41 | 213 | ...evaluation, the attempt to simulate the human behaviour is quite successful. In addition to these kinds of various approaches to the concept of reference, there is a large body of work on Centering Theory [Grosz et al. 1995] which will be described in Section 2.2. In [Walker et al. 1990] and [Walker et al. 1994], a computational treatment of the interpretation of zero and overt anaphora are provided with three aims: 1) to generalize a computational account of the discourse process called Centering, 2) to apply this account to discourse processing in Japanese so that it can be used in computational ... |
| 41 | 214 | ...stochastic grammars [Black et al. 1993] However, these predict only within utterances, while our interest extends to predictions on the scale of several utterances. We might also consider discourse oriented mechanisms such as centering and global focusing models [Grosz and Sidner, 1986] [Walker et al. 1992]; but in fact these are not designed to predict the lexical items that will be seen a bit later in the dialogue. Instead, we propose to permit the flexible definition of windows in a transcribed corpus within which concurrences of morphological or lexical elements can be exa ... |
| 41 | 215 | ...discourse, the more likely it is to be pronominalized by a speaker. The centering model is broadly language independent. The main language dependent parameter is the criteria for ranking the salience of discourse referents. For Japanese, we adopted the saliency ranking proposed by Walker et al. (Walker, Iida, and Cote 1994), given in Table 1. This ranking is based mainly on syntactic function, which in Japanese is indicated explicitly by post positional particles such as wa and ga rather than by word order position as in English. This ranking also takes into account empathy, or the perspective from which a speaker |

| ABS # | Citation # | Citation Context Text |
|-------|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 41 | 216 | ...describes the relationships between consecutive utterances. Though originally formulated as a predictive theory of discourse coherence, recent work has demonstrated the theory's usefulness for resolving pronoun ambiguities and describing constraints on the interpretation of zero topics in Japanese (Walker, Iida and Cote 1994). Centering theory is based on a set of features associated with each utterance in a discourse segment, including the backward looking center, a set of forward looking centers, and a preferred center. The backward looking center (C _b) is a discourse entity that links the current utterance with ... |
| 42 | 217 | ...different task in a new domain (Armstrong Warwick, 1993) While statistical approaches to parsing bear the mentioned advantages, most existing methods (especially in syntactic parsing) use a hand crafted set of contextual features on which probabilistic parsing models are built. The Chill system (Zelle, 1995) represents an approach to learning relevant contextual information (represented as relational knowledge) for the task of disambiguation given complete contexts (i.e. the entire parse state) instead of relying on handcrafting features for parsing. However, the original system builds a ... |
| 42 | 218 | ... higher order representations (e.g. of a meaning representation language) This view clearly states that a simple grammatical ungrammatical distinction or the construction of a syntax tree is of little help for most NLP applications (and is in harmony with recent work on robust text analysis [Zel95] As indicated in the title, the aim of this paper is to give an account of existing methods and techniques. This happens according to a classification that deviates from those chosen by other researchers. Instead of distinguishing, e.g. between syntax and semantics based approaches [Ste92] ... |
| 42 | 219 | ... complete sentences for answering database queries (Zelle Mooney, 1996; Miller, Stallard, Bobrow, Schwartz, 1996; Kuhn De Mori, 1995) 3 CHILL: ILP for Semantic Interpretation Our own research on learning for semantic interpretation has involved the development of a system called Chill (Zelle, 1995) which uses ILP to learn a deterministic shift reduce parser written in Prolog. The input to Chill is a corpus of sentences paired with semantic representations. The parser learned from this data is able to transform these training sentences into their correct representations, as well as ... |
| 42 | 220 | ...of their semantic meanings. The particular representation is determined by the domain at hand and the representation of entire sentences. The initial motivation for learning lexicons was so they could be used to bootstrap a parser acquisition system, Chill (Zelle Mooney, 1993, 1994; Zelle, 1995); they could then be used by the parser to map novel sentences into representations of their meanings. A major assumption of our research has been the assumption of compositionality. This assumption states that the meaning representation of a sentence is composed from the meaning representations ... |
| 42 | 221 | ...description and example in Section 5, some representational issues are discussed in Section 4. Next, some preliminary experimental results are given and discussed. The final three sections discuss future work, related work, and conclusions. 2 Background: Chill Chill (Zelle Mooney, 1993; Zelle, 1995) is a computer system that learns to parse natural language sentences by training over corpora of parsed text. The parsing formalism used is a shift reduce parser. For example, Chill can learn to parse sentences into case role representations when given a sample of sentence case role pairings ... |
| 43 | 222 | ...areas. Table 1 compiles some of them, found in the literature 3 , and uses them to characterize the roles. Other methodologies, such as [Wooldridge et al. 2000] propose a formalization of the roles but it was not deemed necessary for our first prototype. Table 1. Roles Characteristics 3 [Etzioni and Weld, 1995] , Franklin and Graesser, 1996] Nwana, 1996] and [Wooldridge and Jennings, 1995] Social interactions Following the sub societies and roles identification comes the specification of the interactions. Interactions consist of more than the sending of isolated messages and the conversation ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 43 | 223 | ... as a problem of negotiation between the various production cells within the factory [57] Handbook of Software Engineering and Knowledge Engineering 5 Agents in the Internet: Much of the hyperbole that currently surrounds all things agent like is related to the phenomenal growth of the Internet [12, 5]. In particular, there is a lot of interest in mobile agents, that can move themselves around the Internet operating on a user s behalf. This kind of functionality is achieved in the telescript language developed by General Magic, Inc. for remote programming [52] related functionality is ... |
| 43 | 224 | ...Brent Blackstock Attorney at Law, MoreLaw blackstock morelaw.com Abstract Intelligent agents are being deployed in diverse application domains. Both desktop based and Internet based personal assistant agents have been developed to assist users with their information processing chores [1, 2, 6, 5, 3, 7, 10]. In this paper we present an Internet based agent designed to assist legal researchers in retrieving laws and case reports electronically warehoused at a diverse set of databases maintained by local, state, and federal governments. LawBOT is implemented as a collection of agents which are ... |
| 43 | 225 | ...detail. 3.3.1 Intelligent Agents The ticket server architecture employs user and server processes that exhibit intelligent behavior, namely autonomy, communication ability, negotiation, and reasoning. These behaviors are widely agreed upon as being 85 characteristics of intelligent agents [81] [82] [83] User agents act in a semiautonomous fashion (i.e. negotiating within the bounds set by a user) to negotiate for tickets. Server agents act semi autonomously to decide how user claims correlate with the server s view of context and how ticket values should be assigned. Disaster response ... |
| 43 | 226 | ...a human relationship by doing something that another person could do for you. More loosely, Riecken [25] refers to integrated reasoning processes as agents. Others take agents to be computer programs that behave in a manner analogous to human agents, such as travel agents or insurance agents [26] or software entities capable of autonomous goal oriented behaviour in a heterogeneous computing environment [27] while some avoid the issue completely and leave the interpretation of their agents to the reader. Many such other agent definitions can be found in the excellent review by Franklin ... |
| 43 | 227 | ...of the key qualities that can be used to assess agentness . Wooldridge and Jennings also describe a strong notion of agency, prevalent in AI which, in addition to the weak notion, also uses mental components such as belief, desire, intention, knowledge and so on. Similarly, Etzioni and Weld [26] summarise desirable agent characteristics as including autonomy, temporal continuity by which agents are not simply one shot computations, believable personality in order to facilitate effective interaction, communication ability with other agents or people, adaptability to user preferences ... |
| 43 | 228 | ...or observations or instructions from the outside, the more intelligent it is considered to be. Literature provides several examples of applications of the agent programming paradigm: it has been used in the field of intelligent navigation within the Web (Internet) at the University of Washington [4, 5], in order to produce access mechanisms to the information using natural language (MIT) and in order to optimize management systems of satellite images [7] This paper is based on research activity currently in progress at the Universities of Catania, Messina and Turin. Within a distributed ... |
| 43 | 229 | ...dedicated to the information retrieval task. 2 Information Retrieval and Intelligent Agents The techniques currently used to facilitate retrieval operations, in particular finding any information in a reasonable period of time, are based on the use of index files (6) It has been observed in [5] that indexing agents can deliver quick responses, but they have a number of technological limitations. The ideal situation is one in which a generic user gives a high level description of the information she requires and the system, on the basis of heuristic techniques and experience acquired, ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 43 | 230 | ...area in intelligent Web robot research is adaptive Web services. Examples of services include: Ahoy The Homepage Finder, which performs dynamic reference sifting [Shakes et al. 1997] Adaptive Web Sites, which automatically improve their organization and presentation based on user access data [Etzioni, Weld 1995], Perkowitz, Etzioni 1999] and Adaptive Web Page Recommendation Service [Balabanovi c 1997] Balabanovi c, Shoham] Balabanovi c et al. 1995] Discussion and ratings of some of these and other robots are available at several Web sites, e.g. Felt, Scales] Mitchell] Some scientists have ... |
| 43 | 231 | ...familiar with the user and situation exchange knowledge with others who handle the details of how to obtain needed information and services. Consistent with the requirements of a particular problem, each agent might possess to a greater or lesser degree attributes like the ones enumerated in Etzioni and Weld (1995) and Franklin and Graesser (1996) Reactivity: the ability to selectively sense and act . Autonomy: goal directedness, proactive and self starting behavior . Collaborative behavior: can work in concert with other agents to achieve a common goal . Knowledge level (Newell 1982) communication ... |
| 7 | 232 | ...of PriceBots and ShopBots, intelligent agents which can be programmed to implement a particular strategy for a retailer or consumer. Kephart and his colleagues have also studied the role of reinforcement learning, information filtering, and information bundling in an e commerce environment [16]. Intelligent agents technology has also been used for supply chain management and planning and scheduling problems. Sesh Murthy, Richard Goodwin, Pinar Keskon oak, and their colleagues have examined the difficult problem of scheduling multiple machines when there are multiple objectives and ... |
| 7 | 233 | ... on fishmarket auctions [Rodrguez Aguilar et al. 2000] Automatic bidding agents have also been created in this domain [Gimenez Funes, 1998] Outside of, but related to, the auction scenario, automatic shopping and pricing agents for internet commerce have been studied within a simplified model [Greenwald and Kephart, 1999] . FAucS addresses a much more complex scenario than has been previously studied with autonomous bidding agents: the FCC spectrum auctions. Spectrum auctions have been analyzed retrospectively [Weber, 1996; Cramton, 1997] but little is known about them from a theoretical perspective. As ... |
| 7 | 234 | ... following definitions of these terms: A shopbot is a software agent attached to a single user and has the ability to query multiple servers on the network (on behalf of the user) to gather information about prices and other service characteristics, like service quality or expected waiting time [6]. We assume that shopbot and user interests are identical and the shopbot s sole purpose is to serve the user s needs. A pricebot is a software agent attached to a single service provider and has the ability to dynamically change the price of the product service to maximize the provider s ... |
| 7 | 235 | ...that an agent solves a hard optimization problem, or interacts with a busy and expensive human expert. In fact, electronic markets may make the valuation problem more difficult, because of mitigating factors such as decreased aggregation, increased product differentiation, and increased dynamics [1, 4, 5]. In this paper we compare auction performance for agents that have hard local problems, and uncertain values for goods. Just as careful market design can reduce the complexity of the bidding problem, for example by providing incentives for agents to reveal their true value for a good [28] ... |
| 7 | 236 | ... of the supply chain [16, 33, 37, 45] Moreover, in the business to customer area, agent based technologies are exploited in the development of complex systems where agents search for products and services on behalf of a user, compare the solutions offered by different providers, and so forth [19, 22]. In addition to the above issues, agent based technologies can be successfully applied to the enhancement of other features of electronic commerce systems, among which the adaptability of the interfaces to the users needs: the popularity of Web shopping is increasing and very different types of ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 7 | 237 | ...at: www.research.ibm.com/infoecon/researchpapers.html . extent that the categories overlap, there can be direct price competition, and to the extent that they differ, there are asymmetries that again lead to the potential for cyclic price wars. The third model is the Shopbot model described in (Greenwald and Kephart, 1999), which models the situation on the Internet in which some consumers use a shopbot to compare prices of all sellers offering a given product, and select the lowest priced seller. In this model, the sellers products are identical, and their profit functions are symmetric. Myoptimal pricing leads ... |
| 44 | 238 | ...representation. The user can view the state information of the ongoing process, like viewing the shopping cart, inspecting the orders or tracking the delivery process [16] The business models used within these systems are quite static. A proposal for a more flexible shopping model is proposed by [6]. The model assumes an external and independent shopping controller, who provides a specific business workflow to the trading partners. So the merchant can use the external shopping controller if he wants to provide his goods by the shopping controllers business workflow to gain a higher ... |
| 44 | 239 | ...virtue of the IWIM model and MANIFOLD s implementation. 4.3. Integrating different components In this section we show the applicability of control based event driven coordination models such as MANIFOLD for the development of generic interaction frameworks, often referred to as shopping models ([10]) where the interaction and communication part (in other words the program logic) is separated from low level details such as the security or payment mechanisms employed, etc. The top level environment follows the logic of ([10]) and consists of four main components: a merchant handler, a ... |
| 44 | 240 | ...generic interaction frameworks, often referred to as shopping models ([10]) where the interaction and communication part (in other words the program logic) is separated from low level details such as the security or payment mechanisms employed, etc. The top level environment follows the logic of ([10]) and consists of four main components: a merchant handler, a customer handler, a shopping controller and a services controller. The first two are used to intercept and handle the requests, messages and data interchanged between a customer and a merchant, the third one coordinates the interaction ... |
| 44 | 241 | ...flow with its session manager and monitor. Aurora aims to provide coordination support for general purpose applications that fit the service flow paradigm, rather than more customized coordination for automating specific types of transactions. An example of such a specific solution is given in [45] which presents event driven models for different modes of consumer to merchant interaction, and an API to facilitate commerce transactions. Shopping models encapsulate the rules for specific types of commerce transactions and instruct the participants what to do next in the way of ordering, ... |
| 44 | 242 | ...of digital library services and a metadata architecture [6, 21] to maintain metadata. The InfoBus set of interoperability protocols have been augmented with support for customized coordination for automating specific types of transactions. An example of such a specific solution is given in [24] which presents event driven models for different modes of consumer to merchant interaction, and an API to facilitate commerce transactions. Shopping models encapsulate the rules for specific types of commerce transactions and instruct the participants what to do next in the way of ordering, ... |
| 44 | 243 | ...transactions like the purchase of drugs, ensuring auditability by proper authorities. There is, of course, no universal policy. Many different policies are being employed in traditional commerce, and many will have to be employed in electronic commerce as well. For instance, a recent paper [6] identified several examples of policies for purchasing of information, using such things as subscription, pay per view and pre paid voucher, which use different form of payment, employ different interaction protocols, and are suitable for different circumstances. The potential diversity of ... |

| ABS # | Citation # | Citation Context Text |
|-------|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 44 | 244 | ...set of interfaces is obscure, being embedded into the code of the interface. A manually implemented policy is unsafe because it can be circumvented by any participant in a given commercial transaction, by modifying his interface for the policy. The first difficulty has been addressed recently in [6, 1], and the second has been addressed in [5] but we are not aware of any attempt to alleviate both of these difficulties together. In this paper we introduce a mechanism, based on what we call regulated interaction [9] that can support a wide range of commercial policies in a unified and secure ... |
| 44 | 245 | ...Using this facility, it is possible for agents to negotiate about provision of complex services as well as resource allocation. In our approach, the HERMES specification language allows explicit description of capabilities and requirements, and, moreover, of execution plans for compound tasks. [18] presents event driven models for different models of consumer to merchant interaction, and an API to facilitate the construction of electronic commerce applications. A shopping model instructs the participant in a commerce transaction what to do next in the way of ordering, payment, and delivery. ... |
| 45 | 246 | ... Machine learning techniques have been applied to web search (McCallum et al. 1999) Boyan, Freitag, Joachims 1996) Specialized agents that mine the web have been described (Doorenbos, Etzioni, Weld 1997) Light is shed on web search from a different perspective by work on human behavior (Macskassy et al. 1998). Related problems include those of intelligently recommending scientific papers (Basu et al. 1999) and creating digital libraries for efficient indexing and retrieval of scientific documents (Lawrence, Bollacker, Giles 1999) Reviews of work in web searching include (Lawrence Giles 1999) ... |
| 45 | 247 | ...two different professional indexers; as little as 20 of the terms to be indexed may be handled in the same manner by different individuals (as noted in [Korfhage 1997] p. 107) and there is noticeable inconsistency, even by a given individual [Borko 1979] Cooper 1969] Jacoby, Slamecka 1962] [Macskassy et al. 1998], Preschel 1972] Salton 1969] Though not perfect, compared to most automatic indexers, human indexing is currently the most accurate since experts on popular subjects organize and compile the directories and indexes in a way which (they believe) facilitates the search process. Notable ... |
| 45 | 248 | ...needs to be optimal in some way. This, however, runs contrary to the fact that different people have quite different needs with regard to clustering of texts, because they may view the same documents from completely different perspectives (e.g. a business view vs. a technical view; also cf. [9]) Thus, what is needed are subjective criteria that allow for a diversity of views from which to look at the clustering task. Second, text clustering typically is a clustering task working in a high dimensional space where each word is seen as a potential attribute for a text. Empirical and ... |
| 45 | 249 | ... Machine learning techniques have been applied to web search (McCallum et al. 1999) Boyan, Freitag, Joachims 1996) Specialized agents that mine the web have been described (Doorenbos, Etzioni, Weld 1997) Light is shed on web search from a different perspective by work on human behavior (Macskassy et al. 1998). Related problems include those of intelligently recommending scientific papers (Basu et al. 1999) and creating digital libraries for efficient indexing and retrieval of scientific documents (Lawrence, Bollacker, Giles 1999) Reviews of work in web searching include (Lawrence Giles 1999) ... |

Appendix D
Citation Contexts Per Document

| Abstract | Citations | Abstract | Citations |
|-----------------|------------------|-----------------|------------------|
| 1 | 5 | 24 | 5 |
| 2 | 5 | 25 | 5 |
| 3 | 5 | 26 | 3 |
| 4 | 5 | 27 | 4 |
| 5 | 5 | 28 | 6 |
| 6 | 5 | 29 | 5 |
| 7 | 11 | 30 | 5 |
| 8 | 1 | 31 | 8 |
| 9 | 10 | 32 | 5 |
| 10 | 5 | 33 | 10 |
| 11 | 3 | 34 | 5 |
| 12 | 6 | 35 | 5 |
| 13 | 5 | 36 | 6 |
| 14 | 5 | 37 | 5 |
| 15 | 5 | 38 | 5 |
| 16 | 10 | 39 | 4 |
| 17 | 5 | 40 | 6 |
| 18 | 5 | 41 | 5 |
| 19 | 4 | 42 | 5 |
| 20 | 5 | 43 | 10 |
| 21 | 5 | 44 | 8 |
| 22 | 5 | 45 | 4 |
| 23 | 5 | | |

Appendix E
Citation Categorization Data for Raters 1 and 2

| Citation No. | Rater 1 | | Rater 2 | |
|---------------------|----------------------------------------|---------------|----------------------------------------|---------------|
| | Citation | Rating | Citation | Rating |
| 1 | span from | 3 | span from | 3 |
| 2 | work by | 3 | work by | 4 |
| 3 | However ... have not | 1 | research efforts have begun to | 3 |
| 4 | also developed | 3 | also developed | 4 |
| 5 | Provides | 3 | provides | 3 |
| 6 | is related also to | 4 | is related also to | 3 |
| 7 | is directly related | 4 | is directly related | 2 |
| 8 | in a similar vein | 4 | in a similar vein | 4 |
| 9 | are the basis for | 4 | are the basis for | 3 |
| 10 | is possible and is considered by | 4 | is possible and is considered by | 4 |
| 11 | in order to | 5 | among others it can be | 4 |
| 12 | are examples of | 3 | are example of | 4 |
| 13 | all produce | 3 | all produce | 3 |
| 14 | according to the approach described in | 5 | according to the approach described in | 5 |
| 15 | techniques are being used in | 3 | techniques are being used in | 3 |
| 16 | for example | 3 | derived from, the same assumption | 2 |
| 17 | in order to | 2 | | 2 |
| 18 | provides another solution | 2 | provides another solution | 3 |

| | Rater 1 | | Rater 2 | |
|--------------|-----------------------------------------|--------|-----------------------------------------------------------------------------|--------|
| Citation No. | Citation | Rating | Citation | Rating |
| 19 | for example | 2 | for example | 4 |
| 20 | Distinction | 1 | distinction | 4 |
| 21 | recent surveys | 3 | | 0 |
| 22 | See | 2 | | 0 |
| 23 | Overview | 3 | | 0 |
| 24 | Make | 2 | | 0 |
| 25 | related work ... surveys | 3 | | 0 |
| 26 | could benefit from | 1 | | 0 |
| 27 | explicitly address in recent work by | 3 | | 0 |
| 28 | this is surprising, since | 2 | | 0 |
| 29 | Presented | 5 | | 0 |
| 30 | Describe | 5 | | 0 |
| 31 | cue phrase based upon | 5 | based upon these ... his main predictions | 4 |
| 32 | as suggested in | 3 | as suggested in | 3 |
| 33 | provide ... showing | 3 | provide ... showing | 4 |
| 34 | related to | 3 | outside of but related to ... has been studied within a | 3 |
| 35 | Approach | 3 | finally a fourth general approach is | 3 |
| 36 | However | 1 | similar to Ontoberoker are ... however ... they all lack ... from our point | 1 |
| 37 | results are quite counter intuitive | 5 | this question has been studied in a limited number of studies especially | 3 |
| 38 | motivated by this discrepancy | 4 | | 3 |
| 39 | understandable in view of the fact that | 2 | while this may seem low... it is understandable in view of the fact that | 5 |
| 40 | found ... they tested | 2 | found a precision range of | 5 |

| | Rater 1 | | Rater 2 | |
|--------------|-----------------------------|--------|-------------------------------------------|--------|
| Citation No. | Citation | Rating | Citation | Rating |
| 41 | Discuss | 3 | in particular they examine | 4 |
| 42 | work finds that | 3 | | 2 |
| 43 | like in | 4 | | 3 |
| 44 | past work ... has been done | 3 | past work | 3 |
| 45 | new methods | 3 | | 3 |
| 46 | Uses | 3 | | 3 |
| 47 | decades of research | 3 | | 0 |
| 48 | is nearly ubiquitous | 4 | | 0 |
| 49 | Describe | 3 | | 0 |
| 50 | Present | 3 | present | 4 |
| 51 | However | 1 | has been used before in applications like | 3 |
| 52 | but has not | 1 | but has not | 4 |
| 53 | Introduce | 3 | introduce | 4 |
| 54 | Approach | 3 | approach | 5 |
| 55 | proposed by | 3 | proposed by | 3 |
| 56 | But | 1 | most work on ... up to now | 3 |
| 57 | many different approaches | 3 | many different approaches | 3 |
| 58 | a number of | 3 | a number of ... have been developed | 3 |
| 59 | Provides | 3 | | 3 |
| 60 | all these approaches | 3 | and the work of | 3 |
| 61 | well suited | 2 | well suited | 2 |
| 62 | on the other hand | 3 | | 3 |
| 63 | it is claimed | 2 | it is claimed | 4 |
| 64 | reinforces the point | 5 | reinforces the point | 2 |

| | Rater 1 | | Rater 2 | |
|--------------|-----------------------------|--------|-----------------------------------------------------------|--------|
| Citation No. | Citation | Rating | Citation | Rating |
| 65 | however, are not comparable | 1 | in conclusion we present a new approach ... that combines | 2 |
| 66 | most influential works | 3 | is described in ... the ... Dagan and Itai ... uses | 4 |
| 67 | Expand | 5 | expand | 2 |
| 68 | However | 1 | however | 1 |
| 69 | other ... Method | 3 | other ... Method | 4 |
| 70 | proposed by | 3 | the basic idea of Dagan and Itai | 4 |
| 71 | claims to be | 3 | for example ... in both cases | 4 |
| 72 | attempt to | 3 | attempt to | 4 |
| 73 | goals are similar but | 5 | goals are similar but | 5 |
| 74 | rather than | 5 | Early ... such as the ... demonstrated the | 4 |
| 75 | greater emphasis is placed | 5 | went a step further by | 4 |
| 76 | one solution | 3 | one solution | 4 |
| 77 | Uses | 3 | uses | 4 |
| 78 | But | 1 | but | 4 |
| 79 | but ... noticeably absent | 1 | as well as work done | 3 |
| 80 | for example | 2 | can make ... dramatically more | 4 |
| 81 | recently...has been applied | 3 | recently ... has been applied | 3 |
| 82 | recently applied | 3 | has been recently applied | 3 |
| 83 | are examples of | 3 | are examples of | 3 |
| 84 | however ... poor results | 1 | however ... poor results | 1 |
| 85 | However | 1 | the potential to greatly improve ... however | 1 |
| 86 | Similarly | 4 | similarly | 3 |
| 87 | they show | 3 | they show | 4 |
| 88 | has been deployed | 3 | has been deployed | 3 |

| | Rater 1 | | Rater 2 | |
|--------------|---------------------------------------|--------|---------------------------------------------------------------|--------|
| Citation No. | Citation | Rating | Citation | Rating |
| 89 | one finds | 3 | among the tools normally employed | 3 |
| 90 | noted in | 3 | noted in | 2 |
| 91 | recent literature | 3 | recent literature | 3 |
| 92 | have been identified | 3 | some researchers have given | 3 |
| 93 | field of research | 3 | presents a range of problems has become a | 3 |
| 94 | as pointed out by | 2 | is also likely to be effective ... as pointed out by | 4 |
| 95 | for example | 2 | | 3 |
| 96 | is considered to be good | 5 | our evaluation method is considered a simplification | 2 |
| 97 | Because ... or because | 2 | | 3 |
| 98 | seems particularly suitable | 2 | seems particularly suitable | 5 |
| 99 | however, our assessment | 1 | these rates ... are similar to those found in | 2 |
| 100 | Explains | 3 | develops this ... the further and explains the | 4 |
| 101 | neither; our method | 1 | | 3 |
| 102 | well known situation | 3 | well known situation | 4 |
| 103 | has also discussed | 3 | has also discussed | 3 |
| 104 | provided | 3 | provided | 4 |
| 105 | additional attempts | 3 | Huffman's system | 4 |
| 106 | have been developed | 3 | more recently | 3 |
| 107 | differentiated our approach | 5 | previous researches on | 3 |
| 108 | however | 1 | in recent years a variety of ... techniques have been used to | 3 |
| 109 | however | 1 | research efforts have begun to address | 3 |
| 110 | Suggests | 2 | suggests | 4 |
| 111 | recently adopted as new way; compares | 5 | | 3 |
| 112 | we admit | 2 | we admit | 2 |

| | Rater 1 | | Rater 2 | |
|---------------------|---------------------------------------------------------|---------------|---------------------------------------------------------|---------------|
| Citation No. | Citation | Rating | Citation | Rating |
| 113 | performed; examined | 3 | performed; examined | 4 |
| 114 | circumvent this problem | 1 | | 3 |
| 115 | thus ... well accepted | 2 | thus ... well accepted | 3 |
| 116 | however | 1 | the data reported in | 3 |
| 117 | further take the view | 5 | however unlike ... we | 1 |
| 118 | these questions are particularly important ... in which | 2 | these questions are particularly important ... in which | 4 |
| 119 | we take the view that | 5 | unlike...we take the view that | 1 |
| 120 | | 0 | most | 3 |
| 121 | | 0 | | 3 |
| 122 | | 0 | has been used to ... as discussed in | 3 |
| 123 | | 0 | wrote that ... they believe that | 4 |
| 124 | | 0 | common belief holds that | 3 |
| 125 | Trend | 3 | trend | 3 |
| 126 | as was the case | 5 | as was the case | 5 |
| 127 | | 2 | | 3 |
| 128 | many ... have been proposed | 3 | for example | 3 |
| 129 | we propose | 5 | some hybrid approaches | 3 |
| 130 | many schemes ... exist | 3 | many schemes ... exist | 3 |
| 131 | Trend | 3 | trend | 3 |
| 132 | other approaches include | 3 | other approaches include | 3 |
| 133 | See | 2 | as opposed to | 1 |
| 134 | other approaches | 3 | other approaches | 3 |
| 135 | for instance | 3 | more ... techniques | 3 |
| 136 | Uses | 3 | | 4 |

| | Rater 1 | | Rater 2 | |
|---------------------|----------------------------|---------------|-----------------------------------|---------------|
| Citation No. | Citation | Rating | Citation | Rating |
| 137 | some researchers use | 3 | e.g. | 4 |
| 138 | is closely related to | 4 | | 3 |
| 139 | for example | 3 | for example | 3 |
| 140 | share ... but | 4 | but ... but they generally do not | 1 |
| 141 | recent work | 3 | ... also do not provide | 4 |
| 142 | for example | 3 | recent work | 4 |
| 143 | it is our belief | 5 | for example | 4 |
| 144 | | 2 | | 3 |
| 145 | we test | 5 | | 4 |
| 146 | Present | 3 | we test | 5 |
| 147 | in contrast to | 1 | when ... is used | 5 |
| 148 | | 0 | in contrast to | 1 |
| 149 | | 0 | these records allow ... and to | 4 |
| 150 | | 0 | for instance | 4 |
| 151 | | 0 | | 3 |
| 152 | | 0 | others | 3 |
| 153 | | 0 | more recent work takes | 3 |
| 154 | | 0 | | 3 |
| 155 | | 0 | it is better to | 4 |
| 156 | neither; number of problem | 1 | | 4 |
| 157 | are directly taken from | 5 | earlier studies on | 3 |
| 158 | refer to | 3 | analyses ... are directly taken | 5 |
| 159 | contributions of | 3 | from | 5 |
| 160 | Is | 2 | refer to | 4 |
| | | | contributions of | 3 |

| | Rater 1 | | Rater 2 | |
|---------------------|----------------------------------|---------------|-------------------------------------------------|---------------|
| Citation No. | Citation | Rating | Citation | Rating |
| 161 | have addressed ... investigating | 3 | several researchers have | 3 |
| 162 | using the same methodology | 4 | using the same methodology | 5 |
| 163 | MISSING | 2 | the first type is | 4 |
| 164 | also used | 4 | also used | 5 |
| 165 | misleading | 1 | misleading | 1 |
| 166 | used extensively in | 2 | measured by means of this | 5 |
| 167 | emphasized | 2 | emphasized | 5 |
| 168 | many studies | 2 | many studies | 3 |
| 169 | show | 3 | show | 4 |
| 170 | discuss | 3 | that must go into the evaluation of | 5 |
| 171 | for example | 3 | for example | 5 |
| 172 | including | 3 | have found applications in | 3 |
| 173 | | 3 | also been successfully applied | 3 |
| 174 | described in | 3 | described in | 5 |
| 175 | we adopt | 5 | we adopt | 5 |
| 176 | | 3 | is it possible to use it to | 5 |
| 177 | | 2 | all possible... have already been characterized | 4 |
| 178 | for instance | 2 | identified by | 4 |
| 179 | studies | 3 | | 4 |
| 180 | for example | 2 | for example | 3 |
| 181 | have presented | 3 | have presented | 4 |
| 182 | based on those presented in | 5 | based on those presented in | 5 |
| 183 | | 2 | this is a difficult case even for | 4 |
| 184 | achieved | 3 | achieved | 5 |

| | Rater 1 | | Rater 2 | |
|--------------|---------------------------------------------------------|--------|--------------------------------------------------------------------------------------|--------|
| Citation No. | Citation | Rating | Citation | Rating |
| 185 | however; this may be insufficient | 1 | however the results of our system may be suggested that this may be insufficient for | 1 |
| 186 | use | 3 | use | 4 |
| 187 | to date some local solutions have already been proposed | 3 | to date some local solutions have already been proposed | 3 |
| 188 | examples include | 2 | examples include | 3 |
| 189 | e.g. | 2 | | 4 |
| 190 | | 3 | variation of Allan's techniques is missing | 5 |
| 191 | describe | 3 | | 4 |
| 192 | appear in ... work | 3 | we too used | 2 |
| 193 | but; are still needed | 1 | considerable approaches have been developed | 3 |
| 194 | e.g. | 2 | | 4 |
| 195 | suggested | 3 | it seems reasonable that ... could be much more | 2 |
| 196 | discuss | 3 | is not appropriate | 5 |
| 197 | suggest a similar approach | 3 | suggest a similar approach | 2 |
| 198 | | 3 | | 2 |
| 199 | | 3 | | 3 |
| 200 | shows | 3 | | 3 |
| 201 | showed | 3 | preliminary results ... Berghel | 3 |
| 202 | derived from | 5 | we use ... derived from a ... obtained from | 5 |
| 203 | however; not well suited | 1 | has incorporated a ... that employs a | 4 |
| 204 | suggests | 2 | there is some research that suggests | 3 |
| 205 | as a result of | 5 | | 3 |
| 206 | but they are not | 1 | but they are not | 1 |
| 207 | is not | 1 | work with ... that represent ... that gives the | 4 |

| | Rater 1 | | Rater 2 | |
|--------------|--------------------------------------|--------|-----------------------------------------------|--------|
| Citation No. | Citation | Rating | Citation | Rating |
| 208 | | 3 | | 3 |
| 209 | are similar to | 4 | are similar to | 2 |
| 210 | for example | 3 | | 4 |
| 211 | | 2 | | 4 |
| 212 | | 3 | in | 4 |
| 213 | in | 3 | in | 4 |
| 214 | but | 1 | it might also consider | 3 |
| 215 | we adopted | 5 | we adopted | 5 |
| 216 | recent work has demonstrated | 3 | recent work has demonstrated | 4 |
| 217 | however | 1 | represents an approach to ... for the task of | 4 |
| 218 | and is in harmony with | 2 | and is in harmony with | 3 |
| 219 | has involved | 5 | our own research has involved | 5 |
| 220 | used to bootstrap | 5 | the initial motivation for | 3 |
| 221 | background | 3 | background | 5 |
| 222 | found in the literature | 3 | | 3 |
| 223 | | 2 | | 3 |
| 224 | have been developed | 3 | | 3 |
| 225 | are widely agreed upon | 3 | are widely agreed upon | 3 |
| 226 | | 3 | | 3 |
| 227 | similarly | 3 | similarly | 4 |
| 228 | literature provides several examples | 3 | literature provides several examples | 3 |
| 229 | it has been observed | 3 | it has been observed in | 3 |
| 230 | examples ... include | 3 | example of ... include | 3 |
| 231 | enumerated | 3 | enumerated | 4 |

| | Rater 1 | | Rater 2 | |
|---------------------|------------------------------|---------------|-------------------------------------|---------------|
| Citation No. | Citation | Rating | Citation | Rating |
| 232 | have also studied | 3 | have also studied | 3 |
| 233 | studied | 1 | studied | 3 |
| 234 | definitions of these terms | 5 | definitions of these terms | 3 |
| 235 | | 2 | in this paper | 3 |
| 236 | | 2 | more over ... and so forth | 3 |
| 237 | which models...in this model | 3 | which models ... in this model | 4 |
| 238 | is proposed | 3 | is proposed | 4 |
| 239 | often referred to as | 2 | we showed the applicability of | 5 |
| 240 | follows the logic of | 5 | follows the logic of | 5 |
| 241 | an example is given in | 3 | an example is given in | 4 |
| 242 | an example | 3 | an example | 4 |
| 243 | for instance | 2 | for instance | 4 |
| 244 | but | 1 | but | 1 |
| 245 | presents | 3 | presents | 4 |
| 246 | | 0 | | 4 |
| 247 | | 0 | many approaches have been used | 3 |
| 248 | | 0 | the system is fully implemented in | 5 |
| 249 | | 0 | all the experiments were done using | 5 |

Appendix F

Data Collection, Storage, and Calculations

A Microsoft Access database was constructed to hold the documents, citations, abstracts, and data and to calculate metrics. This included identifying information on each citation: similarity metrics, the location of the citation in the document, self-citations, and all the subject response data.

Position and self-citation were entered as mutually exclusive dummy variables. The field “Self-Citing” was yes/no and was recorded in the database as “1” for self-citing extracts and “0” for non-self-citing extracts. Five dummy variables represented the location of a citation in a document: 1 = introduction, 2 = review of the literature, 3 = discussion, 4 = methodology, and 5 = conclusion.

Database

The results of the human ratings and classifications of citations were entered into the database. Refer to Figure F-1, the Entity Relationship diagram of the database that was constructed to hold and manipulate the data for this study. The following is a description of the tables of the database and a brief explanation of how the data was manipulated.

- Main Table: Contains all the general bibliographic information of all 50 documents, along with the entire abstract entered in a memo field of the database.

Complete documents are too large to fit into an Access database, so a hyperlink was created for instant location and viewing of documents when necessary.

- Abstract Terms Table: Contains all the individual words of the abstract, excluding stop words. This is done so that the similarity metric can be calculated from individual word frequencies. Queries generate various metric measures from the data.
- Document Terms Table: Contains all the words, excluding stop words, extracted from the entire document. Queries generate frequency counts and various metric measures from the data.
- Document Citation Table: Contains bibliographic information about the document that has the citations. In addition, this table contains the complete context citations extracted about the 50 selected source documents and fields to identify self-citing and location of the citation within the document.
- Citation Words Table: Contains all the individual words extracted from each citation, excluding stop words. This is to calculate metrics on the citations.
- Citation Category Scores Table: Contains the data generated by the two experts in determining the categorical status of the citations. This table also tracks cue words.
- Document Citation Scores Table: Contains data generated from the subjects' evaluation.

Automating the Selection of Citations as Document Summaries -- Database Diagram

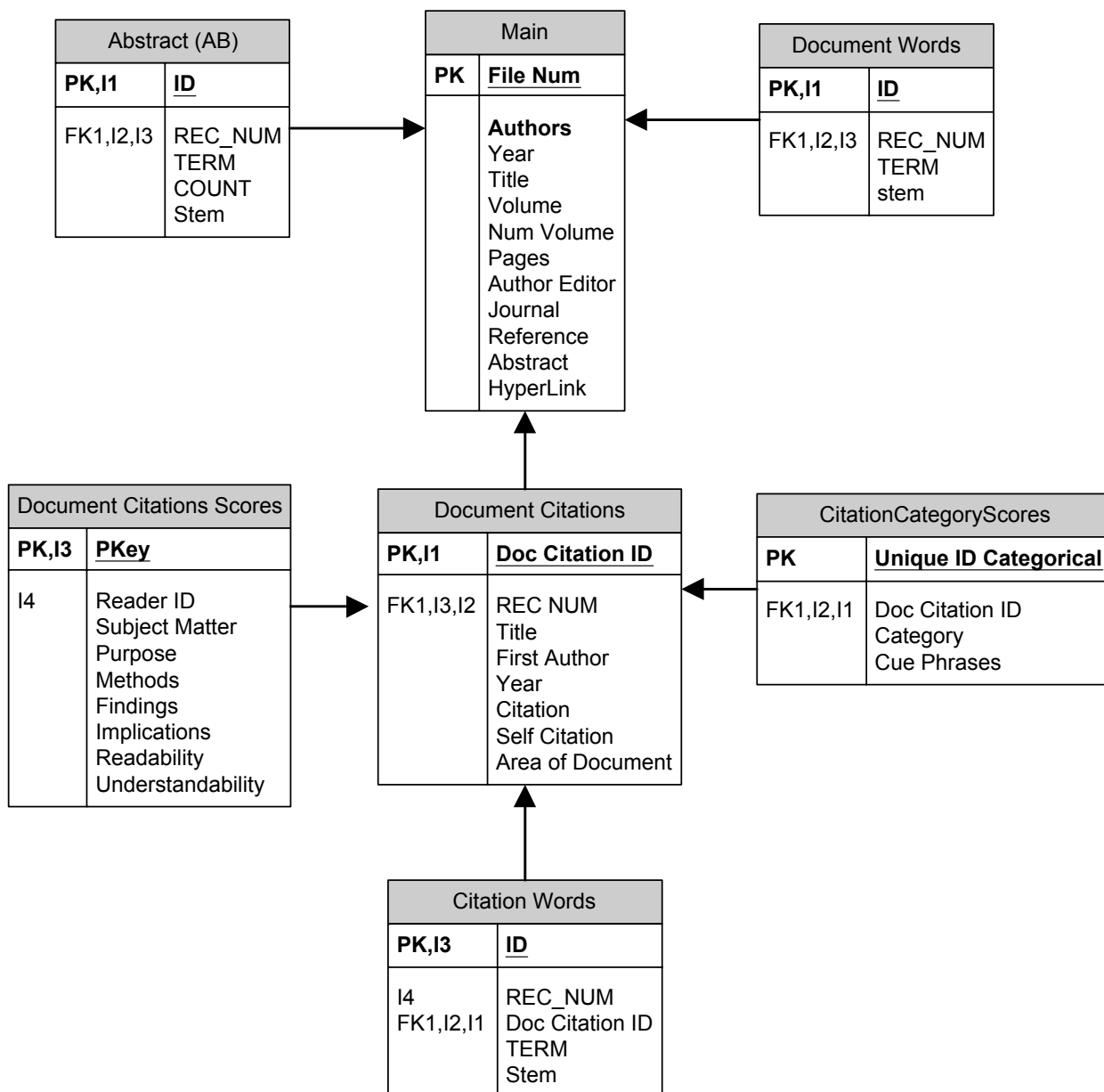


Figure F-1. Database Diagram

Appendix G

Ranking of Citation Context for Expert, Non-expert, and the Combined Groups

The first column is the number of the abstract. The second column is the number of the citation context related to the abstract in the first column. The frequency count column is the number of words in common between the abstract and the citation. Cosine is a similarity metric and agrees with the frequency count column. The next three columns represent the three groups. The grey box indicates the highest ranking for each column. For example, for the first document we can see that citation context 1 has the highest word frequency and cosine related to the first abstract. Experts selected citation 4 and non-experts selected citation 3, based on the average of their survey ratings. The combined ratings in the combined column also selected citation context 3. In this case we can see that neither group agreed with the highest frequency count or cosine.

| Document | Citation | Freq Count | Cosine | Combined | Experts | Non-Experts |
|----------|----------|------------|----------|----------|---------|-------------|
| 1 | 1 | 16 | 0.383131 | 2.50 | 2.68 | 2.38 |
| 1 | 3 | 13 | 0.345349 | 3.49 | 3.29 | 3.62 |
| 1 | 2 | 10 | 0.302891 | 2.89 | 3.32 | 2.60 |
| 1 | 4 | 10 | 0.302891 | 3.46 | 3.43 | 3.48 |
| 1 | 5 | 9 | 0.287348 | 2.73 | 3.14 | 2.45 |
| | | | | | | |
| 2 | 6 | 10 | 0.418854 | 2.47 | 3.14 | 2.02 |
| 2 | 9 | 10 | 0.418854 | 2.17 | 2.61 | 1.88 |
| 2 | 8 | 8 | 0.374634 | 3.14 | 3.61 | 2.83 |
| 2 | 7 | 7 | 0.350438 | 3.29 | 3.68 | 3.02 |
| 2 | 10 | 5 | 0.296174 | 2.69 | 2.71 | 2.67 |
| | | | | | | |
| 3 | 14 | 9 | 0.323498 | 3.10 | 3.61 | 2.82 |
| 3 | 11 | 8 | 0.304997 | 2.55 | 3.36 | 2.08 |
| 3 | 12 | 4 | 0.215666 | 2.51 | 3.11 | 2.16 |
| 3 | 15 | 4 | 0.215666 | 3.09 | 3.25 | 3.00 |
| 3 | 13 | 3 | 0.186772 | 2.39 | 3.07 | 2.00 |
| | | | | | | |
| 4 | 19 | 10 | 0.5 | 2.88 | 3.29 | 2.65 |
| 4 | 18 | 7 | 0.41833 | 3.43 | 3.54 | 3.37 |
| 4 | 16 | 5 | 0.353553 | 2.52 | 3.14 | 2.16 |
| 4 | 20 | 4 | 0.316228 | 2.52 | 2.79 | 2.37 |
| 4 | 17 | 2 | 0.223607 | 2.39 | 2.61 | 2.27 |
| | | | | | | |
| 5 | 25 | 7 | 0.32323 | 2.24 | 2.93 | 1.69 |
| 5 | 23 | 6 | 0.299253 | 3.43 | 3.75 | 3.17 |
| 5 | 24 | 5 | 0.273179 | 2.98 | 3.54 | 2.54 |
| 5 | 22 | 4 | 0.244339 | 3.46 | 4.18 | 2.89 |
| 5 | 21 | 3 | 0.211604 | 3.24 | 3.96 | 2.66 |
| 6 | 29 | 11 | 0.340279 | 3.81 | 3.79 | 3.83 |
| 6 | 30 | 10 | 0.324443 | 3.53 | 3.75 | 3.38 |
| 6 | 28 | 7 | 0.271448 | 2.91 | 3.39 | 2.60 |
| 6 | 26 | 6 | 0.251312 | 2.83 | 3.04 | 2.69 |
| 6 | 27 | 6 | 0.251312 | 3.13 | 3.50 | 2.88 |
| | | | | | | |
| 7 | 31 | 11 | 0.42465 | 3.46 | 3.75 | 3.32 |
| 7 | 34 | 9 | 0.384111 | 2.93 | 3.18 | 2.80 |
| 7 | 233 | 9 | 0.384111 | 2.81 | 3.19 | 2.43 |
| 7 | 232 | 8 | 0.362143 | 3.43 | 3.38 | 3.48 |
| 7 | 35 | 7 | 0.338754 | 2.94 | 2.64 | 3.10 |
| 7 | 236 | 7 | 0.338754 | 3.52 | 3.14 | 3.90 |
| 7 | 33 | 5 | 0.286299 | 3.01 | 3.96 | 2.54 |
| 7 | 234 | 5 | 0.286299 | 3.02 | 2.95 | 3.10 |
| 7 | 235 | 5 | 0.286299 | 2.90 | 3.05 | 2.76 |
| 7 | 32 | 4 | 0.256074 | 2.50 | 2.79 | 2.36 |

| Document | Citation | Freq Count | Cosine | Combined | Experts | Non-Experts |
|----------|----------|------------|----------|----------|---------|-------------|
| 7 | 237 | 4 | 0.256074 | 3.10 | 3.48 | 2.71 |
| 8 | 36 | 8 | 0.281439 | 3.08 | 3.14 | 3.00 |
| 9 | 41 | 15 | 0.453298 | 3.60 | 3.54 | 3.63 |
| 9 | 201 | 12 | 0.405442 | 3.66 | 3.67 | 3.66 |
| 9 | 203 | 12 | 0.405442 | 3.30 | 3.24 | 3.34 |
| 9 | 38 | 11 | 0.388182 | 2.12 | 1.57 | 2.43 |
| 9 | 37 | 10 | 0.370117 | 3.12 | 3.50 | 2.90 |
| 9 | 40 | 9 | 0.351123 | 2.92 | 3.71 | 2.47 |
| 9 | 204 | 8 | 0.331042 | 3.11 | 3.57 | 2.83 |
| 9 | 205 | 7 | 0.309662 | 2.48 | 2.43 | 2.51 |
| 9 | 39 | 6 | 0.286691 | 2.87 | 3.39 | 2.57 |
| 9 | 202 | 6 | 0.286691 | 2.52 | 2.52 | 2.51 |
| 10 | 42 | 13 | 0.447214 | 2.98 | 3.39 | 2.66 |
| 10 | 45 | 12 | 0.429669 | 2.98 | 3.04 | 2.94 |
| 10 | 46 | 9 | 0.372104 | 2.95 | 3.00 | 2.91 |
| 10 | 43 | 7 | 0.328165 | 2.60 | 3.36 | 2.00 |
| 10 | 44 | 5 | 0.27735 | 3.22 | 2.89 | 3.49 |
| 11 | 49 | 14 | 0.385922 | 3.24 | 3.29 | 3.21 |
| 11 | 48 | 10 | 0.326164 | 2.59 | 2.96 | 2.33 |
| 11 | 47 | 9 | 0.309426 | 2.80 | 3.32 | 2.45 |
| 12 | 55 | 13 | 0.457905 | 3.47 | 3.93 | 2.86 |
| 12 | 51 | 11 | 0.421212 | 2.94 | 3.11 | 2.71 |
| 12 | 53 | 10 | 0.40161 | 3.37 | 3.75 | 2.86 |
| 12 | 50 | 9 | 0.381 | 3.24 | 3.75 | 2.57 |
| 12 | 54 | 9 | 0.381 | 2.98 | 3.32 | 2.52 |
| 12 | 52 | 6 | 0.311086 | 3.22 | 3.61 | 2.71 |
| 13 | 56 | 8 | 0.565685 | 2.99 | 2.82 | 3.10 |
| 13 | 58 | 6 | 0.489898 | 3.23 | 3.11 | 3.31 |
| 13 | 57 | 5 | 0.447214 | 2.77 | 2.86 | 2.71 |
| 13 | 59 | 5 | 0.447214 | 2.29 | 2.96 | 1.83 |
| 13 | 60 | 2 | 0.282843 | 2.29 | 2.86 | 1.90 |
| 14 | 65 | 12 | 0.330289 | 2.96 | 3.14 | 2.83 |
| 14 | 61 | 9 | 0.286039 | 3.09 | 3.50 | 2.81 |
| 14 | 62 | 8 | 0.26968 | 2.87 | 3.04 | 2.76 |
| 14 | 63 | 6 | 0.23355 | 2.71 | 3.32 | 2.31 |
| 14 | 64 | 6 | 0.23355 | 2.36 | 2.64 | 2.17 |
| 15 | 66 | 15 | 0.417635 | 3.13 | 3.57 | 2.77 |
| 15 | 67 | 14 | 0.403473 | 3.48 | 3.71 | 3.29 |

| Document | Citation | Freq Count | Cosine | Combined | Experts | Non-Experts |
|----------|----------|------------|----------|----------|---------|-------------|
| 15 | 70 | 10 | 0.340997 | 2.87 | 3.18 | 2.63 |
| 15 | 68 | 9 | 0.323498 | 2.73 | 2.79 | 2.69 |
| 15 | 69 | 9 | 0.323498 | 3.21 | 3.71 | 2.80 |
| | | | | | | |
| 16 | 77 | 14 | 0.32691 | 3.29 | 3.39 | 3.20 |
| 16 | 72 | 13 | 0.315018 | 3.40 | 3.43 | 3.37 |
| 16 | 73 | 12 | 0.30266 | 3.54 | 3.61 | 3.49 |
| 16 | 75 | 12 | 0.30266 | 3.65 | 3.68 | 3.63 |
| 16 | 76 | 12 | 0.30266 | 3.32 | 3.43 | 3.23 |
| 16 | 78 | 11 | 0.289775 | 3.06 | 3.21 | 2.94 |
| 16 | 74 | 9 | 0.262111 | 2.62 | 3.11 | 2.23 |
| 16 | 79 | 5 | 0.195366 | 2.22 | 2.50 | 2.00 |
| 16 | 80 | 5 | 0.195366 | 2.94 | 3.32 | 2.63 |
| 16 | 71 | 3 | 0.15133 | 2.86 | 3.14 | 2.63 |
| | | | | | | |
| 17 | 85 | 7 | 0.336011 | 3.18 | 3.57 | 3.07 |
| 17 | 84 | 6 | 0.311086 | 3.18 | 3.32 | 3.14 |
| 17 | 81 | 4 | 0.254 | 2.61 | 2.93 | 2.52 |
| 17 | 82 | 3 | 0.219971 | 2.78 | 2.86 | 2.76 |
| 17 | 83 | 3 | 0.219971 | 2.83 | 2.64 | 2.88 |
| | | | | | | |
| 18 | 87 | 15 | 0.369274 | 3.69 | 3.50 | 3.79 |
| 18 | 89 | 8 | 0.26968 | 2.69 | 2.96 | 2.55 |
| 18 | 86 | 7 | 0.252262 | 2.90 | 2.93 | 2.89 |
| 18 | 90 | 7 | 0.252262 | 2.38 | 2.96 | 2.09 |
| 18 | 88 | 5 | 0.213201 | 2.52 | 2.50 | 2.54 |
| | | | | | | |
| 19 | 93 | 8 | 0.335673 | 2.83 | 3.39 | 2.45 |
| 19 | 92 | 7 | 0.313993 | 3.03 | 3.32 | 2.83 |
| 19 | 91 | 5 | 0.265372 | 2.54 | 3.54 | 1.88 |
| 19 | 94 | 4 | 0.237356 | 2.64 | 3.18 | 2.29 |
| | | | | | | |
| 20 | 96 | 4 | 0.285714 | 2.98 | 3.00 | 2.97 |
| 20 | 99 | 3 | 0.247436 | 2.44 | 3.18 | 1.86 |
| 20 | 98 | 2 | 0.202031 | 2.33 | 2.36 | 2.31 |
| 20 | 95 | 1 | 0.142857 | 2.57 | 2.89 | 2.31 |
| 20 | 97 | 1 | 0.142857 | 2.35 | 2.93 | 1.89 |
| | | | | | | |
| 21 | 100 | 10 | 0.347105 | 3.07 | 3.68 | 2.77 |
| 21 | 103 | 10 | 0.347105 | 2.95 | 3.43 | 2.71 |
| 21 | 101 | 9 | 0.329293 | 2.75 | 3.00 | 2.63 |
| 21 | 104 | 9 | 0.329293 | 2.70 | 3.46 | 2.32 |
| 21 | 102 | 4 | 0.219529 | 2.50 | 2.96 | 2.27 |
| | | | | | | |
| 22 | 105 | 9 | 0.384111 | 3.38 | 3.57 | 3.27 |
| 22 | 106 | 8 | 0.362143 | 3.10 | 3.11 | 3.10 |

| Document | Citation | Freq Count | Cosine | Combined | Experts | Non-Experts |
|----------|----------|------------|----------|----------|---------|-------------|
| 22 | 107 | 7 | 0.338754 | 2.73 | 3.00 | 2.57 |
| 22 | 108 | 4 | 0.256074 | 2.94 | 3.18 | 2.80 |
| 22 | 109 | 10 | 0.404888 | 3.00 | 3.18 | 2.90 |
| | | | | | | |
| 23 | 111 | 11 | 0.4022 | 3.04 | 3.39 | 2.57 |
| 23 | 113 | 9 | 0.363803 | 3.71 | 4.00 | 3.33 |
| 23 | 112 | 6 | 0.297044 | 2.96 | 3.57 | 2.14 |
| 23 | 110 | 5 | 0.271163 | 2.98 | 3.29 | 2.57 |
| 23 | 114 | 0 | 0 | 2.57 | 2.89 | 2.14 |
| | | | | | | |
| 24 | 117 | 6 | 0.402694 | 3.08 | 3.18 | 3.02 |
| 24 | 119 | 6 | 0.402694 | 3.01 | 3.39 | 2.80 |
| 24 | 116 | 2 | 0.232495 | 2.65 | 2.75 | 2.59 |
| 24 | 118 | 2 | 0.232495 | 2.61 | 2.54 | 2.65 |
| 24 | 115 | 1 | 0.164399 | 2.70 | 3.11 | 2.47 |
| | | | | | | |
| 25 | 127 | 8 | 0.508001 | 3.73 | 3.82 | 3.66 |
| 25 | 125 | 7 | 0.475191 | 3.29 | 3.54 | 3.09 |
| 25 | 126 | 4 | 0.359211 | 2.76 | 2.61 | 2.89 |
| | | | | | | |
| 26 | 128 | 3 | 0.433013 | 3.16 | 3.46 | 2.76 |
| 26 | 129 | 3 | 0.433013 | 2.98 | 3.25 | 2.62 |
| 26 | 131 | 2 | 0.353553 | 3.47 | 3.39 | 3.57 |
| 26 | 130 | 1 | 0.25 | 3.06 | 3.43 | 2.57 |
| | | | | | | |
| 27 | 128 | 3 | 0.433013 | 3.16 | 3.46 | 2.76 |
| 27 | 129 | 3 | 0.433013 | 2.98 | 3.25 | 2.62 |
| 27 | 131 | 2 | 0.353553 | 3.47 | 3.39 | 3.57 |
| 27 | 130 | 1 | 0.25 | 3.06 | 3.43 | 2.57 |
| | | | | | | |
| 28 | 136 | 5 | 0.527046 | 3.87 | 3.82 | 3.90 |
| 28 | 132 | 4 | 0.471405 | 2.61 | 2.93 | 2.40 |
| 28 | 133 | 4 | 0.471405 | 2.79 | 2.75 | 2.81 |
| 28 | 137 | 3 | 0.408248 | 3.00 | 3.29 | 2.81 |
| 28 | 134 | 2 | 0.333333 | 3.06 | 3.29 | 2.90 |
| 28 | 135 | 2 | 0.333333 | 2.90 | 3.43 | 2.55 |
| | | | | | | |
| 29 | 140 | 6 | 0.377964 | 3.04 | 3.21 | 2.95 |
| 29 | 142 | 6 | 0.377964 | 3.48 | 3.93 | 3.25 |
| 29 | 141 | 5 | 0.345033 | 2.96 | 3.89 | 2.50 |
| 29 | 138 | 3 | 0.267261 | 3.42 | 4.07 | 3.09 |
| 29 | 139 | 2 | 0.218218 | 2.64 | 3.18 | 2.38 |
| | | | | | | |
| 30 | 146 | 13 | 0.457905 | 3.29 | 3.19 | 3.33 |
| 30 | 145 | 11 | 0.421212 | 3.11 | 3.24 | 3.05 |
| 30 | 144 | 10 | 0.40161 | 2.73 | 3.33 | 2.43 |

| Document | Citation | Freq Count | Cosine | Combined | Experts | Non-Experts |
|----------|----------|------------|----------|----------|---------|-------------|
| 30 | 143 | 6 | 0.311086 | 2.78 | 3.38 | 2.48 |
| 30 | 147 | 4 | 0.254 | 2.70 | 3.33 | 2.38 |
| | | | | | | |
| 31 | 149 | 11 | 0.353553 | 3.21 | 3.38 | 3.11 |
| 31 | 154 | 10 | 0.3371 | 2.57 | 3.10 | 2.26 |
| 31 | 155 | 10 | 0.3371 | 3.84 | 4.05 | 3.71 |
| 31 | 152 | 8 | 0.301511 | 3.55 | 3.67 | 3.49 |
| 31 | 148 | 6 | 0.261116 | 2.70 | 2.29 | 2.94 |
| 31 | 150 | 6 | 0.261116 | 2.95 | 3.05 | 2.89 |
| 31 | 153 | 6 | 0.261116 | 3.20 | 2.86 | 3.40 |
| 31 | 151 | 4 | 0.213201 | 3.20 | 3.52 | 3.00 |
| | | | | | | |
| 32 | 159 | 9 | 0.358569 | 3.20 | 3.14 | 3.22 |
| 32 | 156 | 7 | 0.316228 | 2.89 | 2.90 | 2.88 |
| 32 | 160 | 7 | 0.316228 | 3.21 | 3.05 | 3.29 |
| 32 | 157 | 6 | 0.29277 | 3.19 | 3.38 | 3.10 |
| 32 | 158 | 5 | 0.267261 | 2.76 | 2.81 | 2.73 |
| | | | | | | |
| 33 | 168 | 8 | 0.365148 | 3.17 | 3.57 | 2.98 |
| 33 | 162 | 7 | 0.341565 | 3.11 | 3.00 | 3.17 |
| 33 | 161 | 6 | 0.316228 | 3.16 | 3.00 | 3.24 |
| 33 | 163 | 6 | 0.316228 | 3.16 | 2.76 | 3.36 |
| 33 | 169 | 6 | 0.316228 | 3.38 | 3.81 | 3.17 |
| 33 | 164 | 5 | 0.288675 | 3.03 | 2.81 | 3.14 |
| 33 | 165 | 3 | 0.223607 | 2.67 | 2.67 | 2.67 |
| 33 | 166 | 3 | 0.223607 | 3.02 | 3.62 | 2.71 |
| 33 | 170 | 3 | 0.223607 | 2.98 | 3.57 | 2.69 |
| 33 | 167 | 2 | 0.182574 | 2.75 | 2.95 | 2.64 |
| | | | | | | |
| 34 | 171 | 10 | 0.355784 | 2.30 | 2.67 | 2.09 |
| 34 | 172 | 7 | 0.29767 | 2.54 | 2.48 | 2.57 |
| 34 | 174 | 7 | 0.29767 | 1.84 | 3.05 | 1.11 |
| 34 | 175 | 7 | 0.29767 | 2.51 | 2.62 | 2.43 |
| 34 | 173 | 6 | 0.275589 | 2.71 | 2.38 | 2.91 |
| | | | | | | |
| 35 | 176 | 7 | 0.394405 | 2.93 | 3.67 | 2.61 |
| 35 | 178 | 7 | 0.394405 | 2.60 | 3.33 | 2.29 |
| 35 | 177 | 5 | 0.333333 | 3.09 | 3.24 | 3.02 |
| 35 | 179 | 5 | 0.333333 | 2.66 | 3.57 | 2.27 |
| 35 | 180 | 5 | 0.333333 | 2.36 | 3.10 | 2.04 |
| | | | | | | |
| 36 | 186 | 9 | 0.437595 | 3.49 | 3.95 | 3.26 |
| 36 | 185 | 7 | 0.385922 | 2.87 | 3.29 | 2.67 |
| 36 | 181 | 6 | 0.357295 | 3.11 | 3.24 | 3.05 |
| 36 | 184 | 4 | 0.29173 | 2.95 | 3.71 | 2.57 |
| 36 | 182 | 3 | 0.252646 | 2.83 | 3.24 | 2.62 |

| Document | Citation | Freq Count | Cosine | Combined | Experts | Non-Experts |
|----------|----------|------------|----------|----------|---------|-------------|
| 36 | 183 | 3 | 0.252646 | 2.94 | 3.57 | 2.62 |
| 37 | 188 | 13 | 0.336219 | 2.67 | 2.43 | 2.78 |
| 37 | 187 | 9 | 0.279751 | 2.90 | 2.81 | 2.94 |
| 37 | 190 | 9 | 0.279751 | 3.11 | 3.29 | 3.04 |
| 37 | 191 | 8 | 0.263752 | 3.00 | 3.67 | 2.71 |
| 37 | 189 | 6 | 0.228416 | 3.11 | 3.48 | 2.96 |
| 38 | 195 | 7 | 0.333333 | 3.55 | 3.48 | 3.60 |
| 38 | 196 | 7 | 0.333333 | 3.71 | 3.57 | 3.80 |
| 38 | 192 | 5 | 0.281718 | 3.61 | 3.57 | 3.63 |
| 38 | 193 | 3 | 0.218218 | 2.64 | 2.71 | 2.60 |
| 38 | 194 | 1 | 0.125988 | 2.36 | 2.81 | 2.09 |
| 39 | 198 | 8 | 0.335673 | 2.82 | 2.95 | 2.71 |
| 39 | 199 | 8 | 0.335673 | 2.69 | 2.90 | 2.54 |
| 39 | 197 | 5 | 0.265372 | 2.47 | 3.33 | 1.82 |
| 39 | 200 | 1 | 0.118678 | 2.10 | 2.52 | 1.79 |
| 40 | 209 | 7 | 0.413197 | 2.87 | 3.00 | 2.81 |
| 40 | 206 | 6 | 0.382546 | 3.00 | 3.00 | 3.00 |
| 40 | 210 | 5 | 0.349215 | 3.08 | 3.19 | 3.02 |
| 40 | 207 | 4 | 0.312348 | 3.00 | 3.62 | 2.69 |
| 40 | 208 | 2 | 0.220863 | 2.63 | 2.76 | 2.57 |
| 40 | 211 | 2 | 0.220863 | 2.68 | 2.33 | 2.89 |
| 41 | 213 | 13 | 0.331918 | 3.08 | 3.57 | 2.89 |
| 41 | 215 | 12 | 0.318896 | 3.45 | 3.19 | 3.55 |
| 41 | 216 | 9 | 0.276172 | 3.31 | 3.52 | 3.23 |
| 41 | 212 | 5 | 0.205847 | 2.21 | 2.62 | 2.05 |
| 41 | 214 | 2 | 0.130189 | 2.70 | 2.76 | 2.68 |
| 42 | 217 | 13 | 0.355266 | 3.38 | 3.62 | 3.26 |
| 42 | 221 | 11 | 0.326797 | 3.30 | 3.52 | 3.19 |
| 42 | 218 | 9 | 0.295599 | 2.83 | 3.00 | 2.74 |
| 42 | 219 | 9 | 0.295599 | 3.46 | 3.71 | 3.33 |
| 42 | 220 | 8 | 0.278693 | 3.19 | 3.10 | 3.24 |
| 43 | 228 | 8 | 0.356348 | 3.27 | 3.38 | 3.20 |
| 43 | 224 | 7 | 0.333333 | 3.29 | 2.86 | 3.54 |
| 43 | 229 | 6 | 0.308607 | 3.46 | 3.52 | 3.43 |
| 43 | 227 | 5 | 0.281718 | 3.23 | 3.95 | 2.80 |
| 43 | 223 | 4 | 0.251976 | 2.96 | 2.86 | 3.03 |
| 43 | 226 | 4 | 0.251976 | 2.86 | 3.14 | 2.69 |
| 43 | 231 | 4 | 0.251976 | 2.86 | 3.10 | 2.71 |
| 43 | 225 | 3 | 0.218218 | 2.82 | 3.10 | 2.66 |

| Document | Citation | Freq Count | Cosine | Combined | Experts | Non-Experts |
|----------|----------|------------|----------|----------|---------|-------------|
| 43 | 222 | 2 | 0.178174 | 2.11 | 1.86 | 2.26 |
| 43 | 230 | 1 | 0.125988 | 3.14 | 3.33 | 3.03 |
| | | | | | | |
| 44 | 242 | 12 | 0.443533 | 2.98 | 3.86 | 2.46 |
| 44 | 241 | 11 | 0.42465 | 3.04 | 3.81 | 2.57 |
| 44 | 245 | 9 | 0.384111 | 3.54 | 3.57 | 3.51 |
| 44 | 239 | 8 | 0.362143 | 2.54 | 3.29 | 2.09 |
| 44 | 238 | 7 | 0.338754 | 2.73 | 3.86 | 2.06 |
| 44 | 244 | 6 | 0.313625 | 2.57 | 3.00 | 2.31 |
| 44 | 240 | 5 | 0.286299 | 2.86 | 2.95 | 2.80 |
| 44 | 243 | 4 | 0.256074 | 2.95 | 3.52 | 2.60 |
| | | | | | | |
| 45 | 247 | 8 | 0.2965 | 3.78 | 3.67 | 3.83 |
| 45 | 246 | 5 | 0.234404 | 2.86 | 3.14 | 2.71 |
| 45 | 249 | 5 | 0.234404 | 2.49 | 3.14 | 2.17 |
| 45 | 248 | 4 | 0.209657 | 3.29 | 3.33 | 3.26 |

Appendix H

Stepwise Regression Tables

Method: Stepwise forward regression of the nine eligible variables (abstract, area, sub, purp, meth, find, impl, read, under); variables were added to the model sequentially. In a given step, the variable that could account for the largest amount of remaining variation (shown in Seq SS in tables) was added if the variable explained a significant amount of the remaining variation at a 10% level.

COMBINED SCORES — Unweighted Cosine

Step History

| Step | Parameter | Action | “Sig Prob” | Seq SS | RSquare |
|------|------------|---------|------------|----------|---------|
| 1 | Abstract | Entered | 0.0000 | 0.687139 | 0.4254 |
| 2 | Sub_comb | Entered | 0.0000 | 0.275139 | 0.5958 |
| 3 | Meth_comb | Entered | 0.0031 | 0.027747 | 0.6130 |
| 4 | Under_comb | Entered | 0.0150 | 0.01818 | 0.6242 |
| 5 | Purp_comb | Entered | 0.0860 | 0.0089 | 0.6297 |

EXPERT SCORES — Unweighted Cosine

Step History

| Step | Parameter | Action | “Sig Prob” | Seq SS | RSquare |
|-------------|------------------|---------------|-------------------|---------------|----------------|
| 45 | Abstract | Entered | 0.0000 | 0.687139 | 0.4254 |
| 46 | Subj_exp | Entered | 0.0000 | 0.180311 | 0.5371 |

STUDENT SCORES — Unweighted Cosine

Step History

| Step | Parameter | Action | “Sig Prob” | Seq SS | RSquare |
|-------------|------------------|---------------|-------------------|---------------|----------------|
| 1 | abstract | Entered | 0.0000 | 0.687139 | 0.4254 |
| 2 | sub_st | Entered | 0.0000 | 0.219909 | 0.5616 |
| 3 | meth_st | Entered | 0.0012 | 0.035681 | 0.5837 |
| 4 | under_st | Entered | 0.0092 | 0.0224 | 0.5975 |

COMBINED SCORES — Frequency Counts

Step History

| Step | Parameter | Action | “Sig Prob” | Seq SS | RSquare |
|-------------|------------------|---------------|-------------------|---------------|----------------|
| 1 | abstract | Entered | 0.0000 | 1287.9884 | 0.4859 |
| 2 | sub_comb | Entered | 0.0000 | 405.6758 | 0.6389 |
| 3 | Meth_comb | Entered | 0.0008 | 51.50368 | 0.6584 |
| 4 | under_comb | Entered | 0.0350 | 19.85951 | 0.6658 |
| 5 | purp_comb | Entered | 0.0504 | 16.83093 | 0.6722 |

EXPERT SCORES — Frequency counts

Step History

| Step | Parameter | Action | “Sig Prob” | Seq SS | RSquare |
|-------------|------------------|---------------|-------------------|---------------|----------------|
| 1 | abstract | Entered | 0.0000 | 1287.9884 | 0.4859 |
| 2 | subj_exp | Entered | 0.0000 | 260.3277 | 0.5841 |

STUDENT SCORES — Frequency counts

Step History

| Step | Parameter | Action | “Sig Prob” | Seq SS | RSquare |
|-------------|------------------|---------------|-------------------|---------------|----------------|
| 1 | abstract | Entered | 0.0000 | 1287.9884 | 0.4859 |
| 2 | purp_st | Entered | 0.0000 | 333.2692 | 0.6116 |
| 3 | meth_st | Entered | 0.0005 | 59.67253 | 0.6341 |
| 4 | under_st | Entered | 0.0617 | 16.74373 | 0.6404 |
| 5 | sub_st | Entered | 0.0332 | 21.43391 | 0.6485 |

JEFF HAND

29 Washington Court
Mullica Hill, NJ 08062

H: (856) 223-9931

E-Mail: Jeff.Hand@drexel.edu

W: (215) 895-0912

Objective:

Associate Professor, Administrator

Academic Preparation:

Ph.D. in Information Systems, College of Information Science and Technology, Drexel University, Philadelphia PA, 2003.

Concentration: Information Retrieval and Statistical Language Processing

Dissertation: Feasibility of Using Citations as Document Summaries

Advisor: Dr. Carl Drott

M.A. in Clinical Psychology, Department of Psychology, West Chester University, 1988

Concentration: Educational Psychology and Psychometrics

Dissertation: Prediction of Harvard Stanford Scale Scores with a Phenomenological Instrument

Advisor: Dr. V. Kumar

B.S. in Electrical Engineering, Department of Engineering, Temple University, 1980

Concentration: Computer Architecture

A.A. Liberal Studies, Bucks County Community College, 1978

Employment History

| | |
|-----------------|----------------------------------------------|
| 12/03 – Present | Assistant Dean |
| 03/02 – 12/03 | Program Manager |
| 03/02 – Present | Adjunct Associate Professor |
| 06/98 – 03/02 | Adjunct Assistant Professor |
| 01/98 – 12/98 | Adjunct Instructor |
| 01/97 – 12/97 | Teaching Assistant |
| 01/96 – 12/96 | Research Assistant |
| 10/91 – 8/95 | President, CleanNet Corporation |
| 1988 – 1991 | Director of Marketing, Biddle Instruments |
| 1986 – 1988 | Director of Technical Marketing |
| 1983 – 1986 | Product Manager, AT&T Information Systems |
| 1980 – 1983 | Product Manager, Commodore Business Machines |

Teaching and Research Interests

The content focal point of my teaching is Information Systems and Business with specializations in Six Sigma, Electronic Marketing, System Analysis and Design, and Statistical Language Processing. My research focus on Education in Business, Information Systems and the Adult Learner.

Honors and Awards

Student Choice Award Teacher of the Year 2002. An annual award given to a teacher as voted by the student body of the University for outstanding classroom instruction and concern for students.

Teaching Assistant of the Year 1999 for the College of Information Science. An annual award given to an outstanding student teacher as selected by the administration and the student body of the College.

Publications

Prediction of Harvard Stanford Scale Scores with a Phenomenological Instrument, Hand, Jeff, Pekala, Ronald, and Kumar, V.K. Australian Journal of Clinical and Experimental Hypnosis. Vol. 23, No. 2. 1995, pp. 124 –134.

Professional Memberships

Institute of Electrical and Electronic Engineers
Association of Computing Machinery

Teaching

Courses Taught

Undergraduate

ISYS 101 INTRODUCTION TO INFORMATION SYSTEMS I

A survey course in information systems and issues associated with their use. Provides a broad-based introduction to computer hardware, software, telecommunications and information.

ISYS 102 INTRODUCTION TO INFORMATION SYSTEMS II

Introduces students to major types of organizational information systems and their development and use.

ISYS 105 INFORMATION EVALUATION, ORGANIZATION, AND USE

Introduction to the users of information systems and the information resources that can be accessed through these systems. Users are considered in terms of their information needs, communication and information seeking behavior, and information processing capabilities. Print and electronic information resources are considered in terms of both their content and structure.

ISYS 110 HUMAN-COMPUTER INTERACTION I

Introduces the student to the study of computer-based user interfaces. Presents a user-centered focus in evaluation of computer interfaces. Teaches the basic principles of user analysis and interface evaluation. Gives a practical introduction to ergonomics.

ISYS 140 INFORMATION SYSTEMS LAB I

Provides hands-on experience with a variety of software products basic to current information systems. Covers products that support personal productivity in organizing, analyzing and presenting information. Addresses both local processing on personal computers and creation and use of information on the Internet.

ISYS 141 INFORMATION SYSTEMS LAB II

Provides hands-on experience with sample, current tools for information systems prototyping and development. Introduces students to basic concepts and facilities for coding, debug, and execution of programs in a software development environment. Provides exposure to creation of simple user interface elements common to many information systems.

ISYS 142 INFORMATION SYSTEMS LAB III

Continues ISYS 141 Information Systems Lab II. Provides hands-on experience with sample, current tools for information systems prototyping and development. Emphasizes prototyping of user interface elements and associated information system behaviors.

ISYS 200 SYSTEMS ANALYSIS I

Study of the principles, practices and tools of information systems analysis and design. Emphasis on learning pragmatic aspects of working as a systems analyst and employing the tools of systems analysis and design.

ISYS 205 STRATEGIC USES OF INFORMATION SYSTEMS

Familiarizes students with basic business problems and operations and provides an understanding of how information systems can be used to benefit organizations. Also introduces students to the pitfalls of developing and implementing information systems in organizations and helps students improve critical thinking skills.

ISYS 210 DATABASE MANAGEMENT SYSTEMS

Focuses on how to design databases for given problems, and how to use database systems effectively. Topics include database design techniques using the entity-relationship approach, techniques of translating the entity-relationship diagram into a relational schema, relational algebra, commercial query languages, and normalization techniques.

ISYS 360 LANGUAGE PROCESSING

Study of the problems and techniques of processing natural language. Introduces theory of spoken language and how it differs from theories of computer-generated natural language. Includes language pattern recognition and syntactic inference, and semantic networks.

Graduate**INFO 503 INTRODUCTION TO INFORMATION SYSTEMS ANALYSIS**

Presents information systems development as a life-cycle process, incorporating problem definition, modeling and analysis, system design, implementation, evaluation, support, and maintenance. Provides an introduction to those modeling and analysis tools and techniques necessary for leveraging information and information technologies to achieve business objectives. Gives students practice in modeling information systems with respect to functions (functional decomposition) processes (dynamic modeling) and data (data-flow diagramming).

INFO 605 DATABASE MANAGEMENT I

A first course in database management systems. Covers database design, data manipulation, and database integrity. Emphasizes concepts and techniques related to the entity-relationship model and relational database systems. Discusses normalization up to the third normal form and commercial query languages.

INFO 608 HUMAN COMPUTER INTERACTION

Focuses on the design and evaluation of human-computer interfaces covering such topics as task analysis techniques for gathering design information, iterative design through prototyping, and formative and summative usability testing; theoretical foundations of HCI and cognitive modeling of user interactions; the integration of HCI techniques into the software development life cycle and the use of user constraints to generate new interaction designs.

| Quarter | Course Name | Number |
|------------------|----------------------------------------------|---------------|
| Winter '95 - '96 | Systems Analysis I | ISYS 200 |
| Spring '95 - '96 | Introduction to Information Systems Analysis | INFO 503 |
| Spring '95 - '96 | Introduction to Information Systems Analysis | INFO 503 |
| Summer '95 - '96 | Introduction to Information Systems Analysis | INFO 503 |

| | | |
|------------------|----------------------------------------------|----------|
| Fall '96 - '97 | Introduction to Information Systems Analysis | INFO 503 |
| Fall '96 - '97 | Systems Analysis I | ISYS 200 |
| Fall '96 - '97 | Information Systems Lab I | ISYS 140 |
| Winter '96 - '97 | Human Computer Interaction | INFO 608 |
| Winter '96 - '97 | Introduction to Information Systems Analysis | INFO 503 |
| Winter '96 - '97 | Information Systems Lab II Visual Basic | ISYS 141 |
| Spring '96 - '97 | Database Management I | INFO 605 |
| Spring '96 - '97 | Information Systems Lab III Visual Basic | ISYS 142 |
| Summer '96 - '97 | Systems Analysis I | ISYS 200 |
| Summer '96 - '97 | Information Systems Analysis II | ISYS 355 |
| Summer '96 - '97 | Information Evaluation, Organization and Use | ISYS 105 |
| Summer '96 - '97 | Information Systems Lab I | ISYS 140 |
| Fall '97 - '98 | Information Evaluation, Organization and Use | ISYS 105 |
| Fall '97 - '98 | Strategic Uses of Information Systems | ISYS 205 |
| Fall '97 - '98 | Human Computer Interaction | ISYS 110 |
| Winter '97 - '98 | Information Systems Lab I | ISYS 140 |
| Spring '97 - '98 | Information Evaluation, Organization and Use | ISYS 105 |
| Spring '97 - '98 | Information Systems Lab I | ISYS 140 |
| Spring '97 - '98 | Human Computer Interaction | INFO 608 |
| Summer '97 - '98 | Information Evaluation, Organization and Use | ISYS 105 |
| Summer '97 - '98 | Information Systems Lab II Visual Basic | ISYS 141 |
| Summer '97 - '98 | Strategic Uses of Information Systems | ISYS 205 |
| Fall '98 - '99 | Information Evaluation, Organization and Use | ISYS 105 |
| Fall '98 - '99 | Human Computer Interaction | INFO 608 |
| Fall '98 - '99 | Introduction to Information Systems Analysis | INFO 503 |
| Winter '98 - '99 | Information Systems II | ISYS 102 |
| Winter '98 - '99 | Information Evaluation, Organization and Use | ISYS 105 |
| Winter '98 - '99 | Information Systems Lab II Visual Basic | ISYS 141 |
| Winter '98 - '99 | Information Systems Lab III Visual Basic | ISYS 142 |
| Spring '98 - '99 | Information Systems Lab II Visual Basic | ISYS 141 |
| Spring '98 - '99 | Information Systems Lab III Visual Basic | ISYS 142 |
| Summer '98 - '99 | Information Systems II | ISYS 102 |
| Fall '99 - '00 | Information Systems I | ISYS 101 |
| Fall '99 - '00 | Information Evaluation, Organization and Use | ISYS 105 |
| Fall '99 - '00 | Information Evaluation, Organization and Use | ISYS 105 |
| Winter '99 - '00 | Information Evaluation, Organization and Use | ISYS 105 |
| Winter '99 - '00 | Information Evaluation, Organization and Use | ISYS 105 |

| | | |
|------------------|----------------------------------------------|----------|
| Winter '99 - '00 | Information Evaluation, Organization and Use | ISYS 105 |
| Spring '99 - '00 | Human Computer Interaction | ISYS 110 |
| Spring '99 - '00 | Human Computer Interaction | ISYS 110 |
| Spring '99 - '00 | Strategic Uses of Information Systems | ISYS 205 |
| Summer '99 - '00 | Introduction to Information Systems Analysis | INFO 503 |
| Summer '99 - '00 | Human Computer Interaction | INFO 608 |
| Summer '99 - '00 | Information Systems II | ISYS 102 |
| Fall '00 - '01 | Information Evaluation, Organization and Use | ISYS 105 |
| Winter '00 - '01 | Human Computer Interaction | INFO 608 |
| Winter '00 - '01 | Human Computer Interaction | ISYS 110 |
| Winter '00 - '01 | Systems Analysis I | ISYS 200 |
| Spring '00 - '01 | Information Systems Lab II Visual Basic | ISYS 142 |
| Spring '00 - '01 | Information Systems Lab II Visual Basic | ISYS 142 |
| Spring '00 - '01 | Systems Analysis I | ISYS 200 |
| Summer '00 - '01 | Human Computer Interaction | INFO 608 |
| Summer '00 - '01 | Human Computer Interaction | ISYS 110 |
| Summer '00 - '01 | Linguistic Processing | ISYS 360 |
| Fall '01 - '02 | Systems Analysis I | ISYS 200 |
| Fall '01 - '02 | Systems Analysis I | ISYS 200 |
| Winter '01 - '02 | Information Evaluation, Organization and Use | ISYS 105 |
| Winter '01 - '02 | Systems Analysis I | ISYS 200 |
| Spring '01 - '02 | Information Evaluation, Organization and Use | ISYS 105 |
| Spring '01 - '02 | Systems Analysis I | ISYS 200 |
| Fall '03 - '04 | Information Systems II | ISYS 102 |
| Fall '03 - '04 | Information Systems II | ISYS 102 |
| Fall '03 - '04 | Information Systems Lab II Visual Basic | ISYS 141 |
| Fall '03 - '04 | Strategic Uses of Information Systems | ISYS 205 |
| Fall '03 - '04 | E-Marketing | CUST 380 |
| Spring '03 - '04 | Information Systems II | ISYS 102 |
| Spring '03 - '04 | Systems Analysis I | ISYS 200 |

