The Use of Automated Document Structuring and Classification Methods in the Legal Domain
By
Edward Charles Bradshaw
B.Sc., The University of British Columbia, 1989
LL.B., The University of British Columbia, 1993

# A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF <br> THE REQUIREMENTS FOR THE DEGREE OF MASTER OF LAWS 

in
THE FACULTY OF GRADUATE STUDIES
(Faculty of Law)
We accept this thesis as conforming to the require Standard

THE UNIVERSITY OF BRITISH COLUMBIA
August 1995
(c) E.C. Bradshaw, 1995

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

## Faculty

Department of $\square$ Low

The University of British Columbia Vancouver, Canada

Date August 30,1995


#### Abstract

This paper presents a review of some of the more innovative and successful projects in the area of automated document classification, along with a practical attempt at document structuring and classification. The focus in selecting projects for this study was on their potential application to a database of legal judgments, though none of the selected projects have actually been applied to legal judgments. The idea was to select a few good projects, focusing on those which have been implemented with some degree of success. Strictly theoretical papers were not considered. Though they need not necessarily be in use commercially, most systems chosen were in fact in use on a daily basis.

The goal of the practical component of this research was an attempt to use the distinctive elements of legal judgments to improve retrieval effectiveness on legal databases. This was to be done by identifying a substructure within a judgment, and then using standard retrieval techniques based on this substructure in addition to the text as a whole. Previous studies have shown that retrieval based on some subdivision of full text documents does indeed show better results. The problem addressed in this project is in identifying this initial substructure.


Table of Contents
Abstract ..... ii
Table of Contents ..... iii
Acknowledgement ..... v
Chapter 1: Automated Document Classification ..... 1
I. Introduction ..... 1
II. Classification and Indexing ..... 2
III. Approaches to Classification ..... 4
IV. Knowledge-Based Representation ..... 5
V. Knowledge-Based Representation Methods ..... 8
A. The Worldviews System ..... 8
B. The CONSTRUE Text Indexing System ..... 10
C. The JULLS Automated Keywording System ..... 12
D. The MedIndEx System ..... 13
VI. Case-Based Representation ..... 16
VII. Case-Based Representation Systems ..... 17
A. A Memory Based Reasoning System ..... 17
B. The PRISM System ..... 19
VIII. Statistical Representations ..... 20
IX. Statistical Representation Systems ..... 20
A. The DR-LINK System ..... 20
X. Classification Based on Document Structure ..... 22
XI. Application to the Legal Domain ..... 24
Chapter 2: Automated Document Structuring ..... 27
I. Introduction ..... 27
A. The Method ..... 28
B. Background ..... 31
II. The Method ..... 33
A. Outline ..... 33
B. Parsing the Text ..... 33
C. The Database ..... 34
D. The Initial Blocks ..... 35
E. Comparing Adjacent Blocks ..... 36
F. Term Weights ..... 37
G. Finding Subtopic Boundaries ..... 39
III. The Results ..... 40
A. Overall Results ..... 40
B. Analysis of Results ..... 42

1. Overview ..... 42
2. Problems with the Method ..... 43
a. General Terms ..... 43
b. Generic Legal Terms ..... 45
c. The Parser ..... 47
3. Problems with the Lexicon ..... 47
a. Cases with no Structure ..... 48
b. Short Paragraphs ..... 49
4. Other Results ..... 50
IV. Alternative Methods ..... 51
Chapter 3: Finding Legal Categories ..... 55
I. Overview ..... 55
II. The Method ..... 56
III. The Results ..... 58
IV. Alternative Substructuring Method ..... 58
V. Results of Alternative Substructuring Method ..... 59
Chapter 4: Conclusion ..... 61
Bibliography ..... 63
Appendix A: Program Code ..... 68
Appendix B: Case Excerpts ..... 99

## ACKNOWLEDGMENTS

I would like to thank J.C. Smith for his help, encouragement and support through this project. I would also like to thank everyone at FLAIR for their help in the beginning stages of this project, and the use of some of their programs.

## Chapter 1: Automated Document Classification

## I. Introduction

This paper consists of a study of automated document classification, along with a practical attempt at document structuring and classification. The first chapter outlines some of the automated classification projects currently in use or being developed. The following chapters discuss an attempt at applying a document structuring and classification method to legal judgments.

The first chapter presents a review of some of the more innovative and successful projects in the area of automated document classification. The focus in selecting projects was on their potential application to a database of legal judgments, though none of these projects have actually been applied to legal judgments. Most research in this area uses business letters, news reports or technical papers for analysis, as these types of documents are most readily available in machine-readable form. They are also the types of documents for which there is the most demand for classification.

This paper is not meant to be a comprehensive overview of current classification research. The idea was to select a few good projects, focusing on those which have been implemented with some degree of success. Strictly theoretical papers were not considered. Though they need not necessarily be in use commercially, most systems chosen were in fact in use on a daily basis.

Evaluation of the success of a document categorization system has proven difficult. Often, a categorization subsystem is included as part of a larger text
retrieval system, and the only test results provided are for retrieval. There is sometimes no way of establishing the success of the categorization system in contributing to the overall retrieval results. For this study, things such as the sophistication of the method, the size of the system, and its commercial application were considered more significant indicators of a successful project than the precision and recall values in retrieval.

## II. Classification and Indexing

Lancaster sums up subject analysis as the presence, identification, and expression of subject matter in document texts, databases, controlled and natural languages, information requests, and search strategies ${ }^{1}$. There are many different terms for the process of organizing and sorting documents, each of which has a slightly different meaning. It is important to specify exactly what is meant by each of these terms at the outset, because many of them are used in slightly different contexts in the projects to be presented.

The myriad of tasks involved in document processing include classification, clustering, retrieval and indexing. Classification involves grouping documents with respect to a set of two or more predefined classes, usually of long term interest ${ }^{2}$. Clustering is similar, except that it does not require predefined classes. Related

[^0]documents are simply grouped into subsets.
Text retrieval can also be defined in terms of classification. It is merely an attempt to sort documents into two classes: Those the user would like to see at the particular moment, and those they would not ${ }^{3}$. Similarly, the difference between classification and retrieval is that in retrieval the classes are not predetermined, but are defined by the user at retrieval time ${ }^{4}$.

Ginsberg defines automatic indexing as the process by which thesaurus entries are automatically assigned to documents as content descriptors ${ }^{5}$. This is similar to classification, in that the indexes are predefined. The essential difference is that indexes tend to be narrower, and more numerous.

The concepts of classification and indexing are sufficiently related that both processes were included in this study. As mentioned above, indexing projects tended to use narrower categories and assigned more categories to each document. There were, however, some projects that used but a few, broad classes and still claimed to be doing indexing. The differences between the indexing and classification projects were found to be no greater than the differences between individual projects within the two categories. When referring to a specific project, I will use the same terminology (classification or indexing) that they employ, otherwise I will refer to all

## ${ }^{3}$ ibid.

${ }^{4}$ Hayes, P.J., Weinstein, S.P., "CONSTRUE-TIS: A System for Content Based Indexing of a Database of News Stories," Innovative Applications of AI 2, The AAAI Press/The MIT Press, Cambridge Ma, 1991, p. 54.
${ }^{5}$ Ginsberg, A., "A Unified Approach to Automatic Indexing and Information Retrieval", IEEE Expert Magazine, v8 n5, Sept. 1993, pp. 47.
subject analysis type processes as classification.

## III. Approaches to Classification

Early work in the area of document classification used frequency of words to find a list of keywords to represent or characterize a document ${ }^{6}$. Frequency was also used to indicate the keyword's degree of significance. More recent work uses sophisticated algorithms, expert systems and database structures in attempts to replicate the process of manual classification techniques.

The key feature of each of the systems covered in this paper is the way in which the classification knowledge is represented. This single component defines the classification method. Other important issues, such as disambiguation, use of thesauruses, semantic analysis and syntactic parsing are all equally important, but they are relatively independent of the knowledge representation.

The classification methods can be grouped into three streams:

1. Knowledge-based Representation
2. Case-based Representation
3. Statistical Representation

While these three categories reflect the major streams of research, they are by no means exhaustive. There were a number of papers discovered using methods which did not fit into any of these classifications, however none of them were found to be effective enough to merit a category of their own. A number of projects were

[^1]found which used the physical structure or layout of a document for classification. Some of these systems appeared to be quite effective, however they were not considered relevant to this study. Legal cases generally vary little in structure, and what differences there are provide few clues to case content. These systems are discussed briefly in a subsequent section. Similarly, projects which extracted document content using headings were also excluded.

The categories listed above are also not mutually exclusive. Most of the systems to be discussed here used aspects of two or sometimes all three of these methods to some extent. There was, however, a primary method that could be distinguished.

## IV. Knowledge-Based Representation

Knowledge-based classification is centred on the idea that knowledge is symbolic, and can somehow be represented or encoded ${ }^{7}$. A knowledge-based indexing system attempts an actual understanding of the content of a document, rather than simply analyzing and matching individual terms. This level of understanding varies widely, from simple textual structure, to indepth semantic knowledge.

Intellectual classification of document collections entails the assignment of

[^2]labels to a document to facilitate their location by subject ${ }^{8}$. In other words, the indexer must construct queries for which the document then becomes relevant ${ }^{9}$. A problem arises because the terms used to index these documents are selected from a thesaurus, so they do not necessarily match the actual terms in the document. This is why intellectual indexing requires concept matching, rather than word matching. In order for knowledge based systems to mirror this concept matching process, they must somehow encode the relationship between concepts and terms in the document.

To encode this relationship, knowledge-based indexing systems incorporate artificial intelligence techniques, typically in a rule-based format, for knowledge representation. The knowledge of an expert indexer is represented using a collection of manually created rules. Defining and creating these rules to work for general documents, or even a single domain, is the major task in knowledge-based indexing.

Most useful knowledge representation systems also require some level of semantic knowledge, often in the form of a dictionary. What separates these systems from a standard information retrieval system with a dictionary or thesaurus is the structure of the dictionary. It is not simply a list of terms, but contains knowledge about combinations of words and what they represent. Processing for such systems requires identifying these combinations of words in a document and acting

[^3]accordingly.
Knowledge-based indexing also generally involves some form of linguistic processing, where the syntactic structure of a document is analyzed ${ }^{\mathbf{1 0}}$. While actual linguistic understanding has proven difficult, current systems are able to effectively extract meaningful single and multi-word phrases. These phrases are further processed for such things as stemming, normalizing verb forms, and eliminating redundancy. The result is a structure of phrases in a standardized form that is highly representative of document content.

The major advantage to knowledge-based systems is the precision with which knowledge can be represented. Both case-based and statistical methods take a more generalized approach. They would have difficulty dealing with cases which are out of the ordinary. With a knowledge-based approach, specific rules can be encoded to deal with any given situation.

The ability to represent knowledge in this much detail, of course, has its price. The major drawback to this type of system is clearly the high cost of creation and maintenance of the knowledge base. While all systems will have associated startup costs, knowledge-based systems are particularly expensive in terms of time and labour because each possible scenario must be dealt with explicitly.

Another serious limitation to knowledge-based systems is that they are

[^4]generally useful only in narrow and carefully defined domains ${ }^{11}$. Not only are they unable to function outside their own limited domain, but they are also unable to recognize when a document is beyond their competence ${ }^{12}$. They are incapable of functioning in a more general domain because there is such an immense amount of background knowledge that goes far beyond the ability of any specialized knowledge base to consider, let alone encode.

## V. Knowledge-Based Representation Systems

## A. The WorldViews System

An example of an automatic indexing system which uses knowledge-based representation is the WorldViews system by Ginsberg ${ }^{13}$. The full system features information retrieval and a sophisticated user interface, in addition to automatic document indexing. It has been tested on electronic news articles, as well as abstracts of technical reports, but it is not currently a commercial system.

The automatic indexing subsystem in Worldviews accepts as input a thesaurus and a set of documents. The output consists of an updated version of the thesaurus where any entry that has been used in a document will now include a pointer to that

[^5]document. The subsystem then traces connections among concepts to draw out a documents implicit conceptual content. For example, a document with terms "wasp" and "mosquito", but not "insect" will still be indexed under "insect". The process also estimates the percentage of the documents content that is relevant to each descriptor.

The knowledge-base in Worldviews is in the form of a lattice-structured thesaurus. It consists of a structured set of subject headings or descriptors, with each thesaurus entry containing pointers to documents for which the entry is an index. Each entry also contains word sense information and alias terms for disambiguation purposes. The nodes in the dictionary are related via broader than (BT) and narrower than (NT) relations.

The Worldviews system uses a constrained form of activation spreading in a semantic network ${ }^{14}$ for matching entries in the dictionary. This process traces connections among concepts to draw out a document's implicit conceptual content based on the explicit concept references it contains. It is based on the idea that if a document contains an explicitly reference to a node, it also contains implicit references to the more general nodes above.

The two phases in the indexing process are as follows:

1. A list of concepts explicitly referenced by the document is generated, consisting of actual terms from the document, syntactic variations, or aliases.

[^6]2. An implied document sublattice is created using explicit concept references and implicit references. The implicit references are found by beginning with the explicit concept nodes and working up, tracking the number of times each implicit node is visited. All implicit nodes visited at least twice are included in the implied sublattice.

The terms in this implied document sublattice are then used for indexing. The totals for the number of times each implicit node is visited are also used as an indication of that term's relevance by the retrieval system.

There were no quantitative values given to measure the success of this system, however it has been tested against traditional keyword-based systems and has proven much more effective. There are plans to compare it with more sophisticated systems, such as the Smart system.

## B. The CONSTRUE Text Indexing System

The CONSTRUE Text Indexing System, by Hayes ${ }^{15}$, is the largest and most successful of the fully operational classification systems using knowledge-based techniques. It is a manually created rule-based system, essentially consisting of a large number of if-then rules. It has been in use by Reuters to classify news reports for online customers since 1989.

The CONSTRUE-TIS system is divided into two steps: Concept recognition and categorization. The concept recognition step uses a concept definition (a set of words and phrases which are indicative of a certain concept) to extract concepts from

[^7]a document. This definition is in the form of a pattern language, where concepts are defined as patterns of words and phrases in context. This allows a single pattern to match many different words and phrases. For example, phrases with gaps for a number of arbitrary words (similar to a proximity requirement), or word order requirements can be defined. This langauge also allows for negative rules, where phrases are rejected when they contain certain words. For example the word "gold" can be defined to indicate a story on mining, but not when accompanied by the word "jewelry".

Each pattern is also given a weight according to how indicative it is of the concept. The number of matches for each phrase is counted and used along with this weight to calculate a score for the strength of the appearance of a concept in a document.

The categorization rules are a set of if-then rules based on a boolean combination of concepts, the strength of these concepts, and their location in the document (eg. in the heading, body, etc.). This allows for flexibility in combining evidence from different concepts, as well as for encoding specialized knowledge related to a particular concept.

A category can be assigned on the bases of the concepts alone, it can be inferred based on the existence of other concepts, or it can be supported by weak evidence from related concepts. For example, the word "lead" on its own is ambiguous, and its presence alone is not sufficient to assign the commodity category "lead" to a document. When accompanied by concepts like "metals" (indicated by
"mining" and "ore") and general commodities (indicated by "metric tons"), however, it can be safely assigned.

The CONSTRUE-TIS system reports totals of $98 \%$ and $99 \%$ for precision and recall in terms of classifying news stories in the general categories of subject country. This drops to $94 \%$ and $96 \%$ for classification in four more general categories. A better indication of the effectiveness of this system can be inferred by the fact that it is in commercial operation.

## C. The JULLS Automated Keywording System

The Joint Universal Lessons Learned System (JULLS) is an automated processing program for documents pertaining to observations and lessons learned from military exercises, wargames and operations. The JULLS Automated Keywording System ${ }^{16}$ (JAKS) is used to index documents in the JULLS system. It uses a rule-base of "insertion" and "deletion" rules to transform a list of phrases selected from a document into a list of key phrases, which can then be used for indexing. The rules are manually created and maintained.

The first step in this system is to scan a document for recognizable phrases, which are matched from a list maintained by the system. A list of recognized phrases, along with their frequency, is passed on to the next step. Phrase position information within the document is also included, as some later rules have proximity or ordering

[^8]requirements.
The next step is to process the deletion rules. The purpose of these rules is to remove ambiguous phrases which are not indicative of content. There are two types of deletion rules: Conditionally weak, and always weak. Conditionally weak rules deal with situations where a term is only ambiguous in a certain context (ie. when combined with certain other terms). Rules that are always weak cover terms that are always ambiguous.

The final step is to process the insertion rules, which specify phrases implied by the existence of certain other phrases. The assumption here is that some phrases found in the text imply or trigger key phrases. There are two types of insertion rules: single and multiple. A multiple insertion rule implies a key phrase if a number of other key phrases are present. For example, if "time", "over" and "target" are all present within 2 words of each other, then the phrase "air warfare" is implied. A single insertion rule is the equivalent one to one relation. The remaining list of phrases from this final step are then used for indexing.

This system was evaluated by comparing its results with those of a manual indexer. In terms of recall, there was no difference between the two. In terms of precision, JAKS selected a higher proportion of incorrect key phrases, and thus had a lower precision.

While this method is simplistic, it seems to work well in the limited domain for which it was designed. The design does claim domain independence as a feature, however I believe that it would only be effective in domains such as military reports,
where indexing can revolve around a relatively small number of key phrases.

## D. The MedIndEx System

Computer assisted indexing differs from automatic indexing in that there is some human involvement at some point in the process. The difference may be slight (eg. human confirmation or random spot checks) or more significant (eg. the system simply provides suggestions or guidelines, or does error checking after the manual process). Although the focus of this paper is on automatic indexing, the two processes are sufficiently similar that a look at a sophisticated system for assisted indexing is relevant. Essentially, the only difference between automatic indexing and some types of assisted manual indexing is the degree to which the automatic results are relied upon.

The MedIndEx system ${ }^{17}$ is used in the indexing of medical literature to help manual indexers in selecting index terms and applying rules. The key aspect of this system for our purposes is the use of frames ${ }^{18}$ in the knowledge-base. A frame is a structure that names a collection of related concepts. A knowledge-based frame is similar to a thesaurus record, except that frames are linked by numerous specific relations, whereas a thesaurus is linked by only three relations: narrower, broader and related. Frames also have procedures (computer programs) associated with

[^9]them and stored within the frame.

The defining feature of a frame-based structure is inheritance. Inheritance provides links between frames which allow a child to inherit both data and procedures from a parent. This is a natural form for representing the hierarchical structure required for subject analysis.

In the MedIndEx system, frames are used to represent indexable entities. These entities are related using slots in the frames. For example, the concept "disease" has a slot called children, in which "neoplasms" (cancers) and "bone diseases" are values. Similarly "neoplasms" and "bone diseases" both have slots called inherits-from which contain the value "disease". With this structure, a procedure can be associated with the children slot which would then apply to all of the child nodes of the "disease" node.

If there were any exceptional diseases which needed to be dealt with differently, they would simply have a different associated procedure, and would not inherit the one from the parent "disease" node. To implement this in a rule-based system, each disease would have to be dealt with explicitly, with its own set of rules.

An indexing frame is an instance of a knowledge base frame which is linked to a specific document. The slots of indexing frames are filled interactively by manual indexers. The data inherited by these indexing frames is used as default values for these slots. The inherited procedures are used to restrict the possible values entered by the indexers.

To make this system fully automated would require encoding rules to convert
document terms into concepts and associate them as procedures with the concepts. Note again that each concept would not need to be explicitly dealt with, only those that are exceptional and cannot be dealt in procedures associated with their parent concept.

## VI. Case-Based Representation

The essential difference between knowledge-based systems and case-based systems is what is represented in the knowledge base. Knowledge-based systems attempt to replicate the process by which documents are manually indexed. Casebased systems are not particularly interested in the actual process, but simply try to replicate the results.

The defining feature of a case-based system is not the way the knowledge is structured, but rather the way it is created. The knowledge base is created by examining a large number of previously classified documents and attempting to determine which features of these documents were determinant in their classification. These features are then used to classify future documents.

The most significant advantage of case-based systems is the ease with which they can be created and maintained. Once the structure has been determined, the knowledge base is essentially created automatically by processing previously classified documents.

This also allows such systems to be largely domain independent. If the representation structure is sufficiently flexible, many different domains can be covered
simply by using different sample documents. This method is also suitable for broad domains, where there is no strong domain theory to support generalizations of the abstract structure derived from the features of a document ${ }^{19}$.

The biggest drawback in this method is the processing power necessary to create the knowledge base. The system by Masand ${ }^{20}$ (to be discussed below) required the resources of a massively parallel supercomputer to process the sample documents and create a knowledge base. Processing at this level is available to very few researchers.

Another potential problem is with the number of pre-classified documents required. It may not be feasible to manually create the huge corpus of sample documents necessary to create an accurate system of this type. Both of the systems outlined below were created using documents which had been manually classified for other purposes.

## VII. Case-Based Representation Systems

## A. A Memory Based Reasoning System

A system by Masand ${ }^{21}$ uses something called memory based reasoning

[^10](MBR) ${ }^{22}$, and a standard text retrieval system called Seeker to classify news stories.
The knowledge base for this system was created using a training database of 50,000 previously coded news stories from Dow Jones press. They were labelled with 350 possible codes in 7 categories.

MBR is basically a variation on the nearest neighbour technique ${ }^{23}$. It solves new tasks by looking up examples of similar tasks and using similarity with these remembered solutions to determine the new solution.

The most difficult task in creating this system was identifying features that allow simple and quantitative comparisons between documents. For this project, single words and capital word pairs were selected, largely because the retrieval system (SEEKER) supports this method.

The steps in MBR are as follows:

1. Find the near matches for each document to be classified. This is done by constructing a relevance feedback query out of the text of the document. This query is applied to SEEKER, a standard text retrieval system, and returns a weighted list of near matches.
2. Assign codes to the document by combining codes assigned to the nearest matches. Weights are also assigned by summing similarity scores for the near matches.
3. The best codes are chosen based on a score threshold.

As stated above, the big advantage of this type of system is its start-up time.
This particular system was created in about two person-months, compared with many

[^11]years for rule-based systems of comparable size. Furthermore, it will actually improve in performance as it increases in size and there are more examples for comparison.

This system was evaluated using n-way cross validation, where each text example is excluded one at a time and classification is performed on it. It achieved a recall of $83 \%$, with precision of $88 \%$ when tested in this manner.

## B. The PRISM System

Another application using case-based representation is the PRISM system ${ }^{24}$. This system has been in daily operation classifying telexes at Chase Manhattan Bank since 1989. It began as a rule-based system with approximately 700 rules, but evolved into its current case-based format because of the difficulties in maintaining and enhancing such a large rule-base. It works by retrieving cases similar to an incoming telex from its case library and using the classification of these telexes as the basis for a classification of the new telex.

The knowledge structure for this system is a directed acyclic graph, with each node consisting of a binary discrimination on the presence or absence of a feature of the case. The features were selected using a credit-assignment algorithm which evaluated the correlation between each feature and the variance in case outcome.

Using 4,000 previously classified telexes, a case library was constructed. Documents were initially represented as a list of individual terms. Indexes were

[^12]generated corresponding to terms which appeared to account for the variance in classification.

The current version of PRISM has three modules:

## 1. Lexical Pattern Matcher

Performs spelling correction, alias substitution and disambiguation tasks returning a set of hierarchically organized symbolic values.
2. CBR Module

The telex is classified by retrieving similar cases from the case library.
3. Router

Sends the telex to the appropriate destination. This module is independent to allow for easier installation at different sites.

The current commercial version of this system is used specifically to identify, classify and route letter-of-credit telexes, which it does with $90 \%$ accuracy.

## VIII. Statistical Representations

The final method of knowledge representation involves using statistical techniques to calculate a value for the similarity between documents. In general, these systems attempt to estimate the probability of correctly associating documents based on document descriptors.

Statistical representations have similar advantages and disadvantages as casebased representations in that they are relatively easy to create and maintain, and can cover a wide domain, but they require a high level of processing power. The difficulty in these systems comes in finding a good formula or algorithm to estimate document similarity.

## IX. Statistical Representation Systems

## A. The DR-LINK System

An example of a system using predominantly statistical techniques for categorization is the DR-LINK project ${ }^{25}$, a subsystem of DARPA'S TIPSTER project. The purpose of this system is to classify documents according to their subject matter and then use this classification to focus the search in the information retrieval stage. This preliminary step is necessary because the query matching step is computationally expensive, therefore it is beneficial to narrow its scope beforehand.

As mentioned previously, there were few projects that used strictly one method of knowledge representation and this one is no exception. Although the focus is on the use of statistical techniques in the matching process, a minimal knowledge structure, in the form of a structured dictionary, was still necessary.

Each word in the document is initially tagged with a subject field code (SFC) from a dictionary. This technique is similar to a controlled vocabulary, as it deals with plurals, synonyms and similar syntactic problems. These codes are summed and normalized so that each document is then represented as a vector of SFC's.

Assigning SFC's is done using "Longman's Dictionary of Contemporary
English" in a structured, machine readable form. This is a general, commercially available dictionary, containing headwords and various senses for over 35,000 English

[^13]langauge words.
The following steps are used to generate vectors to represent each document:

1. Run documents through POST, a probabilistic part of speech tagger ${ }^{26}$, which identifies the appropriate syntactic category of the terms based on the syntax of the neighbouring terms.
2. Retrieve SFC's from the dictionary for each term.
3. Term disabiguation is done using sentence level context heuristics. Single and most frequent SFC's are used as determining factors.
4. The end result is a vector of SFC's and their frequencies for each document, which are normalized for document length.

Document vectors are then clustered using Ward's agglomerative clustering algorithm ${ }^{27}$ to form classes in a document database. The clustering process is as follows:

1. Each document is considered a single cluster to begin the process.
2. A distance matrix is calculated between each document. The distance measure between documents is calculated using WARD'S least "loss of information" distance criteria, where lost information is measured by the error sum of squares. The idea is to join whichever pair of clusters results in the minimum in "within groups" variance.
3. The two clusters most similar according to this distance measurement are joined to form a single cluster.
4. The new cluster is represented by an average of the vectors of both documents.
5. A new distance matrix is calculated.
6. This process repeats until it is observed that dissimilar clusters are being joined. This occurred at 48 clusters during testing for this system.

Evaluation of this system was done by finding the precision at varying levels of
recall. At recall levels of $0.25,0.50,0.75$ and 1.00 , the corresponding precision was

[^14]$1.00,1.00,0.94$ and 0.65 respectively.

## X. Classification Based on Document Structure

The actual structure or layout of a document can often provide clues helpful in the classification process. While this information alone is not particularly useful, it can become meaningful when used in conjunction with other semantic or syntactic clues. For example, the location of names in a business letter can be used to determine which name belongs to the sender and which is the receiver. This type of information is particularly useful in office type environments, where many different document types are processed. Knowledge-based indexing is the logical choice for this type of analysis, as it allows structure definition for different types of documents.

Both of the following projects use document structure for classification, and are oriented toward an office environment. TEXPROS ${ }^{28}$ is a complete text processing system for office documents, including storage, extraction, classification, browsing and retrieval. Eirund's ${ }^{29}$ project is also an office system, but it is still at the prototype stage. Each of these systems use rule-based knowledge representation for all or part of their processing, and both use structure as the primary means to describe and classify documents.

[^15]The structure of a legal judgment does not provide many clues as to its content. The litigants, court and judge can provide some very basic information (eg. whether the area of law is criminal), but not much else. There are other legal documents, such as contracts or wills, where structure may be useful, however they are beyond the scope of this paper. Such documents would have more in common with standard office documents than with judgments.

## XI. Application to the Legal Domain

The most important feature of the legal domain, when it comes to classification, is the fact that it is narrow and well defined. This allows the builder of a classification system to focus specifically on legal terms, and how they can be used to extract classification information. Because of this feature, a rule-based approach to knowledge representation would be best suited.

Of the previously discussed projects, only two would be considered to operate in limited domains (Medlars on medical literature, and JULLS on military reports), and both of these systems are rule-based. Even the CONSTRUE-TIS system, a rulebased system which works in the broad domain of news stories, reduces this domain by limiting classification to the subject country in the story and a few other similarly broad categories.

Both the case-based and statistical systems attempt to classify documents in a much broader domain. This is to be expected, because as the domain grows broader, so too will the classification categories, and the "best guess" nature of both of these
methods will be more accurate. They are most effective in broad domains with broad categories.

Conversely, as the domain grows, it becomes less and less feasible to deal explicitly with every possible situation, which is necessary in a case-based system. This method is most effective in a narrow, well defined domain which requires distinguishing between relatively similar circumstances.

The domain of legal judgments has a largely consistent document structure featuring a limited lexicon of terms. Most cases tend to be similar not only in physical layout, but also in terms of content organization (facts, law, judgment). There is a relatively small lexicon of technical legal terms on which a classification system can potentially operate.

Most legal terms are only used in a limited context within the legal domain. This context can be quite broad (eg. "charged" in a criminal case, or "buyer" in a contracts case), or more narrow (eg. "break and enter", or "collateral contract"). In either case, the term communicates classification information to some extent.

The above examples also demonstrate the inherent hierarchical structure of legal terms in this domain. This would seem to indicate that a latticed-dictionary type structure would be best suited for representing classification knowledge. There are, however, many legal terms which are much more ambiguous. Terms such as "evidence" and "duty" do not narrow the scope of possible categories enough to be of any practical use. Furthermore, there are terms which are only slightly ambiguous within the legal domain (eg. "partner" or "estoppel"), but may be determined using
accompanying terms. These factors indicate that some type of rule-base is also necessary.

It is apparent that a classification system in a legal domain should include a structured dictionary along with an accompanying rule-base. There is an impression, however, gathered from this research and elsewhere, that rule-based systems are becoming outdated. As was discovered with the PRISM system, the cost of maintaining a rule-based system of any appreciable size is prohibitive.

A possible solution can be found with the use of frames. A hierarchical frame structure, with its accompanying procedures, can represent the same knowledge as a dictionary and rule-base, but do it much more efficiently. The inheritance features of this structure allow for general, global rules which may apply to a large section of the structure, but also specific, unique rules for special cases.

## Chapter 2: Automated Document Structuring

## I. Introduction

The ability to retrieve specific pieces of information quickly and efficiently from vast amounts of data is a primary requirement in legal research. With the improvements to information storage and processing brought about by advances in technology, there has been an explosion of information available to legal researchers. Most new cases published in Canada are now available in electronic form. Cases less than fifteen years old are also becoming available, and this coverage is continually expanding into the past. Other new sources of information, such as newspapers, magazines, journals and entire libraries, are coming on-line every day.

Despite the heavy requirements of legal researchers and rapid expansion of available information, there is little in the way of effective retrieval systems for legal research in Canada. Quicklaw, which is the main source for on-line cases, is both inefficient and outdated. Its effectiveness is questionable at best, as the simple boolean search strategy it employs often requires searching through many lines of irrelevant cases
where terms happen to match out of context.
There is a need for more effective methods of information retrieval in the legal domain, which will only grow more severe as the amount of information available continues to expand. The question that remains to be answered is whether legal judgments are sufficiently distinct from other types of text such that they can be processed in a way that will take advantage of these
distinctions.
There are some well recognized patterns in legal judgments. In the most common form, the judge will set out the facts of a case, discuss the applicable law, then apply the law to the given facts and render a decision. Legal text also features a limited lexicon of terms that is highly knowledge intensive, all within a single domain. Belew claims that "Lawyers tend to use more precise and consistent language than the average writer, and it can be argued that legal prose is therefore more amenable to computer analysis ${ }^{30}$." These factors suggest that there are unique features in legal discourse which can be exploited. By creating information retrieval systems designed specifically for legal judgments, the hope is that better and more efficient results can be achieved. Furthermore, systems may eventually be able to suggest legal strategies based on materials retrieved ${ }^{31}$.
A. The Method

The method chosen for this experiment in attempting to exploit the unique features of legal discourse is known as subtopic structuring. It involves imposing a substructure on cases, then using standard retrieval methods based on this substructure. A number of studies have shown that this method shows significantly

[^16]better results than processing entire documents alone ${ }^{32}$.
This research is based on two previous studies which employ a similar method in improving retrieval effectiveness over a standard full-text system ${ }^{33}$. The unifying feature of these two studies is that they both use subsections of the original pieces of text for processing and retrieval. The theory is that improvements can be made in retrieval effectiveness if text passages are split into smaller chunks which are then processed individually, as well as part of the text as a whole. The results of both Hearst's and Salton's work show that there are indeed benefits to be gained employing such a method. The focus of this research is on applying similar methods to the sphere of legal judgments.

While the work of Hearst and Salton focused on a comparison of full-text versus sub-text retrieval, this project will cover only the first step in the process: Defining a subtopic structure. In Salton's research this is a simple step, as he uses sentences, paragraphs and uniform size sections as multiple layers of substructures. The subsequent study by Hearst shows that using subsections based on something more meaningful, such as subtopics, achieves better results. This research will attempt to identify a meaningful subtopic structure in legal judgments.

[^17]The method used in defining this subtopic structure is similar to Hearst's "Textiling" approach, which "approximates the subtopic structure of a document by using patterns of lexical connectivity to find coherent subdiscussions ${ }^{134}$. With this method, text passages are initially divided into very small blocks. Each pair of adjacent blocks is compared, and assigned a value based on the similarity of content between the two blocks. These values are then graphed, resulting in a single line with peaks and valleys. The valleys, which denote the lowest points of similarity, are presumed to identify where the subtopic in the case has changed.

Using excerpts rather than full cases in the retrieval process should provide improvements in both precision and recall. Precision will be improved because large documents in which search terms appear evenly distributed will be ranked lower when processed as a series of text excerpts than as a full document. A sparse but even distribution suggests that the term may be used out of context. The result is that these documents will be ranked lower in the list of retrieved documents, or will not be retrieved at all. Recall will be improved in cases where large documents contain search terms which are concentrated in one excerpt. That one excerpt will be ranked higher than the whole document would have been, which means the excerpt may be retrieved when the entire document would not have been. A local concentration of search terms suggests a highly relevant piece of text.

A side effect of this method will hopefully be to reduce the volume of output

[^18]which the user must sift through. By allowing retrieval of the most relevant excerpts from a case, the user may be able to decide on its usefulness without having to read the entire case.

There are some other benefits that this process may provide, once a document is split into its subtopic structure. For instance, it may be possible to combine those sections of a document with the highest relevance values into a meaningful summary of the document. Note that this would not be a general summary, but would be specific to those parts of the document that are relevant to the query. Different queries would result in different summaries being generated for the same document. If this proves successful, it represents a substantial reduction in the amount of output the user would have to sift through to find what they want.

Another possible use for the subtopic structure of a document is to try to establish the types of sections present. By type, I am referring to the various kinds of paragraphs that are present in most legal judgments, such as factual paragraphs, legal paragraphs, decision paragraphs, etc. If the subtopic structure can be accurately established, then assigning the subtopic type becomes simply a matter of determining which types of terms appear most often in each section.

## B. Background

There have been a number of previous experiments using subsections of full
documents in attempting to improve retrieval effectiveness. An experiment by $\mathrm{Ro}^{35}$ compared full text retrieval with methods using a controlled vocabulary, text abstracts and paragraphs. The results showed that the highest recall was achieved using full text, however this method also produced the lowest precision. The results were also limited by the fact that a boolean retrieval method was employed.

The work of Salton and Buckley ${ }^{36}$ involved breaking documents up into paragraphs and then attempting to find similar paragraphs (for use in hypertext links, for example). They found that the best results were achieved when comparing paragraphs at both the paragraph and sentence level. Though the focus here was on finding similarity within a single document, the results are relevant because they found that the most effective method was to look at both the local and overall similarity.

Experiments by Stanfill and Waltz ${ }^{37}$ compared retrieval on even sized pieces versus full text. They found that the even size pieces where more effective in terms of both precision and recall. Subsequent experiments by Hearst show that using motivated segments, which reflect the substructure of the text, work even better.

[^19]II. The Method
A. Outline

As mentioned above, the approach taken in finding a subtopic structure is similar to that employed by Hearst. The basic method is to break the text down into initial blocks and then delimit subsections where adjacent blocks show a low similarity. This similarity is based on matching elements in the adjacent blocks.

The case is initially parsed into terms made up of single words and phrases, which are then loaded into a database. The list of terms for each case is divided into blocks and processed once again, this time comparing adjacent blocks. A similarity score based on the number of matching terms is calculated for each pair of adjacent blocks. The resulting list of similarity scores is then smoothed and graphed, resulting in a line with peaks and valleys. The subtopic boundaries are chosen at the lowest point of the valleys on the graph.

## B. Parsing The Text

The words and phrases were extracted from the text using a parser specifically designed for legal judgments as part of the FLEXICON ${ }^{38}$ project. This parser not only picks out single words, but also multiple word phrases. This is particularly important when dealing with legal terms, which can have very different meanings depending on surrounding words.

[^20]The input to the parsing program is the text of a legal judgment. It produces a list of terms with their location from the beginning of the text (in bytes), as well as a list of the paragraph breaks (also in bytes). Other information is also produced by this program, however it is not needed in this process and is ignored.

The parser works by first extracting statute and case references using templates, along with some basic rules. Legal terms are then extracted by matching with a legal dictionary. Finally, noise words are removed, leaving what are assumed to be factual terms. Although the FLEXICON parser extracts and identifies four different types of terms (cases, statutes, legal phrases and facts), they are all treated similarly for the purpose of finding subtopic structure.
C. The Database

The database is in a simple B-tree format. While this format worked well for the limited number of cases required for testing purposes, it is not recommended for large scale use. It was chosen because it was relatively simple to implement. It was not particularly efficient, and would be impractical on a larger project. Each term in a case was given a separate entry in the database, consisting of the term itself, a reference to the case in which it appeared, its location in that case, and its paragraph number.

All of the terms in the lexicon were required to be loaded into the database before any matching could be done, because the similarity calculation requires values for the frequency of each term and the total number of terms. Once the database is
loaded, each text passage in the collection is processed again, this time calculating similarity values for each pair of adjacent blocks.

## D. The Initial Blocks

The most difficult problem with this method comes in deciding on the size of the initial blocks. Hearst used the average paragraph size for his research, although he admits that "the block size that best matches the human judgment data is sometimes one sentence greater of fewer. ${ }^{139}$.

In initial tests, I experimented with a number of different possibilities, including average paragraph size, actual paragraphs, and many different line count values. Average paragraph size was judged to be too small to be of any practical use. Legal judgments, unlike general text, contain a relatively high number of very short paragraphs, which greatly reduced the average paragraph size. This resulted in fewer terms being compared for matches, and consequently fewer matches found. In many instances, a similarity score of zero between two blocks of text was more a reflection of few terms being compared rather than any change of topic. For this reason, average paragraph size was discarded as being too small.

There is also a potential problem in making the initial blocks too large. If they were set to range over a number of paragraphs, it would be likely that the actual change in subtopic would occur somewhere in the middle of the block. This would result in terms from different subtopics in the same block. The terms at the

[^21]beginning of the block would match well with the terms from the previous subtopic, and those at the end would match well with the next subtopic. Rather than getting a low similarity value in the block directly before the change in subtopic, the result would be slightly lower values for the blocks both before and after. Thus, the actual boundary would be obscured.

Many experiments were done with a fixed sized initial block based on a set number of lines, anywhere from 15 to 40 . While there was some limited success with this method, there was no single value that worked well for most cases. Nor did there seem to be any correlation between the size that worked well and the average paragraph size.

In the end, the best results were found using actual paragraphs as the initial blocks. This guarantees that the subtopic boundaries will not fall in the middle of the block, assuming subtopics do not change mid-paragraph. The previously mentioned problem with very short paragraphs was dealt with by discarding those paragraphs with less than a minimum number of terms.

## E. Comparing Adjacent Blocks

The process of comparing adjacent blocks of text to determine their similarity was done using a variation of the vector-space model ${ }^{40}$. In this method, blocks are represented as vectors in a multidimensional space, the dimensions of which are the

[^22]terms selected from the text. Each of these terms is assigned a weight based on the statistical distribution of the term in the database. The similarity of each set of adjacent blocks is then determined by comparing the two vectors.

The main advantage of this method is that it assigns different weights to particular terms depending on their perceived importance to the case. This importance is a measure of when and how often the terms appear in the lexicon. Precedence is given to terms which occur infrequently over the entire lexicon, but are clustered together when they do appear. This method is based on the concept of "locality"41, which says that if a set of references to a particular concept occur in close proximity, this is a good indicator of topicality.

## F. Term Weights

The terms in each block are first assigned a weight (W) based on the relative frequency of the term in the block and in the rest of the database:

$$
w=f_{b} \times \log \left(\frac{D}{d_{t}}\right)
$$

where $f_{b}$ is the frequency of the term in that block, $D$ is the total number of blocks in the collection, and $d_{t}$ is the number of blocks which contain the term.

This function results in higher weights for terms which appear frequently in one block and infrequently in the rest of the text. The assumption is that such terms are good content indicators.

[^23]The similarity score (S) for an adjacent pair of blocks $\left(B_{i}, B_{j}\right)$ is then:

$$
S\left(B_{i}, B_{j}\right)=\sum_{i, j=1}^{t}(W i \times W j)
$$

where there are $t$ distinct terms available.
Because of the significant difference in the size of some paragraphs, and consequently the number of terms being compared, it was necessary to normalize the similarity scores. This was done using the cosine coefficient, where:

This function allows blocks of different sizes to be compared accurately.

A small variation on this method was added to make the calculation a better reflection of actual similarity, rather than simply a measure of the number of terms being compared. Despite the fact that very small blocks (less than five terms) were discarded, there were still many instances where no matches between two relatively smaller adjacent blocks were found. This was often due more to the small number of terms being compared than a change in subject.

The solution to this problem was to compare each block to a window of three following blocks. This increased the number of terms available for comparison, which
reduced the chance of no matches being found. This addition should not reduce the effectiveness of the similarity calculation because terms from one subtopic section would be no more likely to match those in the second or third block in the next section than the first block. Therefore, there should still be a relatively low similarity score for blocks in different sections. The only drawback to using a window of blocks for comparison is that it also increases the chance of a coincidental match of terms which do not cohere. This increase, however, would be uniform throughout the entire text passage and so should not affect the results.

As stated previously, those paragraphs which contained less than five terms, none of which matched with the subsequent three paragraphs, were ignored.

Finally, a simple smoothing algorithm (an average calculation with a window of three blocks) was applied to account for small local minima.

## G. Finding Subtopic Boundaries

The final step in this process is to graph the resulting sequence of similarity values. High similarity values will form peaks in the graph, while low similarity values, indicating a change in subtopic, will form valleys.

The smoothing step resulted in a single value for each set of three blocks, so it is now necessary to determine between which of these blocks the boundary should fall. The simple solution would be to find the block with the lowest score and place the boundary after this block. While this would work adequately, it will not always be correct. In theory, the last three blocks preceding the boundary should all have lower
similarity scores, with the last block being the lowest. The following block, the first in the next subtopic section, should then be a much higher value. Rather than assume that the last of the three blocks will actually be the lowest, it would be more effective to use the increase as the boundary indicator. Therefore, the actual boundary should be placed between those blocks for which there is the largest increase in similarity score.

One last point to note here is that the lines in the graph do not always form a clear valley with a distinct lowest point. In some cases, the valley bottom is rounded over a number of blocks. When this occurs, the same process described above is done (finding the largest increase in similarity score), but it is done over all of the blocks which form the lowest part of the valley.

## III. The Results

## A. Overall Results

A total of fifty cases were used to form the lexicon for this test. They were randomly selected to represent a cross section of typical Canadian case law. Thus, they included various court levels, jurisdictions, and areas of law. There was also an attempt made to select a wide range of case sizes, although very short cases were excluded. ${ }^{42}$

In order to evaluate how well this process is able to identify the subtopic

[^24]structure in a case, there must be some "correct" answer to which the results can be compared. For this purpose, each of the cases in the lexicon were manually broken down into sections based on changes in subtopic. While this is admittedly a largely subjective exercise, the subtopics were chosen at a very broad level, where changes were obvious. Real judgment was only required when selecting the exact location of a boundary within a few paragraphs. The effect of this problem was further reduced in the smoothing process, where every three blocks was averaged into one.

Once the manually selected topic boundaries were chosen, evaluation was simply a matter of comparing how often the process found these same boundaries. The window in which the process boundary must occur in order to be considered correct is within three paragraphs of the manually selected boundary, the same size as the smoothing window.

Of the 102 manually selected subtopic boundaries in the lexicon of cases, 43 were successfully found by this process, for an accuracy rating of $42 \%$. There were also 71 extra subtopic boundaries found which were incorrect, making $38 \%$ of the boundaries found correct.

While there was no real threshold to be met for this process to be considered successful, the results achieved are clearly less than impressive. With fewer than half of the subtopic boundaries identified, and close to two-thirds of those identified incorrect, this process is clearly not able to distinguish any sort of accurate substructure from the collection of legal cases.

It is difficult to compare the results of this experiment with those found by

Hearst because the initial substructuring step in Hearst's research is never really evaluated. He simply accepts whatever sections are found by the Textiling process and uses these sections in the retrieval stage. There is no indication of how well these sections reflect the actual substructure of the text.

There is one example given, a graph of a single document overlaid with manually selected boundaries. In this example, there are three good boundaries, a single poor one, and three that may or may not be considered correct depending on how restrictive the criteria is. There is no indication as to whether this example is a typical one.

## B. Analysis of Results

## 1. Overview

The problems with using this method for finding substructure in legal judgments can be attributed to two causes: Those resulting from the method itself; and those resulting from working with the particular lexicon of legal judgments. Problems with the method revolve around the fact that it was not very effective in distinguishing between matching terms which were important to the substructure and those which were not. The vector space model is designed to emphasize terms which occur clustered in one section of text and appear rarely elsewhere in the lexicon. The results show that this effect was accomplished, but the difference in score between clustered and general terms was not large enough. When blocks as small as a single paragraph are being used for comparison, any match is significant. In most cases, the
boundaries found using the vector space model occur where there were no term matches at all, rather than where there were a few insignificant matches. In fact, the results show that there were actually very few insignificant matches.

## 2. Problems With the Method

a. General Terms

There were many instances where general terms that were spread throughout a case affected the boundaries chosen by the process. These terms are not good indicators of subtopic structure because they relate to the case as a whole. Their presence or absence in any part of the text gives no indication as to a subdiscussion. These terms are more likely to be matched simply because they occur more often, but these matches should have less influence on the selected boundaries than matches between rare terms. It appears, however, that these terms exerted a significant influence.

Re Adoption Act Chapter 4 R.S.B.C. $1979^{43}$, for example, is an adoption application in which the term "child" occurs frequently throughout the case. Although this term would not be a good indicator for the purposes of subtopic structure, it clearly had an effect on where the boundaries were placed. The multiple appearances of this term in both paragraph 14 and 15 results in a similarity score sufficiently large enough to miss the boundary between these two paragraphs, despite the fact that it is the only term that matches.

[^25]Similarly in Budai v. Ontario Lottery Corporation ${ }^{44}$, a tort case involving a winning lottery ticket, there is a subtopic break between paragraphs three and four which was not found. This border separates the facts of the case from the beginning of the decision. It was not found by the process because terms like "won", "winner" and "prize" in paragraph three are matched in later paragraphs, resulting in a high similarity score and no boundary. These terms appear throughout the case and offer no indication as to subtopic structure.

The same situation occurs again in Macneill Industrial Inc. v. John Posnikoff ${ }^{45}$, where a boundary between paragraphs four and five is missed because the term "option agreement" matches in paragraph four. This term occurs in 17 of the 35 paragraphs in the case.

In general, terms which are common to a specific case, yet rare in the entire lexicon, are factual terms. This is understandable, as factual terms would be expected to be specific to a single case, and the more important factual terms would be expected to appear repeatedly spread over a case. Legal terms which are spread over a single case would be expected to appear over the rest of the lexicon as well.

For this reason, a possible solution to the above problem would be to simply eliminate all of the factual terms from the process, leaving only case references,

[^26]statutes and legal terms.
This would eliminate the majority of the general terms, leaving the substructure to be formed based on legal subdiscussions, indicated by clusters of common legal terms. The problem with this solution is that once the facts are removed, very few terms will be left for analysis. There is already a problem with similarity scores of zero between adjacent blocks resulting from few terms being compared. This would only get worse once the facts are removed.

This is in fact what happened in trials with the facts removed. There were simply too few terms left to make any sort of meaningful comparison.

## b. Generic Legal Terms

A related problem with this method is its handling of generic legal terms. While general factual terms appear evenly spread throughout a single case, generic legal terms are also common throughout the lexicon. They include words like "evidence", or "duty", which can be used in many types of cases and in many contexts. The offer few real clues as to subtopic structure.

Again, the vector space model as it is implemented here does not do enough to reduce the effect of these terms. They should not have a significant influence on where the subtopic boundaries are placed because they provide no meaningful information. The presence of a cluster of these terms is not an indication of a subdiscussion. It is merely the natural result of a discourse in which certain fundamental terms get a great deal of usage. Unfortunately, they did have an effect
on the subtopic boundaries chosen.
The Central B.C. Planers Ltd. v. Hocker et al. ${ }^{46}$ case, for example, shows a change in topic between paragraphs 10 and 11, where the judge moves from a discussion of the facts of the case to the applicable law. This boundary is not found because the term "evidence" is found in paragraphs before and after. Although it is a very common term and would not be considered a good indicator of content, it is enough to cause the break to be missed.

Similarly, in Burman's Beauty Supplies Ltd. v. Kempster ${ }^{47}$, the border between paragraphs 7 and 8 is missed because of matches on terms like "mortgages", "chattel mortgage" and "solicitor" which appear throughout the lexicon.

A possible solution to this problem would be to increase the size of the initial blocks. This would increase the number of terms being compared, and consequently the number of matches. With more matching terms, the process would be better able to distinguish between matches that are significant (that is, involving a cluster of terms that are otherwise rare) and those that are not. With so few terms being used for comparison, there are no insignificant matches.

As stated previously, basic tests done with larger (fixed size) blocks actually had worse results, as the manually selected section breaks became obscured when they fell in the middle of the initial blocks.

[^27]
## c. The Parser

Another problem with this process is due to the parsing program. The parser was designed not just to simply split sentences into words, but also combine words into single unit phrases where appropriate. While this is important in order to give proper meaning to groups of words, it caused some difficulties for the purposes of this project. The problem was that the use of phrases sometimes made the meaning too precise, such that terms which were generally but not exactly the same did not match. This high level of precision was too much for the purposes of this process. Examples of this can be found in Dixon v. Bank of Nova Scotia et al. ${ }^{48}$, where an incorrect subtopic boundary between paragraphs seven and eight is identified because no matches were found in paragraph seven. Phrases like "purchase of shares", "acquired shares", and simply "shares" are found in these two paragraphs, but none of them were considered matches. Similarly in Re Adoption Act Chapter 4 R.S.B.C $1979^{49}$, an incorrect boundary was found between paragraphs 24 and 25 because there were no matches found between any of "legislation", "gap", "legislature" or "legislative gap test".
3. Problems With the Lexicon

Not all of the problems associated with this research can be attributed to the

[^28]process. Some are clearly attributable to difficulties in dealing with the unique lexicon of legal discourse. When this research began, the hope was that the peculiarities of legal judgments could be exploited to allow more effective analysis than would be possible with plain text. While I believe that this is still possible, it appears that this process is not the one best suited for the job.

## a. Cases With No Structure

One of the problems particular to this lexicon is that some judgments simply do not have a clearly defined substructure. There were many cases in the test group which did not follow the standard facts, law, decision pattern. There were some cases where facts, law and sometimes multiple issues were extensively intermingled, making it difficult to identify a structure even manually. Such cases will never be successfully analyzed by this type of method.
R. v. Kamal Baig ${ }^{50}$ is a good example of this type of case. It is long, with a number of different issues and no real separation of fact and law. It also contains some unusual text (a resume, a business letter and the text of a cross examination) which this process did not handle well. None of the manually selected section borders were found, and those that were assigned seemed to be completely random. An examination of the similarity scores shows that there is a significant increase in value from the paragraph before to the one after the manually selected boundaries, as expected ( 7 of the 9 boundaries showed a meaningful increase). There were,

[^29]however, also many larger increases at other places in the text. This prevented any real pattern from forming in the list of similarity scores, resulting in a seemingly random selection of boundaries.

G \& C Collision Repairs Ltd. v. United Buy \& Sell Service Inc. ${ }^{51}$ is another such case, with five different issues spread throughout the judgment and quotations of dialogue as evidence. This case was not handled well either.

## b. Short Paragraphs

Another problem in working with legal judgments that was never handled completely was dealing with extremely short paragraphs. Paragraphs of two or three lines are much more common in legal judgments than in general text. Despite the measures taken to negate their effect, these paragraphs still had an influence.

The essential problem was that a similarity score of zero cannot be normalized to account for a small number of terms being compared. In many instances, a zero score was more a result of small paragraphs being compared than any lack of coherence.

This was dealt with somewhat by skipping paragraphs which had a similarity score of zero and five terms or fewer, but the small paragraphs continued to have an effect. They were still included in the three paragraph sliding window used for comparison as each paragraph is processed. This meant that the paragraph

[^30]immediately before a group of very short paragraphs would have few terms with which to match.

This was demonstrated in Lawrence Nesis v. Benter Investments Ltd. ${ }^{52}$, where an incorrect section boundary was identified between paragraphs 13 and 14. The boundary was placed there because no term matches were found for paragraph 13 , which was a result of the fact that the next three paragraphs together totalled only six lines.

## 4. Other Results

Despite the overall results, this method of finding substructure did work well for some cases. This was limited, however, to relatively short cases in which a single issue was being decided, and which strictly followed the standard facts, law, decision pattern. The case of Roche Lake Developments v. Urban Systems Ltd. ${ }^{53}$ fits this pattern, and it was handled well by the process. It begins with a chronology of the facts, a discussion of the legal precedents involved, and finally, the decision. There is a distinct point here at which the discussion of facts ends and the analysis of the law begins.

One of the unexpected results that was not reflected in the overall

[^31]performance evaluation in this experiment is the fact that some of the incorrect boundaries found by the process may arguably be correct. As stated previously, the initial manual sectioning was a subjective process, though tempered by the fact that the changes in topic were very general and usually obvious (within a paragraph or two). On examination, there were some places where a manual section boundary could well have been added. In Chand and Chand v. Sabo Bros. Realty Ltd. ${ }^{54}$, for example, there is a seemingly incorrect boundary found between paragraph twelve and thirteen. The case begins with a statement of facts, then a single law paragraph, then moves on to other facts unrelated to the initial ones. This could easily have been chosen as a boundary in the manual sectioning process. While there were a few instances like the above, it is important to note that they were rare, and certainly would not have an impact on the overall success rate.

## IV. Alternative Methods

When the initial results obtained in this experiment proved unsuccessful, a number of alternative methods were explored in attempts to improve the results. While none of these alternatives produced any significant improvement, it may be useful to outline some of them here in order to provide guidance in future research. There are also suggestions for variations which were not attempted.

The biggest variable in this method is the size of the initial blocks. Trials were

[^32]done using the average paragraph size as the initial block (as in Hearst's method) as well as many different set sizes, however the best results were attained using actual paragraphs. The problem with using average paragraph size was that with the number of very short paragraphs usually found in legal judgments, the average size was too small. A small block size means fewer terms available for comparison, resulting in a zero similarity score. This score is not a reflection of actual similarity, but only the few number of terms being compared.

A possible solution to this problem would be to discard the very short paragraphs before this average is calculated. Another possibility would be to use the actual average paragraph size but increase the window of following paragraphs used for comparison. This would increase the number of terms being compared. Neither of these variations were attempted.

The real problem with this method is that it does not sufficiently distinguish between terms which are good indicators of the substructure and those that are not. There are a number of things which can be done towards remedying this problem. Since it is usually factual terms which are repeated throughout a case, they could be eliminated, leaving only legal terms. This was attempted, but was not successful because there were too few terms left in each block once the facts where removed. Perhaps this method, in combination with some other variation which increases the number of terms being compared (such as increasing the comparison window) may be more effective.

Another possible solution would be to discard terms which appear in more
than a certain ratio of paragraphs spread evenly over a case. Based on where and how often they occur, certain terms could be eliminated as poor topic indicators. Removing these terms would reduce the number of matches occurring over subtopic boundaries, thus making them more visible.

An attempt was made to reduce the effect of general factual terms by changing the vector space function slightly. These terms offer no indication of substructure, yet they often obscured it when they matched over boundaries. The vector space function was changed to use only the blocks in a single case as a basis for calculation rather than all of the blocks in the lexicon. This should have had the effect of reducing the influence of factual terms which occurred frequently in a single case, regardless of how often they occurred in the entire lexicon. The problem with this variation is that it removes the frequency of a term in the lexicon as a factor, which makes the problem of generic legal terms even worse. The generic legal terms are usually rare in a single case, but frequent over the entire lexicon. If the similarity calculation is based on each case individually, these terms become much more significant. This solution was attempted without any noticeable improvement, however it may be more effective if used in conjunction with another variation.

A final variation in the method would be to change the parsing program so that terms are reduced to a single word. There were many instances in the cases where similar phrases did not match because of an extra word or a partial change. This would also make more terms available for comparison.

An alternative would be to change the requirements for two terms to be
considered a match. The current method was fairly basic, requiring the first $3 / 4$ of the letters in both terms to match. This ratio could be reduced, or the criteria could be changed so that any common words in the two terms would be considered a match. Either of these changes would increase the number of matches found, although they would also increase the potential for incorrect matches.

## Chapter 3: Finding Legal Categories

## I. Overview

In addition to identifying the subtopic structure of a case, an attempt was made to classify each of the subtopics in terms of a legal category. Initially, identified subsections were simply divided into factual and legal types, where possible. There were some sections which could not be definitively placed in either category, but there were still enough to make this subdivision useful. Since this classification was done using a dictionary of legal terms, there was no attempt to further classify factual subsections, however legal discussions could be further broken done into particular areas of law.

This type of classification could bring about improvements in both retrieval effectiveness and user efficiency. It would allow a user to target, for example, only the factual sections of cases, which would greatly reduce the body of text to be searched. It would also allow searches for legal terms used in a factual context, which would not otherwise be possible.

The ability to target a search based on a certain area of law would be even more useful. Again, the body of text would be reduced substantially, improving the retrieval speed. More significantly, cases which are outside the specified area of law, but which may contain some of the search terms, would not be included. This allows the user to limit search terms to a specific legal context. This ability would be particularly useful when searching for issues that are often minor in cases with multiple issues, such as procedural or evidentiary questions.

Essentially, what this process does is to execute part of the search process in advance. It classifies sections of cases based on some general legal categories, then allows the user to select which of these categories is applicable. Of course, if more general queries were required, searches could be done covering all of the categories.

There are also benefits in terms of the output to the user from classifying substructure. It could be limited to just the facts of the cases retrieved, or only those subsections which cover a certain area of law. Even being able to label the output as to the type of discussion it contains would be helpful.

## II. The Method

The method used to classify subsections of text is based on looking up terms in a legal dictionary. The dictionary is actually a hierarchical list of legal terms, starting with general terms at the top, and becoming more specific as the branches descend. Each legal term has a corresponding general area of law near the top of the hierarchy, which is used to assign categories. The category for a particular block is based on the category found for the most terms in that block. Similarly, a subsection category is based on the most common category among the blocks in that subsection.

The first step in this process is to distinguish between factual and other blocks. As mentioned in the discussion of the original process, the parsing program separates factual, legal, case and statute type terms. The ratio of factual terms to other terms is then used to determine the block type.

A ratio of $3 / 4$ or fewer factual terms to other terms is enough to assume that
the block contains a legal discussion. Additionally, any block which contains a statute or case cite is also considered a legal block. Once a block is identified as legal, the dictionary can be used for further classification.

The dictionary is in hierarchical form, starting with general legal terms at the top, and becoming more specific as the branches descend. This structure provides a simple method for classifying terms, as each parent node provides a more general category of law. The example below shows the vertical path for the term "wrongful dismissal":
wrongful dismissal -> employment law -> contract $->$ private law

Since the top level (private law, public law, etc.) would be too general to provide any useful information, the second to top level (contract, tort, criminal, etc.) is used for categorization. This level is specific enough to offer useful information, yet broad enough to allow terms to be grouped together.

Legal terms, as they were processed for subtopic structure, were also looked up in the dictionary to determine their category type. If a term was found in the dictionary, the general category (at level two in the hierarchy) corresponding to the specific category for the term was retained. The subject for a legal block was determined by the most frequent general category, with the requirement that it occur at least twice. Similarly, the subject for a subsection was determined by the most common category among the blocks in that subsection, again with the requirement
that the subject must appear at least twice.
III. The Results

Because of the highly subjective nature of categorizing subsections, the success or failure of this procedure was difficult to evaluate. Selecting the single best subject is not always possible, even when done manually. To alleviate this problem somewhat, the categories were kept general, with most of them being well defined areas of law, such as tort, or contract. There were also a few more specific categories, such as wills and estates, or banking.

One of the consequences of using such general categories was that the assigned subjects for each subsection in a single case were usually the same. This is too be expected, as the majority of cases involve only a single area of law.

The results of the categorization process were evaluated by simply reading each subtopic section and deciding whether or not the category selected was accurate. Of the 71 subsections labelled by this process, 42 were considered correct, for an accuracy rating of $59 \%$.

## IV. Alternative Substructuring Method

One of the observations made when examining the blocks types for the purpose of finding section subjects was that there tended to be large sections of either factual or legal blocks, and these sections often corresponded to the manually assigned substructure. This is not all that surprising considering that in many cases,
the substructure turned out to be a basic division between the facts, law and decision sections. With this in mind, an alternate method of determining subtopic structure was attempted based on block types.

In this method, blocks were labelled as either legal or factual in the same way as the original process, based on the types of terms in the block. This list of block types was then "smoothed" by forming groups of three blocks and assigning to them the majority type. This helped to eliminate any mislabelled blocks, or single blocks that appear out of place.

The only additional step was to select the exact boundary, within the three "smoothed" blocks, so that it matched more closely with the change in paragraph type. This was done by placing the boundary at the point where the change took place in the original list of paragraph types, before they were "smoothed".
V. Results of Alternative Substructuring Method

The evaluation for this method was the same as for the original, except for the type of boundaries expected to be found. Since this new method distinguishes only between factual and legal discussions, it could not be expected to find, for example, multiple issues. Therefore, the only substructure expected was boundaries between factual and legal discussions. Consequently, the manually selected subsection boundaries only represent this type of structure.

This method of defining subtopic structure showed some improvement over the original method, though it was far from being considered a success. It found 39 of
the 85 manually selected subtopic boundaries for a success rate of $46 \%$, which is slightly higher than the original at $42 \%$. Where is showed significant improvement was in not assigning as many incorrect boundaries. This process had only 39 incorrect boundaries, meaning $52 \%$ of the boundaries chosen were correct, compared with $38 \%$ for the original method.

Part of the explanation as to why this method worked somewhat more effectively can be attributed to the fact that it does not attempt to do as much. The only goal was to find boundaries between factual and legal sections of text, as opposed to identifying any type of subdiscussion, which the original method was attempting to do.

## Chapter 4: Conclusion

The initial goal of this research was to use the distinctive elements of legal judgments to improve retrieval effectiveness on legal databases. This was to be done by identifying a substructure within a judgment, and then using standard retrieval techniques based on this substructure in addition to the text as a whole. Previous studies have shown that retrieval based on some subdivision of full text documents does indeed show better results. The problem is in identifying this initial substructure.

While I believe that most legal judgments do contain subdiscussions that can be identified and delimited, the method used in this experiment is clearly not the one to do it. It is questionable whether any method based on term analysis would be successful in identifying a plausible substructure. The problem is that there are too many terms, both legal and factual, which appear throughout a case and which tend to obscure the subtopic boundaries.

A single judgment is a unit unto itself, involving (usually) two parties in a dispute covered by a certain area of law. All of the terms which relate to these general attributes of a case serve to make identifying subdiscussions within the case more difficult.

Perhaps a better alternative would be to attempt to split a judgment into specific subsections, such as fact and law, as was done in the second experiment. While this provides less detailed information, it is much more likely to be implemented successfully. The experiment done as part of this project was very simple in its analysis, and could definitely be improved with a more sophisticated
method and further testing. It may also be possible to expand the types of subsections identified to include such section categories as "decision" or "damages". Although it is a less ambitious project, it shows much more promise.

## Bibliography

## Chapter 1

Biebricher, P., Fuhr, N., Lustig, G., Schwantner, M., Knorz, G., "The automatic indexing system AIR/PHYS-from research to application," 11th International Conference on Research and Development in Information Retrieval, ACM New York, June 1988, pp. 333-342.

Blosseville, M.J., Hebrail, G., Monteil, M.G., Penot, N., "Automatic document classification: natural language processing, statistical analysis, and expert system techniques used together," SIGIR Forum spec. iss., 1992, pp. 51-58.

Crouch C., "A Cluster Based Approach to Thesaurus Construction", SIGIR Forum spec. iss., 1988, pp. 309-320.

Driscoll, J., Rajala, D. Shaffer, W., "The operation and performance of an artificially intelligent keywording system", Information Processing and Management, v27, n1, 1991, pp. 43-54.

Eirund, H., "Knowledge based document classification supporting content based retrieval and mail distribution", Network Information Processing Systems. Proceedings of the IFIP TC6/TC8 Open Symposium, May 1998, pp. 309-20.

Evans, D.A., Ginther-Webster, K., Hart, M., Lefferts, R.G., Monarch, I.A., "Automatic indexing using selective NLP and first-order thesauri", Intelligent Text and Image Handling. Proceedings of a Conference. RIAO '91, April 1991, pp. 624-643.

Farkas, J., "Neural Networks and Document Classification", Electrical and Computer Engineering, 1993, p. 1

Farkas, J., "IndeXpert: an intelligent indexing system with geometric document ranking", Twelfth International Conference. Artificial Intelligence, Expert Systems, Natural Language, v3, June 1992, pp. 139-149.

Field, B., "Towards Automatic Indexing: Automatic Assignment of Controlled Language Indexing and Classification from Free Indexing", Journal of Documentation, v31, n4, Dec. 1975, pp. 246-265.

Frants, Valery I, "One Approach to Classification of Users and Automatic Clustering of Documents", Information Processing and Management, v29, n2, 1993, pp. 187.

Fuhr, N., "Models for retrieval with probabilistic indexing", Information Processing and Management, v25 n1, 1989, pp. 55-72.

Fuhr, N., Buckley, C., "Probabilistic document indexing from relevance feedback data", Proceedings of the 13th International Conference on Research and Development in Information Retrieval, ACM, New York, Sept. 1990, pp. 45-62.

Gibb, F., "Knowledge-based indexing in SIMPR: integration of natural language processing and principles of subject analysis in an automated indexing system", Journal of Document and Text Management, v1, n2, 1993, pp. 131-153.

Ginsberg, Allen, "A Unified Approach to Automatic Indexing and Information Retrieval", IEEE Expert Magazine, v8, n5, Sept. 1993, p. 46.

Ginsberg, Allen, "Automatic Knowledge Base Refinement for Classification Systems", Artificial Intelligence, v35, n2, Jan. 1988, pp. 197-226.

Goodman, M., "Prism, An AI Case Based Text Classification System", Innovative Applications of AI, R. Smith and E. Rappaport (eds.), 1991, pp. 25-30.

Hao, X., Wang, J.T.L., Bieber, M.P., Ng, P.A., "Heuristic classification of office documents", International Journal on Artificial Intelligence Tools [Architectures, Languages, Algorithms], v3, n2, June 1994, pp. 233-265.

Hayes, P.J., Weinstein, S.P., "CONSTRUE/TIS: A System for Content Based Indexing of a Database of News Stories", Innovative Applications of AI 2, The AAAI Press/The MIT Press, Cambridge Ma, 1991, pp. 49-64.

Hebrail, G., Suchard M., "Classifying Documents: A Discriminant Analysis and an Expert System Work Together", CompStat '90, edited by Momirovic and Midner, Springer-Verlag, 1990, pp. 63-68.

Humphrey, S., "A Knowledge Based Expert System for Computer Assisted Indexing", IEEE Expert Magazine, v4, n3, 1989, p. 25.

Jones, K.P., "Automatic indexing for information retrieval systems", Informatics 11 pp. 51-64, 1991.

Jones, K.S., "Notes and references on early automatic classification work", SIGIR Forum, v25, n1, Spring 1991, pp. 10-17.

Lancaster, F.W., Elliker, C., Harkness Connell, T., "Subject analysis", Annual review on information science and technology, v24, Elsevier Science Publishers, 1989, pp. 35-84.

Lewis, D., "Evaluating Text Categorization", Proceedings of Speech and Natural Language Workshop, Morgan Kaufmann: San Mateo Ca, Feb. 1991, pp. 312-318.

Lewis, D., "An Evaluation of Phrasal and Clustered Representations on a Text Categorization Task", SIGIR Forum spec. iss., 1992, p. 37.

Liddy, E.D., Paik, W., Woelfel, J.K., "Use of subject field codes from a machine-readable dictionary for automatic classification of documents", Advances in Classification Research. Vol.3. Proceedings of the 3rd ASIS SIG/CR Classification Research Workshop, Oct. 1992, pp. 83-100.

Lin, Wei-Chung, Tsao, Chen-Kuo, "Document Classification Using Associative Memories", Neural Networks, 1989 IEEE International Conference, 1989 IEEE Publications.

Macleod, K., Robertson, W., "A Neural Algorithm for Document Clustering", Information Processing and Management, v27, n4, 1991, pp. 337-346.

Markus, R.S., "The RIAO '94 Conference and the Status of Information Retrieval: A Personal View", SIGIR Forum, v28, n2, Fall 1994, pp. 7-17.

Masand B., Linoff G., Waltz, D., "Classifying News Stories Using Memory Based Reasoning", SIGIR Forum spec. iss., 1992, p. 59.

Milstead, J.L., "Methodologies for subject analysis in bibliographic databases", Information Processing \& Management, v28, n3, 1992, p. 407-431.

Schocken, S., Pyun, J., "A Dempster-Shafer Model of Relevance", System Sciences, 1990 Annual Hawaii International Conference, p. 544.

Schuegraf, E.J., van Bommel, M.F., "An automatic document indexing system based on cooperating expert systems: design and development", Canadian Journal of Information and Library Science, v18, n2, July 1993, pp. 32-50.

Tzeras, K., Hartmann, S., "Automatic indexing based on Bayesian inference networks", SIGIR Forum spec. iss., 1993, pp. 22-34.

Wang, J.T.L., Ng, P.A., "TEXPROS: an intelligent document processing system", International Journal of Software Engineering and Knowledge Engineering, v2, n2, June 1992, pp. 171-96.

Watanabe, Toyohide, Nagoya, "Automatic Extraction and Classification of Data Items from Library Cataloging Cards by a Knowledge Based Approach", Proceedings of the International Workshop on Industrial Applications of Machine Intelligence and

Vision, Apr. 1989, pp. 67-71.
Willet P., "Recent Trends in Hierarchic Document Clustering: A Critical Review", Information Processing and Management, v24, 1988, pp. 577-597.

Bibliography

## Chapters 2 to 4

Belew, R.K., "A Connectionist Approach to Conceptual Information Retreival," Association for Computing Machinery, 1987, pp. 116-126.

Debessonet, C.G., Cross, G.R., "An Artificial Intelligence Application in the Law: CCLIPS, A Computer Program that Processes Legal Information," High Technology Law Journal, 1987, p. 329.

Gelbart, D., Smith, J.C., "FLEXICON: An Evaluation of a Statistical Ranking Model Adapted to Intelligent Legal Text Management," Proceedings of the Fourth International Converence on Artificial Intelligence and the Law, 1993.

Hahn, U., "Topic Parsing: Accounting for Text Macro Structures in Full-Text Analysis," Information Processing and Management, v26, n1, pp. 135-170.

Hearst, M.A., Plaunt, C., "Subtopic Structuring for Full-Length Document Access," Proceedings of the sixteenth International ACM SIGIR Conference on Research and Development in Information Retreival, 1993, pp. 59-68.

Ro, J.S., "An Evaluation of the Applicability of Ranking Algorithms to Improve the Effectiveness of Full-text Retrieval," Journal of the American Society for Information Science, v39, n3, 1988, pp. 73-78.

Salton, G., Allan, J., Buckley, C., "Approaches to Passage Retrieval in Full Text Information Retrieval," Proceedings of the sixteenth International ACM SIGIR Conference on Research and Development in Information Retreival, 1993, pp. 49-58.

Salton, G., Buckley, C., "Automatic Text Structuring and Retrieval: Experiments in Automatic Encyclopedia Searching", Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retreival, 1991, pp. 21-31.

Salton, G., Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer, Addison-Wesley Publishing Company, 1989.

Stanfill, C., Waltx, D.L., "Statistical Methods, Artificial Intelligence, and Information Retrieval," Text-based Intelligence Systems: Current Research and Practice in Information Extraction and Retrieval, Jacobs, P.A. (ed), Lawrence Erlbaum Associates, 1992, pp. 215-226.

## Appendix A: Program Code

The following appendix contains program code for the major steps in this process. This includes programs which build the database, calculate similarity scores, and split a case into subsections. There is also an include file listed at the end, which contains commonly used constants and data structures.
*//*
/*/**/*/
BUILD DATABASE ..... */
The following program takes a list of terms and stores */ /* them in a database. This list must be created by the */
/* FLEXICON program prior to running this program. In ..... */
/* addition to the term itself, the following information ..... */
/* is also stored as part of each database entry: ..... */Case identifier - A number which Identifies the case in
/* Case identifier - A number which Identifies the case inwhich the term appeared.- The location (in bytes) of the term*/ ..... */
from the beginning of the case.
from the beginning of the case.
Paragraph number - The number of the paragraph in which*/
the term appeared. ..... */
The above information is used by subsequent programs to */ calculate similarity between paragraphs.
The case identifier is read in as a parameter. The ..... */

offset is included with each term as part of the outputfrom the parsing process. The paragraph number is found */
/* from the parsing process. The paragraph number is found


| \#include | <stdinc.h> |
| :--- | :--- |
| \#include | <constant.h> |
| \#include | <fileio.h> |$\quad$ /* Standard C library */

int main (int argc, char *argv[]) \{
struct $\quad$ db_term_entry
/* Database records for */
/* structure declaration */
struct $\quad$ db_term_entry
*term_ptr;
*new_モerm_ptr;
FILE *termfile; /* Database file */
FILE *log_file; /* Log file */
FILE *in2_file; /* Paragraph boundaries file */
FILE *srt_file; /* List of terms file */
char $\quad l_{0}$ _filename[NAME_LEN]; /* File name strings */
char in2_filename[NAME-LEN];
char srt_filename[NAME_LEN];
char input_line[LINE_LENGTH]; /* Input line for various */


```
term count = OL;
term_entry_size = sizeof(struct db_term_entry);
                                    /* Reserve memory for term structures */
term_ptr = (struct db_term_entry *) malloc (term_entry_size);
```



```
    /* Open database file */
term_file = open_file(db_dir, DB_FILENAME, "r+b");
                            /* Open file with paragraph offsets in bytes */
sprintf(in2_filename, "%s.in2", filename);
in2 file = Open file(text dir, in2_filename, "rt");
see\overline{k}file(in2_file, OL, SEEK_SET);
result = read_line(in2_file, &input_line[0], LINE_LENGTH);
/* Open file with list of terms */
sprintf(srt_filename, "%s.srt", filename);
srt_file = Open_file(keys_dir, srt_filename, "rt");
see\overline{k}_file(srt_fīle, OL, SEEEK_SET);
    /* Get the number of paragraphs */
    /* currently in the DB - stored */
    /* in the first entry */
seek_file(term_file, OL, SEEK_SET);
read_file(term_file, term_ptr, term_entry_size, 1);
tota\overline{l}_paras = 可erm_ptr->o\overline{f}fset;
                                    /* Find end of DB and set pointer */
                                    /* for next new term to be added */
seek_file(term_file, 0, SEEK_END);
new_term_addr = ftell(term_file);
```

```
/* Find the number & end border */
/* in bytes) of the first */
/* paragraph */
```

result = read_line(in2_file, \&input_line[0], LINE_LENGTH);
*input_token $\bar{\equiv}$ strtok (input_line, "T\");
curr_pāra $=$ strtol (*input tōken, NULL, 10);
*input_token = strtok(NULE, "\n");
total_paras++;
/* Find the number \& end border */
/* of the next paragraph */
result = read_line(in2_file, \&input_line[0], LINE_LENGTH);
*input token $\equiv$ strtok(input_line, "T\");
next_pāra = strtol(*input_token, NULL, 10);
*input token $=$ strtok (NULL, " $\backslash$ n" $)$;
next_pāra_border = strtol(*input_token, NULL, 10);
totā̄_parās++;
/* Get the number of the */

```
                                    /* case being processed */
file_num = strtol(srt_filename, NULL, 10);
strcpy(term, "");
    /* MAIN LOOP */
    /* This loop is executed */
    /* once for each term in */
    /* the case */
    /* Get term input line */
result = read_line(srt_file, &input_line[0], LINE_LENGTH);
while (result}\mp@subsup{}{}{-}!= NULL
{
    /* Get term offset (in bytes) from */
                            /* from the beginning of the case */
*input_token = strtok(input_line, "|\\");
offset = strtol(*input_token, NULL, 10);
    /* Get term type - not used */
*input_token = strtok(NULL, "\\");
                                    /* Get term */
*input_token = strtok(NULL, "\\");
strcpy(term, *input_token);
                            /* Truncate term if it is too long */
if (strlen(term) > KEYWORD_LENGTH)
    term[KEYWORD_LENGTH-1] =-EOL;
                            /* Convert to lower case */
for (i=0; i<strlen(term); i++)
    if ((isupper(term[i]) != 0))
        term[i] = term[i] + 32;
                        /* While the current term offset */
                    /* is greater than the paragraph */
                            /* border, get the next paragraph */
while ((offset > next_para_border) &&
                (result != NULD))
{
        total_paras++; /* Increment paragraph counter */
        curr_para = next_para; /* Save current paragraph values */
                            /* Get next paragraph number and */
                            /* offset */
    result = read_line(in2_file, &input_line[0], LINE_LENGTH);
    *input_token = strtok(input_line, "\\");
    next_pära = strtol(*input tōken, NULL, 10);
    *input_token = strtok(NUL\overline{L, "\n");}
    next_para_border = strtol(*input_token, NULL, 10);
}
strcpy(new_term_ptr->term, term); /* Init DB term structure */
new_term_p\overline{tr->doc_id = file_num;}
new_term_ptr->offset = offset;
```

```
new_term_ptr->para = curr_para;
new_term_ptr->prev_addr =- OL;
new_term_ptr->next_addr = OL;
printf("adding term: %s\n", term); /* Log message */
term_count++;
prev_term_addr = OL;
term_addr = OL;
/* Traverse DB tree to find */
/* location for new term */
do
{
    /* Get term from DB */
    seek_file(term_file, term_addr, SEEK_SET);
    read_file(term_file, term_ptr, term_entry_size, 1);
    prev_term_addr = term_addr; /* Save parent DB term address */
    /* Get address of prev or next */
    /* child node in tree depending */
    /* on whether new term comes */
    /* before or after DB term */
    /* alphabetically */
    if (strcmp(term, term_ptr->term) < 0)
        term_addr = term_ptr->prev_addr;
    else
        term_addr = term_ptr->next_addr;
                            /* Continue until a leaf node is */
                            /* reached
                                    */
} while (term_addr != 0);
```

    /* Set child of parent node */
    /* to address of new term */
    if (strcmp(term, term_ptr->term) < 0)
term_ptr->prev_addr = new_term_addr;
else
term_ptr->next_addr $=$ new_term_addr;
/* Write parent node to DB */
seek_file(term_file, prev_term_addr, SEEK_SET);
write_file(term_file, term_ptr, term_entry_size, 1);
/* Write new child node to DB */
seek_file(term_file, new_term_addr, SEEK_SET);
writē_file(term_file, new_term_ptr, term_entry_size, 1);
/* Get address for next new term */
seek file(term_file, 0, SEEK_END);
new_Term_addr $\overline{=}$ ftell(term_file);

```
                                    /* Read next term input line */
    result = fgets( &input_line[0],LINE_LENGTH,srt_file);
}
fclose(in2_file);
fclose(srt_file);
fprintf(log_file, "%s added to DB\n", srt_filename);
                                    /* Save new paragraph total */
                                    /* as first entry in DB */
seek_file(term_file, OL, SEEK_SET);
read_file(term_file, term_ptr, term_entry_size, 1);
term_ptr->offset = total_paras;
seek_file(term_file, OL, SEEK_SET);
writē_file(term_file, term_pt\overline{r}, term_entry_size, 1);
fclos\overline{e}(term_fil\overline{e});
printf("\n");
printf(" terms added: %li\n", term count);
fprintf(log_file, " terms added: %Ii\n", term_count);
fprintf(log_file, "******** END BUILD DATABASE **********\n");
fclose(log_file);
return 0;
}
```

/****************************************************************//
/*
/*
/*
/*
/*

| FILE | *srt_file; /* List of terms file */ |
| :---: | :---: |
| FILE | *rst_file; /* Result file */ |
| FILE | *log_file; /* Log file */ |
| char | in2_filename[NAME_LEN]; /* File name strings */ |
| char | srt_filename[NAME_LEN]; |
| char | rst_filename[NAME_LEN]; |
| char | log_filename[NAME_LEN]; |
| char | input_line[LINE_LENGTH]; $\quad$ /* Input line for */ |
| char | out_line[LINE_LENGTH]; /* Output line */ |
| char | temp_out_line[LINE_LENGTH]; /* Temp output line */ |
| char | prev_input_line[LINE_LENGTH]; /* Previous input line */ |
| char | term[KEYWORD_LENGTH]; /* Term read from list file */ |
| char | term_category ${ }^{\text {c }} 80$ ]; /* Category for current term */ |
| char | temp_term_category[80]; |
| char | *input token[LINE_LENGTH]; /* Input line */ |
| char | *resul̄ |
| char | term_type; /* Type for current term */ |
| int | terms_in_paragraph[200]; /* Number of terms in */ |
| int | /* current paragraph */ <br> total_terms; /* Total terms in database */ |
| long | term_count; /* Number of terms processed */ |
| long | term_entry_size; /* Size of term entry DB structure */ |
| long | new_term_addr; /* Address pointers for DB and */ |
| long | prev_dic $\bar{t}_{\text {_term_addr; /* Dictionary */ }}$ |
| long | dict_term_addr; |
| long | prev_term_addr; |
| long | term_addr; |
| int | dict_search_count; /* Number dictionary searches */ |
| int | dict_term_found; |
| long | dict_term_entry_size; /* Size of entry in Dictionary */ |
| int | curr_para; /* Values (in bytes) for current */ |
| long | curr_para_border; /* and next paragraph boundaries */ |
| int | next_para; |
| long | next_para_border; |
| long | file_num; /* Number of file being processed */ |
| int | term_category_count; /* Number of terms with this */ /* category */ |
| long | offset; /* Offset from beginning of case */ |
| int | i; |
| int | compare_length; |

```
int search_count; /* Number of DB searches */
int
int
int
int stat_count; /* Counters for each para type */
int case_count;
int fact count;
int legaI_term_count;
int
char
double
double
double
double
double
double
double
double
double
long total_paras; /* Total number of paragraphs */
char doc_para[20];
char db_dir[NAME_LEN]; /* Directory names */
char ouEput dir[\overline{NAME_LEN];}
char text_dir[NAME_LEN];
char keys-dir[NAME-LEN];
char fileñame[NAME_LEN];
if (argc < 6) /* Check input parameters */
{ printf("Usage: CALC_SIM database_dir output_dir text_dir
keys_dir file_name\n")
    re\overline{turn(1);}
}
```

```
strcpy(db_dir, argv[1]); /* Get various directories */
```

strcpy(db_dir, argv[1]); /* Get various directories */
strcpy(oūput_dir, argv[2]);
strcpy(oūput_dir, argv[2]);
strcpy(text dír, argv[3]);
strcpy(text dír, argv[3]);
strcpy(keys_dir, argv[4]);
strcpy(keys_dir, argv[4]);
strcpy(fileñame, argv[5]); /* Get case name */
strcpy(fileñame, argv[5]); /* Get case name */
/* Open log file */
/* Open log file */
sprintf(log_filename, "%s.log", filename);
sprintf(log_filename, "%s.log", filename);
log_file = open_file(output_dir, log_filename, "at");
log_file = open_file(output_dir, log_filename, "at");
fprīntf(log_fil\overline{e, "********- START CALCCULATE SIMILARITY}
fprīntf(log_fil\overline{e, "********- START CALCCULATE SIMILARITY}
**********<br>overline{n}}\mp@subsup{}{}{\prime\prime})
**********<br>overline{n}}\mp@subsup{}{}{\prime\prime})
term_count = OL;
／＊dictionary entries＊／
term＿entry＿size＝sizeof（struct db＿term＿entry）； dict＿－term＿entry＿size＝sizeof（struc̄t dic̄t＿term＿entry）；
／＊Reserve memory for＊／
／＊various structures＊／
term ptr $=$（struct $d b$ term entry＊）malloc（term entry size）； new＿左erm＿ptr＝（struc̄ db＿モerm＿entry＊）malloc（̄̄erm＿entry＿size）； dic $\bar{t}$ ter $\bar{m}$＿ptr $=$（struct dict＿tērm＿entry＊）malloc （dic気＿term＿entry＿size）； prev＿可erm＿̄̄tr＝（struct dict＿term＿entry＊）malloc （dic $\bar{t} \_$term＿entry＿size）；
／＊Reserve memory for documents＊／
head＿doc＿list $=$（struct doc＿list＿entry＊）malloc（sizeof（struct doc＿list－entry））；
memset（head＿doc＿list－＞doc＿para，0， sizeof（head＿doc＿list－＞doc＿para））； head＿doc＿list－＞next $=$ NUL̄$;$
／＊Reserve memory for categories＊／ head＿category $=$（struct category＿entry＊）malloc（sizeof（struct category＿entry））； memset（hēad＿category－＞type，0，sizeof（category＿list－＞type））； head＿category－＞count $=0$ ； head＿category－＞next＝NULL；
／＊Open database and dictionary files＊／
term＿file $=$ open＿file（db＿dir，DB＿FILENAME，＂r＋b＂）； dict＿file＝open＿file（db＿dir，DICT＿FILENAME，＂r＋b＂）；
／＊Open file with paragraph offsets＊／
sprintf（in2＿filename，＂\％s．in2＂，filename）；
in2 file $=$ O－pen file（text dir，in2＿filename，＂rt＂）； seē̄＿file（in2＿file，OL，SEEK＿SET）； result $=$ read＿line（in2＿file，\＆input＿line［0］，LINE＿LENGTH）；
／＊Open file with list of terms＊／
sprintf（srt＿filename，＂\％s．srt＂，filename）；
srt＿file＝Open＿file（keys＿dir，srt＿filename，＂rt＂）； seek＿file（srt＿file，OL，SEEK＿SET）；

```
/* Find the number & end border (in */
                    /* bytes) of the first paragraph */
                    /* and the next paragraph */
```

result $=$ read_line(in2_file, \&input_line[0], LINE_LENGTH);
result $=$ read_line (in2 file, \&input line[0], LINE_LENGTH);
*input token $\bar{\equiv}$ strtok(input_line, "T\");
curr_pāra $=$ strtol (*input tōken, NULL, 10) - 1;
*input token $=$ strtok(NULE, "\n");
curr_pāra_border = strtol(*input_token, NULL, 10);

```
next_para = curr_para + PARAGRAPH_WINDOW;
memsét(terms_in_paragraph, 0, 200);
/* The following loop reads through */
/* the list of terms once. Its */
/* purpose is to count the number */
/* of terms in each paragraph. */
/* Get term input line */
result = read_line(srt_file; &input_line[0], LINE_LENGTH);
while (result}\mp@subsup{}{}{-}!= NULL
{
                    /* Get term offset (in bytes) from */
                    /* from the beginning of the case */
    *input_token = strtok(input_line, "\\");
    offset}\mp@subsup{}{}{-}= strtol(*input_toke\overline{n}, NULL, 10)
    while ((offset > curr_para_border) &&
        (result != NULL))
    {
        result = read_line(in2_file, &input_line[0], LINE_LENGTH);
        *input_token = strtok(input_line, "T\");
        curr_pära = strtol(*input tōken, NULL, 10) - 1;
        *input_token = strtok(NUL\overline{L, "\n");}
        curr_pära_border = strtol(*input_token, NULL, 10);
        next_para-= curr_para + PARAGRAP\overline{H_WINDOW;}
    }
                            /* Increment term counter */
    terms_in_paragraph[curr_para]++;
    resul\overline{t =- read_line(srt_\overline{f}ile, &input_line[0], LINE_LENGTH);}
}
    /* Reset input files */
seek file(in2_file, OL, SEEK_SET);
resul̄t = read_line(in2_file, &input_line[0], LINE_LENGTH);
seek_file(srt_file, OL, SEEK_SET);
    /* Open result file */
sprintf(rst_filename, "%s.rst", filename);
rst_file = Open_file(output_dir, rst_filename, "wt");
                            /* Get the number of paragraphs */
                            /* currently in the DB - stored */
                            /* in the first entry */
seek_file(term_file, OL, SEEK_SET);
read_file(term_file, term_ptr, term_entry_size, 1);
```


/* Find end of DB and set pointer */
/* for next new term to be added */
seek_file(term_file, 0, SEEK_END);

```
new_term_addr = ftell(term_file);
    /* Find the number & end border (in */
    /* bytes) of the first paragraph */
result = read_line(in2_file, &input_line[0], LINE_LENGTH);
result = read_line(in2-file, &input_line[0], LINE_LENGTH);
*input_token \equiv strtok(ínput_line, "\\");
curr_pāra = strtol(*input_tōken, NULL, 10) - 1;
*inpút_token = strtok(NUL\overline{L, "\n");}
curr_pāra_border = strtol(*input token, NULL, 10);
next_para-= curr_para + PARAGRAP\overline{H_WINDOW;}
                                    /* Get the number of the */
                    /* case being processed */
file_num = strtol(srt_filename, NULL, 10);
                                    /* Initialize values */
strcpy(term, "");
doc_score = norm_score = 0;
cur\overline{r}_norm_value = next_norm_value = 0;
stat_coun\overline{t = case_coun\overline{t}= fact_count = legal_term_count = 0;}
para-type = ' ';
lega\overline{_}ratio = 0.0;
/* MAIN LOOP */
/* This loop is executed */
/* once for each term in */
/* the case */
/* Get term input line */
result = read_line(srt_file, &input_line[0], LINE_LENGTH);
while (result'!= NULL)
{
                                    /* Get term offset (in bytes) from */
                                    /* from the beginning of the case */
    *input_token = strtok(input_line, "|\\");
    offset = strtol(*input_token, NULL, 10);
    /* Get term type - not used */
    *input_token = strtok(NULL, "\\");
    term_tȳpe = *input_token[0];
                                    /* Get term */
*input_token = strtok(NULL, "\\");
strcpy(term, *input_token);
                                    /* Truncate term if it is too long */
if (strlen(term) > KEYWORD_LENGTH)
    term[KEYWORD_LENGTH-1] = EOL;
```

```
                                    /* Convert to lower case */
for (i=0; i<strlen(term); i++)
    if ((isupper(term[i]) != 0))
        term[i] = term[i] + 32;
printf("processing term: %s\n", term);
term_count++;
prev_term_addr = OL;
term_addr = OL;
sear\overline{ch_count = 0;}
term_found = FALSE;
match_in_next_para = 0;
paras_coñtaining_term = 0;
                    /* If the current term offset is */
                    /* greater than the paragraph border, */
                    /* find the category for the current */
                    /* paragraph. This is done by */
                    /* finding the most frequent term */
                    /* category in the list created as */
                    /* the terms were processed. */
if (offset > curr_para_border)
{
                            /* If no terms in list, set category */
                    /* to blank
                            */
if (head_category->next != NULL)
{
    term_category_count = 1;
    strcpy(term_category, EOL_STR);
    category_li\overline{s}t = head_category->next;
                            /* For each category in the list */
    while (category_list != NULL)
    {
    /* Compare count for each category */
    /* with the current max count */
        if (category_list->count > term_category_count)
        {
                            /* If the current count is higher, */
                            /* save the category and count */
            strcpy(term_category, category_list->type);
            term_category_count = category_list->count;
            }
                    /* If the current count is the same */
                    /* and > 1, save both categories */
        else if ((category_list->count == term_category_count) &&
                        (term_category_count > 1))
        {
            sprintf(temp_term_category, "%s | %s",
            term_category, category_list->type);
            strcpy(term_category, temp_term_category);
```

```
                                    /* Get next category */
        new_category = category_list;
        catēgory_list = category_list->next;
        free(new_category);
    }
    free(category_list);
    head_category=>next = NULL;
}
else
    strcpy(term_category, EOL_STR);
                                    /* Calculate normalized */
                                    /* paragraph score */
if ((curr_norm_value > 0) && (next_norm_value > 0))
    norm_scōre =- doc_score / sqrt(currr_no\overline{rm_value *}
        next_norm_value);
else
    norm_score = 0.0;
                                    /* Calculate average paragraph score */
if (terms_in_paragraph[curr_para] > 0)
{ total terms = 0;
    for (i=curr_para+1; i<=next_para; i++)
        total_terms = total_terms + terms_in_paragraph[i];
    if (total_terms > 0)
        ave_score = doc_score / (terms_in_paragraph[curr_para] *
            tōtal_terms);
    else
        ave_score = 0;
}
else
    ave_score = 0;
                                    /* Output results only if paragraph */
                    /* has > 5 terms or score > 0 */
if ((terms_in_paragraph[curr_para] > 5) ||
            (doc_score > 0))
{
                                /* Find ratio of legal/non-legal terms */
    if (fact_count > 0)
        legal_ratio = (double)legal_term_count /
                                    (double)fact_count;
    else
        legal_ratio = 0;
                                    /* Find paragraph type */
    if ((case_count + stat_count > 0) |
            (legal_ratio > 0.25))
        para_type = 'L';
    else
        para_type = 'F';
```

                                /* Print paragraph score and other */
    ／＊values to result and $\log$ files＊／

```
fprintf(rst_file,"%2i\t%c\t%6.3g\t%6.3g\t%2i\t%2i\t%2i\t%2i\t%s\n",
    curr_para, para_type, ave_score, norm_score,
    fact_count, legal_term_count, case_count, stat_count,
    term_category);
    fprintf(log_file,"%2i\ttype: %c\tfact: %2i\tlaw: %2i\tcase:
        %2i\tstat: %2i\tcategory: %s\n",
        curr_para, para_type, fact_count, legal_term_count,
        case_count, stat_count, term_category);
        fprint\overline{f}(log_file,"陪i\tdoc_score: %%.3g\tave_score:
        %6.3g\tnorm_score: %6.3g\n",
        curr_para, doc_score, ave_score, norm_score);
    }
    stat_count = case_count = fact_count = legal_term_count = 0;
    doc_score = 0;
    norm_score = 0;
    curr_norm_value = 0;
    next_norm_value = 0;
```

                                    /* Find the number \& end border (in */
                                    /* bytes) of the next paragraph */
    while ((offset > curr_para_border) \&\&
                (result != NUL気)
    \{
        fprintf(log_file, "****************** \%li end of: \%i
    *****************菓** $\mathbf{n}^{\prime \prime}$,
curr_para_border, curr_para);
result $=$ read_line(in2_file, \&input_line[0], LINE_LENGTH);
if (result !=-NULL)
\{
*input_token = strtok(input_line, "<br>");
curr_para = strtol(*input_token, NULL, 10) - 1;
*input_token = strtok(NUL工, "\n");
curr_para_border = strtol(*input_token, NULL, 10);
next_para = curr_para + PARAGRAPH_WINDOW;
\}
/* Set border for last paragraph */
else
\{
curr_para = 9999;
curr_para_border = 9999;
next_para = 9999;
\}
\}
\}

```
fprintf(log_file, "%i term: %s %c %li\n", curr_para, term,
    term_type, offset);
                                    /* Increment counters for paragraph type */
switch(term_type)
{
    case 'F': fact count++;
        brea\overline{k};
    case 'P': legal_term_count++;
        break;
    case 'C': case_count++;
        brea\overline{k};
    case 'E': stat_count++;
                break;
}
```

do
\{
/* Get term from DB */
seek_file(term_file, term_addr, SEEK_SET);
read_file(term_file, term_ptr, term_ētry_size, 1);
prev_term_addr = term_addr; /* Save parent DB term address */
/* Find comparison length for term */
search_count++;
if (strlen(term) < strlen(term_ptr->term))
compare_length = strlen(term);
else
compare_length = strlen(term_ptr->term);
if (compare_length >= 5)
compare_length = compare_length * 0.75;
/* Compare ter with current DB entry */
if (strncmp(term, term_ptr->term, compare_length) $==0$ )
\{
/* If terms match and DB entry is */
/* within the window of following */
/* paragraphs, increment match count */
if ((term_ptr->doc_id == file_num) \&\&
(term_ptr->para > curr_para) \&\&
(term_ptr->para <= next_para))
\{
match_in_next_para++;
fprin̄̄f(̄og_file, " match in: \%i\n", term_ptr->para);
\}
/* Add this document to the list of */
/* documents containing this term */

```
            sprintf(doc_para, "%li%i", term_ptr->doc_id, term_ptr->para);
        doc_list = head_doc_list;
        while ((doc_lis\overline{t}->doc_para != doc_para) &&
            (doc_list->nex\overline{t}}!=\mathrm{ NULL))
            doc_list = doc_list->next;
        if ( strcmp(doc_İist->doc_para, doc_para) != 0)
        {
            paras_containing_term++;
            new_dōc_node = (struct doc_list_entry *) malloc
                        (\overline{sizeōf(struct doc_list_entry)});
            strcpy(new_doc_node->doc_para, doc_para);
            new_doc_node->n̄ext = NUL\overline{L};
            doc_lis\overline{t}->next = new_doc_node;
}
    term_addr = term_ptr->next_addr;
        term_found = TRUE;
    }
            /* Get address of prev or next child */
                    /* node in tree depending on whether */
                    /* new term comes before or after */
                            /* DB term alphabetically */
if (strcmp(term, term_ptr->term) < 0)
    term_addr = term_ptr->prev_addr;
else
    term_addr = term_ptr->next_addr;
                                    /* Continue until a leaf node is reached */
} while (term_addr != 0);
                                    /* Free memory for document list */
doc_list = head_doc_list->next;
whi\e (doc_list'!= N
{
    new_doc_node = doc_list->next;
    fre\overline{e}(doc
    doc_list = new_doc_node;
}
head_doc_list->next = NULL;
                                    /* If term was found in database, check */
                                    /* for term in dictionary */
if (term_found == FALSE)
    fprintf(\log_file, "term: %s not found\n", term);
else
{
    prev_dict_term_addr = OL;
    dict_term_addr-= OL;
    dict_search_count = 0;
    dict_term_found = FALSE;
```

```
/* Traverse dictionary tree to find */
/* term
seek_file(dict_file, dict_term_addr, SEEK_SET);
read_file(dict_file, dict_term_ptr, dict_Eerm_entry_size, 1);
dict_search_count++;
prev_dict_tērm_addr = dict_term_addr;
                                    /* Check for match */
if (strcmp(term, dict_term_ptr->term) == 0)
{
    dict_term_addr = 0;
    dict_term_found = TRUE;
    strc\overline{p}y(prēv_term_ptr->term, EOL_STR);
    strcpy(out_line,,EOL_STR);
                                    /* When a match is found, traverse back */
                                    /* up the dictionary tree to find the */
                                    /* parent category for the term */
    while (dict_term_ptr->parent_addr != 0)
    {
        *prev_term_ptr = *dict_term_ptr;
        seek_\overline{file(\overline{dict_file, dīct_tērm_ptr->parent_addr,}}\mathbf{}\mathrm{ ,}
            SEEK SET);
        read_fīle(dict_file, dict_term_ptr, dict_term_entry_size,
            1);
        sprintf(temp_out_line, " | %-15s", dict_term_ptr->term);
        strcat(temp_out_line, out_line);
        strcpy(out_\}ine, temp_out_line)
    }
    if (prev_term_ptr->term == EOL_STR)
        strcpy(term_category, dict_tērm_ptr->term);
    else
        strcpy(term_category, prev_term_ptr->term);
}
                    /* Get address of prev or next child */
                    /* node in tree depending on whether */
                    /* new term comes before or after */
                    /* dictionary term alphabetically */
else if (strcmp(term, dict_term_ptr->term) < 0)
    dict_term_addr = dict_term_ptr
else
    dict_term_addr = dict_term_ptr->next_addr;
                            /* Continue until a leaf node is reached */
} while (dict_term_addr != 0);
```

```
    /* If term was found in dictionary, */
    /* add its category to the category */
    /* list
if (dict_term_found == TRUE)
{
    fprintf(log_file, "found in dictionary - category: %s\n",
        term_category);
    category_list = head_category;
                                    /* Traverse category list until */
                                    /* term category is found or end */
                            /* is reached */
    while ((strcmp(category_list->type, term_category) != 0) &&
                (category_list->next != NULL))
        category_list \equiv category_list->next;
                                    /* If category is found, increment */
                    /* counter for that category */
    if ( strcmp(category_list->type, term_category) == 0)
        category_list->count++;
    else
    {
                    /* If category not found, add it */
        new_category = (struct category_entry *) malloc
            (\overline{sizeof(struct category_entry));}
        strcpy(new_category->type, term_category);
        new_category->next = NULL;
        new_category->count = 1;
        category_list->next = new_category;
    }
}
/* If term was matched, calculate */
/* term weights, otherwise weights */
/* are zero */
if (match_in_next_para != 0)
{
    curr_term_wt = log10((double)total_paras /
        (double)paras_containing_term);
    next_term_wt = (double)match_in_next_para *
        log}10((\overline{double)total_paras / (d
    term_score = curr_term_wt * next_term_wt;
    curr_norm_value = curr_norm_value + pow(curr_term_wt,2);
    next_norm_value = next_norm_value + pow(next_term_wt,2);
    doc_score = doc_score \mp term_score;
    fprintf(log_file, " score: %6.3g\n",term_score);
}
```

\}

```
    /* Get next term */
    result = fgets( &input_line[0],LINE_LENGTH,srt_file);
}
```

                                    /* Close files */
    fclose(in2_file);
fclose (srt_file);
fclose (rst_file);
fprintf(log_file, "\%s added to $D B \backslash n "$, srt_filename);
fclose(term_file);
printf("\n");
printf(" terms added: \%li\n", term_count);
fprintf(log_file, " terms added: \%Ii\n", term_count);
fprintf(log_file, "******** END CALCULATE SIMILARITY
********** $\left.\backslash \mathbf{n}^{\prime \prime}\right)$;
fclose(log_file);
return 0;
\}

```
/*************************************************************************
/* */
/*
/*
/*
/*
/*
/* from the CALC_SIM program and splits a case into subsections */
/* based on these}\mathrm{ similarity values. */
/* The output from this program consists of the text of a */
/* case marked with section boundaries, as well as section */
/* subjects.

```

\#include <stdinc.h> /* Standard C library */
\#include <constant.h> /* Constants \& record structures */
\#include <fileio.h> /* Subroutines for file input and output
*/
int main (int argc, char *argv[])
{
struct
{
int num;
char type;
double score;
double norm_score;
double ave_score;
double norm_ave_score;
char category[80];
} para[200];

```
\begin{tabular}{|c|c|c|}
\hline struct struct struct & ```
category_entry
category_entry
category_entry
``` & \begin{tabular}{l}
Pointers for paragraph data strucure */ \\
*head_category; \\
*category_list; \\
*new_category;
\end{tabular} \\
\hline int & current; & /* Paragraph numbers */ \\
\hline int & prev; & \\
\hline int & next; & \\
\hline int & prev_ave; & /* Average paragraph scores */ \\
\hline int & next_ave; & \\
\hline int & end; & \\
\hline int & start; & /* First and last paragraph numbers */ \\
\hline int & finish; & \\
\hline
\end{tabular}

strcpy(filename, argv[5]);
/* Open \(\log\) file */
sprintf(log_filename, "\%s.log", filename); log_file = open_file(output_dir, log_filename, "at"); fprintf(log_file, "********-START SPIITT **********\n");
/* Open input and output files */
sprintf(end_filename, "\%s.end", filename); sprintf(fin_filename, "\%s.fin", filename); sprintf(par filename, "\%s.p", filename); sprintf(bnd_filename, "\%s.b", filename); end_file = ōpen_file(output_dir, end_filename, "rt"); fin_file = open_file(output_dir, fin_filename, "rt"); par_file = open_file(output_dir, par_filename, "wt"); bnd_file = open_file(output_dir, bnd_filename, "wt");
/* Reserve memory for list of categories */ head_category \(=\) (struct category_entry *) malloc (sizeof(struct category_entry)); memset (head_category->type, 0, sizeof(category_list->type)); head_category->count \(=0\);
head_category->next \(=\) NULL;
/* Initialize */
memset (para, 0, sizeof(para));
memset(split, 0, sizeof(split));
current = 1;
split_count \(=0\);
/* Get result values for first paragraph */
end_result = read_line(end_file, \&end_line[0], LINE_LENGTH);
while (end_result ! = NULL)
\{
/* Parse result values for each paragraph */
/* and store them in an array
*input_token = strtok (end_line, "\t");
para[current].num = atoi( \({ }^{*}\) input token);
*input_token = strtok(NULL, "\t");
para[current].type \(=\) *input_token[0];
*input token \(=\) strtok(NULL, " " 1 "") ;
para[current].score \(=\) strtod(*input_token, NULL);
*input_token = strtok(NULL, "\t");
para[current]. norm_score \(=\) strtod(*input_token, NULL);
*input_token = strモ̄ok(NULL, "\t");
para[current].ave_score = strtod(*input_token, NULL);
*input_token = str̄tok(NULL, "\t");
para[cūrrent]. norm_ave_score = strtod(*input_token, NULL);
*input_token = strच̄ok(NULL, "\t");
*input_token = strtok(NULL, "\t");
```

    *input_token = strtok(NULL, "\t");
    *input_token = strtok(NULL, "\t");
    *input-token = strtok(NULL, "\t");
    *input_token = strtok(NULL, "\n");
    if (*input_token != NULL)
    strcpy(pära[current].category, *input_token);
    else
        strcpy(para[current].category, EOL_STR);
    current++;
    end_result = read_line(end_file, &end_line[0], LINE_LENGTH);
    }
end = current;
current = 1;

```
    /* Set current, previous and next */
    /* paragraph numbers */
current \(=\) AVE_COUNT +1 ;
prev \(=\) curren \(\bar{t}-1\);
next \(=\) current +1 ;
prev_ave = current - AVE_COUNT;
next_ave = current + AVE_COUNT;
                                    /* MAIN LOOP */
                                    /* Process until paragraph number */
                                    /* reaches last paragraph */
while (next_ave <= end)
\{
                    /* If the previous and next scores */
/* are greater than the current
/* score (a valley)
if ((para[prev_ave]. norm_ave_score >=
    para[current]. norm_ave_score) \&\&
        (para[next_ave]. norm_ave_score >=
    para[current]. norm_ave_score))
\{
    /* Find the lowest point in the valley */
    /* Scores varying by less than 0.02 */
    /* are considered to be equal */
    if (para[prev_ave]. norm_ave_score -
        para[curren \(\bar{t}\) ]. norm_ave scōre \(<0.02\) )
        start \(=\) current - \(\bar{A} V E_{-} \bar{C} O U N T ;\)
    else
        start = current;
    if (para[next_ave]. norm_ave_score -
        para[curren̄]. norm_ave_scōre < 0.02)
        finish \(=\) current + 2 * AVE_COUNT - 1;
    else
```

            finish = current + AVE COUNT - 1;
        diff = 0.0;
        para_num = para[finish].num;
                                    /* Find the lowest score among those */
                                    /* at the bottom of the valley
                                    */
        for (i=start; i<finish; i++)
        {
            if ((temp_diff = para[i+1].score - para[i].score) > diff)
            {
                diff = temp_diff;
                para_num = para[i+1].num;
            }
        }
        split[split_count] = para_num;
        split_count++;
    }
    /* increment paragraph counters */
    current = current + AVE_COUNT;
    prev = current - 1;
    next = current + 1;
    prev_ave = current - AVE_COUNT;
    next_ave = current + AVE_COUNT;
    }
split[split_count] = end;
end_split = split_count;
memset(term_categöry, 0, sizeof(term_category));
split_count = 0;
curreñt = 1;
/* Find section category using */
/* paragraph categories */
while (current < end)
{
/* If category found for paragraph */
if (strcmp(para[current],category, EOL_STR) != 0)
{
category_list = head_category;
/* Find category in current list */
while ((strcmp(category_list->type, para[current].category) !=
0) \&\& (category_list->next != NULL))
category_list =- category_list->next;
/* If found, increment counter */
if ( strcmp(category_list->type, para[current].category) == 0)
category_list->count++;
else
{
/* If not found, add it to the list */
new_category = (struct category_entry *) malloc
(\overline{sizeof(struct category_entry)});
strcpy(new_category->type, para[current].category);

```
```

        new_category->next = NULL;
        new_category->count = 1;
        catēgory_list->next = new_category;
    }
    }

```
/* If current paragraph is a section */
    /* boundary, find the category for */
    /* this section. This is done by */
    /* finding the most frequent */
    /* category among paragraphs in the */
    /* section.
        */
if (para[current].num == split[split_count])
\{
    if (head_category->next != NULL)
    \{
        term_category_count \(=1\);
        strcpy (term_cātegory[split_count], EOL_STR);
        category_lis̄t = head_categōry->next;
                        /* For each category in the list */
        while (category_list != NULL)
        \{
                                    /* If this category is more frequent */
                                    /* save it as the section category */
            if (category_list->count > term_category_count)
            \{
                        strcpy(term_category[split_count], category_list->type);
            term_category_count = category_list->count;
        \}
                    /* If it has the same frequency, */
                            /* combine it with the current category */
        else if ((category_list->count \(==\) term_category_count) \&\&
                                    (term_category_count > 1))
            \{
                sprintf(temp_term_category, "\%s | \%s",
                        term_category[split_count], category_list->type);
                        strc̄̄y (term_category\(\left.\left[s p l i t \_c o u n t\right], ~ t e ̄ m p \_t e r m \_c a t e g o r y\right) ; ~\)
            \}
                                    /* Get the next category */
            new_category = category_list;
            category_list = category_list->next;
            free(new_category);
        \}
        free(category_list);
        head_category=>next \(=\) NULL;
    \}
    else
                                    /* No categories in the list */
        strcpy(term_category[split_count], EOL_STR);
        split_count++;
\}
```

    current++;
    }
split[end_split] = 0;
split_count = 0;
fin_result = read_line(fin_file, \&fin_line[0], LINE_LENGTH);
/* The following loop reads through */
/* each line in the case, and inserts */
/* a section subject and boundary */
/* where the above routines */
/* calculated they should appear. */
while (fin_result != NULL)
{
if (fin_line[0]== '<br>')
{
strcpy(temp_line, fin_line);
*input_token = strtok(temp_line, "<br>");
para_num = atoi(*input_token);
/* If current paragraph is a */
/* section boundary
*/
if (para_num == split[split_count])
{
/* Mark boundary in output file */
if (strcmp(term category[split_count], EOL_STR) != 0)
fprintf(par_file, "Section sūbject: %ss\n",
term_category[split_count]);
else
fprintf(par file, "Section Subject: Not Found\n");
fprintf(par_fīle,
"\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#
\#\#\#\#\#\#\#\#\#\#\#\n');
split count++;
fprintf(bnd_file, "p: %i\n", para_num-1);
}
}
if (strncmp(fin_line, "----------", 10)== 0)
{
fprintf(bnd_file, "m: %i\n", para_num);
}
fprintf(par_file, fin_line);
fin_result \equiv read_line\overline{(fin_file, \&fin_line[0], LINE_LENGTH);}
}
if (strcmp(term_category[split_count], EOL_STR) != 0)
fprintf(par_file, "Section Subject: %s\n",
term_category[split_count]);

```
```

else
fprintf(par_file, "Section Subject: Not Found\n");
fclose(end_file);
fclose(fin file);
fclose(par_file);
fclose(bnd_file);
printf("\n");
fprintf(log_file, "******** END SPLIT *********\n");
fclose(log_file);
return 0;
}

```
```

/*****************************************************************/
/*
/* CONSTANTS */
*/
/* */
/* The following file contains constants, global variables
*/
/* and record structures used in other programs.
/* */
/***********************************************************************
\#define DB_FILENAME \#define DICT FILENAME \#define KEYWORD LENGTH \#define LINE LENḠTH
\#define NAME LEN 50
\#define MIN MÄTCH LEN
\#define PARA\overline{GRAPH_WINDOW 3}
\#define AVE_COUNT }\mp@subsup{}{}{-
\#define FALSE 0
\#define TRUE 1
\#define EOL '\x0'
\#define EOL_STR
"\x0"
"TERMS.DB"
"TERMS.DCT"
30
200
5
\#define PAR\overline{A}GRAPH_WINDOW 3
/* Database file entry structure */
struct db_term_entry
{
Char term[KEYWORD_LENGTH];
long doc_id;
long offset;
int
long
long
};
/* Dictionary file entry structure */
struct dict term entry
{
char term[KEYWORD_LENGTH];
long parent addr;
long prev_ad̄dr;
long next-addr;
char temp[6];
};
/* Document list structure */
struct
{
doc_list_entry

```
```

    char
    doc_para[10];
    struct doc_list_entry *ne\overline{x};
    };
/* Category list structure */
struct
{
char
int
struct category_entry
};

```

\section*{Appendix B: Case Excerpts}

The following appendix contains excerpts from cases which were processed for this project. Each excerpt includes the case cite, a number of paragraphs from the case, and a partial listing of the log file created when this case was processed. The selected paragraphs correspond to those mentioned in the section on analysis of the results.

A single dotted line (-----) in the text of a case denotes manually selected subsection boundaries. A double line \((=====\) ) denotes subsection boundaries selected by the program.

The log file lists each term as it is processed. It includes information about whether each term was matched in subsequent paragraphs, as well as whether it was found in the legal dictionary. The log file also includes similarity totals and legal categories for each paragraph (when available).
IN THE MATTER OF THE ADOPTION ACT, BEING CHAPTER 4 OF THE R.S.B.C., AND AMENDMENTS THERETO AND IN THE MATTER OF A MALE INFANT, BRITISH COLUMBIA BIRTH REGISTRATION NUMBER 78-09-024190 B.C.S.C.

\author{
Type: Law
}
1014 In 1979 Manitoba introduced a similar provision in its Child Welfare Act, S.M. 1974, c. 30 by adding s.100(5) [en. 1979, c.22] which states:
Where one of the applicants... dies prior to the making of an order of adoption by the judge, the judge may nevertheless grant the order of adoption in the names of both applicants; and in that case the child shall be deemed for all purposes to have been adopted prior to the death of that applicant.

Type: constitutional law
\(1015 \backslash\) PARENS PATRIAE JURISDICTION
Mr. Eeles, solicitor for the petitioner, submits that because our Adoption Act is silent on the issue, the court's jurisdiction must come from its role as parens patriae.

\footnotetext{
Type: Fact
1016 In his article "The Welfare of the Children and the Jurisdiction of the Court under Parens Patriae" which is found in Connell-Thouez and Knoppers Contemporary Trends in Family Law: A National Prospective (1984), S.I. Bushnell summarizes the description of parens patriae given by Lord Esher M.R. in The Queen v. Gyngall, [1985] 2 Q.B. 232 (C.A.) in the following way:

Lord Esher points out that the words signify a jurisdiction exercised by the Court of
Chancery from time immemorial; it is a "parental" jurisdiction, judicially administered, by which the Court acts on behalf of the Crown as guardian of all infants. It is a prerogative that has been delegated to the Court of Chancery, and, when exercising the jurisdiction the Court acts in the manner of a wise, affectionate, and careful parent for the welfare of the child. The welfare of the child is to be the dominant matter to be considered, with "welfare" to be
}
taken in its largest possible sense. (p. 225)
This passage was cited with approval by Huddart, L.J.S.C. (as she then was) in O'Driscoll v. McLeod (1986), 10 B.C.L.R. 108 at p. 113.

Type: Fact
\(\backslash 017\) There are mainly three situations in which the court may assert its parens patriae jurisdiction. The first are emergency situations in which a child is considered to be in need of protection: Re D.S.; Supt. of Fam. \& Child Service v. R.D., [1983] W.W.R. 618 (B.C.S.C.). The second is judicial review of the Superintendent's exercise of his statutory power: J.D.S. v. W.S. and Supt. of Fam. \& Child Service, supra; W.A.M. and P.L.M. v. Supt. of Fam. \& Child Service (1985), 65 B.C.L.R. 229 (B.C.C.A.). The third, which is relevant to the case at bar, arises when there is found to be a "gap" in the legislation in issue. This test was established by Wilson, J. in Beson v. Dir. of Child Welfare, [1982] 2 S.C.R. 716 (S.C.C.). In that case the child was placed with the Beson family for adoption. After just six months the child was removed from the Beson home by the authorities because of allegations of child abuse which were ultimately unfounded. The Supreme Court of Canada invoking parens patriae overruled the decision of the Director to remove the child and ordered that adoption be granted to the Besons.

Log File for:

\section*{IN THE MATTER OF THE ADOPTION ACT, BEING CHAPTER 4 OF THE R.S.B.C., AND AMENDMENTS THERETO AND IN THE MATTER OF A MALE INFANT, BRITISH COLUMBIA BIRTH REGISTRATION NUMBER 78-09-024190 B.C.S.C.}

14 term: child welfare act E 3622
14 term: en F 3677
14 term: dies prior F 3740
14 term: child F 3916
match in: 16
match in: 17
match in: 17
match in: 17
score: 11.3
14 term: death F 3991
14 type: L fact: 4 law: 0 case: 0 stat: 1 category:
14 doc_score: 11.3 ave_score: 0.0425 norm_score: 1
\(* * * * * * * * * * * * * * * * * * * 4021\) end of: 14 *******************
15 term: jurisdiction P 4043
15 term: solicitor F 4078
15 term: law the adoption act of this E 4134
15 term: silent F 4150
15 term: jurisdiction P 4184
15 term: role F 4216
15 term: parens patriae F 4224
\(* * * * * * * * * * * * * * * * * * 4245\) end of: 15
16 term: article F 4262
16 term: jurisdiction P. 4308
16 term: parens patriae \(F 4515\)
16 term: words signify F 4660
16 term: jurisdiction \(P 4677\)
16 term: exercise \(F 4690\)
16 term: time immemorial F 4731
16 term: parent F 4758
16 term: jurisdiction P 4768
16 term: judicially administered F 4782
16 term: court acts F 4821
16 term: guardian of all infants F 4859
16 term: prerogative F 4894
16 term: delegation P 4921
16 term: exercising F 4968

16 term: jurisdiction P 4983
16 term: court acts F 5000
16 term: affectionate F 5037
16 term: careful parent for the welfare F 5056
16 term: child F 5094
16 term: welfare of the child F 5106
16 term: dominant matter F 5140
16 term: welfare F 5181
16 term: sense F 5227
16 term: 113 F 5376
\(* * * * * * * * * * * * * * * * * * 5386\) end of: 16
17 term: parens patriae F 5465
17 term: jurisdiction P 5480
17 term: emergency F 5509
17 term: child F 5542
17 term: protection F 5580
17 term: exercise F 5728
17 term: statutory power P 5744
17 term: relevant F 5930
17 term: bar F 5954
17 term: gap F 5996
17 term: legislation F 6008
17 term: child F 6156
17 term: family F 6189
17 term: child F 6237
17 term: remove F 6248
17 term: home by the authorities F 6271
17 term: allegations of child abuse F 6307
17 term: unfounded F 6356
17 term: invoking parens patriae F 6397
17 term: remove F 6464
17 term: adoption P 6475
******************* 6539 end of: 1

\section*{BUDAI v. ONTARIO LOTTERY CORPORATION ONT.S.C.}

Type: Fact
\(\backslash 003 \backslash\) On the appeal Ontario Lottery Corporation admitted that it was responsible for the error that resulted in the plaintiff being told that he had won \(\$ 835.40\). In fact 29 ticket purchasers were erroneously informed through the computer that they had won \(\$ 835.40\) when they had won only \(\$ 5\). The other 28 of those erroneously told they were \(\$ 5\) winners learned of the error before they had spent any portion of their anticipated winnings, and were paid their \(\$ 5\) prizes by the defendant.

Type: Law
\(\backslash 004 \backslash\) With respect, I disagree with the learned trial Judge that the plaintiff won \(\$ 835.40\) in the lottery draw because the computer printout so informed him. The evidence is clear that the computer printout was in error and in fact according to the rules of the Ontario Lottery Corporation created under the Ontario Lottery Corporation Act, R.S.O. 1980, c. 344, the plaintiff was the winner of \(\$ 5\) prize only. The lottery winners were determined when the draw took place. Because the plaintiff held one ticket containing three winning numbers, he won \(\$ 5\) in the draw. The fact the computer printout said he won \(\$ 835.40\) did not change the fact he won \(\$ 5\) only.

\section*{Type: Law}
\(\backslash 005 \backslash\) Following the hearing of the appeal I endorsed on the appeal book the following short reasons for varying the judgment appealed from:
"September 17, 1982
By virtue of s. 5(3) of the Ontario Lottery Corporation Act, the board of directors of the Corporation has such powers as are necessary for the purpose of carrying out its objects. It was therefore empowered to determine the rules for the Lottario lottery schemes including the rules of the \(6 / 39\) lottery.

Type: Fact
\(\backslash 006 \backslash\) The plaintiff under those rules was on February 24, 1979, the winner of a \(\$ 5\) prize only.

Log File for:

\section*{BUDAI v. ONTARIO LOTTERY CORPORATION ONT.S.C.}

3 term: responsible for the error F 1692
3 term: won F 1773
match in: 4
match in: 4
match in: 4
match in: 4
score: 24.5
3 term: ticket purchasers F 1798
3 term: erroneously informed F 1821
match in: 4
score: 9.12
3 term: computer F 1855
3 term: won F 1878
match in: 4
match in: 4
match in: 4
match in: 4
score: 24.5
3 term: won only F 1905
3 term: erroneously F 1941
3 term: winner F 1972
match in: 4
match in: 4
match in: 6
score: 17.5
3 term: error before F 1995
3 term: spent F 2017
3 term: anticipated winnings F 2045
3 term: prize F 2090
match in: 4
found in dictionary - category: canadian tax
score: 5.85
3 type: Fact: 13 law: 0 case: 0 stat: 0 category:
3 doc_score: 81.6 ave_score: 0.241 norm_score: 0.873
****************** 2121 end of: 3
4 term: won F 2202
4 term: lottery F 2221
4 term: computer printout F 2246

4 term: evidence P 2287
4 term: computer printout F 2314
4 term: error F 2340
4 term: ontario lottery corporation a E 2435
4 term: winner F 2512
4 term: prize only F 2526
4 term: lottery winners F 2543
4 term: one ticket containing F 2630
4 term: winning F 2659
4 term: won F 2679
4 term: computer printout F 2713
4 term: won F 2740
4 term: won F 2779
****************** 2797 end of: 4
5 term: appeal endorsed on the appeal F 2833
5 term: s. 5(3) ontario lottery corpo E 2990
5 term: ontario lottery corporation a E 3002
5 term: directors F 3055
5 term: powers F 3093
5 term: empowered F 3194
5 term: lottery schemes F 3251
5 term: lottery F 3299
\(* * * * * * * * * * * * * * * * * 3313\) end of: \(5^{* * * * * * * * * * * * * * * * * * * * ~}\)
6 term: winner F 3389
6 term: prize only F 3404

> MACNEILL INDUSTRIAL INC. JULIA RESOURCES CORPORATION ENEXCO INTERNATIONAL LIMITED AND CLOCKTOWER MINES LTD. v. JOHN POSNIKOFF
> B.C.Chamber Application
> Type: Law
> l004\The second issue is whether Clocktower duly remedied the default of non-payment of advance royalties by making such payment. (The plaintiffs other than Clocktower are its assignees. Notice of the assignments was not given to the defendant as provided in the option agreement. The parties therefore agree that as far as the defendant is concerned Clocktower is the only optionee and the only plaintiff eligible to obtain the declaration sought).

Type: Law
\005\ CHRONOLOGY OF EVENTS
3 May 1989 Cariboo Chilcotin Helicopters Ltd. (Cariboo) files a lien claim against some of the mineral claims covered by the option agreement.

Type: Fact
\006\} 1 4 \text { June } 1 9 8 9 \text { Clocktower is provided with written notice } that it has 30 days in which to "cure the default" of failing to keep the property free of liens.

Type: Fact
\(\backslash 007 \backslash 19\) June 1989 Mr. Simpson of Armstrong and Company
(solicitors for the plaintiffs) informs Rod Snow of Davis \& Company (solicitors for the defendant) that settlement discussions are underway which would lead to the removal of the lien claim.

Log file for:

\section*{MACNEILL INDUSTRIAL INC. JULIA RESOURCES CORPORATION ENEXCO INTERNATIONAL LIMITED AND CLOCKTOWER MINES LTD. v. JOHN POSNIKOFF \\ B.C.Chamber Application}

4 term: duly remedied the default F 1266
4 term: non-payment F 1296
4 term: royalties P 1319
found in dictionary - category: intellectual property
4 term: payment F 1344
4 term: assignees F 1400
4 term: assignment P 1427
found in dictionary - category: assigment
4 term: option agreement P 1490
match in: 5
found in dictionary - category: contract
score: 4.73
4 term: only optionee F 1598
4 term: eligible F 1636
4 type: L fact: 6 law: 3 case: 0 stat: \(0 \quad\) category:
4 doc_score: 4.73 ave_score: 0.0404 norm_score: 1
\(* * * * * * * * * \bar{*} * * * * * * * * 1685\) end of: \(4^{* * * * * * * * * * * * * * * * * * * *}\)
5 term: liens P 1808
5 term: mineral claims covered F 1839
5 term: option agreement P 1890
****************** 1913 end of: 5
6 term: written notice F 1968
6 term: cure the default F 2037
6 term: failing F 2079
6 term: property free F 2099
6 term: liens P 2137
\(* * * * * * * * * * * * * * * * * * 2149\) end of: 6

\author{
CENTRAL B.C. PLANERS LTD. KALLWEIT AND BIZICKI (PLAINTIFFS) RESPONDENTS v. HOCKER ET AL (DEFENDANTS) APPELLANTS B.C.C.A. \\ Type: evidence \\ \(\backslash 010 \backslash\) Having regard to this finding on credibility, which I am satisfied cannot be attacked successfully, and to all of the evidence \\ to which counsel referred in their able arguments and which I have studied with care, I am of the opinion that the learned judge's findings of fact quoted above are amply supported by the evidence. In the result, it seems to me that no questions of general application applicable to stockbrokers are really involved in this appeal. A decision of the case depends upon its own particular facts.
}

Type: Law
\(\backslash 011 \backslash\) I think the law to be applied is established clearly as explained in Nocton \(v\). Ashburton (Lord) [1914] AC 932, 83 LJ
Ch 784, which has been referred to with approval by the Supreme Court of Canada in London Loan Savings Co. v. Brickenden [1933] SCR 257, [1933] 3 DLR 161, affirmed [1934] 2 WWR 545, [1934] 3 DLR 465, and by this court in Jarvis v. Maguire (1961) 35 WWR 289, 28 DLR (2d)
666. In the course of his speech in Nocton v. Ashburton (Lord), supra, Lord Shaw of Dunfermline pointed out
(p. 968) the importance in such cases of ascertaining clearly the relation of the parties to each other at the time of the transaction in respect of which the claim for damage, compensation or restitution is made. He said at p. 969:
"Once, my Lords, the relation of parties has been so placed, it becomes manifest that the liability of an adviser upon whom rests the duty of doing things or making statements by which the other is guided or upon which that other justly relies and does arise irrespective of whether the information and advice given have been tendered innocently or with a fraudulent intent."

Type: tort
\012\ I have already stated that the relation of the parties in this case at the relevant time was that of broker and customer in which, to the knowledge of the broker, the customer relied on the good faith, skill and reputation of the broker in relation to the subject matter of the transaction in respect of
which the claim for damage arises. Moreover, the broker undertook to advise the customers to buy and later to hold: Vide Glennie v. McDougall \& Cowans Holdings Ltd. [1935] SCR 257, [1935] 2 DLR 561, reversing 7 MPR 544, [1934] 3 DLR 360. Viscount Haldane, L.C., in Nocton v. Ashburton (Lord), stated the matter thus at p. 948:
"... Although liability for negligence in word has in material respects been developed in our law differently from liability for negligence in act, it is none the less true that a man may come under a special duty to exercise care in giving information or advice. I should accordingly be sorry to be thought to lend countenance to the idea that recent decisions have been intended to stereotype the cases in which people can be held to have assumed such a special duty. Whether such a duty has been assumed must depend on the relationship of the parties, and it is at least certain that there are a good many cases in which that relationship may be properly treated as giving rise to a special duty of care in statement."

Type: Law
\(\backslash 013\) \In my opinion that duty arose in the present case and it has been found, on ample evidence, that Hocker failed to exercise the care which that duty requires. This duty must include
that of ascertaining and transmitting information as to facts
with reasonable accuracy.

Log file for:
CENTRAL B.C. PLANERS LTD. KALLWEIT AND BIZICKI (PLAINTIFFS) RESPONDENTS v. HOCKER ET AL (DEFENDANTS) APPELLANTS B.C.C.A.

10 term: credibility P 6947
found in dictionary - category: evidence
10 term: attacked F 6991
10 term: evidence P 7032
match in: 13
match in: 13
found in dictionary - category: evidence
score: 1.47
10 term: evidence \(\mathbf{P} 7233\)
match in: 13
match in: 13
found in dictionary - category: evidence
score: 1.47
10 term: stockbrokers F 7330
10 type: L fact: 2 law: 3 case: 0 stat: 0 category:
evidence
10 doc_score: 2.93 ave_score: 0.0101 norm_score: 1
****************** 7449 end of: 10 *******************
11 term: nocton v ashburton lord C 7525
11 term: london loan savings co v bric C 7658
11 term: affirmed F 7731
11 term: jarvis v maguire C 7799
11 term: speech F 7877
11 term: nocton v ashburton lord C 7887
11 term: importance F 7974
11 term: transaction F 8086
11 term: compensation F 8132
11 term: damage F 8132
11 term: restitution F 8156
11 term: manifest F 8279
11 term: liability P 8297
11 term: duty P 8345
11 term: advice F 8522
11 term: tendered innocently F 8545
11 term: fraudulent intent F 8575
\(* * * * * * * * * * * * * * * * * * 8600\) end of: 11
12 term: relevant time F 8684

12 term: broker F 8710
12 term: customer F 8721
12 term: broker F 8766
12 term: customer F 8778
12 term: good faith F 8803
12 term: skill F 8803
12 term: reputation of the broker F 8825
12 term: subject matter P 8869
12 term: transaction F 8891
12 term: damage F 8939
12 term: broker F 8969
12 term: undertaking P 8978
12 term: customers to buy F 9002
12 term: vide glennie v mcdougall \& co C 9040
12 term: reversing F 9125
12 term: nocton v ashburton lord C 9194
12 term: liability P 9280
12 term: negligence P 9294
12 term: liability P 9389
12 term: negligence P 9403
12 term: act F 9417
12 term: man F 9458
12 term: duty P 9487
12 term: exercise care F 9495
12 term: advice F 9538
12 term: sorry F 9575
12 term: lend countenance to the idea F 9598
12 term: intended to stereotype F 9663
12 term: people F 9709
12 term: duty P 9763
12 term: duty P 9785
12 term: rise F 9984
12 term: duty of care P 10006
****************** 10039 end of: 12
13 term: duty P 10065
13 term: ample F 10123
13 term: evidence P 10129
13 term: failed to exercise F 10151
13 term: duty P 10190
13 term: duty P 10211
13 term: transmitting information F 10255
\(* * * * * * * * * * * * * * * * * * 10324\) end of: 13 ********************

\section*{BURMAN'S BEAUTY SUPPLIES LTD. v. KEMPSTER ONT.CO.CT.}

Type: property
1007 With respect to Kempster, an experienced solicitor of some nine years' standing, he stated in chief that he asked Cassis if the plaintiff was aware of his mortgage and that he never believed that the plaintiff was relying on him for anything since his only instructions came from Cassis. However, in cross-examination, Kempster admitted knowing that Rose Lee Beauty Lounge Limited was not on sound financial ground in 1969 and 1970 and that the plaintiff's chattel mortgage would be a shaky risk since the chattels were not even worth the amount of his first mortgage. Kempster agreed that he should have been aware that the chattel mortgage contained a clause (standard in all printed chattel mortgages) that the goods and chattels which were the subject-matter of the chattel mortgage were free from liens and encumbrances, although he stated that he never paid any attention to this clause when drafting mortgages. While he attempted to convince the Court that Mrs. Burman knew or must have known about the existence of the prior mortgage, such a statement by him is difficult to accept in light of the fact that he had had previous dealings with Mrs. Burman and must have known her to be an astute businesswoman who would refuse to make a loan on worthless security, particularly since it was known to all parties that Rose Lee Beauty Lounge Limited was on shaky financial grounds.

Type: property
\(\backslash 008 \backslash\) I find, as a fact, that the plaintiff expected to receive a first chattel mortgage and that it had never been informed of the existence of the prior mortgage in favour of Mr. Kempster. I find further as a fact that the defendant Kempster either deliberately or negligently and recklessly omitted to inform Mrs.
Burman of the existence of his mortgage either directly or reciting it in the mortgage which he prepared in favour of the plaintiff. In the light of the foregoing facts it must be determined whether or not the defendant Kempster owed a duty to the plaintiff and, if so, whether by his course of conduct he was in breach of that duty.

Type: debtor and creditor
\(\backslash 009 \backslash\) As stated previously, I found that the plaintiff expected to
receive a first chattel mortgage as security for the loan it made to Rose Lee Beauty Lounge Limited. I found further that the plaintiff was unaware of the existence of the prior Kempster mortgage and, in particular, that neither Cassis nor Kempster informed the plaintiff of its existence. I found further that the plaintiff placed reliance on the clause on p. 1 of its chattel mortgage to the effect that there were no outstanding liens or encumbrances. Of the several cases referred to me by counsel I propose to mention only two. The first is Hedley Byrne \& Co. Ltd. v. Heller \& Partners Ltd., a decision of the House of Lords reported in [1964] A.C. 465. The second is Dutton v. Bognor Regis United Building Co. Ltd. et al., a decision of the English Court of Appeal reported in [1972] 1 All E.R. 462.

Type: Law
\(\backslash 010 \backslash\) Counsel for the defendant contended that there was no solicitor-and-client relationship between the plaintiff and the defendant and, hence, no duty. his contention cannot stand in light of the foregoing decisions to which I have referred.
While neither of these English cases involve a solicitor, they do set forth principles of law which are applicable in the case before me.

Log file for:

\section*{BURMAN'S BEAUTY SUPPLIES LTD. v. KEMPSTER ONT.CO.CT.}

7 term: solicitor F 5529
match in: 10
score: 3.63
7 term: standing F 5560
found in dictionary - category: criminal law
7 term: chief F 5583
7 term: mortgages P 5645
match in: 8
match in: 8
match in: 8
match in: 9
found in dictionary - category: property
score: 9.1
7 term: only instructions F 5744
7 term: admitted knowing F 5821
7 term: sound financial ground F 5886
7 term: chattel mortgage P 5952
match in: 8
match in: 9
match in: 9
found in dictionary - category: debtor and creditor
score: 16.8
7 term: shaky risk F 5981
7 term: chattels F 6002
found in dictionary - category: property
7 term: first mortgage P 6050
found in dictionary - category: property
7 term: chattel mortgage P 6123
match in: 8
match in: 9
match in: 9
found in dictionary - category: debtor and creditor
score: 16.8
7 term: clause standard F 6152
7 term: all printed F 6173
7 term: chattel mortgage P 6185
match in: 8
match in: 9
match in: 9
found in dictionary - category: debtor and creditor score: 16.8
7 term: goods F 6213
7 term: chattels F 6224
found in dictionary - category: property
7 term: subject matter P 6248
7 term: chattel mortgage P 6270
match in: 8
match in: 9
match in: 9
found in dictionary - category: debtor and creditor score: 16.8
7 term: free F 6293
7 term: liens P 6303 match in: 9
found in dictionary - category: debtor and creditor score: 3.25
7 term: encumbrance \(F 6313\) match in: 9
found in dictionary - category: property score: 6.47
7 term: clause F 6388
match in: 9
score: 2.36
7 term: drafting F 6401
7 term: mortgages P 6410
match in: 8
match in: 8
match in: 8
match in: 9
found in dictionary - category: property
score: 9.1
7 term: attempted to convince F 6430
7 term: mortgages P 6537
match in: 8
match in: 8
match in: 8
match in: 9
found in dictionary - category: property
score: 9.1
7 term: light F 6597
match in: 8
match in: 10

\section*{score: 7.99}

7 term: dealings F 6641
7 term: astute businesswoman F 6701
7 term: loan on worthless security F 6749
7 term: shaky financial grounds F 6868
7 type: L fact: 22 law: 10 case: 0 stat: 0 category:
property
7 doc_score: 118 ave_score: 0.127 norm_score: 0.913
\(* * * * * * * * * * * * * * * * * 6898\) end of: 7 ********************
8 term: chattel mortgage P 6972
8 term: mortgages P 7056
8 term: deliberately F 7153
8 term: negligently F 7169
8 term: recklessness P 7185
8 term: omitted F 7196
8 term: mortgages \(\mathbf{P} 7251\)
8 term: reciting F 7280
8 term: mortgages \(\mathbf{P} 7299\)
8 term: light F 7362
8 term: owed F 7451
8 term: duty P 7458
8 term: conduct F 7518
8 term: breach F 7537
8 term: duty P 7552
***************** 7563 end of: 8
9 term: chattel mortgage P 7648
9 term: security for the loan F 7668
9 term: mortgages P 7822
9 term: reliance P 7968
9 term: clause F 7984
9 term: chattel mortgage P 8007
9 term: liens P 8069
9 term: encumbrance F 8078
9 term: hedley byrne \& co ltd v helle C 8185
9 term: dutton v bognor regis united C 8314
\(* * * * * * * * * * * * * * * * * * 8454\) end of: \(9^{* * * * * * * * * * * * * * * * * * * *}\)
10 term: costs P 8516
10 term: duty P 8606
10 term: light F 8644
10 term: solicitor F 8751
****************** 8849 end of: 10

DIXON v. BANK OF NOVA SCOTIA ET AL. B.C.CO.CT.

\author{
Type: Fact
}

1007 Subsequently, Dixon commenced this action and later in an endeavour to mitigate his losses, sold his shares at what he felt, on advice, was as good a price as he might obtain. The difference between the purchase price and the sale price of \(\$ 1018.88\) is the subject-matter of this action.

\begin{abstract}
Type: Fact
Section Subject: Not Found
\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\# \#\#
\(1008 \backslash\) It should be noted that neither Dobie nor Dixon are novices in business. Both have obviously done well but neither have experience to any degree in the purchase of shares in the stock market. Dixon pointed out that his forte was in the field of management administration and that he made his business decisions on the advice of those whose opinions he considered to be valid and accurate; he numbered Collins among them.
\end{abstract}

Type: Fact
1009 In his defence, Mr. Collins stated that he at no time told either Dixon or Dobie that any investigation had been done on Nicola by the Bank or that the Bank was in any way, financially or otherwise, involved with Nicola. He said that he had acquired shares himself in Nicola at the initial offering, that he knew it had a prospect in Highland Valley, that he knew the officers of the company and thought highly of them, that the company was in a good financial position and that he had made a good profit himself. Mr. Collins said he had a recollection, which he said was somewhat hazy, of a meeting with Dixon and Dobie and that he had told them about Nicola with some enthusiasm and about its officers, and his own investment, and about the good reports on the assay results. He denied saying the Bank was financing the company and said he would not have used the term "endorsing" anyway because he said, "this is what goes on the back of a cheque". He said he would not have said and did not say the Bank had investigated or was investigating Nicola Copper and that his own association with Nicola was strictly a personal one. Mr.
Collins stated that he could not recall specifically any second occasion when he spoke with Dixon and Dobie about Nicola. In any case, he said, he felt that both men were sufficiently astute to
weigh any information given to them about mining stocks before acting on it. He said he himself always checks on what he called "hot tips". Mr. Collins said that he could not recall whether or not he recommended that Dobie and Dixon buy stock in Nicola, and he could not recall whether he "told them the stock was speculative or highly volatile". He said he may have told the two men that "the mine was well financed". He said that he could not have given any inference that the Bank was supporting the company, unless it could have been taken incorrectly when he said the company was well financed. Mr. Collins denied that Dixon had spoken to him subsequently in 1976 following Dixon's meeting with the regional supervisor, notwithstanding that shortly thereafter Dixon withdrew all his business except some mortgages from the Bank and that he severed all connection with the Bank as soon as those mortgages were concluded some months later.

\section*{Type: Fact}
\(\backslash 010 \backslash\) On several occasions Mr. Collins was somewhat slow in formulating his answers to questions, and was very hesitant when he finally agreed that he might give unsolicited advice about the purchase of shares. He gave me the distinct impression that he was not being completely forthright in several of his answers and did, on occasion, engage in an exercise in semantics.

Log file for:

\section*{DIXON v. BANK OF NOVA SCOTIA ET AL. B.C.CO.CT.}

7 term: mitigate F 4741
found in dictionary - category: tort
7 term: loss F 4754
7 term: sold F 4754
7 term: shares F 4771
found in dictionary - category: canadian tax
7 term: advice F 4799
7 term: price F 4821
7 term: purchase price F 4874
7 term: sale price F 4897
7 term: subject matter P 4928
7 type: F fact: 8 law: 1 case: 0 stat: 0 category:
7 doc_score: 0 ave_score: 0 norm_score: 0
******************* 4964 end of: 7
8 term: novices F 5023
8 term: business F 5035
8 term: purchase of shares F 5125
8 term: stock market F 5151
8 term: forte F 5193
8 term: management administration F 5219
8 term: business decisions on the adv F 5266
\(* * * * * * * * * * * * * * * * * * 5402\) end of: 8
9 term: investigation F 5500
9 term: financially F 5584
9 term: acquired shares F 5653
9 term: initial offering F 5695
9 term: officers F 5782
9 term: good financial position F 5862
9 term: good profit F 5909
9 term: recollection F 5957
9 term: hazy F 5999
9 term: enthusiasm F 6089
9 term: officers F 6115
9 term: investment F 6137
9 term: assay F 6184
9 term: financing F 6229
9 term: endorsing F 6294
9 term: back of a cheque F 6356

\section*{9 term: investigated F 6437}

9 term: investigating F 6457
9 term: both men F 6719
9 term: astute F 6746
9 term: mining stocks before acting F 6799
9 term: checks F 6861
9 term: hot tips F 6888
9 term: buy stock F 6993
9 term: stock F 7065
9 term: speculative F 7076
9 term: volatile F 7098
9 term: men F 7143
9 term: well financed F 7166
9 term: incorrectly F 7309
9 term: well financed F 7351
9 term: regional supervisor \(F 7472\)
9 term: withdrew all F 7539
9 term: business F 7556
9 term: mortgages P 7577
9 term: severed all connection F 7614
9 term: mortgages P 7669
****************** 7718 end of: 9
10 term: slow F 7771
10 term: hesitant F 7831
10 term: unsolicited advice F 7883
10 term: purchase of shares F 7913
10 term: distinct impression F 7948
10 term: forthright F 8002
10 term: exercise F 8075
10 term: semantics F 8087
******************* 8103 end of: 10

\title{
IN THE MATTER OF THE ADOPTION ACT, BEING CHAPTER 4 OF THE R.S.B.C., AND AMENDMENTS THERETO AND IN THE MATTER OF A MALE INFANT, BRITISH COLUMBIA BIRTH REGISTRATION NUMBER 78-09-024190 B.C.S.C.
}

> Type: Fact
> 1024 In my opinion this does not constitute a gap
> in the legislation. The intent of the legislature is clear - the particulars of the father should not be registered unless the mother consents. The legislature has set out the one way in which a natural father, not married to the mother of his child, can be registered as the child's father: both the father and mother must jointly apply for registration. Using the parens patriae jurisdiction to order registration of the father in this case would, rather than 'fill a gap' in the legislation, override the clear intent of the legislature.

\footnotetext{
Type: declatations
Section Subject: Not Found
\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#
\(1025 \backslash\) Even before Beson v. Dir. of Child Welfare, (supra) the
British Columbia Supreme Court applied a legislative gap test of sorts in Re Chappell (1977), 4 R.F.L. (2d) 3, a decision of Anderson J. (as he then was). That case involved an application to rescind an adoption order. All of the parties consented to the rescission. The Court cited Re M.J.C. (No. 1) (1972), 9 R.F.L. 241 (N.B.C.A.) at p. 18 as follows:

The Court may draw upon this residual fund of power we refer to as its 'inherent jurisdiction' whenever it is just or equitable to do so to ensure the observance of due process of law, to prevent abuse of court processes, or to do justice between parties.

\section*{\026 \(\backslash\) And continued at pp. 18-19:}

I am unable to find any exceptional circumstances in the case at bar. I am unable to find that there has been any breach of equitable principles or any abuse of process or anything of that sort.
}

Type: Fact
1027 The words 'just and equitable' do not mean that the court can rescind a final order merely because all the parties to the original proceeding are desirous of having such order rescinded.

Log file for:
IN THE MATTER OF THE ADOPTION ACT, BEING CHAPTER 4 OF THE R.S.B.C., AND AMENDMENTS THERETO AND IN THE MATTER OF A MALE INFANT, BRITISH COLUMBIA BIRTH REGISTRATION NUMBER 78-09-024190 B.C.S.C.

\section*{24 term: gap F 10867}

24 term: legislation F 10879
24 term: intent of the legislature F 10897
24 term: particulars P 10939
found in dictionary - category: criminal law
24 term: father F 10959
24 term: registered F 10980
24 term: mother consents F 11003
24 term: legislature F 11025
24 term: natural father F 11073
found in dictionary - category: family law
24 term: not married to the mother F 11073
24 term: child F 11123
24 term: registered F 11138
24 term: childs father both the father F 11156
24 term: mother F 11193
24 term: jointly F 11205
24 term: registration F 11224
24 term: parens patriae F 11249
found in dictionary - category: children
24 term: jurisdiction P 11265
found in dictionary - category: constitutional law
24 term: order registration of the fat F 11281
24 term: gap F 11356
24 term: legislation F 11368
24 term: override the clear intent F 11368
24 term: legislature F 11415
24 type: fact: 21 law: 2 case: 0 stat: 0 category:
24 doc_score: 0 ave_score: 0 norm_score: 0
****************** 11433 end of: 24
25 term: legislative gap test F 11541
25 term: rescind F 11696
25 term: consent F 11743
25 term: rescission P 11761
25 term: residual fund of power F 11892
25 term: inherent jurisdiction P 11936

\section*{25 term: equitable \(F 11983\)}

25 term: observance F 12017
25 term: prevent abuse F 12055
\(* * * * * * * * * * * * * * * * * * 12129\) end of: \(25^{* * * * * * * * * * * * * * * * * * * * ~}\)

\section*{26 term: bar F 12231}

26 term: breach of equitable F 12282
26 term: abuse F 12321
****************** 12370 end of: \(26^{* * * * * * * * * * * * * * * * * * * * ~}\)
27 term: equitable F 12397
27 term: not mean F 12411
27 term: rescind F 12440
27 term: original F 12501
27 term: desirous F 12526
27 term: order rescinded F 12550
27 type: \(F\) fact: 6 law: 0 case: 0 stat: 0 category:
27 doc_score: 7.39 ave_score: 0.0373 norm_score: 1
\(* * * * * * * * * \bar{*} * * * * * * * * 12573\) end of: 27

\author{
R. v. KAMAL BAIG
}

\author{
B.C.County Court
}

Type: Law
\(1001 \backslash\) This is an appeal by the Crown against an acquittal of the respondent in provincial court on five counts of offences alleged under the Medical Practitioners Act R.S.B.C. 1979 Chap. 254 and the Psychologists Act R.S.B.C. 1979 Chap. 342, as amended.

Type: Law
\(1002 \backslash\) The five counts alleged in the information are as follows:
"Count 1: Between the 3rd day of November, A.D. 1987 and the 10th day of November, A.D. 1988, at the District of Burnaby in the County of Westminster, Province of British Columbia, practised medicine while not registered under the Medical Practitioners Act, contrary to Section 83 of the Medical Practitioners Act.

Type: Law
\003\ Count 2: Between the 3rd day of November, A.D. 1987 and the 10th day of November, A.D. 1988, at the District of Burnaby, in the County of Westminster, Province of British Columbia, offered to practise medicine while not registered under the Medical Practitioners Act, contrary to Section 83 of the Medical Practitioners Act.

Type: Law
\(\backslash 004 \backslash\) Count 3: Between the 3rd day of November, A.D. 1987 and the 10th day of November, A.D. 1988, at the District of Burnaby, in the County of Westminster, Province of British Columbia, while not registered under the Medical

Practitioners Act held himself out under the title of "Dr." as an occupational designation relating to the treatment of human ailments contrary to Section 86(1) of the Medical Practitioners Act.

Type: Law
Section Subject: Not Found
\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#
\(\backslash 005 \backslash\) Count 4: Between the 3rd day of November, A.D. 1987 and the 10th day of November, A.D. 1988, at the District of Burnaby, in the County of Westminster, Province of British Columbia, while not registered under the Psychologists Act did engage in the practise of psychology and represented himself as a psychologist, contrary to Section 16(1) of the Psychologists Act.

Type: Law
\(\backslash 006\) Count 5: Between the 3rd day of November, A.D. 1987 and the 10th day of November, A.D. 1988, at the District of Burnaby, in the County of Westminster, Province of British Columbia while not registered under the Psychologists Act did represent himself as a psychologist, contrary to Section 16(1) of the Psychologists Act."

Type: evidence
1007 As to the offences alleged in the first four counts, the learned provincial court judge held, in reasons delivered July 17, 1989, that there was no evidence to support those charges. Alternatively, he said there was a reasonable doubt as to the respondent's guilt on counts 1 to 3 . In the result, he dismissed the first four counts for lack of evidence.

> Type: Law
> \(\backslash 008 \backslash\) As to the offence alleged under count 5 of the information, the learned provincial court judge found that the respondent did represent himself as a psychologist while he was not registered under the Psychologists Act. He held, however, that s. 16 of that Act, which prohibits an unregistered psychologist from holding himself out as a "psychologist", and s. 18 of that Act which creates certain exemptions from the application of \(s .16\), were both unconstitutional as infringements of the respondent's right to equality under s. 15 of the Canadian Charter of Rights and Freedoms.

Type: Law
1009 In oral reasons delivered October 26, 1989, the learned provincial court judge held that ss. 16 and 18 of the Psychologists Act were not saved by the provisions of s .1 of the Charter. As a result, he dismissed count 5 of the information as well.

Type: evidence
\(1010 \backslash\) The Crown's appeal is by way of trial de novo. No oral evidence was called, and the appeal was argued upon a transcript of the evidence adduced before the learned provincial court judge on March 30, May 30, June 7 and October 26, 1989, and of his reasons for acquittal delivered July 17 and October 26, 1989.

\section*{Type: Law}
\(1011 \backslash\) Counsel for the Crown says that there was ample evidence to support the offences alleged on all 5 counts of information. And he says that ss. 16 and 18 of the Psychologists Act do not contravene the Charter's guarantee of equality rights.

Type: Fact
\(1012 \backslash\) The respondent, appearing for himself, submitted written and oral arguments dealing with both the evidentiary and constitutional issues. He says the learned provincial court judge was correct in acquitting him.
\(\backslash 013 \backslash\) The respondent also advanced written and oral arguments on his "cross appeal" by which he claims relief under s.24(1) of the Charter for what he says was the provincial government's wrongful dismissal of him.

\section*{II}

Type: Law
\014\THE OFFENCES ALLEGED:
Counts 1 and 2 allege breach of \(s .83\) of the Medical
Practitioners Act. That section specifies the minimum penalties for contravention of \(s .72\) of the Act. It is \(s .72\) which creates the offences with which the respondent is charged in these two counts.

> Type: Law
> \(\backslash 015 \backslash\) Section 72 provides in part:
> "(1) A person who practises or offers to practisemedicine while not registered or while suspended from practice under this Act commits an offence."

Type: Law
\(1016 \backslash\) Section 73 limits the scope of s.72, and provides in part as follows:
"For the purposes of section 72, a person does notpractise or offer to practise medicine who ...
(i) practises psychology while registered under the Psychologists Act; ..."

Count 3 alleges an offence under s. 86 of the Medical Practitioners Act. It provides in part:
"(1) A person not registered under this Act shall notuse, assume, employ, advertise or hold himself out under the title of "doctor", "surgeon", or "physician", or any affix or prefix or abbreviation of those titles as an occupational designation relating to the treatment of human ailments...
(3) Subsection (1) does not prevent the use by a person of the title "doctor" or of the abbreviation "Dr." where the use is authorized by another Act."

Type: Law
\(\backslash 017 \backslash\) Counts 4 and 5 allege offences under s. 16 of the Psychologists
Act. Section 16 reads in part as follows:
"(1) No person shall engage in or carry on the practice of psychology and represent himself as a psychologist, unless he is registered under this Act.
(2) No person shall use, assume, or employ, or advertise or hold himself out under the title of a "registered psychologist" or "psychologist" or any affix or prefix or abbreviation of the title as an occupational designation relating to the practice of psychology, unless he is registered under this Act. ...
(4) A person who contravenes subsection (1),(2) or (3) commits an offence. ...
(7) A person represents himself as a psychologist who, for a fee or reward, monetary or otherwise, act, (sic) represents, holds himself out or advertises as a psychologist, and uses a title or description or words incorporating the word "psychology", "psychological" or "psychologist", or other terms implying training, experience or expertise as a psychologist."

Log file for:

\author{
R. v. KAMAL BAIG \\ B.C.County Court
}

1 term: acquittal F 182
found in dictionary - category: criminal law
1 term: provincial court F 214
1 term: counts of offences alleged un F 239
1 term: medical practitioners act E 277
match in: 2
match in: 3
match in: 4
score: 14.6
1 term: psychologists act E 336
1 term: amendment P 382
1 type: L fact: 3 law: 1 case: 0 stat: 2 category:
1 doc_score: 14.6 ave_score: 0.162 norm_score: 1
\(* * * * * * * * * * * * * * * * * * 396\) end of: 1
2 term: counts alleged F 412
2 term: practised medicine F 739
2 term: not registered under F 764
match in: 3
match in: 4
match in: 5
score: 15.5
2 term: medical practitioners act E 810
match in: 3
match in: 4
score: 9.73
2 term: s. 83 medical practitioners a E 870
match in: 3
score: 9.31
2 type: \(L\) fact: 3 law: 0 case: 0 stat: 2 category:
2 doc_score: 34.5 ave_score: 0.432 norm_score: 0.905
****************** 920 end of: 2 *********************
3 term: offered to practise medicine F 1196
3 term: not registered under F 1231
match in: 4
match in: 5
match in: 6
score: 15.5
3 term: medical practitioners act E 1277
match in: 4
score: 4.86
3 term: s. 83 medical practitioners a E 1337
3 type: fact: 2 law: 0 case: 0 stat: 2 category:
3 doc_score: 20.4 ave_score: 0.299 norm_score: 0.897
****************** 1408 end of: 3
4 term: not registered under F 1690
match in: 5
match in: 6
score: 10.3
4 term: medical practitioners act E 1715
4 term: title F 1810
4 term: occupational designation F 1831
4 term: treatment of human ailments c F 1893
4 term: s. 86(1) medical practitioner E 1954
4 type: L fact: 4 law: 0 case: 0 stat: 2 category:
4 doc_score: 10.3 ave_score: 0.0782 norm_score: 1
****************** 2028 end of: 4
5 term: not registered under F 2310
match in: 6
match in: 8
score: 10.3
5 term: psychologists act E 2335
match in: 6
match in: 8
match in: 8
match in: 8
score: 13.3
5 term: practise of psychology F 2392
5 term: represent F 2440
match in: 6
match in: 8
score: \(\quad 9\)
5 term: psychologist F 2465
match in: 6
match in: 8
match in: 8
score: 10.1
5 term: s. 16(1) psychologists act E 2512
match in: 6
score: 8.26
5 type: L fact: 4 law: 0 case: 0 stat: 2 category:
5 doc_score: 50.9 ave_score: 0.274 norm_score: 0.9
***************** 2578 end of: 5

6 term: not registered under F 2859
match in: 8
score: 5.16
6 term: psychologists act E 2884 match in: 8 match in: 8
match in: 8
match in: 9
score: 13.3
6 term: represent F 2927
match in: 8
score: 4.5
6 term: psychologist F 2950
match in: 8
match in: 8
score: 6.72
6 term: s. 16(1) psychologists act E 2997
6 type: L fact: 3 law: 0 case: 0 stat: 2 category:
6 doc_score: 29.6 ave_score: 0.174 norm_score: 0.84
****************** 3064 end of: 6
7 term: offences alleged F 3081
7 term: count F 3116
7 term: provincial court F 3137
match in: 8
match in: 9
match in: 10
score: 10.2
7 term: delivered F 3177
match in: 9
score: 5.34
7 term: evidence P 3221
match in: 10
match in: 10
found in dictionary - category: evidence
score: 1.44
7 term: charge F 3247
7 term: burden of proof P 3292
found in dictionary - category: evidence
7 term: guilt on counts F 3333
7 term: dismissed F 3376
match in: 9
score: 3.89
7 term: count F 3402
7 term: evidence P 3421
match in: 10
match in: 10
found in dictionary - category: evidence score: 1.44
7 type: L fact: 8 law: 3 case: 0 stat: 0 category:
evidence
7 doc_score: 22.3 ave_score: 0.0677 norm_score: 0.881
\(* * * * * * * * * * * * * * * * * * 3436\) end of: \(7^{* * * * * * * * * * * * * * * * * * * *}\)
8 term: offence alleged under count \(F 3453\)
8 term: provincial court F 3516
match in: 9
match in: 10
score: 6.82
8 term: represent F 3570
8 term: psychologist F 3593
8 term: not registered under F 3619
8 term: psychologists act E 3645
match in: 9
match in: 11
score: 6.63
8 term: psychologists act E 3702
match in: 9
match in: 11
score: 6.63
8 term: prohibits an unregistered psy F 3713
8 term: psychologist F 3784
8 term: psychologists act E 3816
match in: 9
match in: 11
score: 6.63
8 term: certain exemptions F 3835
8 term: both unconstitutional F 3889
8 term: infringements F 3915
8 term: equality under F 3959
8 term: canadian charter of rights an E 3986
match in: 9
match in: 11
score: 6.92
8 type: L fact: 11 law: 0 case: 0 stat: 4 category:
8 doc_score: 33.6 ave_score: 0.0934 norm_score: 1
\(* * * * * * * * * * * * * * * * * 4032\) end of: \(8^{* * * * * * * * * * * * * * * * * * * *}\)
9 term: oral F 4042
match in: 12
score: 5.16

9 term: delivered F 4055
9 term: provincial court F 4096
match in: 10
match in: 12
score: 6.82
9 term: ss.16 F 4129
match in: 11
score: 5.01
9 term: psychologists act E 4149
match in: 11
score: 3.31
9 term: not saved F 4173
9 term: canadian charter of rights an E 4215
match in: 11
score: 3.46
9 term: dismissed count F 4242
9 type: L fact: 6 law: 0 case: 0 stat: 2 category:
9 doc_score: 23.8 ave_score: 0.141 norm_score: 0.952
****************** 4293 end of: 9
10 term: de novo F 4338
10 term: no oral F 4338
10 term: evidence P 4357
match in: 11
match in: 12
found in dictionary - category: evidence
score: 1.44
10 term: transcript F 4411
10 term: evidence P 4430
match in: 11
match in: 12
found in dictionary - category: evidence
score: 1.44
10 term: provincial court F 4466
match in: 12
score: 3.41
10 term: acquittal delivered F 4564
match in: 12
score: 7.56
10 type: L fact: 5 law: 2 case: 0 stat: 0 category:
evidence
10 doc_score: 13.9 ave_score: 0.11 norm_score: 0.961
\(* * * * * * * * * * * * * * * * * * 4619\) end of: 10 *******************
11 term: ample F 4668
11 term: evidence P 4674
match in: 12
found in dictionary - category: evidence
score: 0.721
11 term: offences alleged F 4699
11 term: count F 4725
match in: 14
score: 4.86
11 term: ss. 16 F 4767
11 term: psychologists act E 4787
11 term: not contravene F 4808
11 term: canadian charter of rights an E 4828
match in: 13
score: 3.46
11 term: guarantee of equality rights F 4838
11 type: \(L\) fact: 6 law: 1 case: 0 stat: 2 category:

11 doc_score: 9.05 ave_score: 0.0591 norm_score: 1
******************* 4873 end of: 11 *******************
12 term: oral F 4942
match in: 13
score: 5.16
12 term: evidence P 4979
found in dictionary - category: evidence
12 term: constitutional issues F 4995
12 term: provincial court F 5040
12 term: acquitting F 5079
12 type: F fact: 4 law: 1 case: 0 stat: 0 category:
12 doc_score: 5.16 ave_score: 0.0544 norm_score: 1
****************** 5100 end of: 12
13 term: oral F 5148
13 term: claims relief under F 5198
13 term: canadian charter of rights an E 5234
13 term: provincial governments wrongf F 5267
****************** 5359 end of: 13 *******************
14 term: breach F 5416
14 term: medical practitioners act E 5438
match in: 15
match in: 16
match in: 16
match in: 16
score: 19.5
14 term: section specifies the minimum F 5472
14 term: contravention F 5517
14 term: medical practitioners act E 5546
match in: 15
match in: 16
match in: 16
match in: 16
score: 19.5
14 term: offence F 5582
14 term: charge F 5620
14 term: count F 5641
14 type: \(L\) fact: 6 law: 0 case: 0 stat: 2 category:
14 doc_score: 38.9 ave_score: 0.0685 norm_score: 1
******************* 5654 end of: 14
15 term: s. 72 medical practitioners a E 5661
match in: 16
score: 9.31
15 term: practise F 5719
15 term: offers to practisemedicine F 5732
15 term: not registered F 5765
15 term: suspended from practice under F 5789
15 term: medical practitioners act E 5835
match in: 16
match in: 16
match in: 16
score: 14.6
15 term: commits an offence F 5839
match in: 17
score: 8.26
15 type: L fact: 5 law: 0 case: 0 stat: 2 category:
15 doc_score: 32.2 ave_score: 0.0455 norm_score: 0.867
****************** 5855 end of: 15 *******************
16 term: s. 73 medical practitioners a E 5872
16 term: s. 72 medical practitioners a E 5975
16 term: notpractise F 6001
16 term: offer to practise medicine F 6016
16 term: practises psychology F 6074
16 term: registered under F 6101
match in: 17
match in: 17
score: 13.2
16 term: psychologists act E 6143
match in: 17
match in: 17
match in: 18
match in: 18
match in: 19
score: 16.6

16 term: offence under F 6192
16 term: medical practitioners act E 6218
match in: 19
score: 4.86
16 term: not registered under F 6293
16 term: medical practitioners act E 6319
match in: 19
score: 4.86
16 term: notuse F 6329
16 term: advertise F 6345
match in: 17
score: 6.28
16 term: employ F 6345
match in: 17
score: 5.75
16 term: surgeon F 6399
16 term: title of doctor F 6399
16 term: physician F 6433
16 term: affix F 6452
match in: 17
match in: 18
score: 11.1
16 term: prefix F 6467
match in: 17
match in: 18
score: 15.1
16 term: abbreviation F 6477
match in: 17
score: 8.26
16 term: title F 6499
match in: 17
match in: 18
match in: 19
score: 15.5
16 term: occupational designation F 6512
match in: 17
match in: 19
score: 13.2
16 term: treatment of human ailments F 6559
match in: 19
score: 4.86
16 term: not prevent F 6628
16 term: title doctor F 6678
16 term: abbreviation F 6703
```

    match in: }1
    score: }8.2
    16 term: authorized F }675
16 term: medical practitioners act E 6772
match in: }1
score: 4.86
16 type: L fact: 22 , law: 0 case: 0 stat: 6 category:
16 doc_score: }133\mathrm{ ave_score: 0.0533 norm_score: 0.868
******************* }6783\mathrm{ end of: 16 ********************

```

\section*{G \& C COLLISION REPAIRS (1...) LTD. PLYMOUTH v. ROGER HARTMUT JAHN \\ B.C.S.C.}

Type: Fact
\001\HOOD: In the first action, the plaintiff, G \& C Collision
Repairs (1973) Ltd. ("G \& C") sues the defendant, United Buy \& Sell Service Inc. ("United"), for damages suffered as a result of United's failure to complete the purchase of certain lands and premises situate at 7595 Kingsway, in the City of Vancouver ("the property"), pursuant to the provisions of an interim agreement dated November 20, 1985 as amended. The dealings between the parties, giving rise to the actions, were conducted, in the main, between the owner and president of G \& C, James Jakab ("Jakab"), and two licensed real estate salesmen, Ross Garner ("Garner") and Jack Rozen ("Rozen"), who were employed by G \& C's real estate agent, Montreal Trust, on the one hand, and Roger Hartmut Jahn ("Jahn"), an employee of United, on the other. United's main defence to the action is that its employee, Jahn, had no authority, actual or ostensible, to act on its behalf; and, in particular, to enter into the interim agreement on its behalf.

Type: contract
\(\backslash 002 \backslash\) In the second action, the same plaintiff, G \& C, sues United's said employee, Jahn, for damages for breach of warranty of authority. Both actions were tried at the same time, pursuant to the order of this Court dated March 17, 1988. However, the second action is an alternative one, in that G \& C only intended to proceed against Jahn if the first action was unsuccessful. Since, as these reasons will disclose, G \& C was successful in the first action, I need not deal further with the second action, other than to now dismiss it with costs to the defendant Jahn.

\section*{\003\ THE ISSUES:}

The issues raised in the pleadings or argued are as follows:
1. Did Roger Jahn have (a) actual, or (b) ostensible, authority to execute the interim agreement and subsequent documents on behalf of United?
Type: Fact
Section Subject: Not Found
\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#
\(1004 \backslash 2\). If Jahn had no authority, did United by its
subsequent conduct approve or ratify the agreement?
Type: Fact
\(\backslash 005 \backslash 3\). If the agreement was entered into, are its terms so uncertain as to make it unenforceable?
Type: Fact
\006\4. Did the agreement entered into contain a term, express or implied, that the plaintiff would obtain rezoning to permit the conduct of the defendant's business on the property, as a condition precedent, and if so did the defendant fail to fulfil that condition?
\(\ 007\) 5. Did the plaintiff fail to mitigate its damages?
Type: Law
\(\backslash 008 \backslash\) Both counsel also argued on an issue of estoppel. However, it seems to me that that issue is intertwined with both issues 1 and 2 and can be dealt with under the latter issue, if necessary. Counsel for United also raised in argument the issues that there was a duty on the part of G \& C to enquire as to the authority of Jahn, and, as well, that there was a duty on the part of \(G \& C\) to tender.
Type: evidence
\009\ THE WITNESSES:
Some of the evidence in this case is unsatisfactory. One of the problems may be that the witnesses were giving evidence about conversations and events which had taken place four years earlier. "I don't recall" was a fairly common answer to questions posed.
Type: evidence
\(1010 \backslash\) The witnesses Garner and Rozen both acknowledged that time had perhaps weakened their recollection of the details of some conversations and events. İn some cases "impressions"
instead of exact words were given. However, I believed that they both attempted to give their evidence as truthfully and as accurately as possible and I accept their evidence.

Type: evidence
\(1011 \backslash\) The two main witnesses were Jahn and the president of United, John Volken ("Volken"): Jahn, the defendant in the second action, was subpoenaed to give evidence on behalf of G \& C. He attended the trial only during the time that he was giving his evidence. He was not represented by counsel and it appeared to me that he had not prepared, or had not been prepared, for trial. However, I also believed that he gave his evidence in a truthful manner and as accurately as possible and I accept his evidence. While demeanour did not play a major role in my decision, I did not find Volken's demeanour to be particularly convincing. I found his evidence to be conflicting and much less probable than that of Jahn, whose evidence seemed to be much more consistent with the whole of the evidence. I therefore do not accept the evidence of Volken on crucial matters and, in particular, where it conflicts with that of Jahn.

Type: evidence
\(\backslash 012 \backslash\) The evidence of the other witnesses has played a lesser part in my decision. However, I accept their evidence as well.

Type: Fact
1013\ THE FACTS:
In the fall of 1985, United was looking for a new location, from which to carry on its business as a retail furniture store, to replace one of its stores, the East Hastings Street store. That store was United's smallest store and was considered to be totally inadequate at 4,000 square feet. United did not wish to renew its lease of that store which lease expired in April, 1986.

Type: Fact
\(\backslash 014 \backslash\) United had experienced considerable difficulty locating premises for lease which suited its requirements. The optimal store was approximately 15,000 square feet of bare, semiwarehouse space, with attractive exterior and located on a high count main traffic artery. The property consisted of approximately 33,000 square feet of land, and a 16,000 square
foot single storey concrete block building, with 145 feet of frontage on Kingsway Avenue. The property was considered to be ideal for United's needs.

Type: Fact
\(\backslash 015 \backslash\) The property was listed for sale with Montreal Trust, and had been for sale for some time prior to November 20, 1985, the date of the interim agreement. It was, in fact, being foreclosed by the Bank of Montreal, a mortgagee. Montreal Trust's two licensed salesmen, Garner and Rozen, both gave evidence at the trial. By that time Garner had retired after 15 years in the real estate business, 13 of them being with Montreal Trust in their Commercial Division. Rozen had spent approximately 25 years in the real estate business and had joined Montreal Trust in 1984.

Type: Fact
\016 \(\backslash\) THE EVIDENCE OF ROSS GARNER
Garner said that when the listing for the property was obtained, a sign stating "For Sale - Montreal Trust" was placed on the property. He did not believe that the sign referred to the property being available for leasing, as well as for sale. He was on duty on September 13, 1985 when Volken telephoned to make enquiries about the building and to arrange to see it. Volken seemed to be familiar with the property. He introduced himself as Mr. Volken of United.

Type: Fact
\(\backslash 017\) On the following day, Garner drove his car to United's East Hastings Street store to pick Volken up and take him to see the property. On entering the store, he met Volken, who immediately introduced Jahn to him as United's manager, either of the store or of its whole operation. He said that he may have taken it either way, but that in any event, Jahn was introduced to him as a principal in United.

Log file for:

\section*{G \& C COLLISION REPAIRS (1...) LTD. PLYMOUTH v. ROGER HARTMUT JAHN}
B.C.S.C.

1 term: damages P 533
match in: 2
found in dictionary - category: damages
score: 1.2
1 term: suffered F 541
1 term: fail F 575
1 term: purchase of certain lands F 599
1 term: premises F 630
1 term: property F 697
found in dictionary - category: property
1 term: interim agreement F 743
found in dictionary - category: property
1 term: amendment \(P 788\)
1 term: dealings F 803
1 term: rise F 840
1 term: conduct F 867
match in: 4
score: 3.8
1 term: owner F 903
1 term: president F 913
1 term: license F 965
1 term: estate salesmen F. 979
1 term: employed F 1056
found in dictionary - category: canadian tax
1 term: estate F 1081
found in dictionary - category: canadian tax
1 term: agency P 1088
found in dictionary - category: agency
1 term: one hand F 1119
1 term: employee F 1165
match in: 2
score: 5.13
1 term: employee F 1250 match in: 2
score: 5.13
1 term: no authority F 1270
match in: 4
score: 6.13
1 term: ostensible F 1295
1 term: act F 1310
1 term: interim agreement F 1368
found in dictionary - category: property


2 term: employee F 1482
2 term: damages P 1502
found in dictionary - category: damages
2 term: warranty P 1514
found in dictionary - category: contract
2 term: author F 1537
found in dictionary - category: intellectual property
2 term: only intended F 1717
2 term: unsuccessful F 1780
2 term: disclosure P 1825
found in dictionary - category: contract
2 term: dismiss F 1944
2 term: costs P 1961
found in dictionary - category: costs
2 type: L fact: 5 law: 4 case: 0 stat: 0 category:
contract
2 doc_score: 0 ave_score: 0 norm_score: 0
****************** 1995 end of: 2
3 term: pleadings P 2045
found in dictionary - category: criminal law
3 term: agency by estoppel P 2133
found in dictionary - category: agency
3 term: execute the interim agreement F 2158
3 term: document F 2209
****************** 2251 end of: 3
4 term: no authority F 2275
4 term: conduct F 2324
match in: 6
score: 3.8
4 term: ratify the agreement F 2343
4 type: F fact: 3 law: 0 case: 0 stat: 0 category:
4 doc_score: 3.8 ave_score: 0.105 norm_score: 1
****************** 2376 end of: 4
5 term: agreement F 2395
match in: 6
score: 2.21
5 term: unenforceable F 2476
5 type: F fact: 2 law: 0 case: 0 stat: 0 category:
5 doc_score: 2.21 ave_score: 0.0651 norm_score: 1
6 term: agreement F 2516
6 term: implied F 2572
6 term: rezoning F 2619
found in dictionary - category: municipal law
6 term: conduct F 2642
6 term: business on the property F 2675
6 term: condition F 2712
6 term: fail F 2767
match in: 7
score: 4.86
6 term: condition F 2787
6 type: F fact: 8 law: 0 case: 0 stat: 0 category:
6 doc_score: 4.86 ave_score: 0.0434 norm_score: 1
****************** 2803 end of: 6
7 term: fail to mitigate F 2833
found in dictionary - category: tort
7 term: damages P 2860
found in dictionary - category: damages
****************** 2874 end of: 7
8 term: estoppel P 2921
found in dictionary - category: equity
8 term: intertwined F 2976
8 term: duty P 3148
8 term: enquire F 3177
8 term: author F 3196
found in dictionary - category: intellectual property
8 term: duty P 3247
8 term: tender F 3276
found in dictionary - category: contract
8 type: L fact: 4 law: 3 case: 0 stat: 0 category:
8 doc_score: 0 ave_score: 0 norm_score: 0
****************** 3289 end of: 8 *
9 term: evidence P 3329
match in: 10
match in: 10
match in: 11
match in: 11
match in: 11
match in: 11

\section*{match in: 11}
match in: 11
match in: 11
match in: 11
match in: 12
match in: 12
found in dictionary - category: evidence
score: 8:02
9 term: unsatisfactory F 3354
9 term: evidence \(P 3430\)
match in: 10
match in: 10
match in: 11
match in: 11
match in: 11
match in: 11
match in: 11
match in: 11
match in: 11
match in: 11
match in: 12
match in: 12
found in dictionary - category: evidence
score: 8.02
9 term: conversations F 3446
match in: 10
score: 6.13
9 term: earlier dont recall F 3505
9 type: L fact: 3 law: 2 case: 0 stat: 0 category:
evidence
\(\underset{* * * * * * * * * * * * * * * * * *}{\text { doc_score: } 22.2}\) ave_score: \(0.177 \underset{* * * * * * * * * * * * * * * * * * *}{\text { norm score: }} 0.576\)
***************** 3585 end of: 9
10 term: weakened F 3664
10 term: recollection F 3679
10 term: conversations F 3716
10 term: impression F 3758
10 term: both attempted F 3839
10 term: evidence P 3868
match in: 11
match in: 11
match in: 11 .
match in: 11
match in: 11
match in: 11
match in: 11
match in: 11
match in: 12
match in: 12
found in dictionary - category: evidence
score: 6.68
10 term: evidence P 3941
match in: 11
match in: 11
match in: 11
match in: 11
match in: 11
match in: 11
match in: 11
match in: 11
match in: 12
match in: 12
found in dictionary - category: evidence score: 6.68
10 type: L fact: 5 law: 2 case: 0 stat: 0 category: evidence
10 doc_score: 13.4 ave_score: 0.0597 norm_score: 1
******************* 3956 end of: 10 *
11 term: president F 4004
11 term: subpoenaed F 4098
11 term: evidence P 4117
match in: 12
match in: 12
found in dictionary - category: evidence
score: 1.34
11 term: evidence \(P 4215\)
match in: 12
match in: 12
found in dictionary - category: evidence
score: 1.34
11 term: not represented F 4233
11 term: evidence P 4391
match in: 12
match in: 12
found in dictionary - category: evidence
score: 1.34
11 term: evidence P 4469
match in: 12
match in: 12
found in dictionary - category: evidence score: 1.34
11 term: demeanour F 4486
11 term: not play a major role F 4500
11 term: demeanour F 4564
11 term: convincing F 4593
11 term: evidence P 4618
match in: 12
match in: 12
found in dictionary - category: evidence score: 1.34
11 term: evidence P 4695
match in: 12
match in: 12
found in dictionary - category: evidence score: 1.34
11 term: evidence P 4761
match in: 12
match in: 12
found in dictionary - category: evidence score: 1.34
11 term: evidence P 4802
match in: 12
match in: 12
found in dictionary - category: evidence score: 1.34
11 type: fact: 7 law: 8 case: \(0 \quad\) stat: \(0 \quad\) category: evidence
11 doc_score: 10.7 ave_score: 0.0238 norm_score: 1
****************** 4905 end of: 11
12 term: evidence P 4916
match in: 15
found in dictionary - category: evidence
score: 0.668
12 term: played F 4952
12 term: evidence P 5015
match in: 15
found in dictionary - category: evidence
score: 0.668
12 type: L fact: 1 law: 2 case: 0 stat: 0 category:
evidence
12 doc_score: 1.34 ave_score: 0.0114 norm_score: 1
******************5039 end of: 12
13 term: fall F 5070

13 term: business F 5151
13 term: retail furniture store F 5165
13 term: replace one F 5193
13 term: store F 5212
13 term: store F 5246
13 term: store F 5259
13 term: smallest store F 5278
13 term: inadequate \(F 5327\)
13 term: square feet F 5347
match in: 14
score: 8.43
13 term: renew F 5385
13 term: lease F 5395
13 term: store F 5409
13 term: lease expired F 5422
13 type: F fact: 14 law: 0 case: 0 stat: 0 category:
13 doc_score: 8.43 ave_score: 0.0158 norm_score: 1
\(* * * * * * * * * * * * * * * * * * 5457\) end of: 13
14 term: difficulty locating premises F 5500
14 term: optimal store F 5576
14 term: semi F 5616
14 term: square feet of bare F 5616
14 term: warehouse space F 5616
14 term: exterior F 5677
14 term: located on a high count main F 5690
14 term: property F 5741
match in: 15
match in: 16
match in: 16
match in: 17
found in dictionary - category: property score: 6.92
14 term: square feet of land F 5785
14 term: square foot single storey con F 5819
14 term: feet of frontage F 5880
14 term: property F 5923
match in: 15
match in: 16
match in: 16
match in: 17
found in dictionary - category: property
score: 6.92
14 term: idea F 5954
14 type: F fact: 13 law: 0 case: 0 stat: 0 category:
property
14 doc_score: 13.8 ave_score: 0.0323 norm score: ..... 1
****************** 5985 end of: 14
15 term: property F 5996
match in: 16
match in: 16
match in: 17
match in: 18
match in: 18
found in dictionary - category: property
score: ..... 8.65
15 term: listed for sale F 6009
15 term: sale F 6064
match in: 16 ..... 16
match in: ..... 18
score: ..... 5.65
15 term: interim agreement F 6128
found in dictionary - category: property
15 term: being foreclosed F 6165
15 term: mortgages P ..... 6210
found in dictionary - category: property
15 term: licensed salesmen F 6244
15 term: evidence ..... 6292
match in: 16
found in dictionary - category: evidence
score: 0.668
15 term: retired after F 6340
15 term: estate business F 6376
15 term: spent F 6472
15 term: estate business F 6514
15 type: F fact: 10 law: 2 ..... case: 0 stat: 0
category:
property
15 doc_score: 15 ave_score: 0.0402 norm_score: 0.881
****************** 6575 end of: 15
16 term: evidence P 6586
found in dictionary - category: evidence
16 term: listing for the property F 6642
16 term: sign F 6684
16 term: property F 6744
match in: 17
match in: 18
match in: 18
found in dictionary - category: property
score: ..... 5.19
16 term: sign F 6783
16 term: property being available for F 6805
16 term: sale F. 6859
match in ..... 18
score: ..... 2.83
16 term: duty P 6876
16 term: telephone F 6916
16 term: enquiries F 6935
16 term: building F 6955
match in: 18
score: ..... 2.92
16 term: familiar F 7012
16 term: property F 7030
match in: 17
match in: 18
match in: 18
found in dictionary - category: property
score: ..... 5.19
16 type: F ..... fact: 11 law: 2property
16 doc_score: 16.1 ave_score: 0.0376 norm_score: 0.875
****************** 7093 end of: 16
17 term: drove F 7129
match in: ..... 18
score: ..... 6.42
17 term: car F 7139
17 term: store F 7177
match in: 19
match in: 19
score: ..... 7.03
17 term: property ..... 7226
match in: 18
match in: 18
found in dictionary - category: property
score: ..... 3.46
17 term: store F 7253
match in: 19
match in: 19
score: ..... 7.03
17 term: manager F 7327
17 term: store F 7351
match in: 19
match in: 19
score: ..... 7.03

\section*{17 term: whole operation F 7367}

17 type: F fact: 8 law: 0 case: 0 stat: 0 category:
17 doc score: 31 ave score: 0.125 norm score: 0.961
******************* 7514 end of: 17

\section*{LAWRENCE NESIS v. BENTER INVESTMENTS LTD., JEFF LEE AND HOMELIFE BAY CITY REALTY INC.}
B.C.S.C.

Type: Fact
\(\backslash 013 \backslash\) Each of the above conditions, if any, is, if so indicated, for the sole benefit of the party indicated. Unless each condition is waived or declared fulfilled, by written notice given by the benefiting party to the other party on or before the date specified for each condition, this contract will be thereupon terminated and the deposit returned to the Purchaser.
\014\There are no representations, warranties, guarantees, promises or agreements other than those set out herein;
\(\backslash 015 \backslash\) all of which will survive the completion of the sale.
Type: Fact
Section Subject: contract
\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#
\016\The inspection of the property took place on October 13, 1988
at approximately 5 p.m. That inspection was attended by the
plaintiff, his girlfriend (Ms. Schmidt), Mr. Melo, and Mr. Lee.

Log file for:

\section*{LAWRENCE NESIS v. BENTER INVESTMENTS LTD., JEFF LEE AND HOMELIFE BAY CITY REALTY INC. \\ B.C.S.C.}

13 term: condition F 3729
13 term: sole benefit F 3784
13 term: condition F 3839
13 term: waiver P 3852
found in dictionary - category: contract
13 term: written notice \(F 3891\)
13 term: benefit F 3919
found in dictionary - category: taxation
13 term: date specified F 3978
13 term: condition F 4002
13 term: contracts P 4024
found in dictionary - category: contract
13 term: terminate F 4051
13 type: Fact: 8 law: 2 case: 0 stat: 0 category:
contract
13 doc_score: 0 ave_score: 0 norm_score: 0
\(* * * * * * * * * * * * * * * * * * 4116\) end of: 13
14 term: no representations F 4133
14 term: warranty P 4153
14 term: guarantees F 4165
14 term: promises F 4165
14 term: agreements other F 4195

\section*{******************* 4244 end of: 14}

15 term: will survive F 4264
15 term: sale F 4299
******************* 4310 end of: 15
16 term: inspection of the property F 4321
16 term: inspection F 4410
16 term: girlfriend F 4457
******************* 4511 end of: 16

\title{
ROCHE LAKE DEVELOPMENTS LIMITED v. URBAN SYSTEMS LTD. AND HER MAJESTY THE QUEEN IN THE RIGHT OF THE PROVINCE OF BRITISH COLUMBIA \\ B.C.S.C.
}

Type: Fact
\(\backslash 010 \backslash\) On March 13, 1986, Urban wrote to Mr. Brian Ross, the plaintiff's solicitor, enclosing the drawings for site services and advising him that they had "many reservations with respect to the ability of the existing site services to adequately accommodate existing development." However, Mr. Martin testified at his examination for discovery in March 1988 that he had not instructed Mr. Ross to contact Urban and that he had not seen the letter prior to the Monday preceding the discovery. The Crown submits that, even if the principals of the plaintiff were unaware of that letter, the plaintiff was fixed with the knowledge of its agent, Mr. Ross. In my opinion, that issue should not be determined on this application.

Type: Fact
Section Subject: tort
\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#
\(\backslash 011 \backslash\) The thrust of the Crown's submission is that the loss suffered by the plaintiff, which was characterized by this Court as pure economic loss on the previous application, is not recoverable against either of the defendants.

\section*{Type: Law}
\(\backslash 012 \backslash\) The Crown submits that on the authority of Nigro v .
Agnew-Surpass Shoe Stores Ltd. et al (1977) 82 DLR (3d) 302 (Ont. HC); Bjarnarson et al v. Government of Manitoba (1987) 38 DLR (4th)
32 (Man. QB); and Bjarnarson v. Government of Manitoba (1987) 45
DLR (4th) 766 (Man. CA), the plaintiff is bound by the doctrine of issue estoppel with respect to my earlier ruling as to the nature of the loss sustained.

Type: tort
\013\ The previous application did not raise the issue of private law duties imposed upon a public authority in the discharge of its statutory duties. The Crown concedes that the existence at law of a duty of care to protect another against pure economic loss can differ from one defendant to another. However it submits that in the instant case, "there can be no triable issue in relation to
whether a 'uniquely proximate relationship' between the parties existed so as to impose upon the Crown a duty of care in relation to the pure economic loss sustained."

Log file for:

\section*{ROCHE LAKE DEVELOPMENTS LIMITED v. URBAN SYSTEMS LTD. AND HER MAJESTY THE QUEEN IN THE RIGHT OF THE PROVINCE OF BRITISH COLUMBIA B.C.S.C.}

10 term: enclosing the drawings F 3845
10 term: solicitor F 3845
10 term: site services F 3883
10 term: site services F 3993
10 term: discovery F 4112
found in dictionary - category: discovery
10 term: contact F 4176
10 term: not seen the letter prior F 4206
10 term: discovery F 4261
found in dictionary - category: discovery
10 term: principals F 4310
10 term: letter F 4360
10 term: fixed F 4386
10 term: agency P 4418
found in dictionary - category: agency
10 type: Fact: 11 law: 1 case: 0 stat: 0 category:
discovery
10 doc_score: 0 ave_score: 0 norm_score: 0
******************* 4515 end of: 10 *
11 term: thrust F 4529
11 term: loss suffered F 4574
11 term: characterized F 4617
11 term: pure economic loss P 4649
match in: 13
match in: 13
match in: 14
match in: 14
match in: 14
found in dictionary - category: tort
score: 40.4
11 term: not recoverable F 4700
11 type: F fact: 4 law: 1 case: 0 stat: 0 category:
11 doc_score: 40.4 ave_score: 0.225 norm_score: 1
******************* 4756 end of: 11
12 term: authority F 4796
match in: 14
score: 3.16
12 term: nigro v agnew-surpass shoe st C 4809
12 term: bjarnarson et al v government C 4891
12 term: bjarnarson et al v government C 4973
12 term: doctrine F 5076
12 term: estoppel P 5095
found in dictionary - category: equity
12 term: loss sustained F 5163
12 type: \(L\) fact: 3 law: 1 case: 3 stat: 0 category:
12 doc_score: 3.16 ave_score: 0.00806 norm_score: 1
******************5 5184 end of: 12
13 term: private F 5247
match in: 15
score: 5.6
13 term: duties F 5259
13 term: public authority P 5281
match in: 15
found in dictionary - category: canadian tax
score: 9.12
13 term: discharge F 5305
13 term: statutory duty P 5323
found in dictionary - category: administrative law
13 term: duty of care P 5393
match in: 15
found in dictionary - category: tort
score: 4.17
13 term: protect F 5409
13 term: pure economic loss P 5433
match in: 14
match in: 14
match in: 14
found in dictionary - category: tort score: 24.3
13 term: no triable issue F 5556
13 term: proximate relationship F 5609
13 term: duty of care P 5695 match in: 15
found in dictionary - category: tort score: 4.17
13 term: pure economic loss P 5728
match in: 14
match in: 14
match in: 14
found in dictionary - category: tort

\section*{score: 24.3}

13 term: sustained F 5747
13 type: \(L\) fact: 7 law: 6 case: \(0 \quad\) stat: \(0 \quad\) category: tort
13 doc_score: 71.6 ave_score: 0.0776 norm_score: 0.88
******************* 5764 end of: 13 *******************

\section*{CHAND AND CHAND v. SABO BROS. REALTY LTD., SABO AND GANSKE ALATA.S.C.}

Type: contract
\(1012 \backslash\) But where is the breach of that duty? The defendant
Josephine Ganske, the agent with whom the plaintiffs were dealing, took their signed offer, accepted their \(\$ 1,000\) deposit and had it presented to the vendor. The next day she called Mr. Chand and said that "the Pennos had signed the acceptance." There was an immaterial change in the offer, which the plaintiffs accepted the following day when they were given a copy of the accepted offer, Ex. 3. The plaintiffs read in the following answer given by Mrs. Ganske on her examination for discovery: "I phoned them and told them that evening that it was, their offer was accepted, and I took them for a mortgage the next day." These were true statements of facts.

Type: property
Section Subject: property
\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\#\# \#
\(1013 \backslash\) The negligence of the defendants comes from either the fact that they did not ask the Pennos if they owned the property jointly or that they took their word for the fact that they did.

\section*{Type: Law}
\(1014 \backslash\) Mrs. Ganske and Donald Sabo had taken courses given under the auspices of the Edmonton Real Estate Board. Mr. Thomas G. McCaskill testified that he lectured at such an orientation course, which all new salesmen were required to pass. He says that he always stressed the fact that the provisions of The Dower Act must be complied with by obtaining the consent of the spouse and the acknowledgement of that consent as required by the Act whenever necessary.

\section*{Type: Fact}
\(1015 \backslash\) The defendants Sabo and Ganske say that they were aware of these requirements but accepted the word of the Penno's without searching the title. Their evidence, supported by that of one Mark Dubord, a director and a former president of the Edmonton Real Estate Board, is to the effect that real estate agents do not usually search titles of property listed for sale unless they have doubt as to the vendor's ability to convey title.

Log file for:

\section*{CHAND AND CHAND v. SABO BROS. REALTY LTD., SABO AND GANSKE ALATA.S.C.}

12 term: breach of duty P 6783
found in dictionary - category: tort
12 term: agency P 6841
found in dictionary - category: agency
12 term: signed F 6898
12 term: offer and acceptance P 6905
found in dictionary - category: contract
12 term: vendor F 6971
12 term: signed F 7045
12 term: immaterial change F 7083
12 term: offer and acceptance P 7108
found in dictionary - category: contract
12 term: copy F 7186
12 term: offer and acceptance P 7199
found in dictionary - category: contract
12 term: discovery phoned F 7312
12 term: offer and acceptance P 7386
found in dictionary - category: contract
12 term: mortgages P 7428
found in dictionary - category: property
12 type: L fact: 6 law: 7 case: 0 stat: 0 category:
contract
12 doc_score: 0 ave_score: 0 norm_score: 0
\(* * * * * * * * * * * * * * * * * * 7495\) end of: 12
13 term: negligence \(P 7506\)
13 term: joint ownership P 7604
13 term: property F 7614
****************** 7694 end of: 13
14 term: auspices F 7764
14 term: lectured F 7852
14 term: orientation course F 7872
14 term: salesmen F 7907
14 term: dower act E 8005
14 term: consent of the spouse F 8059
14 term: consent F 8114
14 term: dower act E 8141
****************** 8170 end of: 14
15 term: searching the title F 8302

15 term: evidence P 8330 .
15 term: director F 8381
15 term: president F 8403
15 term: estate agents F 8476
15 term: not usually search titles of F 8494
15 term: sale F 8543
15 term: vendors ability F .8582
15 term: title F 8609
******************* 8621 end of: 15```


[^0]:    ${ }^{1}$ Lancaster, F.W.; Elliker, C.; Harkness Connell, T., "Subject Analysis," Annual review on information science and technology, v24, Elsevier Science Publishers, 1989, p35.
    ${ }^{2}$ Lewis, D., "An Evaluation of Phrasal and Clustered Representations on a Text Categorization Task", SIGIR Forum Spec. Iss., 1992, p. 37.

[^1]:    ${ }^{6}$ Luhn, H.P., "A Statistical Approach to Mechanized Encoding and Searching of Library Information," IBM Journal of Research and Development, v1, 1975, pp. 309-317.

[^2]:    ${ }^{7}$ Lancaster, F.W., Elliker, C., Harkness Connell, T,. "Subject Analysis," Annual Review on Information Science and Technology, v24, Elsevier Science Publishers, 1989, pp. 35-48.

[^3]:    ${ }^{8}$ Humphrey, S., "A Knowledge Based Expert System for Computer Assisted Indexing", IEEE Expert Magazine, v4, n3, 1989, p. 25.
    ${ }^{9}$ Driscoll, J., Rajala, D., Shaffer, W., "The Operation and Performance of an Artificially Intelligent Keywording System", Information Processing and Management, v27, n1, 1991, p. 43.

[^4]:    ${ }^{10}$ Eirund, H., "Knowledge Based Document Classification Supporting Content Based Retrieval and Mail Distribution", Network Information Processing Systems, Proceedings of the IFIP TC6/TC8 Open Symposium, May, 1988, p. 309.

[^5]:    ${ }^{11}$ Lenat, D.B.; Guha, R.V., Pittman, K., Pratt, D., Shepherd, M., "Cyc: Toward Programs with Common Sense", Communications of the ACM, v33, n8, 1990, pp. 133-138
    ${ }^{12}$ Milstead, J.L., "Methodologies for Subject Analysis in Bibliographic Databases", Information Processing and Management, v28, n3, 1992, pp. 407-431.
    ${ }^{13}$ Ginsberg, Allen, "A Unified Approach to Automatic Indexing and Information Retrieval," IEEE Expert Magazine, v8, n5, Sept. 1993, p. 46.

[^6]:    ${ }^{14}$ Cohen, P.R., Kjeldsen, R., "Information Retrieval by Constrained Spreading Activation in Semantic Networks," Information Processing and Management, v23, n4, 1987, pp. 255-268.

[^7]:    ${ }^{15}$ Hayes, P.J., Weinstein, S.P., "CONSTRUE-TIS: A System for Content-Based Indexing of a Database of News Stories," Innovative Applications of AI 2, The AAAI Press/The MIT Press, Cambridge Ma, 1991, pp. 49-64.

[^8]:    ${ }^{16}$ Driscoll, J., Rajala, D., Shaffer, W., "The Operation and Performance of an Artificially Intelligent Keywording System," Information Processing and Management, v27, n1, 1991, pp. 43-54.

[^9]:    ${ }^{17}$ Humphrey, S., "A Knowledge Based Expert System for Computer Assisted Indexing," IEEE Expert Magazine, v4, n3, 1989, p. 25.
    ${ }^{18}$ Minsky, M., "A Framework for Representing Knowledge," The Psychology of Computer Vision, P.H. Winston, ed., McGraw-Hill, New York, NY, 1975, pp. 211-277.

[^10]:    ${ }^{19}$ Hao, X., Wang, J.T.L., Bieber, M.P., Ng, P.A., "Heuristic Classification of Office Documents", InternationalJournal on Artificial Intelligence Tools [Architectures, Languages, Algorithms], v3, n2, June 1994, pp. 233-265.
    ${ }^{20}$ Masand, B., Linoff, G., Waltz, D., "Classifying News Stories Using Memory Based Reasoning," SIGIR Forum Spec. Iss., 1992, p. 59.
    ${ }^{21}{ }^{i b i d}$.

[^11]:    ${ }^{22}$ Waltz, D.L., "Memory-Based Reasoning", Natural and Artificial Parallel Computation, M.A. Arbib and J.A. Robinson eds., The MIT Press, Cambridge, Ma, 1990, pp. 251-276.
    ${ }^{23}$ Dasrathy, B.V., "Nearest Neighbour (NN) Norms: NN Pattern Classification Techniques," IEEE Computer Society Press, Los Alamitos, California, 1991.

[^12]:    ${ }^{24}$ Goodman, M., "PRISM, An AI Case-Based Text Classification System," Innovative Applications of AI, R. Smith and E. Rappaport eds., 1991, pp. 25-30.

[^13]:    ${ }^{25}$ Liddy, E.D., Paik, W., woelfel, J.K., "Use of Subject Field Codes from a MachineReadable Dictionary for Automatic Classification of Documents," Advances in Classification Research, vol. 3, Proceedings of the 3rd ASSIS SIG/CR Classification Research Workshop, Oct. 1992, pp. 83-100.

[^14]:    ${ }^{26}$ Meeter, M., Schwartz, R., Weischedel, R., "POST: Using Probabilities in Language Processing," Proceedings of the Twelfth International Conference on Artificial Intelligence, Sidney, Australia, 1991.
    ${ }^{27}$ Ward, J., "Hierarchical Grouping to Optimize an Objection Function," Journal of the American Statistical Association, v58, 1963, pp. 237-254.

[^15]:    ${ }^{28}$ Wang, J.T.L., Ng, P.A., "TEXTPROS: An Intelligent Document Processing System," International Journal of Software Engineering and Knowledge Engineering, v2, n2, June 1992, pp. 171-196.
    ${ }^{29}$ Eirund, H., "Knowledge-Based Document Classification Supporting Content-Based Retrieval and Mail Distribution," Network Information Processing Systems, Proceedings of the IFIP TC6/TC8 Open Symposium, May 1988, pp. 309-320.

[^16]:    ${ }^{30}$ Belew, R.K., "A Connectionist Approach to Conceptual Information Retrieval," Association for Computing Machinery, 1987, pp. 116-126.
    ${ }^{31}$ Debessonet, C.G., Cross, G.R., "An Artificial Intelligence Application in the Law: CCLIPS, A Computer Program that Processes Legal Information," High Technology Law Journal, 1987, p. 329.

[^17]:    ${ }^{32}$ Hahn, U., "Topic Parsing: Accounting for Text Macro Structures in Full-Text Analysis," Information Processing and Management, v26, n1, pp. 135-170.
    ${ }^{33}$ Salton, G., Allan, J., Buckley, C., "Approaches to Passage Retrieval in Full Text Information Retrieval," Proceedings of the sixteenth International ACM SIGIR Conference on Research and Development in Information Retrieval, 1993, pp. 49-58.

    Hearst, M.A., Plaunt, C., "Subtopic Structuring for Full-Length Document Access," Proceedings of the sixteenth International ACM SIGIR Conference on Research and Development in Information Retrieval, 1993, pp. 59-68.

[^18]:    ${ }^{34}$ Hearst, M.A., Plaunt, C., "Subtopic Structuring for Full-Length Document Access," Proceedings of the sixteenth International ACM SIGIR Conference on Research and Development in Information Retrieval, 1993, pp. 59-68.

[^19]:    ${ }^{35}$ Ro, J.S., "An Evaluation of the Applicability of Ranking Algorithms to Improve the Effectiveness of Full-text Retrieval," Journal of the American Society for Information Science, v39, n3, 1988, pp. 73-78.
    ${ }^{36}$ Salton, G., Buckley, C., "Automatic Text Structuring and Retrieval: Experiments in Automatic Encyclopedia Searching", Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1991, pp. 21-31.
    ${ }^{37}$ Stanfill, C., Waltx, D.L., "Statistical Methods, Artificial Intelligence, and Information Retrieval," Text-based Intelligence Systems: Current Research and Practice in Information Extraction and Retrieval, Jacobs, P.A. (ed), Lawrence Erlbaum Associates, 1992, pp. 215226.

[^20]:    ${ }^{38}$ Gelbart, D., Smith, J.C., "FLEXICON: An Evaluation of a Statistical Ranking Model Adapted to Intelligent Legal Text Management," Proceedings of the Fourth International Converence on Artificial Intelligence and the Law, 1993.

[^21]:    ${ }^{39}$ Hearst, p. 61.

[^22]:    ${ }^{40}$ Salton, G., Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer, Addison-Wesley Publishing Company, 1989.

[^23]:    ${ }^{41}$ Hearst, p. 60.

[^24]:    ${ }^{42}$ Cases of less than 20 paragraphs were not considered to be large enough to contain an identifiable substructure.

[^25]:    ${ }^{43}$ Re Adoption Act [1979] B.C.S.C. See Appendix B for case excerpt.

[^26]:    ${ }^{44}$ Budai v. Ontario Lottery Corporation [1980] Ont. S. C. See Appendix B for case excerpt.
    ${ }^{45}$ Macneill Industrial Inc., Julia Resources Corporation, Enexco International Limited and Clockwater Mines Ltd. v. John Posnikoff [1990] B.C. Chamber Application. See Appendix B for case excerpt.

[^27]:    ${ }^{46}$ Central B.C. Planers Ltd., Kallweit and Bizicki v. Hocker et al [1969] B.C.C.A. See Appendix B for case excerpt.
    ${ }^{47}$ Burman's Beauty Supplies Ltd. v. Kempster [1972] Ont. Co. Ct. See Appendix B for case excerpt.

[^28]:    ${ }^{48}$ Dixon v. Bank of Nova Scotia et al [1977] B.C. Co. Ct. See Appendix B for case excerpt.
    ${ }^{49}$ Re: Adoption Act [1979] B.C.S.C. See Appendix B for case excerpt.

[^29]:    ${ }^{50}$ R. v. Kamal Baig [1990] B.C. Co. Ct. See Appendix B for case excerpt.

[^30]:    ${ }^{51}$ G \& C Collision Repairs Ltd. Plymouth v. Roger Hartmut Jahn [1989] B.C.S.C. See Appendix B for case excerpt.

[^31]:    ${ }^{52}$ Lawrence Nesis v. Benter Investments Ltd., Jeff Lee and Homelife Bay City Realty Inc. [1990] B.C.S.C. See Appendix B for case excerpt.
    ${ }^{53}$ Roche Lake Developments Limited v. Urban Systems Ltd. and Her Majesty the Queen in the Right of the Province of British Columbia [1989] B.C.S.C. See Appendix B for case excerpt.

[^32]:    ${ }^{54}$ Chand and Chand v. Sabo Bros. Realty Ltd., Sabo and Ganske [1977] Alta. S.C. See Appendix B for case excerpt.

