CHAPTER 1

INTRODUCTION

In a common law system, which is currently prevailing in countries like India, England, and USA, decisions made by judges are important sources of application and interpretation of law. The increasing availability of legal judgments in digital form creates opportunities and challenges for both the legal community and for information technology researchers. While digitized documents facilitate easy access to a large number of documents, finding all documents that are relevant to the task at hand and comprehending a vast number of them are non-trivial tasks. In this thesis, we address the issues of legal judgment retrieval and of aiding in rapid comprehension of the retrieved documents.

To facilitate retrieval of judgments relevant to the cases a legal user is currently involved in, we have developed a legal knowledge base. The knowledge base is used to enhance the question given by the user in order to retrieve more relevant judgments. The usual practice of the legal community is that of reading the summaries (*headnotes*) instead of reading the entire judgments. A headnote is a brief summary of a particular point of law that is added to the text of a court decision, to aid readers in interpreting the highlights of an opinion. As the term implies, it appears at the beginning of the published document. Generating a headnote from a given judgment is a tedious task. Only experienced lawyers and judges are involved in this task, and it requires several man-days. Even they face difficulty in selecting the important sentences from the judgment due to its length and the variations in the judgment. In this thesis, a system has been proposed and tested for creating headnotes automatically for the relevant legal judgments retrieved for a user query. The major difficulty of interpreting headnotes generated by legal experts is that they are not structured, and hence do not convey the relative relevance of the various components of a document. Therefore, our system generates a more structured "user-friendly" headnote which will aid in better comprehension of the judgment.

In this introductory section, we motivate the choice of text summarization in a legal domain as the thesis topic. The discussion also covers the scope and objectives of the study and an overview of the work.

1.1 Motivation

Headnotes are essentially the summaries of the most-important portions of a legal judgment. Generating headnotes for legal reports is a key skill for lawyers. It is a tedious and laborious process due to the availability of a large number of legal judgments in electronic format. There is a rising need for effective information retrieval tools to assist in organizing, processing, and retrieving the legal information and presenting them in a suitable user-friendly format. For many of these larger information management goals, automatic text summarization is an important step. It addresses the problem of selecting the most important portions of the text. Moreover, a goal of information retrieval is to make available relevant case histories to the skilled users for quicker decision making. Considering these issues, we have come up with a research design as given in Figure 1.1 depicting the overall goal of our legal information retrieval system. Our aim is to bring out an end-to-end legal information retrieval system which can give a solution to legal users for their day to day activities. There are four different stages of work that have been undertaken to achieve our goal.

1. Automatic rhetorical role identification in order to understand the structure of a legal judgment.
2. Build a legal knowledge base for the purpose of enhancement of queries given by user.
3. Apply a probabilistic model for the extraction of sentences to generate a final summary.
4. Modify the final summary to a more concise and readable format.

The need of the stages (1-4) for retrieval and comprehension of legal judgments for headnote generation is briefly explained here. In recent years, much attention has been focused on the problem of understanding the structure and textual units in legal judgments. We pose this problem as one of performing automatic segmentation of a document to understand the rhetorical roles. Rhetorical roles are used to represent the collection of sentences under common titles. Graphical models have been employed in this work for text segmentation to identify the rhetorical roles present in the document. Seven rhetorical roles, namely, *identifying the case, establishing the facts of the case, arguing the case, history of the case, arguments, ratio decidendi and final decision* have been identified for this process. The documents considered for study in this thesis are from three different sub-domains viz. *rent control*, *income tax* and *sales tax* related to civil court judgments.

One of the most challenging problems is to incorporate domain knowledge in order to retrieve more relevant information from a collection based on a query given by the user. The creation of an explicit representation of terms and their relations (defined as *ontology*) can be used for the purpose of expanding the user requests and retrieving the relevant documents from a corpus. Ontologies ensure an efficient retrieval of legal resources by enabling inferences based on domain knowledge gathered during the construction of knowledge base. The documents which are retrieved in the ontology query enhancement phase will be summarized in the end for presenting a summary to the user.

Many document summarization methods are based on conventional *term weighting approach* for picking the valid sentences. In this approach, a set of frequencies and term weights based on the number of occurrences of the words is calculated. Summarization methods based on *semantic analysis* also use term weights for final sentence selection. The term weights generally used are not directly derived based on any mathematical model of term distribution or relevancy [1]. In our approach, we use a term distribution model to mathematically characterize the relevance of terms in a document. This model is then used to extract important sentences from the documents.

Another major issue to be handled in our study is to generate a "user-friendly" summary at the end. The rhetorical roles identified in the earlier phase have been used to improve the final summary. The extraction-based summarization results have been significantly improved by modifying the ranking of sentences in accordance with the importance of specific rhetorical roles. Hence, *the aim of this work is to design a text-mining tool for automatic extraction of key sentences from the documents retrieved during ontology driven query enhancement phase, by applying standard mathematical models for the identification of term patterns. By using rhetorical roles identified in the text segmentation phase, the extracted sentences are presented in the form of a coherent structured summary.* The research design used in this study is depicted in Figure 1.1

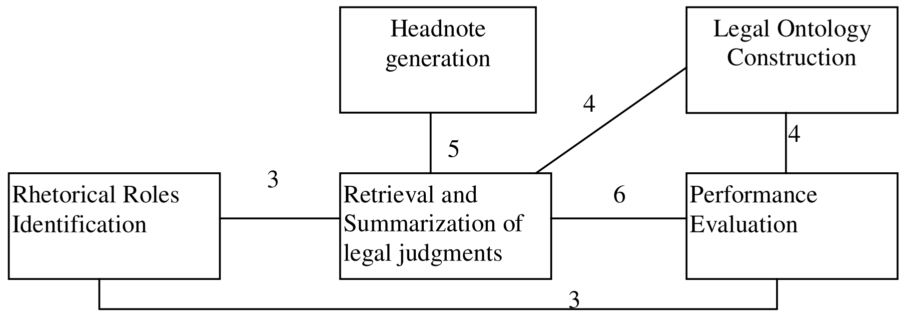


Figure 1.1 Schematic overview of the research design. The number depicted at the relationships in the scheme refer to the chapter in which the relationship is described.

1.2 Text Data Mining

Data Mining is essentially concerned with information extraction from structured databases. *Text data mining* is the process of extracting knowledge from the unstructured text data found in articles, technical reports, etc. *Data mining* [2] or *knowledge discovery in textual databases* [3], is defined by Fayyad, Piatetsky-Shapiro and Smyth (1996) as

> "*The non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data*".

Since the most natural form of storing information is text, text data mining can be said to have a higher commercial potential than those of other types of data mining. It may be seen that most part of the web is populated by text-related data. Specialized techniques operating on textual data become necessary to extract information from such kinds of collections of texts. These techniques come under the name of text mining. Text mining, however, is a much more complex task than data mining as it deals with text data that are inherently not so well structured. Moreover, text mining is a multidisciplinary field, involving different aspects of information retrieval, text analysis, information extraction, clustering, categorization, visualization, database technology, and machine learning. In order to discover and use the implicit structure (e.g., grammatical structure) of the texts, some specific *Natural Language Processing* (NLP) techniques are used. *One of the goals of the research reported in this thesis is on designing a text-mining tool for text summarization that selects a set of key sentences by identifying the term patterns from the legal document collection.*

1.3 Machine Learning

Machine learning addresses the question of how to build computer programs that improve their performance at some task through experience. It draws ideas from a diverse set of disciplines, including artificial intelligence, probability and statistics, computational complexity, information theory, psychology, neurobiology, control theory, and philosophy. Machine learning algorithms have proven to be of great practical value in a variety of application domains [4]. They are now-a-days useful in:

- Text mining problems where large text data may contain valuable implicit regularities that can be discovered automatically;
- Domains where the programs must dynamically adapt to changing conditions;
- Searching a very large space of possible hypothesis to determine the one that best-fit's the observed data and any prior knowledge provided by the experts in that area;
- Formulating general hypotheses by finding empirical regularities over the training examples;
- Providing a highly expressive representation of any specific domain.

In this thesis, application of machine learning algorithms to explore the structure of legal documents has been discussed in the context of identification of the presence of rhetorical roles, which in turn are shown to be helpful in the generation of a concise and cohesive summary.

1.4 Evolution of Legal Information Retrieval

The existence of huge legal text collections has evoked an interest in legal information retrieval research [5]. The issue is how to deal with the difficult Artificial Intelligence (AI) problem of making sense of the mass of legal information. In the late eighties and early nineties, research on logic-based knowledge systems - so-called expert systems - prevailed. Legal information retrieval was regarded as an outdated research topic in comparison with the highly sophisticated topics of artificial intelligence and law. Unfortunately, lack of practical success in the aim of replacing lawyers left the community with a lack of orientation. Now, things are seen differently and to some extent, legal information retrieval has returned to the centre of research in legal informatics. New retrieval techniques come from three different areas: integration of AI and IR, improvement of commercial applications, and large scale applications of IR on the legal corpus.

The impact of improved access to legal materials by contemporary legal information systems is weakened by the exponential information growth. Currently, information retrieval systems constitute little more than electronic text collections with (federated) storage, standard retrieval and nice user interfaces. Improvements in these aspects have to be left to the IR community. This brings the realm of legal information retrieval back into the core of research in legal informatics.

1.5 Ontology as a Query Enhancement Scheme

One of the most challenging problems in information retrieval is to retrieve relevant documents based on a query given by the user. Studies have shown, however, that users appreciate receiving more information than only the exact match to a query [6]. Depending on the word(s) given in the user's query, and with an option to choose more relevant terms which narrow the request, retrieval will be more efficient. An ontology enables the addition of such terms to the knowledge base along with all the relevant features. This will speed up the process of retrieving relevant judgments based on the user's query.

An ontology is defined as an explicit conceptualization of terms and their relationship to a domain [7]. It is now widely recognized that constructing a domain model or ontology is an important step in the development of knowledge based systems [8]. A novel framework has been identified in this study to develop a legal knowledge base. The components of the framework covers the total determination of rights and remedies under a recognized law (*acts*) with reference to *status (persons and things)* and *process (events)* having regard to the *facts* of the case. In this work, we describe the construction of a legal ontology which includes all the above components that is useful in designing a legal knowledge base to answer queries related to legal cases [9]. The purpose of the knowledge base is to help in understanding the terms in a user query by way of establishing a connection to legal concepts and exploring all possible related terms and relationships. Ontologies ensure an efficient retrieval of legal resources by enabling inferences based on domain knowledge gathered during the training stage. Providing the legal users with relevant documents based on querying the ontological terms instead of only on simple keyword search has several advantages. Moreover the user does not have to deal with document-specific representations related to the different levels of abstraction provided by the newly constructed ontology. The availability of multiple supports to ontological terms, like equal-meaning words, related words and type of relations identify the relevant judgments in a more robust way than traditional methods. In addition to these features, a *user friendly interface has been designed which can help the users to choose the multiple options to query the knowledge base. The focus of our research is on developing a new structural framework to create a legal ontology for the purpose of expanding user requests and retrieving more relevant documents in the corpora.*

1.6 Text Summarization – A new tool for Legal Information Retrieval

As the amount of on-line information increases, systems that can automatically summarize one or more documents become increasingly desirable. Recent research has investigated different types of summaries, methods to create them, and also the methods to evaluate them. Automatic summarization of legal documents is a complex problem, but it is of immense need to the legal fraternity. Manual summarization can be considered as a form of information selection using an unconstrained vocabulary with no artificial linguistic limitations. Generating a headnote (summary) from the legal document is the most needed task, and it is of immediate benefit to the legal community. The main goal of a summary is to present the main ideas in a document concisely. Identifying the informative segments while ignoring the irrelevant parts is the core challenge in legal text summarization. The document summarization methods fall into two broad approaches: extract-based and abstract-based. An extract-summary consists of sentences extracted from the document, whereas an abstract-summary may employ words and phrases that do not appear in the original document [10]. In this thesis, an extraction-based summarization has been performed on retrieved judgments based on the user query that have bearing to their present cases. It produces the gist of the judgments specific to their requirements. Thus, the user need not spend too much time by reading the entire set of judgments. *The present work describes a system for automatic summarization of multiple legal judgments. Instead of generating abstracts, which is a hard NLP task of questionable effectiveness, the system tries to identify the most important sentences of the original text, thus producing an extract.*

1.7 Objectives and Scope

The main aim of our study is to build a state-of-the-art system for automatic retrieval and summarization of legal judgments. The present investigation deals with the issues which have not been examined previously. Thus, the objectives of the present work from a technical perspective are to:

1. Apply graphical models for text segmentation by the way of structuring a given legal judgment under seven different rhetorical roles (labels).
2. Investigate whether extracted labels can improve document summarization process.
3. Propose a novel structural framework for the construction of ontology that supports the representation of legal judgments
4. Enhance the query terms mentioned in the user query to minimize the irrelevant responses.
5. Create a well-annotated corpus of legal judgments in three specific sub-domains.

6. Employ suitable probabilistic models to determine the presence of information units.
7. Generate automatic summaries of complex legal texts.
8. Create a generic structure for the summary of legal judgments belonging to different sub-domains.
9. Build an end-to-end legal judgment summarizer.

1.8 Overview of the Work

Earlier studies have shown improvement on text segmentation task by the application of graphical models like Hidden Markov Model and Maximum Entropy. These models have limitations and constraints. Hence, the search for a better method in the text segmentation task is always on. Especially in the legal domain, due to its complexity, we need a better method to understand the structure and perform useful segmentation of legal judgments. Conditional Random Fields (CRFs) model is one of the recently emerging graphical models which has been used for text segmentation problem and proved to be one of the best available frameworks compared to other existing models. Hence we have employed CRFs model for the segmentation of legal judgments. The results show much improvement compared to the standard text segmentation algorithms like SLIPPER and a simple rule-based method. The next step in our work is to help the legal community to retrieve relevant set of documents related to a particular case. For this, we have developed a new legal knowledge base with the help of a novel framework designed for this study. A legal ontology has been generated which can be used for the enhancement of user queries. In the final stage, we have used a term distribution model approach to extract the important sentences from the retrieved collection of documents based on the user query. We have used the identified rhetorical roles for reordering sentences in the final summary to generate a user-friendly summary. The overall system architecture is shown in Figure 1.2.

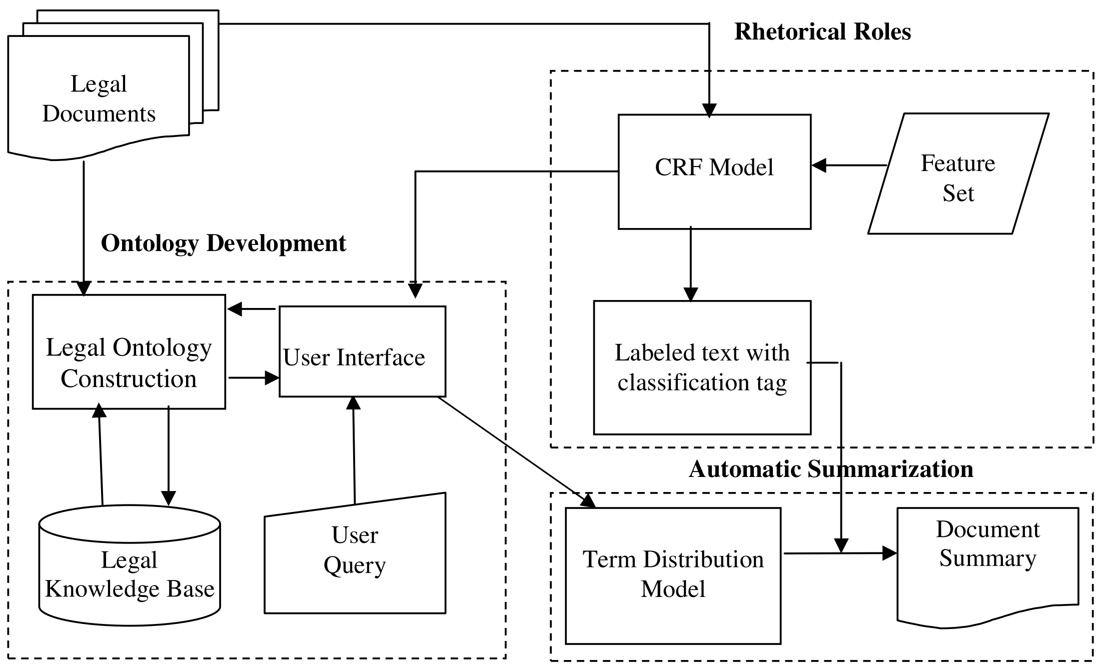


Figure 1.2 Overall system architecture of a Legal Information Retrieval System

The different stages of the proposed model were evaluated on a specific data collection spanning three legal sub-domains. The performances of our system and other automatic tools available in the public domain were compared with the outputs generated by a set of human subjects. It is found that, at different stages, our system-generated output is close to the outputs generated by human subjects, and it is better than the other tools considered in the study. Thus, the present work comprises different aspects of finding relevant information in the document space for helping the legal communities in their information needs.

1.9 Organization of the Thesis

Chapter 2 deals with a review of document summarization which includes the discussion of various types of summarization methods. The statistical approach to document summarization consists of the use of the TF-IDF method and other ad-hoc schemes, whereas, the NLP approach deals with semantic analysis, information fusion and lexical chains. It also discusses text segmentation methodologies, legal document structure identification methods, different ontology-based techniques and possible evaluation methodologies.

In Chapter 3, we discuss the use of graphical models as text segmentation tools in our approach for processing the documents and identifying the presence of rhetorical roles in legal judgments. The discussion also includes the availability of various rule learning algorithms used for text segmentation and our rule-based and CRF-based methods. Finally, our approach to text segmentation is evaluated with human annotated documents and compared with other tools. The chapter ends with a presentation of a sample annotated judgment with the help of labels identified in the text segmentation stage.

In Chapter 4, we discuss the need of a ontology, a new framework for the creation of ontology, and how an ontology is used as a query enhancement scheme. The results of ontology-based information retrieval processing are compared with a publicly available tool for query search and retrieval.

In Chapter 5, an overview of the term distribution models, the methodology adopted for term characterization and issues like the term burstiness, normalization of terms, etc., are discussed. The importance of using K-mixture model for the document summarization task is critically evaluated. The work presented here is a special case of our earlier work on multi-document summarization [11].

Chapter 6 discusses the performance measures of evaluation of an IR system and the results of tests performed to evaluate the proposed system. The probabilistic approach to document summarization method is compared with the other publicly available tools to document summarization. The performance of the auto-summarizers and that of the proposed system are compared with the human-generated summary at different ROUGE levels of summarization. Chapter 7 summarizes the work and concludes with suggestions for future work.

CHAPTER 2

A SURVEY OF SUMMARIZATION AND RETRIEVAL IN A LEGAL DOMAIN

More and more courts around the world are providing online access to judgments of cases, both past and present. With this exponential growth of online access to legal judgments, it has become increasingly important to provide improved mechanisms to extract information quickly and present rudimentary structured knowledge instead of mere information to the legal community. Automatic text summarization attempts to address this problem by extracting information content, and presenting the most important content to the legal user. The other major problem we address is that of retrieval of judgments relevant to the cases a legal user is currently involved in. To facilitate this we need to construct a knowledge base in the form of a legal ontology. In this chapter, we present the methodologies related to single document summarization based on the method of extraction of key sentences from the documents as a general approach. This chapter also explains the importance of statistical approach to automatic extraction of sentences from the documents for text summarization. We also outline the different approaches to the summarization for a legal domain, and the use of legal ontology for knowledge representation of legal terms.

2.1 Introduction to text summarization

With the proliferation of online textual resources, an increasing need has arisen to improve online access to data. This requirement has been partly addressed through the development of tools aimed at the automatic selection of portions of a document, which are best suited to provide a summary of the document, with reference to the user's interests. Text summarization has become one of the leading topics in informational retrieval research, and it was identified as one of the core tasks of computational linguistics and AI in the early 1970's. Thirty Five years later, though good progress has been made in developing robust, domain independent approaches for extracting the key sentences from a text and assembling them into a compact, coherent account of the source, summarization remains an extremely difficult and seemingly intractable problem. Despite the primitive state of our understanding of discourse, there is a common belief that a great deal can be gained for summarization from understanding the linguistic structure of the texts.

Humans generate a summary of a text by understanding its deep semantic structure using vast domain/common knowledge. It is very difficult for computers to simulate these approaches. Hence, most of the automatic summarization programs analyze a text statistically and linguistically, to determine important sentences, and then generate a summary text from these important sentences. The main ideas of most documents can be described with as little as 20 percent of the original text [12]. Automatic summarization aims at producing a concise, condensed representation of the key information content in an information source for a particular user and task. In addition to developing better theoretical foundations and improved characterization of summarization problems, further work on proper evaluation methods and summarization resources, especially corpora, is of great interest. Research papers and results of investigation reported in literature over the past decade have been analyzed with a view to crystallize the work of various authors and to discuss the current trends especially for a legal domain.

2.2 Approaches to text summarization

Generally, text summarization methods are classified broadly into two categories. One category is based on using statistical measure to derive a term-weighting formula. The other is based on using semantic analysis to identify lexical cohesion in the sentences. This approach is not capable of handling large corpora. Both the approaches finally extract the important sentences from the document collection. Our discussion will focus on the concept of automatic extraction of sentences from the corpus for text summarization task. More details of extraction-based methods are given in Section 2.4.

The summarization task can also be categorized as either generic or query-oriented. A query-oriented summary presents the information that is most relevant to the given queries, while a generic summary gives an overall sense of the document's content [12]. In addition to single document summarization, which has been studied in this field for years, researchers have started to work on multi-document summarization whose goal is to generate a summary from multiple documents that cover similar information. Next, our discussion will focus on the importance of considering the basic factors that are needed for generating a single-document summary.

Quality close to that of human-generated summaries is difficult to achieve in general, without natural language understanding. There is much variation in writing styles, document genres, lexical items, syntactic constructions, etc., to build a summarizer that will work well in all cases. Generating an effective summary requires the summarizer to select, evaluate, order, and aggregate items of information according to their relevance to a particular subject or purpose. These tasks can be approximated by IR techniques that select text spans from the document.

An ideal text summary includes the relevant information which the user is looking for and excludes extraneous and redundant information, while providing background matching with the user's profile. It must also be coherent and comprehensible which are the qualities that are difficult to achieve without deep linguistic analysis to handle issues such as co-reference, anaphora, etc. Fortunately, it is possible to exploit regularities and patterns such as lexical repetition and document structure, to generate reasonable summaries in most document genres without any linguistic processing.

There are several dimensions to summarization [13]:

- **Construct**: A natural language generated summary is created by the use of a semantic representation that reflects the structure and main points of the text, whereas an extract summary contains pieces of the original text.
- **Type**: A generic summary gives an overall sense of the document's content, whereas a query-relevant summary presents the content that is most closely related to a query or a user model.
- **Purpose**: An indicative summary gives the user an overview of the content of a document or document collection, whereas an informative summary's purpose is to contain the most relevant information, which would allow the

user to extract key information. An informative summary's purpose would be to act as a replacement for the original text.

- **Number of summarized documents**: A single document summary provides an overview of one document, whereas a multi-document summary provides this functionality for many.

- **Document length**: The length of individual documents often will indicate the degree of redundancy that may be present. For example, newswire documents are usually intended to be summaries of an event and therefore contain minimal amounts of redundancy. However, legal documents are often written to present a point, expand on the point and reiterate it in the conclusion.

- **User task**: Whether the user is browsing information or searching for specific information may impact on the types of summaries that need to be returned.

- **Genre**: The information contained in the genres of documents can provide linguistic and structural information useful for summary creation. Different genres include news documents, opinion pieces, letters and memos, email, scientific documents, books, web pages, legal judgments and speech transcripts (including monologues and dialogues).

2.3 Single Document Summarization

Automatic summarizers typically identify the most important sentences from an input document. Major approaches for determining the salient sentences in the text are term weighting approach [14], symbolic techniques based on discourse structure [15], semantic relations between words [16] and other specialized methods [17, 18]. While most of the summarization efforts have focused on single documents, a few initial projects have shown promise in the summarization of multiple documents. The concept of multi-document, multilingual and cross-language information retrieval tasks will not be discussed in this thesis.

Edmundson's Abstract Generation System (1969) [19] was the trendsetter in automatic extraction. Almost all the subsequent researchers referred to his work, and used his heuristics. At that time, the only available work on automatic extracting system was Luhn's [20] system, which used only high frequency words to calculate the sentence weights. In addition to the relative frequency approach, Edmundson described and utilized cue phrases, titles and locational heuristics, and their combinations. The evaluation is based on the comparison of computer-generated extracts against human-generated target extracts. For a sentence to be eligible for the target extract it was required to carry information about at least one of the following six types: subject matter, purpose, methods, conclusions or findings, generalizations or implications, and recommendations or suggestions. The final set of selected sentences must be coherent, and should not contain more than 20% of the original text.

All these methods are tried singly as well as in combinations. From the above studies, we understand that the automatic extraction systems need more sophisticated representations than single words. The best combination is chosen on the basis of the greatest average percentage of sentences common in the automatic extracts and the target extracts.

In another study, Salton's passage retrieval system [21], SMART, does not produce straight abstracts, but tries to identify sets of sentences (even whole sections or paragraphs), which represents the subject content of a paper. In his report, there is a brief introduction to sentence extracting, and it is stated that retrieving passages is a right step towards better response to user queries. Tombros and Sanderson present an approach to query-based summaries in information retrieval [22] that helps to customize summaries in a way which reflect the information need expressed in a query. Before building a summarization system, one needs to establish the type of documents to be summarized, and the purpose for which the summaries are required. With the above factors in mind, Tombros and Sanderson collected the documents of the Wall Street Journal (WSJ) taken from the TREC (Text Retrieval Conference) collection [23]. In order to decide the aspects of the documents which provide utility to the generation of a summary, title, headings, leading paragraph, and their overall structural organization were studied. Moreover, it was a repetition of Edmundson's work of abstract generation system, but carried out specifically for text summarization system.

Another method to summarization is based on semantic analysis of texts for sentence extraction. Linguistic processing and Lexical chains [16] are the two common approaches discussed in this regard. Linguistic information can prove useful on the basis of looking for strings of words that form a syntactic structure. Extending the idea of high frequency words, one can assume that noun phrases form more meaningful concepts, thus getting closer to the idea of terms. This overcomes several problems of the first single-word method because it can utilize compound nouns and terms which consist of adjective + noun (e.g. computational linguistics), though there is a possibility that one term can be implemented with more than one noun phrase. For example, *information extraction* and *extraction of information* refer to the same concept. But in the method of lexical chains [16], the importance of the sentence is calculated based on the importance of sequence of words that are in a lexical cohesion relation with each other, thus tending to indicate the topics in the document. It is a technique to produce a summary of an original text without requiring its full semantic interpretation, but instead relying on a model of the topic progression in the text derived from lexical chains. The algorithm computes lexical chains in a text by merging several robust knowledge sources like the WordNet thesaurus, a part-of-speech tagger, and a shallow parser. The procedure for constructing lexical chains is based on the following three-step algorithm.

- Select a set of candidate words.
- For each candidate word, find an appropriate chain relying on a relatedness criterion among the members of the chains.
- If it is found, insert the word in the chain and update it accordingly.

Some of the other methods which are in the same purview are given below:

Location method: The leading paragraph of each document should be retrieved for the formation of the summary as it usually provides a wealth of information on the document's content. Brandow *et al.* [24] suggests that,

> "Improvements *(to the auto-summaries) can be achieved by weighting the sentences appearing in the beginning of the documents most heavily*".

In order to quantify their contribution, an ordinal weight is assigned to the first two sentences of each document.

Term occurrence information: In addition to the evidence provided by the structural organization of the documents, the summarization system utilizes the number of term occurrences within each document to further assign weights to sentences. Instead of merely assigning a weight to each term according to its frequency within the document, the system locates clusters of significant words [20] within each sentence, and assigns a score to them accordingly. The scheme that is used for computing the significance factor for a sentence was originally proposed by Luhn [20]. It consists of defining the extent of a cluster of related words, and dividing the square of this number by the total number of words within this cluster.

Query-biased summaries: In the retrieved document list, if the users of IR systems could see the sentences in which their query words appeared, they could judge the relevance of documents better. Hence, a *query score* is calculated for each of the sentences of a document. The computation of that score is based on the distribution of query terms in each of the sentences. This is based on the hypothesis that larger the number of query terms in a sentence more likely those sentences convey a significant amount of information expressed through that query. The actual measure of significance of a sentence in relation to a specific query is derived by dividing the square of the number of query terms included in that sentence by the total number of the terms of the specific query. For each sentence, the score is added to the overall score obtained by the sentence extraction methods, and the result constitutes the sentence's final score.

Query-based summarization: Research on Question Answering (QA) is focused mainly on classifying the question type and finding the answer. Presenting the answer in a way that suits the user's needs has received little attention [25]. A question answering system pinpoints an answer to a given question in a set of documents. A response is then generated for this answer, and presented to the user [26]. Studies have shown however that the users appreciate receiving more information than only the exact answer [6]. Consulting a question answering system is only part of a user's attempt to fulfill the information need: it's not the end point, but some steps along what has been called a 'berry picking' process, where each answer/result returned by the system may motivate a follow-up step [27]. The user may not only be interested in the answer to a question, but also in the related information. The 'exact answer approach' fails to show leads to related information that might also be of interest to the user. This is especially true in the legal domain. Lin et al. [28] show that when searching for information, increasing the amount of text returned to the users can significantly decrease the number of queries that they pose to the system, suggesting that users utilize related information from the supporting texts.

In both the commercial and academic QA systems, the response to a question tends to be more than the exact answer, but the sophistication of their responses varies from system to system. *Exact answer*, *answer plus context* and *extensive answer* are the three degrees of sophistication in response generation [29]. So the best method is to produce extensive answers by extracting the sentences which are most salient with respect to the question, from the document which contains the answer. This is very similar to creating an extractive summarization: in both cases, the goal is to extract the most salient sentences from a document. In question answering, what is relevant depends on the user's question rather than on the intention of the writer of the document that happens to contain the answer. In other words, the output of the summarization process is adapted to suit the user's declared information need (i.e. the question). This branch of summarization has been called query-based summarization [25].

Two other studies related to mathematical approach are discussed here to strengthen the motive of using the probabilistic models in our summarization task.

(1) Neto and Santos [30] proposed an algorithm for document clustering and text summarization. This summarization algorithm is based on computing the value of the TF-ISF (term frequency-inverse sentence frequency) measure of each word, which is an adaptation of the conventional TF-IDF (term frequency – inverse document frequency) measure of information retrieval. Sentences with high values of TF-ISF are selected to produce a summary of the source text. However, the above method does not give importance to term characterization (i.e., how informative a word is). It also does not reveal the distribution patterns of the terms to assess the likelihood of a certain number of occurrences of a specific word in a document.

(2) In the Kupiec's Trainable Document Summarizer [31], which is highly influenced by Edmundson [19], document extraction is viewed as a statistical classification problem, i.e. for every sentence, its score means the probability that it can be included in a summary. This algorithm for document summarization is based on a weighted combination of features as opposed to training the feature weights

using a text corpus. In this method, the text corpus should be exhaustive to cover all the training features of the word occurrence.

The application of machine learning to prepare the documents for summarization was pioneered by Kupiec, Pedersen and Chen [31], who developed a summarizer using a Bayesian classifier to combine features from corpus of scientific articles and their abstracts. Aone et al. [32] and Lin [28] experimented with other forms of machine learning algorithms and their effectiveness. Machine learning has also been applied to learning individual features; for example, Lin and Hovy [26] applied machine learning to the problem of determining how sentence position affects the selection of sentences, and Witbrock and Mittal [33] used statistical approach to choose important words and phrases and their syntactic context. Hidden Markov Models (HMMs) and pivoted QR decomposition were used [34] to reflect the fact that the probability of inclusion of a sentence in an extract depends on whether the previous sentence has been included as well. Shen et al. [35] proposed a Conditional Random Fields (CRFs) based approach for document summarization, where the summarization task is treated as a sequence labelling problem. In our study, we used machine learning technique for segmenting and understanding the structure of a legal document. More related studies in this regard are discussed in Chapter 3.

Alternatively, a summarizer may reward passages that occupy important portions in the discourse structure of the text [36, 37]. This method requires the system to compute the discourse structure reliably, which is not possible in all genres [37]. Teufel and Moens [38] show how particular types of rhetorical relations in the genre of scientific journal articles can be reliably identified through the use of classification. MEAD [39] is an open-source summarization environment available which allows researchers to experiment with different features and methods for the single and multi-document summarization.

2.4 Approaches to automatic extraction of sentences

Automatic summarizing via sentence extraction operates by locating the best content-bearing sentences in a text. Extraction of sentences can be simple and fast. The drawback is that the resulting passage might not be comprehensible. It sacrifices the coherence of the source for speed and feasibility. Hence, we need to apply suitable methods to undertake this problem and present the summary in a more user-friendly manner.

The assumption behind extraction is that there is a set of sentences, which present all the key ideas of the text, or at least a majority of these ideas. The goal is first to identify what really influences the significance of a sentence, what makes it important. The next step is to extract important sentences based on the syntactic, semantic and discourse analysis of the text. Systems built on a restricted domain show promising results.

It is relevant to observe here that many readers usually underline, emphasize with a marker, or circle important sentences or phrases, to facilitate a quick review afterwards. Others may read only the first sentence of some paragraphs to get an idea of what the paper is about, or just look for key words/phrases (also called a scan or speed reading). This leads one to believe that an extraction method does not require a deep understanding of the natural language text.

2.4.1 Extracts vs. Abstracts

The various issues to consider in choosing between an extract-based approach and an abstract-based approach are as follows:

- The sentences of an abstract are denser. They contain implications, generalizations and conclusions, which might not be "expressed" intact in the sentences of main text.
- The language style of an abstract is generally different from the original text, especially in their syntax. Although an extract preserves the style of the writer, an abstract is dense, and is represented in a conventional style.
- The extracted sentences might not be textually coherent and might not flow naturally. It is possible that there will be fragmentary sentences, which will not make sense in the context of the extract, in spite of being important ones. Furthermore, the extract will probably contain unresolved anaphora.
- There is a chance of inconsistency and redundancy in an extract, because sentences with similar content will achieve high scores and will be extracted.

2.4.2 Basic approaches in extraction-based summarization

Typically, the techniques for automatic extraction can be classified into two basic approaches [40]. The first approach is based on a set of rules to select the important sentences, and the second approach is based on a statistical analysis to extract the sentences with higher weight.

Rule-based approach: This method uses the facts that determine the importance of sentence as encoded rules. The sentences that satisfy these rules are the ones to be extracted. Examples of rules are:

- Extract every sentence with a specified number of words from a list containing domain-oriented words.
- Extract every first sentence in a paragraph.
- Extract every sentence that has title word(s) and a cue phrase.

The drawback in this approach is that the user must provide the system with the rules which are specifically tailored to the domain they have been written for. A change of domain may mean a major rewriting of the rules.

Statistical approach: In contrast to the manual rules, the statistical approach basically tries to automatically learn the rules, that predict a summary-worthy sentence. Statistics-based systems are empirical, re-trainable systems, which minimize human effort. Their goal is to identify the units in a sentence which influence its importance, and to learn the dependency between the occurrence of units and the significance of a sentence. In this framework, each sentence is assigned a score that represents the degree of appropriateness for inclusion in a summary.

Statistical techniques for automatic extraction are very similar to the ones used for information retrieval. In the latter, each document is viewed as a collection of indices (usually words or phrases) and every index has a weight, which corresponds to the number of its appearances in the document. The document is then represented by a vector with index weights as elements. In this method, extraction of each document is treated as a collection of weighted sentences, and the highest scoring one is the final extract.

2.4.3 Factors to be considered in a system for automatic extraction

The following are the factors to be considered in the process of automatic extraction of sentences from a document collection [41].

Length of an extract: Morris et al. [42] postulate that about 20% of the sentences in a text could convey all the basic ideas about it. Since abstracts are much shorter than this proportion, the length of extracts should lie between the length of an abstract and the Morris's figure. Following are the ways of describing the length of an extract:

Proportion: The predefined percentage (usually 10%) of the number of sentences of the document should be selected. This technique is good for normally sized documents but will produce long extracts for long documents.

Oracle method: If a target extract is available, select the same number of sentences. In addition, it is intuitive that a computer extract will need more sentences than the perfect extract in order to have a good point of coverage and coherence. An advantage of the oracle method is that the system can be "trained" from the target extracts so that the optimum number of sentences can be predicted from the test documents.

Fixed number of sentences: Here the length of an extract is always the same (typically, 10-15 sentences) regardless of the size of the documents. This technique is closer to human-produced abstracts. It favours shortness, but the problems in the previous methods continue.

Sentences above a certain threshold: For a sentence to be included in the extract, it suffices to have a score which is reasonable enough. This is one way of trade-off between the extremes of the previous methods, but it requires determination of a threshold.

Mathematical formula: The number of extracted sentences is an increasing function of the number of sentences in the text, but it does not grow linearly. Hence, relatively few sentences are added when the text is big, and fewer still for a much bigger one. This is probably one of the best methods as it prevents a size explosion. It caters to huge documents as well.

Length of a sentence: It may be stated that sentences that are too short or too long are generally not ideal for an abstract, and therefore for an extract as well. This is usually referred to [31] as *sentence cut-off feature*. It penalizes short (less than 5-6 words) and long sentences either by reducing their score, or by excluding them completely.

In our work, we focus on single-document sentence extraction method which forms the basis for other summarization tasks and which has been considered as a hot research topic [43].

2.5 Legal document summarization – An overview

Law judgments form the most important part of a lawyer's or a law student's study materials. These reports are records of the proceedings of a court, and their importance derives from the role that precedents play in any common law system, including Indian law. In order to find a solution for legal problems that are not directly covered by the notified laws, lawyers look into previous judgments for possible precedents. Legal users constitute a law jurisprudence precedent from which it is possible to extract a legal rule that can be applied to similar cases. One reason for the difficulty in understanding the main theme of a legal case is the complexity of the domain, specific terminology of the legal domain and legal interpretations of expressions producing many ambiguities. Currently, selected judgments are manually summarized by legal experts. The ultimate goal of legal summarization research would be to provide clear, non-technical summaries of legal judgments.

Legal document Summarization is an emerging subtopic of summarization specific to legal domain. Legal document summarization poses a number of new challenges over general document summarization. The discussion in this section outlines some of the methods used for the summarization of legal documents. The usefulness of these methods and outcomes have also been described.

SUM Project: SUM is an EPSRC research project of the Language Technology Group, based in the Institute for Communicating and Collaborative Systems of Edinburgh's School of Informatics [44]. This project uses summarization to help address the information overload problem in the legal domain. The main focus of this project is the sentence extraction task and methods of structuring summaries. It has been argued that most practically oriented work on automated summarization can be described as based on either *text extraction* or *fact extraction*. In these terms, the Teufel & Moens [38] approach can be characterized as *augmented* text extraction: the system creates summaries by combining extracted sentences, but the sentences in the source texts are first categorized to reflect their role in the rhetorical or argumentative structure of the document. This rhetorical role information is used to guide the creation of the summaries and to permit several summaries to be created for a document, of which each one is tailored to meet the needs of a different class of users. The system performs automatic linguistic annotation of a small sample set. The hand-annotated sentences in the set are used in order to explore the relationship between linguistic features and argumentative roles. The HOLJ Corpus [45] is used in this work which comprise of 188 judgments delivered in the years 2001-2003 taken from the House of Lords website. The entire corpus was automatically annotated with a wide range of linguistic information using a number of different NLP components: part-of-speech tagging, lemmatization, noun and verb group chunking, named entity recognition (both general and domain-specific), clause boundary identification, and main verb and subject identification. The approach used in this study can be thought of as a more complex variant of template filling, where the slots in the template are high-level structural or rhetorical roles, and the fillers are the sentences extracted from the source text using a variety of statistical and linguistic techniques exploiting indicators such as cue phrases. Feature set includes elements such as location of a sentence within the document and its subsections and paragraphs, cue phrases, information on whether the sentence contains named entities, sentence length, average

TF-IDF term weight, and data on whether the sentence contains a quotation or is inside a block quote. Maximum entropy model has been used for sequence labelling framework [44]. The rhetorical roles identified in the study are *Fact, Proceedings, Background, Proximation, Distancing, Framing* and *Disposal*. The details of these roles are given in Chapter 3 in which we also discuss the importance of identifying different set of roles for legal judgments which are relevant to Indian Court judgments.

Summary Finder: This study [46] leverages the repetition of legal phrases in the text by using graph-based approach. The graphical representation of the legal text is solely based on similarity function between sentences. The similarity function as well as the voting algorithm used on the derived graph representation is different from other graph-based approaches (e.g. LexRank). In general, for legal text, some paragraphs summarize the entire text or at least parts of the text. In order to find such paragraphs, this method computes inter-paragraph similarity scores and selects the best match for every paragraph. The system acts like a voting system where each paragraph casts a vote for another paragraph (its best match). The top paragraphs with most votes were selected as the summary. The vote casting can be seen as a similarity function based on phrase similarity. Phrase similarity is computed by looking for phrases that co-occur in two paragraphs. The longer the matched phrase, higher the score will be.

LetSum (Legal Text Summarizer): This is a prototype system [47] which determines the thematic structure of a legal judgment along four themes: *Introduction,* *Context, Judicial Analysis* and *Conclusion.* LetSum is used to produce short summaries for legal decision of the proceedings of federal courts in Canada. This method investigates the extraction of the most important units based on the identification of the thematic structure in the document and the determination of argumentative themes of the textual units in the judgment [47]. The generation of summary is done in four steps: thematic segmentation to detect legal document structure, filtering to eliminate unimportant quotations and noises, selection of the candidate units and production of structured summary. The presentation of the summary is in a tabular form along with the themes of the judgment.

FLEXICON: The FLEXICON project [48] generates a summary of legal cases by using information retrieval based on location heuristics, occurrence frequency of index terms, and the use of indicator phrases. A term extraction module that recognizes concepts, case citations, statute citations, and fact phrases leads to the generation of a document profile. This project was developed for the decision reports of Canadian courts.

SALOMON: Moens [49] automatically extracts informative paragraphs of text from Belgian legal cases. SALOMON extracts relevant text units from the case text to form a case summary. Such a case profile facilitates the rapid determination of the relevance of the case or may be employed in text search. Techniques are developed for identifying and extracting relevant information from the cases. A broader application of these techniques could considerably simplify the work of the legal profession. In this project a double methodology was used. First, the case category, the case structure, and irrelevant text units are identified based on a knowledge base represented as a text grammar. Consequently, general data and legal foundation concerning the essence of the case are extracted. Secondly, the system extracts informative text units of the alleged offences and of the opinion of the court based on the selection of representative objects.

2.6 Legal Ontology – an Overview

The potential of knowledge-based technological support for work in the legal domain has become widely recognized in recent time. In this connection, we discuss different ontology projects available that provides linguistic information for large amount of the legal text.

The CORTE Project: The goal of CORTE [50] is to provide knowledge-based support using techniques from computational linguistics based on a sound theoretical understanding of the creation, semantics, and use of legal terminology. In particular, the project aims at:

- Developing a linguistic model of definitions in legal text
- Building computational linguistic tools for the automatic extraction of such definitions
- Exploring methods for the exploitation of the extracts in terminologies for the legal domain

In this work, a corpus of more than 8 million German legal documents provided by *juris* GmbH, Saarbrücken is used. In order to analyze these documents grammatically, a semantically-oriented parsing system has been developed in the COLLATE project (Computational Linguistics and Language Technology for Real Life Applications, funded by the German Ministry for Education and Research) at the Saarbrücken CL group [50] (initially applied to newspaper texts). The system balances depth of linguistic analysis with robustness of the analysis process, and is therefore able to provide relatively detailed linguistic information for large amounts of text. To deal with the problem of ambiguity it makes use of syntactic under specification. Under certain conditions, it commits only to the established common parts of alternative syntactic analyses. Done this way, later processing steps are enabled to access at least partial information without having to settle for one syntactic reading. The most important fact is that the system is semantically oriented. It not only analyzes the grammatical structure of the input, but also provides an abstract representation of its meaning (a so-called partially resolved dependency structure or PREDS).

For instance, active and passive sentences receive identical representations, so that their common semantic content becomes accessible for further processing. PREDS-parsing system is adapted for the domain of legal documents. Starting off from a collection of definitions compiled relying on legal expert knowledge, an annotation scheme has been devised for marking up the functional parts of these definitions. This scheme has plans for extensions to encode information regarding external relations such as rhetorical and argumentative function of definitions and citation structure, and it will be applied in the collection of further data. At the same time, a detailed linguistic analysis of definition instances has been worked out.

The main aim in this work is to develop taxonomy of definition types according to semantic functions and syntactic realization. The syntactic-semantic information made accessible by the PREDS system will facilitate the automatic recognition and extraction of definitions by providing an additional level of structure besides the syntactic surface. Extracted definitions can then be used to validate the taxonomy. More importantly, the information contained in the PREDS constructed will be used to organize the collected extraction results within a semi-structured knowledge base. In particular it will serve to automatically segment and classify extracted definitions according to the taxonomy developed based on linguistic cues. The resulting knowledge base will contain the extracted text passages along with rich additional information that allows the user to navigate through the collected definitions according to their needs, e.g. sorted by concept defined, grouped by type of definition, or following citations. A very promising part of the work is that it uses the information provided by the PREDS based definition extraction system to actually update and enlarge the existing formalized ontologies. Languages based on description logics (DL) [51] have emerged as the standard framework for the specification of such formalized ontologies.

The central question to be pursued is therefore how to model the semantic effect of definitions within this formalism. Moreover, with the organization of DL knowledge bases around atomic concepts that are incrementally characterized semantically by adding constraints, the framework is especially interesting for the modeling of "open-texture", i.e. under defined or vague concepts and their incremental specification. Building on a linguistically well-founded understanding of definitions together with automatic definition of extraction methods, it will be possible to approach this topic empirically.

Functional Ontology: Valente [52] developed a legal ontology based on a functional perspective of the legal system. He considered the legal system as an instrument to change by influencing the society in specific directions by reacting to social behavior. The main functions can be decomposed into six primitive functions each of which corresponds to a category of primitive legal knowledge

a) Normative knowledge – which describes states of affairs which have a normative status (such as forbidden or obligatory);

b) World knowledge – which describes the world that is being regulated, in terms that are used in the normative knowledge, and so can be considered as an interface between common-sense and normative knowledge;

c) Responsibility knowledge – the knowledge which enables responsibility for the violation of norms to be ascribed to particular agents;

d) Reactive knowledge – which describes the sanctions that can be taken against those who are responsible for the violation of norms;

e) Meta-legal knowledge – which describes how to reason with other legal knowledge.

f) Creative knowledge – which states how items of legal knowledge are created and destroyed.

This ontology forms the basis of a system ON-LINE [52] which is described as a Legal Information Server. ON-LINE allows for the storage of legal knowledge as both text and an executable analysis system interconnected through a common expression within the terms of the functional ontology. The key thrust of this conceptualization is to act as a principle for organizing and relating knowledge, particularly with a view to conceptual retrieval. Two limitations are noted by Valente in this work. The first is practical - that performing the modeling that is required to follow through this conceptualization is very resource intensive. Although the Ontolingua [53] description of the different kinds of legal knowledge seems relatively complete, the domain model constructed within this framework for the ON-LINE system is rather restricted. Valente writes:

> *While it is expected that the ontology is able to represent adequately legal knowledge in several types of legislation and legal systems, this issue was not yet tested in practice.*

Frame Based Ontology: Kralingen and Visser [54] discuss the desire to improve development techniques for legal knowledge systems, and in particular to enhance the reusability of knowledge specifications by reducing their task dependency. This work distinguishes between an ontology which is intended to be generic to all law, and a statute-specific ontology which contains the concepts relevant to a particular legal domain. This ontology has been used as the basis for the system FRAMER which addresses two applications in Dutch Unemployment Benefit Law, one involving a classification task determining entitlement to Unemployment Benefit and the other a planning task, determining whether there is a series of actions which can be performed to bring about a certain legal consequence.

Visser [53] builds a formal legal ontology by developing a formal specification language that is tailored in the appropriate legal domain. Visser commenced by using Kralingen's theory of frame-based conceptual models of statute law [55]. Visser uses the terms ontology and specification language interchangeably, and claims that an ontology must be:

1) Epistemologically adequate
2) Operational
3) Expressive
4) Reusable
5) Extensible

Visser chose to model the Dutch Unemployed Benefits Act of 1986. He created a CommonKADS expertise model [54], specifying domain knowledge by:

i) Determining the universe of discourse by carving up the knowledge into ontological primitives. A domain ontology is created with which the knowledge from the legal domain can be specified.
ii) Domain specification is created by specifying a set of domain models using the domain ontology.

Legal ontology from a European Community Legislative Text: This work [56] presents the building of a legal ontology about the concept of employees' rights in the event of transfers of undertakings, businesses or parts of undertakings or businesses in the European community legislation text. The construction is achieved both by building the ontology from texts by using the semi-automatic TERMINAE method

[56] and aligning it with a top-level ontology. TERMINAE is based on knowledge elicitations from text, and allows creating a domain model by analyzing a corpus with NLP tools. The method combines knowledge acquisition tools based on linguistics with modeling techniques so as to keep the links between models and texts. During the building process [56], it is assumed that:

(1) The ontology builder should have a comprehensive knowledge of the domain, so that she/he will be able to decide which terms (nouns, phrases, verbs or adjectives) are domain terms and which concepts and relations are labeled with these domain terms;

(2) The ontology builder knows well how the ontology will be used. The alignment process takes place during the construction.

Biébow [57] defined ontology alignment as follows: ontology alignment consists in establishing links between ontologies and allowing one aligned ontology to reuse information from the other. In alignment, the original ontologies persist, with links established between them. Alignment usually is performed when the ontologies cover complementary domains. This ontology is structured around two central ontologies DOLCE [58] and LRI-Core [59]. The resulting ontology does not become part of the DOLCE ontology but uses its top-level distinction. The process of ontology alignment was carried out during the ontology construction and was performed mostly by hand, with the TERMINAE tool. TERMINAE provides easy import of concepts among DOLCE but doesn't check whether consistency is maintained after the performed operations. The alignment process in this case included the following activities: the identification of the content that overlapped with the core ontology; the concepts that were at the top level became subclasses of more general concepts. The concepts are defined from the study and interpretation of the term occurrences in the directive. The term properties (structural and functional) are translated into a restricted language. This translation was realized by hand. The linguistic criteria for identifying these properties remain to be defined for automating this process.

The studies discussed above illustrate that the ontologies are developed for particular purposes. Therefore, a new legal ontology should be developed for query enhancement which is considered as an important information retrieval task in our study.

2.8 Summary

Automatic Summarization helps lawyers and persons needing legal opinions to harness the availability of vast legal resources in a more effective way. In this chapter, a review of the automatic text summarization for single documents for legal domain was presented. The issues related to term-weighting and semantic analysis of text, the two main methods of summarization were discussed. Factors considered for the extraction of sentences towards text summarization were also discussed. Legal document summarization related papers were explored to evolve a new method to perform summarization of legal judgments.

Based on the review of the research work presented, we may note that legal documents are having complex structure, and so we need to segment the document to understand the presence of various roles. Also, we may note that term-weighting systems are not directly derived from any mathematical model of term distribution. Moreover, they are not specific in assessing the likelihood of a certain number of occurrences of a particular word in a document collection. Hence, we have attempted some new techniques to produce a coherent and consistent summary. The following procedures are adopted in this task.

- We used CRF model for the identification of rhetorical roles in legal judgments.
- We used term distribution model for the identification of term patterns and frequencies of the terms.
- We have developed a novel structural framework for the construction of legal ontology.
- Extraction-based summarization usually suffers from coherence problems. We used the identified roles during post processing to avoid the coherence problem.
- In order to make the final output more user-friendly and concise, we have generated a table-style structured summary.

The evaluation part of our study deals with the following methods by considering human referenced outputs as gold standard:

- Comparison of our rhetorical role identification method with rule-based and standard segmentation algorithm.
- Comparison of our ontology-based query enhancement scheme with standard query search method.
- Comparison of our summarizer with the public-domain summarizers, and with reference to the human-generated summaries.

- Arriving at a threshold level of summarization with respect to the human-generated summary.

The remaining chapters discuss our work based on text segmentation, creation of legal ontology and the application of a term distribution model for text summarization by focusing on informative summaries using extracts.

CHAPTER 3

IDENTIFICATION OF RHETORICAL ROLES IN LEGAL DOCUMENTS

Automatic identification of rhetorical roles in a legal document is the most important task in our work. It is a part of genre analysis to be carried out to understand the meaningful textual contents. Generally, a document is segmented into coherent paragraphs known as rhetorical roles. For example, *aim*, *basis* and *contrast* are the basic rhetorical roles of scientific articles. Text segmentation problem focuses on how to identify the role boundary, where one region of text ends and another begins, within a document. The current work was motivated by the observations that such a seemingly simple problem can actually prove quite difficult to automate [60] and that a tool for partitioning a stream of undifferentiated text into coherent regions would be needed to understand the structure of a legal document. Legal judgments are complex in nature and it is difficult to track the presence of different topics (rhetorical schemes). Automatic segmentation of legal text focuses on the identification of key roles, so that they may then be used as the basis of alignment of sentences at the time of final summary generation.

In this chapter, we review the state-of-the-art graphical models for segmentation and role identification. The problem of segmenting structured entities from unstructured data is an extensively researched topic. A number of models has been proposed ranging from the earliest rule-learning methods to probabilistic approaches based on generative models like *Hidden Markov Models(HMM)* [61], and conditional models like *Maximum Entropy Markov model(MEMM)* [62]. We employ undirected graphical models for the purpose of automatic identification of rhetorical roles in legal judgments. To accomplish this task, we apply *Conditional Random Fields* (CRFs) which have been shown to be efficient at text segmentation [63]. In this chapter, we present results related to text segmentation task using Conditional Random Fields, and discuss several practical issues in applying CRFs to information retrieval tasks in general. Using manually annotated sample documents pertaining to three different legal sub-domains (*rent control*, *income tax,* and *sales tax*), we train three different CRF models to segment the documents along different rhetorical structures. With CRFs, we provide a framework for leveraging all the relevant features like indicator phrases, named entities, upper case words, etc., even if they are complex, overlapping and not independent. The CRF approach draws together the advantages of both finite state HMM and conditional MEMM techniques by allowing the use of arbitrary, mutually dependent features and joint inferences over entire sequences. Finally, it is helpful in document summarization in the form of re-ordering the ranking in the final extraction-based summary based on the identified roles. This process could generate a single document summary as shown in Figure 3.1. The details of the extraction of sentences using term distribution model will be discussed in chapter 5.

In this chapter, we discuss the need for graphical models and its various types and applications related to the segmentation of a legal text. For the task of segmenting legal documents, rule-based as well as CRF-based methods are employed. Finally, the effectiveness of our approach is established by comparing the experimental results of our proposed methods with those of SLIPPER, which is a standard rule learner method.

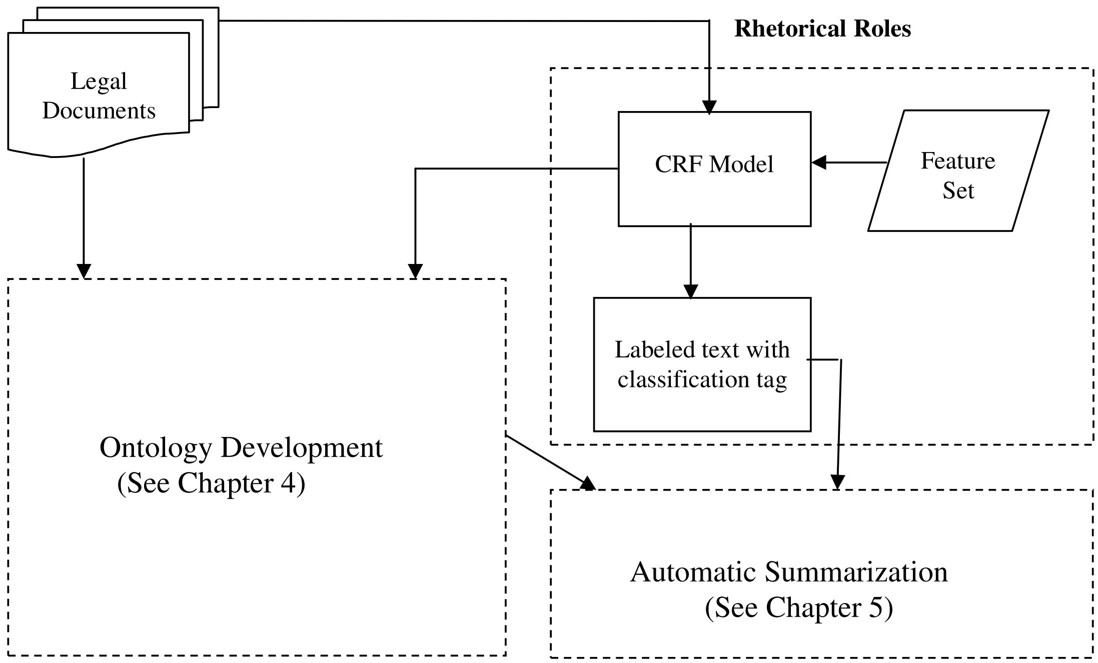


Figure 3.1 System architecture of rhetorical roles identification

3.1 Graphical Models

A *graph* comprises *nodes* (also called *vertices*) connected by *edges* (also known as *links* or *arcs*). In a probabilistic graphical model, each node represents a random variable (or a group of random variables), and the edges express probabilistic relationships between these variables. Probabilistic graphical models are highly advantageous in augmenting the analysis using diagrammatic representations of probability distributions [64]. The other useful properties are:

- They provide a simple way to visualize the structures of a probabilistic model and can be used to design and motivate new models.
- Insights into the properties of the model, including conditional independence properties, can be obtained by inspection of the graph.

- Complex computations, required to perform inference and learning in sophisticated models, can be expressed in terms of graphical manipulations, in which underlying mathematical expressions are carried along implicitly.

Probabilistic graphical models have been used to represent the joint probability distribution p(X, Y), where the variable Y represents attributes of the entities that we wish to predict, and the input variable X represents our observed knowledge about the entities. But modeling the joint distribution can lead to difficulties when using the rich local features that can occur in text data, because it requires modeling the distribution p(X), which can include complex dependencies. Modeling these dependencies among inputs can lead to intractable models, but ignoring them can lead to reduced performance. A solution to this problem is to directly model the conditional distribution p(Y|X), which is sufficient for segmentation. A graphical model is a family of probability distributions that factorize according to an underlying graph shown in Figure 3.2

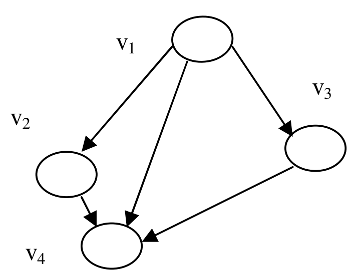


Figure 3.2 Simple graph connecting 4 vertices

The main idea is to represent a distribution over a large number of random variables by a product of local functions each of which depends on a small number of variables.

This section introduces the theory underpinning directed graphical models, in which the edges of the graphs have a particular directionality indicated by arrows, and explains how they may be used to identify a probability distribution over a set of random variables. Also, we give an introduction to undirected graphical models, also known as *Markov random fields,* in which the edges have no directional significance. Finally, we shall focus on the key aspects of Conditional Random Fields model as needed for applications in text segmentation carried out for the identification of rhetorical roles in legal documents.

3.1.1. Directed Graphical Model

A directed graphical model consists of an acyclic directed graph G = (V, E) where $V = \{V_1, V_2, \ldots, V_N\}$ are the set of N nodes belonging to G, and $E = \{(V_i, V_j)\}$ are the directed edges between the nodes in V. Every node V_i in the set of nodes V is in direct one-to-one correspondence with a random variable, also denoted as V_i. We use the common notation in which upper case letters denote random variables (and nodes) while lower case letters denote realizations. A realization of a random variable is a value taken by the variable. This correspondence between nodes and random variables enables every directed graphical models to represent a corresponding class of joint probability distributions over the random variables in V.

The simplest statement of the conditional independence relationships encoded in a directed model can be stated as follows: a node is independent of its ancestors given its parent nodes, where the ancestor/parent relationship is with respect to some fixed topological ordering of the nodes. We see that in equation (3.1) the conditional independence allows us to represent the joint distribution more compactly. We can now state in general terms the relationship between a given directed graph and the corresponding distribution over the variables. The directed nature of G means that every node V_i has a set of parent nodes $V_{\pi i}$, where π_i is the set of indices of parents of node V_i. The relationship between a node and its parents enables the expression for the joint distribution defined over the random variables V to be concisely factorized into a set of functions that depend on only a subset of the nodes in G. Directed graphical models [65] describe a family of probability distributions:

$$p\,(V_1, V_2, \ldots, V_n) = \prod_{i=1}^{n} p(V_i \mid V_{\pi i}) \qquad \ldots\ldots\ldots \; (3.1)$$

where the relation π_i indexes the parent nodes of V_i (the sources of incoming edges to V_i), which may be the empty set. Each function on the right hand side of (3.1) is a conditional distribution over a subset of the variables in V; each function must return positive scalars which are appropriately normalized. An example of directed acyclic graph describing the joint distribution over variables $v_1, v_2, \ldots, v_7$ is given in Figure 3.3. The joint distribution of all 7 variables is therefore given by

$$p(v_1)\, p(v_2)\, p(v_3)\, p(v_4 \mid v_1, v_2, v_3)\, p(v_5 \mid v_1, v_3)\, p(v_6 \mid v_4)\, p(v_7 \mid v_4, v_5) \qquad \ldots\ldots \; (3.2)$$

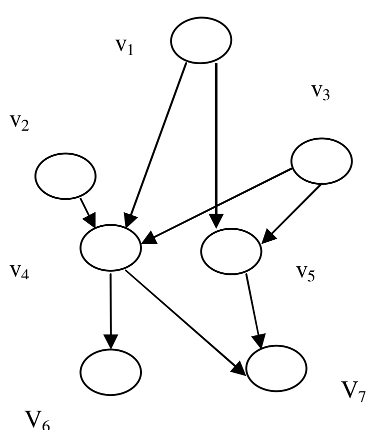


Figure 3.3 Example of a directed graph

Next we discuss two important forms of directed graphical models (Markovian Models) ─ Hidden Markov models [61] and Maximum Entropy Markov models [62] ─ used to express the probability distribution over a sequence of labels.

3.1.2 Hidden Markov Model

Hidden Markov models have been successfully applied to many data labelling tasks including Part of Speech (POS) tagging [66], Shallow parsing [67], Speech recognition [68] and Gene sequence analysis [69]. HMMs are probabilistic finite state automata [70] that model generative processes by defining joint probabilities over observation and label sequences [71]. Each observation sequence is considered to have been generated by a sequence of state transitions, beginning in some start state and ending when a final state is reached. At each state an element of the observation sequence is stochastically generated, before moving to the next state. The states in an HMM are considered to be hidden because of the doubly stochastic nature of the process described by the model. The generation of a set of labels (states) and words (outputs) in HMM are proceeding as follows [68]: the machine selects a start state, and from this state generates an output. It then transitions to another state, where it generates another output. This process continues until it reaches a final state.

In our case, the sequence of emissions forms the observed sequence of sentences, and the state sequence followed can be interpreted as a set of rhetorical roles. A HMM may be expressed as a directed graph G with nodes l_t and s_t representing the state of the HMM (or label) at time t and the observation at time t, respectively. The structure is shown in Figure 3.4.

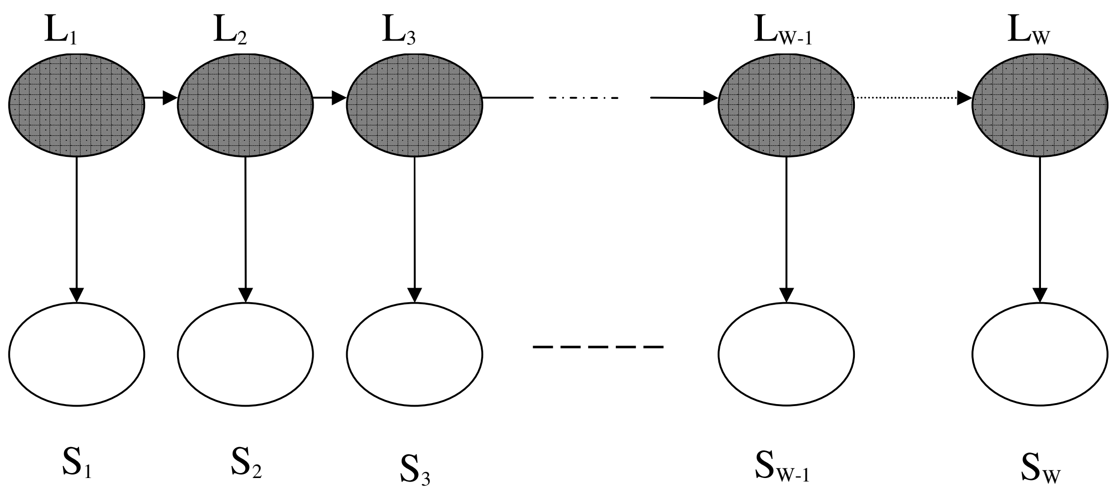


Figure 3.4 Graph structure of first-order HMMs for sequences

Here the state sequence forms a directed chain, with each state, l_t, linked to adjacent states, l_{t-1} and l_{t+1}, and output s_t linked only to l_t. In the text segmentation application, this means that we are considering that the labels related to the roles of each sentence depend only on the label assigned to the previous sentence, and each sentence depends only on the current label. These conditional independence relations, combined with the chain rule of probability, may be used to factorize the joint distribution over a state sequence $L = l_1, l_2, \ldots, l_w$ and observation sequence $S = s_1, s_2, \ldots, s_w$ into the product of a set of conditional probabilities:

$$p(L, S) = p(l_1)\ p(s_1 | l_1) \prod_{t=1}^{w} p(l_t | l_{t-1})\ p(s_t | l_t) \qquad \ldots\ldots.. (3.3)$$

Most commonly, the transition distributions $p(l_t | l_{t-1})$ are assumed to be invariant over time t, and the HMM is said to be homogenous. Text segmentation can be modeled as the task of identifying the rhetorical roles (labels) that best accounts for the observation sequence. This state sequence maximizes the conditional probability distribution of states given the observation sequence which may be calculated from the joint distribution using Bayes's rule:

$$l^* = \operatorname{argmax}_l p(l \mid s) \quad = \operatorname{argmax}_l \frac{p(l, s)}{p(s)} \qquad \text{........} \quad (3.4)$$

As the HMM describes the joint probability over states and outputs, we must represent the required condition in terms of the joint, as shown on the right hand side of (3.3). The optimal state sequence is most efficiently determined using a dynamic programming technique known as Viterbi algorithm [72]. Even though, HMM is a powerful method for labelling sequences, it has two issues. The first issue arises from the implicit independence assumed between an output, s_t, and all other outputs in the sequence. This relation prevents one from crafting features associated with multiple observations without incorporating their specific dependencies into the graph. The second issue faced by HMMs stems from their modeling of the output distribution. Their generative construction means that they not only model the distribution over state sequences, but also that of the observations, i.e., they are trained to maximize the joint likelihood of the data, rather than the conditional one. Despite their widespread use, HMMs and other generative models are not the most appropriate models for the task of labelling sequential data.

Hence, we discuss another model known as *Maximum Entropy Markov model* (MEMM) [62] which is in the form of a conditional model for labelling sequential data designed to address the problems that arise from the generative nature and strong independence assumptions of HMMs. To allow for non-independent, difficult to enumerate observation features, we have moved from the generative, joint probability parameterization of HMMs to a conditional model that represents the probability of reaching a state given an observation and a previous state. In MEMMs, each source state has an exponential model that takes the observation features as input, and a distribution over possible next states as output. These exponential models are trained by an appropriate iterative scaling method in the maximum entropy framework. MEMMs have been applied to a number of labelling and segmentation tasks including POS tagging [66] and segmentation of text documents [62].

3.1.3 Maximum Entropy Markov Model

Like HMMs, MEMMs are also based on the concept of a probabilistic finite state acceptor model. However, rather than generating observations, this model outputs label sequences when presented with an observation sequence. MEMMs consider observation sequences as events to be conditioned upon rather than generated. Therefore, instead of defining two types of distribution - a transition distribution $P(l'|l)$ that represents the probability of moving from state l to state l' and an observation distribution $P(s|l)$ representing the probability of emitting observation s when in state l - a MEMM has only a single set of $|L|$ separately trained distributions of the form

$$P_l(l'|s) = P(l'|l, s) \qquad \text{…….. (3.5)}$$

which represents the probability of moving from state l to l' on observation s. The fact that each of these functions is specific to a given state means that the choice of possible states at any given instant in time $t+1$ depends only on the state of the model at time t. The above state-observation transition functions which are conditioned on the observations mean that the graph structure for a MEMM can be represented in the form as shown in Figure 3.5.

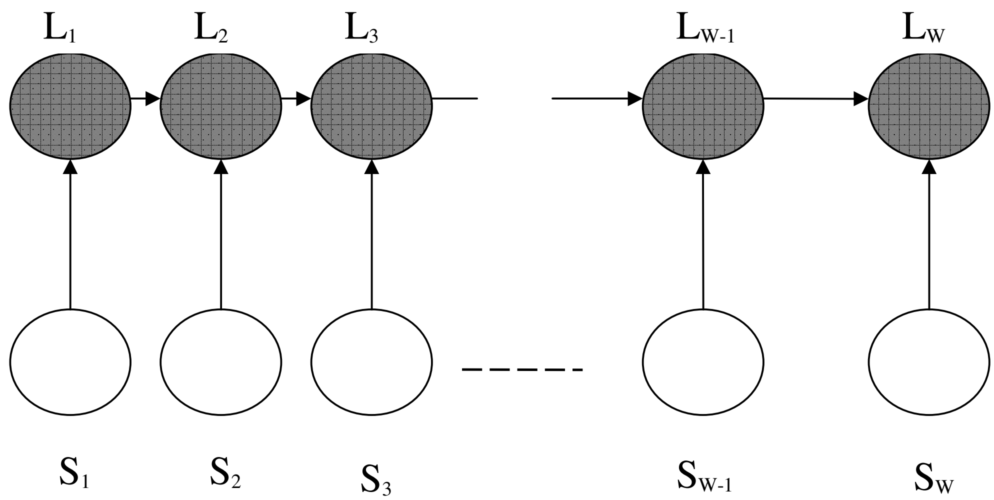


Figure 3.5 Graphical structure of first-order MEMMs for sequence

The constraints applied in this case are that the expected value of each feature in the learned distributions be the same as its average on the training observation sequence $s_1, \ldots s_w$ (with corresponding state sequence $l_1, \ldots .. l_w$). The maximum entropy distribution that satisfies those constraints [73] is unique, and agrees with the maximum-likelihood distribution that has the exponential form:

$$P_l(l'|s) = \frac{1}{Z(l, s)} \exp\left(\sum_a \lambda_a f_a(l', s)\right) \qquad \ldots\ldots.. \ (3.6)$$

where each λ_a are parameters to be estimated, and Z(l, s) is the normalizing factor that makes the distribution sum to one across all next states *l′*. Each f_a is a feature function that takes two arguments, the current observation *s* and a potential next state *l′*. The free parameters of each exponential model can be learned using *Generalized Iterative Scaling* [74]. Moreover, each feature function makes use of a binary feature *v* of the observation which expresses some characteristic of the empirical training distribution that should hold good for the trained model distribution also.

Similar to HMM, MEMMs also has been applied to labelling data by identifying the state sequence that best describes the observation sequence to be labelled [62]. Each state has a label associated with it and so the most probable label sequence for that observation sequence may be trivially identified once the most likely state sequence has been calculated. This state sequence maximizes the conditional probability distribution of states given the observations:

$$l^* = \text{argmax}_l \, p(l \mid s) \qquad \text{……. (3.7)}$$

Like HMM, it is desirable to use some form of dynamic programming algorithm. McCullam et al [3] present a brief overview of a variant on Viterbi algorithm that enables the state sequence to be efficiently identified for the task of text segmentation.

Maximum Entropy Markov Models and other non-generative finite-state models based on next classifiers, such as discriminative Markov models, exhibit undesirable behavior in certain circumstances, termed as *label bias problem* [63]. It describes the transitions on next state classifiers leaving a given state compete only against each other, rather than against all other transitions in the model. In probabilistic terms, transition scores are the conditional probabilities of possible next states given the current state and the observations sequence. This per-state normalization of transition scores implies a "conservation of score mass" whereby all the mass that arrives at a state must be distributed among the possible successor states. An observation can affect which destination states get the mass, but not how much total mass to pass on. This causes a bias toward states with fewer outgoing transitions. In the extreme case, a state with a single outgoing transition effectively ignores the observation. In those cases, unlike in HMMs, Viterbi decoding cannot downgrade a branch based on observations after the branch point, and models with state-transition structures that have sparsely connected chains of states are not properly handled. The Markovian assumptions in MEMMs and similar state-conditional models insulate decisions at one state from future decisions in a way that does not match the actual dependencies between consecutive states. This effect has been discussed in [63, 75] where conditional model structures are shown as leading to accuracy degradation in shift-reduce parsing, part-of-speech tagging and text segmentation.

A MEMM model is expected to suffer more markedly from label bias than a HMM type of generative models, as the MEMM cannot include single label features. These features allow the modeling of the state distribution based on previous results, rather than on bare counts. The net result is lower entropy transition distributions in the MEMM, and therefore an increased prevalence of label bias. Generative models such as HMMs do not suffer from the label bias problem. This is because the Viterbi algorithm which is used to identify the most likely state sequence given an observation sequence is able to down-weight a possible branch of a state sequence on the basis of observations that appear after the branch point. Moreover, HMMs are not required to preserve the probability mass over each transition in the finite-state acceptor; the observation probability at each state can dampen the path probability, and therefore avoid improbable states. For example, in segmenting legal documents, label bias problem occurs during the implementation of MEMMs when we consider the single paragraph which is related to different roles. Other than MEMMs, classical probabilistic automata [76], discriminative Markov models [63], maximum entropy triggers [77], as well as non-probabilistic sequence tagging and segmentation models with independently trained next-state classifiers [78] are all potential victims of the label bias problem.

To overcome the issues discussed earlier in the two different methods, we look into undirected graphical models and especially *Conditional Random Fields*, a sequence modeling framework that has all the advantages of HMMs and MEMMs but also solves the label bias in text segmentation problem. Conditional Random Fields are particularly suited to natural language processing tasks. CRFs are distributions over structured labellings conditioned on some context. This structure fits with many NLP tasks, which require the prediction of complex labellings for spans of text. For example, taggers predict a sequence of labels, one for each word in the text, while segmentation predicts a label for each sentence. The advantages of CRFs over MEMMs is that a MEMM uses per-state exponential models for the conditional probabilities of next states given the current state, while a CRF has a single exponential model for the joint probability of the entire sequence of labels given the observation sequence[63,79]. We will discuss more of these in the following sections.

3.1.4 Undirected Graphical Model

We have briefly reviewed directed models and their problems. Now, we turn to undirected graphical models to set up a base for discussing the CRF model. Undirected graphical models (also known as *Markov Random Fields* or *Markov networks*) are another class of graphical models, which use a different factorization of the joint distribution compared to that of directed models, and also use different conditional independence semantics. An undirected model is a graph G = (V, E), where V = $\{V_1, V_2, \ldots V_N\}$ are the set of N nodes belonging to G and

$E=\{(V_i, V_j): i \neq j\}$ are the undirected edges between the nodes in V. The joint distribution of an undirected graphical model is defined by

$$p(v_1, v_2, \ldots v_N) = \frac{1}{Z} \prod_{c \in C} \psi_{V_c}(v_c) \qquad \ldots\ldots\ldots \ (3.8)$$

where C is the set of maximal cliques in the graph, $\psi_{V_c}(v_c)$ is a potential function (a positive, but otherwise arbitrary, real-valued function) on the clique v_c, and Z is the normalization factor

$$Z = \sum_{v} \prod_{c \in C} \psi_{V_c}(v_c) \qquad \ldots\ldots\ldots \ (3.9)$$

The term clique describes a subset of nodes that are fully connected: every pair of nodes in the subset is connected by an edge. A maximal-clique is a clique that cannot be enlarged with additional nodes while still remaining fully connected. For example, the undirected graph in Figure 3.6 contains the following cliques: $\{V_1\}$, $\{V_2\}$, $\{V_3\}$, $\{V_4\}$, $\{V_5\}$, $\{V_1, V_2\}$,$\{V_2,V_3\}$,$\{V_2, V_4\}$,$\{V_3, V_4\}$$\{V_3, V_5\}$,$\{V_4,V_5\}$,$\{V_2, V_3, V_4\}$, and $\{V_3, V_4, V_5\}$. Of these, only three are maximal cliques; $\{V_1, V_2\}$, $\{V_2, V_3, V_4\}$, and $\{V_3, V_4, V_5\}$, in that these three cliques include all the remaining cliques in the graph. To limit the maximal cliques, it requires a potential function ψ for each non-maximal clique. It is to be incorporated into the potential function of exactly one of its subsuming maximal cliques. Now, we will discuss one of the important undirected models used for text segmentation task in the next section.

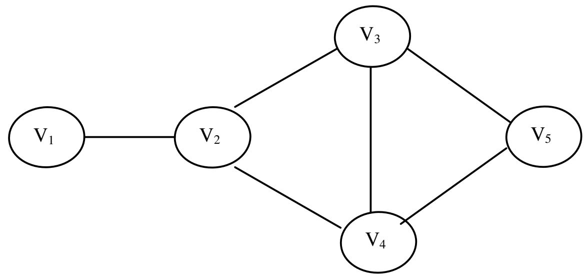


Figure 3.6 Example of undirected graphical model of five random variables.

3.1.5 Conditional Random Fields Model

Conditional Random Fields, as introduced by Lafferty [63], is one form of a conditional model that allows the strong independence assumptions of HMMs to be relaxed, while overcoming the label-bias problem exhibited by MEMMs [62] and other non-generative directed graphical models such as discriminative Markov models [79].

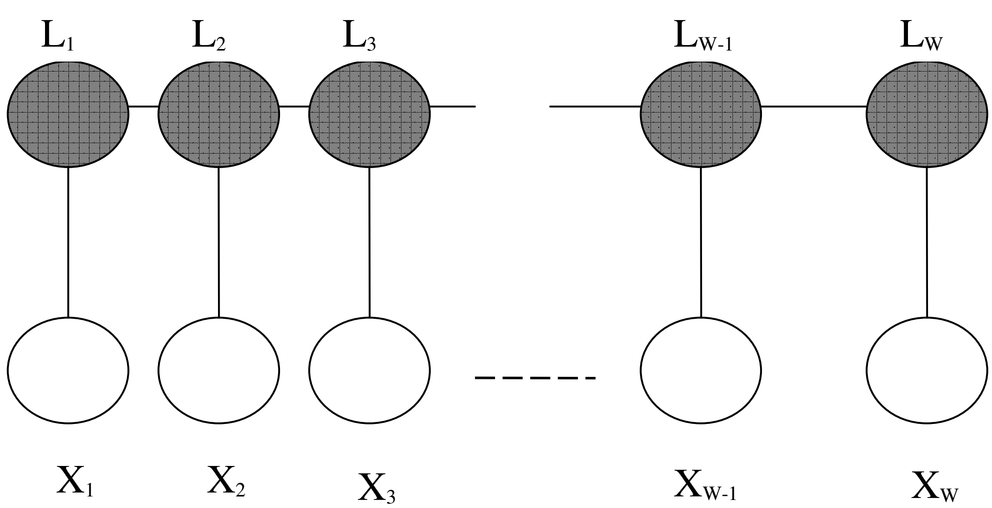


Figure 3.7 Graphical structure of a linear chain CRFs for sequence labeling

Like MEMMs, CRFs are conditional probabilistic sequence models, but they are undirected graphical models. This allows the specification of a single joint probability distribution over the entire label sequence given the observation sequence, rather than defining the per-state distributions over the next states given the current state, as shown in Figure 3.7.

The conditional nature of the distribution over label sequences allows linear chain CRFs to model real-world data in which the conditional probability of a label sequence can depend on non-independent, interacting features of the observation sequence. In addition to this, the exponential nature of the distribution chosen by Lafferty [63] enables the features of different states to be traded off against each other, weighting some states in a sequence as being more important than others.

This thesis deals with Conditional Random Fields [63], which will be shown as a powerful model for predicting the presence of structured roles in legal documents. CRFs are increasingly popular primarily in natural language processing (NLP) and in applications such as computational biology. NLP tasks ranging from document classification, summarization to POS tagging to syntactic parsing can all be considered as structured label identification tasks. CRFs have been previously applied to other tasks such as name entity extraction [80], table extraction [81] and shallow parsing [82]. All of these tasks had been modeled using CRF, often with impressive results [63, 83]. Following are the major reasons for CRFs being particularly suitable for natural language processing applications:

- CRFs allow the joint modeling of complex structured labelling tasks which should not be decomposed into a number of smaller independent tasks. For example, the construction of a parse tree should not be decomposed into a series of independent decision; rather a local decision to label a span of words as a constituent of a given type will affect the decisions for both its parent and child constituents, which recursively affect their children and parents. Therefore the prediction process must consider the complex interdependence between these decisions.

- CRFs provide a rich and largely unconstrained feature representation. For example, a feature might detect the presence of the word 'section' in a sentence which is to be labelled under *argument* related roles, or it may detect the cue phrase 'no merit in the disposition' which is to be labeled as a *final decision*. These features may reference overlapping and non-independent aspects of the data. Associated with each feature is a model parameter which assigns a weight to the feature. CRFs can use very large feature sets, which allow for flexible and expressive modeling power.
- CRFs are probabilistic models which describe a probability distribution over labellings. Therefore they inherently represent ambiguity in their predictions; this can be used to better inform other algorithms which use the model's predictions as input to perform further processing. Complex NLP applications commonly employ the combined architecture where many prediction tasks are performed in series with the results of each task informing all subsequent tasks. In this case, prediction errors in early tasks lead to many more errors in downstream tasks (Sutton [79] demonstrate the effect of cascading errors in POS tagging followed by NP chunking). For this reason, probabilistic models are ideal, in that they can preserve ambiguity in their predictions, thereby limiting one source of errors during the combination.
- The model can be fitted to a training sample using a maximum likelihood estimate. This provides a practical and elegant way of smoothing the model and

thereby reducing overfitting of the training sample. Overfitting is a big problem for CRFs, which are often extremely expressive with thousands or millions of features, and therefore parameters. However overfitting can be limited by using a simple Gaussian prior (or other distribution) which discourages over-reliance on any single feature.

- The model is discriminative, i.e., it predicts the labelling conditioned on some observations. The observations are typically sequences of words, or other contextual data. This setup exactly matches the testing configuration, where the model is supplied with the observations and must predict a labelling. These features are often very difficult to include in other similar models (e.g., Hidden Markov Models).

For these reasons CRFs have been widely adopted by the NLP community. Many of the compelling reasons to use CRFs also apply to other models, although CRFs are one of the few models which combine all these benefits in one model. CRFs combine these strong benefits in a discriminative framework. Discriminative models describe a conditional distribution over labellings given some observations, while their counterparts, generative models, model the joint distribution over the labelling and observations. Both types of model can be used for prediction and other forms of inference. However, discriminative models allow for a more flexible feature representation, as they do not need to model the interdependence between the observations or assume-away the dependence. While CRFs are extremely well suited to language tasks, all their benefits do not come free. The model has two main failings. Firstly, as a consequence of their flexible feature representation, the models are often used with an extremely large feature set and therefore have an equally large parameter space. This allows for a very expressive model which can fit the training sample very precisely, without sufficient inductive bias to ensure strong accuracy. This overfitting can be countered to some extent by smoothing the model – most commonly by including a prior over the parameter values. However, there is compelling evidence that simple priors do not provide sufficient regularization; i.e., the model still could benefit from further smoothing, as demonstrated in Smith and Osborne [84] and Sutton and McCallum [79]. Secondly, CRFs often take a very long time to train especially when compared to similar generative directed models (e.g., Hidden Markov Models). This is because there is no closed form solution for maximizing the data likelihood in a CRF, even for fully observed (supervised) training data. For this reason the likelihood is usually optimized by an iterative process, which is guaranteed to find the optimal values, but often requires many hundreds of iterations to do so. In spite of these two issues, CRFs are successfully implemented for various information retrieval tasks.

3.1.6 Conditional Random Fields for text segmentation task

As a special case in which the output nodes of the graphical model are linked by edges in a linear chain, CRFs make a first-order Markov independence assumption with binary feature functions, and thus can be understood as conditionally-trained finite state machines (FSMs) which are suitable for segmentation and sentence labeling.

A linear chain CRF with parameters $C = \{C_1, C_2, \ldots\}$ defines the conditional probability for a label sequence $L = l_1, l_2, \ldots, l_w$ given an observed input sequence

$S = s_1,\ldots s_w$ to be

$$P_C(L \mid S) = \frac{1}{Z_s} \exp[\sum_{t=1}^{w} \sum_{a} C_a\, f_a(l_{t-1}, l_t, s)] \qquad \ldots\ldots\ (3.10)$$

where Z_s is the normalization factor that makes the probability of all state sequences sum up to one, $f_a(l_{t-1}, l_t, s)$ is a feature function which is generally binary valued and C_a is a learned weight associated with the a[th] feature function. For example, a feature may have the value of 0 in most cases, but given the text "points for consideration", it has the value 1 along the transition where l_{t-1} corresponds to a state with the label *Identifying the case*, l_t corresponds to a state with the label *History of the case*, and f_a is the feature function PHRASE= "points for consideration" belongs to s in the sequence. That is, in a role identification related problem of a legal judgment, we need to define our binary feature in the form of

$$v(s_t) = \begin{cases} 1 & \text{the presence of a word "}\textit{act}\text{" in the sentence} \\ 0 & \text{otherwise.} \end{cases} \qquad \ldots\ldots\ldots\ (3.11)$$

Now we define each feature function f_a as a pair a = (v, l), where *v* is a binary feature of the observation s_t and l_t is a destination state:

$$f_{(v,l)}(l_t, s_t) = \begin{cases} 1 & \text{if } v(s_t) = 1 \text{ and } l_t = l \\ 0 & \text{otherwise.} \end{cases} \qquad \ldots\ldots(3.12a)$$

Similarly, we define feature function for transitions between different label states l and l' as follows:

$$f_{(l',l)}(l_{t-1}, l_t) = \begin{cases} 1 & \text{if } l_{t-1} = l' \text{ and } l_t = l \\ 0 & \text{otherwise.} \end{cases} \qquad \ldots\ldots(3.12b)$$

Large positive values for C_a indicate a preference for such an event, while large negative values make the event unlikely and near zero for relatively uninformative features. These weights are set to maximize the conditional log-likelihood of the labeled sequence for a training set $D = \{(s_{t,} l_t) : t = 1,2,\ldots N)$, written as

$$L_C (D) = \sum_i \log P_C(L_i \mid S_i)$$

$$= \sum_i (\sum_{t=1}^{W} \sum_a C_a f_a (l_{t-1}, l_t. s) - \log Zs_i) \qquad \ldots\ldots\ldots(3.13)$$

The training state sequences are fully labeled and definite, the objective function is convex, and thus the model is guaranteed to find the optimal weight settings in terms of L_C (D). The most probable label sequence for an input sequence s_i can be efficiently calculated by dynamic programming using modified Viterbi algorithm [72]. These implementations of CRFs are done using a newly developed java class library which also uses a quasi-Newton method called L-BFGS to find these feature weights efficiently.

As a novel approach, CRF model has been implemented for role identification in legal domain and the evaluation details are given in section 3.6. In our approach, we have first implemented a *Rule based* approach and extended this method with additional features and a probabilistic model. That is, we are in the process of developing a fully automatic summarization system for a legal domain on the basis of Lafferty's [63] segmentation task and Teufel & Moen's [38] gold standard approaches. Legal judgments are different in characteristics compared with articles reporting scientific research work or other simple domains in so far as the identification of basic structures of a document is concerned. Even skilled lawyers are facing difficulty in identifying the reasons behind the judgment in a legal document. The sentence extraction task forms part of an automatic summarization system in the legal domain. Before implementing CRF model for text segmentation task, we will discuss other algorithms available for segmentation of a given text in the next section.

3.2 Text Segmentation

Documents usually include various rhetorical roles. A summary of a long document can be composed of summaries of the component roles. We look at the problem of identifying the rhetorical roles present in a document as one of text segmentation. That is dividing the document along the different rhetorical roles. A lot of research has been done in text segmentation [85-87]. A major characteristic of the methods used in the above papers is that they do not require training data to segment given texts. Hearst [88], for example, used only the similarity of word distributions in a given text to segment the text. This property is important when text segmentation is applied to information retrieval or summarization, because both tasks deal with domain-independent documents. Another application of text segmentation is the segmentation of a continuous broadcast news story into individual stories [89]. In this application, systems relying on supervised learning [90] achieve good performance because there are plenty of training data in the domain. In our work, we have used training documents in the legal domain to train the text segmentation algorithm for the purpose of improving the role identification results.

We look at two approaches to text segmentation. The first approach is a rule-based one with rule sets tailored for the legal domain. The second approach is based on a Conditional Random Field as described in Section 3.1.6. The CRF uses a rich set of features tuned for the legal domain. This method shows significant improvement over the rule-based method. The description of rule-based methods is given in the following sections.

3.2.1 Rule-based learning algorithms

Rule learning algorithms produce a compact understandable hypothesis. Some popular rule learning systems are CN2 [92], RIPPER [93] or C4.5 rules [94]. However, the rule learning systems that perform best experimentally have the disadvantage of being complex, hard to implement, and not well-understood formally. SLIPPER (for Simple Learner with Iterative Pruning to Produce Error Reduction) [95] is a standard rule-learning algorithm which was taken for comparison with our approaches for the task of text segmentation. There are two important reasons for which we compared our approaches with SLIPPER. The first reason is that it is more scalable and noise-tolerant than other separate-and-conquer rule learning algorithm, such as reduce error pruning (REP) for rules [96], IREP [97], and RIPPER [93]. The second reason is that it is based on the line of research on boosting [91,98], in particular the AdaBoost algorithm [91], and its successor developed by Schapire and Singer [99]. Moreover SLIPPER rule sets are of moderate size, comparable to those produced by C4.5 rules [94] and C5.0 rules.

Most traditional rule learning algorithms are based on a divide-and-conquer strategy. SLIPPER [95] is one of the standard rule learning algorithms used for textual databases in the information retrieval task. In SLIPPER, the ad hoc metrics used to guide the growing and pruning of rules are replaced with metrics based on the formal analysis of boosting algorithms. For each instance, we need to check each and every rule in the rule set for a given sentence. It takes more time for larger corpora compared to other rule learning algorithms even for a two-class problem. If we need to consider more than two classes and to avoid overfitting of ensemble of rules, one has to think of grouping the rules in a rule set, and following some chaining mechanism. Another rule learning algorithm RuleFit [100] generates a small comprehensible rule set which is used in ensemble learning with a larger margin. In this case overfitting may happen if the rule set gets too large and hence some form of control has to be maintained. Our main idea is to find a preferably small set of rules with high predictive accuracy and with marginal execution time.

SLIPPER [95] generates rule sets to repeatedly boost a simple, greedy, rule-builder for text data. SLIPPER is simpler and better understood formally than other state-of-the-art rule learners. In spite of this, SLIPPER scales well on large datasets, and is an extremely effective learner. Cohen [95] claims that on a set of 32 benchmark problems, SLIPPER achieves lower error rates than RIPPER 20 times, and low error rates than C4.5 rules 22 times. The rule sets produced by SLIPPER are also comparable in size to those produced by C4.5 rules. Moreover, the more established rule learning systems like decision trees, neural networks and SVMs are hard to analyze from a statistical point of view [96]. This is due to the combinatorial complexity that is inherent in the setting and the use of heuristics rather than statistically motivated principles in most rule learning algorithms. Due to the above reasons, we have compared our proposed rule-based method with SLIPPER for segmenting a legal document into different rhetorical roles.

3.2.2 Proposed Rule-based Approach

An alternative rule based method that concentrates on grouping of rules in a rule set and which applies a chaining relation depending on each rhetorical role (Table 3.4) has been proposed. A chain relation is a technique used to identify co-occurrences of roles in legal judgments. This has been framed based on the observation of human annotation schemes. In our approach, rules are conjunctions of primitive conditions. As used by the boosting algorithms, a rule set R can be any hypothesis that partitions the set of inputs X along particular role categorization. In the beginning, evaluating rules are taken that describe the original features found in the training set.

```
Procedure Test (X) {

        Read test set
        Read input sentences from sample X
        Apply rules in R (with role categorization by maintaining chain relation)
        For k = 1 to m sentences
         {
              For j = 1 to 7      /* 7 identified roles */
               {
                  Initialize count_j to 0 /* counts the number of rules that
                                           assign L_j to sentence k  */
                  For i = 1 to  no. of rules whose antecedent is L_j
                  {
                      If rule_i fires on sentence_k then increment count_j
                  }
                }
                   j* = argmax_j count_j
                   Label_k = L_j*
                   Adjust labels of sentences k-2, k-1, k based on
                                              applicable chain relations
          }
}
```

Figure 3.8 A rule-based text segmentation procedure

Let $X = (S_1, S_2,\ldots,S_m)$ be a sample document of size m, where each S_i is a sentence. We assume that the set of rules $R = \{r_1,r_2,\ldots\}$ are applied to sample X, where each rule $r_i : X \rightarrow L$ represents the mapping of sentences of X onto a rhetorical role and $L = \{L_1,L_2,\ldots,L_7\}$. Each L_i represents a rhetorical role from the fixed set shown in Table 3.4. The procedure, given in Figure 3.8, starts with examining the sentences of test set one by one for assigning a particular role to each sentence. In this process, for each sentence, a rule set is thoroughly checked for applying the exact rule which prompts it for correct role assignment. In addition to rule checking, the chain relation (co-occurrence of roles in sentences) is also verified. The co-occurrence of roles in sentences is identified based on observing the human annotated document set and measuring the frequency of co-occurrence of roles. Eventually, this relation avoids the introduction of another role in between a sequence of identical roles. The different rhetorical categories used for labeling the sentences are discussed in the next section.

3.3 Exploration of legal document structure

In recent years, much attention has been focused on the problem of understanding the structure and textual units in legal judgments [45]. In this case, performing automatic segmentation of a document to understand the rhetorical roles turns out to be an important research issue. For instance, Farzindar [45] proposed a text summarization method to manipulate factual and heuristic knowledge from legal documents. Hachey and Grover [42] explored machine learning approach to rhetorical status classification by performing fact extraction and sentence extraction for automatic summarization of texts in the legal domain. They formalized the problem as one of extracting the most important units based on the identification of thematic structure of the document and determination of argumentative roles of the textual units in the judgment. They mainly used linguistic features to identify the thematic structures, as already outlined in Chapter 2.

The work on information extraction from legal documents has largely been based on semantic processing of legal texts to explore the structure and applying machine learning algorithms like C4.5, Naïve Bayes, Winnow and SVMs [42]. These algorithms run on features like cue phrases, location, entities, sentence length, quotations and thematic words. For this process, Named Entity Recognition rules have been written by hand for all domain related documents. The recent work on automatic extraction of titles (concepts) from general documents using machine learning methods shows that machine learning approaches can work significantly better than the baseline methods for meta-data extraction [101]. Some of the other works in the area of legal domain concerns information retrieval and the computation of simple features such as word frequency, cue words and understanding minimal semantic relation between the terms in a document. Understanding discourse structures and the linguistic cues in legal texts are very valuable techniques for information extraction systems [42]. For automatic segmentation task, it is necessary to explore more features which are representative of the characteristics of texts in general and legal text in particular.

The genre structure identified in our process plays a crucial role in identifying the main decision part by grouping the sentences in the document into appropriate categories. We start with understanding the different set of contextual rhetorical schemes suggested for different domains. Then, we come out with a new set of rhetorical scheme for a legal domain based on the above studies. It is important for our task to find the right definition of rhetorical roles to describe the content in legal documents. The definition should both capture generalizations about the nature of legal texts, and also provides the right kind of information to enable the construction of better summaries for a practical application. Our model relies on the following basic dimensions of document structure in legal documents. The structure of the legal reports cited in Indian courts generally is as follows

- Titles (Case description)
- Summary (equivalent to the headnote)
- The facts
- The decision (Incorporating the argument, ratio and judgment)

Table 3.1 shows the basic themes which divide the legal decisions into thematic segments based on the work of Farzindar [45]. Table 3.2 shows the rhetorical categories identified in scientific articles [38]. These rhetorical categories have also been tried with legal texts [42]. To maintain the same basic structure of legal judgments cited in Indian courts, we have identified four basic rhetorical roles which are given in Table 3.3 based on Teufel and Moens [38] gold standard approach.

Table 3.1 Description of the basic thematic segments in a legal document

Labels	Description
Introduction	The sentence describes the situation before the court and answers these questions: who did they do to whom?
Context	The sentence explains the facts in chronological order, or by descriptions. It recomposes the story from the facts and events between the parties and findings of credibility on the disputed facts.
Judicial Analysis	The sentence describes the comments of the judge and finding of facts, and the application of the law to the facts as found. For the legal expert this section of judgment is the most important part, because it gives a solution to the problem of the parties, and leads the judgment to a conclusion.
Conclusion	The sentence expresses the disposition which is the final part of a decision containing the information about what is decided by the court. For example, it specifies if the person is discharged or not, or the cost for a party.

Table 3.2 Annotation scheme for rhetorical status

Labels	Description
Aim	Specific research goal of the current paper
Textual	Statements about section structure
Own	Description of own work presented in current paper: Methodology, results discussion
Background	Generally accepted scientific background
Contrast	Statements of comparison with or contrast to other work; weaknesses of other work.
Basis	Statements of agreement with other work or continuation of other work
Other	Descriptions of other researcher's work

Table 3.3 Description of the basic rhetorical roles for a legal domain.

Labels	Description
Facts	The sentences describe the details of the case
Background	The sentence contains the generally accepted background knowledge (i.e., legal details, summary of law, history of a case)
Own	The sentence contains statements that can be attributed to the way judges conduct the case.
Case	The sentences contain the details of other cases coded in this case.

The roles defined in Table 3.3 do not explicitly define the decision and fact parts of general structure of legal documents. Hence, the classifications given in Table 3.3 have been enhanced into seven different labeled elements in our work. Moreover, we also got the endorsement from leading legal personalities and appreciation from the legal communities [102] related to the seven rhetorical categories identified in our study. To identify the labels, we need to create a rich collection of features, which includes all important features like concept and cue phrase identification, structure identification, abbreviated words, length of a word, position of sentences, etc,. The position in the text or in a section does not appear to be significant for any Indian law judgments except for identification of a few roles. Not all judgments follow the general structure of a legal document. There are some categories of judgments (e.g., judgments belonging to sales tax, income tax) where the fact and the decision are discussed in different parts more than once. To overcome this problem, positioning of a word or sentence in a document is not considered for role identification as one of the important features in this work.

Table 3.4 The current working version of the rhetorical annotation scheme for legal judgments.

Rhetorical Status	**Description**
1. **Identifying the case**	The sentences that are present in a judgment to identify the issues to be decided for a case. Courts call them as "Framing the issues".
2. Establishing facts of the case	The facts that are relevant to the present proceedings/litigations that stand proved, disproved or unproved for proper applications of correct legal principle/law.
3. Arguing the case	Application of legal principle/law advocated by contending parties to a given set of proved facts.
4. History of the case	Chronology of events with factual details that led to the present case between parties named therein before the court on which the judgment is delivered.
5. Arguments (Analysis)	The court discussion on the law that is applicable to the set of proved facts by weighing the arguments of contending parties with reference to the statute and precedents that are available.
6. Ratio decidendi (Ratio of the decision)	Applying the correct law to a set of facts is the duty of any court. The reason given for application of any legal principle/law to decide a case is called Ratio decidendi in legal parlance. It can also be described as the central generic reference of text.
7. Final decision (Disposal)	It is an ultimate decision or conclusion of the court following as a natural or logical outcome of ratio of the decision

Identifying the case 1. *We may now consider the question whether the term 'the Government' used in Section 73A includes Central Government.* 2. *The question that arises for considering in this case is whether block assessment under Chapter XIV B of the Income tax, 1961 would fall within the meaning of case.* *3. The short question that arises for consideration of this writ appeal directed against the judgment* *dated December 18, 2002 passed by a learned single judge in O.P.No. 35879 of 2002.*
Establishing facts of the case *1. We find no reason to interfere with the said finding concurrently entered by the authorities below.* *2. Admittedly, statuary formalities have already been complied with in the instant case.* *3. It is thus clear that the assessing authority has proceeded on the basis that the learned single Judge has directed him to grant exemption and then complete the assessment, which he has accordingly done.*
Arguing the case *1. It is the contention of the Standing Counsel for the Railways that in the absence of a definition of Government in the Act, guidance should be obtained from the General Clauses Act 1897.* *2. Sri. C.K. Nair, leading the arguments on behalf of the assessees, has pointed out that the incentive bonus if at all assessable as salary has to be treated as profit in lieu of salary or in addition to salary as contemplated under Section 17 (1) (iv) of the Income Tax Act* 3. *Sri Georgekutty Mathew, learned Government pleader submitted that the order of the Tribunal cannot be sustained since the decision in Stephan's case has subsequently been overruled by the apex court in Commissioner of sales tax v. Stephan & Co. (1988) 69 STC 320.*
Arguments (Analysis) *1. A bench of this court Rabi Umma v. Mukundan (1992 (I) K.L.T.700) held that the working of the sub-clause especially the words "such further period" would permit extension of the period fixed in the original order by a subsequent order passed not only before the expiry of the original period but also after its expiry.* *2. Counsel Submitted that the assessment for the block period in accordance with the provisions contained in Chapter XIV B is different from "proceedings" under the Act in respect of any year, as mentioned.* *3. Apex court had occasion to consider the scope of Section 6A(ii)(a) of Kerala General Sales Tax Act 1963 in Nandanam Construction Company's case supra Overruling the decision of this court in Commissioner of Sales Tax v. Pio Food packers (1980) 46 STC 63 the apex court held as follows.*
Ratio decidendi *(Ratio of the decision)* *1. We are of the view that the order under challenge does not require any interference by this court.* 2. *Looking at the question in the above perspective, we find no infirmity in the order passed by the Chief Commissioner in transferring the case to the assistant Commissioner of Income Tax, Calicut.* 3. *We are clearly of the view that all these statutory remedies could not be allowed to be bypassed and, therefore it was not appropriate for the learned judge to have entered into the merits of the controversy at the state when the Company had been issued a notice under Section 17 of the Act.*
Final decision *(Disposal)* *1. Revision petition lacks merits and it is accordingly dismissed. However, considering the facts and circumstances of the case, tenant is given three months time from today to vacate the premises.* 2. *We therefore answer the question in favour of the Revenue and dismiss the appeal and writ petition.* *3. In the result, the appeal is allowed, judgment of the learned single Judge dated 18th December 2002 set aside and O.P.No. 35879 of 2002 dismissed.*

Figure 3.9 Examples of manual annotation of relevant sentences with rhetorical roles

The approach to explore the elements from the structure of legal documents has been generalized in the fixed set of seven rhetorical categories based on Bhatia's [103] genre analysis shown in Table 3.4.

Figure 3.9 gives an example of the manual annotation of different sentences into the appropriate rhetorical categories. Sample relevant sentences of all rhetorical categories are shown. We thus decided to augment the available corpus with an independent set of human judgments of relevance. We intend to replace the vague definition of relevance often used in sentence extraction experiments with a more operational definition based on rhetorical status. In inherently subjective tasks, it is also a common practice to consider human performance as an upper bound [38].

Our system creates a list like the one in Figure 3.9 automatically. The actual output of the system when run on the sample judgment is shown later in Fig. 3.17. The manual annotation of all sentences in the judgments related to all of the three sub-domains will be considered as a gold standard for later use during system training and system evaluation. We have two parallel human annotations in our corpus: Rhetorical Role Annotation and Relevant Sentence Selection for a document summary. In the first annotation task, each sentence in the judgment was labeled with one of the seven rhetorical categories. During the second annotation task, the experts were asked to select the important sentences to be included in the final summary. Figure 3.10 is related to the distribution of the seven categories in rent control judgments and it shows that it is very much skewed, with 60% of all sentences being classified as *History of the case*. As that segment includes the remaining contents of the document other than those belonging to the six categories it appears larger than other segments

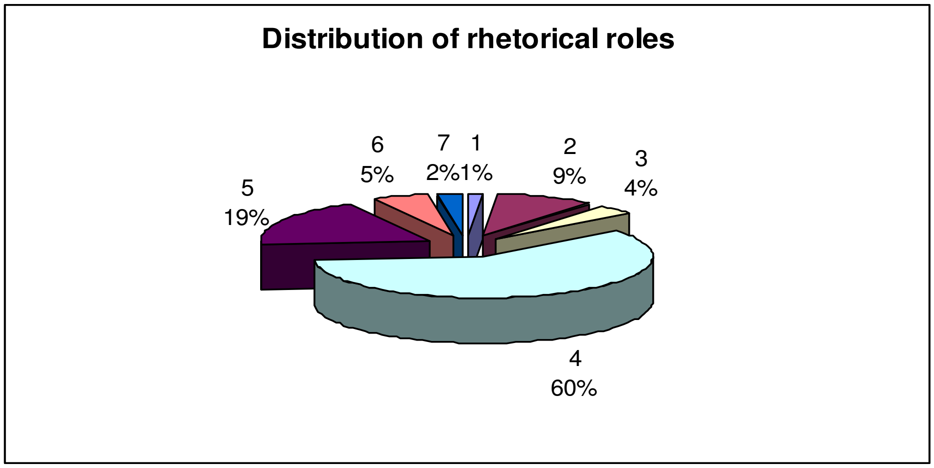


Figure 3.10 Distribution of the seven rhetorical categories in entire documents from the rent control domain

In common law system, decisions made by the judges are important sources of applications and interpretations of law. A judge generally follows the reasoning used by earlier judges in similar cases. This reasoning is known as the reason for the decision (*Ratio decidendi*). The important portion of a headnote includes the sentences which are related to the reason for the decision. These sentences justify the judge's decision, and in non-legal terms may be described as the central generic sentences of the text. Hence, we reckon this as one of the important elements to be included in our genre structure of judgments. Usually, the ratio appears in the decision section, but sometimes may appear in the earlier portion of a document. In our approach, we have given importance to the cues for the identification of the central generic sentence in a law report rather than to its relative position in the text. From the Indian court judgments, we found that the ratio can be found in any part of the decision section of a law report, and that they usually appear as complex sentences. It is not uncommon to find that the experts differ among themselves on the identification of the ratio of the decision in a given judgment. This shows the complexity of the task.

Exploration of text data is a complex proposition. But in general, we can figure out two characteristics from the text data; the first one is the statistical dependencies that exist between the entities related to the proposed model, and the second one is the cue phrase / term which can support a rich set of features that may aid classification or segmentation of given document. The feature set used in this approach is discussed in the next section.

3.4 Feature sets

Feature sets with varying characteristics are employed in order to provide significant improvements in CRFs performance [63] which are already defined in Equations 3.11 and 3.12. Features common to information retrieval, which were used successfully in the genre of different domains, will also be applicable to legal documents. The choice of relevant features is always vital to the performance of any machine learning algorithm. Identifying state transitions of CRF were also considered as one of the important features in any information extraction task [104]. The features with which we have been experimenting for the legal corpus are broadly similar to those used by Teufel and Moens [38] and include many of the features which are typically used in sentence extraction approaches to automatic summarization as well as certain other features developed specifically for rhetorical role identification.

3.4.1 Indicators / cue phrases

The term *cue phrase* indicates the frequently used key phrases which are the indicators of common rhetorical roles of the sentences (e.g. phrases such as "We agree with court", "Question for consideration is", etc.,). Most of the earlier studies dealt with the building of hand-crafted lexicons in which each and every cue phrase was related to different labels.

- *The <u>question for consideration</u> is whether a direction under section 12 (3) of the Act could be given by the Rent Controller or the Appellate Authority during the pendency of proceedings arising under Rule 13 (3) of the Rules.*
- *There is <u>no merit</u> in the revision and the same is dismissed.*
- *<u>We find</u> no reason to disturb the concurrent findings entered by the authorities below.*
- *Revision petition lacks merits and it is accordingly <u>dismissed</u>.*
- *The crucial <u>point to be considered</u> in this case is as to whether daughter-in-law will come within the expression of family dependent of the landlord under section 11(2) of the Act.*
- *<u>If is be so</u> the remedy open to the plaintiff is to sue for damages.*
- *In the above circumstances, <u>we are of the view</u> that in view of the compulsory nature of payments under sub-section (7B) of section 7 which an assessee to pay tax to the State of Kerala.*
- *It may be <u>pertinent to note</u> that the Entry Tax Act has not given any specific definition for furniture.*
- *<u>Therefore I feel that</u> the <u>contention</u> of the petitioners <u>is correct</u> and the respondents have no right to demand entry tax in respect of dental chair brought by them.*
- *He heavily <u>relied on the decision</u> of the Gujarat High court referred to above, which has approved the adoption of the net income after providing for expenses.*
- *<u>We do not find</u> any provision in the <u>Income Tax Act</u>, except <u>Section 10(14)</u>, for allowing deduction towards expenditure of this nature claimed by the assessees.*
- <u>Counsel submitted</u> *when explanation to section 120 was added concept of block assessment was not in vogue.*
- *Looking at the question in the above perspective, we find no infirmity in the order passed by the Chief Commissioner in transferring the case to the assistant Commissioner of Income Tax, Calicut.*
- *<u>We therefore answer the question</u> in favour of the Revenue and <u>dismiss</u> the appeal and writ petition.*

Figure 3.11 Existence of important cue phrases in the source documents

In this study, based on expert suggestions, the initial set of cue phrases has been selected and the associations to labels are learned automatically. If a training sequence contains “No provision in ….act/statute”, “we hold”, “we find no merits” all labeled with *ratio decidendi*, the model learns that these phrases are indicative of ratios. But the model faces difficulty in setting the weights for the feature when the cues appear within quoted paragraphs. This sort of structural knowledge can be provided in the form of rules. Feature function (v) of Equation 3.11 for the rules is set to 1 if they match words/phrases in the input sequence exactly.

Apart from the lexical and other set of features, we noticed that the legal corpus contains a large number of cue phrases which are relevant for the identification of rhetorical roles. Figure 3.11, for instance, shows that the presence of cue phrases indicates the presence of particular roles in the statements.

3.4.2 Named entities

This feature is not considered fully while summarizing scientific articles [38]. However, in this work, we recognize a wide range of named entities and generate binary-valued entity type features which take the value 0 or 1 indicating the presence or absence of a particular entity type like “high court”, “Section 120” , etc., in the sentences.

3.4.3 Upper Case Words

Some proper names are often important and presented through upper-case words, as well as some other words that the legal experts want to emphasize. We use this feature to reflect whether a sentence contains any upper-case words like "Chief Commissioner", "Revenue", etc.

3.4.4 Local features and Layout features

One of the main advantages of CRFs is that the model enables the use of arbitrary features of the input. One can encode abbreviated features; layout features such as position of paragraph beginning, sentences appearing with quotes, etc., all in one framework. We look at these features in the legal document extraction problem, evaluate their individual contributions, and develop some standard guidelines including a good set of features.

3.4.5 State Transition features

In CRFs, state transitions are also represented as features [104]. The feature function f_a defined in Equation. 3.6 is a general function over states and observations. Different state transition features can be defined to form different Markov-order structures. We define state transition features for terms relating to appearance of years attached with *section* and *acts,* thereby indicating periods of time, under the labels *Arguing the case* and *Arguments*. Also the appearance of some of the cue phrases in a label *identifying the case* can be allotted to *Arguments* when they appear within quotes. In the same way, many of the transition features have been added in our model. Here inputs are examined in the context of the current and previous states.

3.4.6 Legal vocabulary features

One of the simplest and most obvious features set is decided using the basic vocabularies from a training data. The words that appear with capitalizations, affixes, and those in abbreviated texts are considered as important features. Some of the phrases that include **v.** and *act/section* are the salient features for *Arguing the case* and *Arguments* categories.

3.4.7 Similarity to Neighboring sentences

We define features to record the similarity between a sentence and its neighboring sentences. Based on the similarity, a common label is allotted to the neighboring sentences also. This may repeatedly happen for the categorization of labels *Arguing the cases* and *Final decision* as some legal points may be reiterated in the force of an argument. For example, in the statements given below:

"This revision is groundless and is liable to be dismissed",

"In the result, the revision is dismissed with costs".

The first statement belongs to *Arguing the cases* and the second statement is on *Final decision.* Due to the influence of similarity feature, we allow both the sentences to be marked under the *Final decision* category.

3.4.8 Paragraph structure

In many documents, paragraphs have an internal structure [105] either starting with a high-level sum-up or providing a summary towards the end. This is very common in the categories of *Arguing the case* and ***Arguments.*** Appropriate feature definition capturing this is helpful.

3.4.9 Absolute location

In newspaper-type documents, location of a sentence can be the single most important feature for selection of that sentence because more will be attempted to be conveyed in the leading sentences [106]. In legal applications, location information, while less dominant, may still be a useful indication for *Ratio of the decision* and *Final decision*.

3.4.10 Citation

There is a strong correspondence between citation behavior and its relevance to rhetorical status, especially for *Arguing the case* and *Arguments* categories, as most of the legal arguments involve citing some precedents. From the manual annotation of legal cases, we observed that this feature could be very helpful, since presence of a citation strongly indicates that a paragraph contains these roles.

In addition to the set of features outlined above, we have also added other features which were perceived based on the annotation report given by legal experts for different sub-domains. The list of cue features is given in Appendix D.

3.5 Legal Corpus

Our corpus presently consists of 200 annotated legal documents related to rent control, income tax, and sales tax act. It is a part of a larger corpus of 1000 documents in different sub-domains of civil court judgments which we collected from Kerala lawyer archive (*www.keralalawyer.com).* Each document in a corpus contains an average of 20 to 25 words in a sentence. The judgments can be divided into exclusive sections like Rent Control, Motor Vehicle, Family Law, Patent, Trademark and Company law, Taxation, Sales Tax, Property, Cyber Law, etc. In this work, we have proposed a generalized methodology of segmentation of documents belonging to different categories of civil court judgments. The header of a legal judgment, containing the information related to a petitioner, respondent, judge details, court name and case numbers, has been removed and stored as a separate header dataset. The annotated corpus has now been made available at http://iil.cs.iitm.ernet.in/datasets.

Even though income tax and sales tax judgments are based on similar facts, the relevant legal sections / provisions are different. The details and the structure of judgments related to rent control domain are not the same for income tax and sales tax domains. Moreover, the roles like *Ratio decidendi* and *Final decision* occur many times spread over the full judgment in sales tax domain, which is comparatively different from other domains. We have implemented the rule-based and CRF methods for rent control domain application successfully, and the results are given in the next section. We then introduced additional features and new set of rules for the income tax and sales tax related judgments. The modification to the rule set and additional features are small in number, but show a good impact on the rhetorical status classification in the sales tax and income tax domains. It is a common practice to consider human performances as an upper bound for most of the IR tasks. Hence, the performance of the system has been evaluated by matching with human annotated documents. The amount of agreement that can be expected between two annotations depends on the number and relative proportions of the categories used. These details are discussed in the next section.

3.6 Intrinsic System Evaluation

We evaluate the system performance in terms of the following components:

- We first report precision and recall values for all the roles, in comparison with human performance. We also compared the performance of SLIPPER especially on the categories of *Ratio decidendi*, *Final decision* and *Identifying the case,* as these three are crucial.
- We use the Kappa coefficient to compare the inter-agreement between sentences extracted by two human annotators for rhetorical role identification in legal judgments.

3.6.1 Evaluating inter-agreement between human annotators

Two remunerated annotators were used for carrying out two parallel annotation tasks. Both are familiar with the reading of legal judgments and understanding the contents of the articles because of their professional experience in the legal field. They were provided with proper written guidelines describing the semantics of the seven rhetorical categories and the relevant information needed for the summary. The evaluations of the annotation were in terms of a formal property of the annotation namely, reproducibility. Reproducibility, the extent to which different annotators will produce the same classifications, is important because it measures the consistency of shared understandings (or meaning) of the two annotators. We used the Kappa coefficient K [107] to measure reproducibility. This measure has been increasingly used in NLP annotation work [108]. In general, precision and recall scores do not take chance agreement into account. The amount of agreement one would expect two annotators to reach by chance depends on the number of relative proportions of the categories used. Kappa has the following advantages over Precision and Recall measure for NLP annotation work [107]:

- It factors out random agreement. (Random agreement is defined as the level of agreement which would be reached by random annotation using the same distribution of categories as that by the real annotators)
- It allows for comparisons between arbitrary numbers of annotators and items.

With P(A) as pairwise agreement, and P(E) as random agreement, the Kappa coefficient controls P(A) by taking into account P(E) which is the agreement by chance, as per the following equation:

$$K = \frac{P(A) - P(E)}{1 - P(E)} \qquad \ldots\ldots\ (3.14)$$

P(A) is the observed proportion of annotators giving identical responses on the legal judgments, and P(E) is the proportion of annotators that would be expected to give identical responses on the basis of chance alone. As per the formula, K=0 when there is no agreement other than what would be expected by chance, and K=1 when the agreement is perfect. If two annotators agree less than expected by chance, Kappa can also be negative. When we used the measure of Kappa to evaluate the human agreement on the legal judgments, it showed that the two annotators labeled the seven categories with a reproducibility of K=0.836, for a sample test population of 16000 sentences. This is slightly higher than that reported by Teufel & Moens [38] and above the 0.80 mark which Krippendorf [108] suggests as the cut-off for good reliability. Since reproducibility is significantly good between the two annotators, we have taken the score of one of the annotators as a gold standard arbitrarily for the evaluation of our automated results.

3.6.2 Overall Results

The results given in Tables 3.5 through 3.7 show that the CRF-based and Rule-based methods perform well compared to SLIPPER in terms of precision and recall of the important categories, like *Ratio decidendi, Final decision,* etc. We use F-measure, defined by Van Rijsbergen [109] as ((2*P*R) / (P+R)) as a convenient way of reporting Precision (P) and Recall (R) as a single value. The F-measures for the different role categories range from 0.30 to 0.98. The recall for some categories is relatively low. As our gold standard (human annotated) can contain some redundant information for the same category, this is not too worrying. However, low precision in some category (eg., *arguing the case* and *arguments*) could potentially be a problem for later steps in the document summarization process. Overall, we find the results encouraging, particularly in view of the subjective nature of the task. Figures 3.12 through 3.14 illustrate that the F-measure performance results of the CRF-based method are closer to the gold standard.

In NLP studies, confusion matrices have typically been used to evaluate annotation tasks [110,111]. A confusion matrix [38] contains information about actual and predicted identifications of roles by our rule based and CRF-based methods. It represents the comparison of automatic and human annotated sentences belonging to seven different roles in this study. Performance of such systems is commonly evaluated using the data in the matrix. It is typically called a matching matrix. Each column of the matrix represents the instances in a predicted role, while each row represents the instances in an actual role. One benefit of a confusion matrix is that it is easy to see if the system is confusing between two identified roles (i.e. commonly mislabeling one as another). Here we use it to represent the sentence count between system-generated and human-generated labels, and to discuss the difficulty of identification of certain roles on legal judgments in our work. In this connection, we have created two different confusion matrices as in Figure 3.15 through 3.16 related to rent control domain for rule-based and CRF-based methods. The numbers represent absolute sentence numbers, and the diagonal (boldfaced numbers) are the counts of the sentences that were identically classified by both system and the annotator.

Table 3.5 Precision, Recall and F-measure for the seven rhetorical roles related to Rent control domain

Rhetorical Roles	Precision			Recall			F-measure		
	SLIPPER	Rule-based	CRF	SLIPPER	Rule-Based	CRF	SLIPPER	Rule-Based	CRF
Identifying the case	0.641	0.742	0.846	0.512	0.703	0.768	0.569	0.722	0.853
Establishing the facts of the case	0.562	0.737	0.824	0.456	0.664	0.786	0.503	0.699	0.824
Arguing the case	0.436	0.654	0.824	0.408	0.654	0.786	0.422	0.654	0.805
History of the case	0.841	0.768	0.838	0.594	0.716	0.793	0.696	0.741	0.815
Arguments	0.543	0.692	0.760	0.313	0.702	0.816	0.397	0.697	0.787
Ratio Decidendi	0.574	0.821	0.874	0.480	0.857	0.903	0.523	0.839	0.888
Final decision	0.700	0.896	0.986	0.594	0.927	0.961	0.643	0.911	0.973
Micro-Average of F-measure							**0.536**	**0.752**	**0.849**

Table 3.6 Precision, Recall and F-measure for the seven rhetorical roles related to Income Tax domain

Rhetorical Roles	Precision			Recall			F-measure		
	SLIPPER	Rule-based	CRF	SLIPPER	Rule-based	CRF	SLIPPER	Rule-based	CRF
Identifying the case	0.590	0.726	0.912	0.431	0.690	0.852	0.498	0.708	0.881
Establishing the facts of the case	0.597	0.711	0.864	0.512	0.659	0.813	0.551	0.684	0.838
Arguing the case	0.614	0.658	0.784	0.551	0.616	0.682	0.581	0.636	0.729
History of the case	0.437	0.729	0.812	0.418	0.724	0.762	0.427	0.726	0.786
Arguments	0.740	0.638	0.736	0.216	0.599	0.718	0.334	0.618	0.727
Ratio Decidendi	0.416	0.708	0.906	0.339	0.663	0.878	0.374	0.685	0.892
Final decision	0.382	0.752	0.938	0.375	0.733	0.802	0.378	0.742	0.865
Micro-Average of F-measure							**0.449**	**0.686**	**0.817**

Table 3.7 Precision, Recall and F-measure for the seven rhetorical roles related to Sales Tax domain

Rhetorical Roles	Precision			Recall			F-measure		
	SLIPPER	Rule-based	CRF	SLIPPER	Rule-based	CRF	SLIPPER	Rule-based	CRF
Identifying the case	0.539	0.675	0.842	0.398	0.610	0.782	0.458	0.641	0.811
Establishing the facts of the case	0.416	0.635	0.784	0.319	0.559	0.753	0.361	0.595	0.768
Arguing the case	0.476	0.718	0.821	0.343	0.636	0.747	0.399	0.675	0.782
History of the case	0.624	0.788	0.867	0.412	0.684	0.782	0.496	0.732	0.822
Arguments	0.500	0.638	0.736	0.438	0.614	0.692	0.467	0.626	0.713
Ratio Decidendi	0.456	0.646	0.792	0.318	0.553	0.828	0.375	0.596	0.810
Final decision	0.300	0.614	0.818	0.281	0.582	0.786	0.290	0.598	0.802
Micro-Average of F-measure							**0.407**	**0.637**	**0.787**

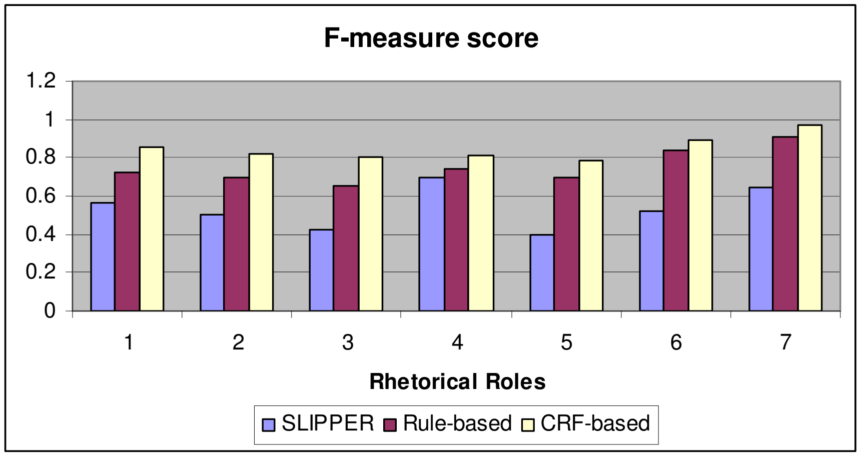


Figure 3.12 **P**erformance as given by F-measure - Rent Control domain

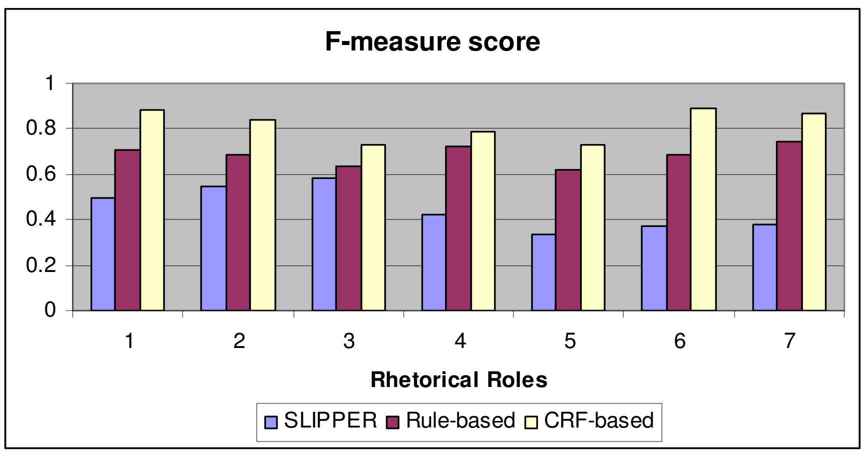


Figure 3.13 Performance as given by F-measure - Income Tax domain

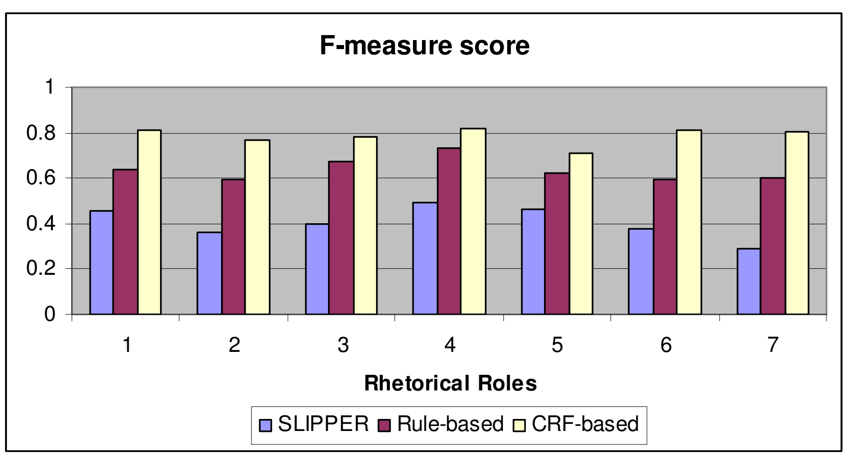


Figure 3.14 Performance as given by F-measure - Sales Tax domain

Figure 3.15 shows a confusion matrix between one annotator and the rule-based system. Given a confusion matrix, where the cells on the matrix diagonal show the agreement of our proposed system results with human annotator. In the confusion matrix given below, of the 823 sample of actual *arguments* role, the system predicted that 83 samples were of the role *arguing the case.* Out of the 159 samples of *arguing the case*, it predicted that 28 samples were *arguments.* We can see from the matrix that the system has trouble distinguishing between the roles *arguments* and *arguing the case.* It also shows a tendency to confuse *history of the case* with other roles. It is able to identify the other roles in the document pretty well.

	Identifying the case	Establishing the facts of the case	Arguing the case	History of the case	Arguments	Ratio of the decision	Final decision	Total
Identifying the case	**49**	2	0	5	5	0	0	61
Establishing the facts of the case	4	**309**	1	28	7	37	18	404
Arguing the case	2	1	**112**	12	28	4	0	159
History of the case	51	84	97	**1826**	289	167	34	2548
Arguments	5	11	83	111	**578**	28	7	823
Ratio decidendi	2	12	0	12	7	**185**	10	228
Final decision	0	0	0	6	3	6	**90**	105
Total	113	419	293	2000	917	427	159	4328

Figure 3.15 Confusion Matrix: human v. automatic annotation - Rule based

Figure 3.16 shows a confusion matrix between one annotator and the CRF-based system. The system is also likely to confuse *arguing the case* and *arguments* roles (e.g. 823 samples of *arguments*, the system predicted that 58 were incorrectly classified as *arguing the case* compare to 83 misclassified samples in Figure 3.15). It also shows a tendency to confuse *history of the case* and *arguments* roles. Comparing the Figures 3.15 and 3.16, we found that there were a lot of improvements in the accuracy level with the CRF-based system. This analysis further emphasizes what was earlier illustrated with precision, recall and F-measure scores in Figures 3.12 through 3.14.

System / Actual	Identifying the case	Establishing the facts of the case	Arguing the case	History of the case	Arguments	Ratio of the decision	Final decision	Total
Identifying the case	**53**	2	0	4	1	1	0	61
Establishing the facts of the case	2	**358**	1	24	7	4	8	404
Arguing the case	2	1	**128**	7	19	2	0	159
History of the case	51	54	20	**2048**	218	143	14	2548
Arguments	2	11	58	40	**697**	13	2	823
Ratio decidendi	1	2	0	10	3	**206**	6	228
Final decision	0	0	0	2	0	2	**101**	105
Total	111	428	207	2135	945	371	131	**4328**

Figure 3.16 Confusion Matrix: human v. automatic annotation - CRF based

3.6.3 System Output: the example judgment

In order to give a better impression of how the figures reported in the previous section translate into real output, we present in Figure 3.17 the output of the system when run on the example judgment (only important roles needed for final summary are given here). The second column shows whether the human annotator agrees with the system decision (a tick for correct decisions, and the human preferred category for incorrect decisions). 11 out of 15 labelled sentences have been classified correctly.

System	Human	
Establishing the facts of the case	History of the case	In the instant case there is no fixed term lease, but the lease deed as only given an option to the tenant to continue on condition on an increase of 10% in the monthly rental amount every three years.
	√	We find no reason to disturb the said finding.
Arguing the case	√	When the matter came up for hearing counsel appearing for the revision petitioners submitted that the Appellate Authority has not properly appreciated the terms of lease deed A7.
	√	Counsel submitted term of the lease is liable to be extended every three years at the option of the tenant an such option has been exercised by the tenant continuously till date and even threafter during the pendency of the present proceedings. In order to establish this contention reference was made by the counsel to the decision of the Apex Court in Laxmidas Babudas, Darbar v. Rudravea, 2001 (3) K.L.T. 324.
	√	Counsel further submitted the Appellate Authority failed to consider the spirit and import of Section 11 (9) of the Act.
	History of the case	Consequently, Rent Control Appellate Authority ought not have ordered eviction on the grounds under Sections 11 (3) or 11 (8) of the Act.
Arguments	√	As held by the Apex court in Nai Bahu v. Lala Ramnarayan, AIR 1978 SC 22 the provisions in the Rent control Act would prevail over the general law of the landlord and tenant
	√	Rent Control Act is a piece of social legislation and is meant mainly to protect the tenants form frivolous eviction.
	Arguing the case	The Apex Court in Muralidhar Agarwal v. State of U.P. AIR 1974 SC 1924 held that an agreement in the lease deed providing that the parties would never claim the benefit of the Act and that the provisions of the Act would not be applicable to the lease deed is illegal.
Ratio decidendi	√	We are of the view that a tenant or landlord cannot contract out of the provision in the Rent Control Act if the building lies within the purview of the Rent Control Act.
	√	We are of the view that clause, as such do not take away the statutory right of the landlord under the Rent Control Act.
	√	We are of the view landlord has made out sufficient grounds in the petition under Section 11 (3) of the Act.
	√	We are of the view in the facts and circumstances of the case, landlord has established the bona fide need for own occupation under section 11 (3) as well as under section 1 (8).
Final decision	√	Revision lacks merits and the same is dismissed in limine.

Figure 3.17 System output compared with human-generated for an example judgment

The above example also shows that the determination of rhetorical status is not always straightforward. For example, the first *establishing the fact of the case* labelled sentence which the system proposes should be classified as *history of the case.* It also shows a tendency to confuse between *arguing the case* and *arguments* roles. An intrinsic evaluation of final summary given in chapter 6 shows that the end result provides considerable added values when compared to earlier sentence extraction method.

3.7 Discussion

We agree that the set of sentences chosen by the human annotator is only one possible gold standard for evaluation of our system results. What is more important is that humans can agree on the rhetorical status of the relevant sentences. Liddy [112] observe that the agreement on rhetorical roles was easier for professional annotators than selection of relevant sentences.

Generally, there is no uniform agreement on which individual sentences should go into an abstract, but there is better agreement on which rhetorical information makes up a good summary. The task for our annotators was to classify the sentences from a set of 200 documents related to three sub-domains namely rent control, income tax and sales tax into seven rhetorical categories. We found that the agreement between annotators is very high as measured by the Kappa coefficient.

We evaluate our work in two steps: first, the evaluation of inter-agreement between two annotators on the rhetorical role identification of the legal judgments; and then compare the evaluation results of CRF-based with other methods. The first step is helpful in defining a gold standard for the human annotation. The evaluation of the second step looks promising enough as we obtained more than 80% correct identification in 5 out of 7 categories (which included the most important *Identifying the case* and *Ratio decidendi*) and nearly 70% in *Arguments* and *Arguing the case* rhetorical roles as shown in Table 3.5 through 3.7. Making use of the precision and recall values for the seven rhetorical categories using CRF model (Tables 3.5 through 3.7), it is verified that the system performs well for *Ratio decidendi* and F*inal decision* which are the main contents for the headnotes generated by human experts. The role *Identification of the case* may not be precisely identifiable for some of the documents. To overcome this difficulty the ratio is rewritten in question format in such cases which improves the readability of the final summary.

3.9. Conclusion

In this chapter, we have presented an annotation scheme for the rhetorical structure of the legal judgments, assigning a label indicating the rhetorical status of each sentence in a specific portion of a document. The annotation model has been framed with domain knowledge and is based on the genre analysis of legal documents.

This chapter also highlights the construction of proper features sets for the efficient use of CRFs in the task of segmentation of a legal document along different rhetorical roles. The identified roles can help in the presentation of extracted key sentences at the time of final summary generation. While the system presented here shows improvement in results, there is still much to be explored. The segmentation of a document based on genre analysis is an added advantage and this could be used for improving the results in the later stages of document processing.

CHAPTER 4

ONTOLOGY BASED QUERY PROCESSING FOR THE LEGAL DOMAIN

This chapter addresses the problem of developing a legal ontology from a given legal corpus in order to facilitate quick access to the relevant legal judgments. The knowledge base developed in this process can be used for enhancing the user query and hence to retrieve more judgments which satisfy the user requirements. The retrieved judgments are used for generating a document summary which will make the user understand the related case histories quickly and effectively. This will be useful to the advocates at the time of handling new cases, and also to the judges who want to write a judgment for new cases. We have proposed a novel comprehensive structural framework for the construction of an ontology that supports the representation of complex legal judgments. The terms defined in the ontology contain the word features as well as various other features like handling of multiple words, synonyms, etc., which can guide the enhancement of queries in retrieving relevant judgments.

In this study, the legal ontology has been constructed by utilizing the legal concepts from a *source ontology* and also from a *case ontology*. The source ontology is a collection of terms built by establishing semantic relationships between the terms selected from legal sources. A case ontology describes various violations, claims etc., in a hierarchical tree structure that is used for determining the legal rights appropriate for a given case.

We evaluate the proposed system using queries generated by legal and non-legal users. In this process, the performance of the proposed system is compared with the human generated search results and also with standard Microsoft Windows search query results [113]. A software environment has been developed to help a legal user to query the knowledge base with his domain experience to extract the relevant judgments.

4.1. An Ontology for Information selection

An ontology is defined as an explicit conceptualization of terms and its relationship for a domain [7]. It is now widely recognized that constructing a domain model or ontology is an important step in the development of knowledge based systems [8]. In this work, we describe the construction of a legal ontology that is useful in designing a legal knowledge base to answer queries related to legal cases [9]. The purpose of the knowledge base is to help in understanding the terms in a user query by way of establishing a connection to legal concepts and exploring all possible related terms and relationships. As a query enhancement methodology, all user queries are expanded with the help of the knowledge base, and relevant documents are retrieved specific to the query terms. The retrieved documents are processed by an extraction-based summarization algorithm to generate a summary for each judgment. The proposed work can assist the legal community to have access to the gist of the related cases that have bearing on their present case instead of having to read the full judgments. The architectural view of the system is shown in Figure 4.1

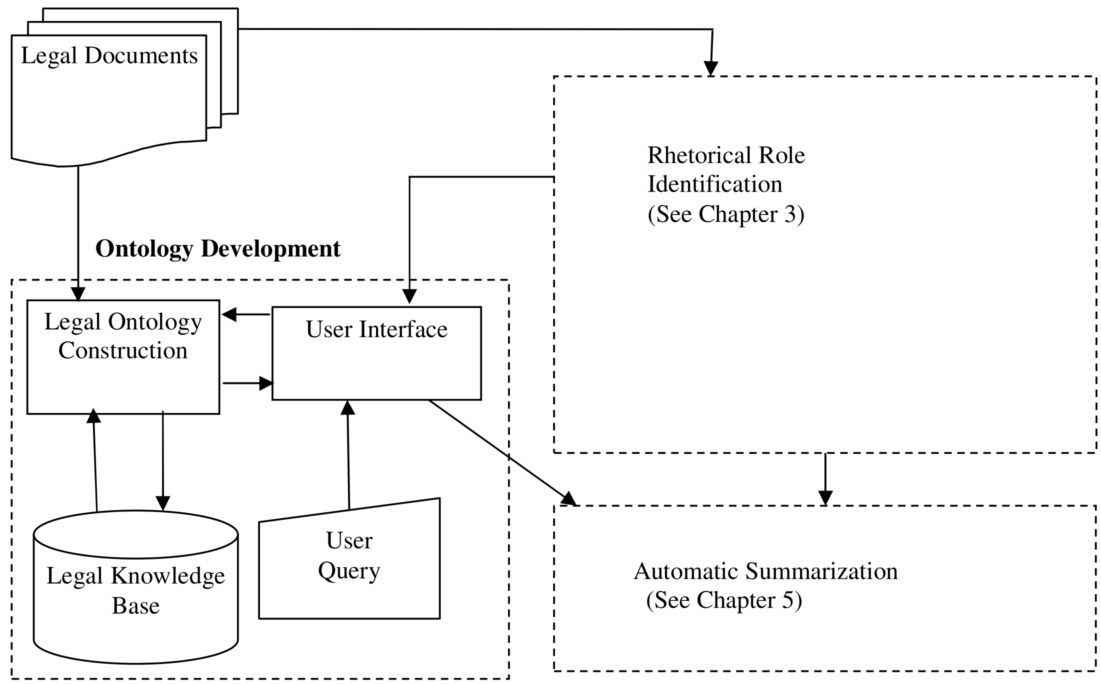


Figure 4.1 System architecture of an ontology-based information retrieval

4.2 Importance of new ontology creation for a legal domain

Legal ontologies are useful to design legal knowledge system, and to solve legal problems [9]. In general, an ontology is used to conceptualize a domain into a machine-readable format [114]. Researchers have been developing different legal ontologies for more than decade. These ontologies were constructed for various projects concerned with the development of legal knowledge system and legal information management. Among the best known legal ontologies the following ones can be mentioned: FOLaw (Functional Ontology of Law) [114], LRI Core [115]; Frame-based Ontology [52], and more recently CLO (Core Legal Ontology) and Jurwordnet [116]. The details of the ontologies were outlined in Chapter 2. These ontologies were developed for computational linguistics studies, sociological studies, etc., or were focused on specific legal environments. The different ideas initiated in the above studies were considered based on which we constructed a new legal ontology. It is for the purpose of query enhancement through the use of semantically and thematically related-terms based on legal theories [116,117]. The previous studies discussed in Chapter 2 confirm that most of the ontology development methods assume manual construction, although, very few methods have been proposed [57] for constructing ontology automatically.

Ontologies are designed for particular purposes [117]. Assessing the adequacy or suitability of an ontology can only be done given the purpose the ontology is created for. The criteria which an ontology must fulfill in order to provide the basis for knowledge representation are far stricter with respect to completeness and the detail required than the one which is merely meant to characterize an approach to legal knowledge systems for the purpose of contextualizing work. Following are the motivations for producing ontologies [117] in the context of:

- Knowledge sharing: For sharing knowledge between different knowledge systems, whatever stored must be make explicit so that there is a common understanding of the represented knowledge.
- Verification of a knowledge base: As the acceptability of knowledge is ascertained by testing, correct behavior of a system is taken to imply that the knowledge base is correct.
- Software engineering considerations: There is a requirement for proper documentation, to guide both end-users and any future maintainers of the related process/resources. An ontology here provides much of the documentation for supplying definitive answers to questions of this sort.

- Knowledge acquisition: It is the process guided by the experts and the skill of the knowledge engineer in eliciting knowledge from them. If the conceptualization is explicit, the knowledge engineer will have a framework by which to guide the knowledge acquisition.
- Knowledge re-use: It is a process to exploit the design of systems in the same or related domain. In the case of the same domain knowledge can simply be adopted and used to supply the vocabulary to design another system, otherwise, it may be necessary to refine the conceptualization to take it to a greater level of detail and extend it for a new application in the domain. In any event the "ontology" of one application will be a reusable component in the design of a new application.
- Domain-theory development: When dealing with radical differences in the way in which the domain is conceptualized, an ontology will greatly facilitate fruitful discussion and comparison of different approaches.

All or any of the above could be a sufficient reason to provide an ontology while designing a knowledge system [117]. All of the above apply with equal or greater force when the knowledge system is in the domain of law: the inter-relation of law makes it a natural area for knowledge sharing; the importance of legal decisions argues for a high level of verification; the rate of change of law argues for readily maintainable systems, a well-known software engineering problem; knowledge acquisition is no less a problem in law than in other domains; the similarity of different branches of law urges the design of re-usable frameworks, and the lack of fundamental theoretical agreement suggests that we should reap whatever insights that are available in the way of domain theory development.

In our work, a methodology to create a legal ontology from a set of legal documents has been proposed. It considered the issues of knowledge sharing, acquisition and re-use. It is based on the following steps:

- Definition of source ontology/ case ontology;
- Identification of concepts referred to in the legal documents and extraction of its properties;
- Identification of relations between the identified concepts;
- Definition of an initial top-level ontology based on our novel framework;
- Creation of an ontology using the identified concepts and relations;
- Addition of different features for identified concepts to enable better query enhancement process;
- Development of a software environment for easy access of knowledge base depending on the user query.

The legal ontology in our context has been defined as an ontology of Indian law tailored for information retrieval tasks. We have considered three sub-domains: rent control, income tax and sales tax, for the legal ontology construction. This can be extended for other sub-domains. We use the codes in Indian law as the basis of legal norms used to infer the legal concepts, and the semantic relations among the concepts. Such an ontology is useful in the information retrieval contexts by providing more judgments relevant to an user query, as well as in the broad access of legal knowledge bases. For the ontology development, a new framework has been designed considering all the top level components and other related components in the form of general conceptual hierarchical structure. The hierarchical structure identified in this process guides the construction and establishment of the relationship between the terms that represent a legal concept. The above discussions emphasize the need for a new legal ontology for query enhancement. In general, the construction of legal ontology is a difficult problem which will be discussed in the next section.

4.3 Difficulties in constructing a legal ontology

Building an ontology for legal information retrieval purposes for such a vast domain leads to some difficulties [118].

1. The difficulty is essentially due to the complexities of the domain: specific vocabularies of the legal domain and legal interpretations of the expressions that can produce many ambiguities.
2. Legal experts may disagree on some points such as determining whether a given legal concept is effectively a part of another legal concept, or in assessing whether a given legal concept is relevant to different legal sub-domains.
3. Large variability of definitions of legal concepts is another issue. The meanings of legal concepts, and the concrete facts or points covered by it (labeled by legal terms) may vary. Also, the legal concept definition can vary depending on the specific period or the judges.

4. Compared with other domains it is more difficult in the legal field to identify the sub-domain terms in view of the large variability, which may even be subjective.
5. Like in all the domains, lexical phenomena such as synonymy, polysemy, etc., are encountered in the legal fields also. Association among the terms with different degrees of similarity causes the synonymy problem, and polysemy is due to the variation of term definitions in the concerned legal sub-domain.
6. The ontology can be created automatically, but it is not possible to create many hierarchical relations among the concepts.

This research work addresses several challenges in ontology creation, maintenance, retrieval from the original documents. To provide valuable, consistent ontology-based knowledge services, the ontologies are manually created with high-quality instantiations. The manual ontology creation is labor intensive and time consuming, while automatic ontology creation is difficult to implement [57]. Some semiautomatic approaches create document annotations and store the results as assertions in an ontology. Other methods add relationships automatically between the instances only if they already exist in the knowledge base; otherwise, user intervention is requested. But no specific method can be followed to do both automatic creation and establishment of the relationship between the concepts without human intervention. According to Gruber [7] "*Automatically populating an ontology from diverse, distributed web resources also presents difficulties, particularly in that of consolidating duplicate information that arises while extracting similar or* *overlapping information from different sources*". The legal ontology designed in our study deals with distinct concepts and its relationships related to three sub-domains, and it is arranged in a hierarchical structure to avoid duplication and irrelevant information. Thus, the construction of an ontology needs a basic framework to include all the variabilities of legal concepts and also to establish a good relationship between the concepts. Our work on ontology construction addresses the following issues:

- Development of a source and a case ontology related to the legal domain which provides semantic and basic information needed for the terms. Key legal concepts are derived based on the terms and their relationships.
- Development of a novel structural framework for the building up of a knowledge base.
- Using the ontology to describe and structure the terms and their relations that need to be probed in three different sub-domains (Rent Control, Sales Tax and Income Tax).
- Presenting a Meta level view on the sub-domains which explore the terms and their relationships from case ontology.
- Enhancing the terms mentioned in the user query to minimize the irrelevant responses using the legal ontology.
- Addition of new documents to the legal ontology through XML representations. This process may be automated later.

4.4 Ontologies Representation

In this section, we consider how to use source and case ontologies in building a legal ontology in order to find out the key concepts. In this process, there are two hard issues which need to be addressed. The first one is to find out the best categorization of a given legal concept in source and case ontologies, and second one is to generalize the concepts defined in the source ontology so as to find out a structural framework to create the legal ontology.

4.4.1 A Source Ontology

In this work, initially a focused domain dictionary that contains more than 1000 words is created. These words were extracted from the legal sources some of which are exclusively related to the sub-domains considered. Each of the words in the dictionary has basic, semantic, and supplementary information associated with it. The basic information corresponds to the meaning of the word and represents the common public knowledge. From the basic information, we proceed to create the semantic information by establishing its relationship to other words in the legal domain. The words in the legal domain represent either a *process* (doing) or a *status* (being). A legal description of a process or a status is known as a concept. The semantic information provides links like *is a*, *kind of*, etc. to different words, each one of which may come under one or more distinct concepts. The detailed descriptions of the various links supported by our ontology are given in Appendix B. Such descriptions are mostly found in statutory enactments or in judicial interpretations which are also known as case law/precedents. The semantic information of a word establishes the distinct context associated with it. Higher level concepts can be defined by grouping together lower level concepts based on the similarity of their descriptions. At the end, the whole schematic converges into a single concept through a hierarchical structure which represents the law (sub-domains) under consideration. For example, the words ‘building’ and ‘eviction’ give the basic information that building is a *property* and eviction is a *forcible act*. The semantic information is shown in Figure 4.2.

Building	**Eviction**
It is a *status* (*things*) It is a *property* In property, it is *tangible* property In tangible, it is immovable property Being immovable property, it is covered by *transfer property act, land acquisition act, urban ceiling act, city tenants protection act, land reforms act, rent control act*, etc.	It is a *process (events)* It is a forcible act In that it can be legal or illegal. It can be covered by *Indian penal code*, *land acquisition act, city tenants protection act, land reforms act, rent control act*, etc.

Figure 4.2 Sample information of source ontology

From Figure 4.2, it can be seen that the concept of transfer of property and the concept of Indian penal code apply to the words ‘building’ and ‘eviction’ respectively, but not to both. So the applicable concepts (sub-domains) cannot be transfer of property act or Indian penal code. By extending the above process of elimination through establishing a semantic relationship to other words in a legal document, we will be able to identify ultimately a unified concept which is the law (sub-domain) applicable. The other features of word like synonyms, related words and the types of relations are considered as supplementary information. The dictionary of words along with associated information is compiled to form the source ontology and details are given in Appendix B.

4.4.2 A Case Ontology

A case ontology depicts the relationships between various legal rights in the form of a tree structure as shown in Figure 4.3. The concept of legal rights comes under different categories like constitutional, civil, criminal, etc. The rights and remedies with reference to a *status* or a *process* are determined by a recognized law. A law is a predetermined set of rules governing relations in a society ensuring safeguard of these rules by punishing any violations. These laws are made through enactments, by customs and practices, or by judicial interpretations. All of these are grouped under the category *acts* in our proposed framework.

A status can be of *groups*, *persons* or *things*. Similarly process relates to *events* which include its consequences. *Facts* of a case may relate to status and process. Rights and remedies cannot be determined without reference to a recognized law. For instance, in a reported case number 03KLC-1058 (www.kerelawyer.com), a tenant was resisting eviction from a building on the ground that a daughter-in-law of the landlord cannot be a family member, and hence eviction on the ground of bona-fide need for own use or use of family member of a landlord is not applicable. It was held by the court that the daughter-in-law came within the scope of family members under rent control act, and hence eviction was applicable. Here, the status is building (things), tenant, landlord (persons), and the process is eviction (event). Facts of the case are *bona-fide own use* and claim by *daughter-in-law* as a *family member* of landlord. Civil right of eviction has been determined under the sub-domain rent control act.

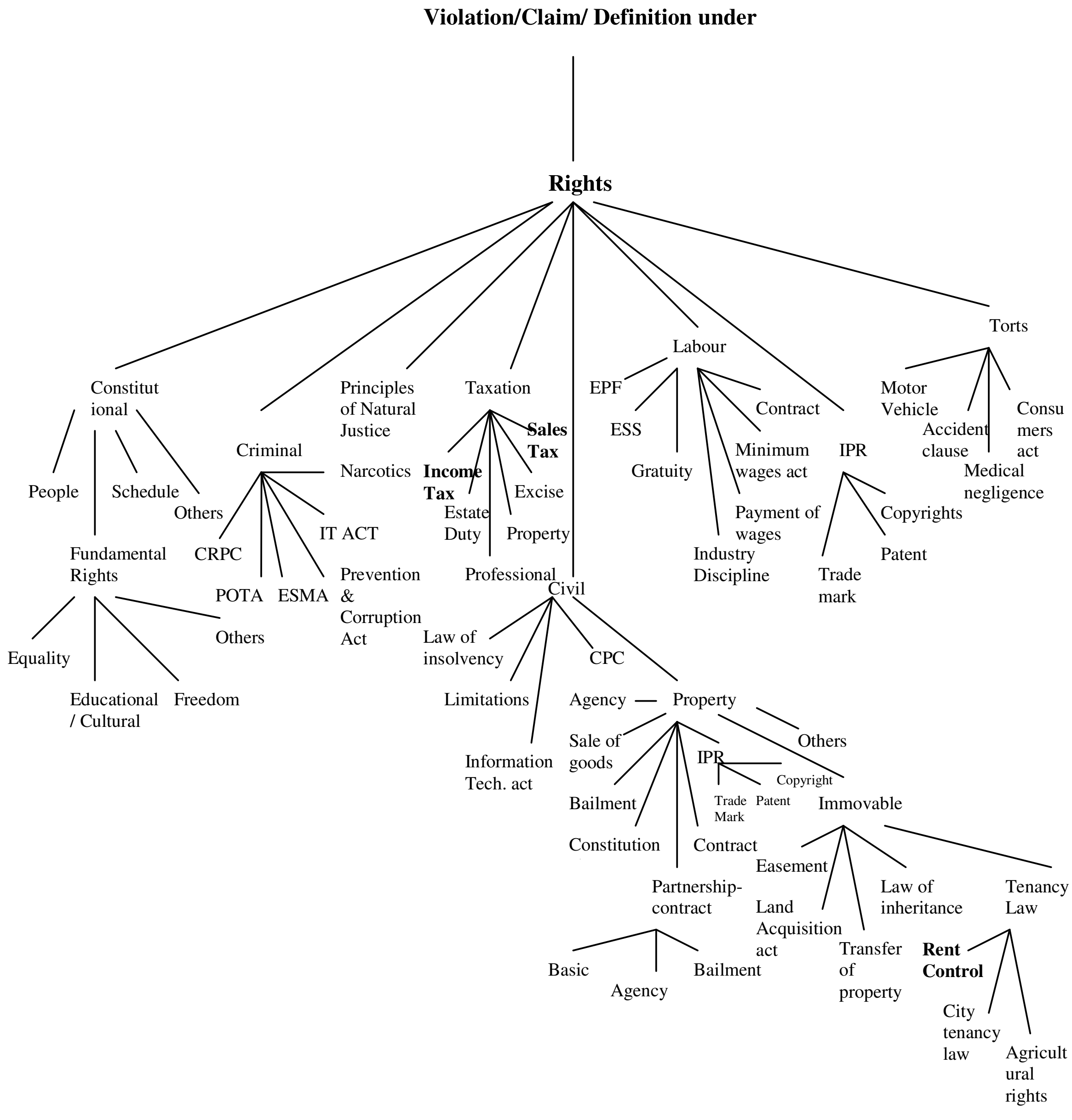


Figure 4.3 A Hierarchy in a case legal ontology

It can therefore be seen that the semantic and basic information provided with words and concepts (from source ontology) lead us to develop a framework for construction of the legal ontology. The current study focuses only on civil statutes in which sub-domains taken for consideration are rent control, income and sale tax acts. These sub-domains are selected having regard to the area of specialization of the annotators engaged in this study. The case ontology structure shown in Figure 4.3 was verified for its appropriateness by a human expert.

4.5 A Proposed Legal Ontology Framework

As already stated, a law is essentially about determination of rights and remedies under a recognized law (*acts*) with reference to *status (persons and things)* and *process (events)* having regard to the *facts* of the case. The five basic components namely *persons, things, events, facts,* and *acts* are used to develop a framework for the construction of legal ontology. In our framework, we first identify the hierarchal structure of legal concepts through source and case ontologies, and fit them into standard basic components. The related terms and their basic components are identified to construct the legal ontology. The components of proposed framework were identified based on several discussions we had with the legal communities related to different sub-domains. A partial view of legal ontology is given in Figure 4.4.

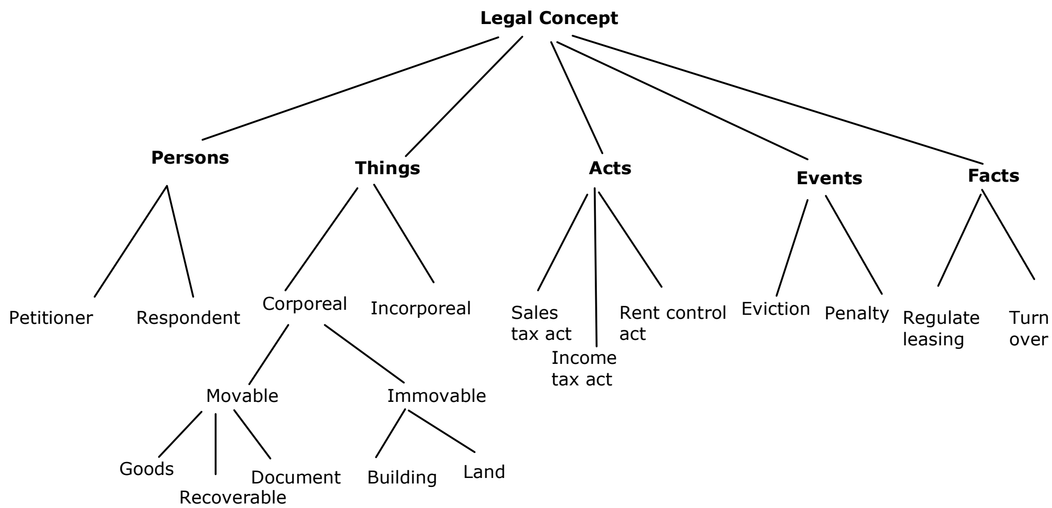


Figure 4.4 An extract of legal ontology framework on selected sub-domains.

There are two main approaches for building an ontology [120]. The first is a bottom-up approach. In this approach, all the elements needed are extracted from appropriate documents to compose an ontology. The main difficulty is that it needs more information to determine which documents are appropriate to compose a reference corpus. The second is the top-down approach. The top-down approach begins by asking domain experts to agree on a unique point of view in their specializations. This unique point of view is taken as the basis for constructing an ontology. The main difficulty is the time duration taken by experts to come to an agreement. But it gives better results compared to the first one for IR-oriented ontology.

Table 4.1 Basic description of the components of ontology framework

Components	*Description of components*
Person	Two contending parties appear before the court to resolve their conflict; one of the parties will be the subject. The object will be either a thing or a person.
Things	An object can be either a thing or animate beings including humans. The thing can be either corporeal or incorporeal. Subject action on the object or with respect to object has an impact or consequence on the other contending party.
Event	The ultimate /last facts in a process (series) of facts that have given rise to the conflict/that have triggered conflict between the contending parties. The conflict is about the duty/obligation (or right) of contending parties.
Facts	This constitutes the process. Subject's actions in relation to an object consequently constitute the process that has or is likely to have or is apprehended to have an impact on the contending parties. Facts and circumstances of the case that are serious / relevant make up the process.
Acts	Courts always deal with an application of law to the given facts. The law applied is extracted from the relevant statutes / provisions / rules / regulations / articles / judicial interpretations that are called acts.

We present a top-down approach in the construction of sub-components by using domain knowledge. A top-down approach starts with the definition of the most general concepts in the domain along with subsequent specialization of those concepts. For example, we start with the class of legal concepts. Then we refine this further by creating some of its subclasses corresponding to top-level components. An initial ontology hierarchy given in Figure 4.4 has five important top-level components (classes) of knowledge in the legal domain – *person, things, event, facts* and *acts* – which are identified and their descriptions are given in Table 4.1.

We can further categorize the *person* class, as for example into *petitioner* and *respondent*, and so on. There are different kinds of relations like *is-a, kind-of, composed-of* etc., which are used in the formation of an entire ontological structure to describe the relationship between this term and the other terms, and the semantics.

Now, we will discuss how to extract the domain terms and establish the relationship between them.

4.5.1 Selecting domain terms and identifying relations among terms

Initially, a list of candidate domain terms is identified from the source ontology. These candidate terms are grouped under different top level components to exploit what is called *discourse structure* suggested by Moens [121]. These terms are then thematically divided into different sub-components. These sub-components are given titles to reflect the underlying themes. We did a simple frequency analysis on this list to prune some uninteresting terms. We consider the remaining terms as domain terms. Some of the most frequent domain terms are *act*, *section*, *goods*, *building*, *receivable* etc. This sub-list allows us to define the initial components of the proposed ontology. A comprehensive work has been performed for the compilation of each top level component. As a result, each component is rationally organized and suffers no repetition. We present briefly the critical points for a successful integration of ontologies in a query enhancement scheme for the information retrieval task.

4.6 IR-Oriented Legal ontology

Ontologies play a central role in the representation of legal concepts and can be used to enhance existing technologies from machine learning and information retrieval. Many of the existing systems using ontologies have typically employed bag-of-words method [119], where each single term in the collection is used as a feature for representing document content. Moreover, the systems using only words as features exhibit a number of inherent deficiencies like the inability to handle synonymy, polysemy, etc. In addition to word features, we have considered other features in this study to handle the user queries and to retrieve related documents from our collection. The additional features considered in this study include the handling of multiple words, different words with same meaning (synonymy), word with multiple meaning (polysemy) and also by considering high level abstraction instead of low level abstraction of terms defined in the ontological structure. Figure 4.5 shows a possible breakdown among the different levels of granularity.

Legal is the most general concept. *Person, Things, Event, Facts* and *Acts* are general top-level concepts. *Corporeal* and *Incorporeal* are considered for the middle-level concepts. *Movable* and *Immovable* are the most specific classes in the hierarchy (or the bottom level concepts). In the top-down approach, we usually start by defining the main components. From the list created during initial ontology generation, we select the terms that describe the concepts having independent existence rather than terms that describe these concepts. These terms will be sub-components in the ontology and will become anchors in the hierarchy structure. Figure 4.6 shows a part of the component hierarchy for the legal ontology. We organize the components into a hierarchical taxonomy by understanding the instance of one sub-component. The sub-component will necessarily (i.e., by definition) be an instance of some other component. This hierarchy satisfies the important class property: *If a class A is a super class of class B, then every instance of B is also an instance of A*. The component hierarchy represents an *is-a* relation: a component X is a sub-component of Y if every instance of X is also an instance of Y. For example, *corporeal* is a sub-component of *Things*.

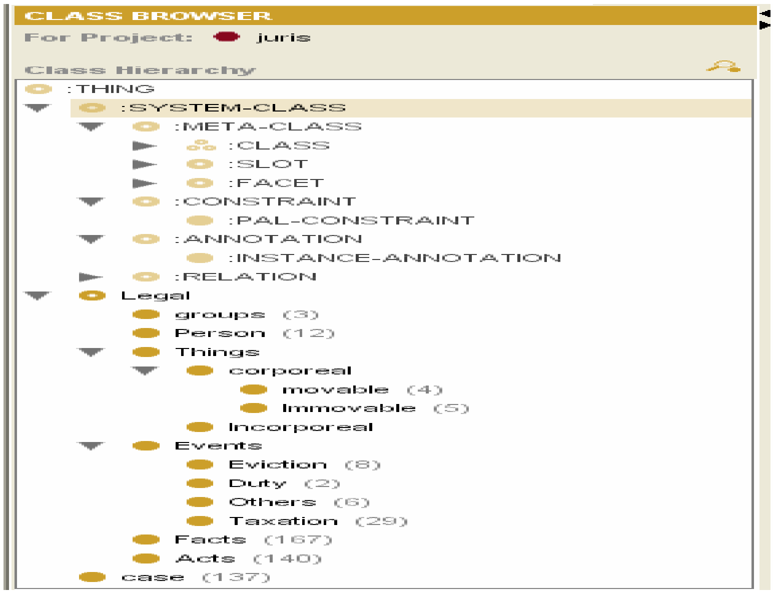
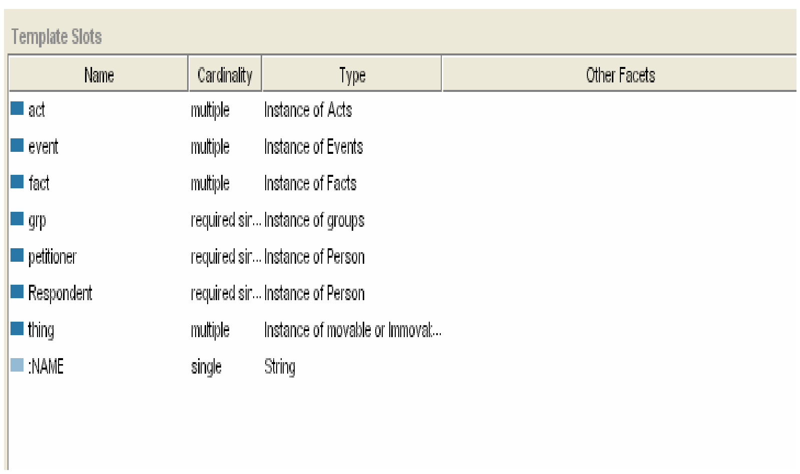


Figure 4.5 Prototype of class hierarchy in a case ontology.


Template Slots

Name	Cardinality	Type	Other Facets
act	multiple	Instance of Acts	
event	multiple	Instance of Events	
fact	multiple	Instance of Facts	
grp	required sir...	Instance of groups	
petitioner	required sir...	Instance of Person	
Respondent	required sir...	Instance of Person	
thing	multiple	Instance of movable or Immovat...	
:NAME	single	String	

Figure 4.6 Properties of components in a legal ontology

Another way to think of the taxonomic relation is as a *kind-of* relation: *Goods* is a kind of *Movable thing*. A *building* is a kind of an *Immovable thing*. We implemented this ontology in protégé (http://protege.stanford.edu), a graphical ontology editor tool that also stores the knowledge base in XML representation. Different types of formats are available to represent the Knowledge base. In this study, we preferred to use XML representation due to the interest of converting the process into a semi-automatic one at a later stage. The idea of using protégé plug-ins is to convert ontological terms into XML representations using class hierarchy structure.


```
<simple_instance>

        <name>01kcl91</name>
        <type>case</type>
        <own_slot_value>
                <slot_reference>petitioner</slot_reference>
                <value value_type="simple_instance">tenant</value>
        </own_slot_value>
        <own_slot_value>
                <slot_reference>Respondent</slot_reference>
                <value value_type="simple_instance">landlord</value>
        </own_slot_value>
        <own_slot_value>
                <slot_reference>thing</slot_reference>
                <value value_type="simple_instance">building</value>
        </own_slot_value>
        <own_slot_value>
                <slot_reference>event</slot_reference>
                <value value_type="simple_instance">eviction</value>
        </own_slot_value>
        <own_slot_value>
                <slot_reference>fact</slot_reference>
                <value value_type="simple_instance">bonafide need</value>
                <value value_type="simple_instance">own use</value>
                <value value_type="simple_instance">necessary
repairs</value>
        </own_slot_value>
        <own_slot_value>
                <slot_reference>act</slot_reference>
                <value value_type="simple_instance">11(2)(b)</value>
                <value value_type="simple_instance">11(8)</value>
        </own_slot_value>
        <own_slot_value>
                <slot_reference>grp</slot_reference>
                <value value_type="simple_instance">rent control</value>
        </own_slot_value>
</simple_instance>
```


Figure 4.7 XML output – A sample instance of annotated legal judgment

The proposed ontology based system moves toward automatic feeding of legal documents to the ontology. XML representation of the training documents belonging to different sub-domains represent information extracted in legal judgments with respect to our ontology framework using tags mapped directly from ontology class and relationship names. Figure 4.7 shows an example (instance) of this XML presentation and how the new structural framework asserts it in the ontology.

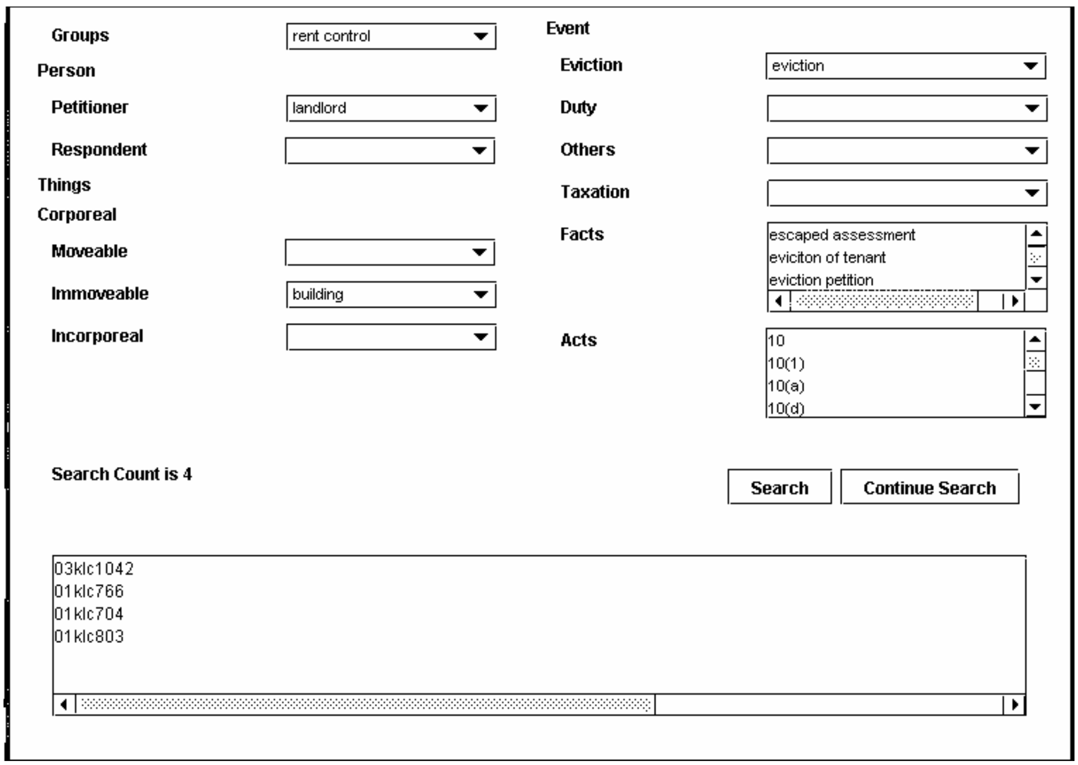


Figure 4.8 Software Environment for Document Retrieval

The user interface is created as a part of system development to make the query enhancement much simpler. The format of user interface is shown in Figure 4.8, which illustrates the expectation of proper usage by legal users to mine the knowledge base to retrieve relevant judgments. It is designed in such a way to help the legal users to choose multiple options to query the knowledge base. Also there is a provision for continuing the search operations by adding more options. For example, a user can begin by choosing an appropriate group rent control. Alternatively a user may chose eviction under events to get all the eviction related cases. This will automatically set the group to rent control. To filter out the eviction cases for particular facts, he can chose the options in the facts column and press the continue search button. It can bring out the filtered information which is more suitable for the user in decision making. In this design, the legal user is able to use the ontology in order to appropriately structure the query so as to retrieve more relevant documents. This is a form of query enhancement that uses the legal ontology as background knowledge. Further enhancement of a query is done by adding the appropriate supplementary information given in the ontology to the query terms. In the next section, we present the results comparing the performance of both levels of query enhancement with the baseline retrieval system.

4.7 Evaluation of Ontology-based query results

Ontology evaluation is an important issue that must be addressed if ontologies are to be widely adopted in information retrieval applications [123]. In general, ontologies have been employed to achieve better precision and recall in the text retrieval systems [124]. In this study, we have measured the effectiveness of ontology-based search results which help the users to judge the relevance of retrieved documents. For this, we have employed measures like Precision (**P**), Recall (**R**) and F-measure (**F**) to evaluate the results of our method with human generated ideal search results. It is also compared with query-based search techniques of Microsoft Windows which has been considered as a baseline in this study [113]. Precision is the ratio of the number of relevant documents retrieved by the system to the total number of documents that human subjects judge as relevant. Recall is the ratio of the number of relevant documents retrieved by the system to the total number of relevant documents in the corpus. F-measure is the weighted harmonic mean of precision and recall. Around 100 queries given by both legal and non-legal users were considered for evaluation. We asked our legal experts to find the set of relevant documents related to the above set of queries. This set has been considered as the "gold standard" for the evaluation of results of windows search and ontology-based methods.

Table 4.2 Precision, Recall and F-measure for comparison of methods

Method/type of user	Legal User			Non-legal user		
	P	**R**	**F**	**P**	**R**	**F**
1.Microsoft Windows Search	0.561	0.724	0.634	0.683	0.772	0.724
2. Ontology based without query enhancement	0.718	0.917	0.805	0.834	0.862	0.837
3. Ontology-based with query enhancement	0.829	0.967	0.893	0.879	0.919	0.899

The findings which given in Table 4.2 show that our ontology-based method yields a relative improvement of 25% compared to the baseline on legal user queries and 15% on non-legal user queries. This difference in success rate is due to the usage of complex legal terms by experts in their query. Comparing the ontology-based methods given in Table 4.2, we find that the one with query enhancement shows better performance than the one without. The integration of many features for a term in the newly developed knowledge base, as discussed in section 4.6, yields excellent improvement in query results compared to the baseline. From Table 4.2, it is clearly seen that there is an increase in precision score in ontology-based with or without query enhancement system. Obviously it is due to the availability of knowledge base. A further improved result with query enhancement procedure is due to the expert usage of our user interface by legal experts. The legal experts are given an option to choose the framework terms which are relevant to their query.

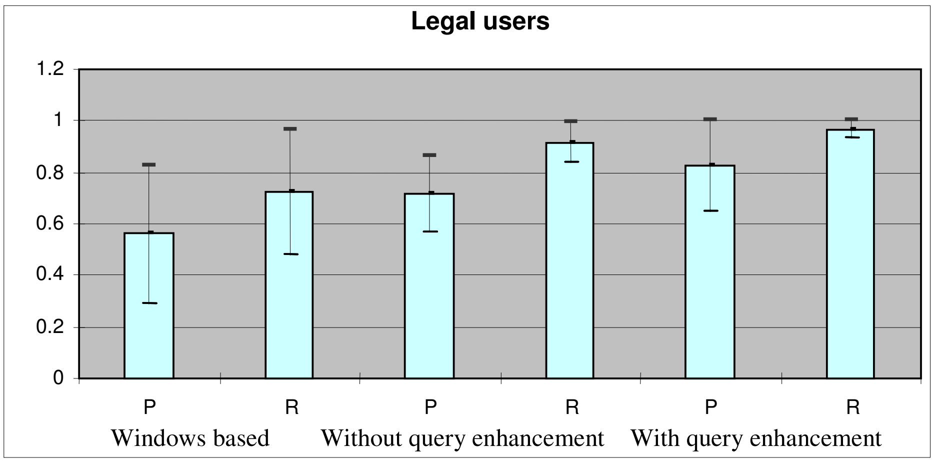


Figure 4.9 Precision and Recall measures based on legal user queries

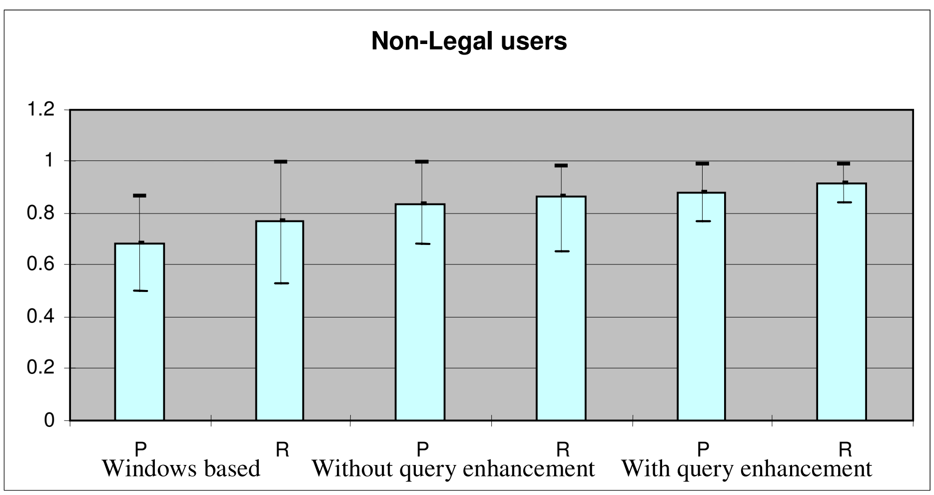


Figure 4.10 Precision and Recall measures based on non-legal user queries

Accuracy is another measure of evaluation to measure with only one answer per question is allowed. That is, there is retrieval of exactly one document for a query. The accuracy measured thus for our ontology-based query enhancement scheme was 96%, which is 18% more than the one for the baseline method. There is an improvement in results for the ontology-based methods over the baseline for all measures, as shown in Figures 4.9 and 4.10.

Table 4.3 Paired t-test values for performance measures of ontology-based with (3) and without (2) query enhancement compared with MS Windows search results (1).

Significance Level for **Performance Measures**	**Precision**		**Recall**		**F-measure**	
	t-value	Sig. Level	t-value	Sig. Level	t-value	Sig. Level
Legal Users						
1 & 2	3.75	$p <. 01$	8.71	$p < .01$	6.04	$p <.01$
1 & 3	6.10	$p < .01$	9.98	$p < .01$	9.70	$p <.01$
Non-Legal users						
1 & 2	4.84	$p < .01$	2.12	$p< .05$	2.80	$p <.01$
1 & 3	7.02	$p < .01$	4.15	$p < .01$	5.08	$p <.01$

To substantiate the significance in the measurement, a paired t-test was applied to the data. The details of paired t-test analysis are given in Appendix C. Table 4.3 show the calculated t-values with the significance level, indicating that the average precision, recall, and F-measure performance measures of ontology-based methods over windows search results and our methods are significantly score higher over the baseline considered at 99% confidence level (Table I of Appendix C).

4.8 Discussion

This aspect of work is concerned with the development of a knowledge based system useful to the legal communities in information search and retrieval. In this study, we have proposed a potentially powerful and novel structural framework for the construction of legal ontology. For developing the framework, we discussed with many legal experts and it can be adapted to other sub-domains also. The overall architectural view of the ontology-based document summarizer is given in Figure 4.1. The documents are retrieved from the knowledge base based on the user query. They will be finally summarized using our sentence ranking algorithm as discussed in the next chapter. Here, we presenting only the *ratio decidendi* as a part of the summary to illustrate the importance of retrieved documents based on a user query given in Figure 4.11.

Our ontology-based query enhancement method consistently outperforms the MS window search method for all the three sub-domains considered in this study. An intuitive explanation for the better performance of our ontology-based system is that it provides a knowledge base which had a huge collection of terms and its relationships and other related features. Microsoft windows search query looks for the exact pattern instead of considering other derived forms of words or phrases.

Query: *Whether the special reserve created has to be maintained continuously to claim the benefit?* Ontology-based query enhancement system returns two documents for the above query which are summarized and here only the ratio decidendi was given for reference
(Before G.Sivarajan P.R. Raman, JJ) Friday, the 14^{th} February 2003/ 25^{th} Magha, 1924 Case No. ITA.No. 191 of 2000 Appellant: Kerala Financial Corporation Respondent: The Commissioner of Income Tax, Cochin.
Ratio decidendi: We find considerable force in the submission made by the learned counsel for the appellant that through the amounts were transferred to "bad and doubtful debts" there was not any existing liability or that there was any known liability. In the absence of any condition that it should be continued to be maintained, there is no warrant to think that the legislature intended to confer the benefit of the provision only if it continued to maintain the reserve. In the above circumstances, we hold that the decision of the Tribunal holding that the assessee is not entitled for the benefit of Section 36(1) (viii) is erroneous of law.
(Before G.Sivarajan K. Balakrishnan Nair, JJ) Monday, the 11^{th} November 2002/ 20^{th} Karthika, 1924 Case No. ITA.No. 161 of 2001 Appellant: The Dhanalakshmi Bank Ltd Respondent: The Commissioner of Income Tax, Cochin.
Ratio decidendi: To make it clear, if the bad debt written off relates to debts other than for which the provision is made under clause (viia), such debts will fail squarely under the main part of clause (vii) which is entitled to deduction and in respect of that part of the debt with reference to which a provision is made under clause (viia), the proviso will operate to limit the deduction to the extent of the difference between that part of debt written off in the previous year and the credit balance in the provision for bad and doubtful debts account made under clause (viia). We are of the view that the matter requires fresh consideration in the light of the said interpretation accordingly, we are of the view that the matter must go back to assessing officer for consideration with reference to the interpretation placed by us in this judgment in the first instance.

Figure 4.11 System outputs (indicative summaries) for a sample query

For example, for a query containing the phrase *rental arrears*, windows can search only the documents with the phrase 'rental arrears' present. But our method checks the documents for rental arrears, rent in arrears, default of payment of rent and such other related phrases. As another example, for a phrase *increase in rent*, our approach can also look for exorbitant rent, and enhancement of rent in addition to the given phrase. Moreover, it can discard the phrases of simple descriptions containing only some word components not relevant to the defined framework. That is, the ontology-based system can consider the next level in hierarchy for a given particular word or phrase.

One of the limitations of this ontology creation is that any errors that may have crept in due to human mistakes, at the time of annotation of documents, will affect the search results. With regard to the difficulties discussed in section 4.3, we make the following observations in our study. The ontology-based system provides significantly more flexibility in retrieving the correct set of judgments for the related query. Achieving something similar with ad-hoc query expansion techniques is difficult due to the enormous number of parameters employed [124]. We have covered semantic variations in the present ontology, but coverage may decrease if the terms cannot be expanded sufficiently. In this approach, we avoid confusion between ontology relation and synonymous names that connect the same components, by specifying appropriate synonyms for the terms in the components and relations to avoid possible ambiguities. Specificity also presents challenges to the ontology creation. For example, we can easily identify a referential entity as a person; but it is harder to deduce whether the person is a *petitioner* or a *respondent*. We could infer that related knowledge is known when we extract more facts about the *person*, such as information about the case *facts* and *arguments*. Likewise, identification of component terms may prove difficult, if extracted sentences contain co-references which usually can be resolved only with the availability of overall contextual details. The user interface created for this application can also play an important role in overcoming some of the difficulties by allowing the legal experts to choose the proper terms and cue phrases based on their query terms.

One significant advantage of working with an ontology framework is that it gives a simple way to integrate other sources of knowledge into the model in an exploratory manner. One could consider, for instance, extending this model to other sub-domains for the retrieval of relevant documents based on a user query, where the user would explicitly model terms and their relationships across several related documents in a given collection. Alternatively, one may fit in the document details belonging to other sub-domains into the same framework for more easy access and query processing, but that will reduce the precision. The user interface developed may have more options thereby enabling the user to reduce the number of terms for query processing.

The representation in XML format of the data can pave the way for an approach to automatically update the details of new documents into the ontology in future. Once the system identifies the concepts present in the new judgment, the document can be converted using a suitable tag set into the XML representation. The newly added judgment in XML format can be updated automatically in a legal ontology using the tools that are available in protégé.

4.9 Conclusion

The present work addresses several challenges in ontology creation, maintenance, information retrieval, and the generation of key information related to legal judgments. User-driven ontology creation tools must avoid duplicating information across judgments. In this work, we have designed a novel structural framework which has guided the development of a legal knowledge base. User queries are enhanced with the rich collection of word features available in the knowledge base to retrieve relevant judgments from the document collection. Finally, the legal ontology which we have proposed plays a decisive role in our summarizer in returning the relevant judgments needed for the legal users. The summarization algorithm is employed to generate a document summary which is discussed in the next chapter.

CHAPTER 5

PROBABILISTIC MODELS FOR LEGAL DOCUMENT SUMMARIZATION

The summary of a judgment, as a compressed but accurate restatement of its content, helps in organizing a large collection of cases and also in finding the relevant judgments for a case. For this reason, the judgments are manually summarized by legal experts. Due to increased diversity of legal document collections, the use of automatic summarization is guaranteed to remain as one of the important topics of research in Legal Information Retrieval Systems. A summary generated from a legal judgment, known as a headnote, enables fast and easy access for arguing a case at hand, with precedents. Extraction of sentences in the generation of summaries of different sizes is one of the widely used methods in document summarization. In addition to the extraction of sentences from the documents, some algorithms automatically construct phrases that are added to the generated summary, in order to make it more intelligible. One problem in this approach is that automatic construction of phrases is a difficult task, and wrongly included phrases will totally degrade the quality of the summary. Hence we have decided to use a purely extraction based approach for generating summary from a legal judgment. Earlier, we had implemented our algorithm in newspaper domain for a multi-document summary, and observed good improvements in the results [11]. Now, we try the same algorithm with changes in parameters suitable for a single document summarization of the legal judgments. Thus, it is a specialization of our earlier work on multi-document summarization.

The drawbacks of some of the existing summarization algorithm were already discussed in Chapter 2. To circumvent those problems associated with summarizers, we pursue a statistical approach that predicts summary-worthy sentences from the input legal documents. Statistical NLP based systems are empirical, re-trainable systems that minimize human efforts [1]. The proportion of terms that are identified for summarization is closely related to the semantic content of the documents. Hence, we applied a probabilistic model, which is a modified version of a term weighting scheme that would improve the performance level of the summarizer. The pre-processed terms in a sentence of a given document, represented by the vector-space model, are further processed by the term distribution model (K-mixture model) that identifies the hidden term patterns, and finally produces the key sentences. The block diagram of the entire system architecture is already given as Fig. 1.1 in Chapter 1.

Probabilistic models of term distribution in documents are getting renewed attention in the areas of statistical NLP and information retrieval [1]. In this chapter, we discuss the initial stages of preprocessing of legal documents. Then we describe the usage of term distribution model, specifically the significance of K-mixture model for the identification of term patterns in the document collection for extracting key sentences and discuss the sentence-ranking algorithm. Finally, the significance of our approach and how the identified roles during text segmentation stage help in the improvement of the final summary generation is discussed.

5.1 Preprocessing

Statistical natural language processing tools are used in the preprocessing stage to filter out stop list words and generate stem words, by avoiding the inflectional forms of terms. The resulting meaningful stems are very useful during the normalization of terms in the term distribution model. One of the major problems in text analysis is that the document size is not known *a priori*. If each of the words in the documents were represented as a term in the vector-space model, the number of dimensions would have been too high for the text summarization algorithm. Hence it is crucial to apply preprocessing methods that greatly reduce the number of dimensions (words) to be passed on to the document summarization process. In addition, it is important that the preprocessing method be robust, i.e., able to cope with noisy text containing grammatical and typographical errors. The proposed system applies a number of preprocessing methods to the original documents, namely *case folding*, *stemming*, *removal of stop words* and *key-phrase identification*. The widely used algorithm for the stemming process is based on the work by Porter [125]. We have already done modifications in the algorithm [126] to get meaningful words (dictionary words) as the stemmer output which was briefly described in our earlier work [11]. As a final step in the preprocessing stage, key-phrases are identified. Each of these preprocessing methods shown in Figure 5.1 is briefly discussed in the next sub-sections.

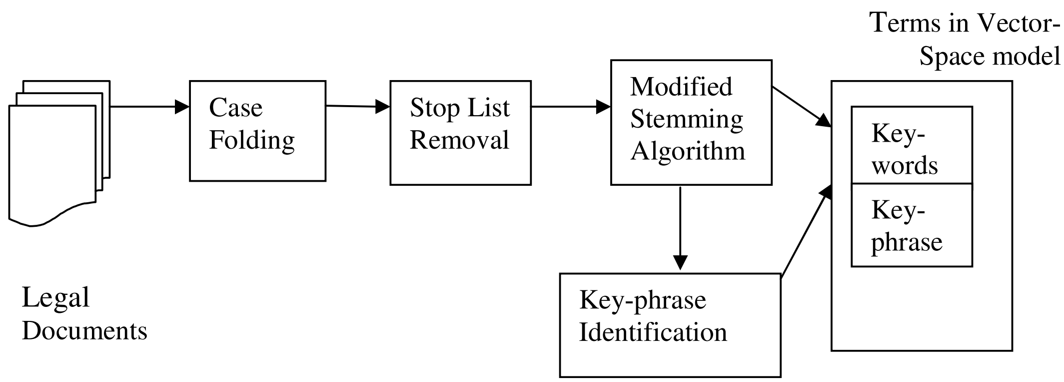


Figure 5.1 Pre-processing tools applied in our system

5.1.1 Case Folding

Case folding consists of converting all the characters of a document into the same case format, either the upper-case or the lower-case format. For instance, the words “act”, “Act”, “aCt”, “ACt”, “acT”, “AcT”, “aCT”, “ACT” will all be converted to the standard lower-case format “act”.

5.1.2 Removal of Stop Words

Stop words are the words occurring very frequently and not conveying independent meaning in a document. For instance, “the”, “would”, “can”, “do” are typical stop words. Prepositions, pronouns, articles, connectives etc. are also considered as stop words. Since they carry very little information about the contents of a document, it is usually a good idea to remove them from the document collections. The frequent occurrences of stop words in any set of documents, in general, imply redundancy. It follows that the stop words should not be included in any statistics and in scoring formulae, since they do not contribute to the relevance and importance of a sentence. Our system uses a list of 455 stop words obtained from the source code of the library BOW, developed by Carnegie Mellon University [127].

5.1.3 Stemming

The process of matching morphologically related terms, fusing or combining them in useful ways is called *conflation*. Conflation can be either manual or automatic. Programs for automatic conflation are called *stemmers* [128]. Stemmers are used in IR to reduce the size of index files. Since a single stem typically corresponds to many full terms, compression factors of over 50% can be achieved by storing stems instead of terms [128], especially in the case of affix-removal stemmers. Terms can be stemmed either at indexing time or at search time. The advantages of stemming at indexing time are efficiency and index file compression. The disadvantage is that information about the full terms will be lost.

Most stemmers today, do not always output root words. Taking note of the fact that linguistic correctness of the stems may become critical to effective retrieval in future, design of a root-word stemmer has been proposed and used in this study. This root-word stemmer is an improved version of the popular affix stemmer, developed by Porter in 1980 [125]. The rule base of Porter's Stemmer has been considerably enhanced so as to give meaningful stems as output in as many cases as possible.

We observe that the addition of more rules in order to increase the performance in one section of the vocabulary may cause degradation of performance elsewhere. Moreover, it is easy to give undue emphasis to cases, which appear to be important, but which turn out to be rather rare. After a detailed analysis, several new suffixes and the context in which they must be removed have been identified and appropriate changes have been made to the Porter's algorithm [125]. In this way, the results of the modified stemming algorithm are very useful in pre-processing stages of term distribution model.

5.1.4 Key-phrase Identification

The meaningful word output from the stemming module is considered for the identification of important phrases in the document space. Considering the occurrences of the word pairs by the relative frequency approach, the system identifies the key-phrases, which will increase the performance of the system. However, key-phrases are treated as single words in our probabilistic approach.

5.2 Proposed Approach to Text Summarization

Many researchers [130-132] have pointed out that the term repetition is a strong cohesion measure. Generally, the term weights are not directly based on any mathematical model of term distribution or relevancy [1]. In our earlier study, we made use of two theoretical models namely, the Poisson and the Negative Binomial distribution. We had discussed this implementation for the application of newspaper domain more thoroughly in our preceding work [11]. In the present work, we have used term distribution models such as Poisson mixtures for the distribution of terms and to characterize their importance in identifying key sentences in a legal document. The usage of term distribution model approach to text summarization will be discussed in section 5.5. As was stated in Chapter 2, our probabilistic approach to document summarization is different from the other related works discussed.

We have adopted a term distribution model which is used for deriving a probabilistically motivated term weighting scheme, assuming the vector-space model for single or multiple documents. This technique makes summarization more meaningful because the proportion of terms that are identified for summarization is closely related to the real content of the document. In our work, we have already explored a novel method of applying CRFs for segmentation of texts in legal domain. Now, we discuss the use of that knowledge for re-ranking of extracted sentences in the generation of a concise and coherent final summary. The summary generated by our summarizer was evaluated with the human generated headnotes which are available with all legal judgments. We find that the results are good, the details of which are discussed in chapter 6.

The major difficulty of headnotes generated by legal experts is that they are not structured, and as such lack the overall details of a document. To overcome this issue, we come out with the detailed structured summary of a legal document. Post processing is done to prepare the summary in a user friendly format. In order to get a readable final summary with consistency we have used rhetorical roles, identified using the CRF model, for grouping and re-ranking the sentences generated from the term distribution model.

5.3 Post Processing

The ranked sentences are available after the application of term distribution model on the judgments. We look at the proportion of the different rhetorical roles of these sentences and compared this with average proportions across all human annotated judgments. Our thesis is that if the distribution of different roles in the system-generated summary matches the average distribution then the quality of the summary would be closer to the human-generated summary. Based on the proportion of sentences observed, we choose more sentences from rhetorical roles that are under represented even if they are not in first 20% of ranked sentences. Likewise, sentences belonging to roles that are abundant are excluded from the final summary, even if they are highly ranked. The sentences finally selected are grouped according to the roles for maintaining the coherency in the final summary. We have understood from the legal experts that the summary generated in this process is more user-friendly.

5.4 Need for Probabilistic Model

The conventional Term Frequency-Inverse Document Frequency (TF-IDF) [30] term weighting approach used in many of the summarizers does not reveal the term characteristics in the related documents. The basic formulae used in term weighting are term frequency, document frequency and collection frequency, as given in Figure 5.2.

Note that $(\mathbf{df_i}) \leq (\mathbf{cf_i})$ and $\Sigma_j (\mathbf{tf_{ij}}) = (\mathbf{cf_i})$**.** The document frequency and collection frequency can be used only if there is a collection. The higher the term frequency (the more often the word occurs) the more likely it is that the word is a good description of the context of the document. A semantically focused word will appear several times in a document, if it occurs at all. Semantically unfocussed words are spread out homogenously over all documents. Another property of semantically focused words is that, if they come up once in a document, invariably they appear several times in the document.

Term Frequency **(tf_{ij})** -- Number of occurrences of words w_i in a document.
Document Frequency(df_i) -- Number of documents in the collection in which w_i occurs.
Sentence Frequency **(sf_i)** -- Number of sentences in which the word w_i occurs.
Collection Frequency (cf_i) -- Total Number of occurrences of w_i in a collection.

Figure 5.2 Basic formulae used for term weighting systems

Even though term weighting approach is a useful method for quantifying the basic information of term occurrence, it is not specific in assessing the likelihood of a certain number of occurrences of a particular word in a document space. Furthermore, it is an ad-hoc weighting approach which is not directly derivable from any mathematical model of term distribution or relevancy.

Another method of automatic extraction based on a user query, such as searching for words in document space, consists of matching the keywords in the query to the index words for all the documents in a given document space. This method is called lexical matching. However, this type of method can be inaccurate [133]. The fundamental inaccuracy of current information retrieval methods is due to the fact that the words in the query often are not the same as those by which the information that users seek are indexed. Hence, we have used probabilistic models as additional support to the conventional TF-IDF weighting method and lexical matching method for the distribution of terms. We use these models to characterize the significance of the terms in the process of legal document summarization.

5.5 Applying Probabilistic Models for Term Characterization

An alternative to the term-weighting method is the development of a model for the distribution of words which characterizes the importance of the words in the process of information retrieval. In particular, we wish to estimate $P_i(k)$ (Ref. Eq. 5.1), the proportion in which word w_i appears exactly k times in a document. In the simplest case, the term distribution model is used for deriving a probabilistically motivated term-weighting scheme, by assuming the vector space representation of terms in the documents. Most term distribution models try to characterize how informative is a word, which is also the information that the inverse document frequency in TF-IDF is trying to derive.

This work addresses the problem of extracting relevant word patterns in text, which is a problem of general interest for many practical applications. As one of the approaches in statistical language modeling, a term distribution approach based on linguistically motivated text characteristics as model parameters has been attempted. The derivation of models to describe word distribution in text is thus based on a linguistic interpretation of the process of text formation. It makes use of the probabilities of word occurrence being a function of linguistically motivated text characteristics. The focus of our study is to model the distribution of content words and phrases (word pairs) on a single document, to identify word occurrence patterns within sentences, and to estimate the corresponding probabilities.

In the next section, we discuss Negative binomial distribution which is a part of mixture of an infinite number of Poisson's [138]. Following our earlier work, we use the distribution which is closest to the negative binomial distribution known as *K-mixture model* in this work [139].

5.5.1 Negative Binomial Distribution (NBD)

Generally, occurrences of words (patterns) vary from genre to genre, author to author, topic to topic, document to document, section to section, and paragraph to paragraph. The proposed mixture of Poisson captures a fair amount of this heterogeneous structure by allowing the Poisson parameter t to vary over documents, subject to a density function $f(k)$. This function is intended to capture dependencies on hidden variables such as genre, role, section, etc. The Negative Binomial distribution is a well-known special case of this Poisson mixture. Poisson mixtures fit the data better than standard Poissons, producing more accurate estimates of the variance over documents [140]. The details of the discussion are available in our earlier work [11].

The negative binomial distribution is like standard Poisson, but the average number of occurrences t of the terms in the members of the document collection, is allowed to vary over documents. This is subject to a density function that models the dependence of t on all possible combinations of hidden variables such as genre, topic, etc. The computation of negative binomial involves large binomial coefficients and it is cumbersome to work with, in practice. Hence, we selected a simpler distribution that fits empirical word distributions as similar to the negative binomial distribution is known as Katz's K-mixture [139].

5.5.2 K-MIXTURE MODEL

The K-mixture model is a fairly good approximation model for term distributions compared to Poisson model [139]. It is described as the mixture of Poisson distribution and its terms can be arrived at by varying the Poisson parameters between observations. The formula used in K-mixture model for the calculations of the probability of the word w_i appearing k times in a document is given as:

$$P_i(k) = (1-r)\ \delta_{k,0} + \frac{r}{s+1}\ \frac{(s)^k}{(s+1)^k} \qquad \ldots\ldots\ldots (5.1)$$

where $\delta_{k,0} = 1$ if and only if $k = 0$, and $\delta_{k,0} = 0$ otherwise. The variables r and s are parameters that can be fit using the observed mean (t) and the observed Inverse document Frequency (IDF) as follows:

$$t = cf_i / N\ ;\ IDF = log_2 N / df_i\ ;\ s = t * 2^{IDF} - 1 = (cf_i - df_i) / df_i\ ;\ r = t / s \ \ldots\ (5.2)$$

where cf_i (collection frequency) refers to the total number of occurrences of terms in the collection, df_i (document frequency) refers to the number of documents in the collection in which the term occurs, and N is the number of documents in the collection. Document frequency (df_i) is closely related to the IDF. IDF is not usually considered as an indicator of variability, though it may have certain advantages over variance. The parameter r used in the formula refers to the absolute frequency of the term, and s used to calculate the number of "extra terms" per document in which the term occurs. The most frequently occurring words in all selected documents are removed by using the measure of IDF that is used to normalize the occurrence of words in the document. In this K-mixture model, each occurrence of a content word in a text decreases the probability of finding an additional term, but the decrease becomes consecutively smaller. The most frequently occurring words in all the selected documents are removed by using the measure of *IDF* that is used to normalize the occurrence of words in the document. This result shows the potential of the method to suggest effective index terms in a set of function words and a set of content words. Word occurrences that tend to cluster together in the same document are likely to be useful as index terms that can be used for summarization.

5.6 SENTENCE RANKING ALGORITHM

The algorithm used to extract the key sentences from the document space by applying the K-mixture model is given below.

Input: Words of the sentence collection from the pre-processing stage.

Output: A set of sentences extracted from a document, arranged in decreasing order of relevance.

Steps:

1. Input the words $\{w_{ij}\}$ for i = 1,2…m, j = 1,2…n, where m refers to the sentence number and n refers to the term number.
2. Compute tf_{ij}, df_i, cf_i, for all i =1,2…….m and j=1,2….n. /* Term frequency, Document frequency and Collection Frequency */
3. Based on the collection frequency (cf_i) and document frequency (df_i), calculate the observed mean and *IDF* of the sentences by using the formula given in equation 5.2.
4. Calculate r and s parameters of the K-mixture model using equation 5.2
5. Compute the probability $P_i(k)$ using equation 5.1
6. Normalize the terms by using the term characterization based on the parameter s.

7. Calculate the sentence weight by summing up the term probability values.
8. Rank the sentences based on the sentence weights.
9. Re-rank the sentences based on evolved roles during CRF implementation
10. Output the sentences in decreasing order of rank.

The distribution of context words among documents as well as the formula for the probabilities of occurrence of context words within documents will be derived using the notion of topicality and pertinent discourse properties. This assumption holds good for non-context words. On the other hand, the frequent occurrence of context words in the document creates a *term clustering* or *burstiness*. The tendency of content word occurrences to cluster is the main problem with the Poisson distribution for words. In the K-mixture model, each occurrence of a content word in a text decreases the probability of finding an additional term, but the amount of decrease becomes consecutively smaller. Moreover, the large number of occurrences of context words points to a central concept of the document. The algorithm given above illustrates that we need to solve complicated non linear equations in this model even for a single document summarization. Hence we adopted an intrinsic measure for evaluation of summary of a document that will be discussed in detail in Chapter 6. We are looking primarily for quality in the system-generated summary. The application of K-mixture model brings out a good extract of highly ranked sentences from the document space which can be used to generate a quality summary. Now we will discuss the method of improvement over the final summary.

5.7 Re-ranking of Final Summary

Tables 3.3 through 3.5 given in chapter 3 show the good performance of CRF model with efficient features sets for text segmentation task. These results can contribute to the generation of structured and efficient summary in the final stage. Use of the identified rhetorical categories can help in modifying the final ranking in such a way as to give more importance to *ratio decidendi* and *final decision*. The proportions of rhetorical roles identified in the ranked sentences in each document are compared with general distribution of rhetorical roles identified in the human annotated documents especially for *ratio decidendi* and *final decision*. One of the most important skills that the lawyers have to acquire is how to identify the ratio decidendi in a legal report for their general reading. Also it is more important for headnote generation. In this case, re-ranking has been performed in such a way as to maintain a good proportion for the above said roles in line. More of this will be discussed in Chapter 6. This will improve the presence of more relevant sentences in our final summary. Finally the extracted key sentences from the legal document using our probabilistic model are compared with headnotes generated by experts in the area. Figure 5.4 shows the results of our system generated summary in unstructured format, using the probabilistic model for important sentence extraction. The summary presented in Figure 5.5 shows the importance of arranging the sentences in a structured manner as it not only improves the readability and coherency but also gives more information like court's arguments to get a comprehensive view of the *ratio* and *disposal* of the case. There is a possibility of replacing some of the sentences in our system-generated summary with low ranked sentences for maintaining the proportion of rhetorical roles. *Identification of the case* may not be precisely identifiable from the corpus, but it is a problem even for human annotators with some of the documents. In our system, to overcome this difficulty, the ratio is rewritten in question format in such cases. The distribution of rhetorical roles given in Figure 3.9 of Chapter 3 demonstrates that 60% of sentences in a legal document belong to the role *History of the case*. The sentences belonging to this role discuss citation to other cases and also general discussions which may not be relevant for the summary. Hence we have not included the sentences belonging to the role *History of the case* in our summary.

Landlord is the revision petitioner. Evictions was sought for under sections 11 (2) (b) and 11 (3) of the Kerala buildings lease and rent control act, 1965. Evidence would indicate that petitioners mother has got several vacant shop buildings of her own. The appellate authority rejected the tenant's case on the view that tenant could not challenge the validity of the sale deed executed in favour of Mohan Lal because the tenant was not a party to it. We do not think this was a correct view to take. An allegation had been made that in reality there was no sale and the sale deed was a paper transaction. We find force in the contention of the counsel appearing for the tenant. The court had to record a finding on this point. This is a case where notice of eviction was sent by the mother of the petitioner which was replied by the tenant by Ext. B2 dated 26.1.1989. The landlady was convinced that she could not successively prosecute a petition for eviction and hence she gifted the tenanted premises to her son. On facts we are convinced that Ext. A1 gift deed is a sham document, as stated by the tenant, created only to evict the tenant. We are therefore of the view that the appellate authority has properly exercised the jurisdiction and found that there is no bonafide in the claim. We therefore confirm the order of the appellate authority and reject the revision petition. The revision petition is accordingly dismissed.

Figure 5.4. Unstructured summary produced by our system; the original judgment has 1250 words and summary is 20% of the source.

(Before K. S. Radhakrishnan & J. M. James, JJ)- Thursday, the 10th October 2002/ 18th Asvina, 1924 - CRP. No. 1675 of 1997(A) Petitioner : Joseph - Respondent: George K. - Court : Kerala High Court	
Rhetorical Status	**Relevant sentences**
Identifying the case	The appellate authority has properly exercised the jurisdiction and found that there is no bonafide in the claim – Is it correct?.
Establishing the facts of the case	We find force in the contention of the counsel appearing for the tenant. This is a case where notice of eviction was sent by the mother of the petitioner which was replied by the tenant by Ext. B2 dated 26.1.1989. The landlady was convinced that she could not successively prosecute a petition for eviction and hence she gifted the tenanted premises to her son.
Arguments	Apex court held as follows: "The appellate authority rejected the tenant's case on the view that tenant could not challenge the validity of the sale deed executed in favour of Mohan Lal because the tenant was not a party to it. We do not think this was a correct view to take. An allegation had been made that in reality there was no sale and the sale deed was a paper transaction. The court had to record a finding on this point. The appellate authority however did not permit counsel for the tenant to refer to evidence adduced on this aspect of the matter. The High Court also did not advert to it. We, therefore, allow this appeal set aside the decree for eviction and remit the case to the trial court to record a finding on the question whether the sale of the building to respondent Mohan Lal was a bonafide transaction upon the evidence on record".
Ratio of the decision	We are therefore of the view that the appellate authority has properly exercised the jurisdiction and found that there is no bonafide in the claim.
Final decision	We therefore confirm the order of the appellate authority and reject the revision petition. The revision petition is accordingly dismissed.

Figure 5.5 Structured summary for example judgment containing title, petitioner, respondent, important rhetorical categories and selected sentences.

5.8 Conclusion

The mathematical model based approach for extraction of key sentences has yielded better results compared to simple term weighting methods. To evaluate the importance of our summary, rather than using simple word frequency and accuracy, we employ an intrinsic measure to be discussed in the next chapter. We have attempted a novel method for generating a summary for legal judgments. We observe that rhetorical role identification from legal documents is one of the primary tasks to understand the structure of the judgments. With the identified roles, the important sentences generated in the probabilistic model will be reordered or suppressed in the final summary. The summary generated by our summarizer is closer to the human generated headnotes. It is hoped that the legal community will get a better insight without the need for reading a full judgment. Further, our system-generated summary may be more useful for lawyers to prepare a case history that has a greater bearing on their present case.

CHAPTER 6

RESULTS AND DISCUSSION

Any development in the realization of a natural language processing application requires systematic testing and evaluation. In the field of automatic summarization, most of the related publications address the problem of evaluation by first stating how hard the problem is and then by applying methods that the developers consider appropriate for the task. The complexity of legal domain makes the task more difficult. In the previous chapters, we discussed the development of our system in three different phases: CRF model for text segmentation, creation of new ontology for query enhancement, and finally use a term distribution model (K-mixture) for the extraction of relevant sentences from the legal document collection. In this chapter, the evaluation of the performance of the system summary and the other results are discussed.

The closeness of the system-generated summary to the human generated headnote (gold standard) is considered as one of the important measures of quality. For evaluation, we have adopted intrinsic methods which are concerned with the quality of summary, produced by considering two techniques. First, we compare the sentences extracted by our system with *reference* summary according to various measures viz., precision, recall, and F-measure. To construct the *reference* summary, a group of human subjects are asked to extract sentences. In our study, two highly experienced human subjects were involved in the task of extraction of sentences from judgments for summarization. Since there was a high degree of agreement between the two human subjects, we considered one arbitrarily chosen expert summary as the reference summary for our evaluation. Also we compared the performance of our system with that of 2 other summarizers and a standard baseline. A paired t-test statistical method [138] has been used to test the significance of the work.

More complex recall-based measures have been used in summarization research to measure how well an automatic system retains important content of original documents [10]. The simple sentence recall measure cannot differentiate system performance more appropriately, as is pointed out by Donaway et al. [141]. Therefore, in addition to pure sentence recall score, we use ROUGE [142] score as a second technique in this study. To make evaluation more comprehensive, we have considered three different methods of summarization for comparison. Out of the three, two of them make use of the publicly available automatic summarizers, and other is a standard baseline considered in many of the relevant studies [47].

6.1 Evaluation Methodology

The effectiveness of the system is evaluated in terms of the standard measures of the information retrieval tasks. A detailed discussion on the measures of evaluation is given in the next section. Our evaluation process has two goals:

1. *Compare the results of the proposed system with the human-generated headnote and also with 2 other summarizers and a baseline, thereby to show the efficiency and effectiveness.*

2. *As attaining maximum-recall is a desirable property of the summarizer, we employ automatic evaluation measure (ROUGE) to compare the closeness of the candidate summary with the human referenced summaries. The same measure will also be used to evaluate other summarizers considered in the study*

The evaluation of our system may be described in terms of the following tasks and methods.

6.1.1 Task

We collected the headnotes generated by human subjects based on their subjective judgments, which we call the *reference summaries.* Taking this as reference, we compared the outputs of our system and those of other automatic summarizers. We consider 20% summarization level for generating a summary as it is one of the levels most widely employed in summarization research [143-145].

The evaluation corpus has been constructed with legal judgments belonging to three different sub-domains, viz., *rent control*, *income tax,* and *sales tax*. The human subjects were given a set of documents for generating summaries. In this process they were neither informed of how their summaries would be used for later processing nor of any number of role breaks to be formed, and also they were not given any clues leading towards a choice. Moreover, the subjects were not constrained by time restrictions. The only demand given to them was to pick the relevant sentences at a compression rate of 20% from the given judgments. They were asked to pick complete sentences and not phrases or fragments.

6.1.2. Extraction Corpus

We used the legal judgments (court cases), available on the Internet from **www.kerelawyer.com**. We did not filter the documents based on the number of sentences or by any other specific means. The corpus consisted of a total of 200 documents grouped into 3 sub-domains with approximately 16000 sentences. It is part of a larger corpus of 1000 documents belonging to different sub-domains of *civil* judgments which we collected from the same source. The entire corpus consists of judgments dated up to the year 2006. We preprocessed the documents as described in Chapter 5. The documents are then segmented based on genre analysis as in Chapter 3. This contributed to the improvement in the consistency and readability of the final summary.

6.1.3 Evaluation method

The methods of evaluation of summarization systems can be broadly classified into two categories, namely, intrinsic and extrinsic methods [146]. Intrinsic methods measure a system's quality; extrinsic methods measure a system's performance for a particular task. We focus on the former technique, since we focused on the quality of the summary. Extrinsic methods have been used in task-based evaluation of programs like TIPSTER and SUMMAC [143]. Moreover in our study, we have applied extrinsic evaluation technique during the evaluation of ontology-based query processing results which are given in Chapter 4. Evaluating the quality of a summary has proven to be a difficult problem, principally because there is no obvious *reference* summary [145]. The use of multiple measures for system evaluation could help alleviate this problem. Most of the evaluations of summarization systems use one intrinsic method or the other. In the intrinsic method, the quality of summaries is determined based on direct human judgment of informativeness, coverage, fluency, etc., or by comparing with reference summary [16, 20, 144, 145]. The typical approach is to create an reference summary, either by professionals or by merging summaries provided by multiple human subjects using methods such as majority opinion, union, or intersection. The output of the system-generated summaries is then compared with the reference summary. In our study, we considered an arbitrarily chosen expert summary as the reference summary for our evaluation. The justifications are given in 6.3.1. The comparison between system and reference summaries is used to measure the quality in the case of extracts in terms of *sentence recall* and *sentence precision*. ROUGE [142] is another automatic performance measure used in our approach to evaluate the extraction-based summaries by comparing it with human reference summaries.

For automatic evaluation, we have compared the sentences produced by our system with: a standard baseline; the commercial automatic summarizers incorporated in Microsoft Word; and *MEAD* [17], a state-of-the-art summarization system available on the web.

6.2 Measures of Evaluation

We use two methods to evaluate the results. The first one is by *precision, recall* and *F-measure* which are widely used in information retrieval tasks [35,38]. For each document, the manually extracted sentences are considered as the reference summary denoted by S_{ref}. This approach compares the candidate (system-generated) summary (denoted as S_{sys}) with the reference summary and computes the precision, recall and F-measure values as shown in equation 6.1, which have been redefined in the context of text summarization along the same lines as given in [35].

$$P = \frac{|S_{ref} \cap S_{sys}|}{S_{sys}} \qquad R = \frac{|S_{ref} \cap S_{sys}|}{S_{ref}} \qquad F1 = \frac{2*P*R}{P + R} \qquad \ldots\ldots (6.1)$$

A second evaluation method is based on measuring maximum recall; we used the ROUGE toolkit, which is based on N-gram co-occurrences between candidate summary and reference human summaries [142]. This tool is adopted by Document Understanding Conferences (DUC) 2001 (http://duc.nist.gov) for automatic summarization evaluation ROUGE scores were found to have high correlation with human evaluations. In this study, we have applied ROUGE-N (N=1, 2) which is relatively simple, and seen to work well in most cases. The score of ROUGE-N is based on the number of n-grams occurring at the *reference summary* side. For example, ROUGE-2 computes the number of two successive words occurring between the candidate summary and reference summary.

Empirical studies of retrieval performance have shown a tendency for precision to decrease as recall increases. In most of the information retrieval tasks, recall curves tend to follow an increasing curve rising from the origin, and a trade-off between precision and recall is inherent [1]. The retrieval of relevant sentences present in the summary increases both precision and recall, while the presence of non-relevant sentences in the summary decreases precision but does not affect recall. Any good summarization system should have both high precision and recall measures. Moreover, higher ROUGE scores means the better performance of the system.

6.3 Comparative Performance Evaluation

For automatic evaluation, we have compared the final sentences generated by our algorithm which is based on the probabilistic approach to single-document summarization with: a baseline, the commercial automatic summaries produced by Microsoft Word and a state-of-the-art summarization system MEAD [17]. The salient features of a baseline referred to as system A and the publicly available summarizer (MEAD) referred as system B, are presented below. We also compared the summaries with those obtained with the third system (System C) which is part of Microsoft Word 2003.

Baseline

A baseline system is a simple reference system with which other systems can be compared. Traditionally for a newspaper domain, a baseline is the first few paragraphs of the text. But, for a legal domain a baseline is formed with the first few paragraphs and last few paragraphs of the text [47]. In our case, we defined a baseline (System A) as the one formed by compressing the source document by a factor of 20% as detailed below:

- Choose 10% of words of the beginning of the judgment. According to our rhetorical role identification methodology, it takes the sentences belonging to the roles *identifying the case* and *establishing the facts of the case.*
- Choose the last 6% of words of the judgment usually related to the roles *ratio of the decision* and *final decision*.

The baseline defined in this study is a standard baseline considered for information retrieval tasks, and it is also an appropriate length for summaries for long and short judgments [47]. Moreover, the sentences typically chosen by these approaches belong to roles that are considered to be the most important for the generation of a worthwhile summary. In the formation of baseline, if the last or first sentence is cut-off because of this limit the whole sentence is included. (*Note that in the earlier synopsis of the work, we considered two baselines, one focusing on beginning of the document and the other on the end. Since then we have adopted the current baseline since it more accurately reflects the structure of the legal judgments.)

System B: MEAD summarizer

The MEAD summarizer [17] was developed at the University of Michigan and at the Johns Hopkins University 2001 Summer Workshop on Automatic Summarization. It produces summaries of one or more source articles. In the initial versions of MEAD, a centroid-based approach is used for summarization via sentence extraction. For each cluster of related documents, a centroid was produced, which specifies key words and their respective frequencies in the set of source articles. Given the input documents and a compression rate, the algorithm chooses sentences with a high number of the key centroid words, since such sentences are considered as central to the cluster's topic.

MEAD is now publicly available as a toolkit for text summarization and evaluation [147]. The toolkit implements multiple summarization algorithm such as position-based, TF-IDF, largest common subsequence, and keywords. MEAD extractive summaries score sentences according to certain features of these sentences.

More recent versions of MEAD use a linear combination of three components: a feature extractor, a sentence scorer and a sentence re-ranker. MEAD first computes a value for user-defined features of each sentence using the feature extractor. The features used in MEAD include position, length (gives more weight to longer sentences) and centroids of clusters of related documents. The position score which assigns higher scores to sentences that are closer to the beginning of the document and lower ones to those further away from the beginning. Once the features are computed, the sentence scorer gives a value to each sentence based on a linear combination of their features. Sentences are then ordered according to their scores. The sentence re-ranker then adds sentences to the summary beginning with the highest scoring sentence. The re-ranker calculates the similarity of the sentence about to be added with all of the sentences already in the summary. If the similarity is above a given threshold, the sentence is not added to the summary and the re-ranker moves on to the next sentence.

Sentences are added to the summary until the amount of sentences in the summary corresponds to the compression rate.

6.3.1 Performance Comparison with Other Auto-summarizers

Tables 6.1 through 6.3 show the mean and standard deviation scores of recall, precision and F-measure of our system along with those of the other methods considered in the study. A higher score means better system performance. This data demonstrates a notable improvement in precision, recall, and F-measure of the proposed summarizer over the other methods. The plots of average measures of precision, recall, and F-measure of the proposed system and the different methods of summarizations mentioned above are shown in Figures 6.1 through 6.3. The graphs show that the proposed summarizer performs better than the other automatic summarizers according to recall, precision, and F-measures, and on all the three sub-domains of *rent control*, *income tax,* and *sales tax*. The result shown in Table 6.1 highlights the better performance of our summarizer on *rent control* domain compared to other methods considered in this study. Similar results occurred in the other two sub-domains like *income tax* and *sales tax* which are shown in Tables 6.2 and 6.3. We can see that the results of MEAD and WORD summaries are below 50% points which is comparatively low, while our summarizer is better in terms of all three evaluation measures.

In the *rent control* domain the F-measure scores have higher values compared to that in sales tax and income tax domains. This clearly indicates that it is harder to predict the basic structural formats in the sub-domains *sales tax* and *income tax* as compared to *rent control* sub-domain. In turn, it makes it difficult to extract the key sentences which are relevant to the cases. In the *sales tax* sub-domain, in particular, the *ratio of the decision*, a key role, may be present in more than one place. This may cause a serious lapse in retrieving relevant sentences needed for a summary, and in turn it affects the performance of the system. In our method, we have not given much importance to the positioning of sentences, and so the results of our summarizer comparatively may have been better than the other methods.

Table 6.1 Precision, Recall and F-measure scores for rent control domain

	Precision		**Recall**		**F-Measure**	
	Mean	Std.dev	Mean	Std. dev	Mean	Std. Dev
System A	0.411	0.14	0.462	0.15	0.420	0.13
System B	0.518	0.07	0.491	0.13	0.494	0.06
System C	0.294	0.06	0.347	0.10	0.309	0.05
Proposed System	0.645	0.08	0.685	0.18	0.654	0.11

Table 6.2 Precision, Recall and F-measure scores for Income tax domain

	Precision		**Recall**		**F-Measure**	
	Mean	Std. dev	Mean	Std. dev	Mean	Std. Dev
System A	0.428	0.19	0.274	0.12	0.349	0.14
System B	0.435	0.13	0.337	0.09	0.377	0.11
System C	0.366	0.11	0.294	0.11	0.323	0.11
Proposed System	0.680	0.09	0.649	0.16	0.657	0.11

Table 6.3 Precision, Recall and F-measure scores for Sales tax domain

	Precision		Recall		F-Measure	
	Mean	Std. dev	Mean	Std. dev	Mean	Std. dev
System A	0.395	0.14	0.281	0.10	0.330	0.11
System B	0.457	0.08	0.426	0.08	0.436	0.06
System C	0.361	0.12	0.325	0.06	0.338	0.08
Proposed System	0.650	0.11	0.600	0.15	0.621	0.13

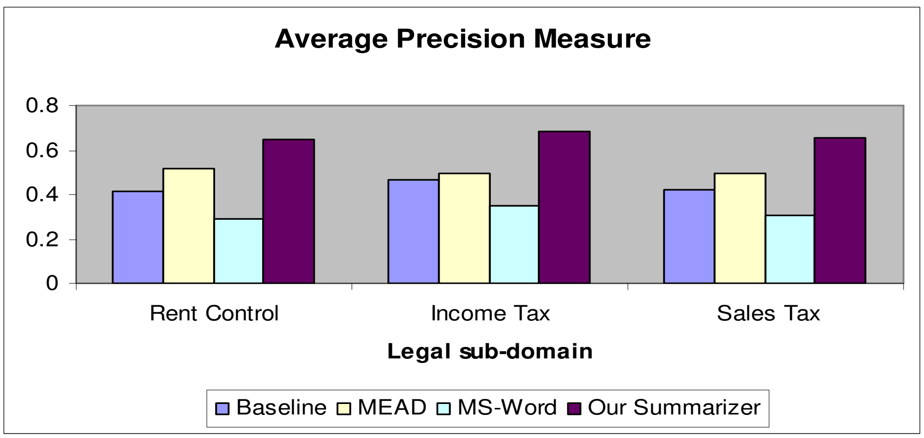


Figure 6.1 Average precision measure of different systems evaluated for the three different sub-domains.

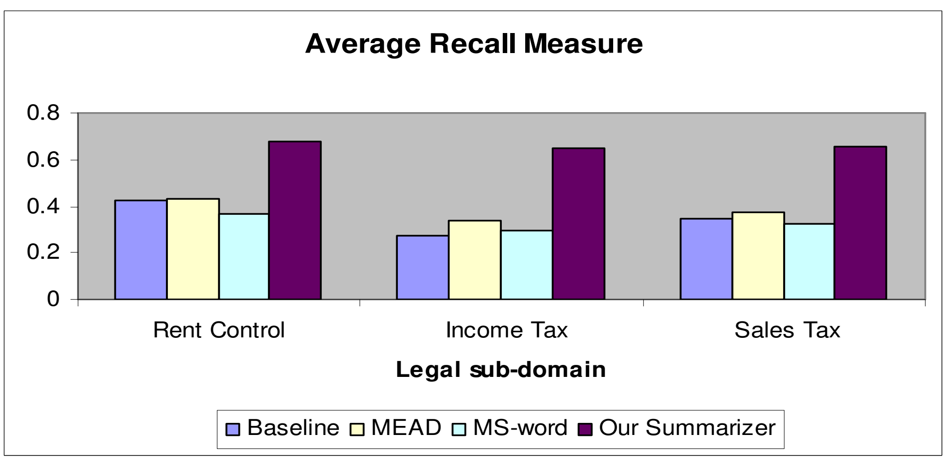


Figure 6.2 Average recall measure of different systems evaluated for the three different sub-domains.

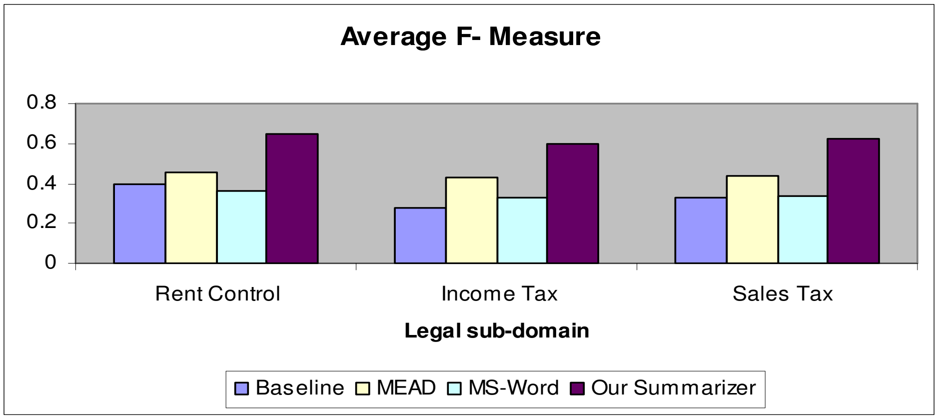


Figure 6.3 Average F-measure of different systems evaluated for the three different sub-domains.

Tables 6.1 through 6.3, and Figures 6.1 through 6.3 illustrate that the resultant summary of the proposed system is very similar to the summary generated by the human subjects. It is clear that the difference between the mean score of proposed system and that of the systems A, B, and C individually is more than 15% points in all sub-domains. This is true for all the three measures reported here. The average scores of precision, recall, and F-measure can be used to compare the performances of the summarizer on a common corpus, but they do not indicate whether the improvement of one summarizer performance over another is statistically significant or not. To substantiate the significance in the measurement, a paired t-test was applied to the data. More detailed analysis will be presented in the next section.

6.3.2 ROUGE: An Automatic Evaluation of Summaries

Traditional evaluation of summarization involves human judgments at different quality metrics. For example [10]:

- Quality evaluation, which involves subjective grading of summary quality with in itself, or comparison against the reference summary
- Informativeness evaluation, which involves comparison of the generated summary against a reference summary
- Fidelity of generated summary to source, or reading comprehension, which compares human's comprehension based on the summary with comprehension based on the source.

However, even simple manual evaluation of summaries on a large scale over selected questions and content requires a lot of human effort. It is expensive and difficult to conduct such evaluations on a frequent basis. As such, how to evaluate summaries automatically has drawn a lot of attention in the summarization research community in recent years. We look at one such automatic evaluation scheme known as ROUGE.

ROUGE, which stands for Recall-Oriented Understudy for Gisting Evaluation, is a package for automatic evaluation of summaries [142]. Following the successful application of automatic evaluation methods, such as BLEU [148] in machine translation and the system used by Saggion et al [149] to measure the similarities between summaries, Lin et al [142] showed that methods similar to earlier methods, but with a more refined approach could be applied to evaluate summaries. It includes measures to automatically determine the quality of a summary by comparing it to other (*reference*) summaries created by humans. The measures count the number of overlapping units such as N-grams, word sequences, and word pairs between the generated summary and the reference summary. ROUGE produces more reliable results if more than one reference summary is used. ROUGE-N is an N-gram recall between a candidate summary and a set of reference summaries and is computed as follows:

$$\text{ROUGE-N} = \frac{\sum\limits_{S \in \{\text{Reference Summaries}\}} \sum\limits_{gram_N \in S} Count_{match}(gram_N)}{\sum\limits_{S \in \{\text{Reference Summaries}\}} \sum\limits_{gram_N \in S} Count(gram_N)} \quad \text{........ (6.2)}$$

where N stands for the length of the N-grams (i.e., $gram_N$), and $Count_{match}(gram_N)$ is the maximum number of N-grams co-occurring in a candidate summary and the set of reference summaries, and $Count(gram_N)$ is the number of N-grams in the candidate summary. It is clear that ROUGE-N is a recall-related measure because the denominator of the equation is the total sum of the number of N-grams occurring in the reference summary side. Note that the number of n-grams in the denominator of the ROUGE-N formula increases as we add more reference summaries. From the earlier results in [150], we found that unigram and bi-gram co-occurrence statistics are good automatic scoring metrics. Longer N-grams tend to score for grammatically rather than content. In general, extraction-based summaries do not really suffer from grammar problems. Hence we have used ROUGE- (1 & 2) measures for automatic evaluation of extraction-based summaries.

We specifically evaluate various system-generated summaries with reference summaries generated by two different annotators. In the reference summaries, annotators capture the different points of law in a case. In this consideration, we note that ROUGE gives high scores for legal judgments with suitable extracts and low scores for those with unsuitable extracts.

Table 6.4 ROUGE scores for rent control domain

	ROUGE-1	ROUGE-2
System A	0.438	0.250
System B	0.482	0.263
System C	0.330	0.201
Our system	0.605	0.386

Table 6.5 ROUGE scores for income tax domain

	ROUGE-1	ROUGE-2
System A	0.275	0.181
System B	0.337	0.226
System C	0.294	0.179
Our system	0.598	0.354

Table 6.6 ROUGE scores for sales tax domain

	ROUGE-1	ROUGE-2
System A	0.302	0.176
System B	0.419	0.218
System C	0.320	0.198
Our system	0.586	0.334

Recall-based evaluation in the above calculations measures the number of reference summary sentences contained in the system-generated summary. Tables 6.4 through 6.6 show the result of this evaluation. Higher ROUGE score means higher performance of the system. Our summarizer evaluation scores comparatively better than other methods. As with the precision and recall study, the significance of the measurement is verified through a paired t-test. The details are given in section 6.4.1.

We compared the automatic evaluation measure (ROUGE) score of our summarizer with those of the publicly available summarizers systems B and C and with baseline system A. System C performs statistical analysis and system B uses statistical and linguistic algorithms to generate a summary. ROUGE performance measures show the proportion of relevant sentences that are retrieved. The results are shown in the form of bar graphs in Fig 6.4.

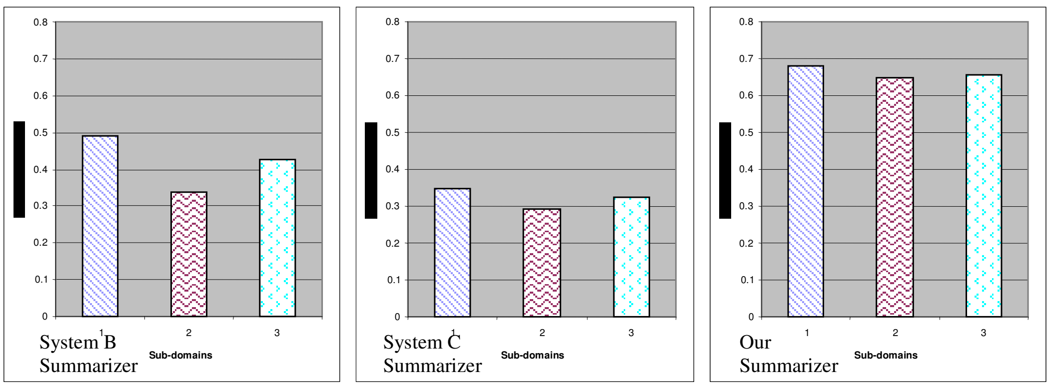


Figure 6.4 ROUGE performance measure for systems B and C, and our system for (1) Rent Control (2) Income Tax and (3) Sales tax sub-domains

We observed that our system performance is not only better but also is the same for the three different sub-domains. This is not the case with the other two systems considered. This shows that there is uniform recall in our system irrespective of the different sub-domains. The system-generated summary and the summaries generated by the other automatic summarizers are given in Appendix A, for a sample source document taken from the Kerala Lawyer archive. The document summary produced by our system, presented in the form of table-style summary would be useful for the legal community.

6.3.3 Agreement among the human subjects

We used two different annotators to create summaries and tested the statistical significance of the agreement among the human subjects. Since, we have followed intrinsic measure of evaluation as an evaluation procedure we need to establish the performance of annotation done by two different annotators, with the help of Kappa Coefficient [107]. The advantage of Kappa coefficient is that it factors out random agreement among the human subjects. We have already given a brief overview of Kappa coefficient in Chapter 3, Section 3.6.1. Our experimental results show that humans extracted summaries with a reproducibility of $K = 0.86$ ($N=16000$; $k=2$, where K stands for the Kappa coefficient, N for the number of sentences annotated and k for the number of annotators). In our study, both the annotators are agreeing on sentences retrieved from judgments in many cases. They differ only in a few cases mainly due to the inclusion of other case histories into the present case causing confusion to the annotators in giving preferences in selecting the sentences. Hence we used one arbitrarily chosen annotator summary as the gold standard for some of our evaluation experiments.

6.4 Statistical Analysis and Results

The analysis of the results based on the intrinsic evaluation method discussed in section 6.3.1, provides a clear picture that the proposed summarizer is better than the other methods considered in this study. In this section, statistical significance tests are discussed which were used to test the significance of the performance measures of the proposed system. This has been done by framing and testing the hypothesis for the above-mentioned objectives at a confidence level of 95% or 99%. A report of a paired t-test is given in Appendix C.

6.4.1 Results and Discussion

In this section, we discuss the statistical significance of the performance measures viz., precision, recall and F-measure of our system compared with the other three methods. A paired t-test was applied to find the significance of the mean score differences of the performance measures, between the proposed system and other automatic summarizers. Tables 6.7 through 6.9 show the calculated t-values with the significance levels, indicating that the average precision, recall, and F-measure performance measures of our system are significantly higher than those of the systems A through C considered at 99% confidence level (Table I of Appendix C). From Tables 6.1 through 6.3, it can be seen that that our system has higher average performance measures than that of other systems.

Table 6.7 Paired t-test values for Precision of our summarizer compared with 3 other systems for three different sub-domains.

Percentage Level for **Precision** Measure	System A		System B		System C	
	t-value	Sig. Level	t-value	Sig. Level	t-value	Sig. Level
Rent Control	11.24	p< .01	7.37	p< .01	10.64	p< .01
Income Tax	9.25	p< .01	6.76	p< .01	9..89	p< .01
Sales Tax	27.18	p< .01	12.71	p< .01	22.11	p< .01

Table 6.8 Paired t-test values for Recall of our summarizer compared with 3 other systems for three different sub-domains.

Percentage Level for **Recall** Measure	System A		System B		System C	
	t-value	Sig. Level	t-value	Sig. Level	t-value	Sig. Level
Rent Control	9.28	p< .01	14.52	p< .01	13.39	p< .01
Income Tax	12.00	p <.01	13.16	p< .01	13.94	p< .01
Sales Tax	17.11	p< .01	14.16	p< .01	16.63	p< .01

Table 6.9 Paired t-test values for F-measures of our summarizer compared with 3 other systems for three different sub-domains.

Percentage Level for **F-measure**	System A		System B		System C	
	t-value	Sig. Level	t-value	Sig. Level	t-value	Sig. Level
Rent Control	11.09	P< .01	13.70	p< .01	13.23	p< .01
Income Tax	10.99	P< .01	7.92	p< .01	10.00	p< .01
Sales Tax	13.75	p< .01	13.18	p< .01	14.36	p< .01

[p < .01 – significant at 99% confidence]

Tables 6.10 and 6.11 show the calculated t-values with the significance level, indicating that the average ROUGE performance measures of our system are significantly higher than those of the other systems A through C considered at 95% to 99% confidence level (Table I of Appendix C). It can be seen that our system has higher ROUGE performance measures than that of most other systems.

Compared to other domains, rent control domain judgments are more structured and most of the important details are present in the beginning and ending of the judgments. Hence, other systems also perform better as indicated by their better recall score.

Table 6.10 Paired t-test values for ROUGE-1 of our summarizer compared with three other systems for three different sub-domains.

ROUGE-1 Measure	System A		System B		System C	
	t-value	Sig. Level	t-value	Sig. Level	t-value	Sig. Level
Rent Control	5.52	P< .01	4.29	p< .01	10.34	p< .01
Income Tax	12.50	P< .01	11.01	p< .01	12.12	p< .01
Sales Tax	12.20	P< .01	7.60	p< .01	12.75	p< .01

Table 6.11 Paired t-test values for ROUGE-2 of our summarizer compared with three other systems for three different sub-domains.

ROUGE-2 Measure	System A		System B		System C	
	t-value	Sig. Level	t-value	Sig. Level	t-value	Sig. Level
Rent Control	4.49	p< .01.	4.29	p< .01	6.95	p< .01
Income Tax	6.70	p< .01	5.40	p< .01	6.98	p< .01
Sales Tax	6.78	p< .01	5.28	p< .01	6.52	p< .01

[*p < .01 – significant at 99% confidence level*]

Therefore, we can conclude that *our summarizer significantly outperforms the other summarization methods for the different evaluation techniques considered in this study.*

6.4.2 Distribution of Rhetorical categories in a human-generated summary

Thus far, we have evaluated our summarizer results with different performance measures. Now, we compare the similarity in the rhetorical structures of human-generated and our system-generated summaries. Figure 6.5 shows the resulting category distribution amongst these 1125 sentences in the chosen human-generated summary, which are far more evenly distributed than the one covering all judgment sentences (Figure 3.10 of Chapter 3). *Ratio of the decision* (label 6) and *Final decision* (label 7) are the two most frequent categories in the sentences extracted from judgments. We see less number of sentences extracted from *History of the case* (label 4). It clearly illustrates the importance of the types of roles considered in the summary in making the final presentation more user-friendly. We already mentioned that *Ratio of the decision* is the general legal principle justifying the judge's decision and, in non-legal terms, we might describe those sentences as the central sentences of the text. From Figure 6.5, we clearly see that ratio contributes more to the final summary. The labels represented in Figure 6.5 denote the rhetorical roles which are defined in Table 3.4 of chapter 3. We can see more closer results in our system-generated summary for the important rhetorical roles like *Ratio of the decision* (label 6) and *Final decision* (label 7), which is shown in Figure 6.6. Based on the distribution of identified roles, we have modified our system-generated summary with a few lower ranked sentences. These minimal modifications improved the readability of our system-generated summary which is shown in Appendix A.

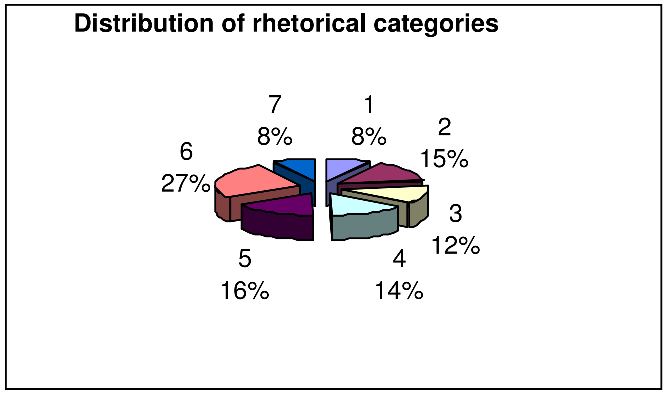


Figure 6.5 Distribution of rhetorical categories (summaries related to rent control domain) – human-generated.

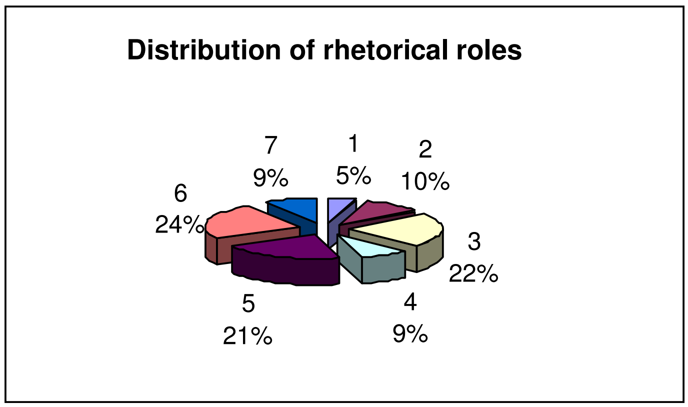


Figure 6.6 Distribution of rhetorical categories (summaries related to rent control domain) – system-generated.

6.5 Summary

Human-quality text summarizers are difficult to design, and even more difficult to evaluate. The results of different research projects are not so easy to compare because the reported results often do not discuss the characteristics of the corpora. In this chapter, we analyzed the results of the experiments carried out to evaluate our text summarization algorithm. We have shown that the proposed summarizer outperforms the three other methods when all of them were compared with the same human reference summaries. We have also demonstrated that our system matches 65% with the reference summary in terms of quality measures viz. precision, recall, and F-measure. Thus, it is shown that the use of term distribution model improves the summarizer performance irrespective of documents selected from different sub-domains. This also highlights the usefulness of statistical NLP in information retrieval applications. ROUGE score of our system summary is relatively better than the other systems considered in this study for all the three sub-domains. Finally, the evaluation results presented here show that the application of our term distribution model for the summarization of legal judgments is a promising approach.

CHAPTER 7

SUMMARY, CONTRIBUTIONS AND FUTURE RESEARCH

7.1 Summary

In this thesis, our ultimate goal was to develop an end-to-end legal judgment retrieval system that will facilitate access to and comprehension of relevant judgments. To this end, first we looked at headnote generation of a legal judgment which is the most important task in facilitating easy understanding of a case and the preparation of its history. In this process, the identification of the structure of a legal document is a crucial problem. Accordingly, in this work, we have presented an annotation scheme for the rhetorical structure of the legal judgments, assigning a label indicating the rhetorical status of each sentence in a portion of a document. Our annotation model has been framed with domain knowledge and is based on the genre analysis of legal documents. Also, we highlighted the construction of proper features sets with an efficient use of CRF for segmentation of a legal document. This is the first-time application of CRF for document segmentation. The identified roles in this process have been shown to aid in the presentation of extracted key sentences at the time of final summary generation. While the system presented here shows improvement over the existing systems, there is still much that remains to be explored.

Second, in the development of a legal ontology, several challenges with creation, maintenance, information retrieval, and the generation of key information related to legal judgments have been addressed. In our work, we designed a novel structural framework which has guided the development of the legal knowledge base. In the development of the framework, we held discussions with many leading legal experts on generalizing the concepts related to the three sub-domains of legal domain. The development of an ontology for intelligent information retrieval system has been analyzed, and a prototype implementation of the system has been described. The legal users are provided with the relevant documents based on their query on ontological terms instead of relying only on a simple keyword search among the queries. This approach has been shown to improve the quality of the results. User queries are enhanced with the rich collection of word features available in the knowledge base to retrieve relevant judgments from the document collection. The integration of several features of a term in the developed knowledge base used in our method has been shown to result in an excellent improvement of query results compared to that obtained by the baseline method. The legal ontology which we have proposed plays a decisive role in our summarizer in returning the more relevant judgments needed for the legal users. All of these have resulted in a knowledge-based system useful to the legal communities in their information retrieval.

Third and finally, the mathematical model based approach for the extraction of key sentences has yielded better results compared to the simple term weighting methods. The K-mixture model that is computationally simpler and closer to negative binomial distribution has been used as the final model in our approach for single document summarization. This is a special case of our earlier work on multi-document summarization carried out for texts related to newspaper domain. With the addition of identified roles in legal texts, the important sentences generated using the probabilistic model will be reordered or suppressed in the generation of the final summary. That is, we have significantly improved the extraction-based summarization results by modifying the ranking of sentences with the additional information on specific rhetorical roles. The summary generated by our summarizer is closer to the human generated headnotes. Thus the legal community can get a better insight without reading a full judgment, and also the system-generated summary is useful in the preparation of a case history that has a greater bearing on their present case.

The summarizers producing human-quality texts are difficult to design, and even so more difficult to evaluate. The results of different research projects are also difficult to compare because the reported results often do not discuss the characteristics of the corpora. Hence, all the results generated in this work have been confirmed for near-agreement with those generated by human subjects, and finally evaluated for statistical significance. The mathematical model based approach for extraction of key sentences has yielded better results compared to those by simple term weighting methods. We have shown that the proposed summarizer outperforms the four other methods considered, while benchmarking all of them against the human generated summaries. Further, it is also observed that irrespective of the sub-domains taken in this study, the summaries generated by the system proposed in this work are uniformly about 60% similar to the desired summaries. We have also demonstrated that our system nearly matches the reference summary in terms of quality measures like recall. We also find from the ROUGE measure that the system-generated summary is closer to the reference summaries. This also highlights the usefulness of statistical NLP in information retrieval applications. We note that the presentation of role-specific table-style summary without redundancy is an additional feature of our summarizer. Thus, a better insight to judgment details is provided to the legal community. Finally, the evaluation results confirm the usefulness of developing a summarization system, given the difficulties of generating a general summary without consideration of user profile and the domain.

7.2 Contributions and Future Research

The goal of our research is to develop models and algorithms for retrieving information from legal document repositories. Improving the quality of summaries generated from legal documents is a worthwhile consideration in the light of information overload. The main contributions of this dissertation include:

- A novel method of applying CRFs for segmentation in the process of structuring a given legal document. The rhetorical role identification from legal documents is one of the primary tasks in understanding the structure of the judgments. CRF model performs much better than the rule based and other rule learning methods in segmenting the text for legal domains. This is the first application of CRFs for document segmentation.
- A potentially powerful and novel structural framework has been proposed for the construction of legal ontology. The top level components considered in our work are *person, things, events, facts* and *acts.* Some of the main features of the ontology include the notion of query enhancement and case histories for legal concepts that change their identity and category through processes; extensive concept hierarchy understanding; notion of understanding the terms which are relevant to the main concepts. Thus, a knowledge engineering approach has been attempted.

- Detailed study of three different sub-domains like *rent control, income tax,* and *sales tax* of legal domain for developing a common knowledge base.
- Implementation of software environment for an intelligent information retrieval system for legal users: our system demonstrates how the developed knowledge base can be used to improve query enhancement system results by using inference rules that employ knowledge contained in the ontology.
- The segmentation of a document based on genre analysis is an added advantage and this could be used for improving the results during the extraction of key sentences by applying term distribution model.
- The output of table style system-generated summary (headnote) is more useful for lawyers to prepare the case history that have bearing to the present case.

There are many possible extensions of this work that can be undertaken as future research projects. Some of them are listed below:

- Extending our method for accessing web-based legal texts to generate a summary online.
- Extensions that incorporate more features to understand the rhetorical roles *Arguments* and *Arguing the case* in a more accurate manner.
- Automate the construction of the ontology by facilitating the addition of new concepts by extracting the relevant terms and their relationships to the existing concepts. This shall be done with a big team of experts and engineers by exploring all the sub-domains related to the legal cases.

- A more detailed study of the inter-relationship of concepts and relations in the proposed framework for the addition of new sub-domains leading to automatically adding new sub-domains to the ontology.
- Development of efficient query processing methods based on this ontology. Application of standard NLP techniques could help in the processing of queries automatically which can be used later for comparison with concepts in legal knowledge base.
- The development of a full-fledged automatic summarization system that can outperform the existing statistical-based systems may be the final goal of any summarization research. We believe that the work presented here is a step in the direction of that goal.

APPENDIX A

SAMPLE SUMMARY

List of sample input file selected from www.keralawyer.com: 01 KLC 292

20% single document summaries generated by different systems

First revision petitioner is a partnership firm and other are its partners.The firm is doing business on commission basis in the sale and export of spices. We have perused the terms of Ext. A7 agreement. We may extract the clause on which reliance was placed by the counsel for the tenant as stated herein. That the Tenant agrees the new tenancy agreement aforesaid shall be for a period of three years from 1-1- 1983. Further it is agreed in case the Tenant wants to continue, they can continue on condition that they give an increase of 10%in the monthly rental amount every three years. Counsel submitted the landlord cannot seek eviction on any of the grounds in the Rent Control Act in view of the above-mentioned clause. As held by the Apex court in Nai Bahu v. Lala Ramnarayan, AIR 1978 SC 22 the provisions in the Rent control Act would prevail over the general law of the landlord and tenant. We are of the view that a tenant or landlord cannot contract out of the provisions in the Rent Control Act if the building lies within the purview of the Rent Control Act. The decision in Laxmidas Babaudas's case cited by the counsel for the revision petitioners, in our view, is not applicable to the facts of this case. In the instant case there is no fixed term lease, but the lease deed as only given an option to the tenant to continue on condition on an increase of 10% in the monthly rental amount every three years. We are of the view that clause, as such do not take away the statutory right of the landlord under the Rent control Act. We are of the view landlord has made out sufficient grounds in the petition under Section 11 (3) of the Act. We are not prepared to say such a claim lacks bona fide. We are of the view, in the facts and circumstances of the case, landlord has established the bona fide need for own occupation under Section 11 (3) as well as uAnder section 11(8). We find no reason to disturb the said finding. Revision lacks merits and the same is dismissed in limine.

Example of 20% unstructured summary produced by our system

(Before K. S. Radhakrishnan & K.A. Mohammed Shafi JJ)- Thursday, the 31st January 2002/11th Magha 1923 Petitioner : M/s Allid Traders & others - Respondent: The Cochin Oil Merchants Association Court : Kerala High Court	
Rhetorical Status	**Relevant sentences**
Identifying the case	Landlord has established the bona fide need for own occupation under section 11(3) as well as under section 11(8) – Is it correct?.
Establishing the facts of the case	In the instant case there is no fixed term lease, but the lease deed has only given an option to the tenant to continue on condition on an increase of 10% in the monthly rental amount every three year. We are not prepared to say such a claim lacks bonafide.
Arguing the case	That the tenant agrees the new tenancy agreement aforesaid shall be for a period of three years from 1-1-1983. Further it is agreed in case the tenant wants to continue, they can continue on condition that they give an increase of 10% in the monthly rental amount every three years.
Arguments	As held by the Apex court in Nai Bahu v. Lala Ramnarayan. AIR 1978 SC 22 the provisions in the Rent control Act would prevail over the general law of the landlord and tenant. The decision in Laxmidas Babaudas's case cited by the counsel for the revision petitioners, in our view is not applicable to the facts of this case.
Ratio of the decision	We are of the view that clause, as such do not take away the statutory right of the landlord under the Rent control Act. We are of the view landlord has made out sufficient grounds in the petition under Section 11 (3) of the Act. We are of the view in the facts and circumstances of the case, landlord has established the bona fide need for own occupation under section 11(3) as well as under section 11(8).
History of the case	First revision petitioner is a partnership firm and other are its partners. The firm is doing business on commission basis in the sale and export of spices. We have perused the terms of Ext. A7 agreement.
Final decision	We find no reason to disturb the said finding. Revision lacks merits and the same is dismissed in limine.

Example of 20% structured summary generated by our system after post-processing

First revision petitioner is a partnership firm and others are its partners. Respondent-landlord has now preferred the present rent control petition under sections 11 3 and 11 8 of the Kerala Buildings Lease and Rent Control Act. Counsel submitted term of the lease is liable to be extended every three years at the option of the tenant an such option has been exercised by the tenant continuously till date and even thereafter during the pendency of the present proceedings. Counsel submitted the landlord cannot seek eviction on any of the grounds in the Rent Control Act in view of the above-mentioned clause. Rent Control Act is a self-contained statute and the right s and liabilities of the landlord and tenant are to be governed by its provisions. AIR 1974 SC 1924 held that an agreement in the lease deed providing that the parties would never claim the benefit of the Act and that the provisions of the Act would not be applicable to the lease deed is illegal. An agreement between a tenant and a landlord by which the tenant undertook to remove the building occupied by him on the expiry of the lease period of 3 years was held not opposed o public policy or the protection given to a tenant under the Act. We are of the view that a tenant or landlord cannot contract out of the provisions in the Rent Control Act if the building lies within the purview of the Rent Control Act. But indefinite continuance of the tenant even after the landlord has satisfied the ingredients of Section 11 of the Act in our view would be defeating the object and purpose of rent control legislation.

Example of 20% summary produced by MEAD system

Revision petitioners were in occupation company. Earlier RCP. 177/81 was preferred by the landlord for eviction of the revision petitioners, which was later compromised. Respondent-landlord has now preferred the present rent control petition under sections 11 (3) and 11 (8) of the Kerala Buildings (Lease and Rent Control) Act. Rent Control Court dismissed the petition on both the grounds. Matter was taken up by the landlord before the Appellate Authority. Counsel submitted term of the lease is liable to be extended every three years at the option of the tenant an such option has been exercised by the tenant continuously till date and even thereafter during the pendency of the present proceedings. As held by the Apex court in Nai Bahu v. Lala Ramnarayan, AIR 1978 SC 22 the provisions in the Rent control Act would prevail over the general law of the landlord and tenant. Rent Control Act is a piece of social legislation and is meant mainly to protect the tenants form frivolous eviction. We are of the view that a tenant or landlord cannot contract out of the provisions in the Rent Control Act if the building lies within the purview of the Rent Control Act. It is true that they can lay down a contractual fixed terms of lease and during the pendency of the term of lease eviction cannot be ordered. The decision in Laxmidas Babaudas's case cited by the counsel for the revision petitioners, in our view, is not applicable to the facts of this case. We are of the view, in the facts and circumstances of the case, landlord has established the bona fide need for own occupation under Section 11 (3) as well as under section 11(8).

Example of 20% summary produced by Microsoft Word system

First revision petitioner is a partnership firm and others are its partners. The firm is doing business on commission basis in the sale and export of spices. Respondents herein is a company owning a double storied building in Jew Town, Mattanchery. They purchased the building in 1972. Revision petitioners were in occupation of the tenanted premises prior to that. A portion of the ground floor as well as the first floor of the building belong to and is in the possession of the respondent company. Earlier RCP. 177/81 was preferred by the landlord for eviction of the revision petitioners, which was later compromised. The claim for eviction was abandoned on the revision petitioners agreeing to pay enhanced rent at the rate of Rs. 500/ per month an don agreeing to surrender a part of the backyard of the building occupied by them in exchange for being given another portion of the backyard by the landlord. A fresh rent deed was executed between the parties on 22.12.1982 incorporating the terms of the fresh tenancy. We are of the view landlord has made out sufficient grounds in the petition under Section 11 (3) of the Act. Landlord is a non trading company constituted by Oil Merchants in the Cochin Area. It controls the trade. It is an occupant of the building adjacent to the petition schedule building. Landlord has been directed by the forward Markets Commission to improve the infrastructure facilities, failing which the recognition was liable to be withdrawn. In order to satisfy the infrastructural facilities insisted by the Forward Markets Commission landlord requires additional space. They are in need of trading hall. We are not prepared to say such a claim lacks bona fide. We are of the view, in the facts and circumstances of the case, landlord has established the bona fide need for own occupation under Section 11 (3) as well as under section 11(8).We find no reason to disturb the said finding. Revision lacks merits and the same is dismissed in limine.

Example of 20% summary produced by Baseline

The firm is doing business on commission basis in the sale and export of spices. Respondents herein is a company owning a double storied building in Jew Town, Mattanchery. Appellate Authority allowed the appeal on both the grounds. Aggrieved by the same this revision petition has been preferred y the tenants. We have perused the terms of Ext. A7 agreement. We may extract the clause on which reliance was placed by the counsel for the tenant as stated herein. "That the Tenant agrees the new tenancy agreement aforesaid shall be for a period of three years from 1-1- 1983. Further it is agreed in case the Tenant wants to continue, they can continue on condition that hey give an increase of 10%in the monthly rental amount every three years." As held by the Apex court in Nai Bahu v. Lala Ramnarayan, AIR 1978 SC 22 the provisions in the Rent control Act would prevail over the general law of the landlord and tenant. The Apex Court in Muralidhar Agarwal v. State of U.P. AIR 1974 SC 1924 held that an agreement in the lease deed providing that the parties would never claim the benefit of the Act and that the provisions of the Act would not be applicable to the lease deed is illegal. But indefinite continuance of the tenant even after the landlord has satisfied the ingredients of Section 11 of the Act, in our view, would be defeating the object and purpose of rent control legislation. In that case there was a contractual fixed term lease. In the instant case there is no fixed term lease, but the lease deed as only given an option to the tenant to continue on condition on an increase of 10% in the monthly rental amount every three years. We are of the view landlord has made out sufficient grounds in the petition under Section 11 (3) of the Act. Landlord has been directed by the forward Markets Commission to improve the infrastructure facilities, failing which the recognition was liable to be withdrawn. We are of the view, in the facts and circumstances of the case, landlord has established the bona fide need for own occupation under Section 11 (3) as well as under section 1 (8). We find no reason to disturb the said finding. Revision lacks merits and the same is dismissed in limine.

Example of 20% human referenced summary

APPENDIX B

Rent control sub-domain

Words/ Phrases	Basic Information	Semantic Information	Supplementary Information	
Building	Status (things) Property	Tangible Property Immovable Property	House, Farm house, Apartments, Godown, Shops, warehouse, hotels, lodge, mansion, hostel	Transfer of Property act Land acquisition act Urban ceiling act City tenants protection act Land reforms act Rent control act.
Landlord	Status (Person) Natural person	Person Petitioner/ Respondent	House Owner, Owner, Land Owner	Transfer of property act, City tenants protection act Land reforms act Rent control act.
Eviction	Process (Events)	Forcible act Legal or illegal	Throw out, Thrown out, Expulsion, Expel, Cleared off	Indian Penal code, Land acquisition act, city tenants protection act, land reforms act, rent control act
Own use	Facts	Self/ Family member	Bonafide need, self occupation, self use	Rent control act, Land reforms act, Transfer of Property act
Rent	Facts	Movable Property (Rent for land, building, ship)	Arrears, Abnormal, default, Lease	Rent control act, Land reforms act, Transfer of Property act
Demolition	Process (Events)	Willful act	Destruction, Pulling down, Knocking down	Rent control act, Indian penal code
Cut off/ withhold Amenities	Process	Willful act	Facilities, Comforts, Easements, Services	Rent control act, Indian penal code
Sublet	Process (Events)	Willful act	Sub lease, sub tenancy	Rent control act, Transfer of Property act
Recovery of possession	Process (Events)	Willful act	Taking back possession/custody	Rent control act, Transfer of Property act, Sale of goods act

Income tax sub-domain

Words/ Phrases	Basic Information	Semantic Information	Supplementary Information	
Assessee	Status	Natural Person Legal entity	Assessed	Income tax act, sales tax act, Customs act, Excise act, property tax act, Land tax act
State	Status	Sovereign entity	Government, Corporation, Punchayat, Assessor, income tax officer, commissioner, Commercial tax officer, Excise Commissioner	Income tax act, sales tax act, Customs act, Excise act, property tax act, Land tax act
Assessment	Process (event)	Value	Appraisal, estimation, evaluation, valuation	Income tax act, sales tax act, Customs act, Excise act, property tax act, Land tax act
Building	Status (thing)	Immovable property	House, Flat, Apartment, Shops	Income tax act, sales tax act, Customs act, Excise act, property tax act, Land tax act, Transfer of Property act
Capital gains	Facts	Tax		Income tax act
Depreciation	Facts	Expenditure Deductible Allowable Reject	Reduction, decline	Income tax act, Companies act
Exemption	Process/ Facts	Allow/disallow/ Claim/Decline/ Reject Willful act	Exclusion, Not included	Income tax act, sales tax act, Customs act, Excise act, property tax act, Land tax act, Transfer of Property act
Deduction	Process/ Facts	Allow/disallow/ Claim/Decline/ Reject Willful act		Income tax act, sales tax act, Customs act, Excise act
Refund	Process/ Facts	Allow/disallow/ Claim/Decline/ Reject Willful act		Income tax act, sales tax act, Customs act, Excise act, property tax act, Land tax act, Transfer of Property act
Penalty	Process/ Facts	Allow/disallow/ Claim/Decline/ Reject Willful act	Fine, Punishment	Income tax act, sales tax act, Customs act, Excise act, property tax act, Land tax act, Transfer of Property act
Perquisite	Fact	Tax	Perk, Benefits, Compensation, Amenities	Income tax act

Sales tax sub-domain

Words/ Phrases	Basic Information	Semantic Information	Supplementary Information	
Assessee	Status	Natural Person Legal entity	Assessed	Income tax act, sales tax act, Customs act, Excise act, property tax act, Land tax act
Trader	Status	Natural Person Legal Entity	Dealer, Buyer, Seller, Agent, Broker, Merchant	Income tax act, sales tax act, Customs act, Excise act
Assessment	Process (event)	Value	Appraisal, estimation, evaluation, valuation	Income tax act, sales tax act, Customs act, Excise act, property tax act, Land tax act
Goods	Status (thing)	Movable property	Supplies, Merchandise, Commodities, wares, Produce	Income tax act, sales tax act, Customs act, Excise act, property tax act, Sale of goods act, Contract act
Inter-state sale	Facts	Movable Property	Interstate, Cross-state border	Sales tax act
Value	Facts	Sale value	Price, cost, rate	Sales tax act, Customs act, Excise act, Transfer of Property act
Exemption	Process/ Facts	Allow/disallow/ Claim/Decline/ Reject Willful act	Exclusion, Not included	Income tax act, sales tax act, Customs act, Excise act, property tax act, Land tax act, Transfer of Property act
Deduction	Process/ Facts	Allow/disallow/ Claim/Decline/ Reject Willful act		Income tax act, sales tax act, Customs act, Excise act
Refund	Process/ Facts	Allow/disallow/ Claim/Decline/ Reject Willful act		Income tax act, sales tax act, Customs act, Excise act, property tax act, Land tax act, Transfer of Property act
Penalty	Process/ Facts	Allow/disallow/ Claim/Decline/ Reject Willful act	Fine, Punishment	Income tax act, sales tax act, Customs act, Excise act, property tax act, Land tax act, Transfer of Property act
Deemed sales/Second sales	Facts	Willful act	Considered	Sales tax act, Customs act, Excise act

APPENDIX C

C.1 Statistical Significance

In this study, a hypothesis has been framed for various measures to test the significance of the performance of the system results. That is, we hypothesize that there is no difference in the effectiveness of the two systems, i.e. the results of the two systems are equally effective. We call hypotheses like these as *null hypotheses* and denote them by H_o. The null hypothesis is used for any hypothesis set up primarily to see whether it can either be rejected or accepted. Accepting or rejecting the hypothesis will be based on the testing of differences in the mean scores through some *significance test*. To approach the problem of hypotheses-testing systematically, the following five steps are considered [138].

1. Formulate a null hypothesis. In case the null hypothesis is rejected, accept the appropriate alternative hypothesis.
2. Specify the level of significance.
3. Construct a criterion for testing the null hypothesis against the given alternative, and the criterion should be based on the sampling distribution of an appropriate statistic.
4. Calculate the statistical value, which will support the decision-making.
5. Decision should be made to reject or accept the null hypothesis, or to reserve judgment.

C.2 Paired t-test

Usually, the two-sample t-test shall be applied to all types of samples, which are independent in nature. We work with (signed) differences of the paired data, and test whether these differences may be looked upon as random samples from a population. We have to apply one-sample t-test to test data, which is referred to as the paired t-test.

A paired t-test is a statistical test, which has been applied to prove the significance of the difference between the mean scores. The procedure involves calculating a *difference score* for each measure. A test statistic called '*t*' is then calculated. This *t* score is a measure of how far apart the average difference score is from zero in standard units. The larger the *t* value, it is more likely that the difference score is not zero, and hence the difference between the means is reliable. Paired t-test [138] should be applied under the following conditions:

- For the purpose of comparing two means
- When the measurements are distributed normally
- When the data is measured on an interval or ratio scale

The equation (C.1) used to calculate the t-statistic is given below:

$$t = \frac{\overline{X} - \mu}{S / \sqrt{n}} \qquad \text{........ (C.1)}$$

where $\overline{X}$ and S are the mean and standard deviation of the differences of n samples respectively, and μ =0 (where μ is the mean of the population of differences sampled).

TABLE C.I TABLE OF CRITICAL VALUES OF '*t*'

Degrees of freedom (df)	Level of Significance				
	0.10	**0.05**	**0.025**	**0.01**	**0.005**
1	3.08	6.31	12.71	31.82	63.66
2	1.89	2.92	4.30	6.96	9.93
3	1.64	2.35	3.18	4.54	5.84
4	1.53	2.13	2.78	3.75	4.60
5	1.48	2.02	2.57	3.36	4.03
6	1.44	1.94	2.45	3.14	3.71
7	1.41	1.89	2.36	3.00	3.50
8	1.40	1.86	2.31	2.90	3.36
9	1.38	1.83	2.26	2.82	3.25
10	1.37	1.81	2.23	2.76	3.17
11	1.36	1.80	2.20	2.72	3.11
12	1.36	1.78	2.18	2.68	3.06
13	1.35	1.77	2.16	2.65	3.01
14	1.35	1.76	2.14	2.62	2.98
15	1.34	1.75	2.13	2.60	2.95
16	1.34	1.75	2.12	2.58	2.92
17	1.33	1.74	2.11	2.57	2.90
18	1.33	1.73	2.10	2.55	2.88
19	1.33	1.73	2.09	2.54	2.86
20	1.33	1.72	2.09	2.53	2.84
21	1.32	1.72	2.08	2.52	2.83
22	1.32	1.72	2.07	2.51	2.82
23	1.32	1.71	2.07	2.50	2.81
24	1.32	1.71	2.06	2.49	2.80
25	1.32	1.71	2.06	2.49	2.80
26	1.31	1.71	2.06	2.48	2.78
27	1.31	1.70	2.05	2.47	2.77
28	1.31	1.70	2.05	2.47	2.76
29	1.31	1.70	2.05	2.46	2.76
30	1.31	1.70	2.04	2.46	2.75
40	1.30	1.68	2.02	2.42	2.70
→ 60	1.30	**1.67**	2.00	**2.39**	**2.66**
120	1.29	1.66	1.98	2.36	2.62
→ **Infinity**	**1.28**	**1.64**	**1.96**	**2.33**	**2.58**

APPENDIX D

List of some of the cue phrases included in our CRF implementation for rent control domain:

1) IDENTIFYING THE CASE :

Question/s / point/s for consideration is /are
Question/s point/s that arise for consideration is /are
Question/s / point/s before us is /are
Question/s point/s that arise before us is /are

We do not find any reasons to interfere with/we do not find anything to interfere with/ we are not in agreement with / we do not agree with -

Order under challenge in that…/ this order is under challenge

2) ESTABLISHING THE FACTS OF THE CASE:

Relevant facts in this case

Facts of the/this case

In the fact of this evidence

On the basis of established/Not established/ failed to establish

/proved/disproved/not proved

Court/lower court/appellate court/ authority found that

We/ this court find/found that

It was found that

"We agree with court/lower court/appellate court/authority"
Supreme court/court says -

If .. so to be avoided.

We find that /found that … if it is presented by/ if that we / in our view

I/ We / this court . find/found/finds/ /do not find / does not find / did not find

3) ARGUING THE CASE:

According to Petitioner/Respondent/Appellant followed by within quotes citing court cases (Case-laws can be detected by the letter "v" or "Vs" or "vs") and "sections"/ "Sec" of "Act".

Petitioner/Respondent/Appellant filed affidavit/counter affidavit followed by strings as above

Petitioner/Respondent/Appellant contended/ argued followed by strings as above.

Indentation paragraphs should be added as a paragraph,

4) HISTORY OF THE CASE:

Petitioner filed against or Appeal/Petition is filed against
Before the trial court/Appellate court/authority allowed/dismissed

Remaining sentences from the document after processing 1,2,3,5,6,7 labels.

5) ARGUMENTS:

ABC v XYZ or ABC Vs XYZ or ABC V XYZ followed by any year starting from 1900 onwards (so anything between 1900 & 2100) …H.C./S.C./Privy Council Court/Lower Court/Appellate Court was of the view/ held We/this court find ….no merits/merits

6) RATIO:

No provision in .. Act/statute
If ….. maintainable (maintained)
We hold
Not valid/legally valid/ legally not valid
We agree with Court….in holding that
We are of the view that
Statute
In view
According
We are also of view
Holding

7) JUDGEMENT:

In the/ Under the circumstances OR consequently the petition/Appeal/ Review/Revision…allowed/dismissed/upheld/ order of the .court upheld.
Dismiss/dismissed/dismissing/sustained/rejected

References

[1] D.M. Christopher and H. Schutze, *Foundations of Statistical Natural Language Processing*, The MIT Press, London, England, 2001.

[2] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, *From Data Mining to Knowledge Discovery: An Overview*, in *Advances in Knowledge Discovery and Data Mining*, U.Fayyad, G.Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, eds., MIT Press, Cambridge, 1996.

[3] R. Feldman and I. Dagan, "Knowledge Discovery in Textual Databases (KDT)", *Proc. of the First International Conference on Knowledge Discovery and Data Mining (KDD-95)*, pp. 112-117, Montreal, Canada, AAAI Press, August, 1995.

[4] T. Mitchell. *Machine learning*, McGraw Hill, USA, 1997.

[5] M. Moens. "An Evaluation Forum for Legal Information Retrieval Systems?", *Proc. of the ICAIL-2003 Workshop on Evaluation of Legal Reasoning and Problem-Solving Systems,* International Organization for Artificial Intelligence and Law, pp. 18-24, 2003.

[6] J. Burger, *Issues, Tasks and Program Structures to Roadmap Research in Question & Answering (Q&A)*, NIST Roadmap Document, 2001.

[7] T.R. Gruber, "Towards Principles for the Design of Ontologies used for Knowledge Sharing", *International Journal of Human-Computer studies,* vol.43, pp. 907-928, 1995.

[8] D.M. Jones, T.J.M. Bench-Capon, and P.R.S. Visser, "Methodologies for Ontology Development", In *Cuena, Journal, (ed) Proc. of XV IFIP World Computer Congress, IT and Knows*, pp.62-75, 1998.

[9] A. Valente and J. Breuker, "Making Ends Meet: Conceptual Models and Ontologies in Legal Problem Solving", *Proc. of the XI Brazilian AI Symposium (SBIA'94)*, pp.1-15, 1994.

[10] I. Mani and M. Maybury, *Advances in Automatic Text Summarization*, Cambridge MA: MIT Press, 1999.

[11] M. Saravanan, T*erm Distribution Model for Automatic Text Summarization of Multiple Documents*, Master Thesis, Department of Computer Science and Engineering, Indian Institute of Technology Madras, India, 2002.

[12] J. Goldstein, M. Kantrowitz, V. Mittal, and J. Carbonell, "Summarizing Text Documents: Sentence Selection and Evaluation Metrics", *Proc. of SIGIR*, pp. 121–128, 1999.

[13] J. Goldstein, *Automatic Text Summarization of Multiple documents*, Ph.D. Thesis proposal, Carnegie Mellon University, Pittsburgh, PA 15213, 1999.

[14] J. Salton and C. Buckley, "Term Weighting Approaches in Automatic Text Retrieval", *Information Processing and Management,* vol. 24, no.5, pp.513-323, 1988.

[15] D. Marcu, "From Discourse Structures to Text Summaries", *Proc. of the ACL 97/EACL-97 Workshop on intelligent scalable Text Summarization*, pp.82-88, Madrid, Spain, 1997.

[16] R. Barzilay and M. Elhadad, "Using Lexical Chains for Text Summarization", *Proc. of the ACL Workshop on Intelligent Scalable Text summarization*, pp. 10-17, Madrid, Spain, 1997.

[17] D.R. Radev, H. Jing, and M. Budzikowska, "Centroid-based Summarization of Multiple Documents: Sentence Extraction, Utility-based Evaluation, and User Studies", *Proc. of ANLP-NAACL Workshop on Summarization*, pp. 21-30, Seattle, Washington, April, 2000.

[18] C.Y. Lin and E. H. Hovy, "The Automated Acquisition of Topic signatures for Text Summarization", *Proc. of the Computational Linguistics Conference,* pp. 495-501, Strasbourg, France, August, 2000

[19] H.P. Edmundson, "New Methods in Automatic Abstracting", *Journal of the ACM* vol.16, no. 2, pp. 264-285, 1969.

[20] H.P. Luhn, "The Automatic Creation of Literature Abstracts", *IBM journal of Research and Development*, vol.2, no. 2, pp. 159-165, 1958.

[21] G. Salton, *The Smart Retrieval System – Experiments in Automatic Document Retrieval*, Englewood Cliffs, NJ, Prentice Hall Inc., 1971.

[22] A. Tombros and M. Sanderson, "Advantages of Query Biased Summaries in Information Retrieval", *Proc. of SIGIR-98,* pp. 2-10, Melbourne, Australia, August, 1998.

[23] D. Harman, "TREC-4", *Proc. of the fourth Text Retrieval Conference*, Washington, DC' GPO, 1996

[24] R. Brandow, K. Mitze, and L.F. Rau, "Automatic Condensation of Electronic Publications by Sentence Selection", *Information Processing and Management,* vol.31, no.5, pp. 675-685, 1995.

[25] Y. Chali, "Generic and Query-based Text Summarization using Lexical Cohesion", In R.Cohen and B. Spencer (eds), *Advances in Artificial Intelligence: 15th Conference of the Canadian Society for Computational Studies of Intelligence(AI 2002)*, pp. 293-302, Calgary, Canada, 2002.

[26] L. Hirschman and R. Gaizauskas, "Natural Language Question Answering: The view from here", *Natural language Engineering,* Vol. 7 (4), pp, 275-300, 2001.

[27] W. Bosma, "Query-based Summarization using Rhetorical Structure Theory", *in T. van der Wouden, M. Po, H. Reckman and C. Cremers (edu), 15th Meeting of CLIN,* LOT*,* Leiben, pp. 29-44, 2005.

[28] J. Lin, D. Quan, V. Sinha, K. Bakshi, D. Huynh, B. Katz, and D.R. Kargerm, "What Makes a Good Answer? the Role of Context in Question Answering", *Proc. of the Ninth IFIP TC13 International Conference on Human Computer Interaction*, Zurich, Switzerland, September, 2003.

[29] H. Saggion, K. Bontcheva, and H. Cunningham, "Robust Generic and Query-based Summarization", *Proc. of 10th Conference of the European Chapter of the Association for Computational Linguistics*, EACL-2003, pp. 235-238, Hungary, 2003.

[30] J. L. Neto and A. D. Santos. "Document Clustering and Text Summarization", *Proc. of 4th International conference on practical applications of knowledge discovery and data mining (PADD-2000)*, pp. 41-55, London, 2000.

[31] K. M. Kupiec, J. Pedersen and F. Chen. "A Trainable Document Summarizer", *Proc. of 18th ACM–SIGIR Conference on Research and Development in Information Retrieval*, pp. 68-73, Seattle, USA, July 1995.

[32] C. Aone, M.E. Okurowski, J. Gorlinsky, and B. Larsen, "A Scalable Summarization System using Robust NLP", *Proc. of the ACL 97/EACL-97 Workshop on intelligent scalable Text Summarization*, pp. 82-88, Madrid, Spain, 1997.

[33] M. Witbrock and V. Mittal, "Ultra Summarization: A Statistical Approach to Generating Highly Condensed Non-extractive Summaries", *Proc. of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 315-316,Berkeley, 1999.

[34] J. Conroy and D. O'Leary, "Text Summarization via Hidden Markov Models", *Proc. of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 406-407, 2001.

[35] D. Shen, S. Jian-Tao H. Li, Q. Yang, and Z. Chen, "Document Summarization using Conditional Random Fields", *Proc. of International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2862-2867, 2007.

[36] K. Ono, K. Sumita, and S. Miike, “Abstract Generation Based on Rhetorical Structure Extraction”, *Proc. of the International Conference on Computational Linguistics*, pp. 344-348, Kyoto, Japan, 1994.

[37] M. Daniel, *The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts*, Ph.D. Thesis, University of Toronto, Toronto, 1997.

[38] S. Teufel and M. Moens, “Summarizing Scientific Articles- Experiments with Relevance and Rhetorical Status”, *Computational Linguistics*, vol. 28, no. 4, pp. 409-445, 2002.

[39] D. Radev, S. Teufel, H. Saggion, W. Lam, J. Blitzer, A. Celebi, H. Qi, D. Elliott, and D. Liu, “Evaluation of Text summarization in a Cross-lingual Information Retrieval Framework”, *Technical Report*, Center for Language and Speech Processing, Johns Hopkins University, Baltimore, June, 2002.

[40] D. Radev, E. Hovy, K. McKeown, Introduction to the Special Issue on Summarization, *Computational Linguistics*, vol. 28, no. 4, Association for Computing Machinery, 2002.

[41] B. Georgantopoulos, *Automatic Summarizing Based on Sentence Extraction: A Statistical Approach*, MSc in Speech and Language Processing Dissertation, University of Edinburgh, 1996.

[42] A.H. Morris, G.M. Kasper, and D.A. Adams, “The Effects and Limitations of Automated Text Condensing on Reading Comprehension Performance”, *Information Systems Research*, vol.26, pp. 17-35, 1992.

[43] Y. Yen-Yuan, K. Hao-Ren, Y. Wei-Pang, and M. I-Heng, “Text Summarization using a Trainable Summarizer and Latent Semantic Analysis”, *Information processing management*, vol. 41, no. 1, pp. 75-95, 2005.

[44] C. Grover, B. Hachey, La. N. Hughson, and C. Korycinski, “Automatic Summarization of Legal Documents”, *Proc. of International Conference on Artificial Intelligence and Law (ICAIL 2003),* pp. 243-251, Edinburgh, Scotland, UK, 2003.

[45] C. Grover, B. Hachey, and I. Hughson, "The HOLJ Corpus: Supporting Summarization of Legal Texts", *Proc. of the 5th International workshop on Linguistically Interpreted Corpora (CLIN'04),* pp.47-54, Geneva, Switzerland, 2004.

[46] F. Schilder and H. Molina-Salgado, "Evaluating a summarizer for legal text with a large text collection", *3rd Midwestern Computational Linguistics Colloquium (MCLC)*, Urbana-Champaign, IL, USA, 2006.

[47] A. Farzindar and G. Lapalme, "Letsum, an Automatic Legal Text Summarization System", in *T.Gorden (ed.), Legal Knowledge and Information Systems, JURIX 2004: The seventeenth Annual Conference*, pp. 11-18, Amsterdam, IOS Press,2004.

[48] J.C. Smith and C. Deedman, "The Application of Expert Systems Technology to Case-based Law", *Proc. of International Conference on Artificial Intelligence and Law,* pp. 84-93, 1987.

[49] M. Moens, C. Uyttendaele, and J. Dumortier, "Abstracting of Legal Cases: the Potential of Clustering based on the Selection of Representative Objects", *Journal of the American Society for Information Science*, vol. 50, no. 2, pp.151-161, 1999.

[50] S. Walter and M. Pinkal, "Computational Linguistic Support for LegalOntology Construction", In *Proceedings of International Conference on Artificial Intelligence and Law (ICAIL* 2005), pp. 242-243, Bologna, Italy, June, 2005.

[51] F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, and P.F. Patel-Schneider, *The Description Logics Handbook: Theory, Implementations, and Applications,* Cambridge University Press, 2003.

[52] A. Valente, *Legal Knowledge Engineering: A Modeling Approach*, Amsterdam, IOS Press, 1995.

[53] T.R. Gruber, "Ontolingua: A Mechanism to Support Portable Ontologies", *Technical report*, Knowledge Systems Laboratory, Stanford University, United States, 1992.

[54] P.R.S. Visser, *Knowledge Specification for Multiple Legal Tasks; A Case Study of the Interaction Problem in the Legal Domain*, Computer/Law Series, No. 17, Kluwer Law International, The Hague, The Netherlands, 1995.

[55] R.W. Kralingen, *Frame-based Conceptual Models of Statute Law*, Kluwer Computer/Law Series, 1995.

[56] S. Despres and S. Szulman, "Construction of a Legal Ontology from a European Community Legislative text", In *T.Gordon (ed.), Legal Knowledge and Information Systems Jurix 2004: The Seventeenth Annual Conference,* pp. 79-88, Amsterdam: IOS Press, 2004.

[57] B. Biébow and S. Szulman, "TERMINAE : A Linguistics-based Tool for Building of a Domain Ontology", In *D. Fensel and R. Studer, editors, Proc. of the 11th European Workshop (EKAW'99)*, LNAI 1621, pp. 49–66. Springer-Verlag, 1999.

[58] N. Aussenac-Gilles, B. Biébow, and S. Szulman, "Revisiting Ontology Design: a Methodology based on Corpus Analysis". *In R. Dieng and O. Corby, editors, Knowledge Engineering and Knowledge Management : Methods, Models, and Tools, Proc. of the 12th International Conference, (EKAW'2000)*, LNAI 1937, pp. 172–188. Springer-Verlag, 2000.

[59] N. Fridman and M.A. Musen, "An Algorithm for Merging and Aligning Ontologies: : Automation and Tool Support", *Proc. of the Workshop on Ontology Management at the Sixteenth National Conference on Artificial Intelligence (AAAI-99)*, pp. 17-27, Orlando, AAI Press, 1999.

[60] D. Beeferman, A. Berger, and J. Lafferty, "Statistical Models for Text Segmentation", *Machine Learning*, vol. 34, no. 1-3, pp. 177-210, 1999

[61] V. Borkar, K. Deshmukh, and S. Sarawagi, "Automatic Segmentation of Text into Structured Records", *Proc. of SIGMOD 2001*, pp. 175-186, Santa Barbara, California, ACM, 2001.

[62] A. McCullam, D. Freitag, and F. Pereira, "Maximum Entropy Markov Models for Information Extraction and Segmentation", *Proc. of International Conference Machine Learning 2000*, pp.591-598, 2000.

[63] J. Lafferty, A. McCullam, and F. Pereira, "Conditional Random Fields: Probabilistic Models and for Segmenting and Labeling Sequence Data", *Proc. of International Conference Machine Learning*, pp.282-289, 2001.

[64] M. Christopher Bishop, *Pattern Recognition and Machine Learning*. Springer, Singapore, 2006

[65] H.M. Wallach, "Conditional Random Fields: An Introduction", *Technical Report MS-CIS-04-21*, Department of CIS, University of Pennsylvania, 2004.

[66] J. Kupiec, "Robust part-of-speech tagging using a hidden Markov model", *Computer Speech and Language,* vol. 6, pp. 225-242, 1992.

[67] A. Molina and F. Pla, "Shallow Parsing using Specialized HMMs", *Journal of Machine Learning* Research, vol. 2, pp. 595-613, 2002

[68] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proc. of the* IEEE, vol. 77, no. 2, pp. 257-285, 1989.

[69] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison, *Biological sequence analysis: Probabilistic models of proteins and nucleic acids*, Cambridge University Press, 1998

[70] J.E. Hopcraft, R. Motwani, and J. Ullman, *Introduction to Automata Theory, Languages, and Computation*, Addison Wesley, 2001

[71] F. Pla, A. Molina, and N. Prieto, "Improved Text Chunking by Means of Lexical-Contextual Information in Statistical Language Models", *Proc. of Fourth Conference on Computational Natural Language Learning ACL-2000*, pp. 148-150, New Jersey, USA, 2000

[72] A. J. Viterbi, "Error Bounds for Convolution Codes and Asymptotically Optimal Decoding Algorithm", *IEEE Transactions on Information Processing,* vol. 13, pp. 260-269, 1967

[73] S.D. Pietra, V.D. Pietra, and J. Lafferty, "Inducing Features of Random Fields", *IEEE transactions on pattern analysis and machine intelligence*, vol. 19, no. 4, pp. 380-393, 1997

[74] J. Darroch and D. Ratcliff, "Generalized Iterative Scaling for Log-linear Models", *The Annals of Mathematical Statistics*, vol.43, pp. 1470-1480, 1972

[75] M. Johnson, "Joint and Conditional Estimation of Tagging and Parsing Models", *Proc. of the 39th Annual Meeting of the Association for Computational Linguistics (ACL-01)*, pp. 322–329, 2001.

[76] A. Paz, *Introduction to Probabilistic Automata*, Academic Press Inc., 1971

[77] A. Ratnaparkhi, "A Maximum Entropy Part-of-speech Tagger", P*roc. of the 1st Conference on Empirical Methods in Natural Language Processing (EMNLP-96*), pp.133–142, Philadelphia, USA,1996.

[78] V. Purnyakanok and D. Roth, "The Use of Classifiers in Sequential Inference", *NIPS 13,* 2001

[79] C. Sutton and A. McCallum, "Piecewise Training for Undirected Models", *Proc. of the 21st Conference on Uncertainty in Artificial Intelligence (UAI-05)*, pp. 568–575, Arlington, Virginia, 2005.

[80] A. McCallum and W. Li, "Early Results for Named Entity Recognition with Conditional Random Fields, Feature Induction and Web-enhanced Lexicons". *Proc.of the 7th Conference on Natural Language Learning (CoNLL-03),* pp.188–191, Edmonton, Canada, 2003.

[81] D. Pinto, A. McCallum, X. Lee, and B. Croft, "Table Extraction using Conditional Random Fields", *Proc. of the 26th ACM SIGIR,* pp. 235-242, Canada, 2003

[82] F. Sha and F. Pereira, "Shallow Parsing with Conditional Random Fields", *Proc. of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL-03)*, pp. 213–220, Edmonton, Canada, 2003.

[83] A. L. Berger, S.A. Della Pietra, and V. J. D. Pietra, "A Maximum Entropy Approach to Natural Language Processing", *Computational Linguistics*, vol.22, no.1, pp.39–71, 1996.

[84] A. Smith and M. Osborne, "Regularization Techniques for Conditional Random Fields: Parameterized versus Parameter-free", *Proc. of the 2nd International Joint Conference on Natural Language Processing (IJCNLP-05)*, pp. 896–907, Jeju Island, Korea, 2005.

[85] A. Buckley, A. Singhal, M. Mitra, and G. Salton, "New retrieval approaches using SMART", *Proc. of TREC-4,* pp. 25-48, 1996.

[86] Y. Nakao, "An Algorithm for One-page Summarization of a Long Text Based on Thematic Hierarchy Detection", *Proc. of the 26th Annual Meeting of the Association for Computational* Linguistics, pp. 302-309, New Jersey, 2000.

[87] H. Kozima, "Text Segmentation Based on Similarity Between Words", *Proc. of the 31st Annual Meeting of the Association for Computational* Linguistics, pp. 286-288, Columbus, OH, 1993.

[88] M. A. Hearst. "Multi-paragraph Segmentation of Expository Text", *Proc. of the 32nd Meeting of the Association for Computational* Linguistics, pp. 9-16, Las Cruces, NM, 1994.

[89] J. Allan, J. Carbonell, G. Doddington, J. Yamron, and Y.Yang, "Topic Detection and Tracking Pilot Study Final Report", *Proc. of the DARPA Broadcast News Transcription and Understanding Workshop,* pp. 194-218, 1998.

[90] Y. Freddy and Y. Choi, "Advances in Domain Independent Linear Text Segmentation", *Proc. of the first conference on North American chapter of the Association for Computational Linguistics*, vol. 4, pp. 26-33, ACM International Conference Proceeding Series, 2000.

[91] Y. Freund and R. Schapire, "Experiments with a New Boosting Algorithm", *Proc. of the 13th International Conference on Machine Learning (ICML-96)*, pp.148–156, Bari, Italy, 1996.

[92] P. Clark and T. Niblett, "The CN2 Induction Algorithm", *Machine Learning*, vol.3, no. 1, pp. 261-283, 1989

[93] W. Cohen, "Fast effective rule induction", *In Machine learning, Proc. of the Twelfth International Conference.* pp.335-342, Lake Tahoe, California, Morgan Kaufmann, 1995.

[94] J. R. Quinlan, *C4.5: Programs for machine learning*, Morgan Kaufmann, 1994.

[95] W. Cohen and Y. Singer, "A Simple, Fast, and Effective Rule Learner", *Proc. of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)*, pp.335-342, AAAI Press, 1999.

[96] C. Brunk and M. Pazani, "An Investigation of Noise-tolerant Relational Concept Learning Algorithms", *Proc. of the Eighth International Workshop on Machine Learning*, pp. 389-393, Ithaca, New York, 1991.

[97] J. Furnkranz and G. Widmer, "Incremental Reduced Error Pruning", *Machine Learning: Proceedings of the Eleventh International Conference*, pp. 70-77, New Brunswick, New Jersey, 1994.

[98] Y. Freund, “Boosting a Weak Learning Algorithm by Majority”, *Information and Computation*, vol. 121, pp. 256-285, 1995

[99] R.E. Schapire and Y. Singer, “Improved Boosting Algorithms using Confidence-rated Predictions”, *Proc. of the Eleventh Annual Conference on Computational Learning Theory*, pp. 80-91, 1998.

[100] J.H. Friedmen and B.E. Popescu, *Predictive learning via rule ensembles*, Technical Report, Stanford University, 2005.

[101] H. Yunhua, H. Li, Y. Cao, D. Meyerzon, and Q. Zheng, “Automatic Extraction of Titles from General Documents using Machine Learning”, *Proc. of the 5th ACM/IEEE-CS Joint conference on Digital Libraries,* pp. 145-154, USA, 2005.

[102] M. Saravanan, B. Ravindran, and S. Raman, “Improving Legal Document Summarization using Graphical Models”, *Proc. of the 19th International Conference on Legal Knowledge and Information Systems JURIX 2006,* pp.51-60, Paris, France, 2006.

[103] V.K. Bhatia, *Analyzing Genre: Language Use in Professional Settings*, Longman, London, 1999.

[104] F. Peng and A. McCullam, “Accurate Information Extraction from Research Papers using Conditional Random Fields”, *Information Processing Management*, vol. 42, no.4, pp. 963-979, 2006.

[105] J.M. Wiebe, “Tracking Point of View in Narrative”, *Computational Linguistics*, vol.20. no.2, pp.223-287,1994

[106] R. Brandow, K. Mitze, and L.F. Rau, “Automatic Condensation of Electronic Publications by Sentence Selection”, *Information Processing Management,* vol. 31, no. 5, pp.675-685, 1995

[107] S. Siegal and N.J. Castellan. *Nonparametric Statistics for the Behavioral Sciences*, McGraw Hill, Berkeley, CA, 1988.

[108] K. Krippendorff, *Content Analysis: An Introduction to its Methodologies*, Beverly Hills: Sage publications, 1980.

[109] C. J. van Rijsbergen, *Information Retrieval*, 2nd Ed, London, Butterworths, 1979.

[110] R. Bruce and J. Wiebe, "Word Sense Distinguishability and Inter-coder Agreement", *Proc. 3rd conference on Empirical methods in Natural Language Processing (EMNLP-98)*, pp. 53-60, ACL SIGDAT, Granada, Spain, 1998.

[111] N. Tomuro, "Tree-cut and a Lexicon based on Systematic Polysemy", *Proc. of Second meeting of North American Chapter of the Association for Computational Linguistics on Language technologies*, pp. 1-8, Pennsylvania, 2001.

[112] E.D. Liddy, "The Discourse-level Structure of Empirical Abstracts: An Exploratory Study", *Information Processing and Management*, vol.27, no.1, pp. 55-81, 1991.

[113] *Enterprise Search from Microsoft: Empower People to Find Information and Expertise*, A Microsoft White Paper, ttp://www.microsoft.com/ enterprisesearch.

[114] B.J. Breuker and R. Winckels, "Use and Reuse of Legal Ontologies in Knowledge Engineering and Information Management", *Proc. of the Workshop on Legal Ontologies & Web based Legal Information Management* (*ICAIL 2003)*, 2003.

[115] A. Gangemi, A. Prisco, M.T. Sagri, G. Steve, and D. Tiscornia, "Some Ontological Tools to Support Legal Regulatory Compliance, with a Case Study", In *Workshop WORM Core*, LNCS, pp. 607-620, Springer Verlag, 2003.

[116] W. Hohfeld, *Fundamental Legal Conceptions as Applied in Legal Reasoning.* Yale University Press, 1996.

[117] T.J.M. Bench-Capon and P.R.S. Visser, "Ontologies in Legal Information Systems; the Need for Explicit Specifications of Domain Conceptualizations", *Proc. of International Conference on Artificial Intelligence and Law (ICAIL-97)*, pp. 132-141, Melbourne, Australia, 1997.

[118] G. Lame, "Constructing an IR-oriented Legal Ontology", *Proc. of the Second International Workshop on Legal Ontologies,10th International Conference on Legal Information Retrieval System, JURIX 2001*, pp. 31-36, Amsterdam, Netherlands, 2001.

[119] A. Hotho, S. Staab, and G. Stumme, "Ontologies Improve Text Document Clustering", *Proc. of IEEE International Conference on Data Mining (ICDM 2003)*, pp. 541-544, 2003.

[120] R. Prieto-Diaz, "A Faceted Approach to Building Ontologies", *Proc. of IEEE international conference on Information Reuse and Integration,* IRI 2003.

[121] M. Moens, *Automatic Indexing and Abstracting of Document texts*. Kluwer Academic Publications, London.

[122] A. Gómez-Pérez. "Ontology Evaluation", *Handbook on Ontologies*, pp. 251-274, 2004.

[123] N. Guarino, "Formal Ontology and Information Systems", *Proc. of FOIS'98*, pp.3 -15, Trento, Italy, June, 1998.

[124] O. Kurland, *Inter-document Similarities - Language Models, and ad hoc Information Retrieval*, Ph.D.Thesis, 2006.

[125] M. Porter, "An algorithm for suffix stripping", *Automated Library and Information Systems*, vol.14, no.3, pp. 130-137, 1980.

[126] M. Saravanan, P.C. Reghu Raj, V.M. Murhty, and S. Raman, "Improved Porter's Algorithm for Root Word Stemming", *Proc. of International Conference on Natural Language Processing (ICON 2002)*, pp. 21-30, Bombay, India, December, 2002.

[127] A. McCallum, *Bow: A Tool Kit for Statistical Language Modeling Text Retrieval, Classification and Clustering*, http://www.cs.cmu.edu/~McCullum/ bow, 1996.

[128] W.B. Frakes and R. Baeza-Yates, *Information Retrieval: Data Structures & Algorithms*, Englewood Cliffs, NJ, Prentice-Hall Inc., Ch.8, 1992.

[129] W. Kraaij and R. Pohlmann, "Viewing Stemming as Recall Enhancement", H.P. Frei, D. Harman, P. Schauble & R. Wilkinson (Eds.), *Proc. of the 17th ACM SIGIR conference*, pp. 40-48, Zurich, August, 1996.

[130] M.A.K. Halliday and R. Hasan, *Cohesion in English*. Longman, London, 1976.

[131] G. Salton, *Automatic Text Processing : the transformation, analysis, and retrieval of information by computer*, Reading, M.A: Addison-Wesley, 1989.

[132] M.A. Hearst, "Texttiling: Segmenting Text in to Multi-paragraph Subtopic Passages", *Computational Linguistics*, vol.23, no. 1, pp. 33-65, 1997.

[133] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman, "Indexing by Latent Semantic Analysis", *Journal of the Society for Information Science*, vol.41, pp. 391-407, 1990.

[138] R. A. Johnson, Miller, and Freund. *Probability and Statistics for Engineers*, 5th Ed, Prentice Hall of India, New Delhi, 1994.

[139] S.M. Katz, "Distribution of content words and phrases in text and language Modeling", *Natural language Engineering*, vol.2, no.1, pp. 15-59, 1995.

[140] K.W. Church and W.A. Gale, "Poisson Mixtures", *Natural Language Engineering,* vol.2, pp. 163-190, 1995.

[141] R.L. Donaway, K.W. Drummey, and L.A. Mather, "A Comparison of Rankings Produced by Summarization Evaluation Measures", *Proc. of the Workshop on Automatic Summarization*, post-conference workshop of ANLP-NAACL 2000, pp.69-78, Seattle, WA, 2000.

[142] C. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries", *Proc. of the workshop on text summarization branches out (WAS 2004)*, pp. 74-81, Barcelona, Spain, July, 2004.

[143] I. Mani, D. House, G. Klein, L. Hirschman, L. Orbsl, T. Firmin, M. Chrzanowski, and B. Sundheirm, "The TIPSTER SUMMAC text summarization Evaluation", *MITRE Technical report*, MTR98W0000138, The MITRE corporation, 1998.

[144] H. Jing, R. Barzilay, K. Mckeown and M. Elhadad, "Summarization Evaluation Methods: Experiments and Analysis", *Proc. of AAAI 98 Spring Symposium on Intelligent Text Summarization*, pp. 60-68, 1998.

[145] M. Hajime and O. Manabu, "A Comparison of Summarization Methods based on Task-based Evaluation", *Proc. of 2nd International conference on language resources and evaluation*, LREC-2000, pp. 633-639, Greece, Athens, 2000.

[146] K.S. Jones and J.R. Galliers, *Evaluating natural language processing review*, New York, Springer.

[147] *MEAD Summarizer*, http://tangra.si.umich.edu/clair/md/demo.cgi.

[148] K.S. Papineni,. T.W. Roukos, and W.J. Zhu, BLEU: a method for automatic evaluation of Machine Translation, *IBM Research Report RC22176 (W0179-022)*, 2001.

[149] H. Saggion, D. Radev, S. Teufel, and W. Lam, "Meta-evaluation of Summaries in a Cross-lingual Environment using Content-based Metrics", *Proc. of 19th Conference on Computational Linguistics*, vol. 1, pp. 1-7, Taipei, Taiwan, 2002.

[150] C.Lin and E. Hovy, "Automatic Evaluation of Summaries Using N-gram Co-Occurrence Statistics", *Proc. of the Human Technology Conference (HLT-NAACL-2003)*, pp. 62-69, Edmonton, Canada, 2003.

LIST OF PAPERS BASED ON THE THESIS

Refereed International Journal

1. **M. Saravanan,** S. Raman, and B. Ravindran, A Probabilistic Approach to Multi-document summarization for generating a Tiled Summary, *International Journal of Computational Intelligence and Applications*, vol. 6, no. 2, pp. 231-243, Imperial College Press,2006.

Refereed International Conferences

1. **M. Saravanan,** S. Raman, and B. Ravindran, "A Probabilistic approach to Multi-document summarization for generating a tiled summary", *Proc. of the International Conference of Computational Intelligence and Multimedia Applications (ICCIMA'05*), pp.167-172, University of Nevada, USA, August, 2005.

2. **M. Saravanan,** B. Ravindran, and S. Raman, "Improving Legal Document Summarization using Graphical Models", *Proc. of 19th International Annual Conference on Legal Knowledge and Information Systems,* JURIX 2006, pp. 51-60, Paris, France, December, 2006.

3. **M. Saravanan,** B. Ravindran, and S. Raman, "Using Legal Ontology for Query Enhancement in generating a document summary", in *Proceedings of JURIX 2007*, 20*th International Annual Conference on Legal Knowledge and Information Systems*, Leiden, Netherlands, 13-15th Dec 2007.

4. **M. Saravanan,** B. Ravindran, and S. Raman, "Automatic identification of rhetorical roles using conditional random fields for Legal Document Summarization", in P*roceedings of IJCNLP 2008*, *International Joint Conference on Natural Language Processing*, Hyderabad, India, 7-12th Jan 2008.

National Conference

1. **M. Saravanan,** B. Ravindran, and S. Raman, "A review of automatic summarizer", *Proc. of the Workshop on Optical Character Recognition with workflow and Document Summarization*, IIIT Allahabad, March, 2005.

Papers communicated

1. M**. Saravanan, B. Ravindran**, **and S. Raman,** "Improving Legal *Information Retrieval Using Ontological Framework*", communicated to *International Journal of Artificial Intelligence and Law, Springer Verlag*

2. M**. Saravanan, and B. Ravindran**, "*Automatic Identification of Rhetorical Roles for Document Segmentation and Summarization*", communicated to *International Journal of Artificial Intelligence and Law, Springer Verlag*