Improving Legal Document Summarization Using Graphical Models

M. Saravanan[a,1], B. Ravindran and S. Raman
Department of Computer Science and Engineering,
IIT Madras, Chennai- 600 036, Tamil Nadu, India.

Abstract. In this paper, we propose a novel idea for applying probabilistic graphical models for automatic text summarization task related to a legal domain. Identification of rhetorical roles present in the sentences of a legal document is the important text mining process involved in this task. A Conditional Random Field (CRF) is applied to segment a given legal document into seven labeled components and each label represents the appropriate rhetorical roles. Feature sets with varying characteristics are employed in order to provide significant improvements in CRFs performance. Our system is then enriched by the application of a term distribution model with structured domain knowledge to extract key sentences related to rhetorical categories. The final structured summary has been observed to be closest to 80% accuracy level to the ideal summary generated by experts in the area.

Keywords. Automatic Text Summarization, Legal domain, CRF model, Rhetorical roles.

1. Introduction

Text summarization is one of the relevant problems in the information era, and it needs to be solved given that there is exponential growth of data. It addresses the problem of selecting the most important portions of text and that of generating coherent summaries [1]. The phenomenal growth of legal documents creates an invariably insurmountable scenario to read and digest all the information. Automatic summarization plays a decisive role in this type of extraction problems. Manual summarization can be considered as a form of information selection using an unconstrained vocabulary with no artificial linguistic limitations [2]. Automatic summarization on the other hand focuses on the retrieval of relevant sentences of the original text. The retrieved sentences can then be used as the basis of summaries with the aid of post mining tools. The graphical models were recently considered as the best machine learning techniques to sort out information extraction issues. Machine learning techniques are used in Information Extraction tasks to find out typical patterns of texts with suitable training and learning methodologies. These techniques make the machine learn through training examples, domain knowledge and indicators.

The work on information extraction from legal documents has largely been based on semantic processing of legal texts [2] and applying machine learning algorithms like C4.5, Naïve Bayes, Winnow and SVM's [3]. 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. Also the above studies mention the need for an active learning method to automate and reduce the time taken for this process. The recent work on automatically extracting titles from general documents using machine learning methods shows that machine learning approach can work significantly better than the baseline methods for meta-data extraction [4]. 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 [5]. For automatic summarization task, it is necessary to explore more features which are representative of the characteristics of texts in general and legal text in particular.

Moreover, the application of graphical models to explore the document structure for segmentation is one of the interesting problems to be tried out. The graphical models have been used to represent the joint probability distribution p(x, y), where the variables y represent attributes of the entities that we wish to predict, and the input variables x represent our observed knowledge about the entities. But

[a] Corresponding Author: Research Scholar, Department of Computer Science & Engineering, IIT Madras, Chennai – 600036, Tamil Nadu, India; Email: msdess@yahoo.com, msn@cse.iitm.ernet.in.
[1] Working as a Faculty in Department of Information Technology, Jerusalem College of Engineering, Pallikkaranai, Chennai-601302, Tamil Nadu, India.

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. The conditional nature of models such as McCallum et al's [6] maximum entropy Markov Models result in improved performance on a variety of text processing tasks. These non-generative finite state models are susceptible to a weakness known as the label bias problem [7]. The label bias problem can significantly undermine the benefits of using a conditional model to label sequences. To overcome this problem, Lafferty et al [7] introduced conditional random fields (CRFs), a form of undirected graphical model that defines a single log-linear distribution over the joint probability of an entire label sequence given a particular observation sequence. Conditional Random Field (CRF) is one of the newly used graphical techniques which have been applied for text segmentation task to explore the set of labels in a given text.

In this paper, we have tried a novel method of applying CRF's for segmentation of texts in legal domain and use this knowledge for setting up importance of extraction of sentences through the term distribution model [8] used for summarization of a document. In many summarizers term weighting techniques are applied to calculate a set of frequencies and weights based on the number of occurrence of words. The use of mere term-weighting technique has generally tended to ignore the importance of term characterization in a document. Moreover, they are not specific in assessing the likelihood of a certain number of occurrences of particular word in a document and they are not directly derived from any mathematical model of term distribution or relevancy [9]. The term distribution model used in our approach is to assign probabilistic weights and normalize the occurrence of the terms in such a way to select important sentences from a legal document. The ideal sentences are presented in the form of a structured summary for easy readability and coherency. Also, a comprehensive automatic annotation has been performed in the earlier stages, which may overcome the difficulties of coherency faced by other summarizers. Our annotation model has been framed with domain knowledge and is based on the report of genre analysis [10] for legal documents. The summary generated by our summarizer can be evaluated with human generated head notes (a short summary for a document generated by experts in the field) which are available with all legal judgments. The major difficulty of head notes 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 a detailed structural summary of a legal document. We have followed an intrinsic evaluation procedure to compare the sentences present in structural summary with sentences appearing in human generated head notes.

In this paper, our investigation process deals with three issues which did not seem to have been examined previously. They are:

1. To apply CRFs for segmentation by the way of structuring a given legal document.
2. To find out whether extracted labels can improve document summarization process.
3. To create generic structure for summary of a legal judgment belonging to different sub-domains.

Experimental results indicate that our approach works well for automatic text summarization of legal documents. The rest of the paper is organized as follows. In Section 2, we identify the rhetorical roles present in legal judgment and in Section 3, we explain the details of CRF and discuss the characteristics of feature sets in Section 4. In Section 5, we describe the proposed system components. Section 6 gives our experimental results and we make concluding remarks in Section 7.

2. Identification of Rhetorical Roles

In our work, we are in the process of developing a fully automatic summarization system for a legal domain on the basis of Lafferty's [7] segmentation task and Teufel & Moen's [11] gold standard approaches. Legal judgments are different in characteristics compared with articles reporting scientific research papers and other simple domains related to the identification of basic structure of a document. The main idea behind our approach is to apply probabilistic models for text segmentation and to identify the sentences which are more important to the summary and that might be presented as labeled text tiles. Useful features might include standard IR measures such as probabilistic feature of word, but other highly informative features are likely to refer conditional relatedness of the sentences. We also found the usefulness of other range of features in determining the rhetorical status of sentences present in a document.

For a legal domain, the communication goal of each judge is to convince his/her peers with their sound context and having considered the case with regard to all relevant points of law. The fundamental need of a legal judgment is to legitimize a decision from authoritative sources of law. To perform a summarization methodology and find out important portions of a legal document is a complex problem. Even the skilled lawyers are facing difficulty in identifying the main decision part of a law report. The genre structure identified in our process plays a crucial role in identifying the main decision part in the way of breaking the document in anaphoric chains. A different set of contextual rhetorical schemes will be taken into considerations which are discussed here with different purview. Table 1 shows the basic rhetorical roles present in a legal document based on Teufel & Moen's [11] and Farzindar [12].

Table 1. Description of the basic rhetorical scheme for a legal domain.

Labels	***Description***
Facts of the case	The sentence describes 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 relatedness	The sentences contain the details of other cases coded in this case.

These general classifications have been enhanced into seven different labeled elements in our work for a more structured presentation. Our current version of labels will explore the different category of contents present in the legal document and this in turn will be helpful at the time of presentation of a final summary. 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 within the section does not appear to be significant for any Indian law judgments, since most of the judgments do not follow any standard format of discussion related to the case. Some of the judgments do not even follow the general structure of a legal document. To overcome this problem, positioning of word or sentence in a document is not considered as one of the important features in our work. Our approach to explore the elements from the structure of legal documents has been generalized in the following fixed set of seven rhetorical categories based on Bhatia's [10] genre analysis shown in Table 2.

In common law system, decisions made by 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 head note includes the sentences which are related to the reason for the decision. These sentences very well justify the judge's decision, and in non-legal terms may describe the central generic sentences of the text. So, we reckon this as one of the important elements to be included in our genre structure of judgments. Usually, the knowledge of the ratio appears in 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 the position of text. From the Indian court judgments, we found that ratio can be found in any part of the decision section of a law report, but usually they appear as complex sentences. It is not uncommon to find that the experts differ among themselves on the identification of 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 field a rich set of features that may aid classification or segmentation of given document. More details of CRFs are given in the next section.

Table 2. The current working version of the rhetorical annotation scheme for legal judgments.

Rhetorical Status	*Description*
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".
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.
Arguing the case	Application of legal principle/law advocated by contending parties to a given set of proved facts.
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.
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..
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.
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

3. Conditional Random Fields

Conditional Random Fields (CRFs) model is one of the recently emerging graphical models which has been used for text segmentation problems and proved to be one of the best available frame works compared to other existing models [7]. In general, the documents involved are segmented and each segment gets one label value might be considered as a needed functionality for text processing which could be addressed by many semantic and rule based methods. Recently, machine learning tools are extensively used for this purpose. CRFs are undirected graphical models used to specify the conditional probabilities of possible label sequences given an observation sequence. Moreover, the conditional probabilities of label sequences can depend on arbitrary, non independent features of the observation sequence, since we are not forming the model to consider the distribution of those dependencies. These properties led us to prefer CRFs over HMM and SVM classifier [13]. In 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 sequence labeling.

Let us define the linear chain CRF with parameters $C = \{C_1, C_2, \ldots\}$ defining a conditional probability for a label sequence $l = l_1, \ldots l_w$ (e.g., Establishing facts of the case, Final decision, etc.) given an observed input sequence $s = s_1, \ldots s_W$ to be

$$P_C(l \mid s) = \frac{1}{Z_s} \exp\Big[\sum_{t=1}^{w} \sum_{k=1}^{m} C_k \, f_k (l_{t-1}, l_t. s, t) \qquad \ldots\ldots\ldots\ldots\ldots \quad (1)$$

where Z_s is the normalization factor that makes the probability of all state sequences sum to one, $f_k(l_{t-1}, l_t, s, t)$ is one of m feature functions which is generally binary valued and C_k is a learned weight associated with 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_k is the feature function PHRASE= "points for consideration" is belongs to s at position t in the sequence. Large positive values for C_k 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 labeled sequence in a training set D = {(s_t, l_t) : t = 1,2,…w), written as:

$$L_C(D) = \sum_i \log P_C(l_i \mid s_i)$$

$$= \sum_i \Big(\sum_{t=1}^{w} \sum_{k=1}^{m} C_k \, f_k (l_{t-1}, l_t. s, t) - \log Z_{s_i} \Big) . \ \ldots\ldots\ldots \quad (2)$$

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 probable labeling sequence for an input s_i can be efficiently calculated by dynamic programming using modified Viterbi algorithm. These implementations of CRFs are done using newly developed java classes which also uses a quasi-Newton method called L-BFGS to find these feature weights efficiently.

4. Feature sets

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. The CRF performance has been improved significantly by efficient feature mining techniques. Identifying state transitions of CRF were also considered as one of the important features in any of information extraction task [13].

In addition to the standard set of features, we have also added other related features to reduce the complexity of legal domain. We will discuss some of the important features which have been included in our proposed model all in one framework:

Indicator/cue phrases – The term ‘cue phrase’ indicates the key phrases frequently used 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 where each and every cue phrases related to different labels. In this study, we encoded this information and generated automatically explicit linguistic features. Our initial cue phrase set has been enhanced based on expert suggestions. These cue phrases can be used by the automatic summarization system to locate the sentences which correspond to a particular category of genre of legal domain. If 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 cannot capture the fact that all phrases are present in a document. 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 functions for the rules are set to 1 if they match words/phrases in the input sequence exactly.

Named entity recognition - This type of recognition is not considered fully in summarizing scientific articles [10]. But in our 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 in the sentences.

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

State Transition features - In CRFs, state transitions are also represented as features [13]. The feature function f_k (l_{t-1}, l_t. s, t) in Eq. (1) 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 corresponding to appearance of years attached with Section and Act Nos. related to the labels ***Arguing the case*** and **Arguments**. Also the appearance of some of the cue phrases in a label ***identifying the case*** can be considered 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. Moreover, in some cases, we need to add a set of previous sentences based on the states.

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 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.

5. The Proposed System

The overall architecture is shown in Figure 1. It consists of different modules organized as a channel for text summarization task.

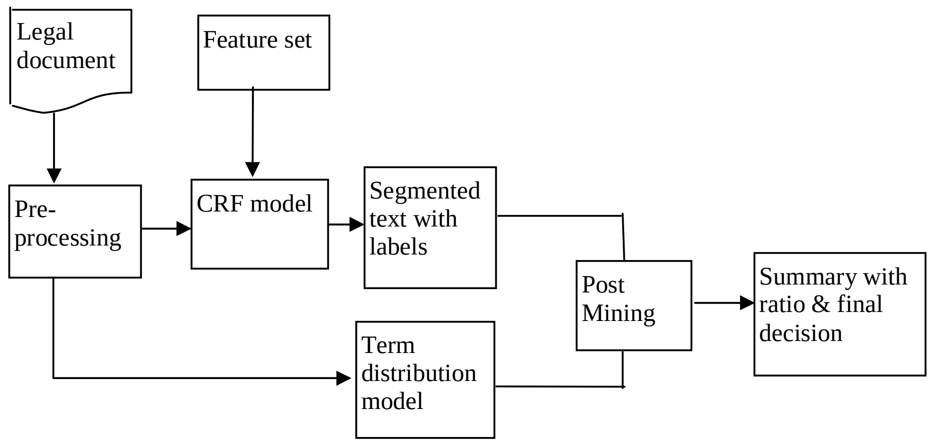

Figure 1. Different processing stages of a system

The automatic summarization process starts with sending legal document to a preprocessing stage. In this preprocessing stage, the document is to be divided into segments, sentences and tokens. We have introduced some of the new feature identification techniques to explore paragraph alignments. This process includes the understanding of abbreviated texts and section numbers and arguments which are very specific to the structure of legal documents. The other useful statistical natural language processing tools, such as filtering out stop list words, stemming etc., are carried out in the preprocessing stage. The resulting intelligible words are useful in the normalization of terms in the term distribution model. The other phase dealing with selecting suitable feature sets for the identification of rhetorical status of each sentence has been implemented with a graphical model (CRFs) which aid in document segmentation.

The term distribution model used in our architecture is K-mixture model [14]. The K-mixture model is a fairly good approximation model compared to Poisson model and it is described as the mixture of Poisson distribution and its terms can be computed 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:

$$P_i(k) = (1-r)\ \delta_{k,0} + \frac{r}{s+1}\ \frac{(s)^k}{(s+1)^k} \qquad \ldots\ldots\ldots (3)$$

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 observed Inverse Document Frequency (*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 (compared to the case where a term has only one occurrence per document). 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. Hence the application of K-mixture model brings out a good extract of generic sentences from a legal document to generate a summary. Post mining stage deals with matching of sentences present in the summary with segmented document to evolve the structured summary.

The sentences related to the labels which have been selected during term distribution model are useful to present the summary in the form of structured way. This structured summary is more useful for generating coherency and readability among the sentences present in the summary. Legal judgment head notes are mainly concentrated on the label *Ratio of Decision*. So, in addition to our structured summary, we are also giving *ratio decidendi* and final decision of a case. This form of summary of legal document is more useful to the legal community not only for the experts but also for the practicing lawyers.

6. Results and Discussion

Our corpus presently consists of 200 legal documents related to rent control act, out of which 50 were annotated. 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 entire corpus consists of judgments dated up to the year 2006. The judgments can be divided into exclusive sections like Rent Control, Motor Vehicle, Family Law, Patent, Trademark and Company law, Taxation, Sales Tax, Property and Cyber Law, etc. Even though it is unique, we had a generalized methodology of segmentation for document belonging to different categories of civil court judgments. The header of a legal judgment contains the information related to a petitioner, respondent, judge details, court name and case numbers which were removed and stored in a separate header dataset. It is a common practice to consider human performance as an upper bound for most of the IR tasks. So in our evaluation, the performance of the system has been successfully tested by matching with human annotated documents.

We evaluate our work in two steps; first the evaluation of correct segmentation of the legal judgments and second, a macro evaluation of a final summary. The evaluation of the first step looks very promising since we have obtained more than 90% correct segmentation in most of the categories (which included the most important *Framing of Issues* and *Ratio Decidendi*) and nearer to 80% in *Arguments* and *Arguing the case* rhetorical schemes shown in Table 3. 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 [15]. The advantage of Kappa over annotation scheme is that it factors out random agreement among the categories. The experimental results shows that humans identify the seven rhetorical categories with a reproducibility of K = 0.73 (N=3816; k=2, where K stands for the Kappa coefficient, N for the number of sentences annotated and k for the number of annotators). Now, we report the system performance with precision and recall values for all seven rhetorical categories using CRF model in Table 3. The system performs well for *Ratio decidendi* and *final decision* which are the main contents for the head notes generated by human experts. *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

Table 3. Precision, Recall and F-measure for seven rhetorical categories.

Rhetorical Category	**Precision**	**Recall**	**F-Measure**
Identifying the case	0.946	0.868	0.905
Establishing facts of the case	0.924	0.886	0.904
Arguing the case	0.824	0.787	0.805
History of the case	0.808	0.796	0.802
Arguments	0.860	0.846	0.853
Ratio decidendi	0.924	0.901	0.912
Final decision	0.986	0.962	0.974
Micro Average	**0.896**	**0.864**	**0.879**

Table 3 shows the good performance of CRF model with efficient features sets for text segmentation task. The above results may contribute to the generation of structured and efficient summary in next phase. Use of these identified rhetorical categories can help in modifying the probabilistic weights of term distribution model in such way as to give more importance to ratio of the decision and others categories. The extracted key sentences from the legal document using probabilistic model should be compared with head notes generated by experts in the area. Figure 2 shows the results of our system generated summary in unstructured format, but it uses CRF results for important sentence extraction. The result of system generated summary is compared with human generated head notes, and it is found that F-measure reaches approximately 80%. We need to improve more on the findings of ratio of decision category to get 100% accuracy in segmentation stage, so as to improve final system-generated summary to maximum level. The summary presented in Figure 3 shows the importance of arranging the sentences in a structured manner as it not only improves the readability and coherency but also gives out more inputs like court's arguments to get a comprehensive view of the Ratio and Disposal of the case.

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. Admittedly the building was rented out by the rent agreement dated 6.10.1980 by the mother of the petitioner. 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 2. Unstructured summary produced by our system, the original judgment has 1100 words and summary is 15% 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 3. System output (Structured summary) for example judgment containing title, petitioner, respondent, rhetorical categories and relevant sentences.

7. Conclusion

This paper highlights the construction of proper features sets with an efficient use of CRF for segmentation and presentation tasks, in the application of extraction of key sentences from legal judgments. While the system presented here shows the 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 during the extraction of key sentences by applying term distribution model. The mathematical model based approach for extraction of key sentences has yielded a better results compared to simple term weighting methods. We have also applied better evaluation metrics to evaluate the results rather than using simple word frequency and accuracy.

We have presented an initial annotation scheme for the rhetorical structure of the rental act sub-domain, assigning a label indicating the rhetorical status of each sentence in a portion of a document. The next phase of our research work will involve refining our annotation scheme for other sub-domains related to legal domain. After completing this phase we plan to develop a legal ontology for querying and generating multi-document summarization in the legal domain.

Acknowledgements. We would like to thank the legal fraternity for the assistance and guidance governs to me. Especially we express my sincere gratitude to Mr. S. Mohan, a retired chief justice of Supreme Court of India for his expert opinions. We also thank advocates Mr. S.B.C. Karunakaran and Mr. K.N. Somasundaram for their domain advice and continuous guidance in understanding the structure of legal document and for hand annotated legal documents.

References

[1] M. Brunn, Y. Chali and C.J. Pinchak, Text Summarization using lexical chains, *Workshop on Text summarization in conjuction with the ACM SIGIR Conference*, New Orleans, Louisiana, 2001.

[2] Claire Grover, Ben Hachey, la n Hughson and Chris Korycinski, Automatic summarization of legal documents, *In. Proceedings of International Conference on Artificial Intelligence and Law*, Edinburgh, UK, 2003.

[3] Ben Hachey and Claire Grover, Sequence Modelling for sentence classification in a legal summarization system, *In Proceedings of the 2005 ACM Symposium on Applied Computing*, 2005

[4] Yunhua Hu, Hang Li, Yunbo Cao, Li Teng, Dmitrly Meyerzon and Qinghua Zheng, Automatic extraction of titles from general documents using machine learning, *Information Processing Management*, 42, PP. 1276-1293, 2006.

[5] Marie-Francine Moens, Caroline uyttendaele, and Jos Durmortier, Intelligent Information Extraction from legal texts, *Information & Communications Technology Law*, 9(1), PP. 17-26, 2000.

[6] Andrew McCallum, Dayne Freitag, and Fernando Pereira, Maximum entropy Markov models for information extraction and segmentation, *In proceedings of International Conference on Machine Learning*, PP. 591-598, 2000.

[7] John Lafferty, Andrew McCallum, and Fernando Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, In *Proceedings of International Conference on Machine Learning* 2001.

[8] M. Saravanan, S. Raman, and B.Ravindran, A Probabilistic approach to multi-document summarization for a tiled summary, *In Proceedings of International Conference on Computational Intelligence and Multimedia Applications*, Los Vegas, USA, Aug 16-18, 2005.

[9] C.D. Manning and H. Schutze, *Foundations of Statistical Natural Language Processing*, The MIT Press, London, England, 2001.

[10] V.K. Bhatia, Analyzing Genre: Language Use in Professional Settings. London: Longman, 1993.

[11] Simone Teufel and Marc Moens, Summarizing scientifi articles – experiments with relevance and rhetorical status, *Association of Computational Linguistics*,28(4), PP. 409-445, 2002.

[12] Atefeh Farzindar and Guy Lapalme, Letsum, an automatic legal text summarizing system, *Legal Knowledge and Information System*, Jurix 2004: The seventeenth Annual Conference, Amsterdam: IOS Press, PP. 11-18, 2004.

[13] Fuchun Peng, and Andrew McCallum, Accurate information extraction from research papers using conditional random fields, *Information Processing Management*. 42(4), PP. 963-979, 2006

[14] S.M. Katz, Distribution of content words and phrases in text and language modeling, *Natural language Engineering*, 2(1), PP. 15-59, 1995.

[15] Siegal, Sidney and N.John Jr. Castellan, *Nonparametric Statistics for the Behavioral Sciences*, McGraw Hill, Berkeley, CA, 2nd Edition., 1988.