LetSum, an Automatic Text Summarization system in Law field

Atefeh Farzindar and Guy Lapalme
RALI, Département d'Informatique et recherche opérationnelle
Université de Montréal, Québec, Canada, H3C 3J7
{farzinda,lapalme}@iro.umontreal.ca

Abstract

This paper presents our work on the development of a new methodology for automatic summarization of justice decision. We describe LetSum (Legal text Summarizer), a prototype system, which determines the thematic structure of a judgment in four themes INTRODUCTION, CONTEXT, JURIDICAL ANALYSIS and CONCLUSION. Then it identifies the relevant sentences for each theme. We have carried our an evaluation of produced summaries with statistical method and also human evaluation based on jurist judgment. The results so far indicate good performance of the system when compared with other summarization technologies.

1 Introduction

Legal experts perform difficult and responsible legal clerical work which requires accuracy and speed. This task includes understanding, interpreting, explaining and researching in a wide variety of legal documents. A summary of a judgement, as a compressed but accurate restatement of its content, helps in organizing a large volume of documents and in finding the relevant judgments for their case. For this reason, the decisions are frequently manually summarized by legal experts. But the human time and expertise required to provide manual summaries for legal resources, makes human summaries very expensive. This problem shows the interest in automated text summarization to process the ever increasing amount of documents.

In this research, within the framework of the project TAPoR at Universit de Montral, in collaboration with the law faculty and department of computer science, we have developed a summarization system, called LetSum, for producing short summaries for legal decision of the proceeding of a court. We have attempted to reproduce the results of human expert reasoning by organizing and manipulating factual and heuristic knowledge.

We present our approach for summarizing the legal record of the proceedings of federal courts in Canada and presenting it as a table-style summary for the needs of lawyers and experts in the legal domain. The FLEXICON (Smith and Deedman, 1987), SALOMON (Moens et al., 1999) and SUM (Grover et al., 2003) projects and (Borges et al., 2003) attest the importance of the exploration of legal knowledge for sentence categorisation and summarisation. Our 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 (Farzindar et al., 2004).

In Canada, the Canadian Legal Information Institute (CANLII) gathers legislative and judicial texts in order to make a virtual library of Canadian law accessible for free on the Internet (www.canlii.org). The large volume of legal information in electronic form creates a need for the creation and production of powerful computational tools in order to extract relevant information in a condensed form.

The lawyers need to process previous legal decisions to find a solution to a legal problem not directly indicated in the law, they look for *precedents* of similar cases. Each decision contains the reasons which justify the solution for a legal problem. They 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 of the work in legal field is the complexity of the domain: specific terminology of the legal domain and legal interpretations of expressions produce many ambiguities. For example, the word *disposition* means nature, effort, mental attitude or property in general English but in legal terms it means the final part of a judgement indicating the nature of a decision: acceptance of an inquiry or dismissal. That is why we have con-

structed our conceptual dictionary containing 200 concepts of legal domain. In this project we collaborate with legal experts of CanLII.

Thematic segments give information which can be used to answer specific questions about the theme of the segment such as: what are the themes of a document? How is a theme used? How are the problem and the facts presented? How does a judge reason to reach a conclusion?

In this paper, we will describe how we deal with the problem of the exploration of structure of document and content selection, according to the themes of a judgement.

1.1 Legal text summarization

Our approach to produce the summary is based on the identification of the thematic structure to find the argumentative themes of the judgement. This approach is a result of our corpus analysis in which we compared model summaries written by humans with the texts of the original judgments. The textual units considered as important by the professional abstractors were aligned manually with one or more elements of the source text. We look for a match between the information considered important in the professional abstract and the information in the source documents.

We extract the relevant sentences for each theme and present them as a table-style summary. Showing the information considered important could help the user read and navigate easily between the summary and the source judgment. If a sentence seems more important for a user and more information is needed about this topic, the complete thematic segment containing the selected sentence could be presented.

The identification of these structures separates the key ideas from the details of a judgment and improves readability and coherency in the summary. Therefore, in the presentation of a final summary, we propose to preserve this organization of the structure of the text in order to build a table-style summary.

Our corpus contains 3500 judgments of the Federal Court of Canada, which are available in HTML on `www.canlii.org/ca/cas/fct`. For some of these decisions, their summaries written by professional legal abstractors are available. We analyzed manually 50 judgments in English as well as their human written summaries. These judgments were suggested by the lawyers of CanLII project, as representing the *standard* judgements with *ideal* summaries.

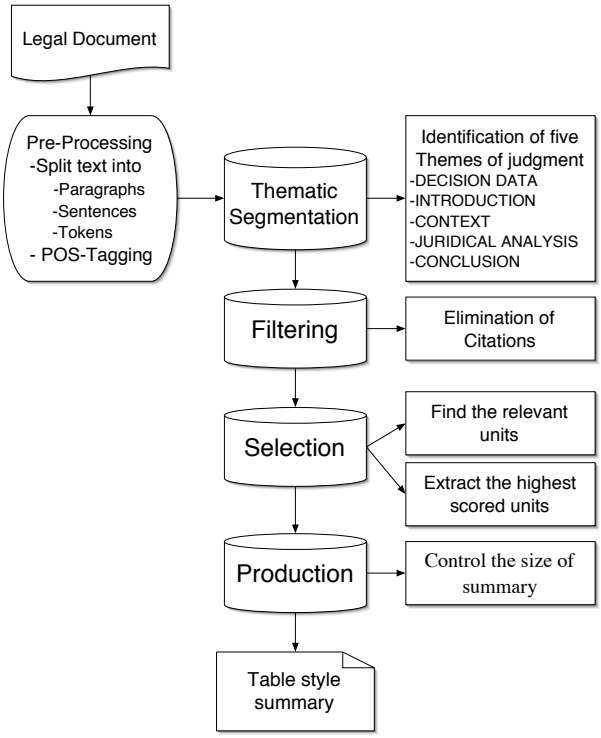

Figure 1: The procedural steps for generating of table-style summary

The rest of the corpus is used for statistical computations.

2 Components of LetSum

To process an input decision to LetSum, the system first does some pre-processing. The summary is built in four phases (Figure 1): thematic segmentation, filtering of less important units such as citations of law articles, selection of relevant textual units and production of the summary within the size limit of the abstract (Farzindar and Lapalme, 2004).

Pre-Processing splits the input judgment into main units. First the body of the text of the decision are identified. Some keywords like *Reasons for order, Reasons for judgment and order* separate the basic data (date, name of court, etc.), placed in the head of document, from the beginning of the judgment. The features used for the end of the decision are the date and place of hearing, name and signature of the judge. Then the document is divided into: section titles, paragraphs, sentences and tokens. To determine the Part-of-Speech tags, the tagger described by (Hepple, 2000) is used.

Thematic segmentation is based on the specific knowledge of the legal field. Accord-

ing to our analysis, the texts of jurisprudence have a thematic structure, independently of the category of judgment. Textual units dealing with the same subject form a thematic segment set. In this context, we distinguish four themes which divide the legal decisions into thematic segments, based on the experimental work of judge Mailhot (Mailhot, 1998):

INTRODUCTION describes the situation before the court and answers these questions: who? did what? to whom?

CONTEXT explains the facts in chronological order, or by description. It recomposes the story from the facts and events between the parties and findings of credibility on the disputed facts.

JURIDICAL ANALYSIS 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 expresses the *disposition* which is the final part of a decision containing the information about what is decided by the court.

For thematic segmentation the following information are used: the presence of significant section titles, the positions of a segment, the identification of direct or narrative style (as the transition of CONTEXT and JURIDICAL ANALYSIS segments), certain linguistic markers. We present the heuristics and some examples of linguistic markers for each theme:

INTRODUCTION presents a short description of a case with cue phrases such as: *application for judicial review, application to review a decision, motion filed by, Statement of Claim*.

This segment is at the beginning of the judgement after a title like: *Reasons for order, Reasons for judgment and order*.

Significant section titles are *Introduction* and *Summary*.

CONTEXT introduces the parties (eg. *The Applicant is a 52 year old citizen of Sri Lanka*).

It describes the application request using markers such as: *advise*, *indicate*, *concern*, *request*, etc. (eg. *Motion concerns Air Canada's failure to provide ground services in the French language at the Halifax airport*).

It explains the situation in the past tense and narration form.

Section titles are *Facts, Background, Factual background and Agreed statement of the facts*.

JURIDICAL ANALYSIS, in which the judge gives his explanation on the subject thus the style of expression is direct using *I*.

Cue phrases are *In reviewing the sections No. of the Act*, *Pursuant to section No.*, *As I have stated*, *In the present case*, *The case at bar is*.

Section titles are *Analysis* and *Decision of the court*.

CONCLUSION contains the final result of the court decision using phrases such as: *The motion is dismissed, the application must be granted*.

This segment is at the end of the judgement before the signature of the judge (judge's name, date, place of hearing, etc.).

Section titles are *Conclusion, Costs* and *Disposition*.

Filtering identifies parts of the text which can be eliminated, without losing relevant information for the summary. In a judgment, citation units (sentence or paragraph) occupy a large volume in the text, up to 30%, of the judgment, whereas their content is less important for the summary (according to our manual alignments between summaries and sources). This is why we remove citations inside blocks of thematic segments. We thus filter two categories of segments: submissions and arguments that report the points of view of the parties in the litigation and citations related for previous issues or references to applicable legislation. In the case of eliminating a citation of a legislation (eg. law articles), we save the reference of the citation in DECISION DATA in the field of authority and doctrine.

The identification of citations is based on two types of markers: direct and indirect. A direct marker is one of the linguistic indicators that we classified into three classes: verbs, concepts (noun, adverb, adjective) and complementary indications. Examples of verbs of citation are:

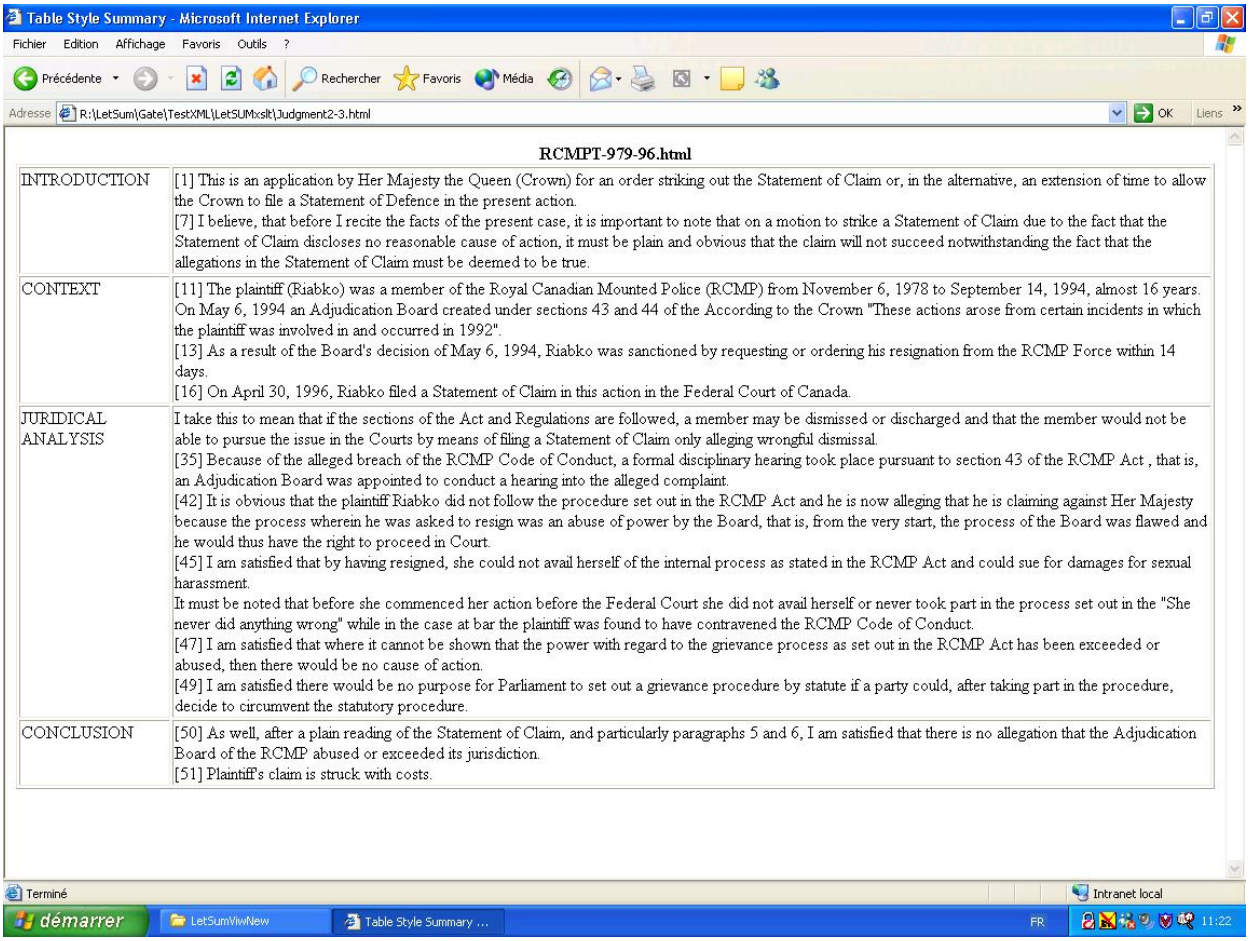

Figure 2: A table-style summary produced by LetSum, the original judgment has 3500 words and the summary is 15% of the source

conclude, define, indicate, provide, read, reference, refer, say, state, summarize. Examples of the concepts are: *following, section, subsection, page, paragraph, pursuant*. Complementary indications include numbers, certain prepositions, relative clauses and typographic marks (colon, quotation marks). The indirect citations are the neighboring units of a quoted phrase. For example, citation segment in the phrase *paragraph 78(1), which reads as follows:* is identified using direct markers but it points to the textual units with no direct marker which are also quotations. We thus identify the enumerated sentences following a quoted sentence for determining a group of citations.

Selection builds a list of the best candidate units for each structural level of the summary. LetSum computes a score for each sentence in the judgment based on heuristic functions related to the following information: position of the paragraphs in the document, position of the paragraphs in the thematic segment, position of the sentences in the paragraph, distribution of the words in document and corpus ($tf \cdot idf$). Depending on the given information in each layered segment, we have identified some cue words and linguistic markers. The thematic segment can change the value of linguistic indicators. For example, the phrase *application is dismissed* that can be considered as an important feature in the CONCLUSION might not have the same value in CONTEXT segment. At the end of this stage, the passages with the highest resulting scores are sorted to determine the most relevant ones.

Production of the final summary controls the size of the summary and displays the selected sentences in tabular format. The final summary is about 10% of the source document. The elimination of the unimportant sentences takes into account length statistics based on our observation from human abstracts. In the IN-

System ID	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4	ROUGE-L
LetSum	**0.57500**	**0.31381**	**0.20708**	**0.15036**	**0.45185**
Baseline	0.47244	0.27569	0.19391	0.14472	0.34683
Mead	0.45581	0.22314	0.14241	0.10064	0.32089
Word	0.44473	0.21295	0.13747	0.09727	0.29652
P. Mining	0.32833	0.15127	0.09798	0.07151	0.22375

Table 1: Result of statistical evaluation with ROUGE, LetSum is the first with the best evaluation scores

TRODUCTION segment, units with the highest score are kept within 10% of the size of the summary. In the CONTEXT segment, the selected units occupy 25% of the summary length. The contribution of the JURIDICAL ANALYSIS segment is 60% and the units with the role CONCLUSION occupy 5% of the summary.

Figure 2 shows an example of a table-style summary generated by LetSum. The summary is 15% of the source judgment.

3 Evaluating LetSum

We evaluate LetSum in two steps; first the evaluation of the modules of the system and second, a global evaluation of produced summaries. The evaluations of components of LetSum are very promising; we obtained 90% correct segmentation for thematic segmentation module and 97% correct detection for filtering stage (correct detection of 57 quoted segments over 60).

The final global evaluation includes the intrinsic and extrinsic tests (Spark-Jones and Galliers, 1995). Intrinsic evaluations test the system in of itself and extrinsic evaluation test the system in relation to some other task.

For intrinsic evaluations, we conducted a pilot study on 10 long judgements (average 8 pages). We compare the summaries produced by machine with the reference summaries written by professional abstractors. This evaluation is Recall-based, which measures how many of the reference summary sentences the machine summary contains. For measuring recall, we used the ROUGE software (Lin, 2004), which determines the quality of a summary by comparing it to ideal summaries created by humans. 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 machine summary and ideal summary. For measuring ROUGE-L, we view a summary sentence as a sequence of words. This evaluation computes the longest common subsequence of words to estimate the similarity between two summaries.

We compared LetSum with some other systems. A baseline system is a simple system which other system can be compared. Traditionally, for a newspaper article a baseline is the first paragraphs of the text. In our case, we defined a baseline with compression rate of 15% of the source document with the following algorithm:

- Choose 8% words of the beginning of the judgment. According our thematic segmentation, it takes the sentences from the themes INTRODUCTION and CONTEXT. If the last sentence is cut with this limit, complete it.
- Choose the last 4% words of the judgment with themes JURIDICAL ANALYSIS and CONCLUSION. If the first sentence is cut with this limit, complete it.

For automatic evaluation, we have compared the produced sentences by LetSum with: the baseline, the commercial automatic summaries produced by Microsoft Word and Pertinence Mining (www.pertinence.net) and a state-of-the-art summarization system Mead (Radev et al., 2003), with the human reference summaries. Table 1 shows the result of this evaluation. The higher score means better score and more performance system. LetSum is ranked first with the best evaluation scores.

The evaluation results show the interest of developing a summarization system for a specific domain because it is more and more difficult to generate a general summary without consideration of the user profile and the domain.

Our extrinsic evaluations will be based on legal expert judgement. We have defined a series of specific questions for the judgment (with the help a lawyer of CanLII), which cover the main topics of the document. If a user is able to an-

swer the questions correctly by only reading the summary, it means the summary contains all the necessary information of the source judgment. We are currently performing our human-based evaluations.

4 Conclusion

LetSum is a one the few systems developed specifically for summarization of legal documents. This system is implemented in an environment such as CanLII which has to deal with thousand of texts and produce summaries for each. We have presented our approach based on the extraction of relevant units in the source judgment by identifying the document's structure and determining the themes of the segments in the decision. The generation of the summary is done in four steps: thematic segmentation to detect the legal document structure in four themes INTRODUCTION, CONTEXT, JURIDICAL ANALYSIS and CONCLUSION, filtering to eliminate unimportant quotations and noises, selection of the candidate units and production of table-style summary. The presentation of the summary is in a tabular form along the themes of the judgment.

The evaluation of the system includes the intrinsic and extrinsic tests. The result of intrinsic evaluation of LetSum is very promising. We are completing the extrinsic evaluation based on legal expert judgments.

Acknowledgements

We would like to thanks LexUM group of legal information-processing laboratory of the Public Law Research Center at the University of Montreal for their valuable suggestions. This work has been developped using infrastructures funded by TAPoR. At Universit de Montral, TAPoR is funded by the Canadian Foundation for Innovation Project 5637 and the Qubec Ministry of Education. This project supported by Natural Sciences and Engineering Research Council of Canada.

References

Filipe Borges, , Raoul Borges, and Danièle Bourcier. 2003. Artificial neural networks and legal categorization. In *The 16th Annual Conference on Legal Knowledge and Information Systems (JURIX'03)*, page 187, The Netherlands, 11 and 12 December.

Atefeh Farzindar and Guy Lapalme. 2004. Legal text summarization by exploration of the thematic structures and argumentative roles. In *Text Summarization Branches Out Workshop held in conjunction with ACL'2004*, pages 27–34, Barcelona, Spain, 25–26 July.

Atefeh Farzindar, Guy Lapalme, and Jean-Pierre Desclés. 2004. Résumé de textes juridiques par identification de leur structure thématique. *Traitement Automatique des Langues (TAL), Numéro spécial sur: Le résumé automatique de texte : solutions et perspectives*, 45(1):26 pages.

Claire Grover, Ben Hachey, and Chris Korycinski. 2003. Summarising legal texts: Sentential tense and argumentative roles. In Dragomir Radev and Simone Teufel, editors, *HLT-NAACL 2003 Workshop: Text Summarization (DUC03)*, pages 33–40, Edmonton, Alberta, Canada, May 31 - June 1.

Mark Hepple. 2000. Independence and commitment: Assumptions for rapid training and execution of rule-based part-of-speech taggers. In *the 38th Annual Meeting of the Association for Computational Linguistics (ACL-2000)*, pages 278–285, October.

Chin Yew Lin. 2004. Rouge: A pakage for automatic evaluation of summaries. In *Text Summarization Branches Out Workshop held in conjunction with ACL'2004*, pages 74–81, Barcelona, Spain, 25–26 July.

Louise Mailhot. 1998. *Decisions, Decisions: a handbook for judicial writing.* Editions Yvon Blais, Québec, Canada.

Marie-Francine Moens, C. Uyttendaele, and J. Dumortier. 1999. Abstracting of legal cases: the potential of clustering based on the selection of representative objects. *Journal of the American Society for Information Science*, 50(2):151–161.

Dragomir Radev, Jahna Otterbacher, Hong Qi, and Daniel Tam. 2003. Mead reducs: Michigan at duc 2003. In *DUC03*, Edmonton, Alberta, Canada, May 31 - June 1. Association for Computational Linguistics.

J. C. Smith and Cal Deedman. 1987. The application of expert systems technology to case-based law. *ICAIL*, pages 84–93.

Karen Spark-Jones and Julia R. Galliers. 1995. *Evaluating Natural Language Processing Systems: An Analysis and Review.* Number 1083 in Lecture Notes in Artificial Intelligence. Springer.