Production automatique du résumé de textes juridiques: évaluation de qualité et d'acceptabilité

Atefeh Farzindar et Guy Lapalme
RALI
Département d'informatique et de recherche opérationnelle
Université de Montréal
C.P. 6128, succursale Centre-ville
Montréal, Québec, Canada H3C 3J7
{farzinda, lapalme}@iro.umontreal.ca

Mots-clefs : résumé automatique, fiches de résumé, textes juridiques, évaluation d'un résumé.

Keywords: automatic text summarization, summary table, legals texts, evaluation of a summary.

Résumé- Abstract

Nous décrivons un projet de production de résumé automatique de textes pour le domaine juridique pour lequel nous avons utilisé un corpus des jugements de la cour fédérale du Canada. Nous présentons notre système de résumé LetSum ainsi que l'évaluation des résumés produits. L'évaluation de 120 résumés par 12 avocats montre que la qualité des résumés produits par LetSum est comparable avec celle des résumés écrits par des humains.

We describe an automatic text summarisation project for the legal domain for which we use a corpus of judgments of the federal court of Canada. We present our summarization system, called LetSum and the evaluation of produced summaries. The evaluation of 120 summaries by 12 lawyers shows that the quality of the summaries produced by LetSum is approximately at the same level as the summaries written by humans.

1 Introduction

La jurisprudence est une référence importante pour les juristes. Pour cette raison les juristes consultent quotidiennement des milliers de documents juridiques. De jour en jour, la masse d'information textuelle sous forme de jurisprudence accessible sur internet ou dans les bases de données des entreprises et des gouvernements ne cesse d'augmenter. Ce qui nécessite le développement des outils spécifiques afin de pouvoir accéder au contenu des textes. Le but d'un résumé d'un jugement est d'abord de livrer l'essence du texte clairement et avec concision pour permettre une consultation facile et rapide; il doit fournir suffisamment d'informations sur le jugement pour permettre au lecteur de décider si celui-ci peut être pertinent à sa recherche. Actuellement, des jugements sont résumés manuellement par les professionnels ce qui est très coûteux.

Notre approche au résumé automatique a l'avantage de fournir des moyens clairs de concevoir des documents juridiques en fonction de résumés courts pour différents types d'utilisateurs: des étudiants, des avocats et des juges.

Le domaine juridique est un domaine ayant un grand besoin de résumés mais avec des exigences spécifiques. Dans ce projet, nous nous sommes intéressés au traitement des décisions des cours judiciaires du Canada. Nous avons collaboré avec les avocats du Centre de Recherche en Droit Public (CRDP), chargés de créer la bibliothèque de droit virtuelle des décisions judiciaires canadiens CanLII [1].

Dans cet article, nous décrivons plutôt les aspects qualitatifs de l'évaluation d'un résumé que la méthodologie de la production de résumé automatique. À la section 2, nous rappelons notre approche de production automatique de résumé de jugements et son l'implantation, LetSum (*Legal Text Summarizer*). La section 3 présente les évaluations effectuées avec LetSum. L'évaluation de 120 résumés automatiques par 12 avocats montre que la qualité des résumés produits par LetSum est excellente. La comparaison des résumés de LetSum avec cinq systèmes de recherche ou commerciaux montre l'intérêt d'utiliser d'un système de résumé spécialisé pour le domaine juridique.

2 Résumé de textes juridiques

Notre méthode a été développée suite à une analyse manuelle de 75 jugements et de leurs résumés rédigés par les résumeurs professionnels. Nous avons déjà présenté la problématique (Farzindar, 2004) et notre méthode pour capturer la structuration thématique des documents et identifier les unités textuelles saillantes (Farzindar *et al.*, 2004). Nous identifions d'abord le plan d'organisation d'un jugement et ses différents thèmes discursifs qui regroupent les phrases autour d'un même sujet. Chaque phrase dans un thème donne des informations complémentaires sur le sujet. Pour les phrases reliées à un thème, nous pouvons en interpréter le sens d'après leur contexte afin d'en extraire les idées clés.

[1] Canadian Legal Information Institute http://www.canlii.org

La création du résumé par LetSum se fait en quatre étapes décrites en détail dans (Farzindar, 2005).

Segmentation thématique qui détermine l'organisation du document original et relie les segments du texte associés avec des sept thèmes suivants:

- DONNÉES DE LA DÉCISION: donne la référence complète de la décision et la relation entre les parties sur le plan juridique.
- INTRODUCTION: qui? a fait quoi? à qui?
- CONTEXTE: recompose l'histoire du litige et l'histoire judiciaire.
- SOUMISSION: présente le point de vue d'une partie sur le problème.
- QUESTIONS DE DROIT: identifie le problème juridique dont le tribunal est saisi.
- RAISONNEMENT JURIDIQUE: décrit l'analyse du juge, la détermination des faits et l'expression des motifs de la solution retenue.
- CONCLUSION: présente la décision finale de la cour.

Selon nos observations, quatre thèmes jouent les rôles principaux: INTRODUCTION, CONTEXTE, RAISONNEMENT JURIDIQUE et CONCLUSION. La présence de ces quatre thèmes dans le jugement et dans le résumé est obligatoire. Dans la structure du résumé, nous préservons ces quatre thèmes et nous extrayons les phrases qui leur appartiennent. Le thème QUESTIONS DE DROIT est optionnel dans le jugement.

Filtrage qui identifie les segments qui peuvent être supprimés dans les documents, sans perdre les informations pertinentes pour le résumé. Dans un jugement, les citations occupent un volume important du texte soit 30% du jugement, alors que leur contenu est moins important pour le résumé. Nous identifions les citations principalement pour les supprimer en ne conservant que leurs références juridiques. En plus, le thème SOUMISSION contenant des discours des avocats, identifié par le segmenteur thématique, sera éliminé dans cette étape.

Sélection des unités textuelles candidates pour le résumé qui construit une liste d'unités saillantes pour chaque niveau structural du résumé en calculant les poids pour chaque phrase dans le jugement. La sélection est basée sur des règles sémantiques et des mesures statistiques.

Production du résumé qui choisit les unités pour le résumé final et les combine afin de produire un résumé représentant au maximum 15% du jugement. Le critère de sélection des unités est basé sur l'importance du segment thématique contenant les unités candidates.

La présentation du résumé final est sous forme d'une fiche de résumé contenant des rubriques homogènes d'informations. Cette fiche présente les informations considérées importantes associées à des thèmes précis, ce qui en facilite la lecture et la navigation entre le résumé et le jugement source. Pour chaque phrase du résumé produit, l'utilisateur peut en déterminer le sujet en regardant le thème associé à son segment thématique. La figure 1 montre un exemple de sortie de LetSum comme une fiche de résumé. Cette fiche de résumé montre les thèmes identifiés dans le jugement qui étaient pertinents pour le résumé. La taille du résumé est de 10% de celle du jugement original (le document source a dix pages).

Dans les prochaines sections, nous présentons l'évaluation des résumés générés par LetSum.

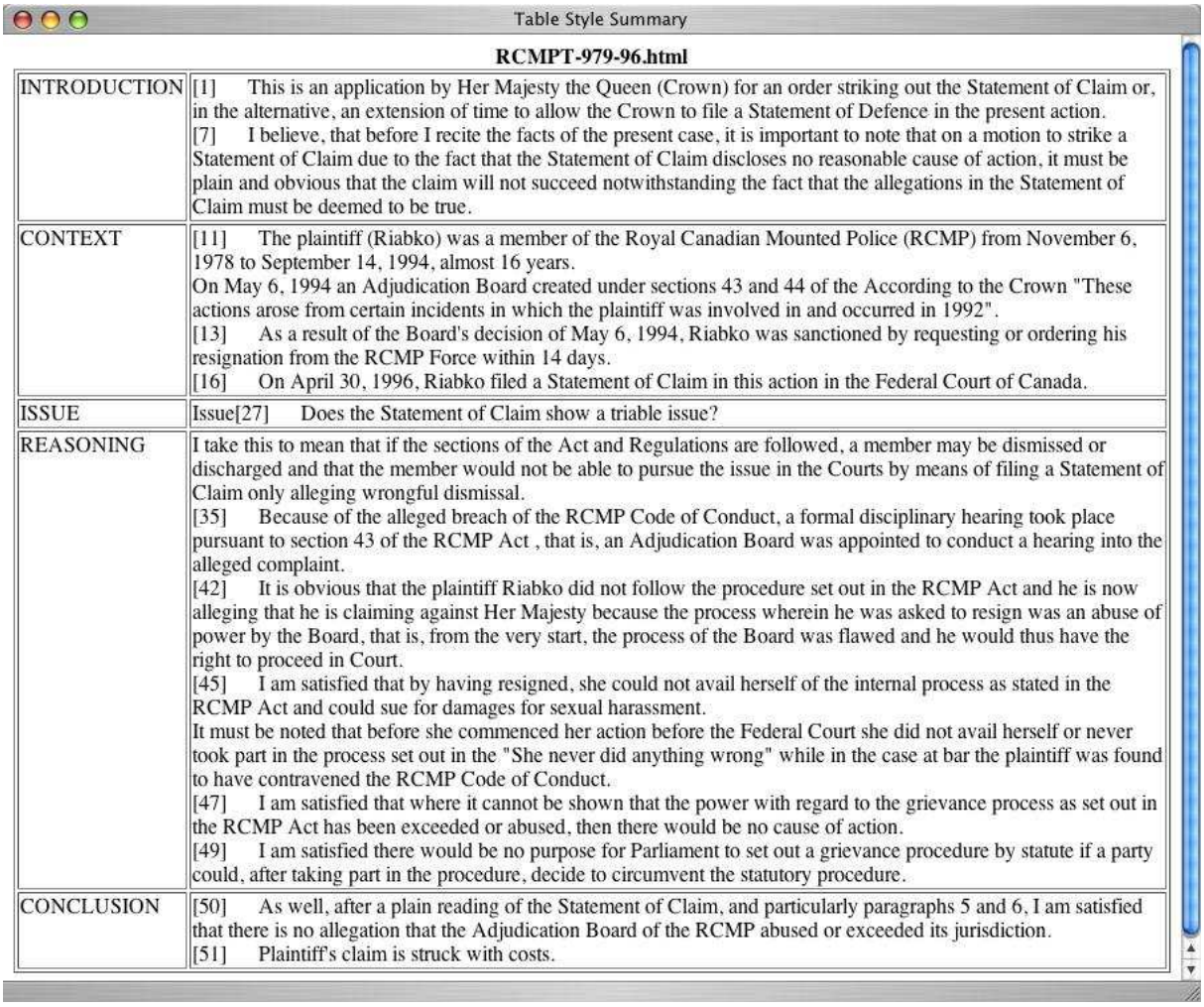

Figure 1: Fiche de résumé produit par LetSum, composé de 350 mots alors que le jugement source avait 4500 mots

3 Évaluation

La comparaison avec un résumé modèle comme référence pour des résumés automatiques est très naturelle, mais des résumés rédigés par des personnes différentes ne sont pas toujours convergents au niveau du contenu. La rédaction d'un résumé demande une analyse du texte pour en dégager les idées, les arguments, le style et les thèmes. Les rédacteurs humains dégagent les affirmations essentielles du document et les expriment dans leur propre style, ce qui donne lieu à plusieurs résumés pour le même document. Il est donc difficile de définir une métrique claire pour juger différents aspects d'un résumé comme la complétude, la thématique et la cohérence.

Plusieurs campagnes d'évaluation de systèmes de résumé comme SUMMAC[2] (Mani *et al.*, 1998) et DUC[3] (organisé par NIST) ont montré l'importance de définir des mesures pour l'évaluation d'un résumé. Spark Jones et Galliers (Spark-Jones & Galliers, 1995) ont proposé de diviser les évaluations en deux types: **intrinsèque** et **extrinsèque**. L'évaluation intrinsèque mesure les propriétés concernant la nature du sujet à évaluer et son objectif, alors que

[2]TIPSTER Text Summarization Evaluation Conference
[3]Document Understanding Conferences http://www-nlpir.nist.gov/projects/duc

l'évaluation extrinsèque mesure les aspects concernant les impacts et les effets de sa fonction. Nous avons évalué LetSum avec ces deux types d'évaluations.

Nos résumés du système ont été évalués en deux étapes: nous avons d'abord évalué les modules du système séparément, ensuite nous avons mesuré la qualité globale des résumés produits. Nous avons également comparé les résumés de LetSum avec des résumés produits par quatre autres systèmes et des résumés manuels.

Pour l'évaluation des modules de LetSum, nous avons utilisé une évaluation intrinsèque à trois niveaux: la qualité des divisions en thèmes par le segmenteur thématique, la détection correcte des citations par le module de filtrage, et le contenu des unités sélectionnées par le module de sélection et production.

Comme évaluation intrinsèque, nous avons utilisé ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*) (Lin & Hovy, 2003). ROUGE est maintenant bien reconnue comme mesure d'évaluation des résumés et a été utilisée pour la première fois dans la compétition de DUC 2004 comme seule mesure de fiabilité pour certaines tâches. ROUGE est basé sur le calcul statistique de co-occurrence de n-grammes. Cette méthode dont les résultats sont bien corrélés avec les jugements humains permet d'optimiser les systèmes et d'accélérer leur évaluation. ROUGE comporte deux méthodes d'évaluation. ROUGE-N, dont le score est basé sur le nombre de n-grammes (normalement $1 \leq n \leq 4$) communs entre le résumé automatique est le résumé modèle. Par exemple, ROUGE-2 calcule le nombre de paires de mots successifs communs entre les résumés candidat et modèle. La deuxième est ROUGE-L, qui considère les phrases comme une suite des séquences des mots. Cette évaluation calcule la plus longue sous-séquence commune des mots afin d'estimer la similarité entre deux résumés.

Pour l'évaluation extrinsèque de LetSum, nous avons demandé à des utilisateurs juristes de juger le contenu des résumés et leur acceptabilité. Pour chaque résumé, le recouvrement du contenu sur les idées clés du document a été évalué par deux avocats.

3.1 Évaluation des modules de LetSum

Nous avons évalué les quatre modules de LetSum séparément. Les deux premiers modules, **segmentation thématique** et **filtrage** sont évalués séparément, alors que les deux autres modules de sélection des unités pertinentes et production ont été évalués dans le cadre de l'évaluation des résumés finals de LetSum. Nous avons comparé les sorties de module de segmentation thématique avec le corpus que nous avons annoté manuellement (avec validation d'un avocat du CanLII).

Pour l'évaluation du module de **segmentation thématique**, nous avons utilisé un corpus de test contenant 10 jugements de la cour fédérale. Ces jugements n'ont pas été utilisés pour entraîner le système, ni servi à la construction du dictionnaire des marqueurs. Pour l'évaluation de ce module les points considérés importants sont: détection des thèmes, degré de pertinence d'un thème pour un segment, couverture des segments thématiques, précision des frontières entre deux thèmes. Pour cette évaluation on peut calculer la précision et le rappel. La précision mesure la proportion des unités pertinentes parmi toutes les unités produites par le système. Le rappel mesure la proportion des unités pertinentes parmi tous les unités pertinentes. F-mesure considère les deux mesures ensemble. Nous avons obtenu une précision de 100% et un rappel de 95% soit F-mesure 99% . Sur 40 thèmes annotés dans le corpus, 38 thèmes ont été identifiés correctement.

Pour l'évaluation du module de **filtrage** de citations, nous avons utilisé 15 jugements de la cour fédérale qui n'ont pas servi à entraîner le module de filtrage. Pour cette évaluation, nous avons comparé les unités de citations identifiées par le **filtrage** avec les citations annotées manuellement dans les jugements. Le résultat d'évaluation du module de filtrage est de 98% pour la précision et 95% pour le rappel ce qui donne 96% pour la F-mesure. Sur 60 cas de citation, 57 unités ont été identifiées correctement. Certaines citations n'étaient pas identifiées correctement à cause la langue de rédaction des références. Dans les jugements canadiens, les juges citent parfois les références de droit tels quels peu importe qu'elles soient en anglais ou en français. Pour les citations en français, lorsqu'il y a des marqueurs d'énumération, le système les identifie mais en absence des marqueurs d'énumération, il ne peut pas les distinguer.

3.2 Évaluation de LetSum par ROUGE

Pour le module de **sélection** et **production**, il faut mesurer les topiques extraits des documents par le système. Il est possible d'aligner automatiquement les unités de deux textes pour comparer la similarité entre les résumés modèles et les résumés produits afin de calculer la fraction du résumé modèle exprimée dans le contenu du résumé produit par le système. Pour cette évaluation des résumés de LetSum, nous avons utilisé ROUGE en les comparant avec des résumés modèles écrits par des humains. Nous avons généré 50 résumés automatiques avec cinq systèmes: système de recherche *MEAD* (Radev *et al.*, 2003), un système de recherche et commercial français *Pertinence Mining* (Lehmam, 1995), un système commercial de *Microsoft Word* (option de résumé dans MS Word) et une méthode *StartEnd* que nous avons définie. Le *StartEnd* est un système basé sur les positions des segments dans le document et LetSum afin de comparer notre système avec d'autres systèmes de résumés. Pour la méthode *StartEnd*, nous avons mis au point cette approche suite à nos analyses du corpus des résumés manuels. Pour définir le *StartEnd* nous avons fait trois expérimentations.

D'après nos études, le début du jugement situé à la fin des DONNÉES DE LA DÉCISION (nom de la cour, lieu de l'audience, date, les références et etc.) est une partie importante qui comprend le début du thème INTRODUCTION. Nous avons défini un baseline qui prend 15% du début du texte. Ce baseline couvre des thèmes INTRODUCTION et CONTEXTE.

Un autre baseline prend 15% de la fin du jugement avant la signature du juge. Ce baseline couvre les unités des thèmes RAISONNEMENT JURIDIQUE et CONCLUSION. Nos expériences avec ROUGE ont montré que le score de premier baseline était plus élevé que le deuxième, ce qui signifie l'importance du commencement du document par rapport à sa fin.

Cette expérience, nous a conduit à définir une approche de résumé avec un taux de compression 15%, basé sur l'algorithme suivant: prendre 8% du début du jugement et en complétant la dernière phrase si cette dernière a été coupée et prendre 4% de la fin du jugement en ajoutant la première phrase complète. Cette dernière approche, que nous avons nommée *StartEnd* est donc assez appropriée pour les documents de style juridique même si son implémentation est assez simple.

Nous avons comparé avec ROUGE les résumés de LetSum, *StartEnd* et ceux de trois autres systèmes avec les résumés humains. Les résultats de l'évaluation sont montrés à la table 1. Un score plus élevé est meilleur et indique un système plus performant. LetSum est classé au premier rang avec les meilleures notes d'évaluation. D'après cette évaluation, le deuxième système est *StartEnd*, ce qui montre l'importance de l'étude sur les documents des domaines

Système	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4	ROUGE-L
LetSum	**0.57500**	**0.31381**	**0.20708**	**0.15036**	**0.45185**
StartEnd	0.47244	0.27569	0.19391	0.14472	0.34683
MEAD	0.45581	0.22314	0.14241	0.10064	0.32089
MsWord	0.44473	0.21295	0.13747	0.09727	0.29652
Per. Mining	0.32833	0.15127	0.09798	0.07151	0.22375

Table 1: Résultat d'évaluation intrinsèque avec ROUGE, LetSum a des meilleurs résultats

spécifiques. Le fait qu'une approche simple puisse dépasser des méthodes complexes de production de résumé met en évidence la différence entre des organisations des documents et elle montre aussi l'intérêt de développer un système spécifique pour un domaine. Il est plus en plus difficile de produire un résumé général pour tous types d'utilisateurs sans prise en compte du besoin des usagers et de la tâche demandée.

3.3 Évaluation extrinsèque de LetSum

L'objectif de cette évaluation est de mesurer l'utilité du résumé automatique par rapport à un résumé écrit par un humain et de comparer la qualité des résumés automatiques générés par différents systèmes. Ce test est basé sur un jugement humain. Cette évaluation est toutefois très coûteuse, parce qu'elle demande des ressources humaines et un temps considérable.

Pour cette évaluation, nous avons utilisé les résumés automatiques produits par cinq systèmes présentés à la section précédente et les résumés écrits par des humains. Il faut noter que dans cette évaluation, nous n'avons considéré que les textes du résumé. Nous n'avons pas généré le format tabulaire d'organisation du résumé comme celui qui est présenté à la figure 1. Nous voulions aussi normaliser l'apparence de la sortie de tous les systèmes pour ne pas influencer les juges. Ce choix pénalise toutefois LetSum car nous ne tenons pas compte de la structure thématique extraite par notre méthodologie. Les évaluateurs ne savaient pas quels résumés avaient été produits par ordinateur et lesquels avaient été écrits manuellement.

Nous avons fait évaluer 120 résumés par les juristes. Le corpus de test contient dix jugements choisis au hasard dans différentes collections de jugements de la Cour fédérale du Canada. Nous avons généré 50 résumés automatiques et nous avons collecté 10 résumés manuels écrits par les arrêtistes de la Cour fédérale. Pour chaque résumé, nous avons répété le test deux fois, ceci nous donne deux avis par résumé. Chacun des 12 avocats du CanLII a évalué 10 résumés sur une période d'une heure sur deux aspects: contenu et qualité.

Pour l'évaluation du **contenu**, nous avons défini sept points importants à retrouver dans un jugement. Si un lecteur peut déterminer les points en question en lisant le résumé, on en déduit que le résumé contient suffisamment d'informations pour couvrir les idées clés d'un jugement. Sept questions (présentées en haut de la table 2) ont été déterminées avec l'aide d'un avocat de CanLII. L'ensemble des réponses de ces questions montre le degré de couverture sur des idées clés du jugement source exprimées dans le résumé. La deuxième partie de l'évaluation portait sur la **qualité** d'un résumé selon trois critères:

Lisibilité : La facilité de distinction et de perception du contenu du résumé qui en facilite la compréhension. Ce critère donne une appréciation globale du résumé. On demande si le

Après avoir lu le résumé peut-on déterminer:
Q1. Qui sont les parties en litige?
Q2. Quel est le problème en litige?
Q3. Les questions de droit soulevées?
Q4. Comment le juge a appliqué le droit aux faits?
Q5. Les motifs couvrent-ils les questions de droit?
Q6. Le résumé contient-il les motifs déterminants pour arriver à la conclusion?
Q7. Le résultat final de la cour?

Résumé	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Moyenne
Humain	55,00	90,00	90,00	70,00	**80,00**	**85,00**	**95,00**	80,71
LetSum	50,00	90,00	80,00	**75,00**	75,00	**85,00**	85,00	77,14
StartEnd	**65,00**	**100,00**	**100,00**	70,00	**80,00**	70,00	90,00	**82,14**
MEAD	55,00	**100,00**	95,00	65,00	50,00	50,00	40,00	65,00
MsWord	30,00	80,00	85,00	60,00	45,00	45,00	60,00	57,86
Per. Mining	25,00	65,00	55,00	35,00	35,00	35,00	45,00	42,14
Moyenne	46,67	87,50	84,17	62,50	60,83	61,67	69,17	67,50

Table 2: Questions juridiques utilisées lors de l'évaluation du contenu du résumé par les juristes et les résultats d'évaluation extrinsèque, les pourcentages des réponses positives pour les sept questions juridiques

résumé est: clair, assez clair, peu clair ou incompréhensible.

Cohérence : La présence simultanée d'éléments qui correspondent au même contenu ou qui s'accordent entre eux, qui s'harmonisent. Ce critère contient le fil conducteur du texte pour en assurer la continuité et la progression de l'information. On demande si la cohérence du texte dans le résumé est: très bonne, bonne, médiocre ou très mauvaise.

Pertinence des phrases : Caractère de ce qui est plus ou moins approprié, qui s'inscrit dans la ligne de l'objectif poursuivi. La pertinence des phrases mesure si les phrases du résumé contiennent un lien clair et direct avec le sujet dont il est question. On demande si le résumé est: très pertinent, assez pertinent, peu pertinent ou non pertinent.

L'évaluation comporte aussi une valeur d'acceptabilité sur la qualité générale du résumé. Nous avons demandé d'attribuer une valeur d'acceptabilité entre 0 et 5 pour chaque résumé (0 pour un résumé inacceptable et 5 pour un texte acceptable) sur la qualité du texte de résumé. Les résumés avec valeur 3 jusqu'à 5 sont considérés acceptables.

Dans la table 2, nous présentons les résultats obtenus pour l'évaluation des 120 jugements où, pour chaque question, nous avons calculé le pourcentage de réponses positives données à cette question. Une réponse positive signifie que le résumé contient assez d'informations sur le point en question. Par exemple dans la deuxième colonne, les résumés produits par le moyenne de toutes les méthodes ont couvert les informations sur la présentation des parties en litige (Q1) dans 47% des cas. LetSum a donc très bien répondu aux exigences des avocats pour des résumés automatiques. Ses résultats sont très proches de ceux des résumés manuels et sa performance est supérieure à celle des autres systèmes commerciaux *Microsoft Word* et *Pertinence Mining*, y compris le système de recherche *MEAD*.

Notre méthode *StartEnd*, basée sur la position des segments, a également donné de bons résul-

Résumé	Lisibilité 0-3	Cohérence 0-3	Pertinence 0-3	Acceptabilité 0-5
Humain	2,00	1,95	2,15	3,43
LetSum	2,30	**2,30**	**2,25**	3,43
StartEnd	**2,40**	2,20	2,10	**3,68**
MEAD	2,15	1,90	2,05	3,23
MsWord	1,65	1,25	1,60	2,63
Per. Mining	1,40	1,10	1,40	2,23
Moyenne	1,98	1,78	1,93	3,10

Table 3: Résultats d'évaluation extrinsèque selon les valeurs qualitatives entre 0 et 3 sur lisibilité, cohérence et pertinence des phrases, valeur d'acceptabilité est entre 0 et 5 sur la qualité générale du résumé

tats. Notre heuristique pour les positions des segments était appropriée, même si elle diffère du baseline utilisé normalement pour les articles journaux. Par le comportement du système *MEAD*, spécialisé pour les articles des journaux, on peut voir que les questions qui possèdent les réponses placées au début du document sont bien répondues alors que le recouvrement des informations clés sur les questions avec réponses dans d'autres positions dans le texte n'est pas satisfaisant. Les systèmes commerciaux comme *Microsoft Word* et *Pertinence Mining* ont les scores les plus faibles dans l'évaluation, car ils produisent des résumés génériques qui ne satisfont pas vraiment les utilisateurs dans un domaine spécifique comme droit.

La table 3 montre les résultats de l'évaluation de la qualité du résumé. Pour les trois critères, lisibilité, cohérence et pertinence des phrases du résumé, les valeurs sont entre 0 et 3. La qualité des résumés produits par LetSum est supérieure à celle des autres méthodes. La lisibilité du résumé de LetSum est jugé clair, la cohérence est évaluée très bonne et les pertinences des phrases sont mesurées très pertinentes pour les besoins des avocats. Les condensés rédigés par un humain sont jugés bons en cohérence (et non pas très bons) parce qu'ils sont en style télégraphique alors que LetSum et les autres systèmes font l'extraction de phrases.

Au point de vue d'acceptabilité du résumé, les résumés de LetSum sont jugés de niveau équivalent à celui des résumés écrits par les arrêtistes des cours. Encore une fois la méthode de positions des phrases dans le jugement des très bons scores pour ce critère d'évaluation. Il faut noter que dans cette partie de l'évaluation il y a peu de différence entre le système StartEnd et LetSum, un système nettement plus élaboré. Ceci peut en partie s'expliquer par le fait que nous n'avons pas considéré le format tabulaire produit par LetSum basé sur l'analyse thématique du texte qui distingue notre méthode.

4 Conclusion

Le domaine juridique est un vaste domaine avec un grand besoin pour le résumé automatique. Au Canada, il y a 30 000 avocats et aux États-Unis plus de 300 000 avocats susceptibles de rechercher de la jurisprudence. Toutes les synthèses de jurisprudence se font manuellement par des juristes. Lors qu'un résumé est disponible, le juriste a une idée du contenu de la décision et il lui est plus facile de savoir si elle a un potentiel de pertinence. Chaque résumé peut sauver, dans la consultation d'une liste de résultats de recherche, deux minutes à la personne qui fait la recherche. Une recherche typique dans CanLII donne plus de trente résultats, on pourrait donc

sauver une heure environ. Comme un avocat-recherchiste facture au moins 100$ de l'heure à son client, et que plusieurs recherches peuvent être requises pour un seul dossier, pour 20 recherches, il y aura donc 20 heures d'économies, 2 000$ sur un seul cas. L'utilisation des résumés automatiques économise du temps, des coûts et des expertises. Ces économies de ressources protègent les intérêts du gouvernement et de la population en tant que des clients attendant de recevoir un service juridique.

Nous avons développé LetSum, le premier système complet pour le résumé de textes juridiques en anglais. Il est basé sur l'identification de la structure thématique et présente le résumé sous forme d'une fiche de résumé augmentant ainsi la cohérence et la lisibilité du résumé. Dans les différentes étapes de notre étude, nous avons cherché à maximiser la précision de notre analyse en vue de diminuer les erreurs, car les textes de lois sont très précieux. L'excellente évaluation de LetSum est le témoin de la validité de notre approche. En faisant ressortir les points essentiels des jugements, nous espérons avoir rendu la justice plus accessible à tous et aussi aider la société.

Remerciements

Nous tenons à remercier l'équipe LexUM du laboratoire d'informatique juridique du Centre de recherche en droit public de la faculté de droit de l'Université de Montréal pour leur collaboration. La recherche présentée ici est soutenu financièrement par le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG).

Références

FARZINDAR A. (2004). Développement d'un système de résumé automatique de textes juridiques. In *TALN-RECITAL'2004*, p. 39–44, Fès, Maroc.

FARZINDAR A. (2005). *Résumé automatique de textes juridiques.* PhD thesis, Université de Montréal et Université de Paris4-Sorbonne.

FARZINDAR A., LAPALME G. & DESCLÉS J.-P. (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), 39–65.

LEHMAM A. (1995). *Le résumé automatique à fragments indicateurs: RAFI.* PhD thesis, Université de Nancy-II, Nancy, France.

LIN C.-Y. & HOVY E. (2003). Automatic evaluation of summaries using n-gram co-occurrence statistics. In *Proceedings of 2003 Language Technology Conference (HLT-NAACL 2003)*, p. 150–157, Edmonton, Canada.

MANI I., HOUSE D., KLEIN G., HIRSHMAN L., ORBST L., FIRMIN T., CHRZANOWSKI M. & SUNDHEIM B. (1998). *The TIPSTER SUMMAC Text Summarization Evaluation.* Rapport interne MTR 98W0000138, The Mitre Corporation, McLean, Virginia.

RADEV D., OTTERBACHER J., QI H. & TAM D. (2003). Mead reducs: Michigan at duc 2003. In *DUC03*, p. 160–167, Edmonton, Alberta, Canada: Association for Computational Linguistics.

SPARK-JONES K. & GALLIERS J. R. (1995). *Evaluating Natural Language Processing Systems: An Analysis and Review.* Number 1083 in Lecture Notes in Artificial Intelligence. Springer.