

Intégration d'une étape de pré-filtrage et d'une fonction multi-objectif en vue d'améliorer le système ExtraNews de résumé de documents multiples

Fatma Kallel Jaoua (1), Lamia Hadrich Belguith, Maher Jaoua (2)
Abdelmajid Ben Hamadou (3)

(1) Laboratoire LARIS-MIRACL, ISG Université de Gabès, Tunisie.
fatma_fseg@yahoo.fr

(2) Laboratoire LARIS-MIRACL, FSEGS, Université de Sfax, Tunisie.
l.belguith@fsegs.rnu.tn, maher.jaoua@fsegs.rnu.tn

(3) Laboratoire LARIM-MIRACL – ISIMS, Université de Sfax, Tunisie.
abdelmajid.benhamadou@fsegs.rnu.tn

Résumé Dans cet article, nous présentons les améliorations que nous avons apportées au système ExtraNews de résumé automatique de documents multiples. Ce système se base sur l'utilisation d'un algorithme génétique qui permet de combiner les phrases des documents sources pour former les extraits, qui seront croisés et mutés pour générer de nouveaux extraits. La multiplicité des critères de sélection d'extraits nous a inspiré une première amélioration qui consiste à utiliser une technique d'optimisation multi-objectif en vue d'évaluer ces extraits. La deuxième amélioration consiste à intégrer une étape de pré-filtrage de phrases qui a pour objectif la réduction du nombre des phrases des textes sources en entrée. Une évaluation des améliorations apportées à notre système est réalisée sur les corpus de DUC'04 et DUC'07.

Abstract In this paper, we present the improvements that we brought to the ExtraNews system dedicated for automatic summarisation of multiple documents. This system is based on the use of a genetic algorithm that combines sentences of the source documents to form the extracts. These extracts are crossed and transferred to generate new ones. The multiplicity of the extract selection criteria inspired us the first improvement that consists in the use of a multi-objectif optimization technique to evaluate these extracts. The second improvement consists of the integration of a sentence pre-filtering step which is based on the notion of dominance between sentences. Our objective is to reduce the sentence number of the source texts. An evaluation of the proposed improvements to our system is realized on DUC' 04 and DUC' 07 corpus.

Mots-clés : Résumé automatique, pré-filtrage de phrases, optimisation multi-objectif, algorithme génétique.

Keywords: Automatic summarization, sentences pre-filtering, multi-objective optimization, genetic algorithm.

1 Introduction

La montée en volume des documents électroniques disponibles en ligne a suscité la production d'outils informatiques dont la tâche principale est de trouver et d'extraire l'information pertinente. Dans ce contexte, les systèmes de résumé automatique de documents semblent être une bonne solution puisqu'ils permettent de reproduire automatiquement une représentation courte d'une collection de documents sources tout en conservant l'information pertinente. Ces systèmes visent, essentiellement, à faciliter la recherche et l'assimilation des informations textuelles pertinentes contenues dans les documents sources.

C'est dans ce cadre que nous avons proposé, dans un travail précédent (Jaoua et al., 2004), une méthode originale de résumé automatique de documents multiples décrivant un même thème. Cette méthode se distingue par le fait qu'elle considère l'extrait en tant qu'unité minimale d'extraction. Elle opère par constitution d'un ensemble d'extraits qui sont, ensuite, évalués et classés en vue de déterminer le meilleur en tenant compte de certains critères statistiques et linguistiques. Cette méthode a été implémentée dans le cadre du système ExtraNews de résumé automatique de documents multiples. Elle a été marquée par l'utilisation d'un algorithme génétique qui assure les mécanismes de génération et de classement d'extraits.

La version du système ExtraNews traitant la langue anglaise a été évaluée dans le cadre de la conférence d'évaluation DUC'04 et a obtenu des résultats intéressants. En effet, il a été classé premier dans la tâche 4 (de résumé de documents multiples traduits de la langue arabe à l'anglais) et troisième dans la tâche 5 (de résumé de biographie humaine) (Jaoua et al., 2004). Toutefois, l'évaluation de ExtraNews dans la conférence DUC'07 n'a pas obtenu les résultats escomptés (i.e. il a été classé seizième dans la tâche principale de résumé guidé par une question utilisateur). Malgré la différence des tâches et des mesures d'évaluations employées dans ces deux sessions d'évaluation, la divergence des résultats met en cause le choix des étapes de mise en œuvre de notre méthode. L'analyse des résultats générés nous a permis de dégager deux maillons faibles dans la chaîne de production d'extraits employée par notre méthode. Le premier maillon se situe au niveau de la fonction utilisée pour l'agrégation des critères de classement des extraits. Le deuxième maillon concerne l'apparition d'une dérive génétique qui s'est traduite par la convergence prématurée de notre algorithme génétique vers des solutions non pertinentes et donc de générer des extraits de qualité relativement "faible". Ce phénomène de dérive a été ressenti essentiellement, dans le cas où la taille de la collection des documents sources est importante. En effet, le nombre de phrases dans une collection de documents est passé de 275 phrases lors de DUC'04 à 720 phrases lors de DUC'07.

Dans cet article, nous allons présenter les améliorations que nous avons apportées à notre système afin de remédier aux problèmes précités. Nous allons d'abord expérimenter une nouvelle technique permettant d'intégrer les critères de classement des extraits. L'idée de cette expérimentation est de vérifier si un classement multi-objectif est plus approprié que la technique initiale utilisée dans le système ExtraNews qui consiste à classer les extraits moyennant une fonction d'agrégation des critères utilisés. La deuxième expérimentation vise à résoudre la dérive génétique qui a été constatée lorsque le système traite des collections de documents de grande taille. Dans ce cadre, nous allons étudier l'utilité d'intégrer une étape de pré-filtrage de phrases qui a pour objectif la réduction de l'espace de recherche exploré par l'algorithme génétique utilisé dans notre méthode.

Le présent article s'articule autour de cinq sections. Cette partie introductive est suivie par la deuxième section qui présente un survol des principaux travaux réalisés dans le domaine du résumé automatique de documents multiples. La troisième et la quatrième section détaillent le principe de base de la méthode proposée ainsi que les améliorations apportées à cette méthode. La cinquième section présente l'évaluation menée avant et après l'intégration de ces améliorations dans le système ExtraNews. Cette évaluation est réalisée sur les corpus utilisés lors de DUC'04 et DUC'07.

2 Etat de l'art

La plupart des travaux réalisés dans le domaine du résumé automatique de documents multiples s'articulent autour du processus de regroupement et de classement des unités textuelles similaires en vue de dégager les unités les plus importantes (Spark Jones, 2007). Ces travaux peuvent utiliser plusieurs types de méthodes : les méthodes statistiques, les méthodes linguistiques et les méthodes par compréhension.

Dans les méthodes statistiques, l'importance des unités textuelles (généralement les phrases) est représentée par un poids ou une probabilité qui leur est assignée en fonction de leur richesse en mots clés, du groupement de termes, de leurs positions dans le document (Lin et al., 2002). Dans d'autres méthodes, cette importance dépend de la diversité que l'unité textuelle apporte à l'extrait généré, tout en gardant un niveau élevé de similarité avec la requête utilisée (Carbonell et al., 1998). D'autres méthodes exploitent la distribution des mots dans les documents sources pour sélectionner les unités qui regroupent les mots jugés pertinents (Mori et al., 2004).

Les méthodes linguistiques exploitent les connaissances sur la langue pour guider le processus d'extraction. Certaines méthodes mènent une analyse linguistique dans le but de regrouper les phrases similaires (McKeown et al., 1999). D'autres travaux exploitent les relations de coréférence (Saggion et al., 2004) ou des relations entreprises par les entités nommées (Fuentes et al., 2003) pour éliminer les prépositions redondantes.

Les méthodes par compréhension opèrent par la construction d'une représentation interne des textes sources en vue de déterminer les composantes importantes qui sont ensuite réduites puis reformulées pour former le résumé final. Cette représentation peut être illustrée par un graphe conceptuel (Mani et al., 1997) ou par des patrons sémantiques (Harabagiu et al., 2005). D'autres méthodes se basent sur l'utilisation de modèles prédéfinis de résumés qui sont instanciés à partir d'informations extraites des documents sources (Radev et al., 1998), (White et al., 2001). Il est à noter que ces méthodes exigent des connaissances préalables du domaine à traiter pour générer des résumés de bonne qualité.

La plupart des méthodes précitées utilisent une phase de regroupement d'unités textuelles similaires pour éviter la sélection d'unités redondantes dans le résumé final. Cette étape est indispensable vu que la granularité d'importance utilisée se limite, généralement, à la phrase. Dans la méthode que nous avons proposée, nous considérons l'extrait en tant que granularité d'importance ; ce qui permet d'éviter la phase de regroupement et d'offrir un moyen pour traiter l'extrait en tant qu'unité à part entière et non pas en tant qu'un ensemble d'unités indépendantes.

3 Méthode d'extraction utilisée dans ExtraNews

Afin d'appréhender le problème d'extraction des phrases pertinentes dans un document, nous avons proposé une nouvelle unité d'extraction qui opère à un niveau plus étendu que la phrase (Jaoua et al., 2003). Cette conviction est guidée par l'idée que pour un niveau englobant la phrase, nous pouvons mieux contrôler les problèmes résultant de la sélection des phrases indépendamment les unes des autres. La granularité d'extraction que nous avons choisie est l'extrait qui est formé à partir de phrases des documents sources (Jaoua et al., 2003). Ainsi le problème d'extraction est abordé comme étant un problème d'optimisation où il s'agit d'effectuer une comparaison entre plusieurs extraits en vue de sélectionner le "meilleur". Il s'agit donc de choisir à partir des textes sources la meilleure partition formée par les phrases issues des documents sources, et qui maximisent certains critères liés à la qualité et à la quantité des informations véhiculées.

Toutefois, la détermination de l'ensemble des partitions d'un document est un problème NP (non polynomial) qui ne peut pas être résolu en un temps raisonnable (Brucker et al., 1978). Afin de résoudre ce problème, les méthodes d'optimisation opèrent par évaluation de solutions intermédiaires en vue de converger vers une solution proche de l'optimale. L'application de ces méthodes pour le problème d'extraction suppose que toutes les solutions intermédiaires représentent des extraits générés en une première étape, puis évalués en fonction des critères utilisés en une deuxième étape.

La méthode d'optimisation que nous avons utilisée dans le contexte d'extraction des phrases s'appuie sur les algorithmes génétiques. Le choix des algorithmes génétiques est essentiellement motivé par l'étendue de l'espace de recherche exploré lors de leur application. En effet, ces algorithmes offrent l'avantage de manipuler plusieurs solutions en même temps, ce qui permet d'explorer un grand espace de recherche. Notons que les algorithmes génétiques se basent sur le principe de génération aléatoire d'une population de génomes qui sera classée en fonction d'une valeur d'adaptation (Goldberg, 1989). Dans notre cas, le génome constitue l'extrait alors que la phrase représente un gène de ce génome. Les meilleurs génomes (extraits) de cette population seront croisés et mutés en vue de générer une nouvelle population qui sera ensuite classée pour retrouver ses meilleurs extraits. Ce processus est réitéré jusqu'à la non-amélioration (stagnation) de la valeur d'adaptation. A chaque itération, il s'agit de classer les extraits moyennant des critères liés à certains aspects statistiques et linguistiques de l'extrait.

Dans notre système ExtraNews, l'évaluation des extraits se base sur cinq critères :

- La taille de l'extrait (C_1) : ce critère permet de fixer la longueur de l'extrait en mots. La valeur de ce critère est égale à 1 si la taille de l'extrait est voisine de la taille désirée. Cette valeur est inférieure à 1 si la taille de l'extrait est inférieure à celle fixée et elle est égale à 0 en cas de dépassement. La pénalisation des extraits longs est due à la troncature, effectuée lors de l'évaluation des extraits dépassant la taille fixée.
- La couverture en mots clés simples (C_2) : un mot clé est un mot non vide dit plein (obtenu après élimination des mots vides tels que les articles, les prépositions, etc.) dont la fréquence est importante. Ce critère est calculé en divisant le nombre de mots clés contenus dans l'extrait par le nombre total des mots clés contenus dans les textes

sources. Il est à noter que les mots clés peuvent apparaître sous forme de synonymes¹ et donc être considérés comme une seule entrée pour le critère de couverture.

- La couverture en mots clés doubles (C_3) : un mot clé double correspond à deux mots adjacents qui se répètent fréquemment dans les textes sources. Il est à noter que deux mots adjacents sont deux mots séparés par un espace ou par des mots vides. Ce critère est calculé de la même manière que le critère précédent et tient compte de la relation de synonymes entre mots.
- La couverture en mots clés de la question utilisateur (C_4) : ce critère a été introduit pour tenir compte de la question de l'utilisateur qui peut guider le système dans la génération du résumé. Le calcul s'effectue de la même manière utilisée dans les deux critères précédents. Tous les mots de la question (à l'exception des mots vides) sont considérés comme étant des mots clés et seront enrichis par leurs synonymes.
- Le poids de l'extrait (C_5) : ce critère correspond à la somme des valeurs $TF*IDF$ des termes contenus dans l'extrait. La valeur de ce critère est normalisée en divisant le poids de l'extrait par celui de l'extrait le plus pondérant de la population courante.

Pour classer les extraits de chaque population générée suite à l'application de l'algorithme génétique, nous avons choisi préalablement une fonction "objectif" qui permet d'agréger les critères précités. La fonction d'agrégation utilisée consiste à multiplier les valeurs associées à ces critères après être normalisées. Si l'extrait dépasse la taille fixée au préalable, sa valeur "objectif" est égale à zéro vu que le critère de longueur vaut zéro, ce qui permet d'éliminer les extraits dont la longueur excède celle désirée : $F(\text{extrait}) = \prod \tilde{\omega}_i$.

Il est à noter que $\tilde{\omega}_i$ désigne le coefficient associé au critère i après être normalisé. Par exemple, la normalisation du critère associé au poids consiste à diviser la valeur de ce critère par le poids maximal obtenu parmi les extraits générés dans la même population.

4 Améliorations apportées

4.1 Classification multi-objectif des extraits

A l'issue de l'évaluation de notre méthode dans les sessions DUC'04 et DUC'07, nous avons prélevé certaines critiques liées à la manière dont la fonction "objectif" agrège les critères précités. En effet, les critères peuvent ne pas avoir le même ordre de grandeur malgré la normalisation effectuée ; ce qui permet de donner à certains critères la possibilité d'augmenter (ou de diminuer) d'une manière significative la valeur de la fonction "objectif". Ce problème persiste même si on attribue des coefficients aux différents critères ; dans ce cas le pouvoir d'un critère reste lié à l'intervalle des valeurs prises par ce critère.

Afin de tenir compte de divers critères lors du classement des extraits, nous proposons d'utiliser un classement multi-objectif. L'optimisation multi-objectif se base sur la notion de dominance qui compare chacun des critères des solutions (Collette et al., 2002).

¹ Afin de chercher les synonymes des mots clés, nous avons utilisé le dictionnaire WordNet.

Formellement, on dit qu'une solution (extrait) X domine une solution (extrait) Y si pour tous les critères C_i , on a :

$$C_i(X) \geq C_i(Y) \text{ et } \exists C_j / C_j(X) > C_j(Y).$$

Avec $C_i(X)$ la valeur attribuée à la solution (extrait) X pour le critère C_i .

Les solutions (extraits) non dominées dans une population sont considérées comme des solutions potentielles, capables de générer des solutions intéressantes. Elles sont sélectionnées pour former la population initiale de l'itération suivante. La stagnation pendant plusieurs itérations des solutions obtenues renseigne sur la convergence de l'algorithme.

Dans le cas de classement des extraits, l'obtention d'un ensemble d'extraits dominants, après la convergence de l'algorithme, ne peut pas résoudre le problème, car le résultat escompté doit être un seul extrait. Pour pallier ce problème, nous avons essayé de classer les critères utilisés en comparant les extraits dominants obtenus avec des résumés de référence rédigés par des experts humains. Le classement des critères permet, dans le cas où il existe plusieurs extraits dominants, de retenir un seul extrait.

Nous avons exploité à cet effet le corpus diffusé lors de la conférence DUC'06, ainsi que la métrique Rouge₂ : Recall-Oriented Understudy for Gisting Evaluation (2_grammes) qui est une métrique de rappel utilisée pour l'évaluation des résumés (Lin, 2004). Cette métrique, qui présente une bonne corrélation avec des évaluations humaines, permet de comptabiliser l'apparition des mots doubles dans l'extrait (ou le résumé) système et les résumés de référence (voir section 5). Le corpus diffusé lors de la conférence DUC'06 est formé de 50 collections de documents. Chaque collection comprend 25 documents, alors que les résumés de référence sont au nombre de 4 pour chaque collection. La procédure de test consiste à déterminer, parmi les extraits dominants, celui qui présente le meilleur indice Rouge₂ et donc la meilleure correspondance avec les résumés de référence. Pour chaque critère nous comparons la valeur attribuée au "meilleur" extrait avec les valeurs attribuées au même critère pour le reste des extraits dominants. Cette comparaison nous a permis de déduire un pourcentage d'importance ainsi qu'un classement de chaque critère par rapport aux autres critères (voir tableau 1).

Critère	Taille de l'extrait (C ₁)	Couverture en mots simples (C ₂)	Couverture en mots doubles (C ₃)	Couverture en mots de la question utilisateur (C ₄)	Poids de l'extrait (C ₅)
Importance	95%	57%	27%	34%	25%
Classement	1	2	4	3	5

Tableau 1 : Classement des critères d'évaluation des extraits

En cas de présence de plusieurs solutions dominantes, la sélection se fait en se basant sur le classement des critères indiqué dans le tableau 1. Dans l'exemple illustré par le tableau 2, seuls les extraits E1, E2 et E4 sont considérés comme des extraits dominants. En effet, E1 possède le meilleur score selon le critère C₅, E2 possède le meilleur score selon les critères C₂ et C₃ et E4 est classé premier selon C₄. Le classement des critères présenté dans le tableau 1 est en faveur de l'extrait E2 qui sera choisi comme étant le meilleur extrait.

Extraits	C ₁	C ₂	C ₃	C ₄	C ₅
E1	0.98	0.41	0.30	0.31	0.93
E2	1	0.42	0.35	0.42	0.87
E3	1	0.37	0.29	0.40	0.86
E4	0.98	0.37	0.30	0.63	0.83
E5	0.95	0.32	0.27	0.24	0.80
E6	0.90	0.15	0.25	0.39	0.75

Tableau 2 : Exemple de classement Multi-objectif des extraits

4.2 Pré-filtrage des phrases

A l'issue de l'évaluation de notre méthode dans les sessions DUC'04 et DUC'07, nous avons détecté un autre problème lié à la taille des collections des documents sources en entrée. En effet, un phénomène de dérive génétique a été constaté lorsque la taille des documents sources est importante. Cette dérive a été signalée par la convergence de l'algorithme utilisé vers des solutions prématurées non pertinentes (i.e. des extraits de qualité relativement faible) vu que l'espace de recherche est très grand.

Afin de réduire l'espace de recherche, nous avons intégré une étape de pré-filtrage de phrases qui a pour objectif de supprimer les phrases redondantes et par suite, de minimiser le nombre de phrases en entrée (Jaoua et al., 2008b). Malgré la présence de critères liés à la couverture des mots clés et qui permet d'éliminer les extraits présentant de la redondance, l'étape de pré-filtrage s'avère indispensable afin de réduire l'espace de recherche et par suite d'éviter la dérive génétique.

L'étape de pré-filtrage utilise la notion de dominance entre phrases qui est une notion inspirée du domaine de l'ordonnancement multi-objectif. On dit qu'une phrase P domine une phrase Q si l'ensemble des mots clés de la phrase Q (ou de leurs synonymes) est inclus dans l'ensemble des mots clés de la phrase P et si la longueur de la phrase P (calculée en nombre de mots) est inférieure ou égale à celle de la phrase Q. Il est à noter la notion de dominance entre phrases utilisée à ce niveau n'implique que les deux critères de longueur et de couverture en mots clés simples car ce sont les deux critères les mieux classés au niveau de l'évaluation des extraits.

L'exemple des deux phrases suivantes extraites du corpus de DUC'07 illustre la dominance de la phrase Q par la phrase P sachant que ces deux phrases renferment les mêmes mots clés (les mots en italique) et que la longueur de P (33 mots) est inférieure à celle de Q (38 mots).

P (APW19991026.0010/D0721E): They saw pictures of the *ponderosa* pine fence where *Shepard* was left to *die*, his hands tied behind his back, and the pool of *blood* caused by *blows* as he fought his attackers.

Q (APW19991026.0133/D0721E): On Monday, jurors saw pictures of the *ponderosa* pine fence where *Shepard* was left to *die*, his hands tied behind his back, and the pool of *blood* caused by more than dozen *blows* as he fought his attackers.

L'étape de pré-filtrage consiste, donc, à comparer les phrases en terme de longueur et de mots clés simples afin de déterminer les phrases dominées qui seront supprimées du pool des phrases initiales utilisé pour générer des extraits. Il est à noter que la sélection des phrases dominantes permet, en outre, d'éliminer les phrases similaires ou synonymes. Il est à signaler que l'élimination des phrases synonymes dans l'extrait a été abordée comme critère de classification d'extraits dans les travaux de Liu (Liu et al, 2006).

5 Evaluation de la nouvelle méthode proposée

Plusieurs types d'évaluations ont été adoptées lors des conférences DUC pour quantifier les performances des systèmes de résumé automatique (Over et al., 2007). Parmi ces évaluations, nous citons l'évaluation Rouge² (Lin, 2004) qui fait intervenir la différence entre la distribution des mots (n_grammes) d'un résumé candidat et celle d'un ensemble de résumés de référence (résumés humains). La formule de calcul des mesures Rouge est la suivante :

$$Rouge_n = \frac{\sum_{C \in \{Référence\}} \sum_{n_gram \in C} correspond(candidat, c)}{\sum_{C \in \{Référence\}} \sum_{n_gram \in C}}$$

Où *correspond (candidat,c)* représente le nombre maximum de n_grammes communs entre le résumé système et le résumé de référence. Le dénominateur de l'équation représente la somme du nombre de n_grammes des résumés de références. Il est à noter que Rouge_n est la formule de base du score Rouge. On peut donc obtenir des mesures de Rouge1 (1_gramme), Rouge2 (2_grammes), etc. Des études de corrélation ont montré que la mesure Rouge2 présente la meilleure corrélation avec les résumés humains (Lin, 2004).

Nous avons évalué notre méthode de génération d'extraits avant et après l'intégration des différentes améliorations présentées dans la section 4. Dans ce qui suit, nous utilisons le terme initiale pour se référer à la méthode avant amélioration. La méthode initiale a été expérimentée dans les trois tâches qui s'intéressent aux résumés de documents multiples dans la session DUC'04 (tâche 2, 4, et 5). Il est à noter que dans cette version, le critère lié aux mots clés de la question utilisateur n'est pas pris en compte par notre méthode car les tâches de résumés expérimentées ne comportent pas de questions. Notre méthode a été en outre expérimentée dans DUC'07 et a participé à la tâche principale dédiée pour le résumé guidé par une question utilisateur.

Après avoir intégré les améliorations précitées, nous avons procédé à une nouvelle évaluation de notre méthode sur le même corpus utilisé lors des sessions DUC'04 et DUC'07. Les résultats obtenus montrent une baisse considérable du nombre de phrases en entrée de notre système (i.e. ce nombre est passé de 720 à 386 phrases en moyenne par collection de documents). Le tableau 3 reporte les résultats obtenus de l'indice Rouge₂ avant et après l'intégration de l'étape de pré-filtrage ainsi que de l'application du classement multi-objectif.

Les résultats du tableau 3 montrent l'intérêt de l'intégration de l'étape de pré-filtrage, et de l'application de la méthode d'optimisation multi-objectif. L'intégration de l'étape de pré-

² Rouge : Recall-Oriented Understudy for Gisting Evaluation <http://berouge.com>

Intégration d'une étape de pré-filtrage et d'une fonction multi-objectif en vue d'améliorer le système ExtraNews de résumé de documents multiples

filtrage a permis d'augmenter la valeur Rouge2 de 2.2 % en moyenne. L'utilisation de l'approche multi-objectif a aussi fait ses preuves et a permis d'améliorer la valeur de Rouge2 de 0.4 %.

Tâche	DUC'04 (identifiant : id=21, 23,24) :						DUC'07 (id=28)	
	Tâche 2		Tâche 4		Tâche 5		Principale	
	Rouge ₂	Rang	Rouge ₂	Rang	Rouge ₂	Rang	Rouge ₂	Rang
Initiale (rang)	0.121	4/14	0.132	1/11	0.118	3/14	0.098	16/32
Initiale + Mutli-objectif	0.122	4/14	0.132	1/11	0.119	3/14	0.103	14/32
Initiale + Pré-filtrage	0.123	3/14	0.132	1/11	0.121	3/14	0.118	4/32
Initiale + Pré-filtrage + Multi-Objectif	0.123	3/14	0.132	1/11	0.122	2/14	0.120	3/32

Tableau 3 : Evaluation selon la métrique Rouge₂ avant et après intégration des améliorations

6 Conclusion

Dans cet article, nous avons présenté les améliorations que nous avons apportées à notre méthode de résumé automatique de documents multiples utilisé dans le système ExtraNews. Les résultats obtenus lors de la session DUC'07 ont permis d'identifier l'importance de l'étape de pré-filtrage de phrases pour faire face à la dérive génétique qui a caractérisé notre méthode. En effet, cette étape a permis de réduire de moitié le nombre de phrases initiales issues des documents sources. Nous avons aussi mis en évidence l'importance d'utiliser une stratégie d'optimisation multi-objectif pour classer les extraits.

L'examen des résultats obtenus montre aussi que l'on peut améliorer davantage les performances de notre système ExtraNews à travers l'application de nouveaux critères de sélection d'extraits. Nous avons, en outre, envisagé l'amélioration de la qualité discursive de l'extrait produit par notre système à travers l'intégration d'une étape de révision en réorganisant les phrases de cet extrait (Jaoua et al., 2008a). Au niveau de cette étape, nous projetons aussi d'intégrer des mécanismes de compression et de fusion visant à améliorer la qualité linguistique de l'extrait final.

7 Références

- BRUCKER P. (1978). On the complexity of clustering problems in Optimization and Operations Research. *Lecture Notes in Economics and Mathematical Systems n° 157*. 45-54.
- CARBONELL J., GOLDSTEIN J. (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 335-336.
- COLLETTE Y., SIARRY P. (2002). *Optimisation multiobjectif*. Paris : Edition Eyrolles.
- FUENTES M., MASSOT M., RODRÍGUEZ H., ALONSO L. (2003). Headline extraction combining statistic and symbolic techniques, *Proceeding of DUC03 Document Understanding conferences Workshop*.

- GOLDBERG D.E. (1998). *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley, New York.
- HARABAGIU S., LACATUSU F. (2005). Topic themes for multi-document summarisation. In *Proceedings of the 28th annual international ACM-SIGIR conference on research and development in information retrieval (SIGIR 2005)*. 202–209.
- JAOUA K.F., BELGUITH H.L., BEN HAMADOU A. (2008a). Révision des Extraits de Documents Multiples Basée sur la Réorganisation des Phrases : Application à la Langue Arabe. Actes de la conférence *IBIMA 08 : Information Management in Modern Organizations*.
- JAOUA K.F., JAOUA M., BELGUITH H.L., BEN HAMADOU A. (2008b). Filtrage de texte pour le résumé automatique de documents multiples à paraître dans les actes du colloque *CORIA'08 : Conférence en Recherche d'Information et Applications*.
- JAOUA K.F., JAOUA M., BELGUITH H.L., BEN HAMADOU A. (2004). Summarization at LARIS laboratory. *Proceeding of the DUC'04 Document Understanding conferences Workshop*.
- JAOUA M., BEN HAMADOU A. (2003). Automatic Text Summarization of Scientific Articles Based on Classification of Extract's Population, *Proceeding of Cicing'03*. 623-634.
- LIN C.Y. (2004). Rouge: a package for automatic evaluation of summaries. *Proceedings of the ACL'04 Workshop*. 74-81.
- LIN C.Y., HOVY E. (2002). Automated multi-document summarization in NeATS. *Proceedings of the DARPA Human Language Technology Conference*. 50–53.
- LIU D., HE Y., JI D., YANG H. (2006). Genetic algorithm based multi-document summarization. *PRICAI'06*. 1140-1144.
- MANI I., BLOEDORN E. (1997). Multi-document summarization by graph search and matching. *Proceedings of the 14th National Conference on Artificial Intelligence*. 622-628.
- MCKEOWN K. KLAVANS J., HATZIVASSILOGLOU V., BARZILAY R., AND ESKIN E. (1999). Towards multidocument summarization by reformulation: progress and prospects. *Proceedings of the 16th National Conference on Artificial Intelligence*. 453-460.
- MORI T., NOZAWA M., ASADA Y. (2004). Multi-answer-focused multi-document summarization using a question-answering engine. *Proceedings of the 20th International Conference on Computational Linguistics (COLING 04)*. 439-445.
- OVER P., DANG H., HARMAN D. (2007). DUC in context. *Information Processing and Management*. 1506–1520.
- RADEV D., MCKEOWN K. (1998). Generating natural language summaries from multiple on-line sources. *Proceeding of the Computational Linguistics*. 469-500.
- SAGGION H., GAIZOUSKAS R.(2004). Multi-document summarization by cluster/profile relevance and redundancy removal. *Proceeding of the Document Understanding Workshop DUC'04*.
- SPARK JONES K. (2007). Automatic summarising: The state of the art. *Information Processing and Management* 43. 1449–1481.
- WHITE M., KORELSKY T., CARDIE C., PIERCE D., NG V., WAGSTAFF K. (2001). Multidocument summarization via information extraction. *Proceedings of the DARPA Human Language Technology Conference*. 143–146.