

Évaluation à DUC2002 d'une adaptation de SumUM, un système de résumé automatique

Atefeh Farzindar et Guy Lapalme
RALI/Université de Montréal
{Farzinda,Lapalme}@iro.umontreal.ca

Horacio Saggion
NLP/University of Sheffield
H.Saggion@dcs.shef.ac.uk

Résumé

Nous présentons SumUM, un système de génération de résumé automatique des textes, et les résultats obtenus lors de l'évaluation d'une version légèrement modifiée utilisée pour la campagne DUC2002 dans le cadre de l'ACL 2002. Malgré le fait que SumUM n'avait été utilisé jusqu'ici que pour produire des résumés d'articles scientifiques, les résultats sur un domaine complètement différent sont parmi les meilleurs de tous les systèmes qui ont participé à DUC 2002.

1. Introduction

SumUM est un système de résumé automatique développé par H.Saggion (2000), qui produit de courts résumés automatiques de longs documents scientifiques et techniques. Il s'agit d'une approche de génération de résumé, qui produit des résumés indicatifs et informatifs. SumUM produit le résumé en deux étapes : l'utilisateur reçoit d'abord un résumé indicatif où il identifie les topiques de document qui l'intéressent; le système génère ensuite un résumé informatif qui élabore les topiques choisis par l'utilisateur. L'implémentation est basée sur une analyse syntaxique et sémantique, l'identification conceptuelle et la régénération des textes.

Pour la participation au colloque DUC 2002 portant sur l'évaluation des systè-

mes de génération de résumé automatique, nous avons modifié SumUM sans en changer l'algorithme ou les *templates* qui avaient été développés manuellement à partir de l'étude de corpus des résumés écrits par les résumeurs professionnels.

2. Description du système

SumUM est un système de génération de résumé indicatif-informatif des articles scientifiques. L'exécution de notre méthode se fonde sur la sélection des types particuliers d'information du texte source; l'instanciation de différents types de *templates*; la sélection des *templates* afin de produire un abstrait indicatif; la régénération d'un nouveau texte court qui indique les topiques du document et; l'expansion du texte indicatif avec l'élaboration des topiques.

L'entrée de SumUM est un article scientifique en anglais (souvent de plusieurs pages), contenant les éléments structuraux suivants: Titre de l'article, l'information d'auteur et l'affiliation, section d'introduction, section principale sans structure prototypique, références, et sur option, résumé d'auteur, liste de mots-clés, et remerciement.

- **Prétraitement et Interprétation**

Le texte source est étiqueté et transformé dans une représentation structurée, permettant aux processus suivants d'accéder à la structure du texte (mots, groupes de

mots, titres, phrases, paragraphes, sections, et ainsi de suite). Par la suite il faut identifier des groupes nominaux et des groupes de verbe. Les étiquettes sémantiques, les relations de repérage de domaine de discours et les concepts sont ensuite ajoutés aux différents éléments de la structure.

- **Sélection indicative**

Sa fonction est d'identifier des phrases de type indicatif pour construire le contenu de résumé indicatif, qui extrait l'information appropriée du texte et identifie des topiques potentiels du document.

- **Sélection informative et génération**

Ce processus détermine les phrases de type informatif. Il élabore les topiques calculés par la sélection indicative, selon l'intérêt de l'utilisateur. La sortie est un résumé informatif composé de phrases régénérées.

3. Évaluation de SumUM à DUC 2002

Avant de présenter nos résultats, nous décrivons les corpus et les procédures d'évaluation de Document Understanding Conference 2002 (DUC 2002).

DUC est une nouvelle série d'évaluation soutenue par National Institute of Standards and Technology des États-Unis (NIST) sous le projet TIDES, pour progresser plus loin dans la production de résumé automatique et pour permettre à des chercheurs de participer aux expériences à grande échelle. Il y avait trois tâches définies comme suit en 2002.

(1) Résumé entièrement automatique d'un seul document (*single-document summarization*).

(2) Résumé entièrement automatique des documents multiples: étant donné un ensemble de document sur un sujet simple, des participants doivent créer 4 résumés

génériques de l'ensemble avec approximativement 50, 100, 200, et 400 mots.

(3) Un ou deux projets pilotes avec l'évaluation extrinsèque.

NIST a produit 60 ensembles de référence d'une dizaine de documents. Ces ensembles proviennent des données des disques de TREC utilisés dans les questions et réponses dans TREC-9. Chaque ensemble contient des documents courts définis selon quatre types de critères : un événement simple de catastrophe naturelle, un événement simple dans n'importe quel domaine, des événements distincts multiples du même type, et des informations biographiques sur un même individu.

Nous avons utilisé SumUM pour participer à la piste *single-document summarization* de DUC 2002. SumUM a produit des résumés de 100 mots ou moins. 13 systèmes ont participé à cette tâche. Les corpus de tests ont été distribués en fin mars 2002 et les résultats ont été soumis au NIST pour l'évaluation le 15 avril.

4. Adaptation de SumUM pour DUC 2002

Nous avons modifié SumUM afin de produire des résumés pour DUC 2002 afin de tenir compte du contexte différent des textes source et d'en identifier la structure. Nous avons changé le mode interactif du système et de son interface. Dans la version originale de SumUM, l'utilisateur doit guider interactivement par des menus chaque étape du pré-traitement, de l'interprétation, de la sélection indicative, de la sélection. Dans notre version, tous les processus sont automatiques: il prend un ensemble de documents de DUC 2002 comme l'entrée

et il produit les résumés de tous les dossiers inclus dans l'ensemble.

Pour chaque ensemble de 10 documents, nous exécutons un seul programme permettant au système de faire les processus suivants automatiquement et enregistrer le résultat avec les soumissions demandées par DUC :

1) Traitement de document source

Nous avons un programme de Perl pour prendre les documents de DUC, enlever les balises et les adapter avec la structure d'entrée reconnaissable par SumUM.

2) Prétraitement et interprétation

Ce processus identifie la structure du document et interprète les phrases du texte.

3) Processus principal de SumUM

SumUM produit d'un résumé court du texte. Pour les corpus de tests de DUC, le système n'est pas interactif et automatiquement génère les résumés, sans intervention d'utilisateur.

4) Sortie du système

Un programme de Perl prend le résumé, produit par SumUM et l'adapte, avec la soumission demandé pour DUC 2002.

5. Résultats de l'évaluation

NIST a effectué deux types d'évaluation des résumés : une faite par des humains (pour les résumés générés), et une autre faite automatiquement (pour les extraits) par NIST. Pour évaluer la qualité des résumés, NIST a utilisé douze questions qualitatives portant sur des aspects tels la grammaticalité, la cohérence et l'organisation des résumés. La table 1 montre les résultats obtenus pour l'évaluation des 13 systèmes participants à la tâche *single-document*. Les résultats de SumUM sont parmi les meilleurs. Ceci montre la performance d'identification des éléments

importants d'un texte par SumUM et l'intérêt d'avoir des phrases régénérées et bien formées.

6. Conclusions

Nous avons présenté notre adaptation de SumUM, un système de résumé automatique, pour la tâche de résumé de courts documents lors de l'évaluation 2002. SumUM avait été développé pour le résumé automatique de longs documents scientifiques et techniques qui sont très différents du corpus DUC dont le corpus de test comportait des textes très courts dans un style différent des textes scientifiques. Les résumés produits par SumUM ont été jugés excellents par rapport à ceux produits par d'autres développés expressément pour ce genre de données.

References

- Barzilay, R. and M. Elhadad, "Using Lexical Chains for Text Summarization", In Proceedings of the Intelligent Scalable Text Summarization Workshop (ISTS'97), ACL, Madrid, 1997.
- Hardy H., N. Shimizu, T. Strzalkowski, L. Ting, X. Zhang, G. B. Wise. (2002), "Cross-document summarization by concept classification". In *Proceedings of the SIGIR Conference*, August, Tampere, Finland: 121-128.
- Lin, C.Y. (2001), "SEE - Summary Evaluation Environment". <http://www.isi.edu/~cyl/SEE/>
- Lin, C.-Y. and E.H. Hovy. (2000), "The Automated Acquisition of Topic Signatures for Text Summarization". In *Proceedings of the COLING Conference*. Saarbrücken, Germany.
- Minel, J.-L., J.-P. Desclés. (2000) "Résumé Automatique et Filtrage des textes, in ingénierie des langues", (sous la direction de J.-M. Pierre) Paris, Editions Hermès, p.253-270.
- Radev, D. and McKeown, K. "Generating natural language summaries from multiple on-line sources". *Computational Linguistics*, 24(3):469-500, September 1998.
- Saggion, H. and G. Lapalme, (2002), "Generating Informative and Indicative Summaries with SumUM", *Computational Linguistics*, Special Issue on Automatic Summarization.

Group	Mean Error Quality Questions	Group	Count of Quality Questions	Group	Mean Length-Adjusted Coverage
SumUM	0.408	Microsoft	0.582	BBN	0.339
Microsoft	0.425	LCC	0.698	SumUM	0.299
LCC	0.448	SumUM	0.758	LCC	0.293
CCS-NSA	0.537	U.Nijmegen	0.885	Microsoft	0.272
U.Ottawa	0.551	U.Ottawa	0.986	NTT	0.272
NTT	0.552	Imperial College	0.997	CCS-NSA	0.261
<i>Moyen</i>	0.642	<i>Moyen</i>	1.275	<i>Moyen</i>	0.259
U.Nijmegen	0.561	CCS-NSA	1.013	U.Leuven	0.251
Imperial College	0.565	NTT	1.014	U.Nijmegen	0.247
U.Michigan	0.644	U.Lethbridge	1.153	U.Lethbridge	0.24
U.Leuven	0.660	U.Leuven	1.210	U.Ottawa	0.232
U.Lethbridge	0.676	U.Michigan	1.441	Imperial College	0.228
BBN	1.040	BBN	2.637	ISI/GLEANS	0.220
ISI/GLEANS	1.281	ISI/GLEANS0	3.200	U.Michigan	0.214
<i>Humain</i>	<i>0.354</i>		<i>0.505</i>		<i>0.336</i>
<i>Baseline</i>	<i>0.490</i>		<i>0.718</i>		<i>0.255</i>

Table 1: Les résultats d'évaluation de 13 systèmes, qui ont participé à la tâche de résumé automatique de *Single-Document Summarization* à DUC 2002