

Résumer automatiquement en ligne : démonstration d'un service web de résumé multidocument

Valentin Nyzam Nathan Gatto Aurélien Bossard
LIASD, Université Paris 8 - IUT de Montreuil, 140 rue de la Nouvelle France,
93100 Montreuil, France

valentin.nyzam@iut.univ-paris8.fr,
nathan.gatto@free.fr,
aurelien.bossard@iut.univ-paris8.fr

RÉSUMÉ

Nous proposons une démonstration d'un webservice de résumé automatique multidocument. Ce webservice s'appuie sur un outil ouvert qui implémente plusieurs algorithmes reconnus de résumé automatique, et permet de résumer des documents en utilisant des configurations différentes.

ABSTRACT

Summarizing Automatically Online :

We propose a demonstration of an automatic multidocument summarization web service. This web service relies on an open source summarization tool that implements several known summarization algorithms, and allows to summarize documents using different configurations.

MOTS-CLÉS : résumé automatique, service web.

KEYWORDS: automatic summarization, web service.

1 Introduction

Le résumé automatique multidocument par extraction, comme beaucoup d'autres tâches du traitement automatique du langage, nécessite au-delà des algorithmes d'extraction des phrases, des ressources externes. Les systèmes développés par les chercheurs de la communauté s'appuient parfois sur des bibliothèques externes.

Ainsi, effectuer des expérimentations sur des corpus en utilisant des algorithmes reconnus peut se révéler fastidieux. Un service web de résumé automatique proposant des méthodes de l'état de l'art peut être une solution efficace à ce problème. Aujourd'hui, quelques services web existent en ligne. Certains sont payants (par exemple <https://resoomer.com>), d'autres gratuits (SweSum¹, Text Compactor², ou encore Open Text Summarizer³) mais ceux-ci sont parfois datés, voire n'annoncent pas les méthodes qu'ils utilisent.

1. <http://swesum.nada.kth.se/index-eng.html>

2. <http://textcompactor.com/>

3. <https://www.splitbrain.org/services/ots>

2 Présentation du service web

Nous présentons un service web de résumé automatique qui s'appuie sur un outil libre qui implémente plusieurs méthodes reconnues de résumé automatique multidocument. Ce service web permet d'une part d'envoyer le corpus que l'on souhaite résumer, d'autre part de choisir la configuration complète de lancement du résumé automatique. Le service web donne également au client le choix entre plusieurs méthodes, et permet pour chacune des méthodes, de choisir des options propres à celle-ci.

A titre d'exemple, les méthodes ou algorithmes suivants peuvent être lancés par le service web :

- MMR (Carbonell & Goldstein, 1998);
- ICSISUMM : résumé par programmation linéaire en nombres entiers (Gillick & Favre, 2009b);
- Algorithme évolutionnaire (Bossard & Rodrigues, 2015).
- Centroid (Radev *et al.*, 2004);
- LexRank (Erkan & Radev, 2004);
- LDA (Blei *et al.*, 2003);
- LSA;
- *Word embeddings* (Zhang *et al.*, 2015)

Ainsi, la constitution du corpus à envoyer ainsi que la configuration pas à pas nécessitent une prise en main et n'est sûrement pas optimale malgré nos efforts.

3 Objectifs de la démonstration

La démonstration a pour objectif de présenter à la communauté notre service web. Nous proposerons également un tutoriel pour apprendre à constituer les corpus à envoyer et pour comprendre les configurations de lancement de l'outil proposées par le service web.

4 Remerciements

Les travaux décrits dans cet article ont bénéficié du soutien de l'ANR (projet Jeunes Chercheurs/Jeunes Chercheuses ASADERA - convention ANR-16-CE38-0008)

Références

BLEI D. M., NG A. Y. & JORDAN M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, **3**, 993–1022.

BOSSARD A. & RODRIGUES C. (2015). Une approche évolutionnaire pour le résumé automatique. In *TALN 2015 - 22ème Conférence sur le Traitement Automatique des Langues Naturelles*.

CARBONELL J. & GOLDSTEIN J. (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *SIGIR'98 : Proceedings of the 21st ACM SIGIR Conference*, p. 335–336.

ERKAN G. & RADEV D. R. (2004). Lexrank : Graph-based centrality as salience in text summarization. *Journal of Artificial Intelligence Research (JAIR)*.

GILLICK D. & FAVRE B. (2009b). A scalable global model for summarization. In *Proceedings of the Workshop on Integer Linear Programming for Natural Language Processing*, p. 10–18 : Association for Computational Linguistics.

RADEV D. R., JING H., STY M. & TAM D. (2004). Centroid-based summarization of multiple documents. *Information Processing Management*, **40**, 919–938.

ZHANG Y., ER M. J. & ZHAO R. (2015). Multi-document extractive summarization using window-based sentence representation. In *2015 IEEE Symposium Series on Computational Intelligence*, p. 404–410.