

La structure du contenu textuel a-t-elle un impact sur les modèles linguistiques pour le résumé automatique ?

Eve Sauvage^{1,2} Sabrina Campano¹ Lydia Ould-Ouali¹ Cyril Grouin²

(1) EDF RD, 7 boulevard Gaspard Monge, 91120 Palaiseau, France

(2) LISN, Campus Universitaire bâtiment 507, Rue du Belvédère, 91400 Orsay, France

eve.sauvage@edf.fr, sabrina.campano@edf.fr, sabrina.ould-ouali@edf.fr,
cyril.grouin@lisen.fr

RÉSUMÉ

Le traitement de séquences longues par des modèles de langues reste un sujet à part entière, y compris pour le résumé automatique, malgré des améliorations récentes. Dans ce travail, nous présentons des expériences de résumé automatique d'articles scientifiques à l'aide de modèles BART, prenant en compte les informations textuelles provenant de passages distincts des textes à résumer. Nous démontrons que la prise en compte de la structure du document améliore les performances des modèles et se rapproche des performances de LongFormer en anglais.

ABSTRACT

Does the structure of textual content have an impact on language models for automatic summarization ?

The processing of long sequences with models remains a subject in its own right, including automatic summary, despite recent improvements. In this work, we present experiments on the automatic summarization of scientific articles using BART models, taking into account textual information coming from distinct passages from the long texts to be summarized. We demonstrate that taking into account document structure improves the performance of state-of-the-art models and approaches the performance of LongFormer on English.

MOTS-CLÉS : Résumé automatique, Séquences longues, Structure des textes.

KEYWORDS: Automatic Summarization, Long context, Text structure.

ARTICLE : **Accepté à** Annual Meeting of the Association for Computational Linguistics (Volume 4 : Student Research Workshop).

<https://aclanthology.org/2024.acl-srw.25.pdf>