

Apport de la structure de tours à l'identification automatique de genre textuel: un corpus annoté de sites web de tourisme en français*

Rémi Cardon¹ Trang Pham Tran Hanh^{1,2}
Julien Zakhia Doueïhi¹ Thomas François¹
(1) CENTAL, IL&C, Université catholique de Louvain, Belgique
prenom.nom@uclouvain.be,
(2) Département de français, Université de Hanoï, Vietnam

RÉSUMÉ

Ce travail étudie la contribution de la structure de tours à l'identification automatique de genres textuels. Ce concept semble être peu exploité dans l'identification automatique du genre. Nous décrivons la collecte d'un corpus de sites web francophones relevant du domaine du tourisme et le processus d'annotation avec les informations de tours. Nous menons des expériences d'identification automatique du genre de texte avec notre corpus. Nos résultats montrent qu'ajouter l'information sur la structure de tours dans un modèle améliore ses performances pour l'identification automatique du genre, tout en réduisant le volume de données nécessaire et le besoin en ressource de calcul.

ABSTRACT

Contribution of Move Structure to Automatic Genre Identification : an Annotated Corpus of French Tourism Websites

The present work studies the contribution of the moves structure to automatic genre identification. This concept - well known in other branches of genre analysis - seems to have little application in automatic genre identification. We describe how we collect a corpus of websites in French related to tourism and annotate it with move structure. We conduct experiments on automatic genre identification with our corpus. Our results show that informing a model with move structure can increase its performance for automatic genre identification, and reduce the need for annotated data and computational power.

MOTS-CLÉS : identification automatique du genre de texte, analyse de genre, corpus, annotation.

KEYWORDS: automatic genre identification, genre analysis, corpus, annotation.

Références

CARDON R., PHAM T. T. H., ZAKHIA DOUEIHI J. & FRANÇOIS T. (2024). Contribution of move structure to automatic genre identification : An annotated corpus of French tourism websites. In N. CALZOLARI, M.-Y. KAN, V. HOSTE, A. LENCI, S. SAKTI & N. XUE, Éd.s., *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, p. 3916–3926, Torino, Italy : ELRA and ICCL.

*. Ce document est un résumé de l'article (Cardon *et al.*, 2024).