Actes de CORIA-TALN 2026
Nantes, France, 29 juin au 3 juillet 2026

Un LLM pour guide : une approche d'extraction non supervisée de relations économiques dans les documents administratifs français.

Thomas Sebbag[1, 2] Solen Quiniou[1] Emmanuel Morin[1]
Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France
Explore, Carquefou, France
Emails : thomas.sebbag@univ-nantes.fr, solen.quiniou@univ-nantes.fr, emmanuel.morin@univ-nantes.fr

RÉSUMÉ

L'efficacité de l'extraction de relations (ER) à partir de textes non structurés est essentielle, en particulier lorsque les relations cibles ne sont pas connues à l'avance. Dans cet état d'imprévu permanent, il est possible d'exploiter les grands modèles de langue (LLM) pour accomplir cette tâche. Dans cet article, nous présentons une approche où un LLM joue le rôle de guide pour l'identification de relations économiques afin d'accompagner des experts métier. Cette méthodologie innovante, fondée sur le regroupement de phrases et l'utilisation d'un LLM, permet d'identifier des relations économiques jusque-là inconnues dans des documents administratifs français. Elle répond au défi consistant à extraire des connaissances exploitables sans disposer d'étiquettes de relations prédéfinies. Nous évaluons notre approche sur des jeux de données d'extraction de relations en français et en anglais, en démontrant une précision et un rappel élevés dans la détection de relations inconnues. Nos résultats suggèrent que les méthodes combinant le regroupement et les LLM peuvent découvrir et catégoriser efficacement des relations économiques, avec des applications potentielles sur des corpus d'entreprises.

ABSTRACT

LLM as a Guide : an Approach for Unsupervised Economic Relation Discovery in Administrative Documents

Effective relation extraction (RE) from unstructured text is critical, especially when the target relations are unknown. In this case, we can leverage Large Language Models (LLMs) to perform this task. In this paper, we present an LLM as a guide approach for identifying economic relations. This novel methodology, based on sentence clustering and an LLM, is used to identify previously unknown economic relations in French administrative documents. It addresses the challenge of extracting actionable knowledge without predefined relation labels. We evaluate our approach on French and English RE datasets, demonstrating high precision and recall in detecting previously unknown relations. Our results suggest that clustering and LLM-based methods can effectively discover and categorize economic relations, with potential applications to private corpora.

MOTS-CLÉS : document administratif, LLM, extraction relations ouvertes, identification de relations, économie.

Actes des 33ème Conférence sur le Traitement Automatique des Langues Naturelles.
Volume 2 : articles déjà publiés, pages 50–51.
Cette œuvre est mise à disposition sous licence Attribution 4.0 International.

KEYWORDS: adminstrative documents, LLM, open relation extraction, relaiton identification, economy.

ARTICLE : **Accepté à** The 10th Workshop on Financial Technology and Natural Language Processing (FinNLP acceuilie lors de EMNLP 2025)

https://aclanthology.org/2025.finnlp-2.1/ .