

Découverte de relations spatiales dans l'encyclopédie de Diderot & d'Alembert

Fabien Duchateau, Ludovic Moncla (prénom.nom@liris.cnrs.fr)

Contexte. De nombreuses connaissances sont actuellement stockées dans des documents textuels, ce qui les rend difficilement exploitables directement par les machines [2]. L'extraction d'information est un domaine de recherche visant à identifier, à partir d'un texte, des entités et leur(s) relation(s). Par exemple, la phrase "Lyon est le chef-lieu de la région AURA" permet d'extraire l'information (Lyon, chef-lieu, Auvergne-Rhône-Alpes). Les entités extraites peuvent être liées à une base de connaissances externe (comme Wikidata ou Geonames) pour lever toute ambiguïté (e.g., [geonames:Lyon](#)). Mais la relation nécessite également d'être associée à un type de relation plus général pour être exploitable (e.g., *chef-lieu* peut se rapporter au type de relation [capital-of](#)).

Un [prototype \(notebook python\)](#) permet d'extraire les entités et leur relation mentionnées dans l'encyclopédie de Diderot & d'Alembert. Dans ce projet, nous voulons associer les relations extraites à un type de relation. Certaines relations, insolites ou peu fréquentes, pourront être identifiées afin d'enrichir la taxonomie des types de relation.

Objectifs. Ce projet cherche à atteindre les objectifs suivants :

- Construire une taxonomie des types de relation [1, 3], en s'inspirant des ontologies spatiales existantes ([ontologie de Wikidata](#), [relations spatiales DE-9IM](#));
- Développer un algorithme (e.g., de classification) qui associe les mentions de relation à un type de relation de la taxonomie;
- Évaluer la qualité de l'algorithme en construisant un jeu de données expertisé.

Bibliographie

- [1] C. Claramunt. Ontologies for geospatial information: Progress and challenges ahead. *Journal of Spatial Information Science*, 1(20):35–41, 2020. [Lien JSIS2020](#).
- [2] X. Han, T. Gao, Y. Lin, H. Peng, Y. Yang, C. Xiao, Z. Liu, P. Li, M. Sun, and J. Zhou. More data, more relations, more context and more openness: A review and outlook for relation extraction. *JCNLP*, 2020. [Lien JCNLP2020](#).
- [3] F. Zhang, Q. Lu, Z. Du, X. Chen, and C. Cao. A comprehensive overview of rdf for spatial and spatiotemporal data management. *The Knowledge Engineering Review*, 36:e10, 2021. [Lien KER2021](#).