Reconnaissance d'entités géographiques dans des documents textuels

Contact : Fabien Duchateau (prénom.nom@univ-lyon1.fr)

Contexte. Les données cartographiques sont au cœur de nombreuses applications, comme le suivi de véhicules ou de personnes, la gestion de catastrophes naturelles ou encore l'aménagement du territoire. Des référentiels comme Geonames ou OpenStreetMap fournissent des informations détaillées sur les lieux, mais il existe peu d'informations sur les relations entre lieux. Pourtant ces relations sont présentes dans des documents textuels (e.g., pages web). Par exemple, la phrase *"la basilique de Fourvière se trouve à Lyon"* décrit une relation entre deux lieux qui sont chacun représentés par une mention. L'identification d'un lieu dans un texte est un processus nécessaire pour extraire des informations, comme la relation qui existe entre deux lieux. Ce processus de reconnaissance d'entités nommées (NER) est aujourd'hui étendu à l'entity linking, qui consiste à relier la mention d'un lieu vers son entité correspondante selon un référentiel. Dans l'exemple précédent, il s'agirait de relier la mention *basilique de Fourvière* à l'entité *8015555* sur le référentiel Geonames.

Objectifs. Ce sujet de TER a pour objectif de découvrir dans un document textuel les liens vers les entités géographiques du référentiel Geonames. De nombreuses approches existent [1], et l'idée est de comparer ces approches pour sélectionner celle qui obtient les meilleurs résultats en termes de qualité et de performance pour le domaine géographique.

Organisation. Pour répondre à cet objectif, le TER sera organisé ainsi :

- Lire un article survey sur l'entity linking [1]
- Implémenter au moins deux approches d'entity linking du survey
- Définir un benchmark (dataset et métriques d'évaluation) pour le domaine spatial
- Comparer avec le benchmark les approches implémentées

Bibliographie

[1] W. Shen, J. Wang, and J. Han. Entity linking with a knowledge base: Issues, techniques, and solutions. *Knowledge and Data Engineering, IEEE Transactions on*, 27(2):443–460, 2015. Link to TKDE14-entitylinking.pdf.