

Intégration spatiale entre Geonames et OpenStreetMap

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

Contexte. Plusieurs fournisseurs de services proposent des informations sur les lieux, et ce sous différentes formes. Par exemple, [Geonames](#) est une base de données géographique (plus de 10 millions de lieux¹) tandis qu'[OpenStreetMap](#) est une base de données cartographique (des milliards de points/lignes/surfaces, dont certains représentent des lieux²). Ces deux fournisseurs possèdent de nombreuses entités spatiales qui représentent le même lieu du monde réel (e.g., la Basilique de Fourvière, [chez Geonames](#) et [chez OpenStreetMap](#)). L'alignement de schémas ou d'entités est un processus qui consiste à découvrir des correspondances entre des sources de données hétérogènes [1], et qui permet entre autre l'intégration de données.

Objectifs. Ce sujet a pour objectifs d'étudier l'intégration entre les données de Geonames et celles de OpenStreetMap. Les challenges pour l'alignement de schémas et d'entités concernent la présence d'attributs librement nommés (sous OpenStreetMap), des entités aux représentations différentes (points/lignes/surfaces), le choix de la combinaison des mesures de similarité (pour une qualité d'alignement optimale) et le passage à l'échelle (des millions d'entités à aligner).

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

- Lire un article sur l'intégration de données [1] et son application au domaine spatial [2, 3]
- Proposer une approche d'alignement entre les données Geonames et OpenStreetMap, en se focalisant notamment sur la méthode qui combine les scores des mesures de similarité. Un jeu de données expertisé sera créé au fur et à mesure pour évaluer l'approche proposée
- Optimiser l'approche proposée pour le passage à l'échelle

Bibliographie

- [1] P. A. Bernstein, J. Madhavan, and E. Rahm. Generic schema matching, ten years later. *Proceedings of the VLDB Endowment*, 4(11):695–701, 2011. [Link to VLDB2011.pdf](#).
- [2] G. McKenzie, K. Janowicz, and B. Adams. A weighted multi-attribute method for matching user-generated points of interest. *Cartography and Geographic Information Science*, 41(2):125–137, 2014. [Link to CaGIS2013.pdf](#).
- [3] G. Recchia and M. Louwerse. A comparison of string similarity measures for toponym matching. *Proceedings of ACM SIGSPATIAL CoMP*, 13, 2013. [Link to CoMP2013.pdf](#).

¹Statistiques Geonames

²Statistiques OpenStreetMap