

Développement d'un outil d'aide à la construction d'un jeu de données cartographiques

Contact : Fabien Duchateau (fabien.duchateau@univ-lyon1.fr)
Bilal Berjawi (bilal.berjawi@liris.cnrs.fr)

Type de sujet : recherche / développement

Le projet UNIMAP a pour objectif la production d'une carte unifiée à partir de différents fournisseurs cartographiques (e.g., *Google Maps*, *OpenStreetMap*, *ViaMichelin*, *Geonames*). Pour obtenir cette carte unifiée, il est nécessaire d'intégrer, pour chaque fournisseur, les points d'intérêt (POI) qui représentent la même entité. Ainsi les deux POIs affichés par *Google Maps* et *Here* (voir Figure 1), bien que possédant quelques valeurs différentes au niveau des attributs, représentent la même entité (l'hôtel Tenor à Saint-Etienne). Une typologie des erreurs (e.g., attribut manquant, valeur erronée) qui recense les différences entre deux POIs représentant la même entité est en cours d'élaboration. Afin de tester l'efficacité des algorithmes d'intégration, il est nécessaire de disposer d'un « benchmark », c'est à dire un jeu de données expertisé. La construction manuelle d'un tel jeu de données étant fastidieuse, les chercheurs du projet UNIMAP ont besoin d'un outil pour faciliter la création du jeu de données. Cet outil permettra à la fois de générer les correspondances entre entités équivalentes, mais aussi de détecter le(s) type(s) d'erreur potentiels d'une entité et de récupérer d'autres informations sur Geonames (e.g., type détaillé). Ce TER a donc pour objectif le développement de cet outil d'aide à la construction du jeu de données expertisé. Plus précisément, il s'agira de :

- lire un article sur la mise en correspondance d'entités géospatiales (« entity matching »);
- développer un outil qui permet aux chercheurs de visualiser un même POI chez différents fournisseurs cartographiques (en utilisant un algorithme simple de « matching »), et de valider les correspondances entre chaque entité représentant ce POI ;
- proposer et développer des algorithmes pour détecter chaque type d'erreur recensé dans la typologie. Chaque erreur détectée sera validée par les chercheurs ;
- évaluer l'efficacité et les performances de ou des algorithmes proposés.

Pré-requis : MIF18

Bibliographie :

Entity resolution in geospatial data integration.

Vivek Sehgal, Lise Getoor, and Peter D Viechnicki. 2006. In *Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*
<http://doi.acm.org/10.1145/1183471.1183486>
<http://cgis.cs.umd.edu/Grad/scholarlypapers/papers/sehgal.pdf>

Projet UNIMAP

<http://liris.cnrs.fr/unimap/>

Base de données géographique Geonames

<http://www.geonames.org/>

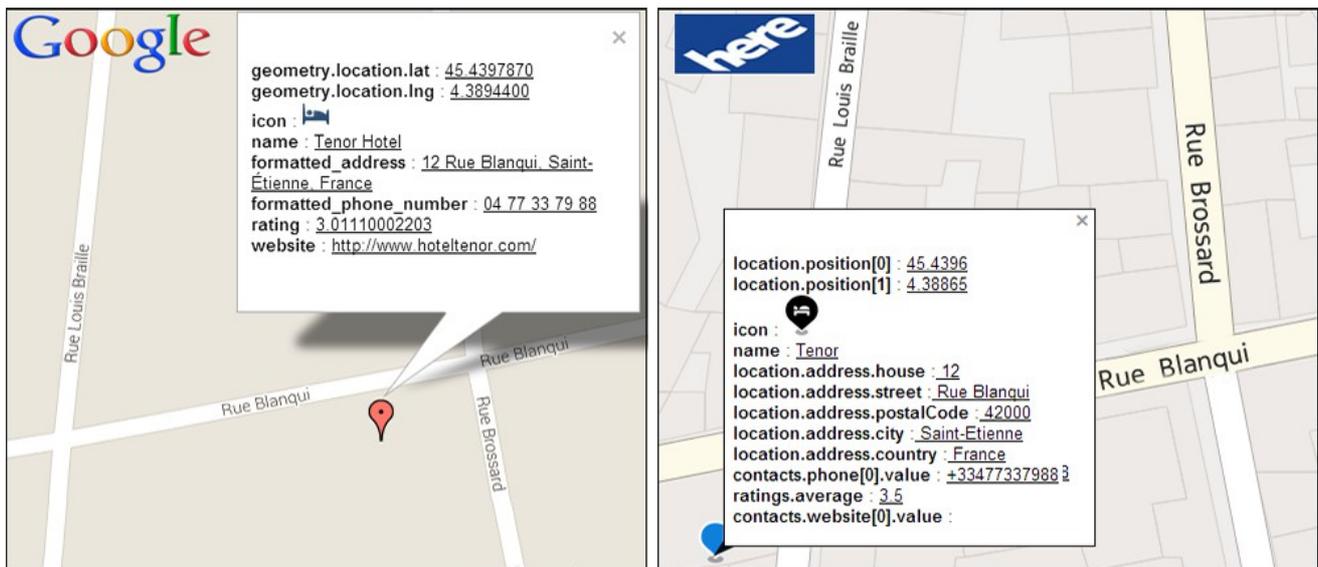


Figure 1. Visualisation du POI « Tenor Hotel, Saint-Etienne » chez deux fournisseurs, Google Maps et Here.