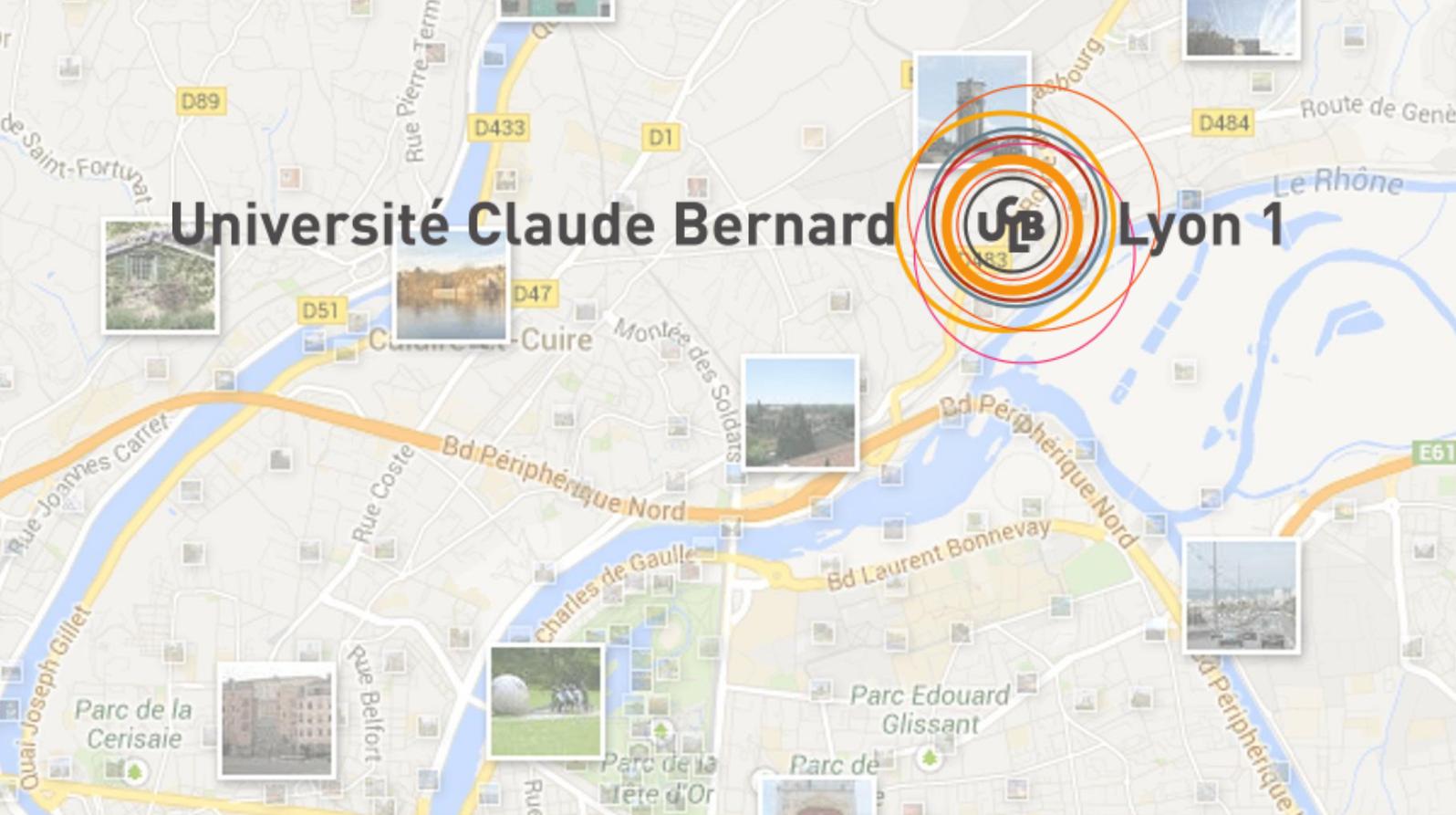


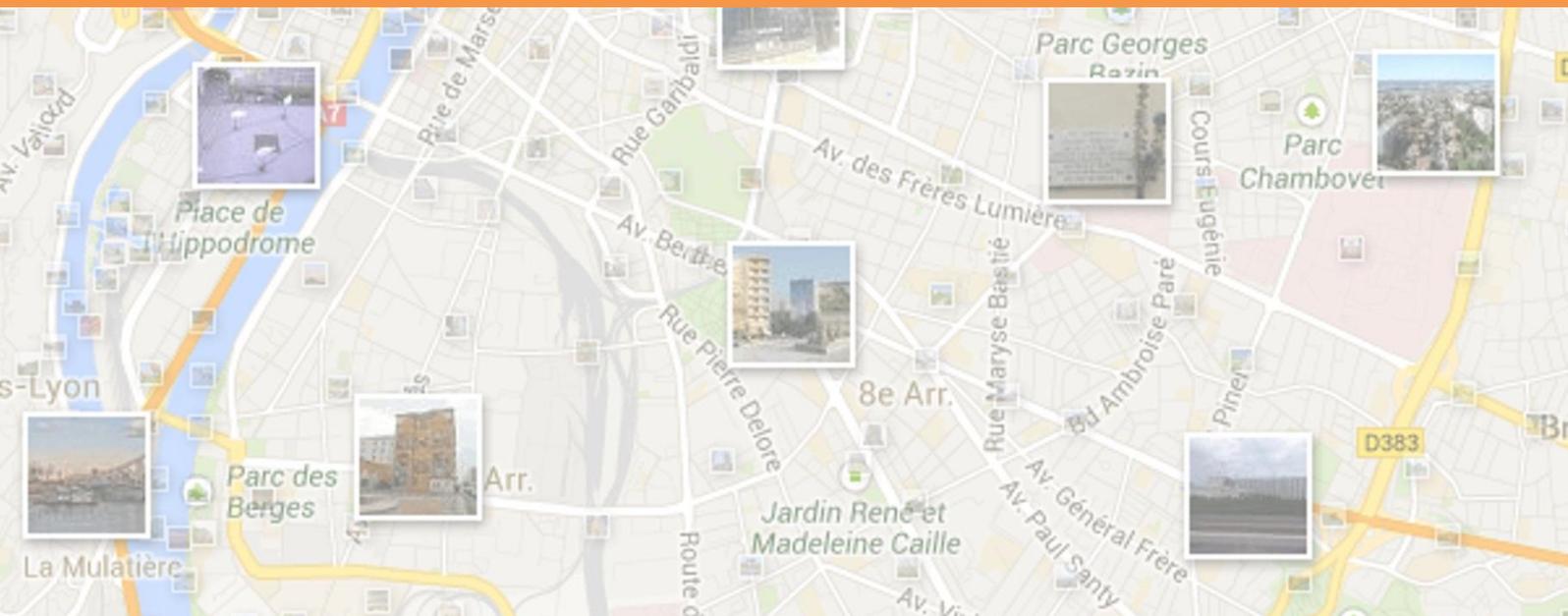
Université Claude Bernard Lyon 1



Projet de recherche

DEVELOPPEMENT D'UN OUTIL D'AIDE A LA CONSTRUCTION D'UN JEU DE DONNEES CARTOGRAPHIQUES

(Cahier des charges)



Réalisé par :
MORANA Anthony 10908950
MOREL Thomas 11005517

Encadré par :
DUCHATEAU Fabien
BERJAWI Bilal

I. Présentation du projet

1. Le contexte

Ce TER s'inscrit dans le cadre de l'UE MIF20 « Projet de recherche », et porte sur le projet de recherche UNIMAP [1]. Ce dernier a pour but de créer une carte unifiée à partir de différents fournisseurs cartographiques (e.g. Google Map, Open Street Map). Chaque fournisseur représente un point d'intérêt¹ (POI) par une ou plusieurs entités. Ainsi, les informations concernant un même point d'intérêt peuvent différer selon le fournisseur. Par exemple, on peut voir sur les figures 1 et 2 que les informations spatiales et sémantiques d'un même POI sont différentes selon le fournisseur. Pour parvenir à l'unification de plusieurs cartes, deux phases sont nécessaires. La première consiste à découvrir des correspondances entre les entités (alignement) [2]. La deuxième étape a pour objectif l'intégration des entités correspondant à un même POI.

La découverte de correspondances entre les différentes cartes peut être fastidieuse à réaliser à la main. L'un des objectifs du projet UNIMAP est donc de proposer une approche de découverte de correspondances semi-automatique. Pour valider les performances qualitatives obtenues par cette approche, il est nécessaire de disposer d'un benchmark (jeu de données expertisé), qui sera construit grâce à un outil élaboré pendant ce TER.

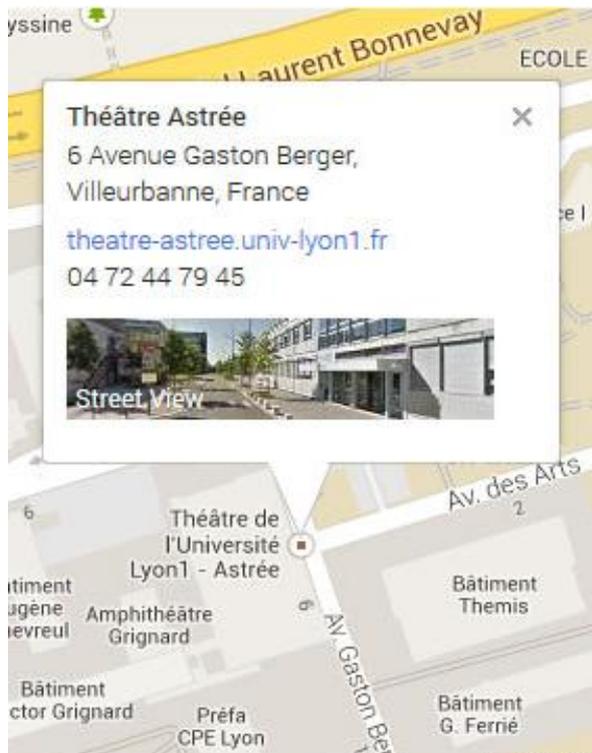


Figure 1 : Le POI « Théâtre Astrée » est représenté par le fournisseur Geonames

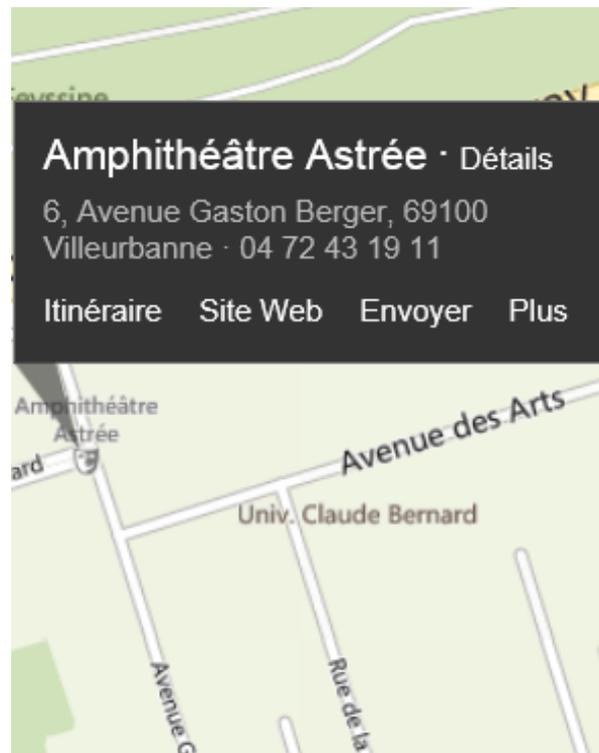


Figure 2 : Le POI « Amphithéâtre Astrée » est représenté par le fournisseur BingMap

¹ Entité nommée dans le monde réel

2. Objectifs

Notre objectif au cours de ce TER est de produire un jeu de données expertisé permettant de vérifier des algorithmes d'alignement. Pour cela nous devons concevoir un outil permettant à la fois de proposer des correspondances entre entités potentiellement équivalentes et de détecter les différences entre les entités.

Dans un premier temps, nous proposerons un algorithme qui permet, pour une entité géographique donnée chez un fournisseur source, de trouver les entités correspondantes chez les autres fournisseurs cibles [3,4]. Un second algorithme aura pour but d'ordonner l'ensemble d'entités correspondantes retournées par un seul fournisseur selon leur pertinence. Enfin, un troisième algorithme détectera les différences entre entités, telles qu'elles sont définies dans la typologie fournie par UNIMAP. Un expert validera chaque correspondance proposée ainsi que chacune des différences détectées. A la fin de la résolution de conflit, l'outil devra stocker les informations sur les correspondances expertisées.

L'algorithme d'ordonnement des entités correspondantes pourra être évalué avec des expérimentations.

II. Description des besoins et contraintes

L'outil développé devra permettre à l'expert de sélectionner une entité sur une carte (idéalement Geonames). Il pourra également renseigner un paramètre optionnel (nommé K) indiquant le nombre d'entités à proposer sur chaque carte.

Une fois les paramètres saisis, la découverte de correspondances avec les autres cartes (minimum deux, Google Maps et Here) est lancée. L'utilisateur devra indiquer pour chaque correspondance proposée si elle est correcte ou non. Pour chaque expertise effectuée par l'utilisateur, celle-ci doit être enregistrée dans une base de données, quelle que soit la pertinence de la correspondance entre les deux entités (correcte ou incorrecte). De plus il faut pouvoir sauvegarder les différences entre les deux entités (e.g., nom ou type différents).

En ce qui concerne les POI, nous ne traiterons que ceux représentés par un point. Les points d'intérêts représentés par un polygone (comme par exemple le contour d'une ville) ne font pas partie de notre étude à moins que ceux-ci puissent être représentés par des points (centre d'une ville).

Enfin, l'interface de l'outil devra être intuitive pour que l'expert ne perde pas de temps.

III. Livrables

Ci-dessous la liste des livrables qui seront rendus à la fin du TER :

- Le présent cahier des charges.
- Un algorithme qui en fonction de l'entité prise en entrée (sur le fournisseur Geonames) retourne les correspondances potentielles chez les différents fournisseurs. Une heuristique pour ordonner les correspondances selon leur pertinence. Enfin, un algorithme qui détecte les différences entre l'entité Geonames et celles des autres fournisseurs.
- L'implémentation des algorithmes ci-dessus.
- Une interface pour que l'expert puisse indiquer si la correspondance est correcte ou non.
- Un rapport final détaillant le travail réalisé durant ce TER.

IV. Déroulement du projet

| Durée | Tâches à effectuer |
|------------|---|
| ½ Semaine | Etat de l'art sur l'alignement d'entités |
| ½ Semaine | Etude et comparatif des différents API |
| 1 Semaine | Proposition d'algorithmes (alignement, tri et détection de différences) |
| 2 Semaines | Développement |
| 1 Semaine | Tests/Expérimentations |
| 1 Semaine | Rédaction du rapport et préparation de la soutenance |

V. Références

[1] Projet UNIMAP

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

[2] Frameworks for entity matching: A comparison. Hanna Köpcke, Erhard Rahm. In Data & Knowledge Engineering. 2009.

[3] Entity resolution in geospatial data integration. Vivek Sehgal, Lise Getoor, and Peter D Viechnicki. In Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems. 2006.

[4] Finding Corresponding Objects when Integrating Several Geo-Spatial Datasets. Catriel Beeri, Yerach Doytsher, Yaron Kanza, Eliyahu Safra, Yehoshua Sagiv. In Database Applications—Spatial databases and GIS. 2005.