

UNIVERSITÉ CLAUDE BERNARD LYON 1

---

## Cahier des charges

---

GEOBENCH v2

Du 20 novembre 2018 au 6 juin 2019



LIRIS, Laboratoire d'Informatique en Image et Systèmes d'information  
8 Boulevard Niels Bohr  
69622 Villeurbanne

Étudiante Nelly Barret (Master 1 Informatique, UCBL1)  
Tuteurs Fabien Duchateau (Maître de conférences, LIRIS)  
Franck Favetta (Maître de conférences, LIRIS)

# 1 Contexte

L'application GeoBench, développée en 2014, a pour objectif la production de données cartographiques plus complètes en exploitant différents fournisseurs (Google Maps, Geonames et Here). Sa première version avait pour objectif de créer un benchmark cartographique pour que les utilisateurs fournissent la réalité terrain des correspondances entre les POI (*Point Of Interest*) issus de différents fournisseurs. Les utilisateurs expriment cette réalité terrain des correspondances par attributs grâce à l'interface de cette application. Pour se faire, l'application avait un algorithme de matching ainsi qu'un algorithme de fusion simples et fonctionnels. Cela nécessite la comparaison d'entités (e.g. un hôtel, un parc, un musée...). La figure 1 illustre la comparaison de deux entités qui représentent le même point d'intérêt (POI) mais qui proviennent de fournisseurs différents. Ainsi, il est nécessaire de détecter les entités équivalentes et éventuellement de fusionner leurs informations. Aujourd'hui, l'application GeoBench ne fonctionne plus (principalement à cause des mises à jour et des limitations chez les fournisseurs).

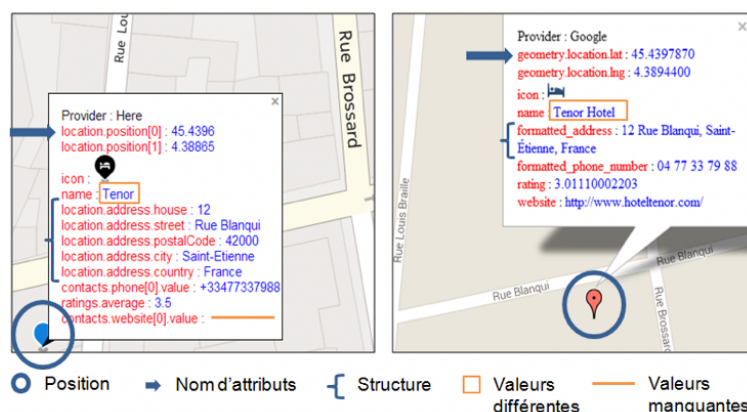


FIGURE 1 – Capture d'écran de l'application GeoBench, avec le POI de l'hôtel Tenor à Saint-Étienne représenté par une entité Here (gauche) et une entité Google Maps (droite). Les différences entre ces entités (e.g., de position, de nommage d'attributs ou de structure) ne facilitent pas leur comparaison.

Dans le cadre de ce projet, Messieurs Fabien Duchateau et Franck Favetta, maîtres de conférences au LIRIS (Laboratoire d'InfoRmatique en Image et Systèmes d'information) et membres de l'équipe « Bases de données », proposent le sujet de stage POM (Projet d'Orientation en Master) : « Geobench v2 ».

## 2 Version actuelle de GeoBench

### 2.1 Système médiateur

GeoBench est un système médiateur, il sert d'interface entre plusieurs fournisseurs cartographiques. Les correspondances entre le schéma médiateur et les fournisseurs sont connues. Un des avantages du système médiateur est de pouvoir facilement ajouter de nouveaux fournisseurs. Les fournisseurs actuels sont [Google Maps](#), [Here](#) et [Geonames](#).

### 2.2 Algorithmes de linking et de fusion

GeoBench implémente un algorithme d'entity linking simple. Cet algorithme est composé d'une phase de blocking (basée sur les API des fournisseurs) et d'une phase de matching (appariement ou comparaison, basé sur les mesures de similarité et un seuil). L'application implémente aussi un algorithme de fusion simple qui permet d'avoir une entité qui regroupe toutes les informations provenant des différents fournisseurs.

### 2.3 Limitations

La version actuelle de GeoBench a une interface peu intuitive, ce qui rend l'expérience utilisateur un peu confuse. En effet, l'application est centrée sur la création d'un benchmark donc le linking est une étape avec laquelle on ne peut pas vraiment interagir.

L'application ne comporte pour l'instant que trois fournisseurs. De plus, c'est le même fournisseur qui est interrogé en premier à chaque fois, ce qui pose deux problèmes. Le premier est que ce fournisseur a plus d'impact sur le résultat par rapport aux deux autres. Le second est que si l'entité n'existe pas pour ce fournisseur, l'étape de linking n'est pas faite. Enfin, une des raisons pour lesquelles GeoBench ne fonctionne plus sont les modifications d'API faites par ce fournisseur.

Enfin, les POI utilisés sont des points alors qu'il existe maintenant des POI plus précis sous forme de lignes ou polygones par exemple.

## 3 Objectifs

L'objectif général de ce POM sera de mettre à jour et de faire évoluer l'application GeoBench. Plusieurs sous-objectifs sont identifiables :

- État de l'art des fournisseurs composé de :
  - Limitations (nombre de requêtes en gratuit/commercial), types de requêtes, disponibilités, technologies...
  - Schémas des types de POI des fournisseurs
- Redéfinition des cas d'utilisation pour :
  - Mode « linking » où les liens entre entités équivalentes sont proposés
  - Mode « fusion » pour produire une seule entité unifiée
- Reprise et mise à jour du code :
  - Reprise du code pour qu'à l'avenir les fournisseurs puissent être changés/ajoutés plus facilement
  - Montée de version de Bootstrap et de JQuery ainsi que l'affichage des cartes
- Développement de nouvelles fonctionnalités :
  - Amélioration de l'interface utilisateur, e.g. selon les cas d'utilisation, possibilité de configurer l'application et les algorithmes
  - Ajout de nouveaux fournisseurs comme [Foursquare](#), [OpenStreetMap](#), [Wikimapia](#)...
  - Implémentation d'algorithmes de linking et de fusion, notamment basés sur l'apprentissage
  - Prédiction de valeurs manquantes
  - Stockage et export en [GeoJSON](#)

## 4 Diagramme de Gantt

En vue de ces différents objectifs, le diagramme de Gantt ci-dessous propose une organisation pour ce POM.

