# Définition d'une ontologie géographique pour les quartiers de Lyon

**Étudiants** : Florian BRAUNSTEIN et  Kévin GUTIERREZ

Dpt Informatique

Lyon 1

**Encadrants** : Fabien DUCHATEAU et Franck FAVETTA

## Contexte du projet :

La métropole de Lyon, peuplés de 1 639 558 habitants, est un pôle majeur attirant de nombreux projets innovants dans les domaines bancaires, financiers, commerciaux, technologiques, les arts et les divertissements. Devant ces nombreux enjeux économiques, la métropole de Lyon s'est donc vu dans l'obligation d'innover dans la recherche de logements et la satisfaction des biens immobiliers présentés à l'acheteur, qui est devenue une partie importante et crucial, afin de facilités chaque mutation. Le projet, bien qu'ayant un public ciblé assez large, s'intéressera surtout aux besoins des nouveaux étudiants et des employés ayant eu une mutation professionnelle ou universitaire dans la métropole de Lyon.

Le projet Home In Love, financé par le Labex Intelligence des Mondes Urbains rassemblant des chercheurs en informatique, psychologie et sociologie, a pour objectif de faciliter la recherche et le choix d'un bien immobilier. L'idée de ce projet est de récolter des informations relatives aux biens immobilier afin de pouvoir recommander suivant de nombreux paramètres du loueur ou acheteur extrait de ses attentes.

Ce projet POM s'inscrit dans ce contexte et vise à concevoir une application web intégrant des informations sur les quartiers de la métropole de Lyon (INSEE, ….) et recommandant des logements à l'utilisateur.

## Problématique :

Plusieurs difficultés se présentent quant à la réalisation de ce projet : Le choix de représentation des donnée pour l'ontologie du projet et celle de l'ontologie spatial, l'intégration et le format des documents existant et pouvant être utilisé et enfin choisir l'algorithme de recommandation le plus performant et cohérent avec notre projet.

Les ontologies traitant de la recherche de logements étant inexistante, le projet vise donc à s'inspirer des autres ontologies traitant de sujet différents pour se faire une idée de notre ontologie, de trouver les relations importantes entre les différents lieux de la métropole pour chaque quartier, en s'inspirant des concepts des ontologies éxistantes telles que Geofla, Geonames ontology ou encore LinkedGeoData.

Concernant l'intégration de données, il faut peupler notre ontologie avec les informations existantes sur les quartiers tels que sa description (noms, limites géographiques, population), des statistiques (par exemple, nombre de personnes de moins de 25 ans, nombre de salariés entre 40 et 50 ans, nombre d'écoles). D'autres statistiques pourront être estimées (e.g., nombre de commerces, surface d'espaces verts), mais aussi détecter des quartiers qui pourront être décrit comme calme, dynamique ou bruyant suivant des cartes des bruits de la métropoles ou d'autre données significatives.

Enfin, le dernier objectif est de recommander un quartier à partir d'un profil. Ces profils sont actuellement disponibles sous forme semi-structurée (quelques propriétés, dont certaines sont associées à un texte potentiellement long). Un algorithme d'apprentissage sera utilisé pour rapprocher un profil à un ou plusieurs quartiers.

## Exemple de profils traités :

* "**ID** : 583ea2d45f288a000fcef607"
  "**Nombre de pièces** : 6"
  "**Surface** : 105m²"
  "**Budget** : 1200 euros"
  "**Description du bien recherché** : maison ou appartement contemporain(e), sans travaux avec min 3 chambres avec une 100 aine de M2 et un petit jardin, salle d'eau / cuisine équipée ouverte sur salon. garage"
  "**Description du futur quartier** : quartier calme résidentiel / pas trop bruyant avec des commodités pas loin ( boulangerie / supérette, écoles maternelle et primaire pas loin du lieu de résidence )"
  "**Projet client, attentes** : Mutation de Dunkerque à FOS SUR MER via ARCELORMITTAL j'ai 2 enfants : 2 ans et 6 ans,je vis avec mes 2 enfants et la maman habite à Paris"

* "**ID** : 583c0b1c5f288a000fcef5ae"
  "**Nombre de pièces** : 3"
  "**Surface** : 75m²"
  "**Budget** : 900 euros"
  "**Description du bien recherché** : Au minimum 80m², 2 chambres, immeuble propre et absolument un garage pour nos 3 motos + la voiture (donc soit un garage double type box soit une place de parking + 1 box fermé),une cave et bcp de rangement serait l'idéal. Un appartement ou maison F3 avec minimum une surface à minima proche des 75/80 M2 et max 90 M2. Coût total loyer charges comprises 900/950 € maximum. Un garage (box car motos) double ou un box et une place de parking. Chauffage individuel gaz ou électrique. Si possible balcon ou terrasse, cave et cuisine équipée. Pas de RDC avec accès direct sur rue"
  "**Description du futur quartier** : proche commerce et surtout transport (bus, métro, tram). Les secteurs favoris seraient jean macé mais pas gerland,

proche part dieu type dauphiné lacassagne, reconnaissance balzac c'est à dire proche du tram, oullins proche métro caluire proche but ,nous n'aimons pas l'est lyonnais type meyzieu, décines. Avec proximité métro, tram proche de la part dieu ou de saxe-Gambetta ou accessible à pied( 30 minutes maximum). Quartiers : Vaise, Lyon 7ème (Jean-Macé) 8ème (Sans-Souci, Monplaisir) 3ème (Montchat entre autre, Lacassagne), Villeurbanne , Oullins, Caluire"
"**Projet client, attentes** : Nous venons de Dijon et sommes mutés à Lyon pour la société Enedis. Je suis cadre RH en charge de la mobilité sur Lyon. A ce jour, nous sommes dans un appart hotel situé à saxe gambetta. Mon entreprise Enedis est situé rue duguesclin, Mon mari, travaillant en logistique, est à ce jour en intérim et est très mobile, nous vivons à ce jour dans un grande maison que nous venons de vendre et avons 3 motos. Nous venons en ville pour des raisons professionnelles mais désirons profiter des infrastructures urbaines."

* "**ID** : 5849a7915f288a000fcef69e"
"**Nombre de pièces** : 2"
"**Surface** : 35m²"
"**Budget** : 500 euros"
"**Description du bien recherché** : Appartement avec une chambre, non meublé situé en centre-ville"
"**Description du futur quartier** : Pleins de commerce à proximité, quelques bars, station vélov"
"**Projet client, attentes** : Sympathique, célibataire, étudiant, "

* "**ID** : 593abc1f5d86cf000f237640"
"**Nombre de pièces** : 6"
"**Surface** : 120m²"
"**Budget** : 1300 euros"
"**Description du bien recherché** : Maison récente avec du terrain et 3 chambres et un bureau et idéalement avec une piscine "
"**Description du futur quartier** : Résidentiel calme proche écoles et du bourg"
"**Projet client, attentes** : Mariée avec 1 fille de 4 ans à scolariser donc bien proche des écoles. On aime le sport et vivre en dehors de la ville "

## Intégration :

Pour l'intégration des données, nous avons choisis de simplifier l'ontologie en un simple fichier JSON, que nous avons préalablement convertis en json grâce à un script python, car cela simplifiera les manipulation diverse à venir dans le cadre de ce projet.

Ensuite il a donc fallu peupler notre base de données, avec des informations existantes sur les quartiers de la métropole de LYON (description du quartier) et les différentes statistiques existantes sur les quartiers afin d'en extraire de nouveaux concepts tel que la dynamique ou non d'un quartier, si celui-ci est calme, etc ...
Ces informations concerne les IRIS INSEE, que nous avons choisis comme "pieds d'édifice" pour notre modèle d'intégration.

Concernants les données, la plupart proviennent de jeux de données INSEE que nous avons dû convertir pour les intégrer à la base de donnée JSON tout en résolvant les problèmes suivants :

- Conflits d'identité avec les données déjà intégrées
- Suppression des données superflues, redondantes ou inutiles
- Données non existantes dans les jeux de données INSEE

## Recommandation :

L'intégration des données réalisés, l'étape de la recommandations de quartiers (IRIS) selon le profil de l'utilisateur à partir des résultats obtenus pouvait commencer.

Nous avons donc définis des vecteurs de données correspondant à des critères que nous avons retrouvés dans les profils tel que les loisirs, l'éducation ou la sécurité, et qui sont déterminés par des statistiques de la base de donnée.
Par exemple, un IRIS peut être plus ou moins sécurisé selon le nombre de poste de police qu'il comporte , dynamique si il y a des commerces aux alentours, ou bruyant si il y a des sites touristiques, des bars ou encore des restaurants.

L'algorithme de recommandation fonctionne suivant ce cheminement :

- On utilise des profils existants dont on connaît le quartier qui a été choisi
- On utilise ce résultat pour entraîner l'algorithme avec plusieurs itérations
- On détermine donc à la fin quel est le ou les meilleurs quartier à recommander