

Définition d’une ontologie géographique pour les quartiers de Lyon

Florian BRAUNSTEIN, Kévin GUTIERREZ

Contexte

- Concevoir une **application** qui **intègre** des informations sur les quartiers de de Lyon et alentours
- Recommander** un ou plusieurs quartiers pertinents pour un profil d'utilisateur donné.
- Initié par le projet **Home In Love** qui rassemble des chercheurs en informatique, psychologie et sociologie, et a pour objectif de **faciliter** la recherche et le choix d'un bien immobilier
- Demander à l'acheteur ou au loueur de se **décrire** et de **décrire le bien** recherché, puis de faire des **recommandations** de biens en se basant sur ces informations (voir Figure 1)

Problématique

- Choix du modèle de données** pour représenter les informations des quartiers et des biens immobiliers
- Intégration** de sources de données dans une nouvelle ontologie pour visualiser les quartiers de Lyon et sa métropole tout en réutilisant les concepts des **ontologies géographiques déjà présentes** telles que **Geofla**, **Geonames ontology** ou encore **LinkedGeoData**.
- Algorithme de **recommandation** d'un quartier grâce à l'intégration des données, des profils avec le résultat du quartier choisi, et à des algorithmes de recommandation existants

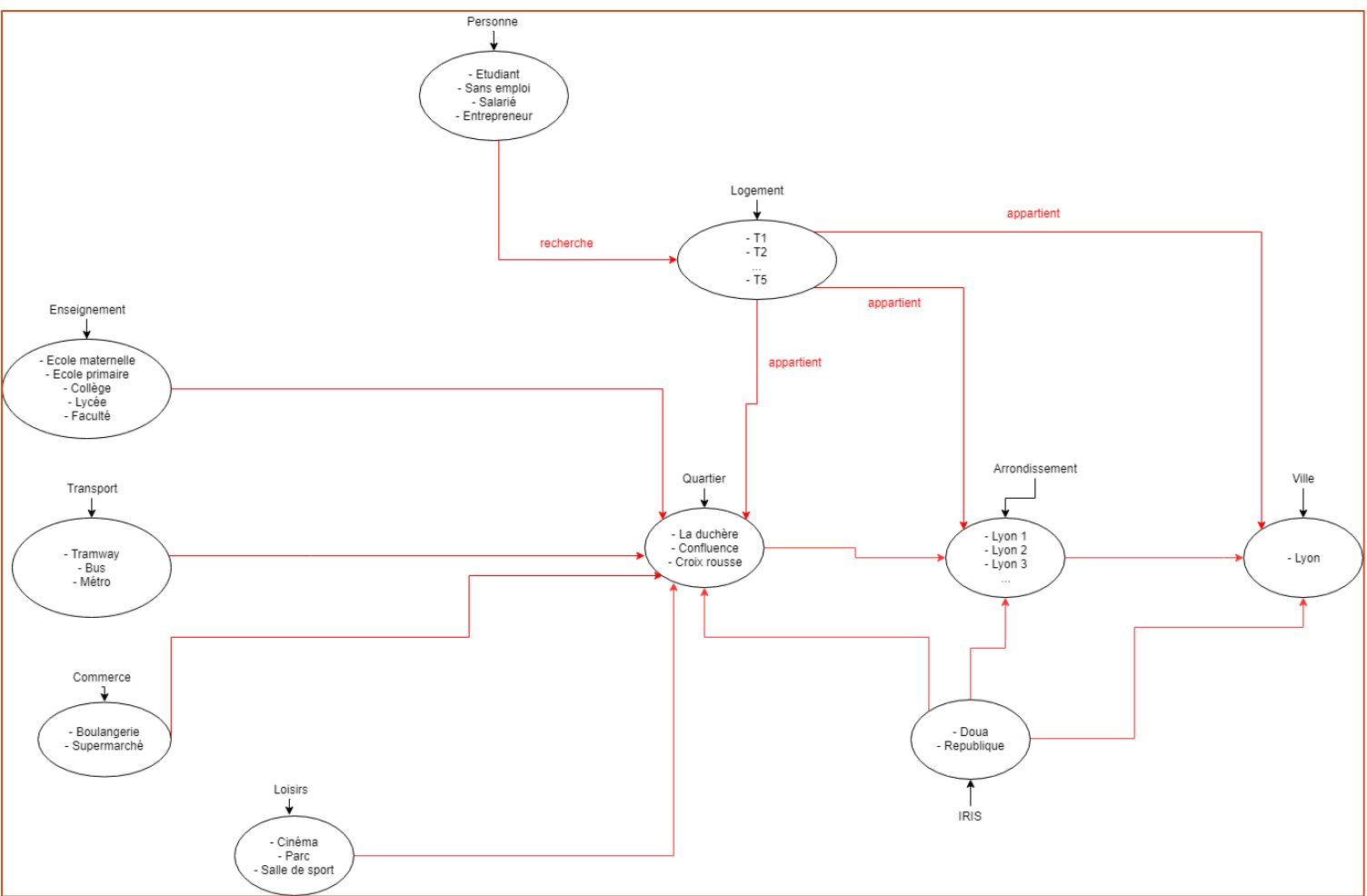


Figure 2. Modèle initiale de l'ontologie

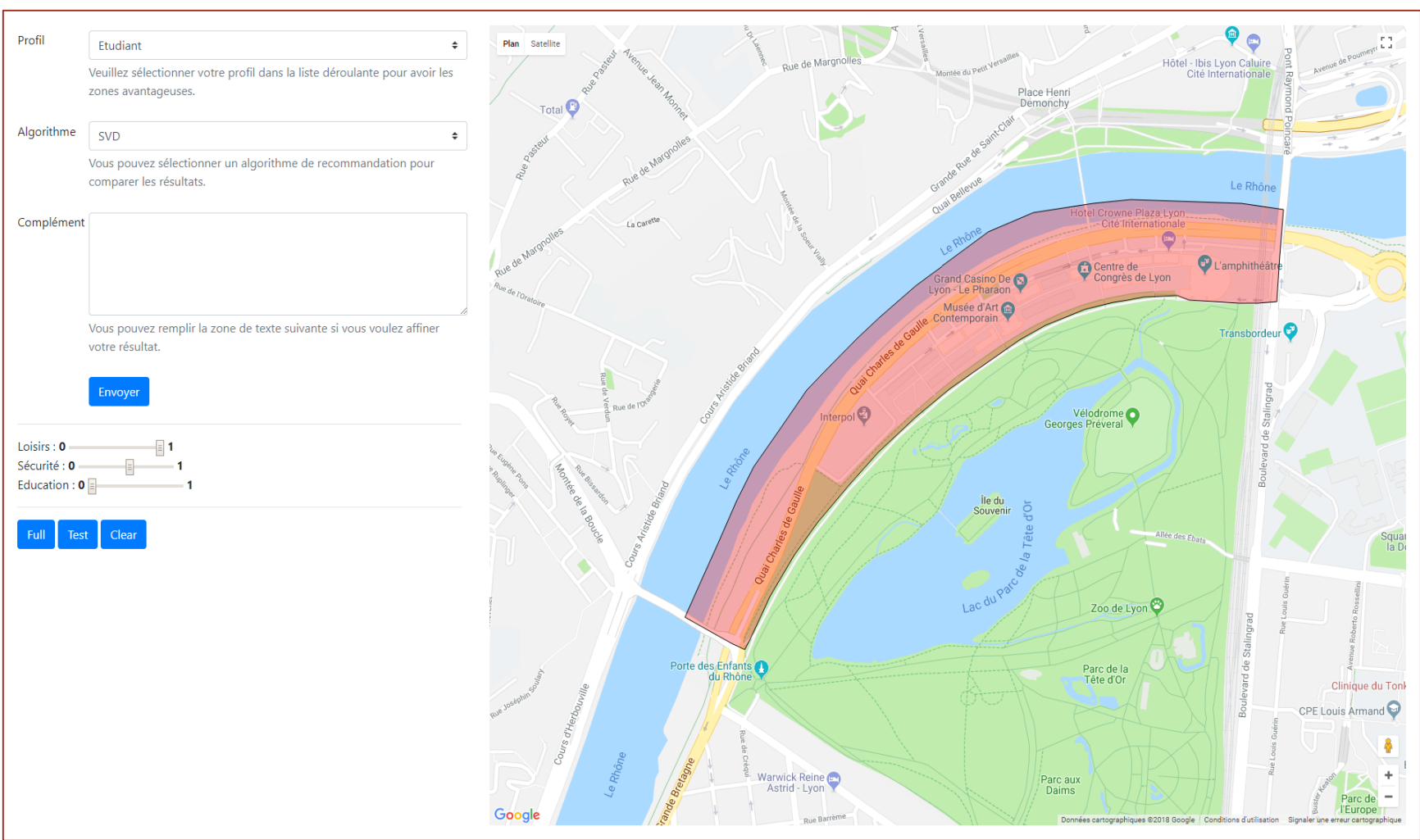


Figure 3. Prototype de l'application

```
{
  "ID": "596f7bdb329bbd000fcec69a",
  "Nombre de pièces": 2,
  "Surface": "30m²",
  "Budget": "700 euros",
  "Description du bien recherché": "30 M² minimum\nLoyer entre 600 et 700€ maximum \nBalcon ou terrasse si possible\nMeublé ou non meublé\nWifi\nEmménagement pour le 15 août ou 28/08 maximum ",
  "Description du futur quartier": "Etudiant à l'INSA et entreprise à Bron : trouver un compromis pour faciliter ces déplacements\nLogement sur ligne métro D ",
  "Projet client, attentes": "Mobilité de Bouc Bel Air à Lyon\nEtudiant en alternance"
}

{
  "ID" : "599acefe20ee1e000fe4c173",
  "Nombre de pièces" : 1,
  "Surface" : "25m²",
  "Budget" : "400 euros",
  "Description du bien recherché" : "Je recherche studio qui soit meublé (je ne compte rester qu'un an) si possible neuf ou du moins récent et très lumineux. La taille n'a pas beaucoup d'importance du moment que l'endroit est agréable à vivre. \n\nLe must serait d'avoir de quoi cuisiner correctement (cela ne reste qu'un plus).",
  "Description du futur quartier" : "Je cherche un quartier dans le centre, à proximité de mon travail car je me déplace en transport en commun. J'aimerais qu'il soit tranquille et me sentir en sécurité."
```

Figure 1. Exemple de profils utilisateurs

Intégration

- Proposition d'une ontologie (voir Figure 2) simplifiée ensuite au format JSON, plus facile à manipuler dans le cadre de ce projet.
- Peupler cette base de données avec des informations existantes sur les quartiers telles que des informations sur la description du quartier ou des statistiques.
- IRIS (INSEE) choisi comme "brique" de base pour notre modèle d'intégration.
- Sources** des données : principalement jeux de données INSEE convertis pour les intégrer à la base de donnée JSON
- Résolution des problèmes** d'intégration 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

- Définitions de **critères** selon les données intégrées : **loisirs**, **sécurité** et **éducation** (voir panel de gauche sur Figure 3)
- Annotation** des profils existants selon ces critères
- Implémentation** (voir Figure 3) :
 - Utilisation de la librairie python **Surpriselib**
 - Définition de **vecteurs de données** selon les critères établis
 - Utilisation de **profils existants** dont on connaît le **quartier choisi** afin d'**entraîner l'algorithme** avec plusieurs **itérations**
 - Résultats des itérations **précédentes** utilisées afin d'entraîner l'algorithme pour les itérations **suivantes**
 - Utilisation d'algorithme existants tel que **SVD**, **NMF** et **Slope One**
- Résultats** peu concluants, sur **13** recommandations attendues, seulement **4** ont un résultat acceptable