

Sujet 4 : Agent conversationnel pour l'analyse géographique

Description :

La thématique principale de ce projet est de développer un outil interactif qui permet d'interroger et de visualiser des données géographiques complexes (RGPD, limites de communes, ...) en utilisant le langage naturel. L'outil traduira les requêtes de l'utilisateur en cartes, en graphiques ou en tableaux pour en faciliter la compréhension.

Les étudiants devront mettre en place une interface de visualisation solide, capable de gérer de grandes quantités de données géographiques.

- **Ingestion et préparation des données** : Téléchargez et traitez les données du Registre Graphique Parcellaire (RPG), ainsi que d'autres sources comme les limites des communes (INSEE) ou les données de consommation de pesticides. C'est un défi en soi, car ces données sont massives et nécessitent une bonne maîtrise de bibliothèques comme GeoPandas.
- **Cartographie interactive** : Créez une application web (avec Streamlit ou Dash) qui affiche une carte interactive. Utilisez des bibliothèques comme Folium ou Pydeck pour superposer les différentes couches de données (parcelles, frontières communales, etc.).
- **Filtres avancés** : L'interface doit permettre de filtrer les données par année, par type de culture, par commune ou par d'autres attributs.

Une fois l'interface en place, il devra être possible d'interagir avec en langage naturel.

Objectif :

L'objectif est de concevoir une plateforme pour :

1. Préparation des données et création de l'interface de base.
2. Compréhension du langage naturel : développez un module capable d'analyser et de détecter le type de requête de l'utilisateur :
 - Requête de visualisation : l'utilisateur demande à voir des données sur une carte ("Affiche-moi les parcelles de vignobles dans la région de Colmar").
 - Requête d'agrégation : l'utilisateur demande des statistiques ou un graphique ("Combien de parcelles de vignobles par communes dans le Haut-Rhin ?").
 - Requête de comparaison : l'utilisateur demande une analyse comparative ("Quelle est l'évolution des parcelles de vignobles entre 2015 et 2020 dans le Bas-Rhin ?").
3. Traduction en code : après analyse, le système traduira la requête en langage naturel en script Python capable de générer le résultat. Par exemple, la requête "Affiche-moi les parcelles de vignobles" serait traduite en une requête de base de données ou un filtre qui ne sélectionne que les parcelles de vigne.
4. Génération de cartes, de graphiques et d'analyses : en fonction du type de requête détecté, le système génère le résultat approprié : cartes interactives pour affichage, statistiques, graphiques, ...
5. Analyse et explication : le système pourra également générer une courte analyse en langage naturel pour accompagner les graphiques et cartes. Par exemple, si l'outil montre une baisse du nombre de vignobles, l'agent pourrait résumer les chiffres clés et attirer l'attention de l'utilisateur sur les zones les plus touchées.