

IFT-6005 Rapport de suivi d'avancement

Agent conversationnel Open Food Facts

Alain Boisvert
Matricule: 994 029 313

13 février 2025

1 Rappel du projet

Le projet vise à développer un agent conversationnel permettant d'interroger la base de données Open Food Facts. L'objectif est de permettre aux utilisateurs de poser des questions en langage naturel sur les produits alimentaires, sans avoir besoin de connaissances en SQL ou en bases de données.

Le système repose sur une architecture modulaire comprenant :

- Un module de dialogue basé sur un LLM
- Un convertisseur de langage naturel vers SQL
- Un connecteur de base de données DuckDB
- Un générateur de réponses naturelles

2 Rappel des tâches planifiées

Selon la planification initiale, les tâches à réaliser sont regroupées en cinq phases principales :

1. Préparation et configuration (30h)
 - Configuration de l'environnement
 - Préparation de la base de données
 - Création du jeu de test
 - Implémentation des métriques d'évaluation
2. Développement du système de base (75h)
3. Rapport de mi-session (25h)
4. Optimisations et fonctionnalités avancées (60h)
5. Documentation et finalisation (25h)

3 Avancement des tâches

3.1 Tâches complétées

- Mise en place de l’environnement de développement (Visual Studio Code, Python, L^AT_EX, GitHub)
- Création de la base de données DuckDB pour Open Food Facts à partir du fichier Parquet
- Exploration et analyse de la structure de la base de données
- Développement d’un premier agent conversationnel utilisant smolagents
 - Capable de répondre à des requêtes simples
 - Outil pour interroger une base de données DuckDB
 - Outil pour chercher des informations sur le site Web du Guide alimentaire canadien

3.2 Tâches en cours

- Documentation détaillée de la base de données pour améliorer la compréhension du modèle
- Apprentissage de smolagents
- Optimisation des prompts pour réduire les coûts d’utilisation

3.3 Tâches à débiter

- Création du jeu de test de 100 questions
- Implémentation des scripts d’évaluation (EX, TCM, TRM)
- Lire plus attentivement les articles sélectionnés dans la revue de littérature et identifier les points importants à retenir pour le présent projet

4 Problèmes rencontrés

4.1 Gestion des données volumineuses

La taille importante du fichier Parquet source a posé des défis techniques pour le traitement sur mon MacBook Pro M1 16 Go RAM. La gestion efficace de ce volume de données nécessitait une approche optimisée.

4.2 Sélection des technologies

Le choix du framework pour l’agent conversationnel a nécessité une analyse comparative approfondie, notamment entre CrewAI et smolagents. Les critères de sélection incluaient :

- La facilité d’intégration
- Les capacités d’exécution de code
- La flexibilité du système

4.3 Performance des modèles

Les premiers tests avec des modèles plus légers (Qwen2-7B-Instruct) ont révélé des limitations dans l'interprétation des prompts et l'interaction avec DuckDB, nécessitant l'utilisation de modèles plus puissants mais plus coûteux.

5 Solutions proposées

5.1 Gestion des données

Adoption de DuckDB comme solution de base de données, offrant :

- Une excellente performance pour les requêtes SQL
- Une gestion efficace de la mémoire
- Une intégration simple avec Python

5.2 Framework d'agent

Choix de smolagents de Hugging Face pour :

- Sa simplicité d'utilisation
- Son approche optimisée pour les "agents de code", c'est-à-dire les agents qui génèrent et exécutent du code Python

5.3 Optimisation des modèles

- Utilisation temporaire de Claude Sonnet pour le développement initial
- Travail en cours sur l'optimisation des prompts pour permettre l'utilisation de modèles plus légers
- Documentation améliorée de la base de données pour faciliter la compréhension du modèle

6 Plan d'action

Pour les deux prochaines semaines, les priorités sont :

1. Création du jeu de test
 - Développer 100 questions de référence
 - Assurer une diversité des cas d'utilisation
 - Inclure des questions en français et en anglais
2. Implémentation des métriques d'évaluation
 - Précision d'exécution (EX)
 - Taux de couverture des données manquantes (TCM)
 - Temps de réponse moyen (TRM)

3. Optimisation des prompts
 - Réduire la dépendance aux modèles LLM coûteux
 - Améliorer la précision des réponses
4. Documentation de la base de données
 - Compléter la documentation des colonnes de la table sur les produits alimentaires
 - Optimiser les instructions données au LLM pour les requêtes SQL
5. Revue de la littérature
 - Lire attentivement les articles sélectionnés
 - Identifier les points importants à retenir pour le projet

7 Conclusion

Le projet progresse selon le plan initial. Les défis rencontrés sont gérés au fur et à mesure.

Les prochaines étapes se concentreront sur la création du jeu de test et l'implémentation des métriques d'évaluation. L'optimisation des prompts et la documentation approfondie de la base de données permettront d'améliorer les performances tout en réduisant les coûts d'exploitation.

La base technique établie jusqu'à présent devrait fournir une base solide pour le développement des fonctionnalités plus avancées prévues dans les phases ultérieures du projet.