

Plan de Travail pour le Projet LLM Éducatif

1 Récupération des Relations de Prérequis Conceptuels à partir des Dépendances Universitaires

1.1 Objectif

L'article traite de la récupération des relations de prérequis entre concepts dans le cadre des cours universitaires. L'objectif est de construire un graphe des concepts qui montre les notions devant être maîtrisées avant d'aborder d'autres sujets. Cette approche utilise des sources telles que Wikipédia pour identifier ces relations entre les concepts partagés universellement.

1.2 Dépendances entre les Cours Universitaires

- Les chercheurs utilisent Wikipédia pour identifier les relations de prérequis entre concepts de manière universelle.
- L'objectif est de construire un graphe des concepts en relation avec les dépendances entre les cours (par exemple, un cours d'algèbre linéaire est un prérequis pour un cours de géométrie analytique).

2 Problème Posé : CPR-Recover (Concept Prerequisite Relation Recovery)

L'article introduit le problème CPR-Recover, qui consiste à reconstruire un graphe des relations de prérequis entre des concepts enseignés dans différents cours universitaires.

2.1 Entrées et Sorties

- **Entrées :**
 - Une liste de cours avec leurs dépendances (par exemple, Cours 2 dépend de Cours 1).
 - Une représentation des concepts enseignés dans chaque cours.
- **Sortie :**
 - Un graphe des concepts montrant les relations de prérequis entre eux.

3 Hypothèses de Base du Modèle

La méthode repose sur deux hypothèses fondamentales :

- **Causalité** : Si un cours dépend d'un autre, les concepts enseignés dans le premier sont nécessaires pour comprendre le second.
- **Parcimonie** : Le graphe des concepts doit rester simple, avec un nombre réduit de relations de prérequis par rapport au nombre total possible de paires.

3.1 Pourquoi ces hypothèses sont-elles importantes ?

- **Causalité** : Elle permet d'établir des liens logiques entre les concepts, facilitant ainsi la construction du graphe.
- **Parcimonie** : Elle permet d'éviter d'ajouter des relations inutiles, assurant la simplicité et la lisibilité du graphe.

4 Méthodologie : Fonctionnement de CPR-Recover

- Chaque cours est représenté par un vecteur de concepts dérivés des articles Wikipédia.
- Le modèle cherche à minimiser une fonction mathématique pour créer le graphe des concepts le plus simple.
- **Contraintes utilisées** :
 - Parcimonie
 - Causalité

5 Utilisation de l'API Groq pour la Gestion des Réponses des PDFs

L'API Groq, utilisée pour le graphe de connaissances, peut être intégrée à ce projet pour :

- Extraire des informations pertinentes à partir des documents PDF téléchargés.
- Connecter des concepts issus des textes et générer des réponses ou recommandations basées sur ces relations.

5.1 Exemple d'application

L'API permet de traiter les questions des utilisateurs sur des documents académiques et de fournir des réponses en utilisant les relations conceptuelles extraites des PDFs.

6 Stockage des Embeddings dans Chroma DB

Chroma DB est utilisé pour stocker les embeddings des textes extraits des PDFs. Un embedding est une représentation vectorielle d'un morceau de texte, permettant des recherches rapides et pertinentes.

- Ces embeddings permettent de trouver des passages pertinents lorsque l'utilisateur pose une question.
- L'API peut interroger la base de données pour obtenir les résultats les plus proches en fonction de l'embedding de la question posée.

6.1 Exemple d'Embedding

Texte : *Le chat mange.*

En Chroma DB, l'embedding du texte pourrait être :

$(0.25, 0.42, -0.10, 0.56, -0.31)$

Ce vecteur permet une recherche rapide la prochaine fois que des informations similaires sont demandées.

7 Intégration du Prompt Engineering

Le prompt engineering joue un rôle essentiel dans la création de modèles LLM efficaces. Pour optimiser les performances, il est important de concevoir des prompts adaptés aux tâches spécifiques.

- Créez des **prompts clairs et spécifiques** pour chaque tâche (par exemple, questions, quiz, générateurs d'exemples, etc.).
- Utilisez des **instructions détaillées** pour guider le modèle dans la génération de réponses pertinentes et précises.

7.1 Exemples de bonnes pratiques en prompt engineering

- Spécifiez le format attendu de la réponse (par exemple, liste, texte, ou tableau).
- Utilisez des contraintes dans les prompts pour guider le modèle vers des réponses plus adaptées au contexte éducatif.