

Contexte

Dans le cadre d'un projet d'assistance intelligente à la formulation en boulangerie et pâtisserie, nous disposons d'un ensemble de fiches techniques d'ingrédients et d'additifs (enzymes, améliorants, agents oxydants, etc.).

Ces fiches ont déjà été :

- Converties en texte
- Découpées en fragments (chunks)
- Transformées en embeddings
- Stockées dans une base PostgreSQL

Votre mission consiste à développer le module de **recherche sémantique** permettant d'interroger cette base vectorielle.(RAG)



Ressources Fournies

Les participants recevront :

Base de données PostgreSQL

Table : `embeddings`

Structure :

- `id` (Primary Key)
- `id_document` (int)
- `texte_fragment` (text)
- `vecteur` (VECTOR(384))

La table contient déjà les embeddings des fiches techniques (ex : enzymes, agents oxydants, améliorants de panification).

Modèle d'embedding officiel

Modèle imposé :

`all-MiniLM-L6-v2`

Bibliothèque : `sentence-transformers`

Dimension : 384

⚠ L'embedding de la question doit être généré avec ce modèle.

Paramètres imposés

- Méthode de similarité : **Cosine Similarity**
 - Nombre de résultats à retourner : **Top K = 3**
 - Langage recommandé : Python
-

Travail demandé

Développer un module capable de :

1. Recevoir une question utilisateur
 2. Générer l'embedding correspondant
 3. Calculer la similarité cosinus entre :
 - o L'embedding de la question
 - o Les embeddings stockés en base
 4. Classer les résultats par score décroissant
 5. Retourner les **3 fragments les plus pertinents**
 6. Afficher :
 - o Le texte du fragment
 - o Le score de similarité
-

Exemple de question

Améliorant de panification : quelles sont les quantités recommandées d'alpha-amylase, xylanase et d'Acide ascorbique ?

Résultat attendu (format indicatif)

Résultat 1

Texte : “Dosage recommandé : 0.005% à 0.02% du poids de farine.”

Score : 0.91

Résultat 2

Texte : “Alpha-amylase : utilisation entre 5 et 20 ppm selon la farine.”

Score : 0.87

Résultat 3

Texte : “Xylanase : améliore l'extensibilité de la pâte...”

Score : 0.82