

LAMA-WeST Lab

Artificial Intelligence
Natural language processing
Semantic web



FAISS (Facebook AI Similarity Search)

INF8460 - Traitement automatique de la langue naturelle

Karou Diallo
Polytechnique Montréal

22 octobre 2024

Présentation

Utilisation de FAISS



Présentation

Utilisation de FAISS



► **Définition :**

- ※ FAISS (Facebook AI Similarity Search) est une bibliothèque optimisée pour rechercher rapidement des vecteurs similaires dans de grandes bases de données.
- ※ FAISS résout les limitations des moteurs de recherche traditionnels qui sont optimisés pour les recherches basées sur des tables de hachage
- ※ FAISS offre des fonctions de recherche de similarité plus évolutives.



La bibliothèque d'indexage FAISS est utilisé dans un pipeline ayant trois phases :

1. Le modèle de plongements (embedding-model) qui fournit les vecteurs de plongements d'un document
2. La bibliothèque FAISS qui est utilisé pour sauvegarder et indexer ces vecteurs
3. La recherche de similarité : le modèle de plongements qui fournit le plongement d'une requête reçue et le passe à l'index qui retourne les k documents les plus pertinents



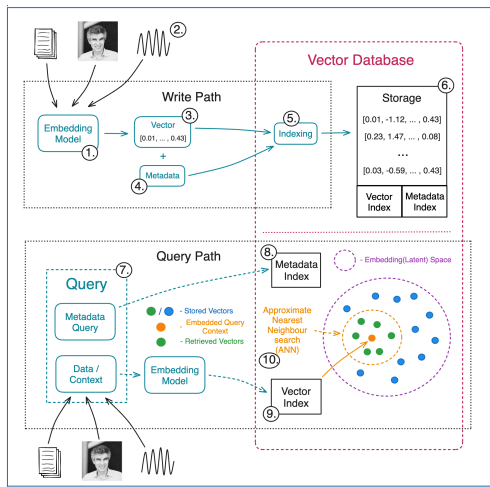


Figure – Vue globale de système de recherche de documents pertinent pour une requête en utilisant l’indexage. Source de l’image originale.



```
[ ]: import pandas as pd
import numpy as np
import faiss
from sentence_transformers import SentenceTransformer

[5]: data = [
    ['What are the programs offered by Polytechnique Montreal?', 'registrar'],
    ['Does PolyMTL offer funding for student projects?', 'finance'],
    ['Recommend me a good Italian restaurant near the PolyMTL campus', 'food'],
    ['How can I reset my password on Moodle?', 'tech support'],
    ['Tell me an engineer joke', 'entertainment'],
    ['What are the symptoms of the flu?', 'health'],
    ['What is the score of the Polytechnique football game?', 'sports'],
    ['I feel motivated for studies today', 'personal emotion']
]

df = pd.DataFrame(data, columns=['text', 'category'])
df.head()
```

Figure – Import des bibliothèques et initialisation des documents



```
[5]:
```

	text	category
0	What are the programs offered by Polytechnique...	registrar
1	Does PolyMTL offer funding for student projects?	finance
2	Recommend me a good Italian restaurant near th...	food
3	How can I reset my password on Moodle?	tech support
4	Tell me an engineer joke	entertainment

```
[ ]: text = df['text']  
encoder = SentenceTransformer("paraphrase-mpnet-base-v2") # bert-base-nli-mean-tokens  
embeddings = encoder.encode(text)
```

```
[7]: embeddings.shape
```

```
[7]: (8, 768)
```

Figure – Encodage des documents en utilisant un modèle de plongements




```
[8]: vector_dimensions = embeddings.shape[1]

index = faiss.IndexFlatL2(vector_dimensions)
faiss.normalize_L2(embeddings)
index.add(embeddings)

•[9]: search_text = 'are all PhD programs fully funded at Polytechnique Montreal ?'
search_vector = encoder.encode(search_text)
new_vector = np.array([search_vector])
faiss.normalize_L2(new_vector)

[13]: distances,ann = index.search(new_vector,k=3)
results = pd.DataFrame({'distances': distances[0], 'indexes': ann[0]})
df_merged = pd.merge(results,df,left_on='indexes',right_index=True)

[14]: df_merged.head()
```

	distances	indexes	text	category
0	0.665727	0	What are the programs offered by Polytechnique...	registrar
1	0.765357	1	Does PolyMTL offer funding for student projects?	finance
2	1.477638	2	Recommend me a good Italian restaurant near th...	food

Figure – Initialisation → Indexation → Recherche de similarité



► Références

- ※ <https://engineering.fb.com/2017/03/29/data-infrastructure/faiss-a-library-for-efficient-similarity-search/>
- ※ <https://ai.meta.com/tools/faiss/>
- ※ <https://www.newsletter.swirlai.com/p/sai-notes-07-what-is-a-vector-database>

