

Participation à ReNeuIR@SIGIR24

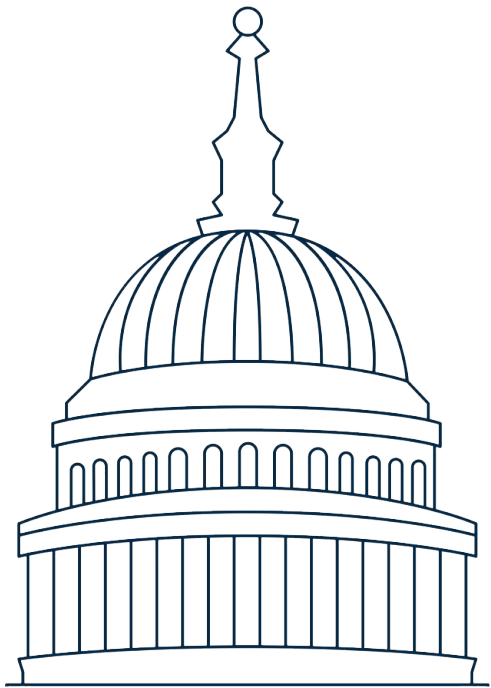
Soutenance de TER

Auteur : Nouh CHELGHAM
Encadrant : José MORENO
Binôme : Alexandra Villon

Contenue

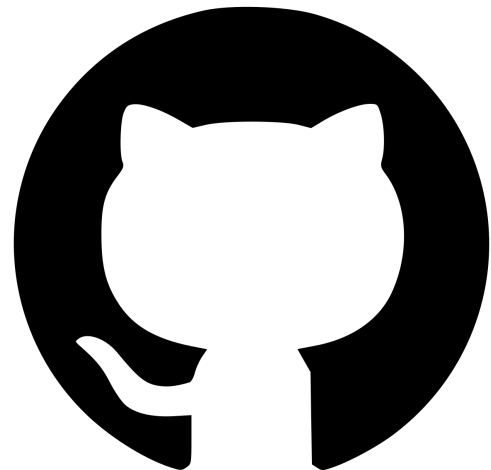
- Introduction
- Indexation
- Retrieve
- Re-Rank
- Recherche et Pipeline Final
- Conclusion

ReNeuIR@SIGIR24



**SIGIR
2024**

Washington, D.C.





● INDEXATION

- Introduction
- **Indexation**
- Retrieve
- Re-Rank
- Recherche et Pipeline Final
- Conclusion



Prétraitement

SpaCy extrait les tokens, lemmes, et stopwords pour structurer les documents avant l'indexation

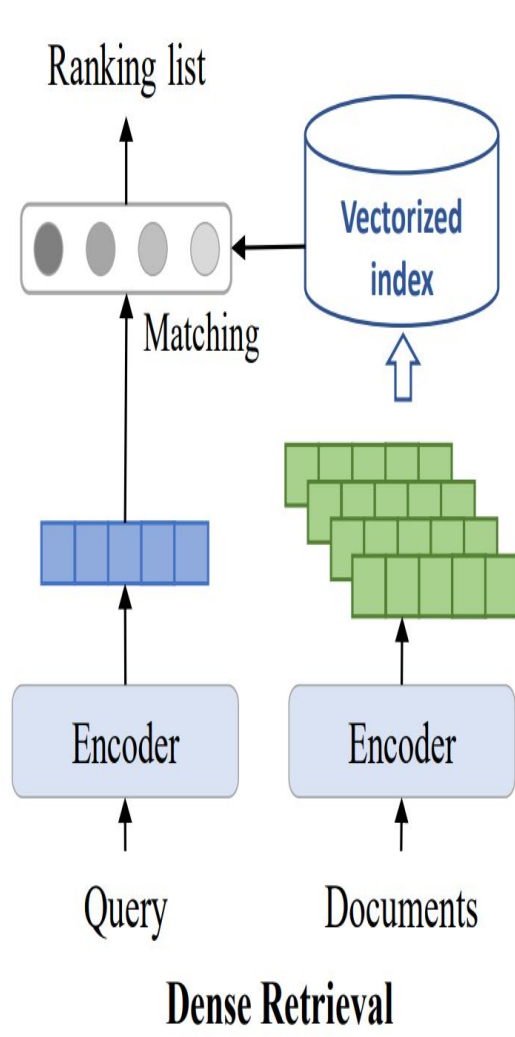
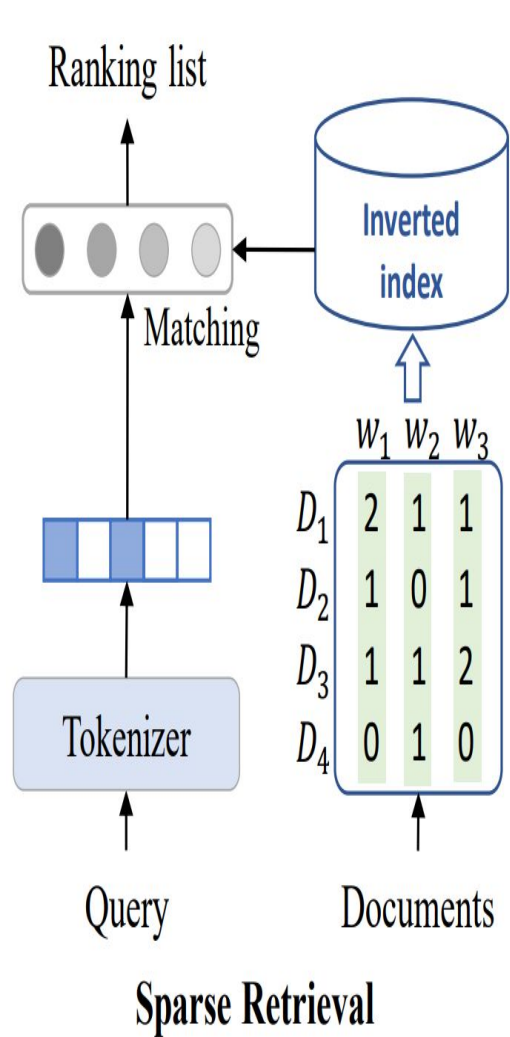


Index Inversé

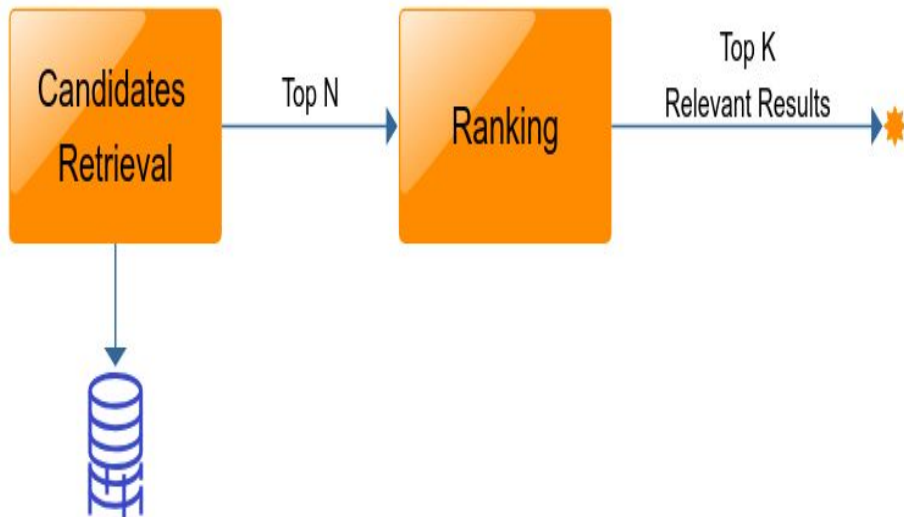
PyTerrier crée un index inversé en mappant les termes aux documents pour une recherche rapide

● Retrieve

- Introduction
- Indexation
- Retrieve
- Re-Rank
- Recherche et Pipeline Final
- Conclusion



la méthode sparse utilise des vecteurs de mots et un index inversé pour trouver des documents en fonction de la correspondance exacte des termes, tandis que la méthode dense encode les requêtes et documents en vecteurs continus, évaluant leur similarité dans un espace vectoriel pour un matching plus complexe.



Pipeline

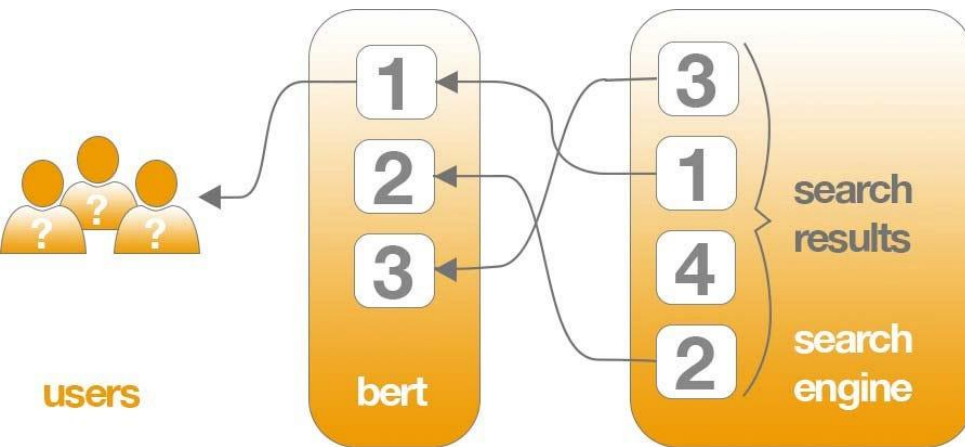
Une couche sparse initiale pour récupérer les candidats, puis une deuxième couche dense pour les mieux organiser.

● Re-Rank

- Introduction
- Indexation
- Retrieve
- Re-Rank
- Recherche et Pipeline Final
- Conclusion

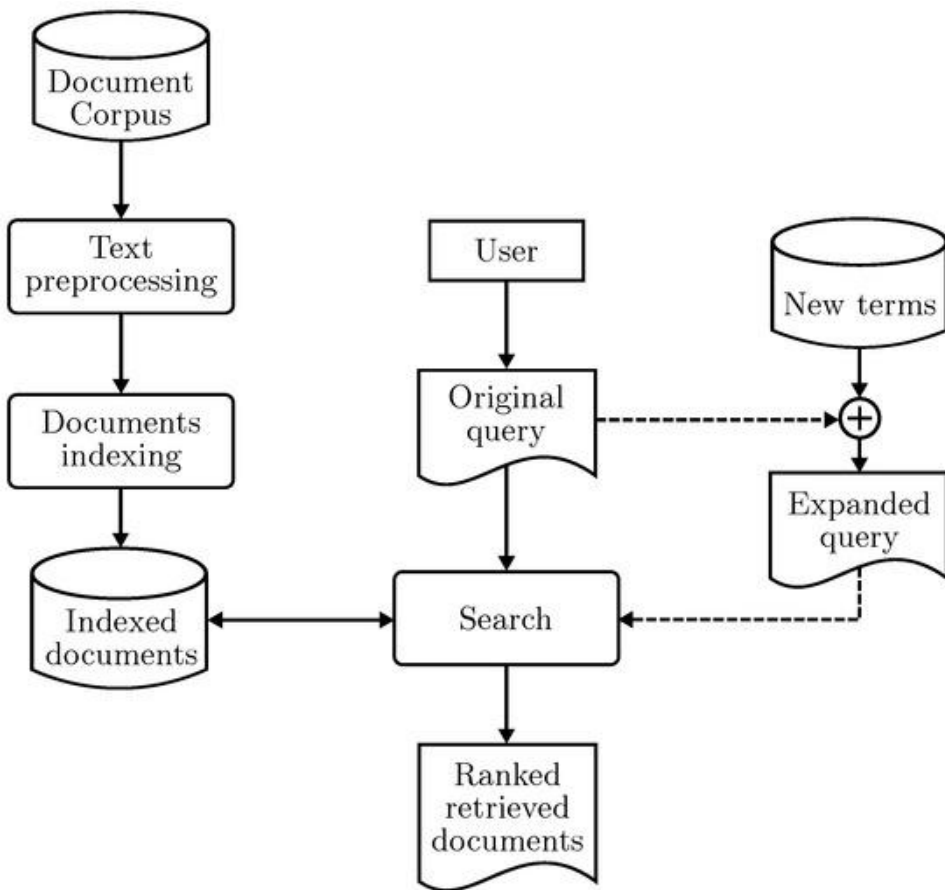


Hugging Face



PLM

Utilisation de modèles
pré-entraînés de Hugging Face
pour classer et retourner les
résultats les plus pertinents.



Query Expansion

Processus d'expansion de requête : enrichir la requête originale avec de nouveaux termes pour améliorer les résultats

● CHERCHE et PIPELINE FINAL

- Introduction
- Indexation
- Retrieve
- Re-Rank
- Cherche et Pipeline Final
- Conclusion

CHERCHE



Résultats obtenus

Méthodes utilisées avec CHERCHE	DL1000	
Lunr	0.4851	
TF-IDF	0.4423	
Lunr + paraphrase-albert-small-v2	0.5280	
Lunr + All-mini-lm-6-v2	0.6564	2H23
Lunr + LaBSE	0.6564	
Lunr + all-mpnet-base-v2	0.6767	11H17
Lunr + bert	0.5312	
Plaid-X + All-mini-lm-6-v2	0.6824	15mins

CONCLUSION

Merci de votre attention