

LO17 : Indexation et Recherche d'information

TD7 : Moteur de recherche

Printemps 2025

Travail à réaliser

Vous avez parcouru les différentes étapes de préparation, d'analyse, de lemmatisation et d'indexation d'une archive ADIT. Ce TD vise à mettre en œuvre un prototype complet de moteur de recherche et à évaluer ses performances.

1. Mise en place du moteur

Implémentez un script `moteur.py` qui :

1. Lit une requête en langage naturel (on se limite aux requêtes proposées dans le TD précédent).
2. Identifie et extrait les composants structurés (mots-clés, dates, rubriques, opérateurs booléens,...).
3. Applique le correcteur orthographique (on suppose qu'une erreur ne peut être faite que sur les mots-clés).
4. Traduit la requête en opérations sur les index inversés.
5. Retourne la liste de documents renvoyés par la requête selon le modèle de recherche booléen ou booléen classé.

2. Évaluation expérimentale

1. Préparez un jeu de test :
 - Sélectionnez 10 requêtes variées (dates, rubriques, champs multiples).
 - Pour chaque requête, définissez manuellement la liste de documents « pertinents ».
2. Calculez pour chaque requête la précision et le rappel.
3. Mesurez le temps de réponse moyen du moteur sur 100 exécutions.
4. Présentez les mesures sous forme de tableaux et de graphiques.

3. Bonus : Interface utilisateur

- Préparez une interface utilisateur interactive de votre choix.
- Affichez pour chaque résultat :
 - Identifiant du document, titre, date, rubrique.
 - Extrait contextuel (snippets) où apparaissent les termes.
- Prévoyez une option de tri : par pertinence, date croissante/décroissante.