

TP10 : Recherche DSL

Il est conseillé d'installer Kibana pour profiter de sa Dev Console et donc effectuer les requêtes plus facilement

10.1 Requêtes DSL

Effectuer les requêtes DSL sur la base documentaire :

- Document PDF triés par date
- Documents dont le champ *content* répond à « administration »
- Documents dont le champ *content* ou *title* répond à « administration »
- Documents dont le champ *content* ou *title* répond à « administration » et dont la date de création est comprise dans un intervalle
- Documents PDF dont le champ *content* ou *title* répond à « administration » et dont la date de création est comprise dans un intervalle
- Documents dont le champ *content* répond à « Administration » ou à « Oracle »
- Documents dont le champ *content* répond à « Administration » et éventuellement à « Oracle »

Effectuer les requêtes DSL suivantes sur l'index des accès Apache:

- Documents dont le code retour est 404
- Documents dont le code retour est différent de 200
- Documents dont le code retour est différent de 200 et le verb POST
- Documents dont la taille est supérieure à 200.000 octets
- Document dont le timestamp est compris entre 2 dates

10.2 Contrôle du score

Effectuer des recherches avancées avec le paramètre *explain* :

- En utilisant le boosting, récupérer les Documents dont le champ *content* ou *title* répond à « administration ». Les documents ayant « administration » dans le champ *title* apparaissant en premier
- Même requête en utilisant des clauses *should*, comparer les scores
- Utiliser un autre mode de calcul avec *dis_max*

10.3 Matching partiel

- Préparer un nouvel index qui utilise une indexation multiple pour le champ *attachment.title* :
 - Type *keyword*
 - Type *text* avec standard analyseur
 - Type *text* avec un analyseur utilisant *edge_ngram*
- Effectuer des requêtes de matching partiel en utilisant un des 3 champs du mapping. Comparer les résultats et les temps de réponse

10.4 Phrases

Effectuer des requêtes avec des phrases :

- Récupérer les documents contenant la phrase « framework java », autoriser une distance de 5 mots
- Récupérer les documents dont le titre commence par « administration j »

10.5 Synonymes, Fuzzy

- Effectuer des recherches fuzzy avec des fautes de typo
- Définir une nouvelle liste de synonymes : Drools équivalent à BIRT
Tester les synonymes sur un nouvel index

10.6 Surbrillance

Effectuer les requêtes précédentes en ajoutant la surbrillance sur le champ content (limiter la taille du fragment)