[Date]

Rapport Laboratoire n°1

MA-WEM – Printemps 2022

Clercq Maxime Demeusy Jean Riondet Guillaume

# Fonctionnement code source

* Dans votre rapport, détaillez la syntaxe utilisée par ElasticSearch pour les faire des querys simples. Comment fait-on pour rechercher uniquement dans certains champs ? Comment fait-on pour « booster » certains champs par rapport à d’autres.

Pour « booster » certains champs par rapport à d’autres, il faut attribuer un score plus élevé à ces champs. Ainsi, s’ils apparaissent dans la requête, ils seront mis plus en avant.

Voici comment faire une querie simple dans ElasticSearch depuis notre système :

Une image contenant texte

Description générée automatiquement

Avec python il faut lancer le module queries en lui donnant les paramèters suivant :

-t :

# Questions théoriques

* Veuillez expliquer quelle(s) stratégie(s) il faut adopter pour indexer des pages dans plusieurs langues (chaque page est composée d’une seule langue, mais le corpus comporte des pages dans plusieurs langues). A quoi faut-il faire particulièrement attention, quels sont les avantages et inconvénients de celle(s)-ci ? Veuillez expliquer la(les) démarche(s) que vous proposez.
* *ElasticSearch* permet par défaut de faire de la recherche floue (fuzzy query). Veuillez expliquer de quoi il s’agit et comment *ElasticSearch* l’a implémenté. Certains prénoms peuvent avoir beaucoup de variation orthographiques (par exemple Caitlin : Caitilin, Caitlen, Caitlinn, Caitlyn, Caitlyne, Caitlynn, Cateline, Catelinn, Catelyn, Catelynn, Catlain, Catlin, Catline, Catlyn, Catlynn, Kaitlin, Kaitlinn, Kaitlyn, Kaitlynn, Katelin, Katelyn, Katelynn, etc). Est-il possible d’utiliser, tout en gardant une bonne performance, la recherche floue mise à disposition par *ElasticSearch* pour faire une recherche prenant en compte de telles variations ? Sinon quelle(s) alternative(s) voyez-vous, veuillez justifier votre réponse.