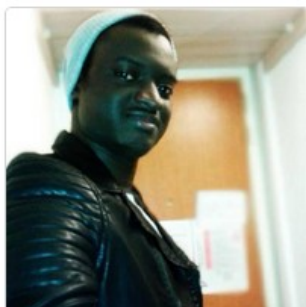




SERIES SEARCH

Un moteur de recherche 100 % séries TV !



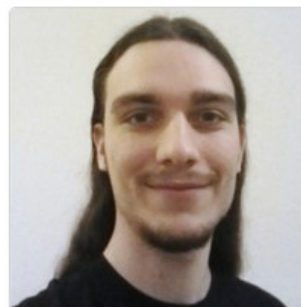
Macky Dieng

dieng444



Julien Foucault

JulienNaufaulo



Baptiste Vannesson

Badacadabra



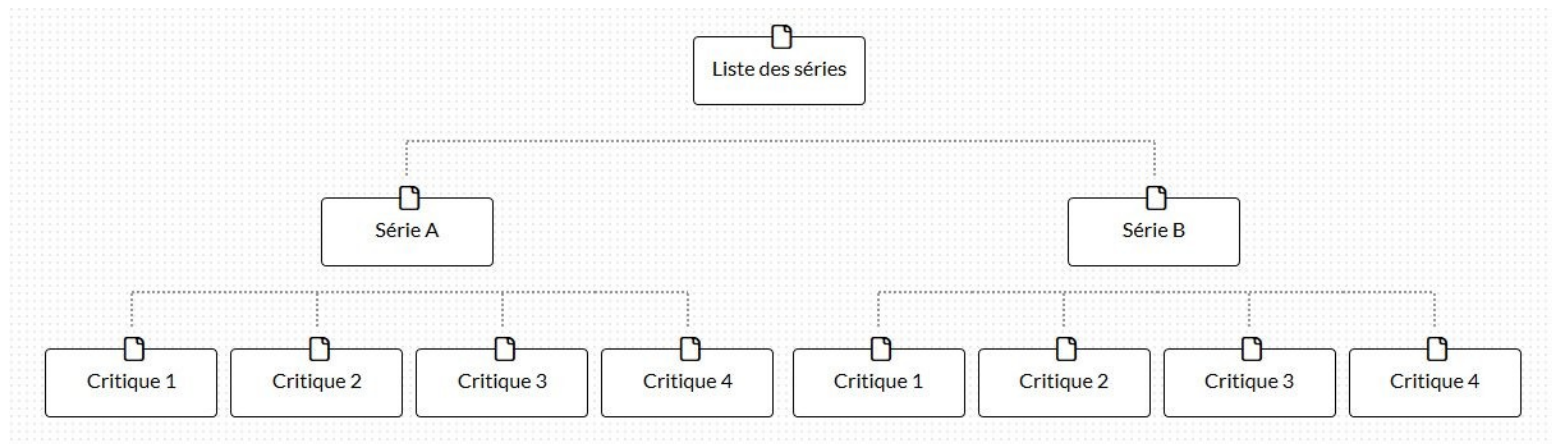
CRAWLING

TECHNOLOGIES



ALGORITHME

DFS pour un crawl en profondeur...



IR* & ARCHIVAGE

- ✓ « Waiting time » de 3 secondes entre chaque page
→ **politesse**
- ✓ Récupération de données spécifiques (titre de la série, noms des acteurs, etc.) via des requêtes **XPath**
- ✓ Stockage des infos récupérées dans une BDD **MySQL**
- ✓ Enregistrement d'une copie de la page **HTML** en dur pour pouvoir la consulter ultérieurement

* Information retrieval



SEARCH ENGINE

DÉMARCHE

✓ Création de l'index

- ✓ Tokenisation du texte
- ✓ Création d'un dictionnaire de tokens avec, pour chaque token, le calcul de :
 - ✓ $DF \rightarrow$ Document Frequency
 - ✓ $TF \rightarrow$ Term Frequency

✓ Requête utilisateur

- ✓ Tokenisation de la requête
- ✓ Récupération des 10 résultats les plus pertinents (selon le TF-IDF)



BACK-END

TECHNOLOGIES

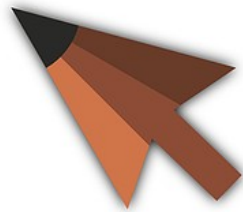


ENJEUX

- ✓ Recherche **temporelle**

FROM:startDate TO:endDate

- ✓ Valorisation des **entités** réelles avec le « **knowledge graph** »
- ✓ Dynamisation des diagrammes de présentation de données → **Dataviz**



FRONT-END

TECHNOLOGIES



HTML



CSS



JS



FONCTIONNALITÉS

- ✓ **Design immersif** (TV et ambiances)
- ✓ **UX persistante** (cookies)
- ✓ **Navigation ludique** (télécommande)
- ✓ **Dataviz** (graphe, timeline, statistiques)
- ✓ **Données structurées** (microdata)



~ MERCI ~



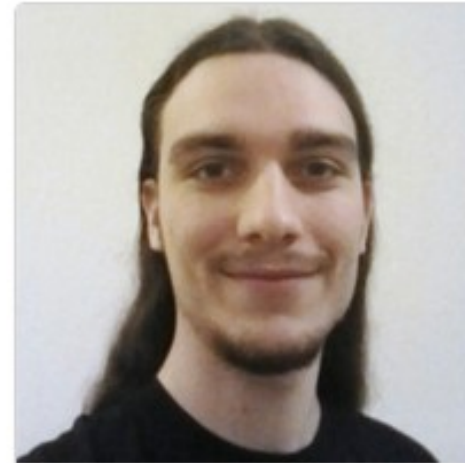
Macky Dieng

 dieng444



Julien Foucault

 JulienNaufaulo



Baptiste Vannesson

 Badacadabra