Classification de texte :

Catégories : cinéma, scène (théâtre/opéra), musique, Arts(sculpture, peinture etc.), littérature, tv

Tokenisation

Listes des champs lexicaux

Dictionnaire : {mot : {cat : oc}} => {acteur : {cinéma : 4}, {peintre : {peinture : 1}, {spectateur : {cinema : 2}, {théâtre : 2}}}

La catégorie ayant le plus haut score est la catégorie du texte.

Donner un poids plus important au titre si un mot d’une des catégories y est présent

Sortie :

Dossier dans lequel on aura classé tous les textes

Pour chaque texte : afficher le pourcentage d’appartenance à une catégorie

Les champs lexicaux :

Cinéma :

scène (théâtre/opéra) :

musique : fête, chanteur, pianiste, guitariste, guitare, auteur-compositeur, enregistrement, enregistrer, studio, concert, groupe, musicale, musique, pop, rock, reggae,

Arts(sculpture, peinture etc.) : œuvre, (filtre), galerie, musée, exposition, moulage, sculpture, peinture, peintre, sculpteur, histoire de l’art, art

Littérature :

tv :