Projet : On attend de nous qu’on fasse une application permettant de regarder un corpus. Obtenir des statistiques, faire des requêtes, comparer des corpus. Comparer les nuages de mots clés par exemple.

Gérer correctement les auteurs.

If auteur is not in dico:

A = nouvel auteur

Dico[auteur] = a

Sinon a = dico[auteur]

Il y a beaucoup d’informations dans notre société. Big data = V de volume, V de vélocité (elles évoluent), V de variété (texte, image, son, etc).

Bibliothèques interdites.

TP7

Construire un dictionnaire de mots (vocab)

Pour chaque document, construire un représentation binaire bas » sur le dico (0/1 selon la présence, ou nombre d’apparition).