

Exercice : sous python

1. Créez une base de 15 textes à partir du web avec leurs classes respectives :
 (5 textes de sport, 5 textes d'économie, et 5 textes de politique)
 2. Créez une fonction qui crée une liste L de tous les mots figurant dans cette base_(sans répétition).
 3. Créez une première fonction T2V qui transforme chaque texte T en un vecteur d'occurrences des mots de la liste L.
 4. Créez une liste de mots vides MV (pronoms, prépositions, conjonctions) et utilisez la pour réduire la dimension de L. Quel est le taux de réduction obtenu ?
 5. Créez une fonction ELIM qui élimine les mots de la liste L ayant figuré dans les 3 classes et garde uniquement ceux qui figurent dans une ou deux classes. Quel est le taux de réduction obtenu ?
 6. Créez une fonction DIST qui calcule la distance euclidienne entre deux vecteurs.
 7. Proposez un texte pour lequel : (en utilisant T2V)
 - Parmi les 15 textes de la base, la DIST minimale correspond à la classe économie.
 8. Proposez un texte pour lequel : (en utilisant T2V + MV + ELIM)
 - Parmi les 15 textes de la base, la DIST minimale correspond à la classe sport.
 9. Proposez un texte pour lequel :
 - La classe de la DIST minimale (en utilisant T2V) est différente de la classe de la DIST minimale (en utilisant T2V + MV + ELIM)
 10. Quel modèle d'apprentissage artificiel avez-vous utilisé dans les questions 7 à 9 ?
 Donnez 1 avantage et 1 inconvénient de ce modèle.
-