

Projet Analyser un texte sans le lire. . .

Idée générale

Adapté l'un TP guidé donné auparavant en début d'année en projet réel demandé plutôt en fin d'année.

But et Objectif

Le but de ce projet est simple :

- ▶ Vous avez un fichier .txt contenant le texte que vous voulez analyser.
- ▶ Vous allez devoir récupérer des mots liés à différents champs lexicaux : les mots de la guerre, de l'amour, de la science. . . Ces mots seront initialement également dans un fichier txt par champ lexical.

Votre but :

- ▶ Lire les mots du fichier un par un et créer des compteurs liés aux différents champs sémantiques. A chaque fois qu'un mot est un mot appartenant à un champ sémantique, on incrémente le compteur de ce champ.
- ▶ Une fois l'analyse terminée, votre programme d'analyse devra :
 - ▶ stocker la valeur associée à chaque champ sémantique ainsi que son pourcentage de présence.
 - ▶ afficher les 3 premières valeurs ainsi que le champ dont elles dépendent.

Prérequis

- ▶ Utilisation des fonctions, boucles et conditionnelles
- ▶ Utilisation des tableaux et des dictionnaires
- ▶ Lecture dans un fichier texte
- ▶ Recherche sur le Web ou utilisation d'IA générative

Déroulé

- ▶ Etape A : Réflexion préalable : découpe de la tâche à réaliser en sous-tâche en imaginant qu'on dispose bien :
 - ▶ d'un fichier texte
 - ▶ de n fichiers correspondant à n champs lexicaux.
 - ▶ nommer chaque fonction intervenant dans la réalisation de la tâche et expliquer en français ce qu'elle doit faire.
- ▶ Etape B : Prototyper plus finement les fonctions.
- ▶ Etape C : Faire une recherche Web ou utiliser une IA générative pour produire les différents fichiers contenant les mots de chaque champ lexical.
- ▶ Etape D : Se partager les tâches et réaliser concrètement le projet.

Revue de projet

- ▶ Chaque partie A-B-C devra être expliquée sur une feuille indépendante des autres.
- ▶ Le déroulé de la partie D devra être explicité sur le fichier README de votre projet et rempli à chaque séance de projet. On doit y voir qui a fait quoi ce jour là.

Prolongement

- ▶ Récupérer en plus les mots qui apparaissent le plus.
- ▶ Stocker le résultat dans un fichier sous le forme suivant : un résultat par ligne de la forme

```
'nom;champ1:valeur1; champ2:valeur2;champ3:valeur3
```

```
'
```