

Avant tout, s'être assuré d'avoir bien installé :

requests

beautifulsoup

pandas

Exo 1 :

Utiliser des boucles pour parcourir les cinq premières pages du site. Utiliser requests pour faire des requêtes HTTP et beautifulsoup pour parser le contenu HTML.

Comparer les tags extraits avec les tags "love", "inspirational", "life", "humor" et ne garder que les citations correspondantes.

Sauvegarder les citations et tags filtrés dans un fichier CSV en utilisant pandas (le csv est result.csv)

Exo 2 :

Utiliser requests.session() pour maintenir une session persistante.

Faire une requête GET pour obtenir la page de login et extraire le token CSRF nécessaire pour la connexion.

Faire une requête POST avec les identifiants et le token CSRF pour se connecter au site.

S'assurer que la connexion est réussie en vérifiant la présence du bouton de déconnexion ("Logout") dans la réponse. Extraire le cookie de session (csrftoken) après la connexion réussie et l'inscrire dans result.csv

Extraction des citations avec le tag "books" et filtrage des doublons :

Parcourir les deux premières pages du site pour les citations avec le tag "books".

Utiliser les mêmes techniques que précédemment pour extraire les citations et tags.

Charger les citations existantes depuis result.csv et fusionner les nouvelles citations en éliminant les doublons. Enregistrer les citations mises à jour dans le fichier CSV.