

# Récapitulatif

## 1. Scraping des 5 premières pages de citations :

Mise en place une boucle pour parcourir les 5 premières pages du site de citations  
Filtration des citations par rapports aux top premiers tags "love", "inspirational", "life", et "humor".

## 2. Connexion au site et récupération du token :

On accède à la page de connexion pour récupérer le token renvoyé par la requête Post

Nous avons envoyé les informations de connexion avec le token pour authentifier la session.

Après la connexion, nous avons extrait le cookie de session contenant le token de connexion.

## 3. Ajout des citations avec le tag "books" :

Grâce à la connexion, on peut scrap les 2 premières pages de citations avec le tag "books"

Ensuite stockage de ces citations dans les citations

Filtrage des doublons :

Utilisation de panda à partir de la liste des citations pour supprimer les doublons basés sur le texte des citations.

## 4. Écriture des résultats dans un fichier csv :

Écriture des citations filtrées et sans doublons dans un fichier csv au nom de results.csv

Ajout grâce au script du token de connexion à la fin du fichier csv

Ajout de l'étape pour écrire le token de connexion dans le fichier Csv.