

Description de code de collection des données de twitter

Le code de dossier twitter est un code Python qui collecte des données à partir de comptes Twitter à l'aide de la bibliothèque Selenium. Décortiquons la stratégie et les étapes à suivre

Importation des bibliothèques nécessaires :

- datetime et timedelta du module datetime : Utilisées pour travailler avec des dates et des différences de temps.
- pytz : Une bibliothèque pour travailler avec les fuseaux horaires.
- By de selenium.webdriver.common.by : Utilisé pour localiser des éléments sur une page web.
- webdriver de selenium : Utilisé pour automatiser les interactions avec le navigateur web.

Définition de la classe TwitterClass :

- Cette classe contient deux méthodes : Onetweet(tweet) et Accounttwitter(self, driver).
- La méthode Onetweet(tweet) prend un élément tweet et extrait les données pertinentes telles que le texte, les images, les hashtags, les liens, la date, le nombre de réponses, le nombre de like et le nombre de retweet. Elle renvoie un dictionnaire contenant ces informations.
- La méthode date_is_within_last_seven_days(date_string) vérifie si une date donnée se situe dans les sept derniers jours. Elle renvoie True si c'est le cas, et False dans le cas contraire.

Création d'une instance WebDriver :

- Le code crée une instance Firefox WebDriver en utilisant webdriver.Firefox().
- Définition d'une liste d'URL de comptes Twitter :
- La liste des comptes contient les URL des comptes Twitter dont les données seront collectées.

Passer en revue les comptes et collecter les données :

Pour chaque URL de compte dans la liste des comptes, le code effectue les étapes suivantes :

- Navigue vers l'URL du compte à l'aide de driver.get(account).
- Attendre le chargement de la page avec driver.implicitly_wait(10).
- Appelle la méthode Accounttwitter de la classe TwitterClass pour collecter les données du compte.
- Sérialise les données collectées au format JSON en utilisant json.dumps(data, indent=4).
- Enregistre les données JSON dans un fichier nommé d'après le nom d'utilisateur du compte dans le répertoire "Accountsdata".
- Afficher un message indiquant que le fichier JSON a été enregistré.
- Quitter le WebDriver :
- Après avoir collecté les données de tous les comptes, le code ferme l'instance du WebDriver avec driver.quit().

Globalement, la stratégie consiste à utiliser Selenium pour automatiser le navigateur web, naviguer vers les pages des comptes Twitter et extraire les données pertinentes des tweets. Les données collectées sont ensuite sérialisées au format JSON et enregistrées dans des fichiers distincts pour chaque compte. Le code utilise plusieurs bibliothèques et outils tels que Selenium, datetime, pytz et JSON.