

Documentation du Web Scraping des Données de Vols

Hamza Elhaiki

April 1, 2024

1 Introduction

Ce document présente la documentation du processus de web scraping des données de vols à partir du site Kayak. L'objectif de ce script est de collecter les informations sur les vols entre un aéroport de départ et plusieurs destinations pour une période spécifiée.

2 Objectif de Projet

L'objectif principal de projet est de fournir aux utilisateurs un moyen efficace de collecter des données sur les vols disponibles entre un aéroport de départ donné et plusieurs destinations, pour une période de temps spécifiée. Les données collectées peuvent être utilisées à diverses fins, notamment :

- Comparer les prix des billets d'avion entre différentes compagnies aériennes et destinations.
- Analyser les tendances des prix des billets pour planifier des voyages à des dates moins chères.
- Suivre les changements dans les horaires de vol et les disponibilités des compagnies aériennes.
- Explorer les options de vol pour des destinations spécifiques afin de faciliter la planification de voyages.

3 Fonctionnement du Script

Le script utilise Selenium pour automatiser l'interaction avec le site Kayak. Il récupère les informations sur les vols disponibles pour chaque destination spécifiée et les stocke dans un fichier CSV.

4 Paramètres du Script

- **Aéroport de Départ** : L'aéroport de départ est spécifié comme argument du script.
- **Destinations** : Une liste des destinations est fournie au script.
- **Dates de Départ et de Fin** : Les dates de départ et de fin spécifient la période pour laquelle les vols sont recherchés.

5 Données Collectées

Les données collectées pour chaque vol comprennent les éléments suivants :

- Heure de départ
- Heure d'arrivée
- Aéroport de départ
- Aéroport d'arrivée
- Type de vol (direct ou avec escale)
- Durée du vol
- Compagnie aérienne
- Prix du billet
- Classe du billet
- URL du vol sur Kayak

6 Fichier de Sortie

Les données collectées sont enregistrées dans un fichier CSV avec un nom de fichier unique basé sur la date et l'heure d'exécution du script.

7 Exemple d'Utilisation

Pour exécuter le script, vous pouvez utiliser la commande suivante dans votre terminal :

```
python web_scraping_flight_data.py CMN PAR,JFK,ROM  
2024-04-26 2024-04-29
```

Dans cet exemple, les paramètres spécifiés sont les suivants :

- **Aéroport de Départ** : CMN (Casablanca, Maroc)
- **Destinations** : PAR (Paris, France), JFK (New York, États-Unis), ROM (Rome, Italie)
- **Dates de Départ et de Fin** : 2024-04-26 et 2024-04-29

La liste des destinations comprend les codes IATA (International Air Transport Association) des aéroports de destination à travers le monde. Voici la signification des symboles pour chaque destination :

- PAR : Paris, France
- JFK : Aéroport John F. Kennedy, New York, États-Unis
- ROM : Aéroport de Rome Fiumicino, Rome, Italie
- FRA : Aéroport de Francfort-sur-le-Main, Francfort, Allemagne
- YYZ : Aéroport Pearson de Toronto, Toronto, Canada
- IST : Aéroport d'Istanbul, Istanbul, Turquie
- ABJ : Aéroport Félix Houphouët-Boigny, Abidjan, Côte d'Ivoire
- ZRH : Aéroport de Zurich, Zurich, Suisse
- JED : Aéroport King Abdulaziz, Djeddah, Arabie Saoudite
- LON : Londres, Royaume-Uni (représente plusieurs aéroports de Londres)
- SYD : Aéroport Kingsford Smith de Sydney, Sydney, Australie

Ces symboles sont utilisés pour spécifier les destinations des vols que vous souhaitez rechercher à partir de l'aéroport de départ spécifié (CMN dans cet exemple).

8 Conclusion

Ce script de web scraping permet de collecter rapidement et efficacement des données sur les vols à partir du site Kayak pour une analyse ultérieure.