

Books Online

Les exigences du système de surveillance des prix

Ce document contient les exigences et les instructions pour élaborer une version bêta du système de surveillance des prix pour Books Online.

Notes:

- Au cours du projet, veuillez à enregistrer votre code dans un repository GitHub et à effectuer des commits réguliers accompagnés de messages de commit clairs.
- N'oubliez pas que vous devez enregistrer un fichier requirements.txt sans enregistrer votre environnement virtuel dans le repository lui-même.
- Vous ne devez pas non plus enregistrer vos fichiers CSV.
- Veuillez également à prendre le temps d'écrire un fichier README.md, que vous ajouterez dans le repository afin que je puisse exécuter le code correctement et produire quelques données !

Phase 1

Choisissez n'importe quelle page Produit sur le site de [Books to Scrape](#). Écrivez un script Python qui visite cette page et en extrait les informations suivantes :

- product_page_url
- universal_product_code (upc)
- title
- price_including_tax
- price_excluding_tax
- number_available
- product_description
- category
- review_rating
- image_url



Écrivez les données dans un fichier CSV qui utilise les champs ci-dessus comme en-têtes de colonnes.

Phase 2

Maintenant que vous avez obtenu les informations concernant un premier livre, vous pouvez essayer de récupérer toutes les données nécessaires pour toute une catégorie d'ouvrages.

Choisissez n'importe quelle catégorie sur le site de [Books to Scrape](#). Écrivez un script Python qui consulte la page de la catégorie choisie, et extrait l'URL de la page Produit de chaque livre appartenant à cette catégorie.

Combinez cela avec le travail que vous avez déjà effectué dans la phase 1 afin d'extraire les données produit de tous les livres de la catégorie choisie, puis écrivez les données dans un seul fichier CSV.

Remarque : certaines pages de catégorie comptent plus de 20 livres, qui sont donc répartis sur différentes pages (« pagination »). Votre application doit être capable de parcourir automatiquement les multiples pages si présentes.

Phase 3

Ensuite, étendez votre travail à l'écriture d'un script qui consulte le site de [Books to Scrape](#), extrait toutes les catégories de livres disponibles, puis extrait les informations produit de tous les livres appartenant à toutes les différentes catégories. Vous devrez écrire les données dans un fichier CSV distinct pour chaque catégorie de livres.

Phase 4

Enfin, prolongez votre travail existant pour télécharger et enregistrer le fichier image de chaque page Produit que vous consultez.