

Bonjour Sam,

Je suis heureux de vous aider à mettre en place un système de surveillance de prix pour Books to Scrape. Je suis disponible pour suivre les étapes décrites dans le fichier de spécifications et créer une version bêta qui sera limitée à un seul revendeur.

En pièces jointes, vous trouverez le lien vers le dépôt GitHub contenant le code, un fichier compressé des données générées, un requirements.txt et un README.md complet qui vous aideront à exécuter le code avec succès et à produire des données. Veuillez noter que le dépôt ne contient pas les données et images extraites.

Je peux également vous fournir une explication sur la façon dont nous pouvons utiliser ce code pour établir une pipeline ETL et vous aider à présenter cette idée à votre responsable. Dans un premier temps, nous allons extraire les données en utilisant BeautifulSoup. Les informations extraites seront les suivantes : product_page_url, universal_product_code (upc), title, price_including_tax, price_excluding_tax, number_available, product_description, category, review_rating, et image_url.

Ensuite, nous allons transformer les données en supprimant les caractères spéciaux et les lettres des chiffres, afin de les rendre plus faciles à utiliser dans la base de données en format CSV. Enfin, nous allons charger les données pour les écrire dans la base de données CSV.

Une fois que nous avons ces fichiers CSV, nous pourrons utiliser la bibliothèque pandas pour effectuer une analyse des données et surveiller les prix.

Si vous avez des questions, n'hésitez pas à me contacter. Je suis à votre disposition.

Cordialement,

Théo Pidoux