

De : EL-WALID EL-KHABOU (alkhabou@gmail.com)
A : SAM (Mon responsable d'équipe)

Objet : Description de l'application pour le pipeline ETL

Paris le 04 Avril 2023

Cher Sam,

J'espère que vous allez bien. Je souhaiterais vous présenter une version bêta de l'application que j'ai développée suite à votre demande, et qui peut être utilisée pour établir un pipeline ETL efficace.

L'application est un script Python qui permet de scraper le site web de vente de livres [Books to Scrape](#) et d'extraire des données spécifiques, telles que le code UPC, le titre, le prix, la disponibilité, la description du produit, la catégorie, la note de l'examen et l'URL de l'image associée. Ces données sont ensuite stockées dans des fichiers CSV, un par catégorie de livre, pour faciliter l'analyse et le traitement ultérieurs.

Le processus ETL commence par l'extraction des données à partir du site web en utilisant la bibliothèque requests et la bibliothèque de parsing HTML, BeautifulSoup. Les données sont ensuite transformées pour correspondre aux spécifications souhaitées et stockées dans les fichiers CSV avec la bibliothèque csv. Enfin, les images associées sont téléchargées et stockées dans des dossiers dédiés à chaque catégorie de livre.

Le script Python peut être exécuté périodiquement pour extraire de nouvelles données à partir du site web et ajouter des entrées aux fichiers CSV. Cela permet de maintenir les données à jour et de les utiliser pour des analyses ultérieures.

En conclusion, cette application peut être utilisée pour extraire et transformer des données à partir de différents sites web et les stocker dans des fichiers CSV pour être utilisées dans des pipelines ETL. Elle peut également être automatisée pour exécuter périodiquement la tâche d'extraction des données et mettre à jour les fichiers CSV.

Je suis prêt à discuter de cette application en personne ou à répondre à toute question que vous pourriez avoir. Merci pour votre temps et votre attention.

Cordialement,

EL-WALID EL-KHABOU