

Bonjour Sam

Je souhaite vous fournir un guide d'utilisation du code que j'ai développé pour établir un pipeline ETL. Ce code est conçu pour extraire les données des livres chez [Books to Scrape](#), les transformer en données prêtes à l'emploi et les charger dans des fichiers csv et des dossiers pour les images. Le but est d'automatiser le suivi des prix des livres de cette librairie en ligne.

Voici les étapes suivies:

1. Extraction des données du site :

Dans le module extraction vous trouverez toutes les fonctionnalités qui s'occupent d'extraire les données des livres :

- `extraire_les_donnees_d_un_livre(url_un_livre)`

Elle retourne un dictionnaire qui contient les informations d'un livre.

- `extraire_les_livres_d_une_page(url_une_page)`

Elle retourne une liste qui contient les urls des livre d'une page

- `extraire_les_livres_d_une_categorie(url_d_une_categ)`

Elle retourne une liste qui contient les urls d'une catégorie entière

- `extraire_les_livres_de_toutes_les_categories()`

Elle retourne une liste qui contient les noms et urls de toutes les catégories comme leurs définitions l'indique chaque fonction à un rôle précis.

2. Transformation des données :

Dans le module transformation vous trouverez par exemple la fonctionnalité qui permet de transformer le nombre d'étoiles de chaîne de caractère à un nombre :

- `transformer_le_nombre_d_etoiles_str_en_nombre(nombre_etoiles_str)`

Elle retourne un nombre d'étoiles numérique

3. Chargement des données :

Dans le module chargement vous trouverez les fonctionnalités qui s'occupent de charger les informations des livres dans des fichiers CSV et les images dans des dossiers. Par Exemple :

- `charger_infos_et_image_d_un_livre_d_une_categorie(categorie, infos_livre)`

Elle permet de charger les informations d'un livre d'une catégorie spécifique ainsi que son image.

Pour lancer le programme il faut exécuter le main qui contient le programme principale et qui utilise toutes les fonctionnalités des trois modules ETL.