

Bonjour,

Je vous contacte concernant le système du site , j ai fini toute la partie scrapping du site et toute les données sont bien repartis dans des fichiers csv par catégorie et les images dans un dossier images .

Pour la partie extraction des données j ai utilisé la librairie beautiful soup qui permet de recuper les éléments du une balise HTML avec son id .

Il y a 4 fonction principale de scrapping :

- Scrap\_book pour récupérer toute les infos d un seul livre
- Scrap\_category pour récupérer tout les URL livres d une category cette fonction fait appel à la fonction scrap\_book pour scrap tout les URL de livre récupéré
- Get\_all\_pages permet de passer a la page suivante d une catégorie si cette page existe et puis appel la fonction scrap\_category
- Get\_all\_categories la fonction principale qui va récupérer les URL de toute les catégories puis appliquer la fonction Get\_all\_pages qui va faire appel à scrap\_category etc

Voilà pour ma partie extraction quand a la transformation des données j ai juste eu a réaliser quelques fonctions (clean price par exemple) pour nettoyer les données récupérer du site des informations inutiles (par exemple le symbole £) pour que les données soient conformes à vos exigences

Finalement pour le chargement des données (load )

J ai créé une fonction mkdir pour créer un répertoire scrapeddata puis une autre fonction pour créer un fichier csv avec le nom de la catégorie dans lequel il y aura tout les livres de cette même catégorie avec les données organisée

Pour la partie image il y a un dossier image dans lequel sont téléchargé route les images du site

.

Si vous avez besoin de plus de renseignements merci de me recontacter par mail ou par téléphone ,

Cordialement,

Shmuel BITAN.