

WEB SCRAPING



Projet Scraping/API

STRUCTURE : Sonatel Academy		
EQUIPE DU PROJET : Ibrahima khalil CISSE, Pape Modou DIOP et Daour Samba NDAW		
ENCADREUR : Mr MBAYE		
<input checked="" type="checkbox"/> TRAVAIL EN COURS	<input type="checkbox"/> TRELLO	<input type="checkbox"/> GITHUB

Introduction :

Le Web Scraping, également connu sous le nom d'extraction de données web ou de collecte de données web, est une technique utilisée pour extraire automatiquement des informations à partir de sites web. Il s'agit d'un processus automatisé qui parcourt les pages web, extrait les données pertinentes et les organise dans un format structuré, généralement une base de données, une feuille de calcul ou un fichier JSON.

Objectifs :

- Sélectionner un sujet, puis de trouver des sites Web proposant du contenu sur ce sujet afin de les analyser et d’extraire les données.
- les données extraites doivent être conservées dans une base de données PostgreSQL.
- Ajoutez-en plus si nécessaire

Daily meeting 1: (04/07/2023)

Résumé de l'objectif du jour précédent :

L'objectif principal du jour précédent était de finaliser les tâches suivantes :

1. Terminer la mise en place du Trello pour la gestion de projet.
2. Terminer la configuration du GitHub avec chacun sa branche.
3. Finaliser le scraping avec NodeJS pour récupérer les données nécessaires.

Progression réalisée aujourd'hui :

Nous avons essayé d'accomplir les objectifs fixés pour la journée. Voici un résumé de notre progression sur chaque tâche :

1. Trello:

- Nous avons créé un tableau Trello pour le projet, avec des colonnes définissant les étapes clés du processus.
- Chaque membre de l'équipe a été invité à rejoindre le tableau et à attribuer les tâches qui leur sont assignées.
- Nous avons importé les tâches existantes dans Trello et les avons triées selon leur priorité.

2. GitHub:

- Nous avons créé un repository GitHub dédié au projet.
- Les membres de l'équipe ont été ajoutés en tant que collaborateurs au dépôt et on a trois branches qui représentent nos trois services.
- Nous avons effectué le premier commit initial en important le code existant dans le dépôt.

3. Scraping avec NodeJS:

- Nous avons finalisé le script de scraping en utilisant NodeJS.
- Le script est maintenant capable de se connecter au site cible, extraire les données spécifiques et les enregistrer dans un format Json.

NB: Y'a que deux membre du groupe qui etait présent du coup chacun a scrappé que le service qu'il doit faire. Donc il reste un service qu'on doit scrapper avec nodeJS.

Défis rencontrés:

Nous avons rencontré quelques défis mineurs au cours de la journée :

- Nous avons eu un problème sur le trello car on ne connaissait pas encore les etapes à suivre pour la mise en place de l'API REST.

Points à discuter:

Nous souhaitons discuter des points suivants lors du prochain Daily Meeting :

- Partage des connaissances : organiser une session de partage des connaissances pour familiariser tous les membres de l'équipe avec les technologies et les outils utilisés.
- Résolution rapide des problèmes : identifier les obstacles potentiels à notre progression et déterminer les actions à entreprendre pour les surmonter rapidement.

Daily meeting 2: (05/07/2023)

Résumé de l'objectif du jour précédent :

L'objectif principal du jour précédent était de finaliser les tâches suivantes :

1. Faire le scraping avec Python pour récupérer les données nécessaires.
2. Identifier un API traitant le même sujet et extraire des données.

Progression réalisée aujourd'hui :

Nous avons essayé d'accomplir les objectifs fixés pour la journée. Voici un résumé de notre progression sur chaque tâche :

1. Scraping avec Python:

- Nous avons fait le script de scraping en utilisant Python.
- Le script est maintenant capable de se connecter au site cible, extraire les données spécifiques et les enregistrer dans un format Json.

2. API:

- Seul un membre de l'équipe a trouvé un API et extraire des données, les autres membres n'ont pas encore trouvé d'API.

Défis rencontrés:

Nous avons rencontré quelques défis mineurs au cours de la journée :

- Nous avons eu un problème pour récupérer des données sur le site Expat Dakar

Points à discuter:

Nous souhaitons discuter des points suivants lors du prochain Daily Meeting :

- Partage des connaissances : organiser une session de partage des connaissances pour familiariser tous les membres de l'équipe avec les technologies et les outils utilisés.
- Résolution rapide des problèmes : identifier les obstacles potentiels à notre progression et déterminer les actions à entreprendre pour les surmonter rapidement.

Daily meeting 3: (06/07/2023)

Résumé de l'objectif du jour précédent :

L'objectif principal du jour précédent était de finaliser les tâches suivantes :

1. Trouver une api traitant les sujets qui restaient .

2. Récupérer les données de l'api.
3. Faire le scraping avec nodejs pour le service qui restait (immobilière).

Progression réalisée aujourd'hui :

Nous avons essayé d'accomplir les objectifs fixés pour la journée. Voici un résumé de notre progression sur chaque tâche :

1. Api traitant les mêmes sujets :
 - Api pour avoir les données Electronique sur [Rapid Api](#)
2. Récupérer les données de l'api:
 - Pour l'api de données électronique.

Défis rencontrés:

Nous avons rencontré quelques défis mineurs au cours de la journée :

- Nous avons eu un problème pour trouver un api offrant des immobilier au Sénégal.
- Api pour les données électronique n'avait pas toutes les données dont nous avons besoin

Points à discuter:

Nous souhaitons discuter des points suivants lors du prochain Daily Meeting :

- Création de notre base de données.
- Partage des connaissances : organiser une session de partage des connaissances pour familiariser tous les membres de l'équipe avec les technologies et les outils utilisés.
- Résolution rapide des problèmes : identifier les obstacles potentiels à notre progression et déterminer les actions à entreprendre pour les surmonter rapidement.

Daily meeting 4: (07/07/2023)

Résumé de l'objectif du jour précédent :

L'objectif principal du jour précédent était de finaliser les tâches suivantes :

1. Discussion sur notre schémas pour la base de données (les tables, les attributs).
2. Créer la base de données .
3. Insérer les données récupérer lors du Scraping et les insérer dans la base de données.

Progression réalisée aujourd'hui :

Nous avons essayé d'accomplir les objectifs fixés pour la journée. Voici un résumé de notre progression sur chaque tâche :

1. Schémas de données:

Nous avons crée les tables suivantes:

- Electronique (id, tittle, price, location, image)
- Electromenager (id, tittle, price, location, image)
- Immobilier (id, tittle, price, location, image)
- Utilisateur (id, full_name, email, password, role)
- Proposition (id, type, status, id_electronique ,id_electromenager, id_immobilier, id_utilisateur)

2. Création de la base de données :

- Nous avons tous créé la même base .
- Création d'un fichier sql contenant les différentes tables en respectant la syntaxe postgres.
- Et nous avons tous importer le fichier dans la base de données.

3. Insertion des données :

- Nous avons écrit un code python qui nous permettait de récupérer les fichiers json créer lors du scraping puis créer un fichier sql avec les insertions.
- Chacun de nous a importer le fichier dans sa base.

Défis rencontrés:

Nous avons rencontré quelques défis mineurs au cours de la journée :

- Problème lors de insertions des données causer par la syntaxe postgres (ou on devrait mettre les guillemets double et les guillemets simple).

Points à discuter:

Nous souhaitons discuter des points suivants lors du prochain Daily Meeting :

- Partage des connaissances : organiser une session de partage des connaissances pour familiariser tous les membres de l'équipe avec les technologies et les outils utilisés.
- Résolution rapide des problèmes : identifier les obstacles potentiels à notre progression et déterminer les actions à entreprendre pour les surmonter rapidement.

Daily meeting 5: (10/07/2023)

Résumé de l'objectif du jour précédent :

L'objectif principal du jour précédent était de finaliser les tâches suivantes :

1. Se documenter
2. Discussion sur notre structure pour la mis en place de notre api et créations des dossiers et fichiers dont nous avons besoin.

Progression réalisée aujourd'hui :

Nous avons essayé d'accomplir les objectifs fixés pour la journée. Voici un résumé de notre progression sur chaque tâche :

1. Documentation:

Nous nous sommes documenter sur les plateformes suivants:

- [OpenClassroom](#)
- [Chat gpt](#)
-
-

2. Structure de l'api :

Au niveau de dossier api nous avons les la structure suivants

- Dossier Routes avec les fichiers `electroniqueRoute` (defintitions de toutes les routes), `electromenagerRoute`, `immobilierRoutes` .
- Dossier Controllers avec les fichiers `electroniqueController` (defintions des midlewares), `electromenagerController`, `immobilierController`.
- Dossier Services avec les fichiers `electroniqueService` (defintioions des interactions avec la base de donnees), `electromenagerService`, `immobilierService`.
- Fichier `Server.js` (les endpoints).

Défis rencontrés:

Aucune

Points à discuter:

Nous souhaitons discuter des points suivants lors du prochain Daily Meeting :

- Partage des connaissances : organiser une session de partage des connaissances pour familiariser tous les membres de l'équipe avec les technologies et les outils utilisés.
- Résolution rapide des problèmes : identifier les obstacles potentiels à notre progression et déterminer les actions à entreprendre pour les surmonter rapidement.