**Projet Web Scraping**

**NB :**

**Date de presentation des travaux : Lundi 29 juillet 2024 à 10h**

1. Introduction au Projet

- Objectif du Projet: Collecter des données spécifiques à partir de sites web contenant des informations sur le Senegal pour une analyse ultérieure.

- Importance: Apprendre les techniques de web scraping, comprendre les aspects légaux et éthiques, et manipuler les données pour en tirer des informations.

2. Formation et Préparation

- Cours Préliminaires: Introduction au web scraping, présentation des outils et des bibliothèques (BeautifulSoup, Scrapy, Selenium, etc.). Un exemple de notebook et de projet vous a été remis.

- Exemples de Projets: Analyse de projets existants pour comprendre les bonnes pratiques et les erreurs courantes.

- Environnement de Travail: Configuration de l'environnement de développement (installation de Python, des bibliothèques nécessaires, IDE recommandé).

3. Définition du Sujet et des Objectifs

- Choix du Sujet: Sélectionner un domaine d'intérêt commun (par exemple : scraping des blogs gérés par des sénégalais ou des sites e-commerce senegalais….).

- Objectifs Précis: Définir clairement ce que l'on veut obtenir avec le web scraping (par exemple : obtenir une base de données de prix au Senegal, recueillir des articles de blog….).

4. Recherche et Planification

- Identification des Sources: Trouver les sites web pertinents à scraper.

- Analyse des Sites Web : Comprendre la structure HTML des pages web cibles.

- Planification du Scraping: Établir un plan pour scraper les données( les données qui vous intéressent et comment vous allez les présenter).

5. Développement du Scraper

- Choix de la Bibliothèque: Sélectionner la bibliothèque ou le framework approprié (BeautifulSoup, Scrapy, Selenium).

- Écriture du Code: Développer le code pour scraper les données.

- Extraction de Données: Utiliser BeautifulSoup ou Scrapy pour extraire les données des pages web.

- Automatisation : Utiliser Selenium pour scraper des sites nécessitant une interaction avec l'utilisateur.( Bonus)

- Gestion des Erreurs: Mettre en place des mécanismes de gestion des erreurs (par exemple : gestion des temps de réponse, des pages manquantes).

6. Stockage des Données

- Choix du Format de Stockage: Déterminer le format de stockage des données (CSV, JSON, base de données SQL/NoSQL).

- Implémentation du Stockage : Écrire le code pour stocker les données scrapées dans le format choisi.

7. Analyse et Visualisation

- Préparation des Données: Nettoyer et préparer les données pour l'analyse.

- Analyse : Utiliser des outils de data science (Pandas, NumPy) pour analyser les données collectées.

- Visualisation: Créer des visualisations des données (matplotlib, seaborn) pour illustrer les résultats de l'analyse.

8. Aspects Légaux et Éthiques

- Respect des Conditions d'Utilisation: Vérifier et respecter les conditions d'utilisation des sites web cibles.

- Éthique du Scraping: Discuter des implications éthiques du web scraping, y compris la protection des données personnelles.

- Ne copiez pas les travaux des autres

9. Documentation et Présentation

- Documentation du Code: Documenter le code et le processus de scraping.

- Rapport Final : Rédiger un rapport détaillé sur le projet, les méthodes utilisées, les résultats obtenus, et les défis rencontrés avec un notebook.

- Présentation : Préparer une présentation pour partager les résultats et les apprentissages avec la classe.

10. Évaluation et Feedback

- Évaluation du Projet: Le projet sera évalué sur la base de critères définis (qualité du code, pertinence des données collectées, analyse des données, etc.).

- Feedback: Recevoir des retours des pairs et des enseignants pour améliorer les compétences.