

Énoncé de Projet: Développement d'un Web Scraper pour l'Analyse de Marché

Contexte:

Une entreprise souhaite analyser les tendances du marché et les comportements des consommateurs en collectant des données depuis des sites web d'actualités, de critiques de produits, et de forums. Votre objectif est de développer un web scraper en Python pour extraire ces données de manière efficace et structurée.

Objectifs du Projet:

Identification des Sources de Données: Sélectionner trois sites web différents (par exemple, un site d'actualités, un site de critiques de produits, et un forum) pour l'extraction des données.

Développement du Web Scraper:

Utiliser Python pour écrire des scripts capables d'extraire des données spécifiques (comme les titres d'articles, les avis des utilisateurs, les notes de produits, etc.).

Gérer les problèmes de navigation dans les pages et de pagination.

Nettoyage et Structuration des Données:

Nettoyer les données extraites pour enlever les balises HTML inutiles et les données non pertinentes.

Structurer les données nettoyées dans un format utilisable (comme CSV ou JSON).

Analyse de Base des Données:

Réaliser une analyse de base, comme le comptage de mots fréquents, l'identification des produits les plus mentionnés, ou les tendances des avis des utilisateurs.

Rapport et Visualisation:

Préparer un rapport décrivant les méthodes, les défis rencontrés, et les résultats de l'analyse.

Utiliser des bibliothèques de visualisation de données pour représenter les résultats.

Outils et Technologies Nécessaires:

Python: Langage de programmation principal.

Bibliothèques Python:

BeautifulSoup pour le parsing HTML.

requests pour faire des requêtes HTTP.

pandas pour la manipulation et l'analyse des données.

matplotlib ou seaborn pour la visualisation de données.

Environnement de Développement: Jupyter Notebook ou un autre IDE de votre choix.

Gestion de Version: Git pour le contrôle de version et GitHub pour le stockage du code.

Livrables:

Code Source: Scripts Python documentés pour le scraping, le nettoyage, et l'analyse des données.

Jeux de Données Collectés: Fichiers de données (CSV ou JSON).

Rapport du Projet: Document décrivant les méthodes, les défis, et les résultats, accompagné de visualisations.

Évaluation:

Qualité du code et respect des bonnes pratiques de programmation.

Efficacité et fiabilité du scraper.

Clarté et utilité de l'analyse et des visualisations.

Complétude et professionnalisme du rapport.