

Projet de Web Scraping – Binôme / Trinôme

Prérequis : remplir la fiche d'information

[https://docs.google.com/spreadsheets/d/1JE2n0zvGrVOtVXP7A0v9HWmBgIcBZa4dN4UE5JxzEio/edit?usp=drive link](https://docs.google.com/spreadsheets/d/1JE2n0zvGrVOtVXP7A0v9HWmBgIcBZa4dN4UE5JxzEio/edit?usp=drive_link)

1. Objectif du projet

Ce projet vise à mettre en œuvre un pipeline complet de traitement de données issues du web : analyse, extraction, traitement (incluant l'export), visualisation et interprétation. Le travail est collectif mais les responsabilités sont individualisées.

2. Organisation du travail

Le projet est réalisé en binôme ou en trinôme. Les étudiants doivent se mettre d'accord sur la répartition des tâches et la documenter clairement dans le rapport et le fichier README.

3. Répartition des responsabilités

Trinôme :

- Analyse & Extraction : Étudiant 1
- Traitement & Export : Étudiant 2
- Visualisation & Interprétation : Étudiant 3

Binôme :

- Analyse & Extraction : Étudiant 1
- Traitement & Export : Étudiant 2
- Visualisation & Interprétation : Travail commun

4. Rapport écrit (≈ 3 pages)

Parties communes (note commune) :

- Introduction,
- Motivation de la thématique choisie
- Liste des sites web choisis
- Légalité et éthique
- Conclusion suite à l'analyse
- Résultats et interprétation

Parties individualisée/responsabilisée

- Analyse et extraction
- Traitement et export
- Visualisation et Interprétation

La participation des autres membres doit être mentionnée. La prise de recul à chaque étape est indispensable.

5. Code et implémentation

Le code est évalué indépendamment du rapport.

Il doit inclure : extraction, traitement & export, visualisation et interprétation.

Le choix des librairies doit être justifié et cadré sur Java ou Python

Outil de visualisation libre : excel, powerBI, D3js, etc

N.B : exemple complet vu en cours <https://www.zenrows.com/blog/java-html-parser-jsoup#export-data-to-csv>

6. Fichier README

Le fichier README doit contenir :

- Description du projet
- Étapes pour reproduire le projet
- Librairies utilisées
- Instructions d'exécution
- Répartition des rôles

7. Règles importantes

Le rapport est noté.

Le code est noté.

L'entraide est obligatoire.

Le manque d'implication impacte la note individuelle.

Tout plagiat entraîne une sanction.