Sarah-Maude Gagné   
& Heidi Lavoie

Techniques de l’informatique

Groupe : 04318

Projet 2

Travail présenté à

M. Nicolas Payre

Département d’informatique  
pour le cours

*Exploration de nouvelles technologies*

Cégep de Sherbrooke  
20 octobre 2025

Table des matières

[Description du site et des données choisies 3](#_Toc210637893)

[Outil de webscraping utilisé 3](#_Toc210637894)

[Liens 3](#_Toc210637895)

[Base MongoDB Atlas 3](#_Toc210637896)

[Dépôt GitHub 3](#_Toc210637897)

[Méthode d’importation des données dans MongoDB Atlas 4](#_Toc210637898)

[Schéma des documents dans votre collection 5](#_Toc210637899)

[3 requêtes de sélection 6](#_Toc210637900)

[Conclusion 7](#_Toc210637901)

[Tableau indiquant la contribution de chaque coéquipier 8](#_Toc210637902)

# Description du site et des données choisies

**Site choisi :** Best Buy, site de vente de produits électroniques, d’électroménagers, de divertissement… (Produit divers)

**Données choisies :** Les données choisies sont la liste des accessoires de téléphone.

**URL cliquable :** <https://www.bestbuy.ca/fr-ca/categorie/accessoires-pour-cellulaires/27236>

# Outil de webscraping utilisé

**Outil choisi :** Scrapy

**URL cliquable :** <https://www.scrapy.org/>

**Fonctionnement de l’outil :**

1. Création d’un projet Scrapy
2. Exécution d’un « *Scrapy spider* » (pour récupérer les données du site)
3. Exporter les données récupérées (format JSON, CSV ou XML)
4. Utiliser « *Scrapy Shell* » (pour tester et débugger la logique *Scraping*)

Scrapy est un *framework* Python *open-source* qui extrait des données de sites web en envoyant des requêtes HTTP, en téléchargeant le contenu des pages, en le parsant avec des sélecteurs CSS ou XPath pour en extraire les informations (utilisant les « *Spiders* »), puis en structurant et en stockant ces données via des pipelines. Son fonctionnement est dirigé par un moteur qui coordonne les araignées (« *Spiders »*) et l'ordonnanceur pour gérer le flux de requêtes et de réponses, offrant ainsi un moyen efficace et extensible de réaliser du web *scraping* pour des projets de toute envergure.

Source : <https://datascientest.com/scrapy-web-scraping-tout-savoir>

# Liens

## Base MongoDB Atlas

## Dépôt GitHub

# Méthode d’importation des données dans MongoDB Atlas

# Schéma des documents dans votre collection

# 3 requêtes de sélection

|  |  |
| --- | --- |
| **1.** | Requête :  Explication : |
| **2.** | Requête :  Explication : |
| **3.** | Requête :  Explication : |

# Conclusion

Impression des technologies utilisées :

# Tableau indiquant la contribution de chaque coéquipier

|  |  |  |
| --- | --- | --- |
|  | **Tâches réalisées** | **% participation** |
| **Heidi Lavoie** | Création et rédaction du rapport  Extraction de 2500 éléments  Exporter les données dans un fichier CVS | 40% |
| **Sarah-Maude Gagné** | Recherche du site  Créer une base de données dans MongoDB Atlas  Importer le fichier CSV dans une collection | 40% |
| **Ensemble** | Recherche de l’outil de webscraping  Production de 3 requêtes de sélection (find)  Schéma des documents dans la collection | 20% |