Samuel Cloutier   
William Lauzon  
Jacob Perreault  
Antoine Plamondon  
Techniques de l’informatique  
groupe : 04320

JAWS

Projet 2 présenté à  
M. [Nicolas Payre](https://cegepsherbrooke-estd.omnivox.ca/estd/aens/InfoProf.ovx?P=RUVyUWN0SzA3Y0RVdkEvNUF3empxRUN5T2JveGFMMEFoVEVZc0doWGRDaz0_&AnSes=20253Ref=083924126736&C=SHR&L=FRA)   
pour le cours  
*Exploration de nouvelles technologies*

Cégep de Sherbrooke  
7 octobre 2025

# Présentation du site

Dans le cadre de notre projet, nous avons retenu le site de la compagnie Best Buy, spécialisée dans la vente de produits électroniques. Nous avons plus particulièrement choisi d’extraire et d’analyser les données relatives aux ordinateurs de jeu, tant de bureau que portables.

Lien vers le site : [*Site de Best Buy*](https://www.bestbuy.ca/fr-ca?cmp=knc-c-1291706821&g=&gad_campaignid=1291706821&gad_source=1&gclid=CjwKCAjwisnGBhAXEiwA0zEOR79nCW4z8hItUzZ0xEDS3sU2exRq3LyNIQWJ63WCIN0SRN1wrp39nBoC-ugQAvD_BwE&gclsrc=aw.ds&ordinateurs+best+buy=&utm_creativeformat=standard-text&utm_id=240815chw917&utm_medium=troas&utm_source=google)

# Présentation de l’outil

Web Scraper est une extension gratuite pour Google Chrome qui permet d’extraire automatiquement des données de sites web sans connaissances avancées en programmation. Elle est particulièrement utile pour collecter des informations structurées comme des fiches produits, des prix ou des images, par exemple pour faire de la veille concurrentielle ou constituer des jeux de données en vue d’analyses. Son fonctionnement repose sur la création d’un “sitemap”, c’est-à-dire un plan d’extraction qui définit la structure du site à parcourir et les éléments à récupérer. L’utilisateur sélectionne visuellement sur la page les données qu’il souhaite extraire (texte, liens, images, attributs, etc.) grâce à une interface simple utilisant des sélecteurs CSS ou XPath. Une fois le sitemap configuré, l’outil parcourt automatiquement les différentes pages (y compris la pagination) et stocke les données récoltées localement. À la fin du processus, ces données peuvent être exportées en CSV ou en JSON pour être exploitées ailleurs. Web Scraper se distingue par sa simplicité et sa compatibilité avec les sites dynamiques, mais il présente aussi certaines limites : il ne contourne pas les systèmes anti-bot complexes, peut être lent sur de très gros volumes, et ne fonctionne que sur les navigateurs basés sur Chromium.

# Méthode d’importation des données

Pour importer les données du site web choisi vers notre base de données, nous avons utilisé Web Scraper afin d’extraire les informations du site. Une fois l’extraction terminée, il faut télécharger le fichier CSV généré par l’extension.

Ensuite, ouvrez votre compte MongoDB, accédez au cluster souhaité, puis sélectionnez ou créez une collection. Créez ensuite la base de données ainsi que les champs nécessaires.

Enfin, choisissez l’option « Insérer un document », cliquez sur l’icône {}, puis copiez-collez le contenu du document dans le champ d’insertion de données. Terminez en enregistrant le tout pour finaliser l’importation.

# Description de la table

Voici les éléments qui sont compris dans notre sélection d’ordinateur de jeux électronique :

* ID : int
* Catégorie : string
* Sous-catégorie : string
* Marque : string
* Modèle : string
* Processeur : string
* Carte graphique : string
* Mémoire RAM (Go) : int
* Stockage (Go) : int
* Type de stockage : string
* Écran (pouces) : int
* Prix (CAD) : int
* Note des clients sur 5 : int
* Poids (kg) : int
* Stock disponible : int
* Couleur : string
* Système d’exploitation : string
* Réseau : string
* Ports USB : int
* Année du modèle : int
* Description : string
* Pays : string
* Devise : string
* Date de génération : string
* Nom du produit : string
* Prix du produit : string
* URL du produit : string
* URL vers l’image du produit : string
* Expédition du produit : string

# Requêtes de sélection

**Requête 1 :**

Nom : {SELECT marque, COUNT(\*) AS nombre\_items FROM produits GROUP BY marque ORDER BY marque ASC; }

A screenshot of a computer

AI-generated content may be incorrect.

Explication : Cette requête sert à rapidement avoir une idée de l’inventaire pour chaque marque d’ordinateur. Elle nous permet de voir le nombre d’item par marque.

**Requête 2 :**

Nom : SELECT marque, COUNT(\*) AS nombre\_items, AVG(prix\_cad) AS prix\_moyen, AVG(note\_clients\_sur\_5) AS note\_moyenne FROM produits GROUP BY marque ORDER BY prix\_moyen DESC;

Une image contenant texte, capture d’écran, ligne, Police

Le contenu généré par l’IA peut être incorrect.

Explication : Cette requête permet d’avoir, pour chaque marque, le nombre d’item, le prix moyen des items et aussi la moyenne de la note sur 5 données par les clients.

**Requête 3 :**

Nom : SELECT carte\_graphique, AVG(mémoire\_ram\_go) AS moyenne\_ram, AVG(stockage\_go) AS moyenne\_stockage, COUNT(\*) AS nb\_modeles FROM produits GROUP BY carte\_graphique ORDER BY moyenne\_ram DESC;

A screenshot of a computer

AI-generated content may be incorrect.

Explication : Cette requête permet d’avoir des renseignements sur les cartes graphiques, rams et stockages et s’avoir combien de modèle ont ça.

# Lien du projet

Lien vers notre dépôt GitHub : [Dépôt GitHub](https://github.com/allo57/The-Webscrappers)

Lien vers notre base de données MongoDB :

Lien vers le site web utilisé : [*Site de Best Buy*](https://www.bestbuy.ca/fr-ca?cmp=knc-c-1291706821&g=&gad_campaignid=1291706821&gad_source=1&gclid=CjwKCAjwisnGBhAXEiwA0zEOR79nCW4z8hItUzZ0xEDS3sU2exRq3LyNIQWJ63WCIN0SRN1wrp39nBoC-ugQAvD_BwE&gclsrc=aw.ds&ordinateurs+best+buy=&utm_creativeformat=standard-text&utm_id=240815chw917&utm_medium=troas&utm_source=google)

# Conclusion

Division des tâches

|  |  |  |
| --- | --- | --- |
| **Nom de la tache réalisé** | **Temps de travail (heures)** | **Pourcentage de participation** |
| **Section Samuel** | | |
| Scrapper le site BestBuy | 5 | 100% |
| Section présentation de l’outil (Rapport) | 0.5 | 100% |
| Rapport | 1 | 25% |
| **Section William** | | |
| Création du cluster MongoDB Atlas et configuration du projet | 2 h | 100% |
| Création de la base de données et des collections The-Web Scrapers et Produit | 1h | 100% |
| Importation des données JSON dans MongoDB Atlas et validation des documents | 2h | 100% |
| Rédaction du rapport (section William et documentation technique) | 0.4 | 100% |
| Rapport | 1 | 25% |
| **Section Jacob** | | |
| Création du dépôt GitHub | 0.1 | 100% |
| Rapport | 1 | 25% |
|  |  |  |
| **Section Antoine** | | |
|  |  |  |
|  |  |  |
| Rapport | 1 | 25% |