|  |  |
| --- | --- |
|  |  |
| Certification bloc 1 |  |
|  | Greta Centre-Val de Loire |
| Réaliser la collecte, le stockage et la mise à disposition des données d’un projet d’intelligence artificiellePrésenté par Maximilien PROUST  |  | | --- | |  | |  | |  |
|  |  |

Table des matières

### Introduction :

Ffffffffffffffffff

### Scrapping :

#### Les spécifications techniques :

Il est important de comprendre quel peut être l’impact des substances chimiques présents dans les polluants. Pour cela, je me suis appuyé sur le site web sagepesticides[[1]](#footnote-1). En effet, ce site web autorise le scrapping comme le mentionne son robot.txt qui affiche la mention user : \* allow : /. Pour l’extraction, j’ai utilisé un script python avec comme bibliothèques : requests, BeautifulSoup, json, re et logging. L’objectif d’utiliser la bibliothèque json est de permettre un enregistrement des données récupérées dans un fichier json et ainsi de pouvoir le rentrer dans une base Mongodb. Enfin, j’ai ajouté dans mon script un délai d’une à deux secondes afin de ne pas surcharger le serveur et éviter ainsi les éventuelles saturation serveur.

#### Les données obtenues :

Au niveau des données récupérés, j’ai choisi de conserver tout ce qui pouvait avoir un intérêt pour la compréhension des molécules chimiques et leurs effets néfastes sur l’humain, puis sur l’environnement. En Annexe 1, vous pouvez retrouver une capture d’écran du site web complet d’une substance recherchée. Le premier bloc, va contenir des informations générales sur la substance. Ce qui va m’intéresser ici va être la description de la substance, indiquant notamment de façon générale l’utilisation de celle-ci. Pour récupérer cette information, j’ai simplement utilisé les balises html présentes.

Annexes :

Annexe 1 : Visualisation du résultat de la recherche d’une substance chimique sur le site web :  https://www.sagepesticides.qc.ca/

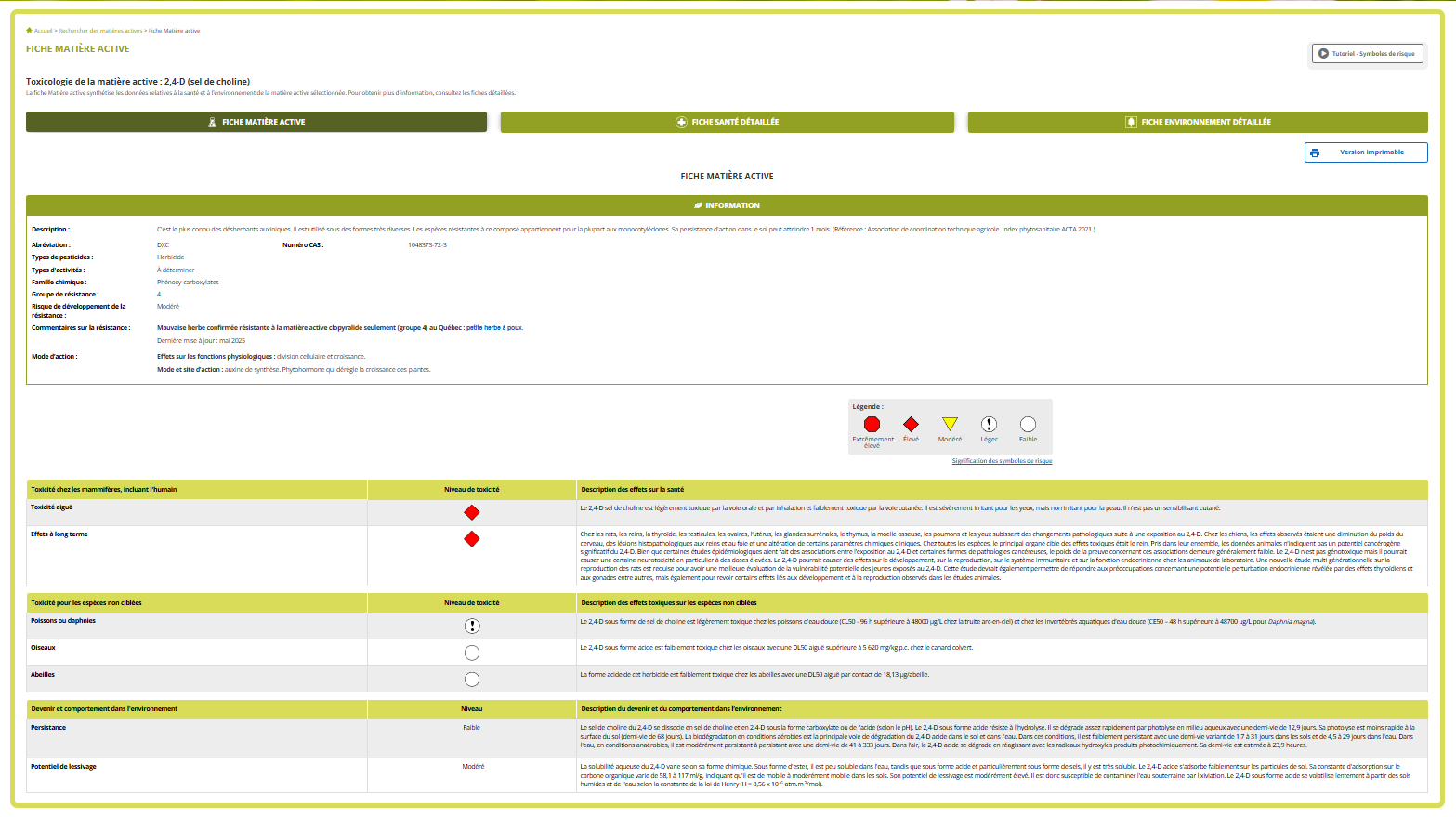


Figure : Visualisation des données obtenues sur la recherche d'une substance

1. « https://www.sagepesticides.qc.ca/ ». [↑](#footnote-ref-1)