

Projet python

*Agrégateur d'informations sur
les cryptomonnaies*



Membres du groupe :

Hakim SAGHIR

Toky Cedric ANDRIAMAHEFA

Thomas REMY

(3IABD – GP2)

Contexte :

Dans le cadre du cours de Scripting Python, on nous a demandé de développer une application en python afin de mettre en pratique toutes les techniques vues en cours. Nous avons choisis le thème des cryptomonnaies.

Description du projet :

Le sujet du projet, est de faire une application qui scrappe des données de cryptomonnaies sur le site <https://coinmarketcap.com/>, des articles sur le site <https://cryptonaute.fr/> les sauvegarder sur une base de données, ensuite de les mettre en valeur (graphes, tableaux ...) et de les présenter sur une page web.

Technologies utilisées :

- Langage : Python
- Base de données : MongoDB
- Frameworks : Streamlit, Pandas, Scrapy, Matplotlib, Pymongo

Sources des données :

- <https://coinmarketcap.com/>
- <https://cryptonaute.fr/>

Fonctionnement de l'application :

- On peut lancer l'application via le main à la racine du projet directement, dans ce cas la base de données de sauvegarde sera tronquée, puis alimentée avec les nouvelles données scrapées, ensuite le serveur de streamlit se lancera et affichera la page WEB.
- On peut aussi lancer les scraper individuellement, pour cela il faut utiliser la commande suivante :

```
$ scrapy runspider chemin_vers_le_spider/nom_du_spider.py
```

Ou

(Nécessite d'être dans le dossier du projet)

```
$ scrapy crawl nom_du_spider
```

- On peut tester les balises HTML/CSS avec lesquelles on scrape nos site en entrant dans le shell de scrapy avec la commande suivante :

```
$ scrapy shell url_du_site_à_scaper
```

On arrive au shell de scraper, on rentre par exemple :

```
>> response.css('.balises_CSS ::type_de_valeur_à_extraire').extract()
```

Ou en HTML, par exemple :

```
>> response.xpath('balise_HTML/text()')
```

References :

- <https://scrapy.org/>
- <https://streamlit.io/>
- <https://stackoverflow.com/>
- <https://pandas.pydata.org/>

Répertoire git :

- <https://github.com/TokyC/PythonCrypto>

Difficultés rencontrées :

- Choix du site à scraper : plusieurs sites génèrent automatiquement les balises HTML/CSS, donc il nous fallait bien choisir les bonnes balises afin de garantir le fonctionnement de l'application à long terme.
- Utiliser des nouveaux frameworks avec lesquels on est pas très familier : Pandas, Streamlit, Scrapy.

Conclusion :

Ce projet, nous a permis de renforcer et de mettre en pratique les techniques vues en cours et découvrir encore plus de techniques et de technologies.