Documentation :

Scraping project at Elba :

This directory contains a web scraper for the website [welcometothejungle](https://www.welcometothejungle.com/) and a processing script for the results of the scraper

If you want to launch the scraper, all commands are reminded in *“Scrapers/COMMANDS.txt”*

-Go to the Scraping directory, in my case on the Desktop

Une image contenant texte

Description générée automatiquement

-Go to the virtual environment directory

Une image contenant texte

Description générée automatiquement

-Activate the virtual environment

Une image contenant texte

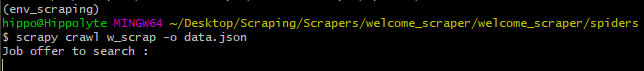
Description générée automatiquement

-Go to the spider directory

Une image contenant texte

Description générée automatiquement

-Run the spider and download the data

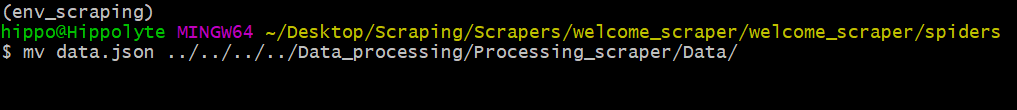


-Check if your data were well saved

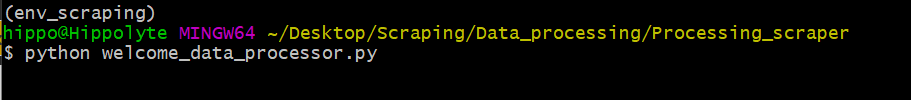
Une image contenant texte

Description générée automatiquement

-Move data for processing



-Finally execute the *Processing\_scraper\_data* file in the *Data Processing* directory or if you do not work on windows execute the python script this way :



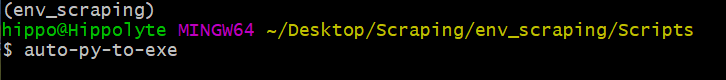
-You will find the results of your analysis in the *Results* directory

If you want to know which settings were changed or added for the scraper, check the *spider settings* in the *files* directory.

If you don’t want to use chrome, you will have to change the path towards the user agents json file in the settings file of the spider. Other lists of user-agents are in *“files/User agents”*

You will also have to modify the w\_scrap.py file, replacing all chrome features by the chosen web driver’s

Then if you want to generate a new .exe file, go back to your terminal, activate the virtual environment and type this command :



It will open this window :

Une image contenant texte

Description générée automatiquement

-Enter the script location, which is *“Scraping/Data\_processing/Processing\_scraper/welcome\_data\_processor.py”*

-Click on *“One Directory”*

-Click on *“Console Based”*

-In *“Additional Files”, click on “Add Folder”* and add the *“Data”* and *“Results”* folders

-In *“Settings”* select the *“Output Directory”*

-Finally click on *“Convert .PY to .EXE”*