DK Crawler Instruction

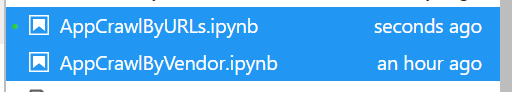
# Installation – install.ps1

Install the necessary packages by running install.ps1 (double click or right click and run with PowerShell). This process will take about 15 minutes and only needs to be done once in the beginning. But it should not break anything if you run it more than once.

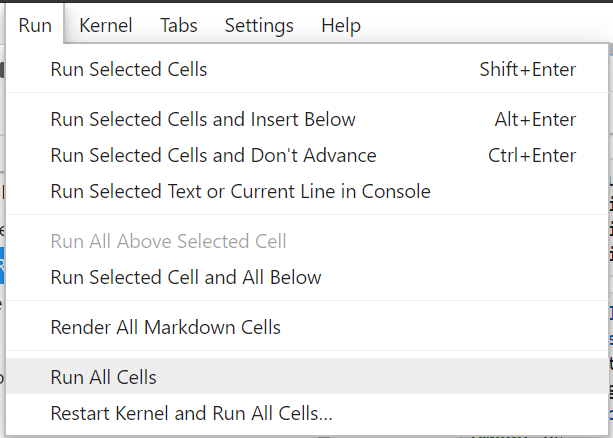
# Start Jupyter Lab – start.ps1

Run start.ps1 to start jupyter lab. This will start a powershell window that hosts the jupyter lab application. A new browser window will pop up with jupyter lab.

Once Jupyter Lab is opened, double click the “.ipynb” files to open the App.

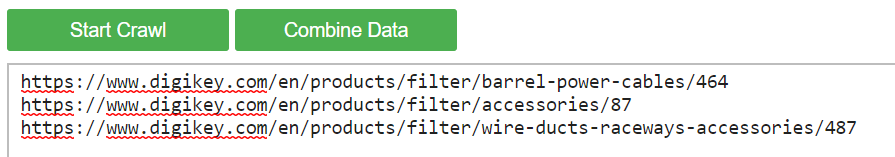


To run the app, you can select Runn All Cells. Or select each cell and press shift+enter.



Follow the instruction to input the required URLs and click “Start Crawl”.

The click “Combine Data” after the crawl has finished:



# Restore and Update – update.ps1

Note that any files with “.py”, “.ps1” or “.ipynb” extensions should not be modified by user. If any modification has occurred and broken the program. Run update.ps1, it will restore the project to the unmodified state and then update the project to the latest version.