

建立開發環境

郭耀仁

Agenda

- Python 開發環境
- 命令列
- 虛擬環境
- 爬蟲所需模組與 Chrome 外掛

Python 開發環境

- 瀏覽器
 - Google Colaboratory
- 本機
 - Anaconda

直接在瀏覽器使用

- Google Colaboratory (<https://research.google.com/colaboratory/faq.html>).

操作步驟

- 登入 Google 帳號，開啟雲端硬碟
- 點選「新增」
- 點選「連結更多應用程式」
- 搜尋「Colaboratory」
- 點選「連結」
- 新增 Google Colaboratory
- 完成瀏覽器的開發環境建置

操作步驟截圖

建立瀏覽器的開發環境 (<https://www.datainpoint.com/python-essentials/01-quick-start.html#%E5%BB%BA%E7%AB%8B%E7%80%8F%E8%A6%BD%E5%99%A8%E7%9A%8C>)

安裝 Anaconda

- Anaconda (<https://www.continuum.io/downloads>).

安裝 Anaconda (2)

- 我為什麼推薦直接使用 Anaconda (<https://www.anaconda.com/download/>)?
 - 有 Python 直譯器
 - 有編輯器 Jupyter Notebook
 - 有套件與環境管理工具 conda
 - 預先安裝好許多模組套件

安裝 Anaconda (3)

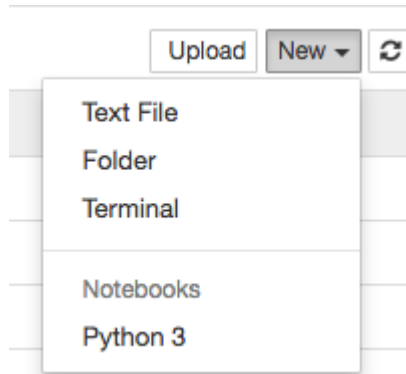
- 注意事項
 - 將 Python 加入路徑變數

啟動 Jupyter Notebook

- 在命令列輸入

```
$ jupyter notebook
```

- 新增一個 Notebook



補充資訊

- 命令列
- 虛擬環境

命令列

命令列

- Windows 使用者請下載 cmdr (<http://cmdr.net/>).
- macOS 使用者請啟動 Terminal 終端機應用程式 (Spotlight search -> "Terminal")

命令列是什麼？

- 電腦並不是一開始就有圖形化介面（Graphic User Interface, GUI）
- 也不是一開始就有滑鼠
- 因此跟電腦的溝通工具是命令列

基礎命令列指令

- pwd
- ls
- cd
- mkdir
- touch
- mv
- cp
- exit

虛擬環境

建立適當的虛擬環境

- 為了避免套件版本衝突，可以建立資源隔絕的虛擬環境

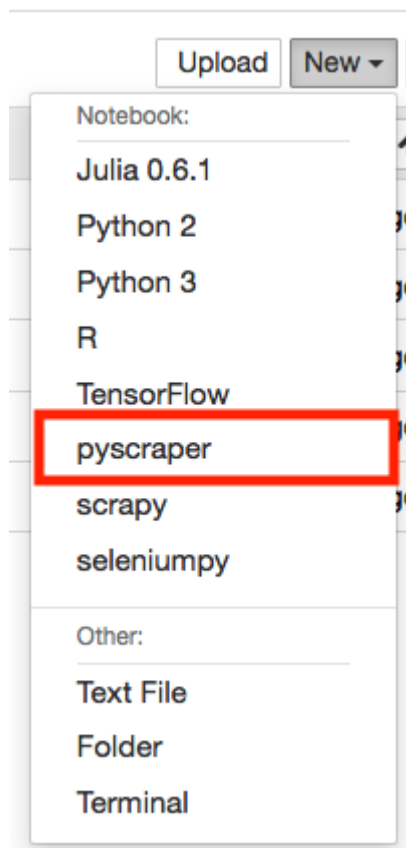
```
conda create -n pyscraper python=3  
conda activate pyscraper  
pip install ipykernel
```

建立對應虛擬環境的 Kernel

- 連結 Jupyter Notebook 與虛擬環境

```
python -m ipykernel install --user --name pyscraper --display-name "pyscraper"
```

新增一個 pyscraper 的 Notebook



移除 Kernel

```
jupyter kernelspec list
```

```
jupyter kernelspec remove pyscraper
```

移除虛擬環境

```
deactivate
```

```
conda env list
```

```
conda remove --name pyscraper --all
```

隨堂練習

自行操作一次新增虛擬環境與 Kernel

爬蟲所需模組與 Chrome 外掛

爬蟲 Python 模組

模組名	功用
requests	對網站發送請求獲取資料
json	解析 JSON 資料，將 JSON 寫入 Python 成為 dict；將 dict 寫出成為 JSON
pyquery	對網站發送請求獲取資料與解析 HTML 資料
BeautifulSoup	解析 HTML 與 XML 資料
selenium	自動化瀏覽器

爬蟲 Python 模組文件連結

- requests (<http://docs.python-requests.org/en/master/>).
- json (<https://docs.python.org/3/library/json.html>).
- pyquery (<https://pythonhosted.org/pyquery/>).
- BeautifulSoup (<https://www.crummy.com/software/BeautifulSoup/bs4/doc/#>).
- selenium (<https://selenium-python.readthedocs.io/>).

爬蟲 Chrome 外掛

外掛名	功用
Quick Javascript Switcher	關閉 JavaScript 功能
JSONView	讓 JSON 資料格式在瀏覽器上呈現得比較漂亮
SelectorGadget	協助 CSS Selector 定位
XPath Helper	協助 XPath 定位

爬蟲 Chrome 外掛連結

- Quick Javascript Switcher (<https://chrome.google.com/webstore/detail/quick-javascript-switcher/geddoclleiomckbhadiapdggiiccfje>).
- JSONView
(<https://chrome.google.com/webstore/detail/jsonview/chklaanhfefbnpohckbnefhakgolr>
- SelectorGadget
(<https://chrome.google.com/webstore/detail/selectorgadget/mhjhnkcfbdhnjickkkdbjoen>
- XPath Helper (<https://chrome.google.com/webstore/detail/xpath-helper/hgimnogjllphhkhklmebbmlgjoejdpjl>).