**試題1**

1. 確認網頁文本格式：

<https://building-management.publicwork.ntpc.gov.tw/bm_list.jsp> 為HTML文本

1. 網頁是使用POST請求，確認需傳送的內容

{

rt: BM

PagePT: 0

A1:

A2V: "%A8%CF%A5%CE%B0%F5%B7%D3" # 職照類型: 使用執照

A2: 3

A3:

B1:

B2:

B3:

K1:

C1:

D1V: "%B7s%A5\_%A5%AB%AAO%BE%F4%B0%CF" # 建築地址-行政區

D1: 220

D2:

D3: "%A4T%A5%C1" # 建築地址-路街段

D4:

D5:

D6:

D7:

D8:

D9:

D10:

D11:

E1V:

E1:

E2V:

E2:

E3:

E4:

F1:

J1:

J2V:

J2:

J3:

Z1: 5506 #驗證碼號碼

}

1. 先取得驗證碼，將驗證碼圖片儲存，手動輸入於payloadData的Z1中
2. 使用urllib.request 套件，取得網頁原始碼，並以 Big5 解碼
3. 使用Beautiful Soup 模組解析html文本
4. 利用pandas處理表格資料，並儲存成Excel檔，或儲存其他檔案格式

進一步增加變化：

* 將傳送的資料，例如:D1、D3等等資訊設為變數，亦即傳送的城市、地區、路名資料設為變數，讓可抓取的資料更有變動性
* 全台灣的路名可從政府資料開放平台-全國路名資料中取得

驗證碼參考影片: <https://www.largitdata.com/course/60/>

政府資料開放平台-全國路名資料: <https://data.gov.tw/dataset/35321>

遇到的卡關：

無法順利抓取到附有表格的資料，卡在搜尋頁面的網頁資料。