

程式語言 HW2 報告

陳俊宇 F64051114

● 執行環境步驟

此次作業我在 ubuntu 16.04 環境中執行，因為使用到 python 繪圖套件 matplotlib，若該環境是第一次使用此套件需開啟終端機輸入以下指令：

```
>>sudo apt-get install python3 python3-pip
```

```
>>pip3 install numpy
```

```
>>python3
```

```
>>>import numpy as np
```

```
>>sudo apt-get install python3-matplotlib
```

而程式執行辦法，就是讓使用者輸入欲查找的作者名稱(如 lan+Goodfellow)，即可得到對應的輸出結果。

● 程式碼說明

此次作業有兩小題，我分成兩個檔案撰寫，分別是 Q1.py(第一題)與 Q2.py(第二題)。

而兩檔案的程式碼架構基本上大同小異，都是以 list 資料結構實作。透過抓取欲查找字串的前後文相同處，並用 regular expression 方式來表示前後文相同處的 pattern，接著使用 re.findall()來查找整篇網頁有此 pattern 的 string (re.findall() would return list type)，再透過一些字串處理技術(切字串、合併字串、擷取自串)，達到需要的結果(年份或是作者名稱)，並將該結果存入對應的 list entry 內(我的 list 是如此設計：

[[‘1997’,1],[‘1600’,8]...])，最後使用 list.sort()將年分(作者名稱)做排序。個人認為字串處理方面，Python 比 C 強大許多，使用起來相當直觀方便。

需特別注意的是有些搜尋結果不一定是一頁可以呈現的，故需實作翻頁功能。以 input lan+Goodfellow 為例，我是先將第一頁上方”Showing 1–50 of 63 results for author: lan Goodfellow”的”63”抓取下來，轉成數字後，除以 50(一頁有 50 個結果)並且無條件進位，用該值搭配迴圈的方式，將 url 末端加上”&start=0(50、100...)”，使 re.findall()結果全部存在一個 list 變數，接著再進行上一段所描述的工作。