## 商管程式設計期末專案報告

# **Programming of Business Computing final project**

Instructor: 孔今傑 教授

### Objective

在這次的期末報告之中,我們選擇了四個類型之中的小幫手,希望能夠從貼近學生生活的議題-實習,為學生提供有效的幫助。找實習佔據了管院學生生活的一大部分,然而,因為職缺散落在各個網站,而不同網站的資訊又有不同的格式,因此求職者往往需要花費大量時間與精力在找實習職缺上。



以上圖的 104 與 1111 為例,可以從圖發現,各家人力資源廠商的搜尋職缺 頁面有差異,不只是欄位而已,搜尋處的篩選條件也有許多不同,而排序方式也 是如此。

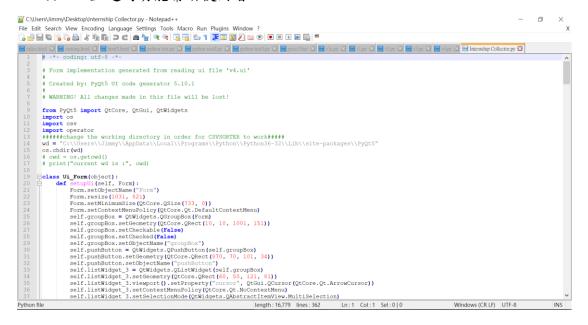
此外,就職缺的詳細內容而言,每個職缺都有相應的頁面,若僅有少數幾個 職缺並不會花費太多時間,但如果有超過百則以上的職缺,光是點開對應的網址, 便會花費大量的等待時間。

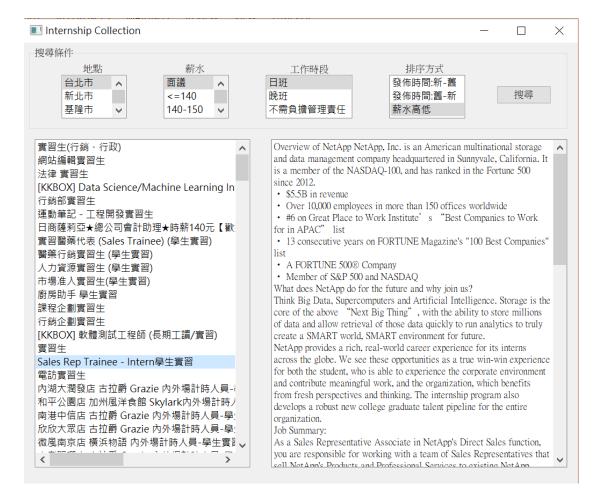


因此在這次的期末報告之中,我們希望能夠結合爬蟲、資料清理、桌面應用程式、搜尋引擎、排序等概念,寫出一個應用程式讓使用者可以有效率的搜尋職缺,並且也有清楚的職缺詳細內容呈現。

### Content

- **爬蟲**:針對提供職缺的目標網站104,1111,我們使用 Beautiful Soup 套件爬下網頁內容,主要的產出為: Date, job\_name, company\_name, Location, Job\_link, Job Description and Job Content 等欄位。我們怕對目標網站造成影響,也有使用random相關函數結合sleep,盡量減少打擾或是影響網站運作的可能。
- 資料整理:這個部份主要就是透過上課所教的字串處理以及活用資料型態觀念,舉例而言我們使用replace函數來刪除空格、空行、段落間距等,也調整了欄位的內容。此外,我們也整理了兩個目標網站之間的歧異,並且從中整理出一個適用所有資料的欄位,整理兩個網站的資料後,再進行整合統一格式。
- 桌面應用程式:我們使用PyQt5來設計我們的桌面應用程式,包含多個下拉式 選單讓使用者依其條件進行選擇,舉例而言,我們使用薪水、地區等作為條 件篩選,另外,也可以依照職缺貼出的日期進行排序,並且為了解決點開連 結的等待問題,我們將職缺詳細內容的版位整合在同一個頁面上,如此一來, 透過條件篩選、日期排序、顯示頁面優化等,應用程式會呈現符合標準的職 缺,以上述的功能幫助使用者。





#### Future work

- 廣度:這邊的廣度指的是爬蟲對象的廣度,這次我們主要爬了兩個網站,光 是這兩個網站就發現在資料欄位有許多無法統合的地方,我們原先還有爬下 yourator,但是後來發現yourator跟前兩者的欄位差距真的太大,無法使用。 因此,我們認為就廣度而言,要整理出一個共通的欄位並且整合進更多網站, 就算有些網站缺乏某些欄位,也應該要能夠做到這樣的標準化,因為只有規 模夠大才有使用的價值。
- 深度:就深度而言,我們想在桌面應用程式加入搜尋欄位,使用者可以在搜尋欄位中輸入關鍵字。依據特定規則進行排序,這裡指的不只是依照時間、薪水高低等基本的排序,我們可能會將規則延伸至更進階的優化,譬如說針對工作內容(job description)的相似度進行排序,或是依照多個關鍵字的先後順序進行排序等。