

商管程式設計期末書面報告

專屬於你的咖啡廳

第74組

B06703002 財金四 黃羿綺

B07610062 生傳四 羅敏

B07610063 生傳三 陳暄穎

B08701248 工管三 唐皓楷

B09701169 工管二 王誼宸

指導教授:孔令傑、盧信銘

壹、專案介紹

- 創作理念：去咖啡廳已形成一種趨勢。無論是背著一台電腦，想找個有插座又安靜的環境，讓自己專注於手邊的事；抑或是三五好友想找一間咖啡廳，品嚐甜點、聊聊近況，度過愉快的週末等等。
- 專案目標(What)：你曾為了選擇咖啡廳而煩惱嗎？曾經想找個安靜的空間認真讀書，卻去到一間非常吵雜的咖啡廳呢？或是曾經背著筆電，想找個有插座的空間，卻誤入歧途呢？「專屬於你的咖啡廳」致力於協助使用者依照自己的需求，從插座、wifi、安靜程度、到咖啡廳風格都能任君挑選，找出專屬於你的咖啡廳！
- Why：由於google map上查詢咖啡廳時只會出現評價、營業時間等等的資訊，若需要無wifi、插座、或是特定風格的其他相關資訊，則需要自行進入咖啡廳的網站或從客人評論中查詢，因此為了節省使用者挑選咖啡廳的時間，我們希望透過整合多種需求選項，讓系統篩選後推薦出符合使用者需求的咖啡廳。
- How：咖啡廳資料的搜集來自cafe nomad和愛食記兩個網站，cafe nomad中有咖啡廳的wifi穩定度、安靜程度及有無插座的資料；愛食記則是有評價星等、顧客評論、營業時間和地址。另外我們訓練愛食記評論抓風格關鍵字，從顧客評論中找到特定關鍵字作為咖啡廳風格的分類，將以上資料利用爬蟲存為csv檔後，結合GUI輸入和輸出，讓使用者輸入需求選項，最終呈現推薦的咖啡廳。

貳、執行計畫

1. GUI輸入輸出

(1) 讀取爬蟲下來的CSV檔，並將每個Column做初步的處理(包刮地區文字處理、風格、判斷能否轉型成float以利後續結果的輸出)

(2) 利用tkinter建立GUI

- 利用grid去做佈置
- 例用Radiobutton、Label、Combobox、Button、Text等功能
- 透過rowconfigure和columnconfigure調整視窗大小
- 建立各個按鈕的function
- 把視窗內的文字及輸出都調整成微軟正黑體

(3) 輸出

- 根據使用者按鈕所設定的條件去找出相對應的結果，並且按照評分高到低做排序，另外用f-string調整輸出的格式。

2. 爬蟲：

(1) Cafe Nomad：

- 爬蟲目的：蒐集全台各縣市咖啡廳的wifi穩定度、安靜程度、插座有無等資料
- 函式庫：利用requests套件擷取資料，並用pandas套件轉換資料結構
- 爬蟲步驟：
 - a. 使用requests套件中的get()函數取得Café Nomad的API資料，並以.json方法擷取之，得到一個包含所有咖啡廳資料的list，list中各咖啡廳的資料則以dict呈現。
 - b. 利用迴圈讀取每一筆咖啡廳的資料，分別將咖啡廳名稱、wifi、插座、安靜程度及限時與否共五項資料存入相對應的list
 - c. 建立新的dict，將上述五項資料設為key，value則是相對應的list
 - d. 使用pandas套件，將dict轉換為data frame的資料結構
 - e. 最後使用.to_csv方法，將資料寫出成csv檔

(2) 愛食記：

- 爬蟲目的：

抓取各店家資訊：店名、評分、連結。
其中連結為提供Instagram截圖程式碼利用。
- 函式庫：

利用Python套件requests及BeautifulSoup獲取網址資料。
- 爬蟲步驟：
 - a. 搜索方式：

利用愛食記依照地區分類的搜索條件尋找全台各縣市咖啡廳，一次搜索一個縣市。以下以「台北市」為例。
 - b. 翻頁處理：

複製網址貼至程式，進行爬蟲程序。將台北市的咖啡廳csv檔打開，最後將咖啡廳資訊存進去。愛食記翻頁後的網址有明顯的規律，為「?page=」加上頁碼，將各頁網址加進清單中。
 - c. 店名資訊：

將各頁網址的清單利用for迴圈一頁一頁處理，並找出一頁所有的店家連結，利用for迴圈一次只處理一個店家的資訊。
建立空白清單「one_shop_info」，將存進店家資訊：店名、評論、連結。

店家連結的後半部網址在店名的href標籤裡，利用["href"]獲取資訊。

一個店家的網頁內有店名資訊，但用此標籤會印出兩次店名，利用迴圈及break函數只取一次，並加一個店家的資訊清單。

d. 店家評分資訊：

取店家評分標籤資訊會出現「4.3 · 3則評論」的字串，但只需評分數字，所以將其加入清單處理，只取評分分數的字串。

e. 資訊存入檔案：

利用write函數將一家店家的資訊存入檔案。

3. 訓練愛食記評論抓風格關鍵字：

(1) 使用Beautifulsoup、request、selenium套件來獲取並解析網頁HTML碼以及操作瀏覽器。

(2) 在愛食記上搜尋「咖啡廳」共會出現3000多家咖啡廳，將每一頁搜尋結果設為一個迴圈，每頁搜尋結果中的每個店家也設為一個迴圈，一個店家點進去中的每一則評論也是一個迴圈。

(3) 透過beautifulsoup套件解析並以評論中的句子為單位。

(4) 我們將咖啡廳大致分為三大風格，分別是文青風、工業風與網美風。

以下為我們為不同風格所設置的關鍵字：

* 文青風：老宅、復古、古典、古董、老屋、文青

* 工業風：工業、公寓

* 網美風：網美、拍照、熱門、質感、打卡、超人氣、美、漂亮

(5) 接著在每間店所獲得的評論迴圈中檢查文章內是否含有我們所需要的關鍵字

(6) 若該店的評論內文符合「文青風」，就將該店的店名append到文青List內，以此類推

(7) 最後蒐集完三種不同風格的店家後，在原先爬下來的csv檔中新增一欄風格的欄位，將該咖啡廳所對應到的風格填入其csv檔中。

4. IG截圖

a. 使用requests、BeautifulSoup、lxml套件擷取愛食記網站的資料夾

b. 利用BeautifulSoup中的find()及find_all()函數，搜尋網站上的IG相片(共九張)，並用get()函數取得相片網址，將網址存入list中

c. 利用迴圈讀取存有相片網址的list，個別用requests中的get()函數取得相片資料後，建立並打開一png檔，將相片資料寫入

- d. 定義crop_center函數, 依照指定邊長從正中央裁剪相片, 避免後續組合相片時長寬無法對齊, 以及避免從邊緣裁減會將相片重點裁掉
- e. 利用Image套件打開九張相片, 將相片丟入crop_center裁剪後, 建立一新圖檔, 將九張相片貼上存成九宮格
- f. 將最終的九宮格照片依咖啡廳名稱存成png檔

參、分工方式

- GUI介面: 唐皓楷
- 爬愛食記評論: 羅敏、王誼宸、陳暄穎
- 訓練愛食記評論抓風格關鍵字: 陳暄穎
- 爬cafe nomad、IG截圖: 黃羿綺

肆、組員心得感想

1. 黃羿綺

沒想過經由這堂課初學程式的我, 能在期末和組員們共同產出完整的專案作品, 很謝謝組員們努力不懈地嘗試各種方法爬取資料, 在發現當初設想的執行方向不可行時, 馬上想出替代方案並成功地解決爬蟲上遇到的困難, 讓專案能順利進行。我在專案中也學習到了很多從未接觸過的程式套件, 尤其是在IG截圖的部分。原先預想是透過selenium操作, 登入IG後進入地標頁面截圖, 但速度太慢且始終無法截到滿意的圖, 因此改成擷取愛食記上的IG相片, 後來也分別使用selenium和BeautifulSoup, 順利地寫出兩種不同的截圖方法。雖然稱不上困難, 但寫出來的當下真的很有成就感! 畢竟也是花很多時間上論壇找相關資料和教學才完成的成果啊。最終來不及將相片這部分整合進我們的程式, 感到非常可惜, 但至少過程中有所收穫, 而且看著寫出來的程式能讓瀏覽器自動輸入帳號密碼和搜尋咖啡廳真的很有趣, 突然覺得寫了好幾個小時都值得了。感謝這學期老師及助教的教導, 讓我從零基礎到現在對Python有基本的理解, 且未來也能自主學習未接觸過的程式! 也謝謝所有組員的努力合作, 共同完成了這次的期末專案, 辛苦了!

2. 羅敏:

看完老師的課程後, 以為可以很快解決爬蟲的部分, 結果一開始就撞牆。原本要用Google地圖抓大部分的資料, 結果發現Google地圖的API需要付費, 後

來就改用愛食記的資料。這中間的過程有掙扎了一下，上網查了很多國內外的論壇，也花蠻久時間的，這過程中了解到爬蟲有一定的難度，各式各樣的網站有不同的編排方式等。感謝組員說要直接換網站爬資料，要不然我會自己在那邊研究個半天，也沒產出什麼有用的資訊。藉由這次經驗，也了解到設立停損點的重要性，讓自己更有效率，不一定要追求原本預想的那個樣子，隨機應變會讓專案有比較好的結果。

找愛食記評論的時候一直卡住，後來才發現是最前面的程式碼出錯，出錯的程式碼是直接沿用別人的，自己也沒有檢查到那個部分。團體合作撰寫程式最常遇到的應該就是這類的問題了，debug還是要檢查一下別人寫好的，要不然再沿用下去可能會出更多的錯。

做完專題之後可以用爬蟲搜索網路資料做報告等等，超級實用，最重要的是我可以跟大家說我會爬蟲，聽起來很厲害.....

3. 陳暄穎：

身為程式初學者，同時也是咖啡廳愛好者，在期末可以和組員做出一個自己也覺得很實用的專案，把興趣和所學結合在一起的感覺真的好充實！這學期一路從list、string、dictionary、class，到期末專案開始學會使用套件、自己時做了聽了很久的名詞「爬蟲」，真的可以明顯感受到自己的進步！非常感謝自己的組員都超級厲害，在需要幫忙的時候都會提點我或是一起解決問題。回想到一開始我們決定要使用Google評論，但卻卡在Google API要付費或是很多程序問題，光爬蟲就卡了好多天，最後在向助教諮詢過後也學到要怎麼量力而為，以我們可以做得出來的程式為主，希望未來自己可以持續精進，再回過頭來優化「專屬於你的咖啡廳」！謝謝老師、助教整學期用心的教導，也超感謝組員的付出！大家都辛苦了～

4. 唐皓楷

很難想像以前討厭寫程式的我，盡然會因為這門課，重新點燃我對程式的熱情，且在期末能和組員們一同做出以往看似很難的Project！在挑戰每週的作業中，從Wrong Answer改到Accepted的那種興奮感，實在無法言喻。另外期末專案和組員一起討論並且做出完成品的過程，不僅讓我學會如何爬蟲、整合大家的程式碼、到最後GUI的輸出，更難能可貴的是大家一起互相激盪出不同的想法並且解決我們所沒遇過的問題。這學期非常感謝老師、助教及組員們，大家都辛苦了～

5. 王誼宸：

有別於過去有題目和測資的作業，這次的專案需要自己設計一個主題並做出來，我認為是一大挑戰。過去幾週的作業我都需要花相當多的時間才能完成，因此一開始有點擔心我能不能完成這次的專案，幸好有學長姐的大力幫助，大家不斷討論和修改，最後才能順利完成。藉由這次的專案，配合過去整個學期循序漸進的課程，我才真正了解到所謂的「爬蟲」，並運用許多從沒學過的套件，例如：selenium、beautifulsoup等等來實際操作，真的獲益良多，希望未來我能夠精進自己的能力，寫出更好的程式。這學期感謝老師及助教的協助，也謝謝學長姐在期末專案的互相幫忙！