

## 108-1 商管程式設計期末專案

組別： 第 23 組

主題： 電影查詢小幫手

組長： 馬愷若 B06703094

組員： 鄭皓天 B06703129

吳培瑜 B06703009

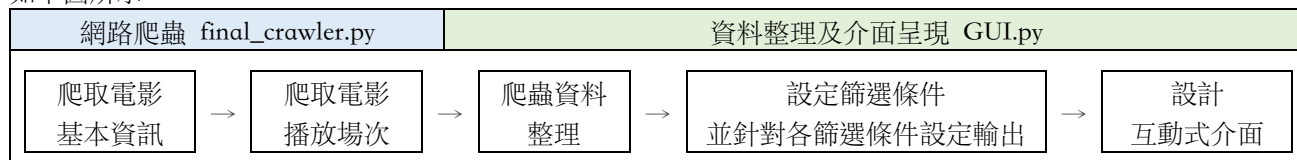
蔡曉淇 B06703128

杜 昕 B06702026

吳芃蓓 B06702053

## 一、專案簡介

「電影查詢小幫手」係藉由 python 從網路上取得台灣電影的相關資訊，進一步進行較複雜的資料分析與篩選，最後再由 python 設計出 GUI 的介面供使用者進行電影相關資訊及放映場次之查詢，整體系統運作步驟大致如下圖所示：



以下將介紹電影查詢小幫手之系統設計及使用者介面呈現，最後附上組內成員於本次期末專案之心得。

## 二、系統設計——網路爬蟲

於 final\_crawler.py 中，我們利用 python 中的 selenium 作為爬蟲的工具。Selenium 是一種自動化測試工具，因為與網頁的交互能力極佳，可以模擬登錄、滑鼠點擊、鍵盤輸入等，所以適合抓取動態頁面的資料數據，而此特性很適合爬取電影相關資訊，因為爬取時需要不停點擊與退出並尋找下一部電影的資訊及各個日期的時間，而它的缺點則是爬取速度較慢，因為需要等待整個頁面加載完成後才能去爬取頁面元素，接著利用 selenium 的定位元素方法，如 find\_elements\_by\_xpath()、find\_elements\_by\_name() 等等去尋找所需要的資訊，因此需要花費較多的時間。

而在爬取電影資料時，我們使用的網站是 yahoo 奇摩電影(<https://movies.yahoo.com.tw>)，在爬蟲時共分成以下幾個步驟：↓

爬取本週排行榜	進入 yahoo 奇摩電影後，第一件事是爬取首頁側邊欄的本週前十名排行榜，並作為最後輸出檔案的第一行資料，並將檔名依當天日期命名(movie yyyy-mm-dd)
↓	
爬取電影網址	接著，我們利用迴圈，讓瀏覽器每一次都直接把網頁中的頁數加一並 get 到上映中的網址，並利用 find_elements_by_name() 找到符合的元素再利用 get_attribute('href') 得到各自的網址，若網址開頭為 <a href="https://movies.yahoo.com.tw/movieinfo_main/">https://movies.yahoo.com.tw/movieinfo_main/</a> 代表是電影相關網址，就加入清單中。至於本週新片也是類似的方法，只是在抓取網址前還要另外先抓取電影上映日期，並與今天日期比較，若比今天晚才會做後續動作。
↓	
爬取清單電影之 資料與播放場次	得到所有電影的網址之後，便依序點擊各個網址，利用 find_elements_by_xpath() 抓取我們所需的各項資料，如電影長度或電影類型等等，並依序點擊不同天的時刻表，若有我們選的那八間電影院的場次就抓下來。 每部電影抓取的資訊會有五行，第一行為電影名稱、電影類型、上映時間、電影長度、IMDb 分數，第二到五行分別為當日及之後三日的那八家電影院的電影時刻，若那天沒有那八家的場次則寫 N，若那天的時刻表不能點擊（可能時刻表還沒出來），顯示 no info。

### 三、系統設計——爬蟲資料整理與篩選

於 GUI.py 中，我們先將整理使用 selenium 取得的電影資料，將其整理成以便進行篩選及呈現之資料，再資料整理時共整理成以下三個重要的 list，作為之後查詢及輸出的依據：

rank_list	將檔案第一行之排行榜資料以列表儲存
movie_list	將每一部電影第一行之電影資料儲存，包括電影名稱、電影類型、上映時間、電影長度、IMDb 分數，但若該電影之二至五行皆沒有任何播放場次，則該電影不會被記錄於列表中
movie_time_list	movie_time_list 是整份程式碼最主要的列表，裡面共有多個列表，每個列表裝著一部電影當日及之後三日之所有播放資料，因此共包含相當多的列表及 datetime 資料組成

將電影資料整理成以下三個 list 之後，接下來將用這些資訊進行電影篩選，而在電影查詢小幫手中，共有以下三種篩選方式，其操作原理方別如下：

搜尋電影資料	input：電影名稱，output：電影相關資訊 將會使用 movie_list 作為搜尋的依據
搜尋電影類型	input：電影類型，output：該類型上映中的電影名稱 將會使用 movie_list 作為篩選的依據
搜尋電影場次	input：欲搜尋資料，output：搜尋結果所有符合場次 在此項目一共有五個篩選條件，分別是電影名稱、電影院、日期、地址、時間，每一項都是非必填，以增加程式的活用程度，但也因此增加了不少處理難度。 於電影名稱、電影院、日期、時間的部分我們將以 movie_time_list 作為篩選依據，將不符合篩選條件的播放資料自列表中剔除，其中時間的部份以差距兩個小時為限(例如：輸入 20:00，將會搜尋 18:00 至 22:00 之資料) 而地址的部分則是附加的條件，若輸入地址，則輸出之電影資訊將會以電影院距離的遠近排序，在此我們使用的是 googlemaps 的模組，此模組可和 google maps 連接，我們預先輸入八間電影院的地址，於使用者輸入地址時，程式會將輸入地址和電影院之地址一一計算開車所需的時間，並將結果依時間由短至長排列

### 四、系統設計——GUI 介面設計

我們使用 python 內建的 tkinter 函式庫來設計使用者互動式介面，運用到的函式有 Canvas、Label、Entry、Combobox、Button、Text、Scrollbar 等功能。在介面呈現上，我們將所有固定的項目設計成背景圖片，再使用 Canvas 將其匯入，如右圖所示。

介面各項目運作形式方別如下：

排行榜	透過讀取 rank_list 中的資料後，再寫成 Label 的樣式呈現。
電影資訊查詢	即上述的搜尋電影資料及搜尋電影類型，用 movie_list 中的資料建立 set 選單(即 Combobox)，以省去使用者輸入上的麻煩，亦避免使用者輸入與我們建立的資料庫有出入。 按下該方框中的"搜尋"後，會彈跳出新的視窗顯示出搜尋結果，以 Label 函式顯示
電影類型查詢	
電影時刻表查詢	即上述的搜尋電影場次，其中電影名稱、電影院、搜尋時間為下拉式選單，地址和輸入時間是輸入式空格，而五項皆為自由輸入，但有以下規定： ✓不得同時輸入電影院及地址 ✓必須輸入日期才能輸入時間 ✓不得皆不輸入直接搜尋，或只輸入地址 按下該方框中的"搜尋"後，若違反上述任何一項規定，會彈跳出新的視窗顯示出搜尋錯誤，若輸入正確時，會彈跳出新的視窗顯示出搜尋結果，以 Text 加上 Scrollbar 函式顯示



## 五、實作結果 (資料日期：2020-01-02)

movie 2020-01-02

1 葉問4：完結篇, 2 白頭山：半島浩劫, 3 STAR WARS：天行者的崛起, 4 霹靂嬌娃, 5 冰雪奇緣2, 6 野蠻遊戲：全面晉級, 7 CATS 貓, 8 倒數計死, 9 兔嘲男孩, 10 賽道狂人

霹靂嬌娃, 動作 冒險, 片 長：01時58分, 上映日期：2019-12-27, IMDb分數：4.1

美麗華 09:50 14:30 16:50 19:10 21:30 23:50, 絕色 10:20 12:35 14:50 17:05 19:20 21:35 23:50, 信義威秀 11:20 13:35 15:50 18:05 20:20 22:35 00:50, 百老匯 10:00 12:10 14:20 16:30 18:40 20:50 23:00, 秀泰 09:50 12:10 14:30 16:50 19:10 21:30, 京站威秀 10:45 13:00 15:15 17:30 19:45 22:00 00:25, 梅花 10:00 15:00 20:00,

美麗華 09:30 14:00 18:30 00:10, 信義威秀 10:10 15:20 19:55 22:15 00:30, 京站威秀 10:20 14:40 17:00 19:20 23:40, 梅花 12:30 21:30,

美麗華 09:30 14:00 18:30 00:10, 信義威秀 10:10 15:20 19:55 22:15 00:30, 京站威秀 10:20 14:40 17:00 19:20 23:40, 梅花 12:30 21:30,

美麗華 09:30 14:00 18:30 00:10, 信義威秀 10:10 15:20 19:55 22:15 00:30, 京站威秀 10:20 14:40 17:00 19:20 23:40, 梅花 12:30 21:30,

潛艦追緝, 動作 劇情 懸疑/驚悚, 片 長：01時56分, 上映日期：2019-12-27, IMDb分數：7.0

信義威秀 13:25 21:55 00:35,

信義威秀 10:40 01:10,

信義威秀 17:10 01:10,

信義威秀 17:10 01:10,

舞力決戰, 愛情 音樂/歌舞, 片 長：01時25分, 上映日期：2019-12-27, IMDb分數：5.5

信義威秀 14:30 18:05,

N

N

N

竊竊屍語, 恐怖 懸疑/驚悚, 片 長：01時50分, 上映日期：2019-12-27, IMDb分數：6.2

N

N

N

N

圖一：final\_crawler.py 執行結果

哈囉你要找電影嗎

## 電影查詢小幫手

台北週末票房榜TOP10

No. 1 葉問4：完結篇

No. 2 白頭山：半島浩劫

No. 3 STAR WARS：天行者的崛起

No. 4 霹靂嬌娃

No. 5 冰雪奇緣2

No. 6 野蠻遊戲：全面晉級

No. 7 CATS 貓

No. 8 倒數計死

No. 9 兔嘲男孩

No.10 賽道狂人

電影資訊查詢

電影名稱:

查詢

輸入電影類型查詢

電影類型:

查詢

電影時刻表查詢

電影名稱:

電影院:

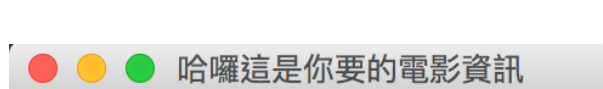
搜尋時間:

輸入地址:

輸入時間:

查詢

圖二：GUI.py 執行介面



**電影名稱：葉問4：完結篇**  
**電影種類：動作 劇情**  
**電影長度：01時45分**  
**上映日期：2019-12-19**  
**IMDb分數：7.6**

圖三：「葉問4：完結篇」的電影資訊查詢 output



**動作電影 共有6項搜尋結果：**

霹靂嬌娃  
 潛艦追緝  
 白頭山：半島浩劫  
 葉問4：完結篇  
 STAR WARS：天行者的崛起  
 賽道狂人

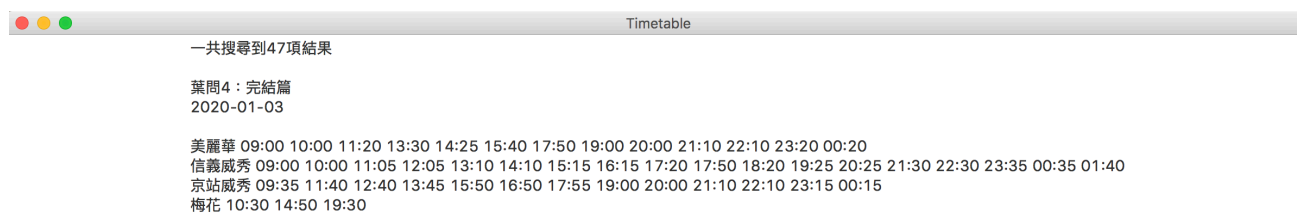
圖四：「動作電影」的電影類型查詢 output

Timetable	
一共搜尋到222項結果	
葉問4：完結篇	
2020-01-02	
美麗華	11:40 13:45 15:50 16:40 17:55 20:00 21:30 22:10 00:15
西門國賓	10:30 12:30 15:10 16:50 19:30 21:10 22:40
絕色	10:40 12:40 13:40 14:40 16:40 18:40 19:40 20:40 21:40 22:40 23:40
信義威秀	11:05 12:10 13:10 14:15 15:15 16:20 17:20 17:50 18:25 19:25 20:30 21:30 22:35 23:35 00:40
百老匯	09:50 10:30 11:00 11:45 12:25 12:55 13:40 14:20 14:50 15:35 16:15 16:45 17:30 18:10 18:40 19:25 20:05 20:35 21:20 22:00 22:30 23:15 23:55
秀泰	10:20 12:30 14:40 16:50 19:00 21:10 23:20
京站威秀	10:20 11:25 13:30 15:35 16:35 17:40 19:50 20:55 21:55 23:00 00:00
2020-01-03	
美麗華	09:00 10:00 11:20 13:30 14:25 15:40 17:50 19:00 20:00 21:10 22:10 23:20 00:20
信義威秀	09:00 10:00 11:05 12:05 13:10 14:10 15:15 16:15 17:20 17:50 18:20 19:25 20:25 21:30 22:30 23:35 00:35 01:40
京站威秀	09:35 11:40 12:40 13:45 15:50 16:50 17:55 19:00 20:00 21:10 22:10 23:15 00:15
梅花	10:30 14:50 19:30
2020-01-04	
美麗華	09:00 10:00 11:20 13:30 14:25 15:40 17:50 19:00 20:00 21:10 22:10 23:20 00:20
信義威秀	09:00 10:00 11:05 12:05 13:10 14:10 15:15 16:15 17:20 17:50 18:20 19:25 20:25 21:30 22:30 23:35 00:35 01:40
京站威秀	09:35 11:40 12:40 13:45 15:50 16:50 17:55 19:00 20:00 21:10 22:10 23:15 00:15
梅花	10:30 14:50 19:30
2020-01-05	
美麗華	09:00 10:00 11:20 13:30 14:25 15:40 17:50 19:00 20:00 22:10 23:20 00:20
信義威秀	09:00 10:00 11:05 12:05 13:10 14:10 15:15 16:15 17:20 17:50 18:20 19:25 20:25 21:30 22:30 23:35 00:35
京站威秀	09:35 11:40 12:40 13:45 15:50 16:50 17:55 19:00 20:00 21:10 22:10 23:15 00:15
梅花	10:30 14:50 19:30

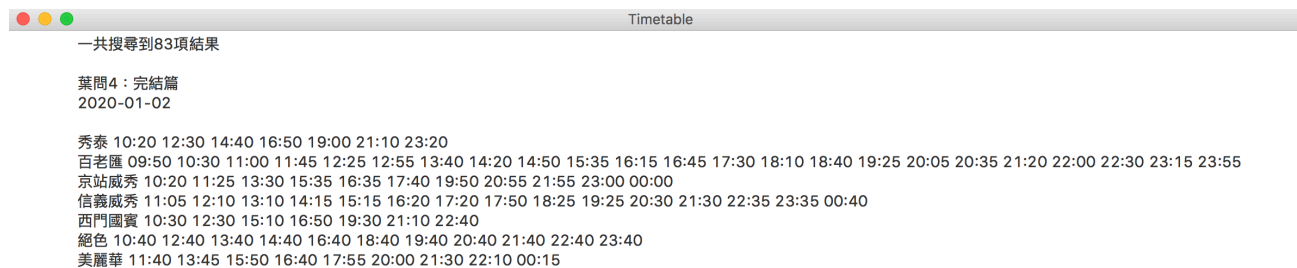
圖五：「葉問4：完結篇」的電影時刻表查詢 output

Timetable	
一共搜尋到47項結果	
葉問4：完結篇	
美麗華	
2020-01-02 11:40 13:45 15:50 16:40 17:55 20:00 21:30 22:10 00:15	
2020-01-03 09:00 10:00 11:20 13:30 14:25 15:40 17:50 19:00 20:00 21:10 22:10 23:20 00:20	
2020-01-04 09:00 10:00 11:20 13:30 14:25 15:40 17:50 19:00 20:00 21:10 22:10 23:20 00:20	
2020-01-05 09:00 10:00 11:20 13:30 14:25 15:40 17:50 19:00 20:00 22:10 23:20 00:20	

圖五：「葉問4：完結篇」、「美麗華」的電影時刻表查詢 output



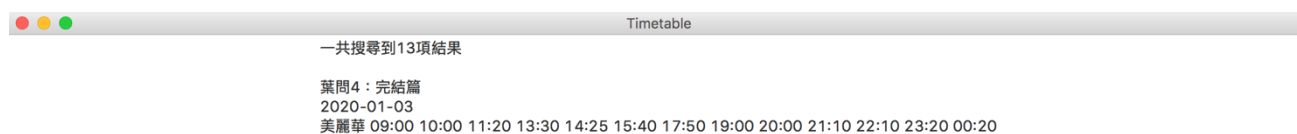
圖六：「葉問 4：完結篇」、「2020-01-03」的電影時刻表查詢 output



圖七：「葉問 4：完結篇」、「2020-01-03」、「台北市大安區羅斯福路 1 號」的電影時刻表查詢 output



圖八：「美麗華」、「2020-01-03」的電影時刻表查詢 output



圖九：「葉問 4：完結篇」、「美麗華」、「2020-01-03」的電影時刻表查詢 output



圖十：「葉問 4：完結篇」、「2020-01-03」、「14:00」的電影時刻表查詢 output



圖十一：「葉問 4：完結篇」、「美麗華」、「2020-01-03」、「14:00」的電影時刻表查詢 output

## 六、組內分工與心得

馬愷若 網路爬蟲、GUI 設計(組長)

這次的期末專案我爬蟲跟 GUI 都接觸了一些，這是我一次不藉由老師的上課、自己上網查詢要怎麼使用這個功能，雖然學到的內容還不夠深，但是也讓我滿有成就感的。這次期末專案我遇到的困難是 debug 跟優化我們的程式，這兩個都是需要先看懂其他人寫程式的邏輯跟語法才能做下一步動作，真的謝謝組員們在寫程式時有適當的註解跟取易懂的名字，讓我在閱讀或修改時方便很多；而且組員們都超級 carry、效率很高，即使是在期末準備時期，遇到問題也會馬上回覆跟解決，根本沒有擔心過我們期末專案會爆掉或做不完的問題。總之，這次期末專案真的很幸運有一群超級棒的組員，我不只學習到不同組員的程式邏輯、新函式庫的使用，我還學習到了團隊合作的重要，比起之前的作業或是考試都是一個人的挑戰，期末專案真的受益更多、更有趣、又不失挑戰！再次謝謝我的組員讓我在初學程式的路上有個開心的回憶！

鄭皓天 網路爬蟲

這次的期末專案我做的滿有成就感的，因為自己是個相較喜歡自學的人，所以面對這樣的專案比較不會有手足無措的緊張感。這次主要負責的部分是爬蟲，也剛好是我較有興趣的部分，因此研究起來很容易就忘記時間，只會想要繼續鑽研並且解決目前發生的問題。

碰到的最大困難則是因為剛好在作業期間內住院開刀，導致有近兩個禮拜的空窗期無法做任何事。幸好有完成階段性的任務才前往治療，放下了心中的重擔。更慶幸的是組員真的都很有效率，付出的努力與實作的經驗技巧也讓我打破對一般分組組員的印象。

總結來說，很感謝這堂課能讓我有機會學習到 python，我的視野也更開闊了一點，這次的專案也讓我對未來自己能夠透過 python 能做到的事更有輪廓和方向，不過未來可能就不會有相對這麼好的組員了，可能就要更習慣一個人做事的能力！

吳培瑜 網路爬蟲

這次的期末專案跟之前作業不一樣的地方在於很多相關的主題（如爬蟲、GUI）都是老師講了一個大概的架構跟方向，至於更仔細的部分就要自己上網查！另一個不一樣的地方是專案不像作業都是自己寫自己 debug，而是大家一起寫並一起找問題！但我常常看不太懂別人的邏輯就會看不懂程式碼，也就無法 debug，甚至常常把別人的程式碼越改越奇怪 QQ 還好組員們都超級 carry！每次都有人超級快就找出問題並解決，然後再發現新的問題，並無限循環。經過一次次的修正，看著程式越來越完整或是介面越來越漂亮真的是很有成就感（雖然都是辛苦的組員再改哈哈 XD）大家真的都超負責又超高效率！真的是很開心跟他們一組，也超級感謝他們！



## 蔡曉淇 GUI 設計

在這次的專案裡，我主要是負責將程式碼視覺化。過程中遇到最大的困難是，在將第三個搜尋的篩選條件套到介面中會不停地出現問題。然而，我卻常常 de 不出 bug，最後是靠著組員們解決的！在這次專案中我學到自學能力的重要性，在寫 GUI 的過程中，跟以前寫作業很不一樣的是，作業如果遇到不會只要扣緊老師教過的東西就一定寫得出來。但是在寫 GUI 時並沒有一定的標準答案，很多時候函式感覺雖然差不多，但在實際呈現上卻會有差異，甚至會影響使用者的使用，所以花了很多時間在找尋適合的函式、套件。此外，我覺得最大的收穫是感受到團隊合作的美好吧！因為我從以前就很不會 debug，所以常常花很多時間寫作業寫到很想停修，但這次得專案雖然比作業花了更多時間，可是卻不會有心想停修的感覺，真的很謝謝組員們讓我對 python 的最後印象是好的！

## 杜 昕 爬蟲資料整理、資料篩選、書面報告及簡報製作

我負責的部分主要是負責爬蟲資料整理和資料篩選，原本以為這是一份比較簡單、跟平常作業差不多難度的工作，但實際做下來跟想像中差非常多，光是資料整理、資料篩選就做了整整一個假日，因為電影場次表本身的資料量就遠大於之前的所有作業，輸出的部分更是讓人頭痛的地方，第三個篩選條件真的花了很多時間，因為我們想要呈現的是不用輸入所有條件也能搜尋，雖然能對於使用者比較友善，但卻讓篩選跟輸出變得更加困難，所以最後又花了很多時間在 debug 還有優化 output 呈現上，也常常因為寫程式所以兩三點才睡，不過最後的成果也讓人覺得相當值得。

此外我覺得還有一個跟平常作業不一樣的地方在於，期末報告不像作業一樣，對就是對、錯就是錯，只要不知道對不對丟上 PDOGS 就有解答，所以常常以為自己的輸出沒有問題，結果卻在組員們嘗試之下發現一堆需要解決的 bug，最後只能硬著頭皮熬夜解決，也是其他管院的課不太會有的經驗。也讓我們知道，即使這門課的時間有限，所以只能教我們最基礎的邏輯，但可以透過小組期末專案，探索到更多 python 的模組、函數等等，在修這門課之前，我從沒想過我們可以在短短一學期內做出這樣的報告，到現在我們靠著自己的力量做出了一個實用的工具，真的很有成就感！

## 吳芃蓓 資料篩選

在這次的期末專案裡，我負責用 googlemap 模組將搜尋結果依照距離遠近排列。我覺得最困難的部分應該是去了解到底要怎麼使用此模組，還有怎麼應用模組給出的結果轉換到我們自己的程式裡。因為 googlemap 的模組跑出的 output 會是非常大一串，看了 google 官網很久的說明才了解我們需要的資訊到底在 output 的哪裡，而且 output 的形式又是 list 跟 dictionary 層層疊疊。尤其到後來來要應用 googlemap 的 output 結合進另外一位同學的程式碼又是一大難關，一直出現 bug，但是很謝謝跟我合作的組員，很仔細跟我說明他的程式碼，讓我可以很快地就想出如何串接兩部分程式碼跟找到 bug 在哪裡！這次的期末專案很謝謝我的組員們，他們每個人都對自己的部分好負責任做事效率又超快！而且大家這次專案根本都是接觸了自己從來沒學過的東西，覺得大家能一起產出一個完整的程式真的好厲害～