

## 1. Name: 交大參考書搜尋系統

## 2. Motivation and description:

每學期一開始，就是要選課，而此時遇到的第一大問題是要選什麼課程和選那一位老師，而現在已經有選課系統和 NCTU+ 等平台可以使用，接下來會遇到的另一個麻煩問題是教授所指定的參考書要如何取得？因此我預計寫一個能透過使用者給定當期課號後，便能自動搜索並整理出指定參考書相關資訊、各大銷售通路的工具，以此省去我查閱、尋找電子書或二手書的時間。

## 3. Plan

最理想狀態是搭配模擬選課系統進行使用，將你選的課指定的參考書資料整理給你，然而選課系統為現有存在，我就只處理取得當期課號後的動作。

### 3.1. 取得教授指定用書資料

每位教授會在課程資訊的參考用書部分附上這門課可能會用到的參考書，我預計透過網路爬蟲，依據課號、參考資料等標籤，找到該欄位並取得 ISBN 碼或書名等資料。

### 3.2. 透過 Google Books API 取得書籍完整資料

教授提供書籍資料不一定完整，因此先透過 Google Books API 先取得該書籍完整資料，如 ISBN 碼、書名、版本、中文翻譯版等資訊。

### 3.3. 尋找書籍可能取得方式

#### 3.3.1. 麗文書局

#### 3.3.2. 交通大學浩然圖書館

#### 3.3.3. 網路商城，如：博客來、蝦皮……等。

#### 3.3.4. FB 社團，如：清交二手大貨倉、交大二手書……等。

透過各平台提供 API (如無則實際去爬)，取得書籍資料。以 FB 為例，透過 FB 提供之 API 搜索各社團文章，並以書名為關鍵字尋找近期文章，再將文章資料通知使用者。

### 3.4. How a user would interact with this?

目前試想過以 Telegram bot、Web、App……等形式呈現，目前偏好使用 Telegram bot 進行呈現，試想過可以儲存想找書籍資料，如二手版有你想的書籍資訊，可以第一時間透過 Telegram 通知使用者。

使用者預計會有一個輸入當期課號的簡單介面，之後程式運行完後理應跳出幾個搜索到的可能為正確書籍的資料，由使用者選擇後，再請使用者選擇想搜索的平台，再將搜索到的資訊以列表形式呈現給使用者，並提供使用者連結以前往。

## 4. Timeline of the project

### 4.1. 爬書籍資料 12/20

### 4.2. 取得完整書籍資訊 12/20

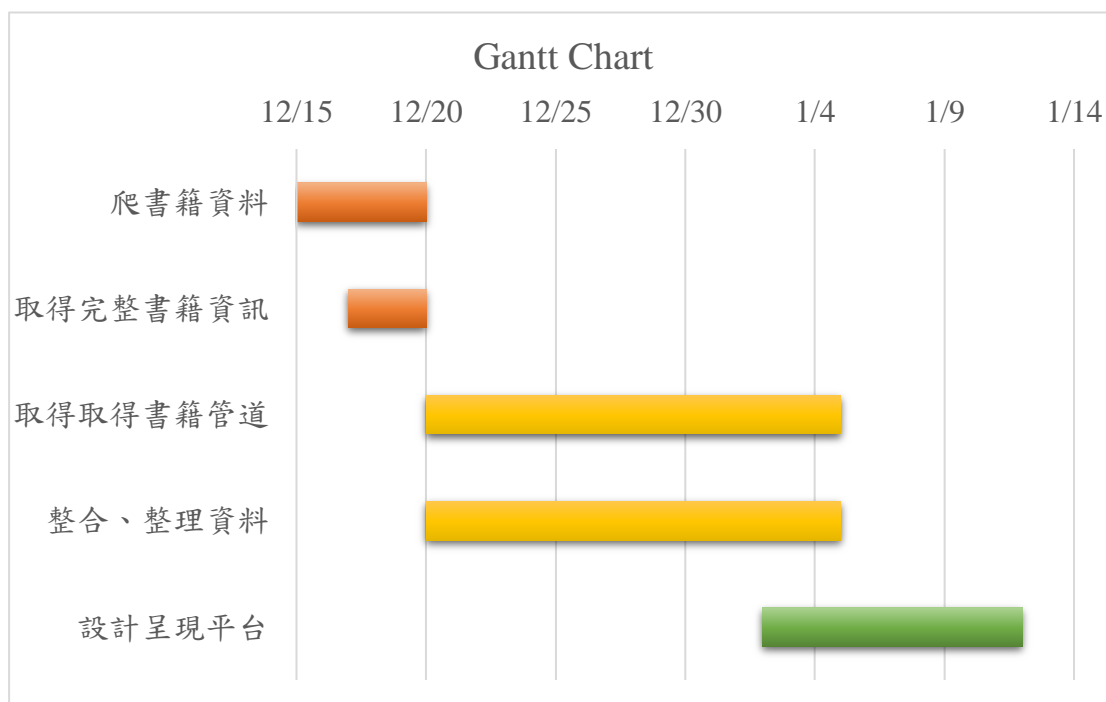
此兩項作業預想可以相對容易完成。

### 4.3. 取得取得書籍管道、並整合整理 01/05

此項預想為最麻煩部分。

### 4.4. 設計呈現平台 01/12

最終使用者與資料的橋樑設計，預想可能有些麻煩，但該段時間比較有空。



## Update 1

### 1. Use requests to get html (false)

會爬出如下圖的 html，於此會發現沒有任何資料。

```
43 <body>
44 <div class="crsoutline-params">
45 <input type="hidden" name="acy" value="109">
46 <input type="hidden" name="sem" value="2">
47 <input type="hidden" name="cos_id" value="1152">
48 <input type="hidden" name="lang" value="zh_TW">
49 <input type="hidden" name="user" value="">
50 <input type="hidden" name="_token" value="3d33e28b371f91d3f82501100937a500819e7db5060197c386f27f2947dd056e1959b69778ce53">
51 </div>
52
53 <div id="crsoutline-edit-container" class="crsoutline-show-container">
54
55 </div>
56 </body>
```

### 2. Try to lean with 交大預排課表 (too hard)

首先於網路查詢得知，此類狀況為 JavaScript 於網頁執行後才會將對應內容填入，因此先去參考下列參考資料後發現，其解決方法為自行寫 API 的概念，給予網站要的參數去跟網站要想要的資料，於大致理解完後，由於參考書是下列參考資料沒有去要的，自己去嘗試理解該頁面的 JavaScript code，但不熟悉該語言，最終放棄。

參考資料：

1. <https://github.com/CSY54/nctu-timetable-plus>
2. <https://github.com/lys0829/NCTUPreTimetable>

### 3. Use selenium to get html

於各方查找資料後，找到 Python 透過瀏覽器真實去爬網頁的方法，這樣即可等頁面加載完後再取 html，實際執行狀況如下，左側為執行中的程式，右側為 Firefox 自動搜尋網址的狀況。

The screenshot shows a Windows PowerShell window on the left and a Firefox browser window on the right. The PowerShell window displays a Python script that uses Selenium to interact with the NCTU Pre-Timetable website. The script sets up a WebDriver, navigates to the website, and extracts course information from the HTML. The Firefox window shows the website's content, including a table of courses and a sidebar with navigation links.

**Course Table:**

課程名稱	學分	上課時間	地點	備註
Ch15 Multiple Integrals (•虛為選教)	3	11 hours		
•Ch16 Vector Calculus (•虛為選教)	3	0 hour		

**教學要點描述:**

1. 學期作業、考試、評量：  
期末成績分會考佔30%
2. 教學方法及教學相關配合事項(如助教、網站或圖書及資料庫等)：

**師生晤談:**

排定時間	地點	連絡方式

**每週進度表:**

週次	上課日期	課程進度、內容、主題

備註：  
一、請遵守智慧財產權觀念及勿使用非法影印教科書。  
二、請參閱本系網頁及公告欄，瞭解各項活動、講座、研討會等資訊。

[illegible]

於是加上了中文辨識和輸出的部分，並嘗試調快 delay，第一次真的跑太久了。第二次結果如下，速度變快了，但少了 1000 筆左右的資料...，且也無法保證前一次真的有全部抓下來。

	Second try_2hours.json	
575	"9908": "1. 田勝華 (2015) 。生涯諮商與輔導。台北：學富文化。(主要教材) 2. 金樹人(2011)。生涯諮商與輔導(重修版)。台北：東華書局。3. 吳文達 (2000)。生涯諮商與諮詢：理論與實務。嘉義：吳石大。4. 吳文達 (2000)。生涯探索與創見：我的生涯手冊。臺南市：豐榮。5. 林榮志、田秀蘭、張郁廉(2010)。生涯輔導。台北：心理。6. 林清文(2000)。大學學生生涯發展與規劃手冊。台北：心理。7. 盧怡玲、劉素雲、敬世興、尤雅敏、邱淑媛、張淑麗 (2015) 。「適性發展-活出個人生命價值」方案介入臺南市中區生涯輔導之行動研究。輔導與諮商學報。37(1)-57-77。 8. 盧怡玲、王碧聰、劉靜芬、張怡婷、吳瑞鴻、陳瑞珍(2013)。華人生涯網際無異量表：縱橫在醫學生涯資訊與輔導系統中的線上施測與解釋學效。台灣心理諮商季刊。5(1)-16-9。劉素雲、盧怡玲、洪瑞鳳、楊介成(2013)。在世間活出人樣特異性：產生現象學取向的生涯觀與其在地化實踐設計理念的落實。輔導季刊。49(4)-2-16。相關資源教育部學生適性輔導專案資訊網： <a href="https://career.moe.gov.tw/">https://career.moe.gov.tw/</a> 行政院青年輔導委員會生涯資訊網： <a href="http://web1.nyse.gov.tw/mylearning/ncfife.aspx#生涯網">http://web1.nyse.gov.tw/mylearning/ncfife.aspx#生涯網</a> ； <a href="http://caringheart.net/">http://caringheart.net/</a> 生涯發展教育網： <a href="http://cde.nsysu.edu.tw/">http://cde.nsysu.edu.tw/</a> 國中生涯網際學習平台： <a href="http://www.heart.net.tw/careerweb/1.htm">http://www.heart.net.tw/careerweb/1.htm</a> 生涯技藝科育資訓網： <a href="http://140.122.103.235/career_5.php">http://140.122.103.235/career_5.php</a> 。	
576	"9909": "教科書：楊文顯、黃靜月(譯)(2016)。導考：幫助高中生輕鬆對考(原著作：P. Thorbiert)。臺北市：親子天下。謝國惠 (2019)。圖解認知心理學(更新版)。臺北市：易博士出版社。教材料：李冠慧、宋淑欣、何繼德、呂志明、洪一隆、阮啟弘(2011)。以認知神經科學取向探討兒童注意力的發展和學習障礙。輔導與諮商學報。42(3)-517-541。黃庭毅 (2018)。專一原則：「一心」用，戰勝拖延症，提高效率，改善習慣。臺灣商務印書館出版(原作者：D. Zick)。臺北市：寶島出版公司。 參考資料：「生涯規劃」教學成效調查及改善策略。臺灣商務印書館出版(原作者：D. Zick)。臺北市：寶島出版公司。 「生涯規劃」：生涯規劃(二版)。臺北：五南(2018)。李耀輝(2018)。輔導與諮商學。教師及家長實用參考用(四版)。臺北：五南。 統编版普通中學教科書(四版)。臺北：五南(2018)。李耀輝(2018)。輔導與諮商學。教師及家長實用參考用(四版)。臺北：五南。 9911": "生涯規劃。鄭志平、陳虹君、周文英、胡瑛、謝信君、李聖鈴、江潔芳、顧文珍、田仲開、黃美蓮、黃漢洋(2020)。特殊教育概論：現況與趨勢(二版)。新北市：心理。", 9912": "自編教料。" 9913": "1.許龍美,許龍玉(2003)。老師如何跟學生說話。心理出版社。2.林壽福,吳如昭(2009)。數學奧術：27個數學概念奇蹟。尖端出版社。3.國高中數學教科書4.十二年國民基本教育數學領域課程綱要	
581	"9914": "Richards, J. C., & Rodgers, T. S. (2014). Approaches and methods in language teaching (3rd Edition). UK: Cambridge university press.",	
582	"9915": ""	
583	"9916": ""	
584	"9917": "1.\t杜鰲照等(1995)。科學學習心理學(第三、四章)。心理出版社2.tshe, H.C.(2002)。Concepts of higher hierarchical level required more dual situational learning events for conceptual change: A study of students' conceptual changes on air pressure and buoyancy. International Journal of Science Education, 24(9), 981-996."	
585		

但仍無法保證是否為全部的資料...

Third\_try\_8hours.json

http://web1.nyc.gov.tw/nyclife/NYCLifehome.asp華人生涯網: <http://careereng.heart.net.tw/>生涯發展教育網: <http://cde.nsysu.edu.tw/>國中生涯規劃教學區: <http://www.heart.net.tw/careernew/11.htm>生涯及技藝教育資訊網: [http://140.122.103.235/career\\_5.php](http://140.122.103.235/career_5.php),"

1637 "9909": "教科書: 權文惠、李靜月 (譯) (2016) , 專生力: 幫助孩子更輕鬆有效的學習 (原作者: P. Thorbrietz) , 臺北市: 親子天下, 謝惠恩 (2019) , 圖解認知心理學(更新版) , 臺北市: 易博文化出版社, 林宜楓、李冠華、宋玖珍、柯淑嫻、曾宏明、洪 馨、阮欣弘(2011) , 以認知神經科學向探討兒童注意力的發展和學習之關聯, 教育心理學報, 42(3) , 517-541.黃美娥 (譯) (2018) , 專生力原則: "一心一用", 克服拖延提高效率改善關係, 臺灣機構與生活時間主權 (原作者: D. Zack) , 臺北市: 寶鼎出版.",

1638 "9910": "孟瑛如 (2010) , 不要比較, 只要教養/親職教育貼心手冊, 新北市: 心理, 孟瑛如(2013) , 看見特殊, 看見希望/特殊生教師家長貼心手冊(二版) , 新北市: 心理, 孟瑛如(2019) , 資源教室方案-班級經營與補救教學(四版) , 臺北: 五南, 孟瑛如(2019) , 學習障礙與補救教學-教師及家長實用手冊(四版) , 臺北: 五南.",

1639 "9911": "孟瑛如、陳志平、陳虹君、周文豪、胡瑛、謝淑君、李麗玲、黃國英、江源泉、簡齡文、田佐順、黃俊偉、陳國龍、黃澤洋(2020) , 特殊教育概論: 現況與趨勢(二版) , 新北市: 心理.",

1640 "9912": "自編教材",

1641 "9913": "1.許麗美,許麗玉(2003).老師如何跟學生說話.心理出版社. 2.林壽福,吳如皓(2009).數學魔術: 27個數學概念奇蹟. 尖端出版社. 3.國高中數學教科書4.十二年國民教育數學領域課程綱要",

1642 "9914": "Richards, J. C., & Rodgers, T. S. (2014). Approaches and methods in language teaching (3rd Edition). UK: Cambridge university press.",

1643 "9915": "教科書1.Vtlois Rhea Land and Mary Ann Vaughan. Music in today's classroom: creating, listening, performing. HB3, June 1978.2.Vatterbury, Betty W. & Richardson, Carol P. The Experience of Teaching General Music. McGraw-HILL, Inc. 1995.3.Vtlois Choksy, Robert M. Abramson, Avon E.Gillespie, David Woods, Frank York. Teaching Music in the 21th - century. Prentice-Hall, Inc. 2001.4.台灣各級之高中音樂課本. 參考書目: 1.vtIsa Findlay. Rhythm and Movement : Applications of Dalcroze Eurhythmics. Summy-Birchard Inc.19712.vtSuzuki Shinichi. Nurtured by Love. Miami : Summy-Birchard Inc.. 1983.3.vtBennett Reimer. A Philosophy of Music Education. Prentice-Hall, Inc. 1989.4.vtWierner, Brigitte. Orff-Schulwerk: Applications for the Classroom. Prentice Hall,1991.5.vtlois Choksy. The Kodaly Method I : Comprehensive Music Education. New Jersey : Prentice-Hall, Inc. 1999.參考資料教育部所頒布的最新年度高中音樂課程綱要",

1644 "9916": " """,

1645 "9917": "1.vt熊昭弟等 (1995). 科學學習心理學(第三、四章). 心理出版社2.vtShe, H.C. (2002). Concepts of higher hierarchical level required more dual situational learning events for conceptual change: a study of students' conceptual changes on air pressure and buoyancy. International Journal of Science Education, 24(9), 981-996."

1646 }

## 5. Get IDs of courses first

後來想到可以先爬所有課號，確認課程總量，且此資訊透過 requests 即可取得，不會有 delay 太短的問題。

```
main.py
1 import requests
2 import json
3 from bs4 import BeautifulSoup
4 data = list()
5 headers = {'user-agent': 'Mozilla/5.0 (Macintosh Intel Mac OS X 10_13_4) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/66.0.3359.181 Safari/537.36'}
6 for num in range(9000, 10000):
7     cos_num = "\n".join(['{:04}'.format(num) ])
8     url = 'https://timetable.nctu.edu.tw/?r=main/crsoutline&Acy=109&Sem=2&CrsNo='+cos_num+'&lang=zh-tw'
9     html = requests.get(url, headers=headers).text
10    soup = BeautifulSoup(html, 'html.parser')
11    tag1 = soup.find_all('input')
12    for x in tag1:
13        try:
14            if x['name'] == 'cos_id':
15                if x['value'] == '':
16                    break
17                else:
18                    data.append(cos_num)
19        except:
20            continue
21
22    with open("cos_id.json", "w") as f:
23        json.dump(data, f, indent = 4)
24
```

```
cos_id.json
1676  "9915",
1677  "9916",
1678  "9917"
1679 ]
```

## 6. Use IDs of courses to get books of courses

最後再拉大 delay 時間，由於總量縮小 9 倍，耗時也不會太久。完整 code 及最終 1679 筆資料如下。

```
main.py
1 #-*- coding: UTF-8 -*-
2 from selenium import webdriver
3 from bs4 import BeautifulSoup
4 import time
5 import json
6 path = 'D:/geckodriver.exe'
7 driver = webdriver.Firefox(executable_path=path,)
8 data = dict()
9 cos_id = json.loads(''.join(x for x in open('D:/cos_id.json') ))
10 for cos_num in cos_id:
11     url = 'https://timetable.nctu.edu.tw/?r=main/crsoutline&Acy=109&Sem=2&CrsNo='+cos_num+'&lang=zh-tw'
12     driver.get(url)
13     time.sleep(1)
14     driver.execute_script("window.scrollTo(0,document.body.scrollHeight)")
15     time.sleep(3)
16     html = driver.page_source
17     soup = BeautifulSoup(html, 'html.parser')
18     tag1 = soup.find_all('span')
19     for x in tag1:
20         try:
21             if x['name'] == 'col_textbook':
22                 data[cos_num] = x.get_text()
23                 break
24         except:
25             continue
26
27     driver.close()
28     with open("D:/data.json", "w", encoding='UTF-8') as f:
29         json.dump(data, f, indent = 4, ensure_ascii = False)
30
```

```
data.json
1675  "9914": "Richards, J. C., & Rodgers, T. S. (2014). Approaches and methods in language teaching (3rd Edition). UK: Cambridge university press.",
1676  "9915": "教科書1.\tLois Rhea Land and Mary Ann Vaughan . Music in today's classroom: creating, listening, performing. HBJ, June 1978.2.\tAtterbury, Betty W. & Richardson, Carol P. The Experience of Teaching General Music. McGRAW-HILL, Inc. 1995.3.\tLois Choksy, Robert M. Abramson, Avon E.Gillespie, David Woods, Frank York. Teaching Music in the 21th - century. Prentice-Hall, Inc. 2001.4.\t各版本之高中音樂課本・參考書目: 1.\tIsla Findlay. Rhythm and Movement : Applications of Dalcroze Eurhythmics. Summy-Birchard Inc.19712.\tSuzuki Shinichi. Nurtured by Love. Miami : Summy-Birchard Inc.. 1983.3.\tBennett Reimer. A Philosophy of Music Education. Prentice-Hall, Inc. 1989.4.\tWarner, Brigitte. Orff-Schulwerk: Applications for the Classroom. Prentice Hall,1991.5.\tLois Choksy. The Kodaly Method I : Comprehensive Music Education. New Jersey : Prentice-Hall, Inc. 1999.參考資料教育部所頒布的最近年度高中音樂課程綱要",
1677  "9916": "",
1678  "9917": "1.\t熊昭弟等 (1995). 科學學習心理學(第三、四章), 心理出版社2.\tTshe, H.C. (2002). Concepts of higher hierarchical level required more dual situational learning events for conceptual change: A study of students' conceptual changes on air pressure and buoyancy. International Journal of Science Education, 24(9), 981-996."
1679 ]
```



## 7. Check

再與先前所有課程對比即可確認是否仍有遺漏。

```
main.py
1 import json
2 cos_id = json.loads('".join(x for x in open('cos_id.json'))')
3 data = json.loads('".join(x for x in open('data.json'))')
4 for x in cos_id:
5     if x not in data:
6         print(x)
```

## 8. Timeline of the project (Update 1)

### 8.1. 爬書籍資料 12/24 (complete)

此項作業意外耗費許多時間處理，不可預期之狀況，嚴重 delay。

### 8.2. 取得完整書籍資訊 01/04

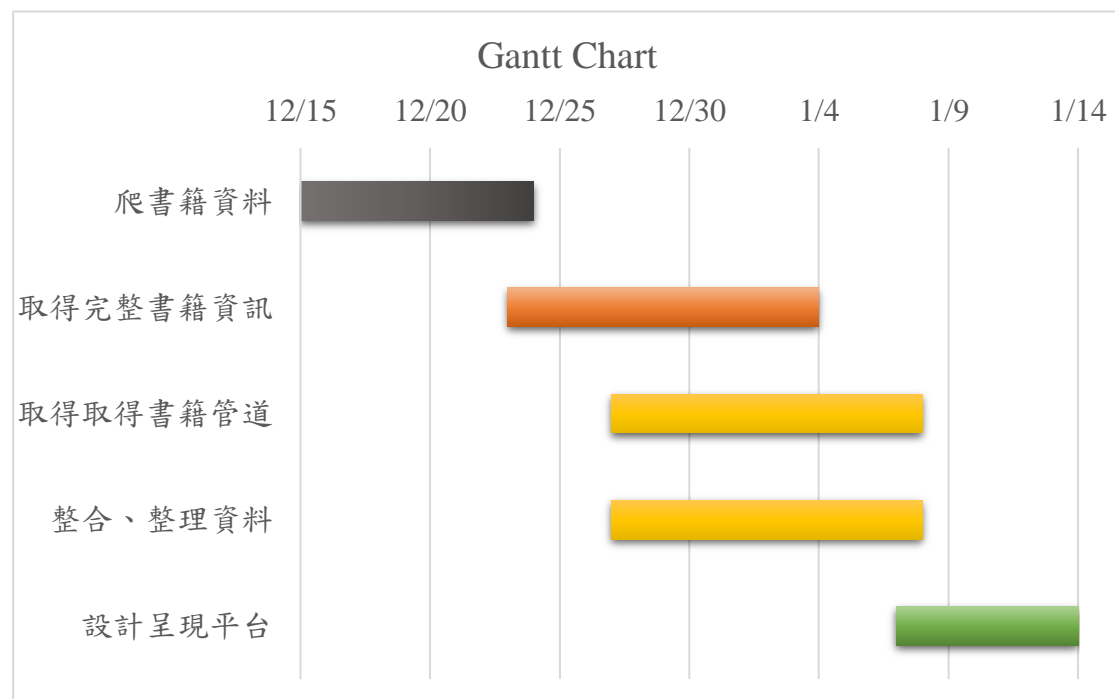
(期末比預想更加忙碌，進度嚴重 delay)

### 8.3. 取得取得書籍管道、並整合整理 01/08

此項預想為最麻煩部分。

### 8.4. 設計呈現平台 01/14

最終使用者與資料的橋樑設計，預想可能有些麻煩，但該段時間比較有空。



## Update 2

### 1. 分析資料

由下列幾分特定資料，我們可以發現有些課程沒有參考書，且表示形式各有異同，於是先透過一些處理整理資料捨去無用的，或暫時用不到的，由於此處過程過於繁雜，我也思考過太多問題，其中也有關於 Google books API 的精度限制問題造成的，其過程約略使用下方 code 及部分工人智慧進行整理。

```
data.json
{
  "0311": "Calculus(Early Transcendentals), James Stewart, 8th Edition",
  "0511": "\"Physics for Scientists and Engineers with Modern Physics\", 10th edition, Serway & Jewett",
  "1024": "無",
  "1025": "",
  "1026": "Lecture edited by Tsai, Chungyong",
  "1032": "\"Biology, (2016) 12th Edition. S. Mader and M. Windelspecht Eds., McGraw-Hell, New York, NY, USA. 藝軒圖書 02-2918-2288",
  "1033": "K. H. Rosen, \"Discrete Mathematics and Its Applications,\" 8th edition, 2019, McGraw-Hill, ISBN: 978-1-260-09199-1"
}

main.py
# -*- coding: UTF-8 -*-
import json
data = 'data.json'
new_data = dict()
with open(data, 'r') as fp:
    book = json.load(fp)
for x, y in book.items():
    if type(y) == list:
        new_data[x] = y
    else:
        if y.isdigit():
            new_data[x] = [y]
        ...
for x, y in book.items():
    if type(y) == list:
        continue
    if y.find('ISBN') != -1 or y.find('isbn') != -1:
        pattern = re.findall(r'ISBN: ([0-9]*)?(?:.+)', y)
        new_data[x] = pattern
for x, y in book.items():
    if y == '' or y[0] == '無':
        continue
    else:
        new_data[x] = y
...
with open("isbn.json", "w", encoding='UTF-8') as f:
    json.dump(new_data, f, indent = 4, ensure_ascii = False)
```

結論而言，只保留有附 ISBN 編碼的參考資料進行這次的操作。

詳細思考過程：

1.1. 透過 Google books API 來取得完整正確之書籍資料。(先略過 API 實際操作方法)

為此須要有好的關鍵字，或者說需要精準度過高的搜尋方式。

1.2. 移除無用資料：如空、'無'、'note'.....等雜七雜八資料。

1.3. 整段搜尋：0311 整段資料拿去 API 查找資料，結果：正確找到資料。

0511 整段資料拿去 API 查找資料，結果：錯誤。

1.4. 任意嘗試分割關鍵字搜尋：0511 仍無法掌握正確書籍資料。

1.5. ISBN 搜尋：精準，無疑問。

1.6. 部分資料據有多項書籍資料：將書籍存成 list 以區隔多本書籍。

1.7. 由於無論如何嘗試：都無法保障精準搜尋 ISBN 以外關鍵字來取得正確書籍，存在絕對的個別差異，且若將書籍進行區分，少去 ISBN 則難以分辨書籍的分塊。

1.8. 暫時結論：先確保教師有提供 ISBN 編碼的書籍。(ISBN 資料如下)

```

isbn.json
1  [
2    "1027": [
3      "1292222824",
4      "0136117260",
5      "0766838943"
6    ],
7    "1033": [
8      "9781260091991"
9    ],
10   "1077": [
11     "0071257098"
12   ],
13   "1085": [
14     "1292061227"
15   ],
16 ]

```

## 2. Google books API

經過一段時間研究後，其功能包含管理個人書籍等，而我須使用之功能如下：

[q](#) – 搜尋 (q=關鍵字+下列標的)

- **intitle**: Returns results where the text following this keyword is found in the title.
- **inauthor**: Returns results where the text following this keyword is found in the author.
- **inpublisher**: Returns results where the text following this keyword is found in the publisher.
- **subject**: Returns results where the text following this keyword is listed in the category list of the volume.
- **isbn**: Returns results where the text following this keyword is the ISBN number.
- **lccn**: Returns results where the text following this keyword is the Library of Congress Control Number.
- **oclc**: Returns results where the text following this keyword is the Online Computer Library Center number.

首先須取 API key，前往 [google APIs](#) 註冊並啟用 Google books API 即可取得。接著透過以上搜尋方法進行搜尋，由資料整理部分得知於此處嘗試過許多搜尋方式但結論而言暫時使用 isbn 實施精準搜尋。再將取得資料存入，以取得書籍詳細資訊，於此處發現 Google books 有提供部分取得通路。

```

main.py
1  #-*- coding: UTF-8 -*-
2  import requests
3  import json
4  file = 'isbn.json'
5  new_data = dict()
6  with open(file, 'r') as fp:
7      book = json.load(fp)
8      APIKey = '取得個人API key'
9      for x, y in book.items():
10         new_list = list()
11         for isbn in y:
12             url = 'https://www.googleapis.com/books/v1/volumes?maxResults=1&q='+isbn+'+isbn&key='+APIKey
13             html = json.loads(requests.get(url).text)
14             new_list.append([isbn,html])
15         new_data[x] = new_list
16 with open("books_data.json", "w", encoding='UTF-8') as f:
17     json.dump(new_data, f, indent = 4, ensure_ascii = False)
18

```



```
books_data.json
1  {
2    "1027": [
3      {
4        "1292222824",
5        {
6          "kind": "books#volumes",
7          "totalItems": 713,
8          "items": [
9            {
10             "kind": "books#volume",
11             "id": "T25pBwAAQBAJ",
12             "etag": "sy/82YB/Wdo",
13             "selflink": "https://www.googleapis.com/books/v1/volumes/T25pBwAAQBAJ",
14             "volumeInfo": {
15               "title": "Problem Solving with C++, Global Edition",
16               "authors": [
17                 "Walter Savitch"
18               ],
19               "publisher": "Pearson Higher Ed",
20               "publishedDate": "2015-02-27",
21               "description": "For the C++ introductory programming course Problem Solving with C++ continues to be the most widely
22               used textbook by students and instructors in the introduction to programming and C++ language course. Through each
23               edition, hundreds and thousands of students have valued Walt Savitch's approach to programming, which emphasizes active
24               reading through the use of well-placed examples and self-test examples. Created for the beginner, this book focuses on
25               cultivating strong problem-solving and programming techniques while introducing students to the C++ programming
26               language.",
27               "industryIdentifiers": [
28                 {
29                   "type": "ISBN_13",
30                   "identifier": "9781292071572"
31                 },
32                 {
33                   "type": "ISBN_10",
34                   "identifier": "0321943877"
35                 }
36               ]
37             }
38           ]
39         }
40       ]
41     }
42   }
```

### 3. Timeline of the project (Update 2)

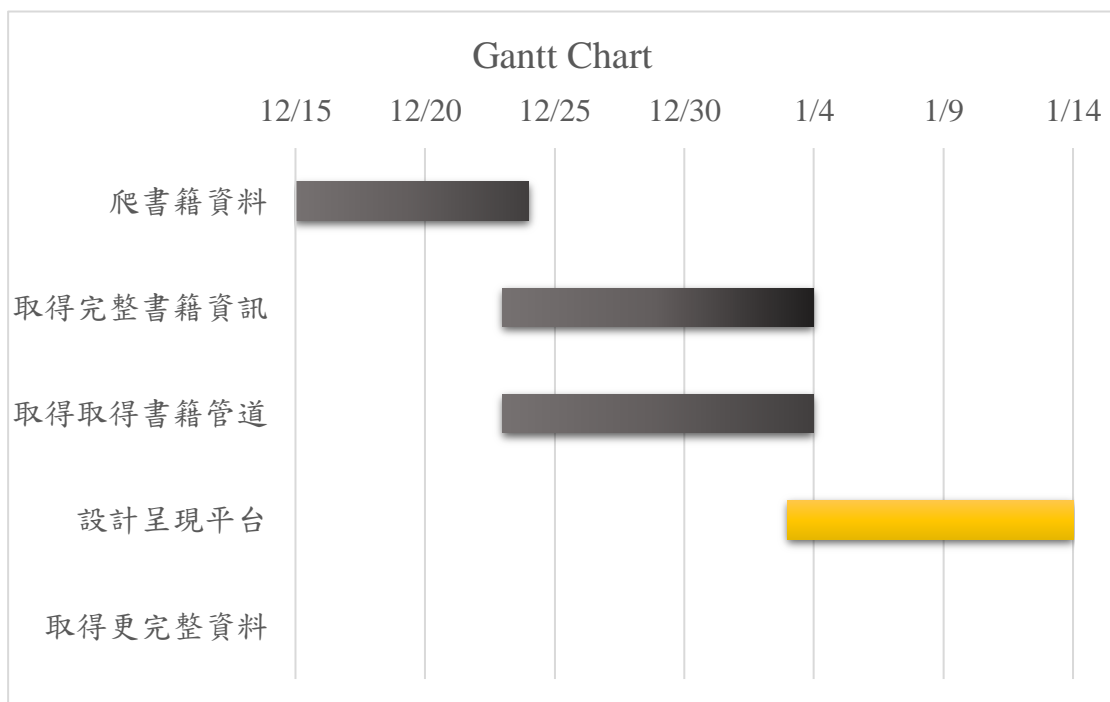
#### 3.1. 取得完整書籍資訊、取得取得書籍管道 01/08

資料整理略顯困難，由於課程資料過於混亂。

#### 3.2. 設計呈現平台 01/14

目前預計使用 Telegram API 來完成。

#### 3.3. 取得更完整資料 未來計畫 or 有多餘時間之計畫



## Update 3

### 1. 呈現方式抉擇

最後思考出較好的兩種方式。

#### 1.1. Github.io

由 github 提供的 page 服務架一個靜態網站，原定最優先是此方案，因此抓取資料時，多以存成 json 檔為目標，原定預先取好所有資料，再丟到 github 上並設計一個靜態網頁來取資料，但由於時間因素以及 JavaScript 等撰寫網頁語言之不熟悉，不選擇此方案。

#### 1.2. Telegram bot API

由 Telegram 的聊天軟體作為呈現介面，同時也是剛好可以作為一個 API 之因素選擇此方案，且 Telegram bot 可以很方便使用者進行使用，只需輸入課號即可操作。

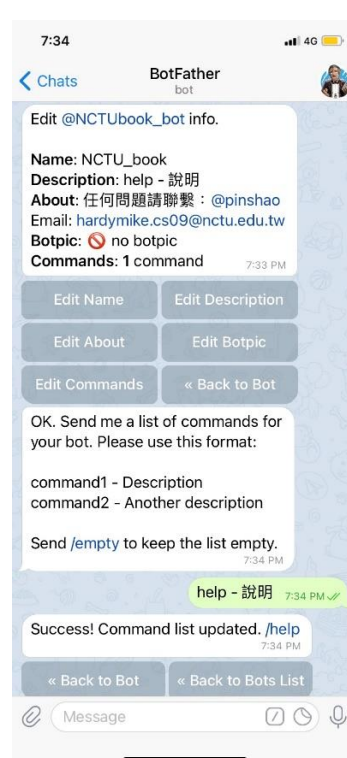
### 2. Telegram bot 基本設定



下載 Telegram，並加入此 bot 帳號。



創建新的 bot，並命名。  
取得 API token。



新增 bot command。

### 3. 接受回復訊息

經過長時間學習使用 Telegram 後，了解 telegram bot 透過 python 的撰寫方式。

```
main.py
1 import telepot
2 import json
3 from telepot.loop import MessageLoop
4 with open('books_data.json', 'r') as fp:
5     books = json.load(fp)
6 API = 'your telegram api token'
7 bot = telepot.Bot(API)
8 def help(ID):
9     bot.sendMessage(int(ID), '輸入書籍編號以取得參考書籍資訊\n須輸入四位數字\n部分編號暫無資料')
10 def infor(ID, TEXT):
11     book = books[TEXT]
12     for data in book:
13         try:
14             name = data[1]["items"][0]["volumeInfo"]["title"]
15             photo_url = data[1]["items"][0]["volumeInfo"]["imageLinks"]["thumbnail"]
16             link = data[1]["items"][0]["volumeInfo"]["infoLink"]
17         except:
18             continue
19         letter = '書名'+name+'\n'+ISBN: '+data[0]+'\n'+photo_url+'\n'+書籍資料: '+link
20         bot.sendMessage(int(ID), letter)
21 def error(ID):
22     bot.sendMessage(int(ID), '無法辨識訊息\n或課程書籍暫無資料\n有任何問題請聯繫 @pinshao\nEmail: hardymike.cs09@nctu.edu.tw')
23 def handle(msg):
24     ID = msg['from']['id']
25     TEXT = msg['text']
26     if TEXT == '/help':
27         help(ID)
28     elif TEXT.isdigit() and TEXT in books:
29         infor(ID, TEXT)
30     else:
31         error(ID)
32 MessageLoop(bot, handle).run_as_thread()
```

3.1. 首先開啟先前得 books\_data 取得所有資料。

3.2. 由 telepot 來處理 telegram 的操作，第 6, 7 行可標定自己的 bot。

3.3. 接著要接受所有傳給 bot 的即時訊息

由 from telepot.loop import MessageLoop 的函式可監聽即時訊息(第 32 行)。

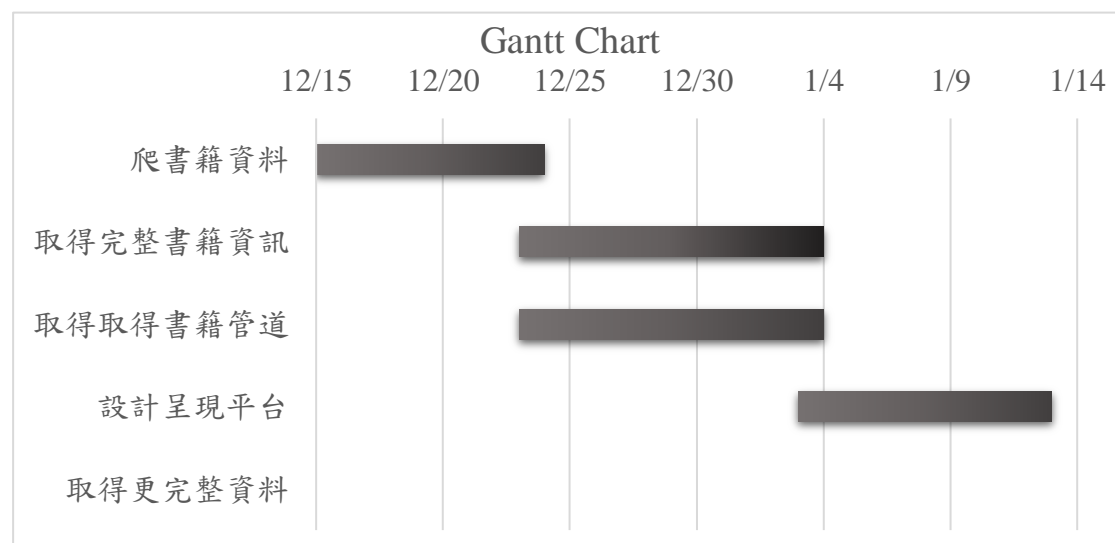
3.4. 交由 hadle function 來確認第一手訊息，再分別交由各函式處理回傳訊息。

### 4. 實際執行狀況

4.1. 可於 Telegram 加入 @NCTUbook\_bot 來使用此軟體，由於 python 須於執行狀態才有用，暫時仍在研究 Heroku 的使用方式。

4.2. 完整操作過程: <https://youtu.be/nlmQhXfH-AA>

### 5. Timeline of the project (Update 3)



## 完整操作流程

(各執行須 import 皆於程式碼中)

(各 API key 皆不提供，如有這樣有問題請聯繫 [hardymike.cs09@nctu.edu.tw](mailto:hardymike.cs09@nctu.edu.tw) )

1. 執行 Project\_take\_cos\_id 資料夾的 main.py 以取得所有課程存成 cos\_id.json。
2. 下載 Fire fox geckodriver.exe 檔案。
3. 執行 Project\_take\_books\_data 資料夾的 main.py 以從 cos\_id.json 來用 slenium 透過 Fire fox 取得所有課程參考書資料存成 data.json。
4. 執行 Project\_book\_clean 資料夾的 main.py 來整理 data.json 資料 (還須些許工人智慧)存成 isbn.json。
5. 執行 Project\_google\_books 資料夾的 main.py 來依照 isbn.json 透過 Google books API 取得所有書籍完整資料存成 books\_data.json。
6. 執行 Project\_telegram 資料夾的 main.py 使用 books\_data.json (須處於執行中)。
7. 開啟 Telegram 使用 NCTUbook\_bot。

## 未來計畫 (中途放棄之功能)

1. 整理好課程資料，以透過關鍵字即可取得正確書籍資料。
2. FB API 以取得即使貼文是否有相關書籍二手拍賣資料(由於大部分社團屬於秘密社團，故無法爬取資料；清交二手大貨倉 2.0 雖然屬於公開社團，卻不活躍)。
3. Github.io 架設網頁版介面。
4. 非官方誠品網路書店 API  
<https://github.com/FlowerEatsFish/eslite-com-api>
5. Selenium 爬取個大銷售通路資訊 (蝦皮、博客來等)。
6. 交大圖書館館藏查詢 (目前除了透過 selenium 以外沒有其他可行之想法)。
7. 雲端執行程式，由於大部分用於回雲端持續執行的大部分只支援 JS，而我對該語言不熟悉，而 python 雲端執行卻異常難處理，因此能在研究中。

## 心得

一開始預估應該部會太難處理，預計做一個有用的、滿意的 Project，不過第一個就用到問題了，對於網頁瞭解還不夠清楚，尤其是 Java Script 這個東西不太熟悉造成的，第一個動態更新的網頁一開始就搞得我頭很痛，沒有頭緒。

後來掌握 selenium 後，感覺它好強大，應該有很多東西都可以靠它處理，不過也有極為明顯的缺點，會花很長的一段時間處理搜尋，我的很多資料沒有 API 的未來應該也都是用 selenium 去抓下來，起馬是我目前能想到的唯一方法了，基於時間緊迫之因素，只能未來再嘗試。

後來去研究 FB API、Google books API、Telegram API，都發現學懂它怎麼用其實滿麻煩的，需要花一些時間去了解，不過搞懂後，可以很簡單的去使用。再來就是 Heroku 搞不太定吧，因為要持續執行，除了電腦一直讓它跑外，就是丟雲端，而 Heroku 是我唯一覺得能用的東西，目前搞不懂，不過也許搞懂之後，它其實也很容易。

整份 Project 最大的關鍵：我不會 JS，它影響了我好多決定、時間。之後應該花時間了解一下，它應該不會太難才對。經過這次 Project 對於整個網路等實際層面的東西多了一些認識，大部分東西一開始都覺得很難用，也許之後要找到從什麼地方切入學習會比較好的方法。不過後來大約理解後，都算容易使用，之後應該會再想辦法完整它，並嘗試設計其他自己想要的工具。