

# 爬蟲成果報告書

## 1. 爬蟲流程說明：

- I. 首先在三民書局網站搜尋 AI 人工智慧
- II. 確認共有 161 筆資料，共 9 頁
- III. 對 9 頁的 url 進行 get 請求並取得資料
- IV. 確認書籍資料在 html 中的位置
- V. 使用 BeautifulSoup 套件取得所需資料
- VI. 將資料印出來並儲存為 txt 和 csv 檔

## 2. 題目各點要求之程式解說

### I. 抓取人工智慧書籍頁面資料

```
SanminBooks.py > ...
1  import csv
2  import requests
3  from bs4 import BeautifulSoup
4
5  books = [] # 儲存所有書的資料
6
7  for i in range(1, 10):
8      url = 'https://www.sanmin.com.tw/search/index?ct=k&k=AI%e4%ba%ba%e5%b7%a5%e6%99%ba%e6%85%a7&ls=sd&fu=0&vs=list&pi=' + str(i)
9      # 來對 url 發送一個 GET 請求，回應的內容是一個字串
10     html = requests.get(url).text
11     # 創建一個 BeautifulSoup 物件，可以更好操作 html 的結構
12     soup = BeautifulSoup(html, 'html.parser')
```

### II. 抓取每本書的相關資訊(包含書名, 作者, 價錢)

```
14     products = soup.find_all('div', 'Info') # 抓取一本書籍資料 (書名、作者、價格)
15
16     for product in products: # 處理資料
17         name = product.find('h3', 'Title').a
18         # 抓取書籍資料 (書名)
19         author = product.find('div', 'Author').find('span', class_='text-green').a
20         # 抓取一本書籍資料 (作者)
21         price = product.find('span', 'Price')
22         # 抓取一本書籍資料 (價格)
23         if name is not None and author is not None and price is not None:
24             # 判斷是否有抓到資料
25             book = [] # 儲存單一本書的資料
26             name_element = "書名:" + name.text + " " # 抓取書名資料轉為text
27             author_element = "作者:" + author.text + " " # 抓取作者資料轉為text
28             price_element = "價格:" + price.text + "元" # 抓取價格資料轉為text
29             price_value = int(price.text) # 提取價格的數值部分並轉換為int
30             book.append(name_element) # 本書資料加入串列
31             book.append(author_element)
32             book.append(price_element)
33             books.append((book, price_value)) # 將價格數值與書籍資料一起存儲
```

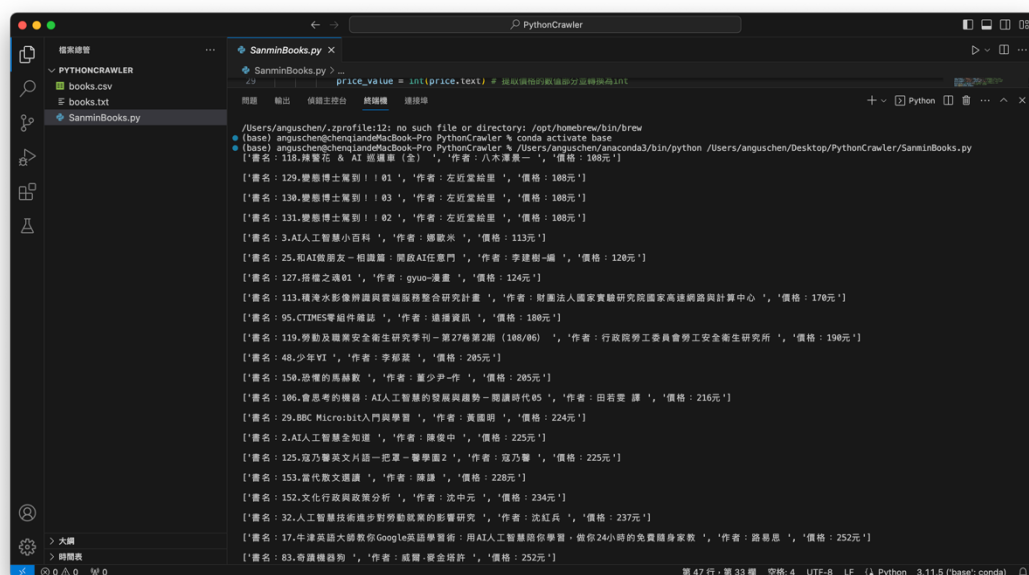
### III. 每筆數據排列請依據價格由小到大排列

```
35     # 按價格進行排序
36     sorted_books = sorted(books, key=lambda x: x[1]) # 利用串列的價格資料(x[1])排序
37     for item in sorted_books:
38         print(item[0], "\n") # 分行印出資料
```

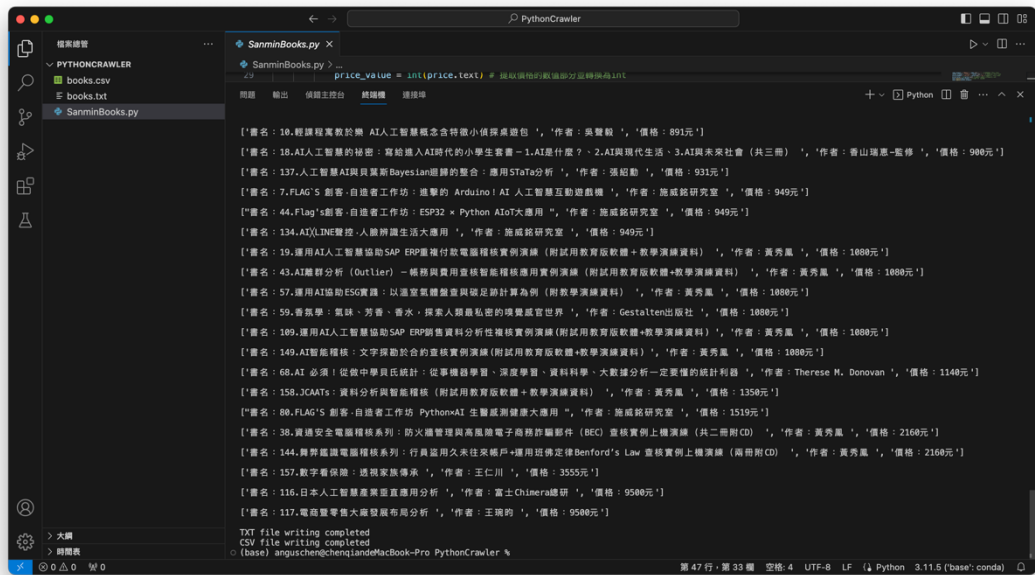
#### IV. 以兩種形式 (txt, csv) 儲存資料

```
40 # 檔案建立或讀取地址
41 directory = "/Users/anguschen/Desktop/PythonCrawler/"
42 text_file_path = directory + "books.txt"
43 csv_file_path = directory + "books.csv"
44
45 textfile = open(text_file_path, 'w') # 新建檔案若存在則清空並寫入資料
46 textfile.write("學號[B1043003]姓名[陳麒安]+\n")
47 for item in sorted_books: # 寫入資料
48     textfile.write(item[0][0] + "\n") # 依序寫入書名、作者、價格
49     textfile.write(item[0][1] + "\n")
50     textfile.write(item[0][2] + "\n")
51     textfile.write("\n")
52 textfile.close() # 關閉檔案
53 print("TXT file writing completed")
54
55 with open(csv_file_path, "w", encoding='utf-8') as csvfile:
56     # 打開名為 "books.csv" 的 CSV 檔案，並使用 UTF-8 編碼寫入
57     csv_writer = csv.writer(csvfile)
58     # 寫入CSV文件的標題行
59     csv_writer.writerow(["書名", "作者", "價格"])
60     # 寫入書籍資訊
61     for book in sorted_books:
62         csv_writer.writerow(book[0])
63 print("CSV file writing completed")
```

#### 3. 程式執行畫面剪貼描述



執行程式並印出排序後資料



程式執行完成，並印出 txt 和 csv 檔案寫處完成提示



Txt 檔案

常用 插入 繪圖 頁面配置 公式 資料 校閱 檢視 自動化 操作說明搜尋					
貼上	新組明體 (本文)	12	A <sup>+</sup> A <sup>-</sup>	自動換行	通用格式
B I U				條件式格式設定	格式化為表格
				儲存格樣式	插入刪除格式
				排序與篩選	尋找與連結
					聯結度
A1	書名				
	A	B	C	D	E
1	書名	作者	價格		
2	書名：118.辣警花 & AI 巡邏車 (全)	作者：八木澤譽一	價格：108元		
3	書名：129.變態博士駕到！！01	作者：左近堂繪里	價格：108元		
4	書名：130.變態博士駕到！！03	作者：左近堂繪里	價格：108元		
5	書名：131.變態博士駕到！！02	作者：左近堂繪里	價格：108元		
6	書名：3.AI人工智慧小百科	作者：鄭敏米	價格：113元		
7	書名：25.和AI做朋友－相識篇：開啟AI任意門	作者：李建樹-編	價格：120元		
8	書名：127.搭檔之戰01	作者：gyuo-漫畫	價格：124元		
9	書名：113.機考水影像辨識與雲端服務整合研究計畫	作者：財團法人國家實驗研究院國家高速網路與計算中心	價格：170元		
10	書名：95.CTIMES零組件雜誌	作者：遠播資訊	價格：180元		
11	書名：119.勞動及職業安全衛生研究季刊－第27卷第2期 (108/06)	作者：行政院勞工委員會勞工安全衛生研究所	價格：190元		
12	書名：48.少年vt	作者：李郁蓀	價格：205元		
13	書名：150.恐懼的馬赫數	作者：董少尹-作	價格：205元		
14	書名：106.會思考的機器：AI人工智慧的發展與趨勢－閱讀時代05	作者：田若雯 譯	價格：216元		
15	書名：29.BBC Micro:bit入門與學習	作者：黃國明	價格：224元		
16	書名：2.AI人工智慧全知道	作者：陳俊中	價格：225元		
17	書名：125.皮乃魯英文片語－把單－醫學圖2	作者：皮乃魯	價格：225元		
18	書名：153.當代散文選讀	作者：陳謙	價格：228元		
19	書名：152.文化行政與政策分析	作者：沈中元	價格：234元		
20	書名：139.人工智慧技術進步對勞動就業的影響研究	作者：沈紅兵	價格：237元		
21	書名：17.牛津英語大師教你Google英語學習術：用AI人工智慧陪你學習，做你24小時的免	作者：路易思	價格：252元		
22	書名：83.奇蹟機器狗	作者：威爾·麥金塔許	價格：252元		
23	書名：93.我的AI同學01：AI的發展史	作者：陳靜芬-作	價格：261元		
24	書名：94.我的AI同學02：AI的應用領域	作者：陳靜芬-文	價格：261元		
25	書名：36.拒當AI時代的局外人：面對機器人開始搶飯碗，你準備好了嗎？	作者：邱文仁	價格：270元		
26	書名：71.用mBlock玩mBot機器人互動程式設計 (最新加強版)	作者：王麗君	價格：270元		
27	書名：89.給下一輪太平盛世的備忘錄	作者：伊塔羅·卡爾維諾	價格：276元		

## CSV 檔案