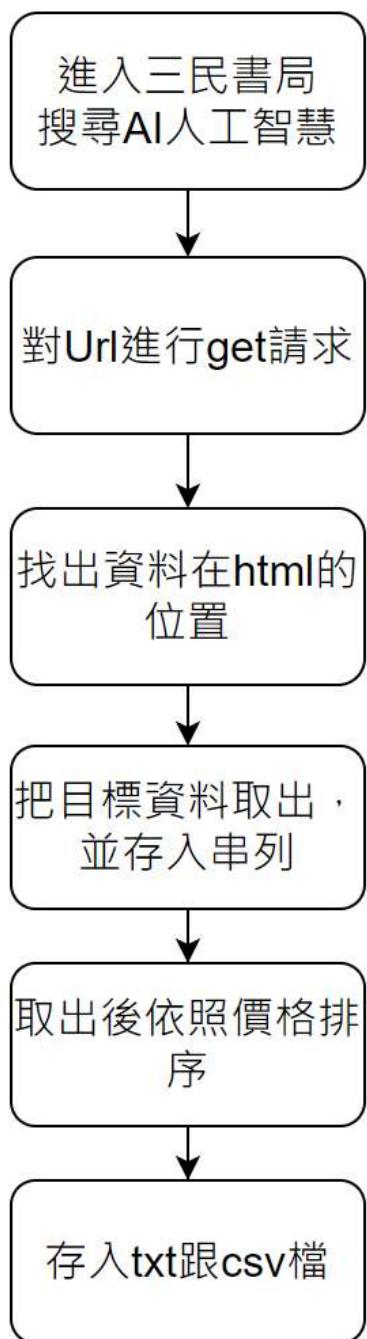


# 資工三 B1043001 許良亦

## 一、爬蟲流程：

1. 到三民書局網站搜尋 AI 人工智慧，並確定商品數跟頁數(161 個商品 9 頁)
2. 獲取 Url，並對 Url 進行 get 請求
3. 利用 Beautiful soup 方法把書名、作者、價格取出，取出後存入 list
4. 進行翻頁重複步驟(直到完成九頁)
5. 使用 list 來進行排序
6. 把排序後的 list 印出
7. 把爬蟲結果寫入 txt 檔跟 csv 檔



圖一、爬蟲流程圖

## 二、題目各點要求之程式解說

The screenshot shows the homepage of Sanmin Bookstore (www.sanmin.com.tw) with a search bar at the top containing 'AI 人工智能'. The search results page displays two book entries:

- 1 AI 人工智能**  
作者: 並美原 出版社: 普林斯頓國際 出版日: 2022/05/01 裝訂: 平裝  
台語於教育部在 108 學年度，開始實行十二年國民基本教育課程綱要，培養學生具備 21 世紀所需的科技素養，推動普及高級中等以下學校新與利技之認知計畫。為讓學生有機會體驗 **AI 人工智能**，且知道 **AI** 的應用與對自己未來及生活的影響。本書特別規劃適用於國內高等教育課程，來銜接 **AI** 教育之基本教材。全書內容規劃，特別將坊間 **AI** 教材裡難澀的數學、專業的電腦程式編碼、臺灣的
- 2 AI 人工智能全知道**  
作者: 陳俊中;曹賴倫 出版社: 世一 出版日: 2021/12/13 裝訂: 平裝  
你知道**AI 人工智能**是什麼嗎？你有聽過機器人可以幫忙打掃家裡嗎？以**AI 人工智能**、大數據、物聯網與機器人等技術為代表的第四次工業革命，正悄悄地大規模進行著！本書透過淺顯易懂的說明與圖片，希望讓孩子們建立相關的基本概念，並且與未來社會在生活中發現的應用相連結，進一步理解與適應新科技帶來的變化，得到許多收穫！

圖二、此圖為進入三民書局並搜尋 AI 人工智能的畫面，左上角為此網站之 URL

The screenshot shows the Sanmin Bookstore website's homepage. At the top, there is a navigation bar with links for '企業採購', '加入會員', '會員登入', '會員專區', '購物車', '紅利兌換', '客服中心', and '門市專區'. Below the navigation is a banner for the '三民網路書店' (Sanmin Online Bookstore) featuring a '萬聖魔幻舞台' (Halloween Magic Stage) event from 10.1 to 11.15. A search bar contains the text 'AI人工智慧'. The main content area displays a grid of book covers under various categories like '外文童書' (Foreign Children's Books), '中文' (Chinese), '外文' (Foreign), etc.

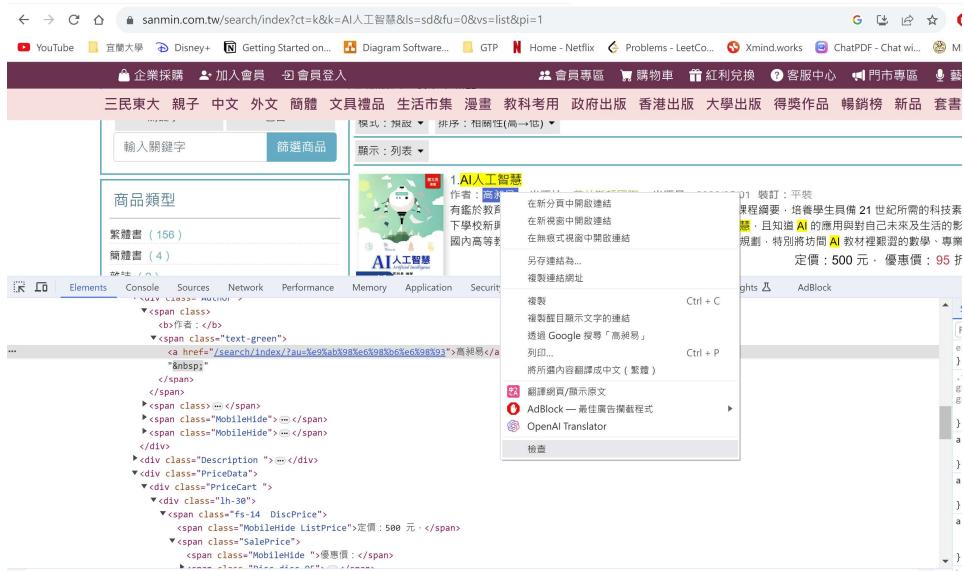
圖三、用 F12 確認網站架構，屬於 html 結構

This screenshot shows the Sanmin Bookstore website with the browser's developer tools (F12) open, specifically the Elements tab. The focus is on a product detail page for a book titled 'AI 人工智慧全知識'. The page includes a sidebar with categories like '商品類型' (Product Type) and '繁體書' (Traditional Chinese Books). The main content area shows the book's cover, title, and a detailed description. A context menu is open over the book's title, providing options like '複製' (Copy), '複製目顯示文字的連結' (Copy link to displayed text), and '翻譯網頁/顯示原文' (Translate page/show original text).

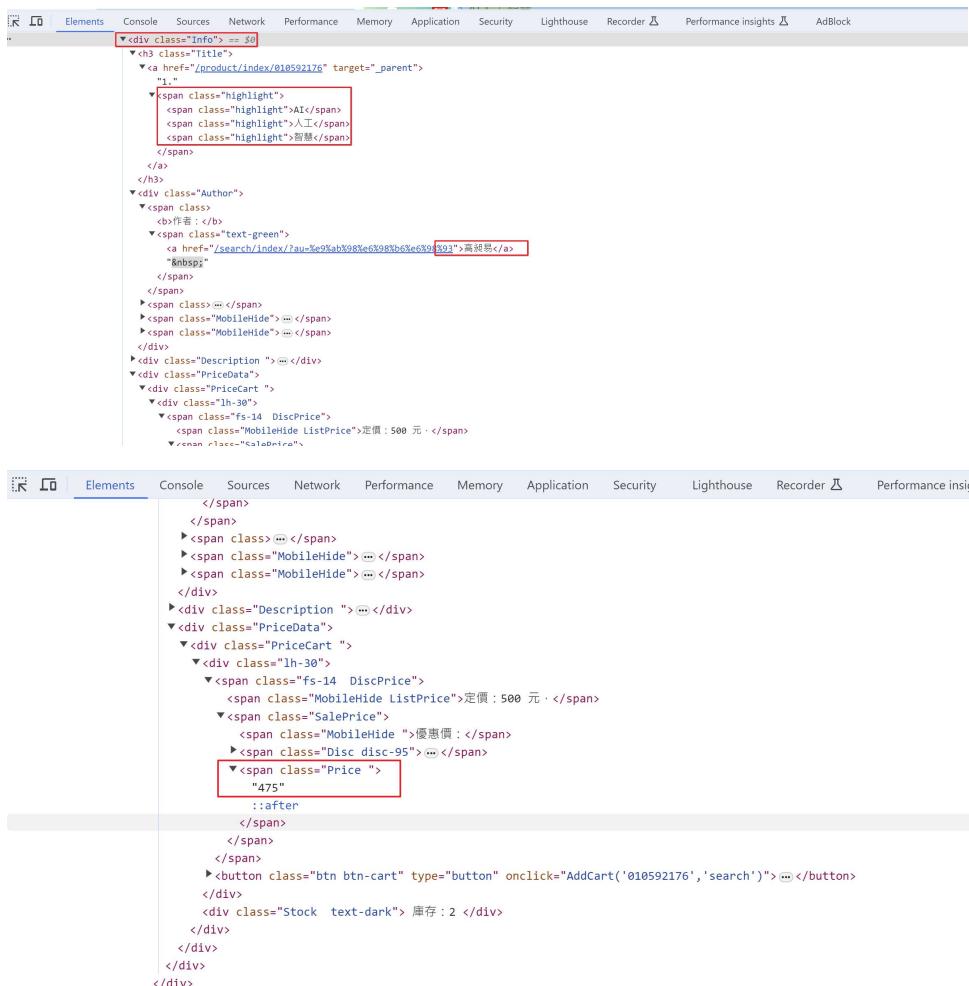
圖四、選取書名按檢查，找出書名的位置

This screenshot shows the Sanmin Bookstore website with the developer tools open again, focusing on the same product detail page for 'AI 人工智慧全知識'. The Elements tab highlights the specific element for the book's title. A context menu is open over the title, showing options like 'Ctrl + C' for copying and 'Ctrl + P' for printing. The right side of the screen shows the expanded CSS and JS code for that specific element, allowing for detailed inspection of the styling and structure.

圖五、選取價格按檢查，找出價格的位置



圖六、選取作者按檢查，找到作者位置



圖七、找到包含書名、作者、價格的最小範圍 div

```

import csv
import requests
from bs4 import BeautifulSoup

books = [] # 儲存所有書的資料

#總共九頁
for i in range(1, 10):
    #用+str(i)來對網頁進行翻頁
    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)
    # 來對 url發送一個 GET請求，回應的內容是一個字串
    html = requests.get(url).text
    # 創建一個 BeautifulSoup物件，可以更好操作 html的結構
    soup = BeautifulSoup(html, 'html.parser')

    goods = soup.find_all('div', class_='Info') # 找到一個div包含書籍資料（書名、作者、價格）

```

圖八、建立一個 books list 來存取所有的書籍資料，利用 for 迴圈來對每頁資料進行爬蟲，總共九頁，用 str(i)來進行翻頁，用 goods 來儲存 div 爬到的資料

```

#總共九頁
for i in range(1, 10):
    #用+str(i)來對網頁進行翻頁
    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)
    # 來對 url發送一個 GET請求，回應的內容是一個字串
    html = requests.get(url).text
    # 創建一個 BeautifulSoup物件，可以更好操作 html的結構
    soup = BeautifulSoup(html, 'html.parser')

    goods = soup.find_all('div', class_='Info') # 找到一個div包含書籍資料（書名、作者、價格）

    # 利用迴圈抓取資料裡的書名、作者、價格
    for good in goods:
        name = good.find('h3', class_='Title').a
        # 抓取書籍資料（書名）
        author = good.find('div', class_='Author').find('span', class_='text-green').a
        # 抓取一本書籍資料（作者）
        price = good.find('span', class_='Price')
        # 抓取一本書籍資料（價格）
        if name is not None and author is not None and price is not None:
            # 判斷是否有抓取到資料
            book=[] # 儲存每本書的資料(書名+作者+價格)
            name_element = "書名：" + name.text + " " # 取出書名資料的text
            author_element = "作者：" + author.text + " " # 取出作者資料的text
            price_element = "價格：" + price.text + "元" # 取出價格資料的text
            price_value = int(price.text) # 把價格的數值部分並轉換為int(用來排序)
            book.append(name_element) # 把抓到的書名、作者、價格存到book
            book.append(author_element)
            book.append(price_element)
            books.append((book, price_value)) # 把書籍資料book和價錢(用來排序)一起存儲books

```

圖九、再利用一個 for 迴圈對爬到的 div 內容裡抓取書名、作者、價格的位置，書名包在<h3 class=” Title ” 裡面的<a 裡，作者則是在<div class=” Author ” 裡的<span class=” text-green ” 裡面的<a 位置，價格在<span class=” Price ” 的位置，然後再利用

if 判斷確認有抓到相對應的位置，如果有，則建立一個 book 的空 list 儲存每一本書的資料(書名、作者、價格)，把前面找到的位置 name、author、price 加上. text 提取對應的內容並存放到 book 裡 並另外用 int(price. text) 把價格轉為 int 用來排序，最後再把 book(每本書的書名、作者、價格)跟 price\_value(每本書自己的價格數字)，一起存入到 books。

```
# print(books)
# 按價格進行排序
sorted_books = sorted(books, key=lambda x: x[1]) # 在books裡用price_value也就是x[1]來排序
for item in sorted_books:
    print(item[0],"\n") # 印出排序的結果
```

圖十、Print(books)，確認每本書的詳細資料(書名、作者、價格)和 int 的價格數字都有印出，然後利用 sorted 搭配 key=lambda x:x[1]，針對每本書的 int 價格數字去做排序(由小到大)存入 sorted\_books 裡，再用一個 for 迴圈，把排序後的 sorted\_books 的書籍詳細資料印出。

```
# 用text儲存資料
textfield = open("books.txt", 'w', encoding='utf-8') # 新建檔案若存在則清空並寫入資料，並使用 UTF-8 編碼寫入
textfield.write("學號[B1043001]姓名[許良亦]\n") # 寫入學號姓名並換行
textfield.write("\n") #換行
for item in sorted_books: # 寫入資料
   textfield.write(item[0][0] + "\n") # 依序寫入書名
   textfield.write(item[0][1] + "\n") # 寫入作者
   textfield.write(item[0][2] + "\n") # 寫入價格
   textfield.write("\n") # 换行
textfield.close() # 關閉檔案
print("TXT file 建立完成!")
```

圖十一、用 Python open 語法，創建一個 books. txt，並用 utf-8 編碼寫入，然後寫入學號姓名，然後用 for 迴圈依序寫入書名、作

者、價格，全部寫完後，關閉檔案，並印出建立完成。

```
# 用csv儲存資料
with open("books.csv", 'w', encoding='utf-8') as csvfile:
    # 打開名為 "books.csv" 的 CSV 檔案，並使用 UTF-8 編碼寫入
    csv_writer = csv.writer(csvfile)
    # 寫入CSV文件的標題行
    csv_writer.writerow(["書名", "作者", "價格"])
    # 寫入書籍資訊
    for book in sorted_books:
        csv_writer.writerow(book[0])
print("CSV file 建立完成!")
```

圖十二、用 with open as file 語法，創建一個 books.csv，並把

標題(書名、作者、價格)寫到標題行，用 for 迴圈把資料寫入，寫

入完後，印出建立完成。



```
[...]
[*書名: 7.FLAG'S 創客·自造者工作坊：進擊的 Arduino！AI 人工智慧互動遊戲機，'作者: 施威銘研究室'，'價格: 949元']
[*書名: 44.Flag's創客·自造者工作坊：ESP32 × Python AIoT大應用，'作者: 施威銘研究室'，'價格: 949元']
[*書名: 134.AI×LINE聲控·人臉辨識生活大應用，'作者: 施威銘研究室'，'價格: 949元']
[*書名: 19.運用AI人工智慧協助SAP ERP重複付款電腦稽核實例演練(附試用教育版軟體+教學演練資料)，'作者: 黃秀鳳'，'價格: 1080元']
[*書名: 43.AI離群分析(Outlier)－帳務與費用查核智能稽核應用實例演練(附試用教育版軟體+教學演練資料)，'作者: 黃秀鳳'，'價格: 1080元']
[*書名: 57.運用AI協助ESG實踐：以溫室氣體盤查與碳足跡計算為例(附教學演練資料)，'作者: 黃秀鳳'，'價格: 1080元']
[*書名: 59.香氛學：氣味、芳香、香水，探索人類最私密的嗅覺感官世界，'作者: Gestalten出版社'，'價格: 1080元']
[*書名: 109.運用AI人工智慧協助SAP ERP銷售資料分析性複核實例演練(附試用教育版軟體+教學演練資料)，'作者: 黃秀鳳'，'價格: 1080元']
[*書名: 149.AI智能稽核：文字探勘於合約查核實例演練(附試用教育版軟體+教學演練資料)，'作者: 黃秀鳳'，'價格: 1080元']
[*書名: 67.AI 必須！從做中學貝氏統計：從事機器學習、深度學習、資料科學、大數據分析一定要懂的統計利器，'作者: Therese M. D onovan'，'價格: 1140元']
[*書名: 158.JCAATS：資料分析與智能稽核(附試用教育版軟體+教學演練資料)，'作者: 黃秀鳳'，'價格: 1350元']
[*書名: 79.FLAG'S 創客·自造者工作坊 Python×AI 生醫感測健康大應用，'作者: 施威銘研究室'，'價格: 1519元']
[*書名: 38.資通安全電腦稽核系列：防火牆管理與高風險電子商務詐騙郵件(BEC)查核實例上機演練(共二冊附CD)，'作者: 黃秀鳳'，'價格: 2160元']
[*書名: 144.舞弊鑑識電腦稽核系列：行員盜用久未往來帳戶+運用班佛定律Benford's Law 查核實例上機演練(兩冊附CD)，'作者: 黃秀鳳'，'價格: 2160元']
[*書名: 157.數字看保險：透視家族傳承，'作者: 王仁川'，'價格: 3555元']
[*書名: 116.日本人工智慧產業垂直應用分析，'作者: 富士Chimera總研'，'價格: 9500元']
[*書名: 117.電商暨零售大廠發展布局分析，'作者: 王琬鈞'，'價格: 9500元']
TXT file 建立完成!
CSV file 建立完成!
```

圖十三、程式執行之結果，印出排序後的書籍資料，並把資料寫入

txt 檔跟 CSV 檔。

學號：B1043001，姓名：許良亦

書名：118. 辣警花 & AI 巡邏車（全）  
作者：八木澤景一  
價格：108元

書名：129. 變態博士駕到！！01  
作者：左近堂繪里  
價格：108元

書名：130. 變態博士駕到！！03  
作者：左近堂繪里  
價格：108元

書名：131. 變態博士駕到！！02  
作者：左近堂繪里  
價格：108元

書名：3.AI 人工智慧小百科  
作者：娜歐米  
價格：113元

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

書名：127. 搭檔之魂01  
作者：gyuo-漫畫  
價格：124元

書名：113. 積淹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

書名：95.CTIMES零組件雜誌  
作者：遠播資訊  
價格：180元

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

書名：48. 少年VI  
作者：李郁棻  
價格：205元

### 圖十三、txt 檔寫入結果

書名	作者	價格
書名：118. 辣警花 & AI 巡邏車（全）	作者：八木澤景一	價格：108元
書名：129. 變態博士駕到！！01	作者：左近堂繪里	價格：108元
書名：130. 變態博士駕到！！03	作者：左近堂繪里	價格：108元
書名：131. 變態博士駕到！！02	作者：左近堂繪里	價格：108元
書名：3.AI 人工智慧小百科	作者：娜歐米	價格：113元
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
書名：127. 搭檔之魂01	作者：gyuo-漫畫	價格：124元
書名：113. 積淹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
書名：95.CTIMES零組件雜誌	作者：遠播資訊	價格：180元
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
書名：48. 少年VI	作者：李郁棻	價格：205元
書名：150. 恐懼的馬赫數	作者：董少尹-作	價格：205元
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：田若斐譯	價格：216元
書名：29.BBC Micro:bit 入門與學習	作者：黃國明	價格：224元
書名：2.AI 人工智慧全知道	作者：陳俊中	價格：225元
書名：125. 寇乃馨英文片語一把罩－醫學園2	作者：寇乃馨	價格：225元
書名：153. 當代散文選讀	作者：陳謙	價格：228元
書名：152. 文化行政與政策分析	作者：沈中元	價格：234元
書名：32. 人工智能技術進步對勞動就業的影響研究	作者：沈紅兵	價格：237元
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	作者：路易思	價格：252元
書名：82. 奇蹟機器狗	作者：威爾·麥金塔許	價格：252元
書名：93. 我的AI同學02：AI的應用領域	作者：陳靜芬-文	價格：261元
書名：94. 我的AI同學01：AI的發展史	作者：陳靜芬-作	價格：261元

### 圖十四、此圖為 CSV 檔寫入的結果