**淡江大學資訊工程學系 學年度**

**專題實驗成果報告**

**教師專業課程推薦系統**

**-以淡江大學Azure系統為例**

**專題學生：**

**410411176簡亦芃**

**410410905呂峻逸**

**410410277張智凱**

**目錄**

**一、摘要-----------------------------------------------2**

**二、研究動機-------------------------------------------2**

**三、研究過程與方法-------------------------------------2**

**四、研究結果與討論------------------------------------13**

**五、展示海報------------------------------------------13**

**六、參考資料------------------------------------------15**

一、摘要

　　教師的開課系統會將全部的課程展示出來，選課的過程繁瑣。本專題藉由推薦系統，讓使用者在選擇課程時可以更加快速且精確的找尋到理想的課程。

二、研究動機

　　聽聞現在的授課選課系統並不方便，選課時會就只是讓教授自己寫出想開的課程。便打算做一套推薦系統，能夠對不同的教授推薦其高可能性會選擇的課程列表，使選課過程更加便捷。

三、研究過程與方法

1. **歷史資料獲取**

　　要訓練課程推薦系統首先需要有歷年的課程資料，藉由過去的資料分析出每位教師會開的課程。因此將目標放到可以查詢到歷屆開課資料的校務行政資訊查詢系統。

1. **分析校務網站的資料架構**

進入查詢系統後，首先會出現使用<select>做成的年度以及課程選項(圖1-1)，因為此專題對於資料的需求是每年度學期中全部的課程，因此選則所有課程、所有學院、全部系所、全部年級進行查詢（圖1-2）。

　　進入查詢頁面後就可以看到排列整齊目標資料了（圖1-3）。

1. **利用Selenium抓取目標資料**

**Selenium**可以開啟網站瀏覽器，並模擬使用者的動作。剛剛已經作為使用者到達了目標資料，現在要編寫一個能模仿剛才動作流程的執行程式。

　　首先回到校務網站並打開開發者工具，可以找到控制學年度到年級的按鈕都是使用<select>製作的，每個選項的value是‘學年度’接上‘學期序’（圖1-4），而課程、學院、所系名稱及年級都需要選擇‘全部’

，也就是value是：課程=1（圖1-5）、學院=%（圖1-6）、所系=\*

（圖1-7）、年級＝\*（圖1-8），最後按下id=Button1的查詢按鈕（圖1-9）就可以到達目標資料位置。

到達目標位置後，可以知道資料的排放方式。每個元素與同列的包裹在一起（圖1-10），每列又再被包裹成整個頁面（圖1-11）。因此讓程式對於頁面查找每個列，再對於每個列查找元素並打包，就可以完成一個頁面的資料抓取。抓取完一個頁面後，還有不少的頁面需要抓取，因此在抓取完此頁面後需要讓程式按下id=Button4的下一頁按鈕（圖1-12）繼續抓取資料。程式會持續執行，當到達最末頁時，下一頁按鈕會出現無法互動的標籤（1-13），因此當偵測到按鈕擁有此標籤時程式便不會繼續翻頁，而是將資料輸出至檔案後停止。

　　調整學年度與學期序後便可以抓取當期全部的開課資料。

1. **推薦系統運行**
2. **利用MYSQL儲存以及前處理**

將上述獲取的資料匯入MYSQL資料庫，用以選擇性得讀取資料以減輕記憶體負擔，並在MYSQL WORKBENCH中將各學期資料統整為一個表，同時將總資料進行前處理，包含去除空值資料和去除助教資料。

1. **使用TF-IDF轉換資料**

首先利用TF-IDF將教授的歷史課程紀錄中的課程名稱轉換為數值表示，TF-IDF可以分析一個詞在文件中的重要程度，將每個課程名稱視為一個文件，而教授的課程紀錄則視為文本集合，主要計算每個詞在文本集合中的出現頻率(TF)以及在多少文件中出現(IDF)，從而獲取每個課程名稱的重要性。

1. **教授相似度計算**

我們將每個教授的課程紀錄表示為向量後，向量的每個維度對應一個課程，其值為該教授對該課程的TF-IDF權重。之後先輸入教授名稱當作目標教授並使用餘弦相似度(Cosine Similarity)計算所有教授與目標教授的相似度，最後依據相似度排序，找出與目標教授相似度高的其他教授。

1. **課程推薦**

最後找到相似度前幾高的教授的課程資料，將其的TF-IDF值再與目標教授的課程資料的值做計算並排序，如此就能得出根據相似度值降冪的推薦課程列表。

1. **結果輸出**

利用PyQt6做推薦系統的前端UI，利用Qt Designer做大略的排版再用PyQt6中的pyuic模組將.ui檔轉為.py檔與後端程式做連結

按下查詢按鈕後將單行輸入框中的名字存入全域變數中放到後端進行運算，再把執行結果傳回前端ui並以表格方式呈現

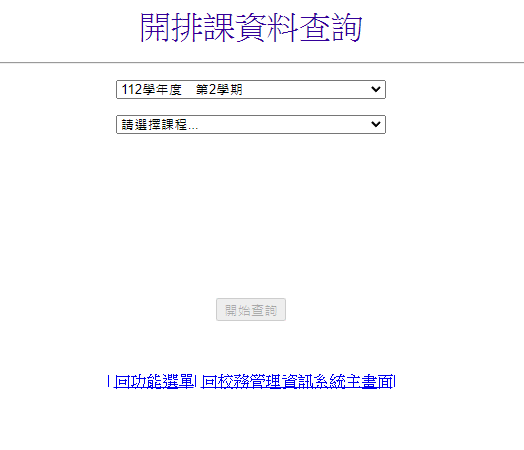


圖1-1

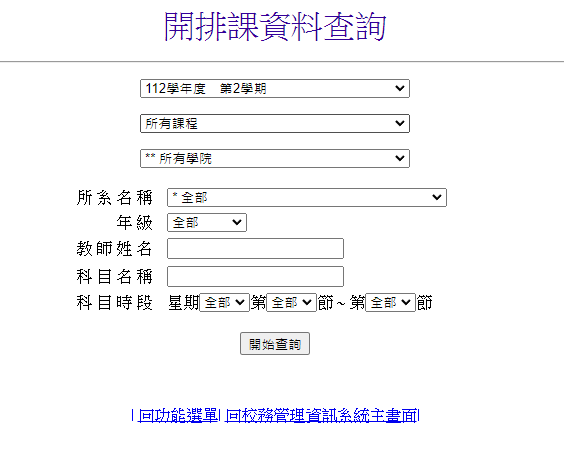


圖1-2



圖1-3



圖1-4



圖1-5



圖1-6



圖1-7

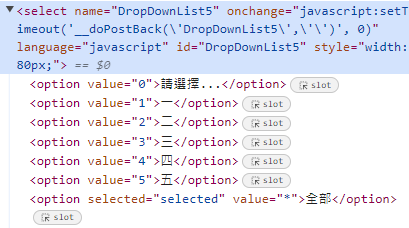


圖1-8



圖1-9



圖1-10

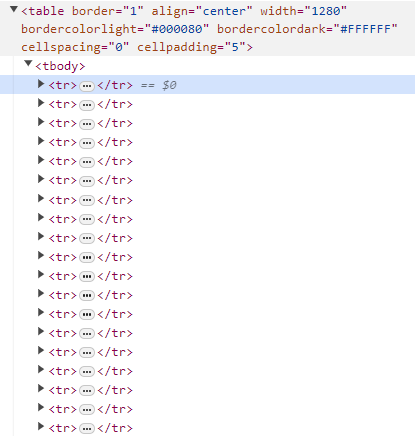


圖1-11



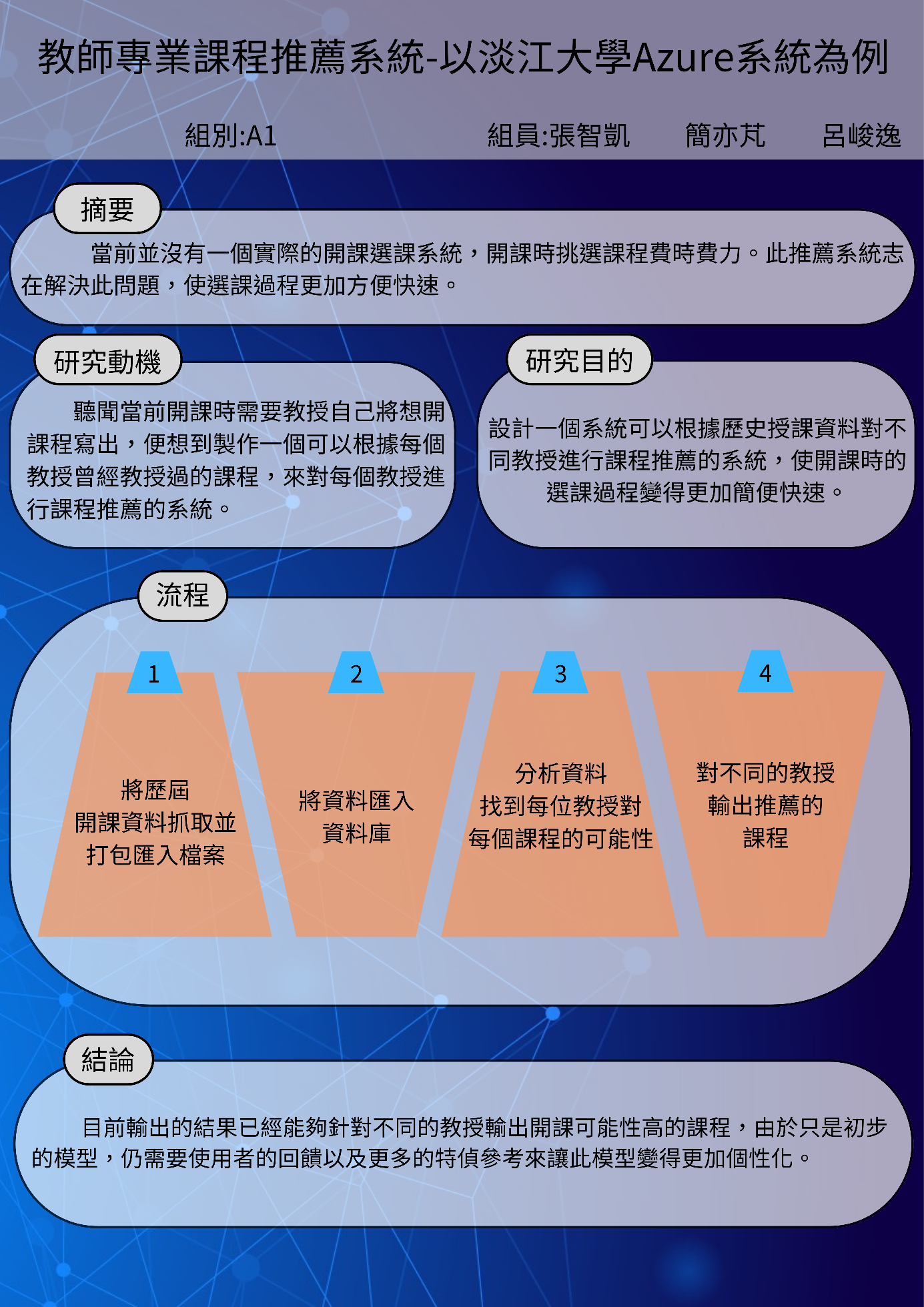
圖1-12



圖1-13

四、研究結果與討論

　　　　目前輸出的結果已經能夠針對不同的教授輸出開課可能性高的課程，由於只是初步的模型，仍需要使用者的回饋以及更多的特偵參考來讓此模型變得更加個性化。

五、展示海報

**六、參考資料**

動態網頁爬蟲第一道鎖 — Selenium教學：如何使用Webdriver、send\_keys(附Python 程式碼) <https://medium.com/marketingdatascience/selenium%E6%95%99%E5%AD%B8-%E4%B8%80-%E5%A6%82%E4%BD%95%E4%BD%BF%E7%94%A8webdriver-send-keys-988816ce9bed>

透過 Selenium 操作下拉式選單 (Select) <https://jzchangmark.wordpress.com/2015/03/05/%E9%80%8F%E9%81%8E-selenium-%E6%93%8D%E4%BD%9C%E4%B8%8B%E6%8B%89%E5%BC%8F%E9%81%B8%E5%96%AE-select/>

PyQt6 教學<https://steam.oxxostudio.tw/category/python/pyqt6/index.html>

Day 13 - [語料庫模型] 01-TF-IDF與餘弦相似性

<https://ithelp.ithome.com.tw/articles/10271604>