

以交友軟體形式建構台大課程搜尋網站

1 簡介

1.1 動機與目的

1. 前人所做的「紅樓夢 台大課程網關鍵字查詢系統」([連結](#))雖然已可利用關鍵字搜尋課程大綱，但仍只能搜尋該關鍵字，無法模糊搜尋。因此希望開發一個網頁，使用如交友軟體的頁面與操作方式，推薦使用者其他與關鍵字相關的課程，讓選課不再局限於自己已知的課程範圍中。

1.2 目標

1. 輸入一關鍵字後，能夠找出與其出現頻率最相關的其他「標籤」，並利用該標籤與關鍵字找出課程，最後以網頁呈現，利用簡單的按鈕選擇喜歡/不喜歡。而喜歡的課程可以匯入課程網中「我預計要選的課程」列表。

2 方法

2.1 資料取得 ([crawler.py](#))

1. 利用 Python 中的正則表達式 (re 套件)，夾取所需資訊。欄位包含「課程名稱」、「開課學期」、「授課對象」、「授課教師」、「課號」、「課程識別碼」、「班次」、「學分」、「全年/半年」、「必修/選修」、「上課時間」、「上課地點」、「備註」、「ceiba 課程網站」、「課程大綱」(syllabus)、「課程大綱網址」，共 16 項。
2. 將資料儲存成 JSON 檔。
3. 註：crawler.py 為上述所提紅樓夢專案中的爬蟲檔案，因已有現成程式碼故不另重寫。檔案在 github 上開源 ([連結](#))，已取得原作者同意再利用。

2.2 原始碼運作說明

2.2.1 資料前處理

1. 使用以下套件：library(tidyverse), library(jiebaR), library(quanteda)
2. 選取流水號、課名、授課教師以及課程大綱這四欄作為主要資料處理對象
3. 針對「課程大綱」這一欄，利用 JiebaR 套件斷詞。斷詞結果如下所示

```
> head(df_course, 2)
# A tibble: 2 x 2
  course_id content
  <int> <chr>
1      108-1 1 課程 資訊 課程名稱 學士 論文 上 Thesis b s 1 開課 學期
2      108-2 2 課程 資訊 課程名稱 台灣 人口 變遷 專題 Seminar on Demog
raphic Population Changes in Taiwan 開課 學期 108-...
```

4. 使用 Quanteda 套件轉換成可供分析的語料庫

5. 使用 Quanteda 套件建立 Document-Feature Matrix
6. 刪除 Stop words
7. 利用 Quanteda 套件建立 Co-Occurrence Matrix，結果如下所示

```
> head(fcm_course)
Feature co-occurrence matrix of: 6 by 6 features.
6 x 6 sparse Matrix of class "fcm"
      features
features  學士 論文 thesis 土木工程 學系 葛宇
學士      10   1    2         3    8    1
論文       1  15   2         1    1    1
thesis     2   2   1         2    2    2
土木工程   3   1   2         1    4    1
學系       8   1   2         4    0    1
葛宇       1   1   2         1    1    0
```

2.2.2 關鍵字輸入的自訂函數

1. 整理資料形式、新增欄位
2. 使用 Dplyr 套件篩選與關鍵字共同出現頻率大於一定數字的標籤
3. 回傳排序前十的標籤

2.2.3 輸出與標籤相關課程的自訂函數 (select_course.R)、測試(0104mark.rmd)

1. 整理資料形式、新增代數與欄位名稱等
2. 新增一欄位、計算關鍵字與標籤出現次數的加權平均
3. 利用該分數來排序，並輸出前十名的課程編號

2.2.4 網頁呈現 (app.R)

1. 利用 Shiny 套件實現即時互動頁面
2. 輸入一關鍵字後，會回傳與關鍵字語意網路最為相關的課程結果
3. 結果可由頁面上之 Like 或是 Dislike
4. 來選擇是否加入台大課程網選課名單，或是跳過繼續尋找下一門課程

3 結果

3.1 網頁呈現（連結：https://yeutong.shinyapps.io/r_class_shiny_app/）



3.2 搜尋範例

3.2.1 以「音樂」為關鍵字進行搜尋



3.2.2 以「科技」為關鍵字進行搜尋



3.3 例外處理

若輸入不在詞庫內的關鍵字，則會顯示：該關鍵字無效，請重新輸入。如下所示。



4 討論

4.1 貢獻

此 App 可以讓學生在選課之餘，多了一種新的方式選擇課程，不同於以往需要逐條尋找課程，或是只能以課程標題搜尋課程。利用語意相關性及其建構的語意網路，學生可以享受選課的樂趣，更進一步探索未知但卻和自身興趣相符的課程。

4.2 限制

此 App 最主要的限制有三點。第一點，課程內容皆為 108-1，還為更新至最新的 108-2 版本。第二點，因為資料太過龐大(一萬筆課程)，因此在運行時會造成運行時間過長。第三點，此 App 中的資料僅為一萬筆中的 100 筆，語意相關的代表性可能不足，而影響最終結果，其原因為第二點所致。

要改善此情況，第一點只需要將 108-2 的課程資料重新爬蟲更新，就可解決。第二點和第三點，則可需要另外建置一個資料庫，以改善前端、後端的檔案輸入、輸出的時間差。

5 附錄

5.1 分工

海報內容製作、程式碼撰寫、以及期末報告的產出，皆由陳博煒，劉禹彤以及柯逸鈞三人共同接力產出。