



第十組、分解茶 建構交友軟體形式的台大課程搜尋網站

劉禹彤 台大經濟系
柯逸均 台大經濟系
陳博煒 台大外文系
涂穎珊 台大工管系

製作流程與架構

1 資料取得

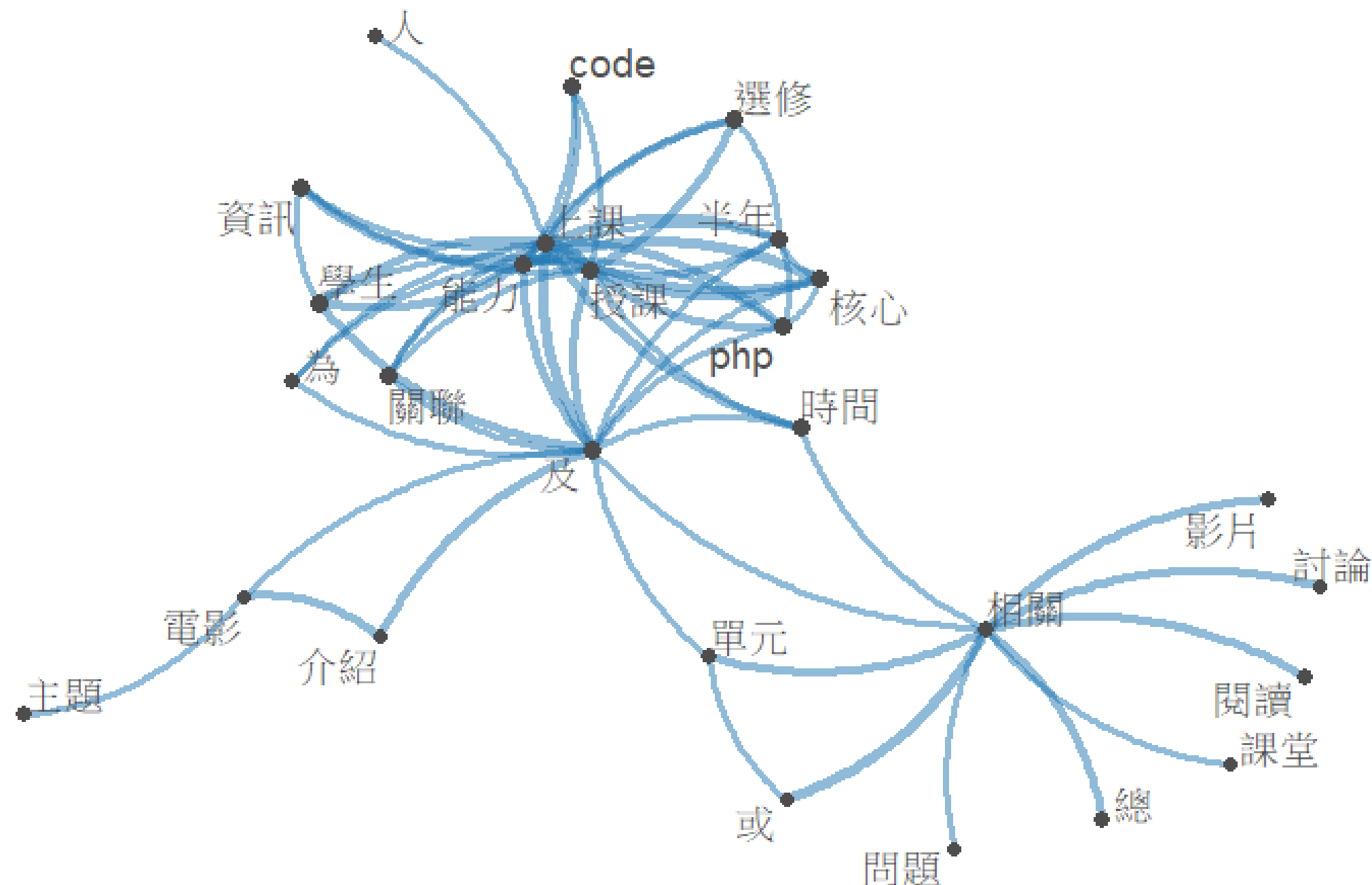
- 1) 利用 Python 中的正則表達式 (re 套件), 夾取所需資訊
- 2) 將資料儲存成 JSON 檔
- 3) 最後選取流水號、課名、授課教師以及課程大綱這四欄作為主要資料處理對象

2 資料前處理

所使用的套件有: Tidyverse, JiebaR, Quanteda

- 1) 將文字檔讀入, 並利用 JiebaR 套件斷詞
- 2) 使用 Quanteda 套件轉換成可供分析的語料庫
- 3) 使用 Quanteda 套件建立 Document-Feature Matrix
- 4) 刪除 stop words
- 5) 利用 Quanteda 套件建立 Co-Occurrence Matrix

語意網路示意圖



3 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

4 輸出與標籤相關課程的自訂函數

- 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
- 4) 輸出前十名的課程編號

5 網頁呈現

- 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 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

動機與目的

- 1) 前人所做的「紅樓夢」課程網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。
- 2) 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。

目標

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。

概念流程圖

取得課程網站資料

前處理

– 斷詞、詞頻分析、共詞分析

自訂函數

– 輸入關鍵字
– 輸出標籤、相關之課程

網頁呈現

結果

範例一: 以「音樂」為關鍵字進行搜尋

音樂	音樂	音樂
課程流水號: 76148	課程流水號: 84599	課程流水號: 56439
課程名稱: 電影美術概論	課程名稱: 音樂作品中的愛	課程名稱: 現代華語文學英譯
授課老師: 陳慧	授課老師: 蔡振家	授課老師: 陳榮彬
dislike like	dislike like	dislike like

範例二: 以「科技」為關鍵字進行搜尋

科技	科技	科技
課程流水號: 23528	課程流水號: 37916	課程流水號: 96674
課程名稱: 資料礦掘	課程名稱: 醫學工程導論	課程名稱: 建築構造解析
授課老師: 張瑞益	授課老師: 宋孔彬	授課老師: 張盈智
dislike like	dislike like	dislike like



試試看吧!

