

**P1**

**P2~P3**

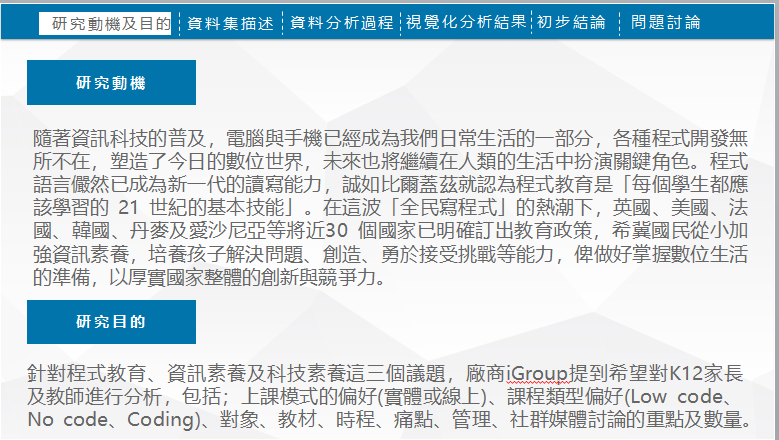
**P4~P5**

**P6~P9**

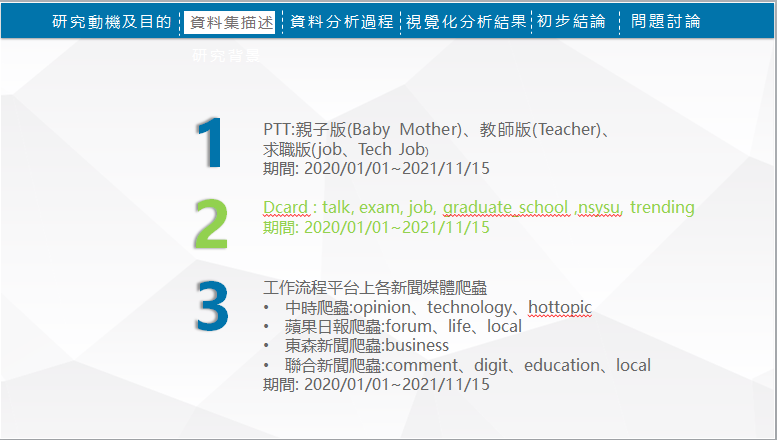
**P10~P12**

**P13**

1. ***研究動機及目的***



1. ***資料集描述***
   1. 資料集區間: 2020/1/1~2021/11/15



我們使用工作流程平台的PTT、Dcard、中時爬蟲元件、蘋果日報、東森新聞及

聯合新聞網爬蟲等六個元件蒐集2020/01/01到2021/11/15相關討論版(例如:

親子版、教師版、求職版、opinion、technology and education等)蒐集資料，

蒐集資料的關鍵字包括iGroup總經理在演講簡報中所提到的:程式設計/程式教

育/資訊素養/科技素養/CSS/JavaScript/Scratch/Minecraft/Python/R語言

no code/low code…等，我們小組列為關鍵字。

* 1. 資料集區間: 2019



由於廠商希望比較疫情前年度(2019)及疫情後(2020~2021)的社群討論焦點，本

組因此進一步分析2019的新聞與討論。另外，由於2019是108新課綱公布，

故加入八卦版資料。而因為Dcard和蘋果日報在工作流程系統出現起始日期無資

料的錯誤訊息，我們因此使用PTT、中時爬蟲元件、東森新聞及聯合新聞網爬蟲   
 等四個元件蒐集2019/01/01到2019/12/31相關討論版(例如:教師版、求職版、

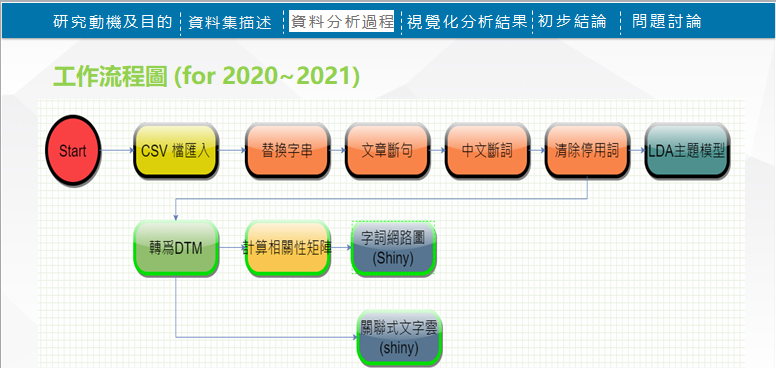
opinion、technology and education等)蒐集資料，蒐集資料的關鍵字包括

iGroup總經理在演講簡報中所提到的:課綱、資訊科技、科技素養、程式語言，

並加入小學生經常用來學習程式的軟體Scratch。另外，為了聚焦與資訊科技有

關的教育和課綱，因此排除的關鍵字為國文、歷史、英文、經濟和投資。

1. ***資料分析過程***
   1. 資料集區間: 2020/1/1~2021/11/15 的分析



對於前述爬蟲元件所抓到的資料轉出CSV檔，並人工彙整為單一CSV檔匯入，  
 先進行資料清理、斷句、斷詞、清除停用詞，並將清除停用詞之後的結果，  
 轉為DTM，並使用LDA主題模型元件進行資料分析。使用轉為DTM元件找出  
 前100個詞彙，然後產生計算相關性矩陣，並產生字詞網路圖及關連式文字雲。

* 1. 資料集區間: 2019 年的分析



2019資料的工作流程的分析元件安排和上一頁相同，但為了想了解2019因為新課綱發布而討論熱度高的關鍵詞，而加入ngrams和詞頻計算。

* 1. 本組將前述爬蟲資料合併為CSV檔，

2020~2021的資料總共匯入5481篇文章

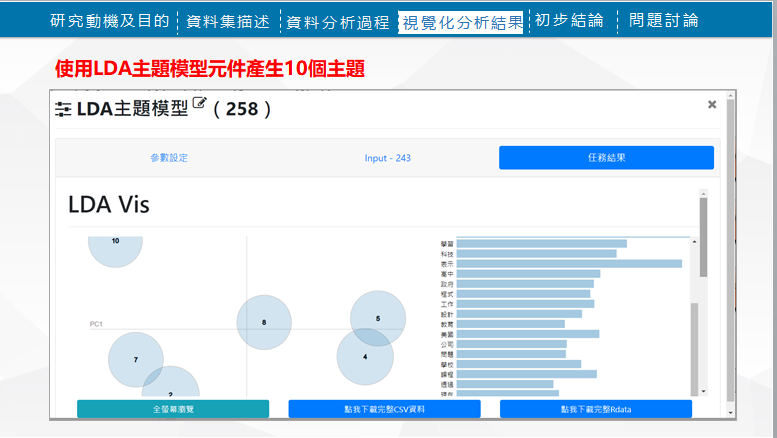


* 1. 而2019聚焦在新課綱後的文章，則總共匯入770篇文章



1. ***視覺化分析結果***

(1) 使用LDA主題模型元件產生10個主題



(2) 我們使用了LDA主題模型元件，產生10個主題，參數設定都先使用平台預

設的值。觀察主題1前五大詞彙分別是大學、科技、工作、公司及google，

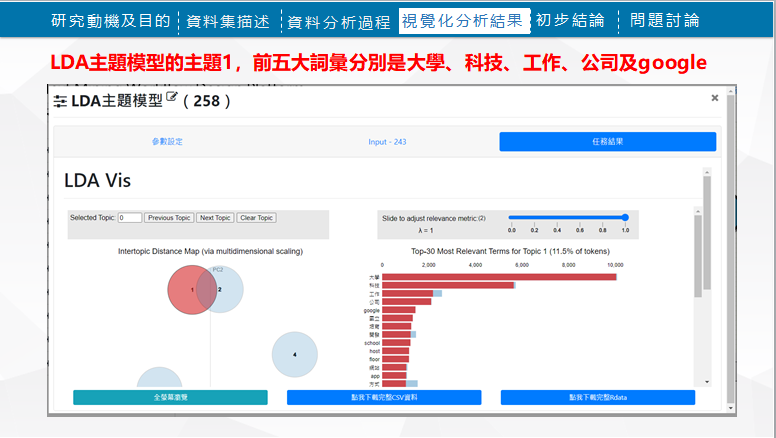
顯示在網路上各大媒體及討論版，大家對於程式教育及科技素養，討論的重

點大都是大學裡有關科技的議題，這個科技的議題可能包括程式語言的學

習、資訊的相關課程，然後這些程式設計的學習會影響到未來工作，因為這

些資訊相關的技能，都是求職時，公司的徵才條件，例如google這個關鍵

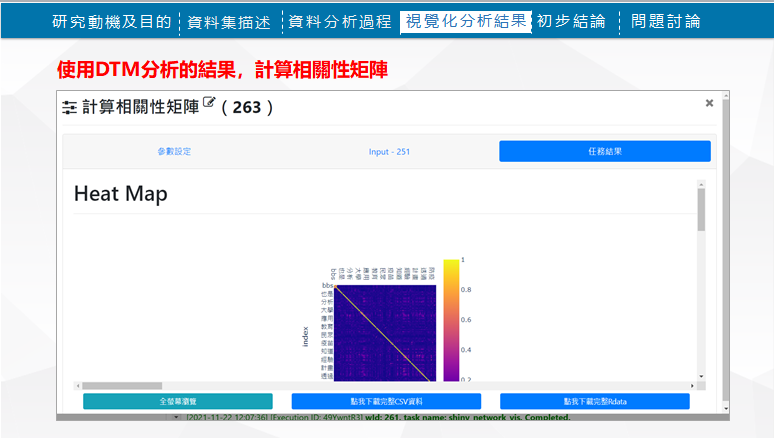
字，很多網友都有提到，而且熱烈討論。



(3) 使用清除停用字的斷詞轉為DTM



(4) 使用清除停用字的斷詞轉為DTM



(5) 從計算相關性矩陣產出字詞網路圖，從這個字詞網路圖可以觀察出一些

訊息，這部分我們在後面初步的結論會再擷取細部的圖跟大家做個報告。



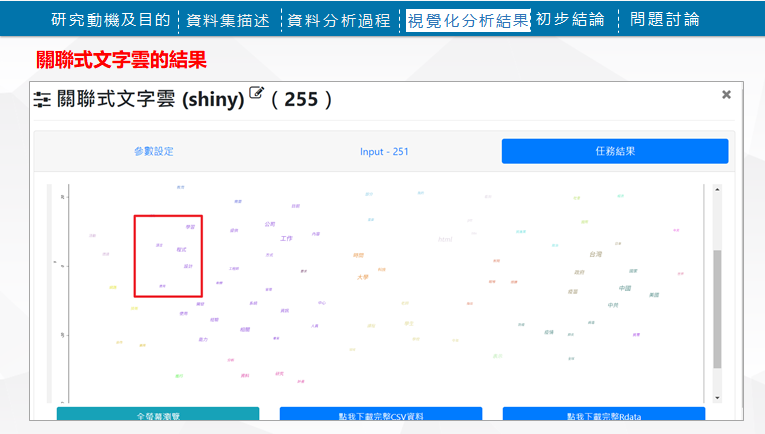
(6) 將DTM的輸出，使用關聯式文字雲元件繪圖



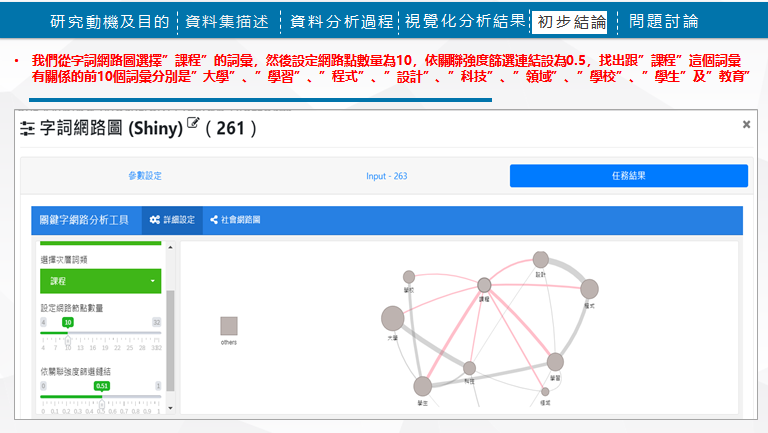
(7) 從關聯式文字雲的分析結果，發現網友提到學習程式設計，都會關心

語言及應用，至於是何種程式語言，例如Python、R語言、HTML5

或CSS，可能需要看應用的領域

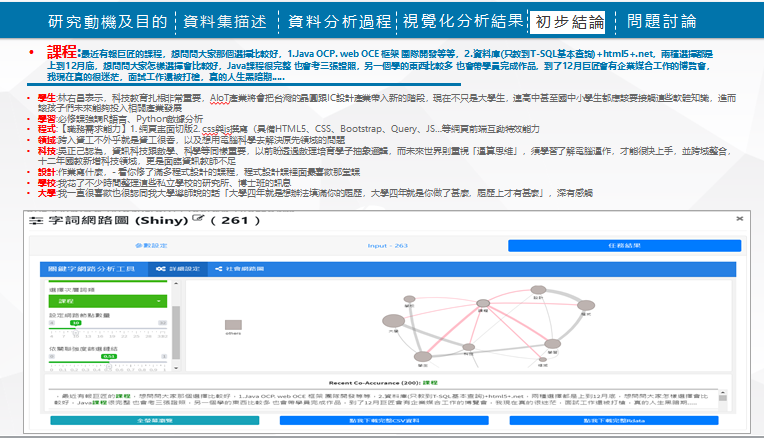


1. ***初步結論***
   1. 我們從字詞網路圖選擇”課程”的詞彙，然後設定網路點數量為10，依關聯強度篩選連結設為0.5，找出跟”課程”這個詞彙有關係的前10個詞彙分別是”大學”、”學習”、”程式”、”設計”、”科技”、”領域”、”學校”、”學生”及”教育”



* 1. 我們分別點選”課程”、”學生” 、學習”、”程式”、”設計”、”科技”、”領域”、”學校” 、”教育”及 ”大學”，看看大家對於這些詞彙討論的詳細內容，例如”課程”，可看到有位網友說：*他最近有報巨匠的課程，想問問大家那個選擇比較好，1.Java OCP. web OCE 框架 團隊開發等等，2.資料庫(只教到T-SQL基本查詢)+html5+.net，兩種選擇都是上到12月底，想問問大家怎樣選擇會比較好，Java課程很完整 也會考三張證照，另一個學的東西比較多 也會帶學員完成作品，到了12月巨匠會有企業媒合工作的博覽會，我現在真的很迷茫，面試工作還被打槍，真的人生黑暗期.....*

“領域”，則看到有位網友提到：*跨入資工不外乎就是資工很香，以及想用電腦  
 科學去解決原先領域的問題*



* 1. 除了前述提到的字詞網路分析外，對於2019 新課綱發布那年，體驗也是很重要的關鍵詞彙，這個詞連結了科技、Ai、技術和智慧。它的相關訊息可以看到許多正在進行以及未來會發展的資訊科技教育趨勢，包括學校成立科技社團、跨界學習、政府成立科技中心，與業界合作及成立科技教育的STEAM聯盟等。這些訊息將有助廠商評估未來跨界及和學校或學校的社團合作等方案。(如下頁的字詞網路圖)



* 1. 另外，針對2019資料的詞頻分析，前50名討論熱度高詞組，討論最多的是學習歷程，同時教育部長潘文忠也出現在高頻名單，顯示課綱發布當年，媒體和民眾也很在乎教育部長說了什麼。依據分析結果，廠商若可以向學生或家長能提出有助於學習歷程建立及整合資訊科技的學習方案，將能進入這個新課綱改變帶來的新教育市場。



1. ***問題討論***

由於iGroup公司對於本小組以中山大學工作流程平台進行程式教育/資訊素養/科技素養主題分析，期望我們產生的報表能夠呈現依年齡層分、依性別分、依縣市別分、依討論族群分，這部分的需求，我們還在思考如何解決資料欄位取得的問題。



目前本小組想出的對應辦法:

(1) 既然沒有資料，那麼就自己生資料。透過Deep learning，自己訓練一個模型。

(2) 依自行標記判斷資料集裡的文章字串分別是屬於哪個身分族群，例如:家長或是老師或是教育機構業者。也可以判斷其孩子或本身是處理哪一個階段的教育層級，例如:幼稚園、小學、中學、或大學、或碩士...等等。

**※待進一步釐清問題：對於依性別與依縣市的資料分群需求，是否能從文字內容判斷？本組成員目前尚無明確解方，計畫於課堂上向老師及助教請益。**