**社群媒體分析期末專案報告**

主題：語言學習的趨勢分析

組員：

N074220001鍾承玫

N094220005謝宜蓁

N094220010吳貞儀

N094220016林映汝

# 分析摘要一(聯合新聞網)

* 1. **資料區間**：2017/1/1 – 2021/12/31
  2. **資料來源**：使用工作流程平台內建資料「聯合新聞網」
  3. **新聞類別**：education
  4. **搜尋關鍵字**：英文學習、英語學習、語言學習、雙語教育、雙語國家
  5. **任務結果**：7個欄位數、517筆資料
  6. **資料清理做法**：中文斷詞、清除停用詞、詞頻計算、計算ngrams、詞頻計算(bi-grams)
  7. **情緒分析**：Lexicon Based、Corpus Based、LIWC情緒分析
  8. **文辭和文件分析**：轉為DTM、關聯式文字雲、計算相關性矩陣、計算共現矩陣、單中心網路圖、字詞網路圖
  9. **主題模型：**LDA

# 分析摘要二 (PTT)

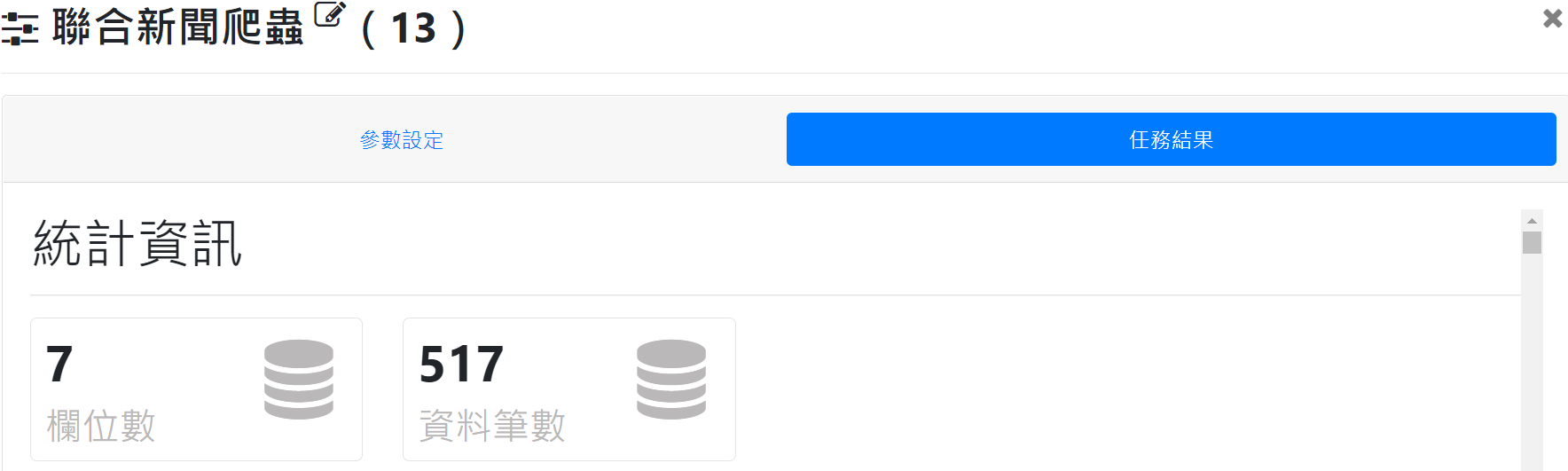
1. **資料區間：2017/1/1-2021/12/31**
2. **資料來源**：使用工作流程平台內建資料PTT
3. **資料看板**：BabyMother、Teacher、learnyf
4. **搜尋關鍵字**：英文學習、英語學習、語言學習、雙語教育、雙語國家
5. **任務結果**：10個欄位數、220筆資料
6. **資料清理做法**：文章斷句、替換字串、中文斷詞、清除停用詞、計算ngrams、詞頻計算
7. **情緒分析**：Lexicon Based、Corpus Based、LIWC情緒分析
8. **文辭和文件分析**：轉為DTM、計算相關性矩陣、計算共現矩陣、單中心網路圖、關聯式文字雲
9. **主題模型**：LDA

# 工作流程一 (聯合新聞網)



# 聯合新聞爬蟲





## 中文斷詞

定義詞彙如下，權重皆為100

108課綱、108新課綱、雙語國家、雙語教育、英文學習、英語學習、雙語教學、雙語學校、雙語師資、雙語學習、2030雙語國家、雙語實驗學校、教學研究中心、全民英檢、英語檢定、台北科技大學、彰化師範大學、台灣師範大學 、美國耶魯大學、亞東技術學院、澎湖科技大學、清華大學、銘傳大學 、元智大學、輔仁大學、成功大學、景文科大、樹德科大、正修科大、北科大、台師大、萬豪酒店、全國教師工會、文化教育基金會、韓國瑜、吳榕峯

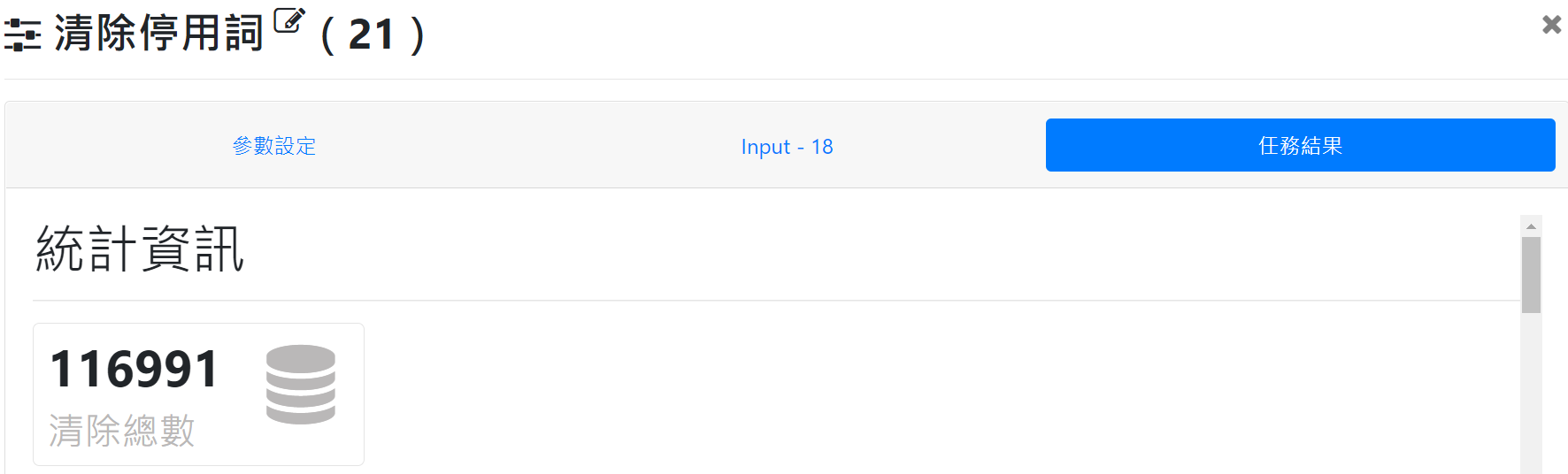




## 清除停用字

使用工作流程內建的參數設定，共清除116,991個字詞。



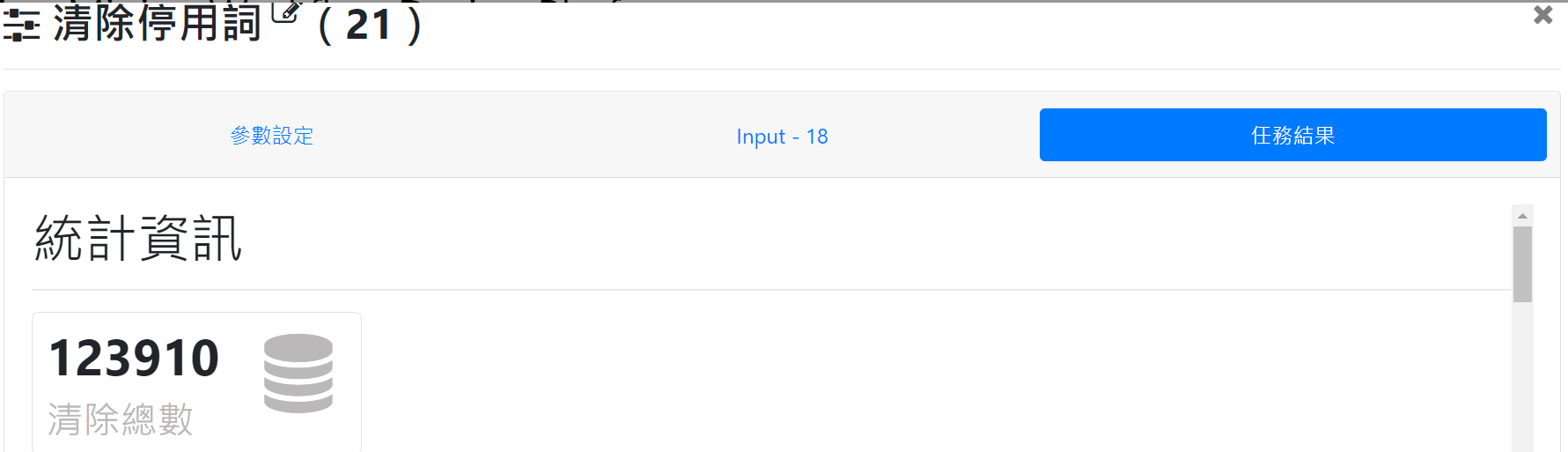


接著，反覆透過「詞頻計算」及「bi-gram詞頻計算」的結果，持續增加停用詞，最終結果如下所述：

共清除123,910個停用詞。

自定義停用詞如下：

表示、成為、目前、一次、指出、今天、上午、今年、分享、facebook、記者、qa、以下、提供、徐如宜、攝影、孩子、神人、實境秀、進行、透過、可請、致用、圖卡、家變、協助、參加。



## 詞頻計算



## 計算ngrams



## 詞頻計算(bi-grams)

一張含有 文字 的圖片

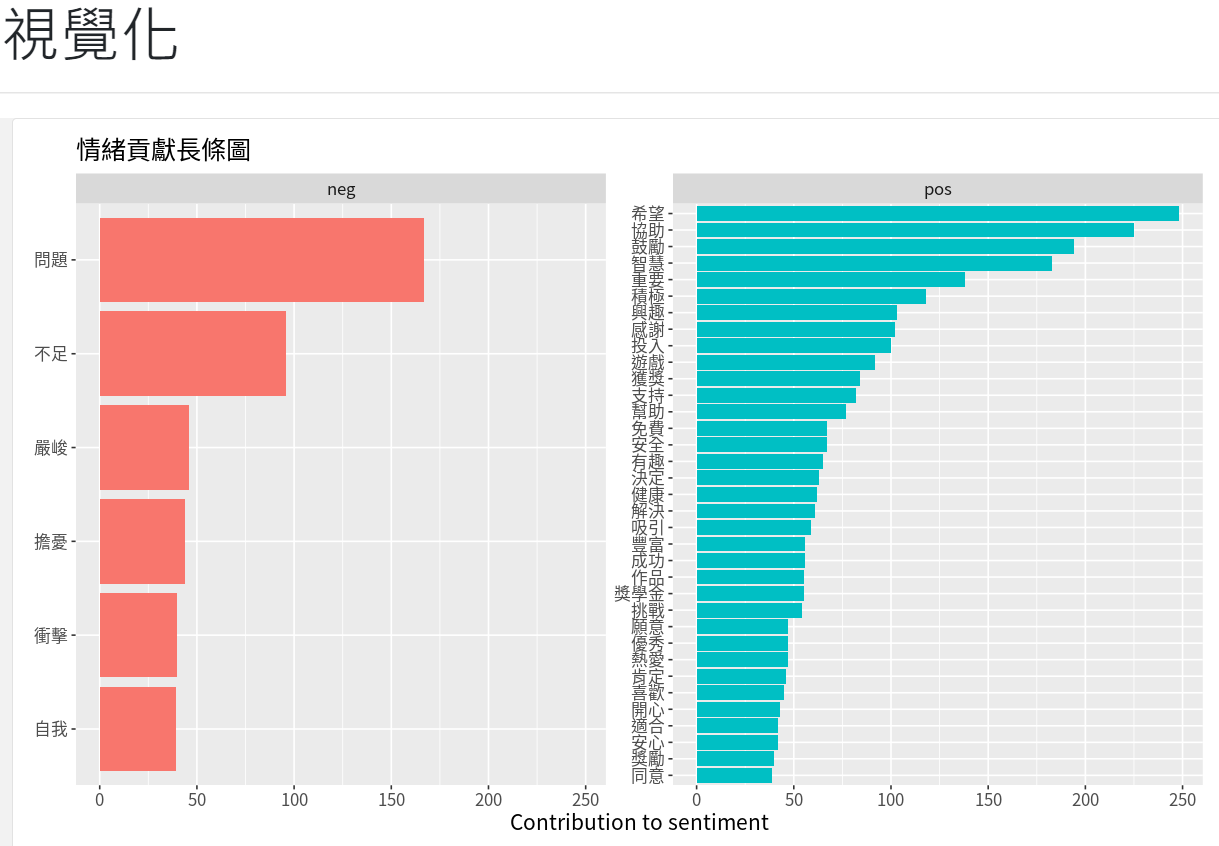
自動產生的描述

# 情緒分析

## Lexicon based

一張含有 文字 的圖片

自動產生的描述



## Corpus based



正向情緒數：447

## LIWC





## 轉為DTM

一張含有 文字 的圖片

自動產生的描述

## 關聯式文字雲

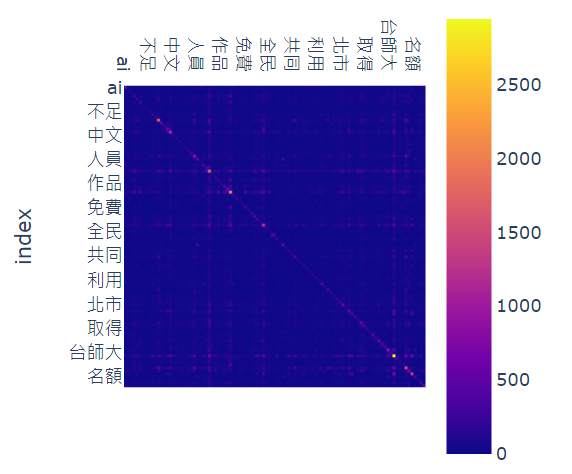
一張含有 文字 的圖片

自動產生的描述

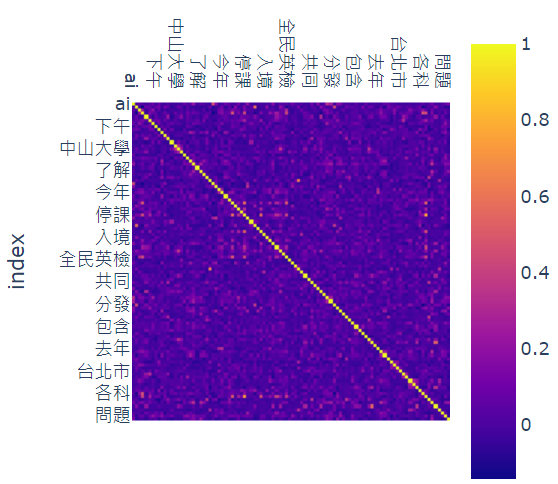
一張含有 地圖 的圖片

自動產生的描述

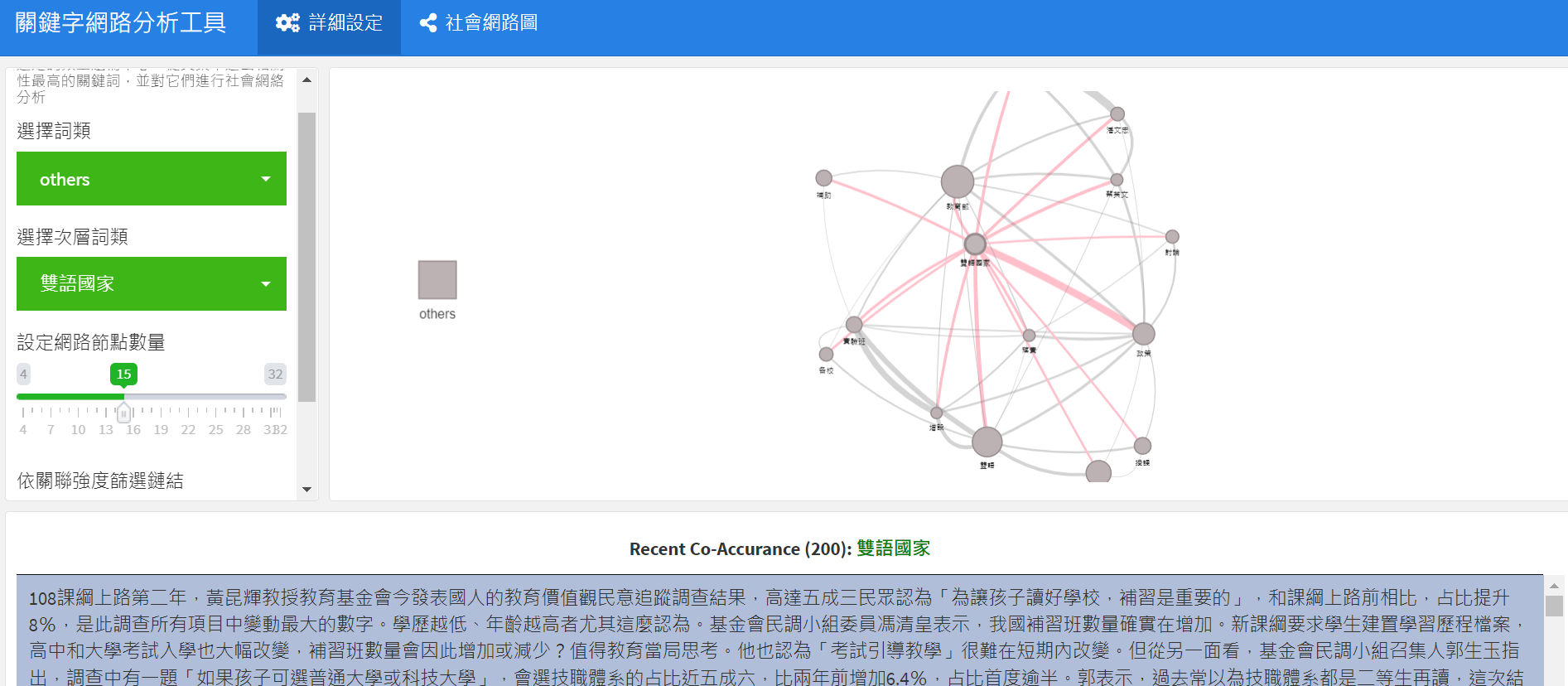
## 計算共現矩陣



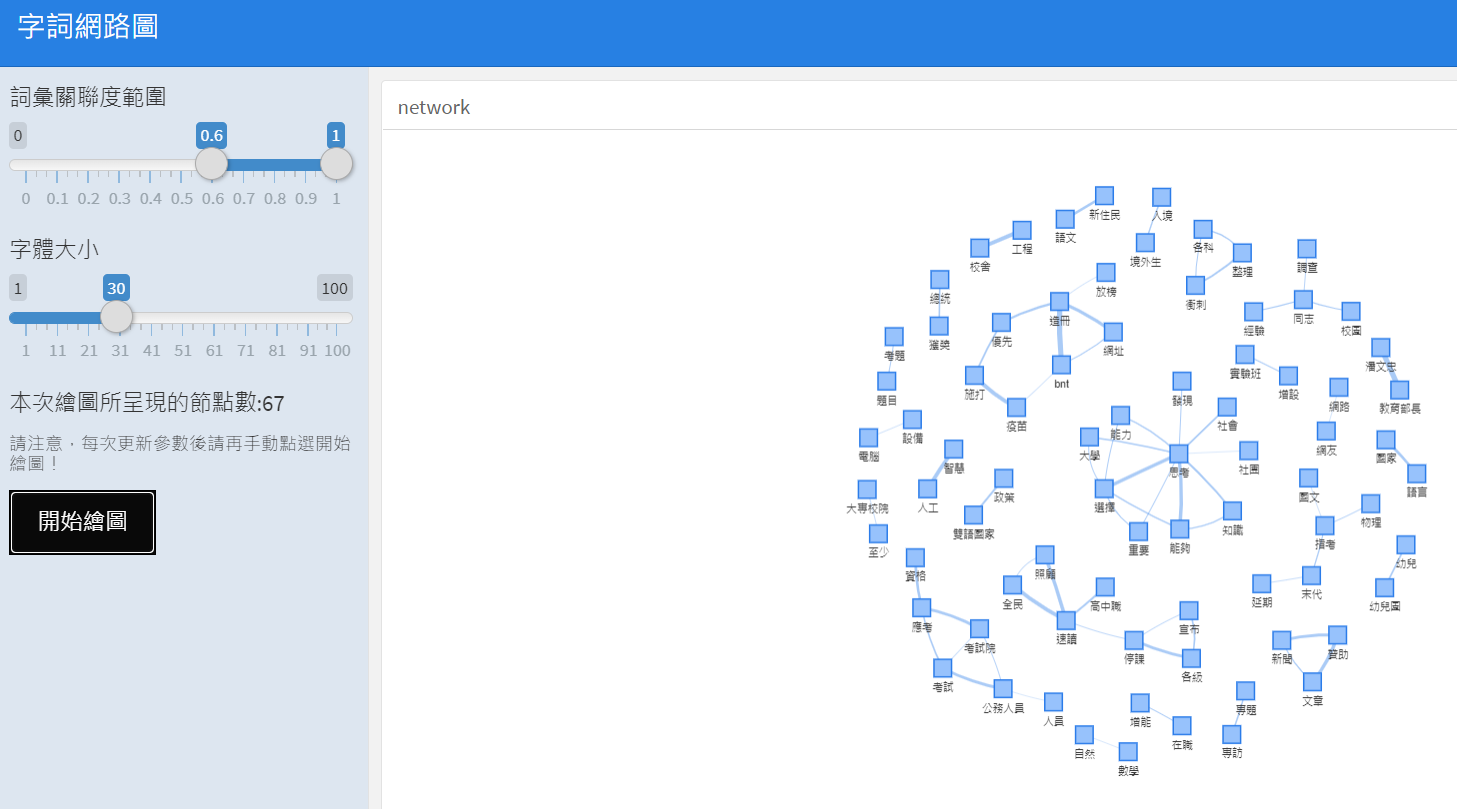
## 計算相關性矩陣



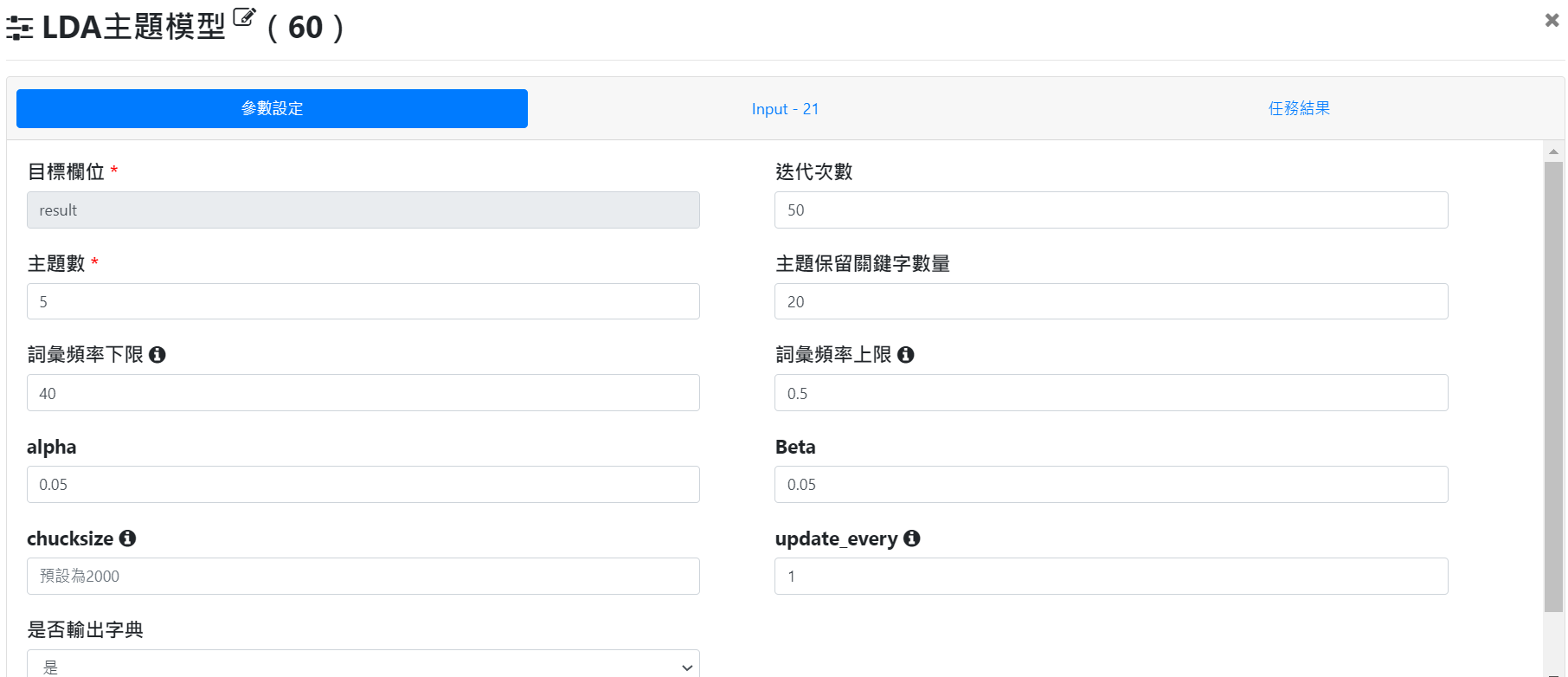
## 單中心網路圖



## 字詞網路圖



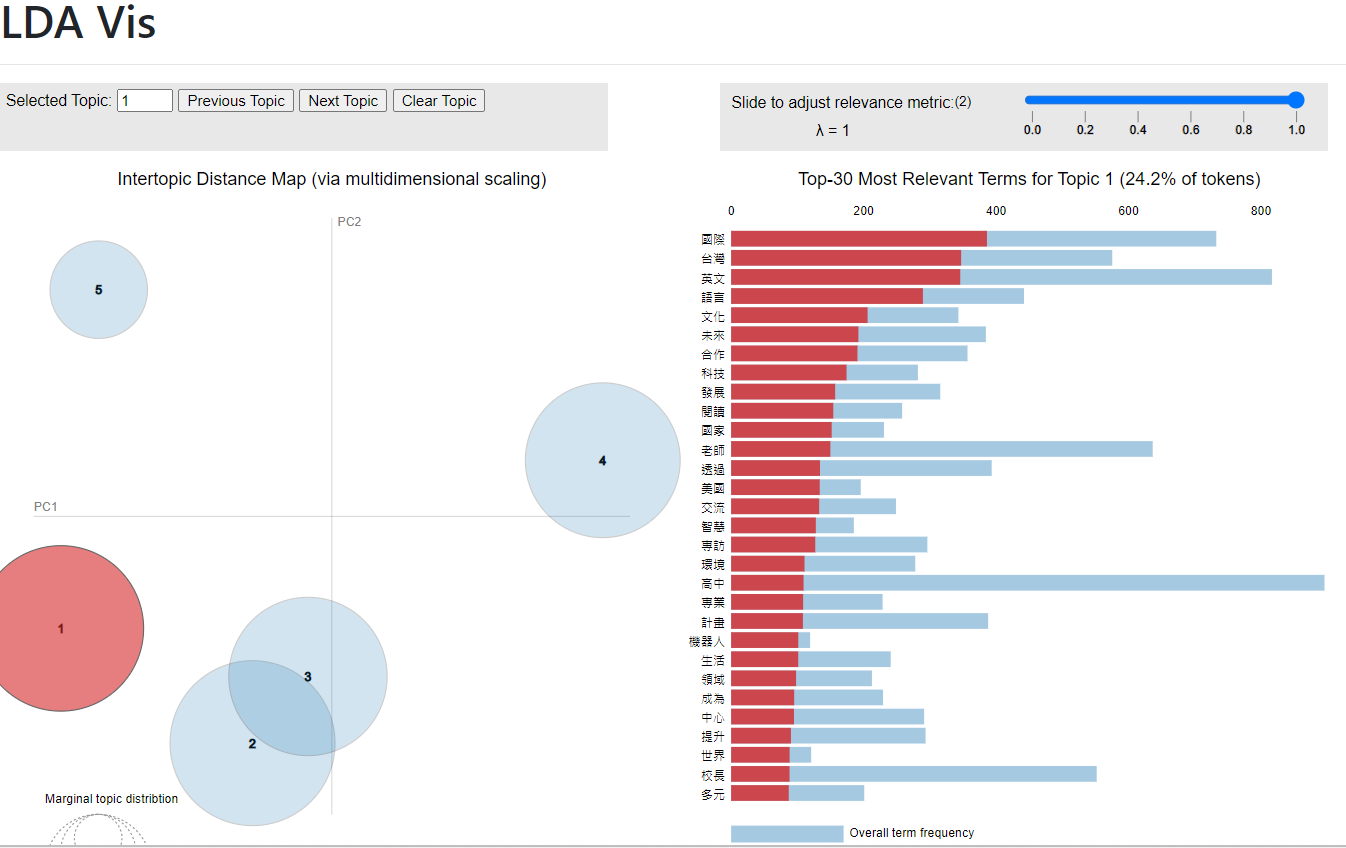
# 主題模型



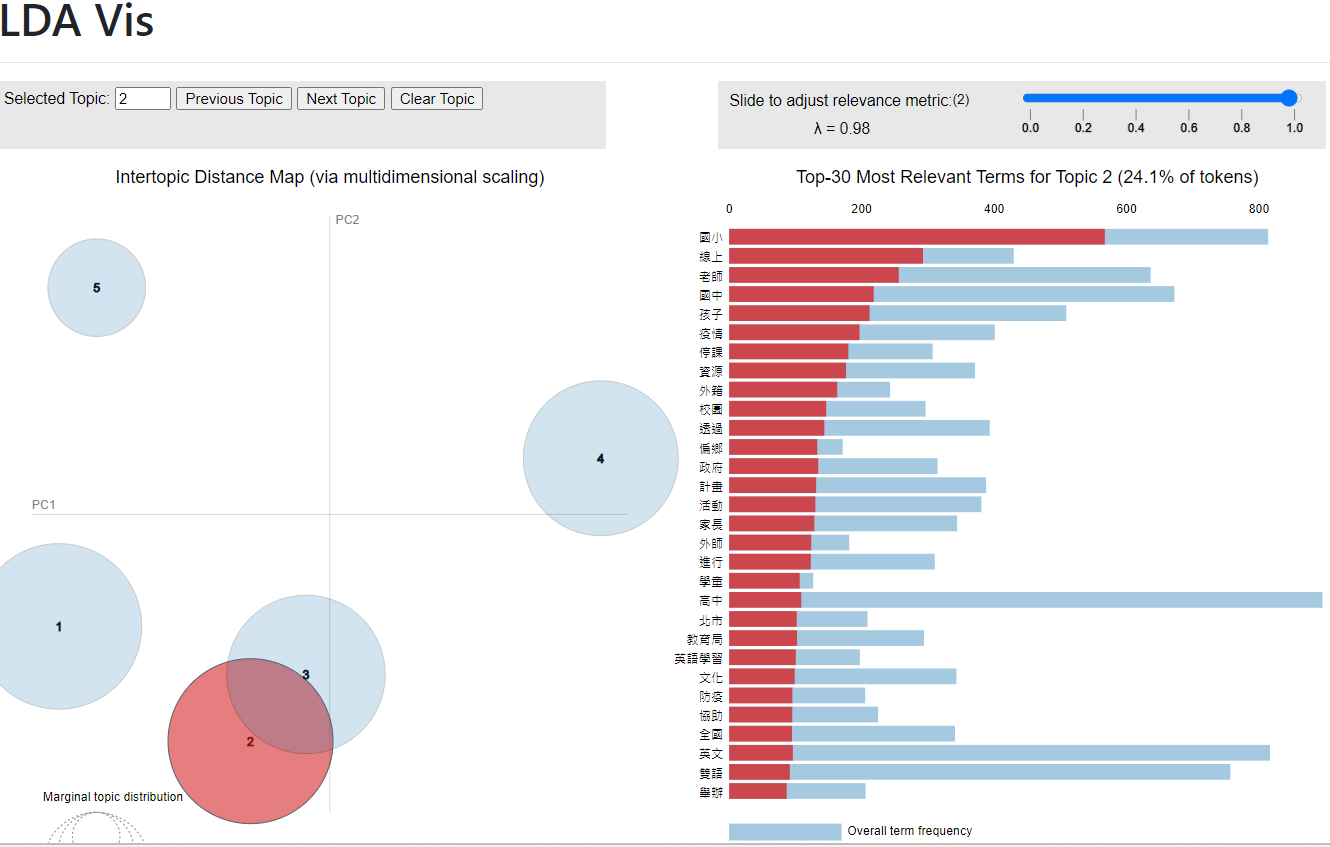
## 主題分群有五組:

國際接軌，疫情停課，高雄市雙語教育，雙語國家，課綱。

主題1：國際接軌



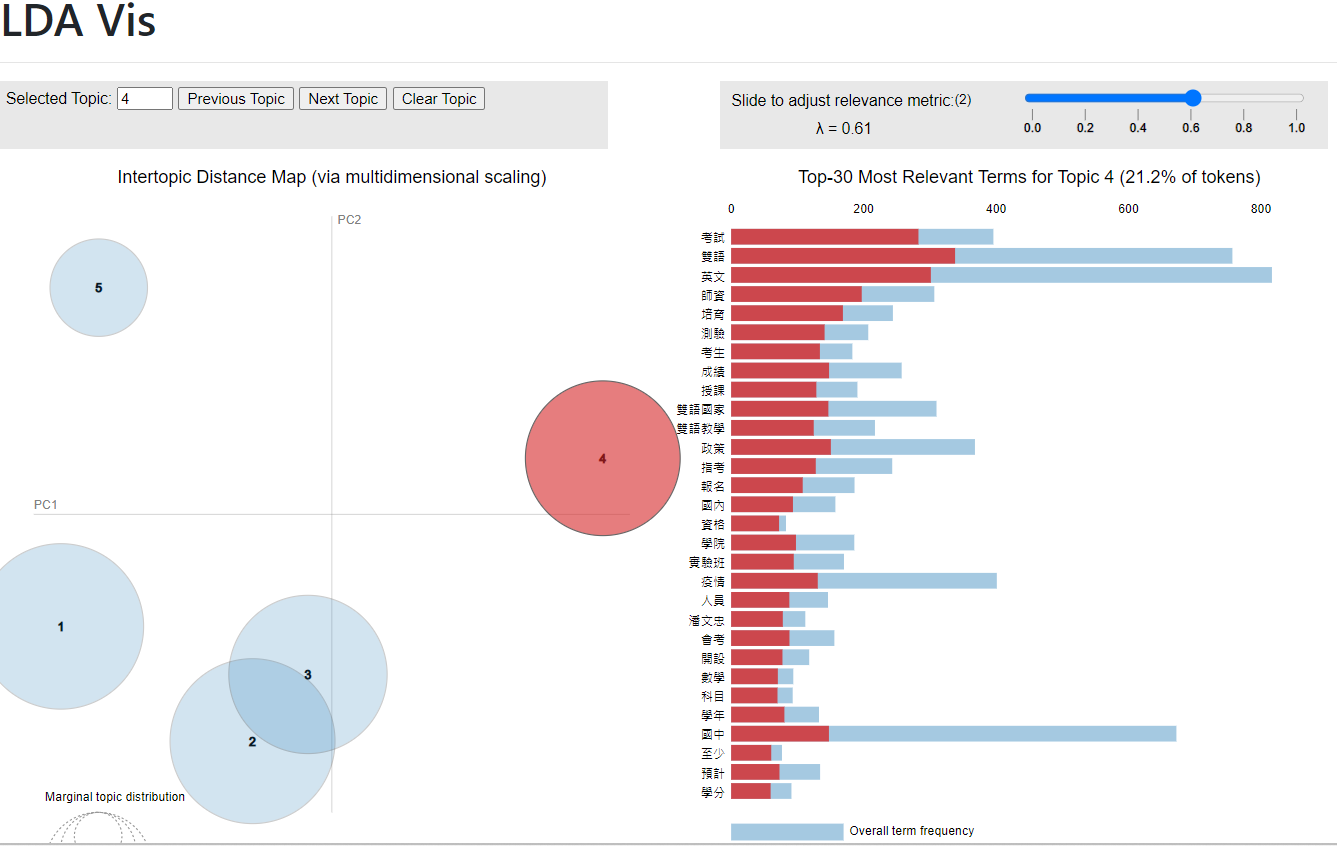
主題2：疫情停課



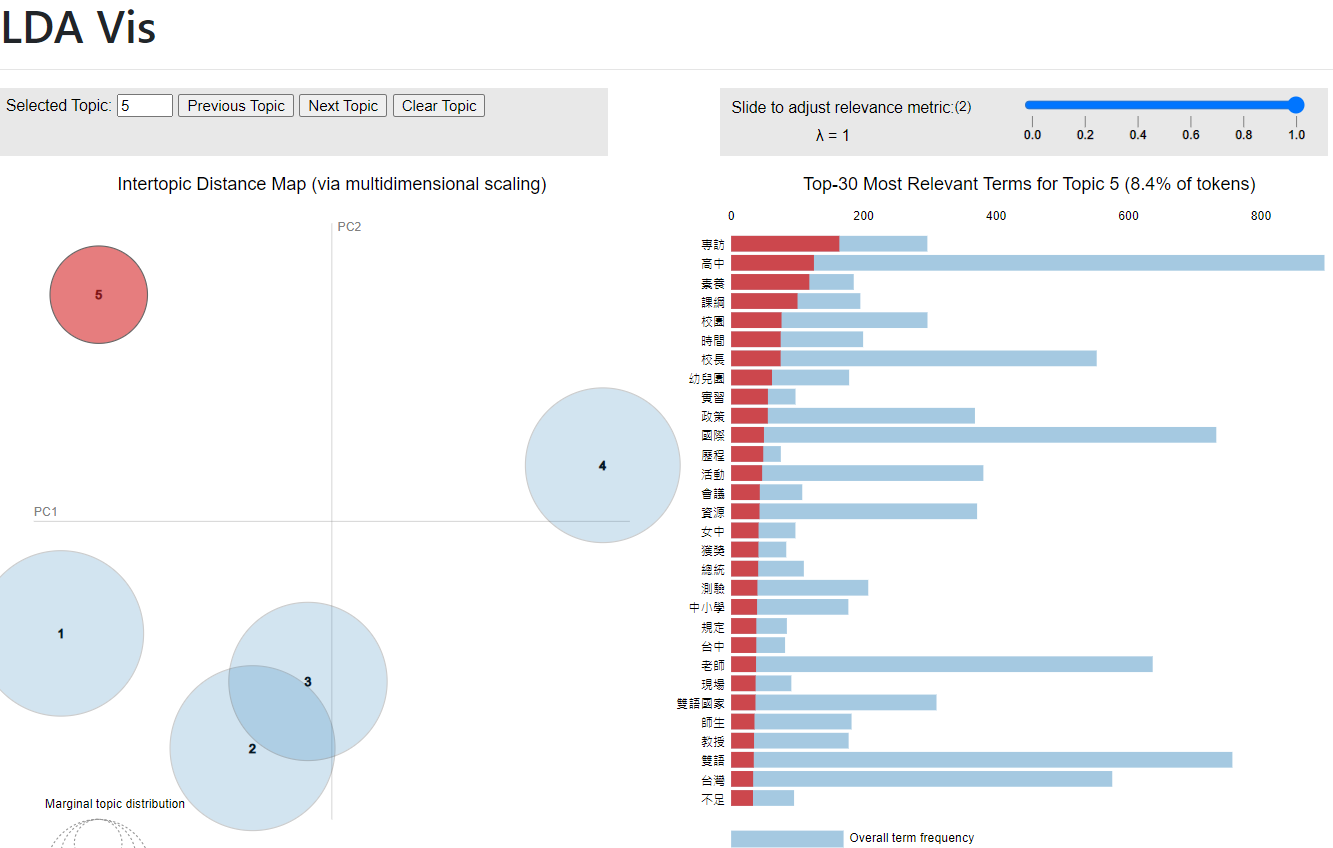
主題3：高雄市雙語教育



主題4：雙語國家



主題5：課綱



# 工作流程二(PTT爬蟲)

* 清理資料過程

1. 步驟一：先進行「文章斷句」與「替換字串」

文章斷句：參數設定以「\n\n、\n、\n\n\n、\n\n\n\n\n\n\n、\n\n\n\n」為主，共切割出10233筆。

替換字串：參數設定以「y>>歲、Y>>歲、M>>個月」為主，共替換掉432個字詞。

一張含有 文字 的圖片

自動產生的描述

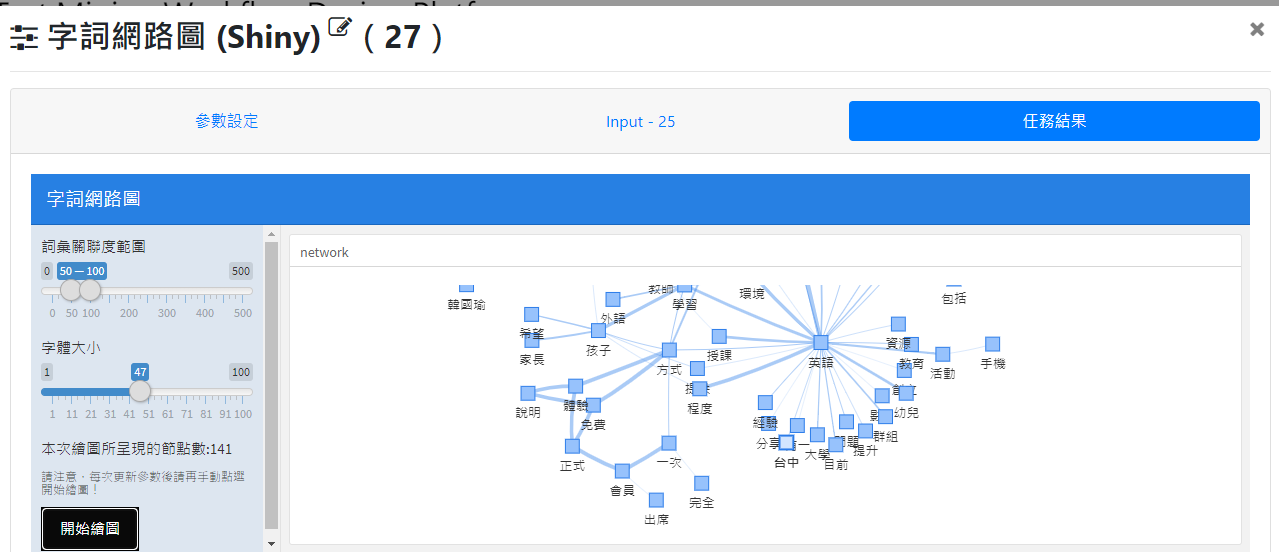
1. 步驟二：使用網路上所提供「中英文停用詞列表」做初步的字詞清理，共清出119398個字詞。

一張含有 文字 的圖片

自動產生的描述

1. 步驟三：檢視關聯式文字雲，出現「blogspot、com、events、facebook、https、imgur、jpg、pec、pecrob、pectw、talk、tw、www」，故將上述字詞納入清理字詞中，進行第二次停用詞整理。

一張含有 桌 的圖片

自動產生的描述

1. 步驟四：共清出124494個字詞，與第一次相比多清出5096個字詞。檢視字詞網路圖，出現「有一、一次、方式、包括、目前」，故將上述字詞納入清理字詞中，進行第三次停用詞整理。
2. 步驟五：共清出125934個字詞，與第二次相比多清出1440個字詞。

文字雲

一張含有 文字, 個人, 螢幕擷取畫面 的圖片

自動產生的描述

一張含有 桌 的圖片

自動產生的描述Ngramg數量：1,000,090

文字雲字詞：線上、英文、語言、學習

* Lexicon Based情緒分析

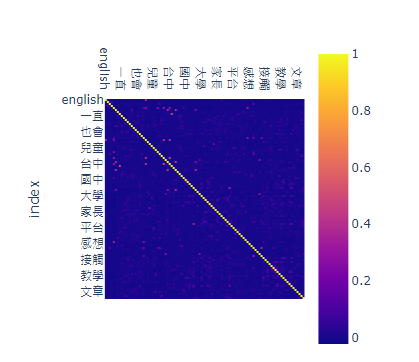
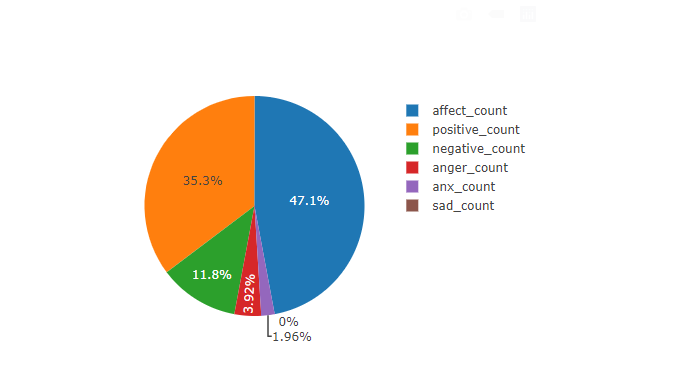
最大正向情緒6、最大負向情緒4

負向關鍵字詞：問題、擔心、強迫

正向關鍵字詞：分享、朋友、希望

* LIWC情緒分析

最大正向情緒6、最大負向情緒4



* 計算相關性矩陣
* 字詞網路圖

一張含有 地圖 的圖片

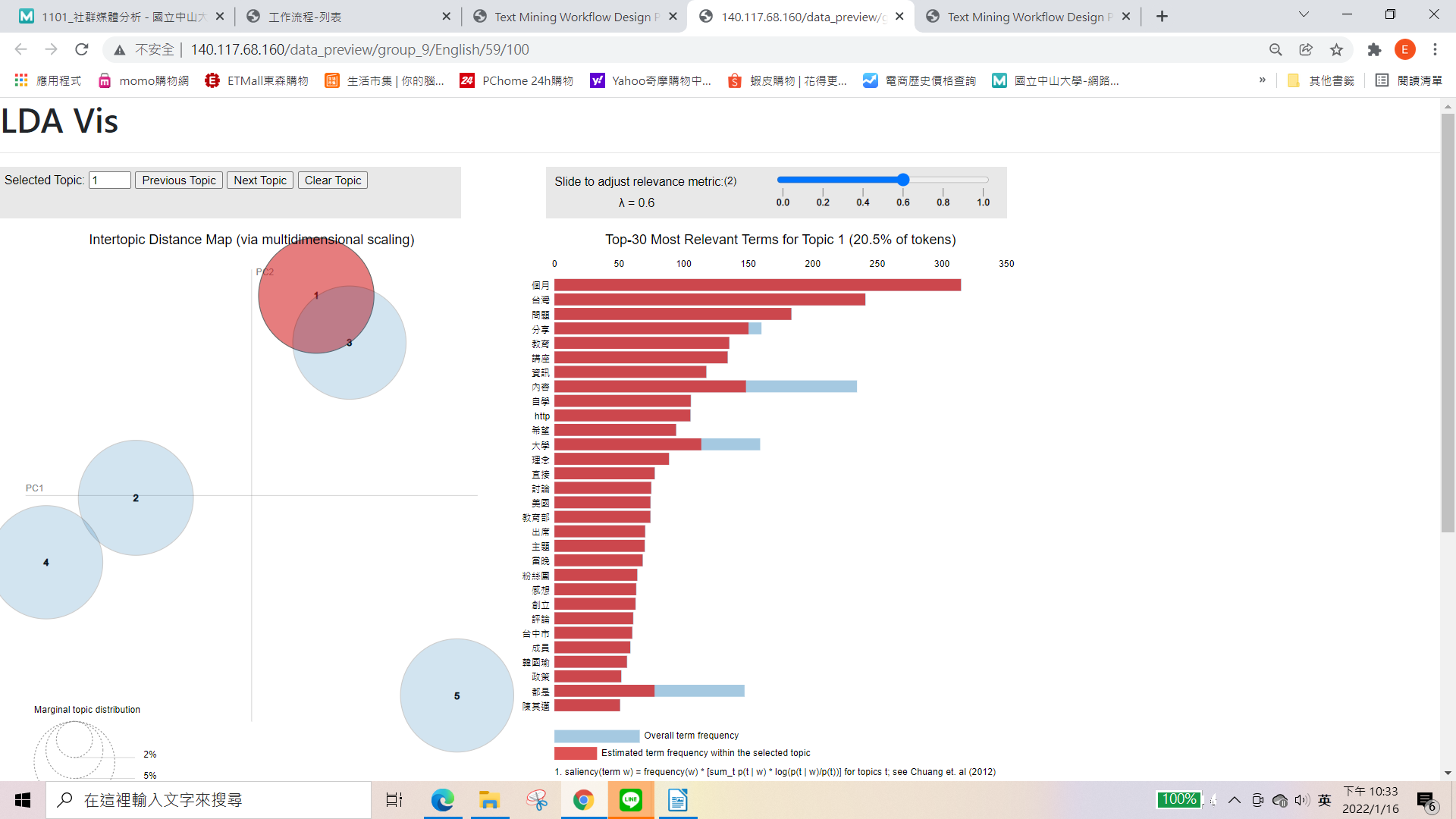
自動產生的描述

* 一張含有 文字 的圖片

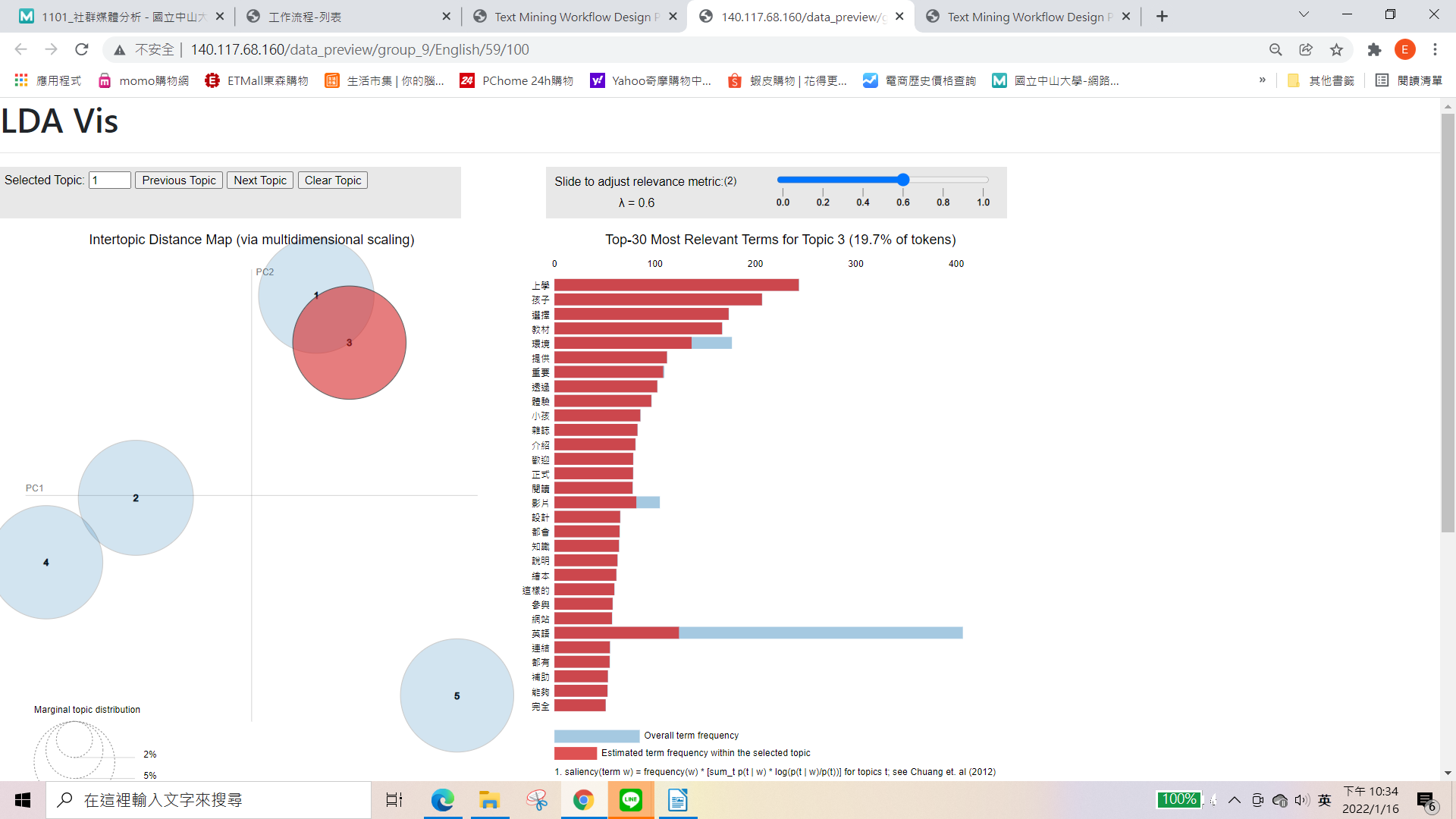
  自動產生的描述關聯式文字雲
* LDA主題模型

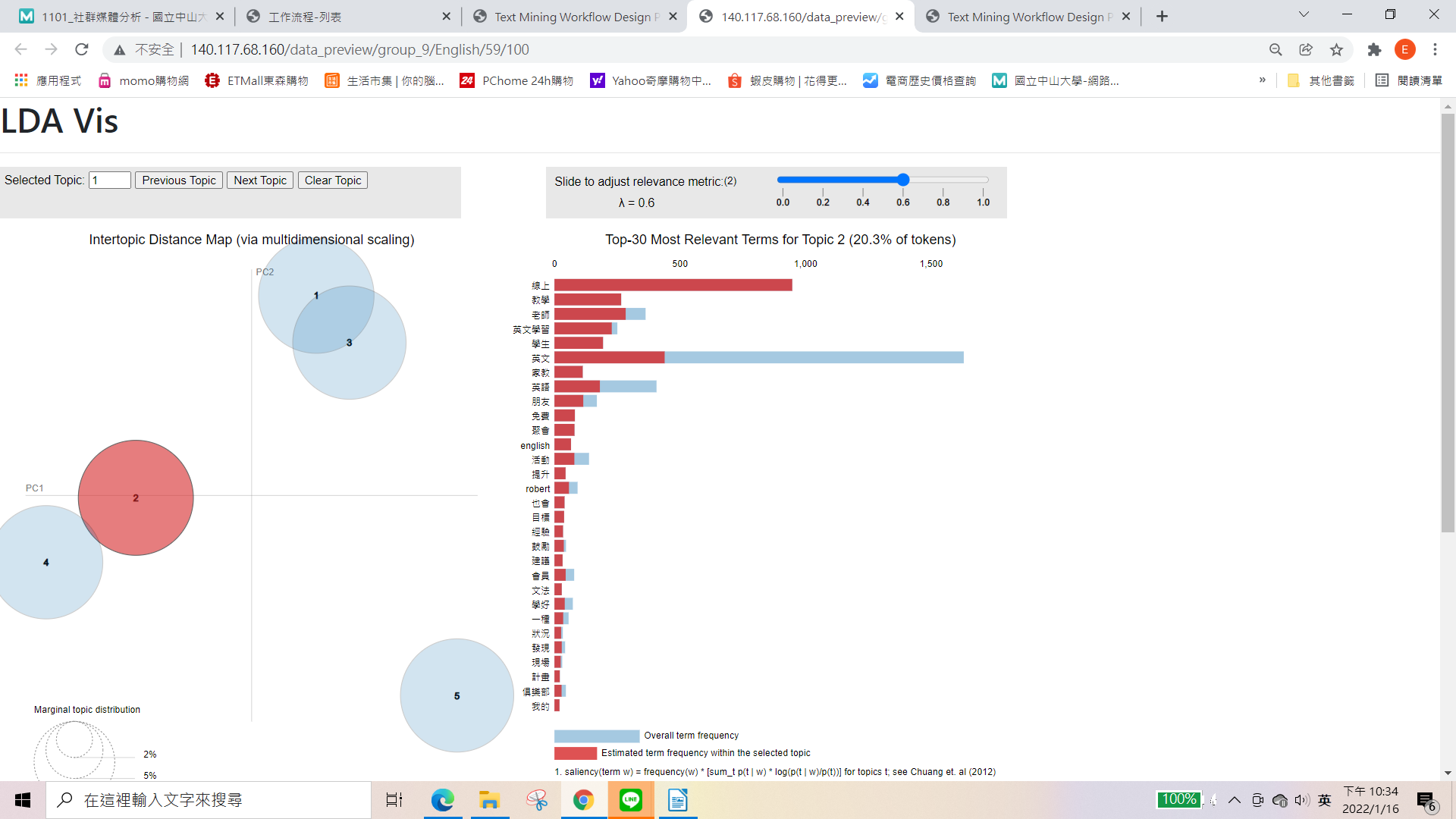
主題分群有五組: 線上課程學習，語言學習環境，英文考試，英文教材，英文社群媒體

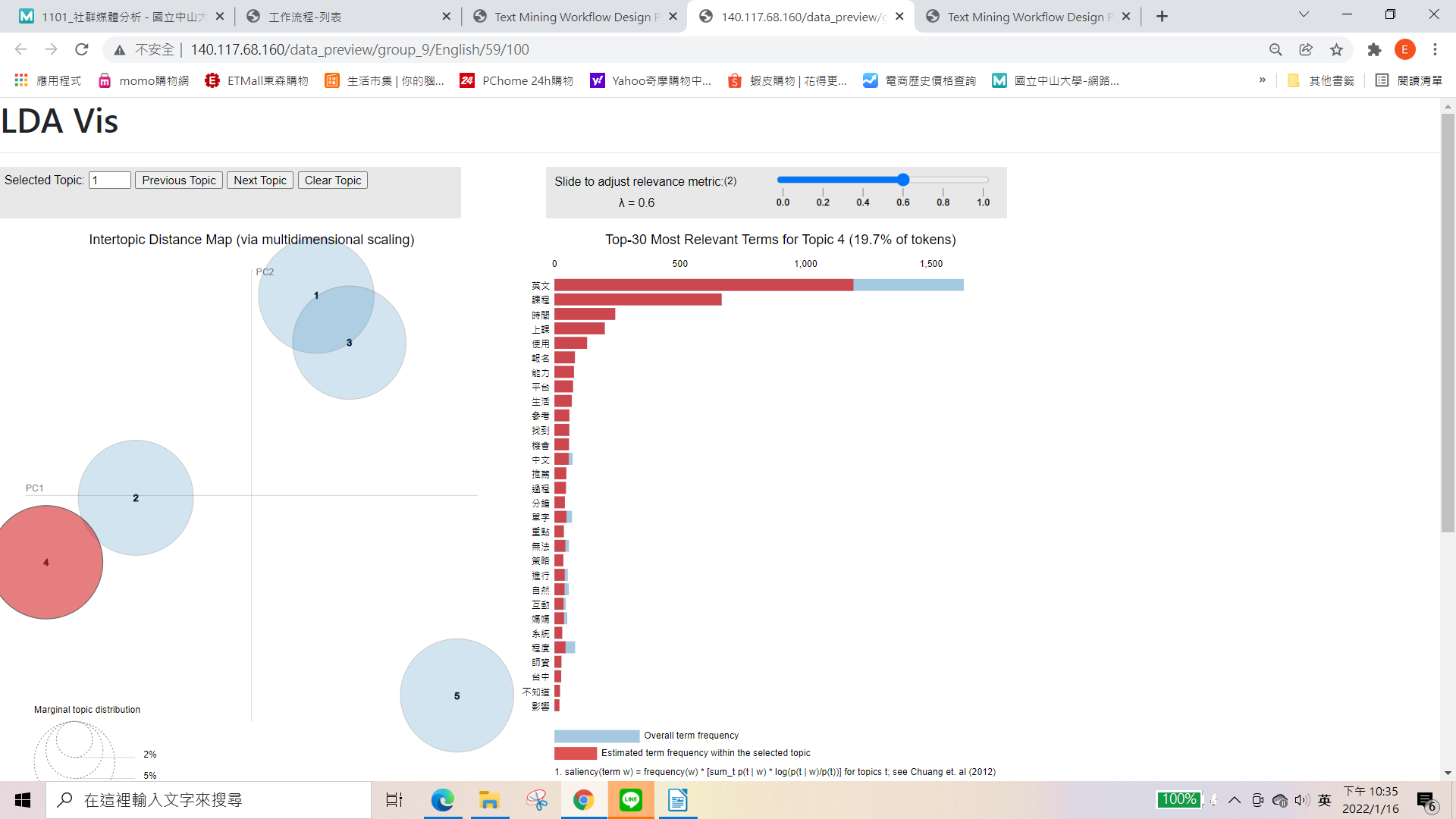
1. 主題一：與「線上課程學習」



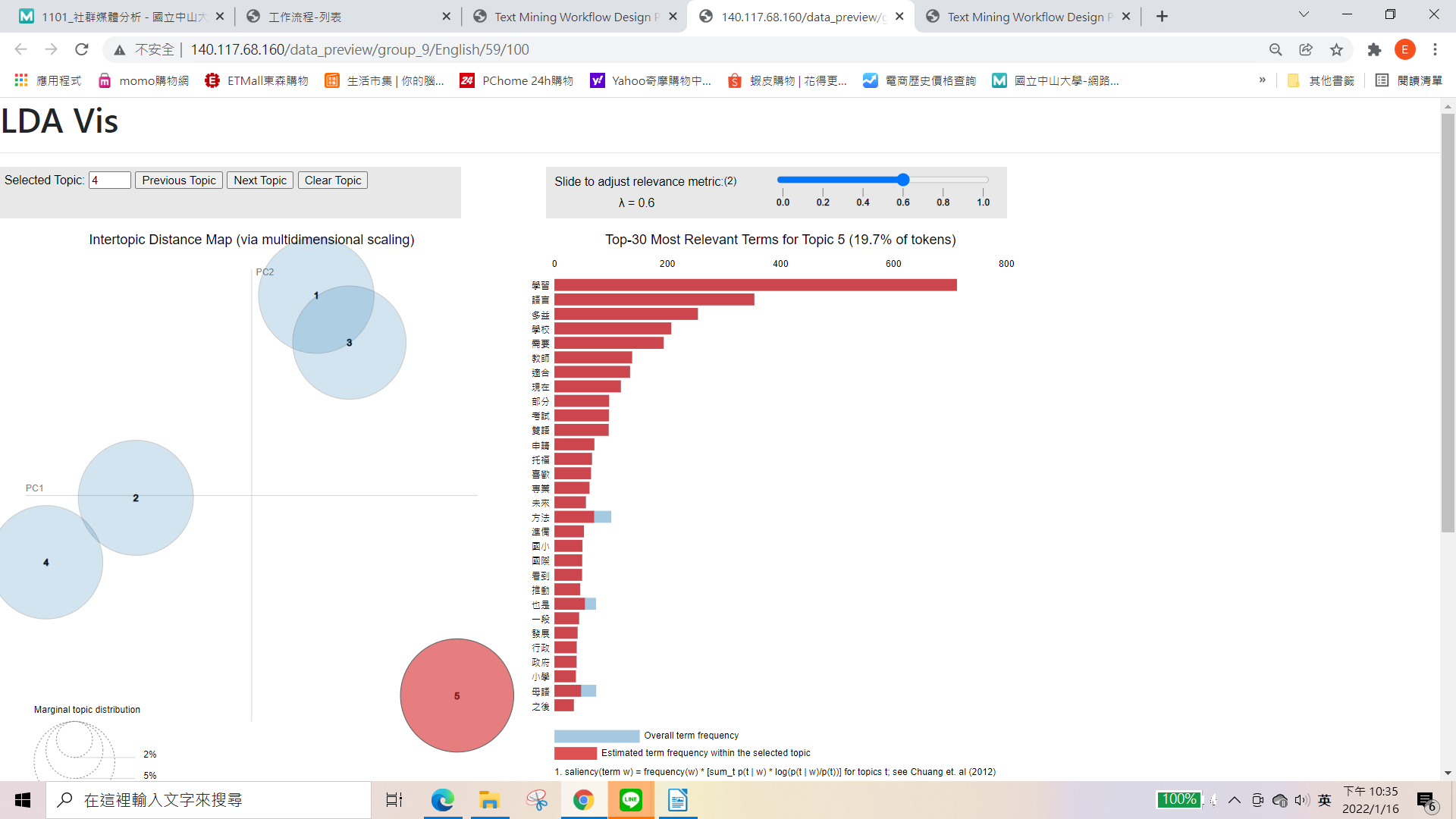
1. 主題二：與「語言學習環境」

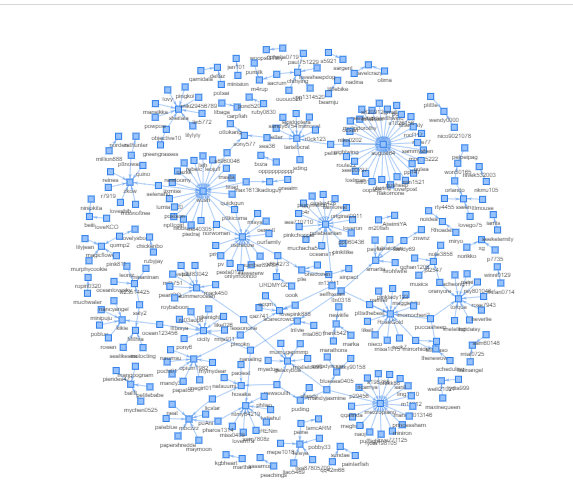


1. 主題三：與「英文考試」
2. 主題四：與「英文教材」



1. 主題五：與「英文社群媒體」



* 社會網絡圖

有數個發言者(RoseGold、summerocean、mezzopiano等)所撰寫的文章在語文學習主題獲得共鳴，並吸引其他留言者至該文章進行內容回覆、留言。

1. **資料分析結果及心得**

此次使用工作流程系統所提供的資料集進行相關分析，由於無法從資料集區分出使用者的年齡層、性別、居住地區以及所屬討論族群，故較無法針對品牌需求進行探討。但從工作流程系統所分析的資料內容，可以歸納出以下結論：

【新冠肺炎疫情】

* 近年來受到新冠肺炎疫情的影響，多所學校為了防疫考量並降低學生染疫風險，全面採取遠距教學，讓學生在家進行線上學習的同時也能確保其安危。由於學生待在家的時間比平時多出許多，造成線上學習課程的需求急速增加。
* 線上英語學習的品質是近年來討論重點，以分析結果可以得知網友進行線上學習時，比起自己使用英文平台學習英文的模式，更偏向與老師進行真人對談，藉此營造出全英文的學習環境。

【國際趨勢】

* 受到網際網路影響，國家與國家之間的關係更加緊密。透過全球緊密的貿易連結，一個國家的經濟發展往往可以牽動到許多國家的經濟產業。為了與其他國家進行交流，國與國之間所能夠使用的共通語言顯得格外重要。所以教育部藉由推動雙語教育來提升學生的語言能力，學校對於語言的要求，也連帶影響家長對於學校師資、英語學習環境的要求。
* 能夠證明其英文能力的語言考試包括多益、托福...等，亦成為網路討論重點。許多人希望自己的英文學習有良好的成效，所以能夠幫助自身順利通過語言考試的英語學習平台，也成為網友選擇使用平台的原因之一。

綜合上述分析結果，初步推論品牌要架設的英文學習平台，需具備下述論點：

* 考量使用者自主學習的時間需配合其個人需求，推出任何時段都可學習的課程內容。
* 依據學習族群不同的學習目標，提供相對應的推薦。例如: 英文檢定相關證照、出國留學、工作需求等。
* 透過虛擬實境應用，在情境模擬上更加寫實逼真，並藉此運用於不同情境上的語言學習和互動體驗，以提升學習效果和趣味性。
* 配合教育部的雙語教育、課綱，推出相對應的英語學習內容，滿足家長與學童的需求。