THE ISSUE OF

一張含有 字型, 文字, 書法, 筆跡 的圖片

自動產生的描述霸凌



指導老師:黃三益教授

第12 組 期末報告

N104020027 林欣穎

N104020005 李姿儀

N104020024 林品均

N104020023 林義行

Agenda :

1. 議題簡介
2. 工作流程
3. 資料清理
4. 文章主題分析
5. 社會網路圖
6. 視覺化工具Gephi應用
7. 結論

一、議題簡介

**時間：**

　　以近三個月時間作為主要分析對象。

**背景與動機：**

　　霸凌是社會存在已久的議題，對於被霸凌者來說這可能是一個難以被抹去的傷痕，韓國心輔書籍《以為長大就會好了 》寫道：

幸運的人用童年治癒一生, 不幸的人用一生治癒童年；

不是所有的痛，時間久了就會痊癒；

不是所有的記憶，時間過去了就會遺忘。

霸凌不只對受害者當下有深刻身心靈的傷害，對於社會本身也是一種重大的挑戰，近日非常火紅Netflex原創韓劇-黑暗榮耀，就是講述著校園中霸凌的故事，這樣的真實案例在日常生活中是層出不窮的，霸凌本身包括言語及身體上的侮辱、攻擊或威脅等行為，這些行為所帶來的傷害，往往造成受害者自尊心創傷，心理上的痛苦，嚴重者甚至導致憂鬱、自殺等極端行為產生，日後更可能衍生許多社會問題，本組在做議題選擇時，對於黑暗榮耀劇情當中霸凌帶來的傷害特別感到難受，便選擇其作為本次報告的主題，我們相信文字具有強烈的力量及意涵，希望藉由文字分析的產出，得到一點對於自己或他人有些啟發的結果。

**分析原因：**

翻拍自韓國校園真實霸凌案例的韓劇 -黑暗榮耀，描述女主角因小時候遭遇校園集體霸凌，從而踏上復仇的人生道路，在戲劇播出後即引發各界熱烈討論，近來幾件關於霸凌的社會案件也引起大量民眾的關注，不限於校園而包含職場、軍旅等都可能有霸凌事件發生，我們認為分析民眾在網路上對於霸凌相關議題的討論，能夠瞭解大眾對於霸凌的想法與見解，且有其作為參考的可能性。

分析期間：

2023/03/01~2023/05/31，共796筆資料。

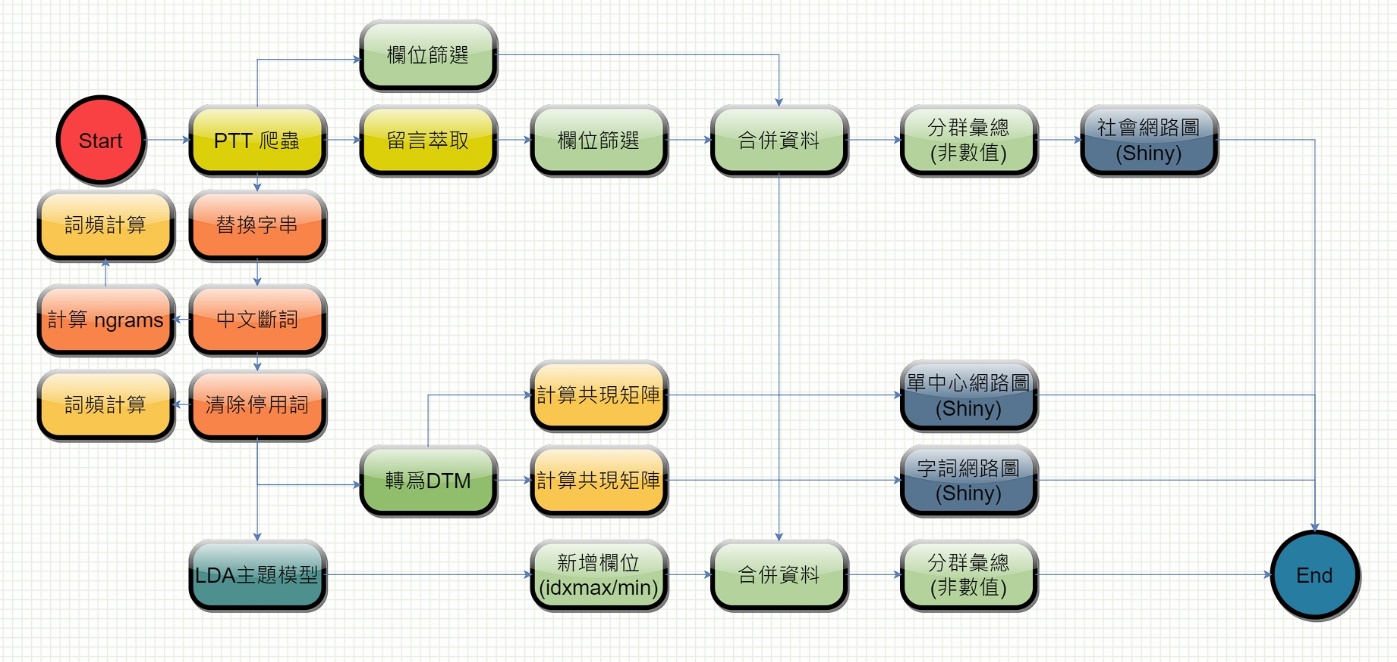
資料來源：

PTT爬蟲>>八卦版、工作版、國中版、軍旅版、憂鬱版、精神疾病、職場版、高中版

使用工具：

文字分析工作流程系統

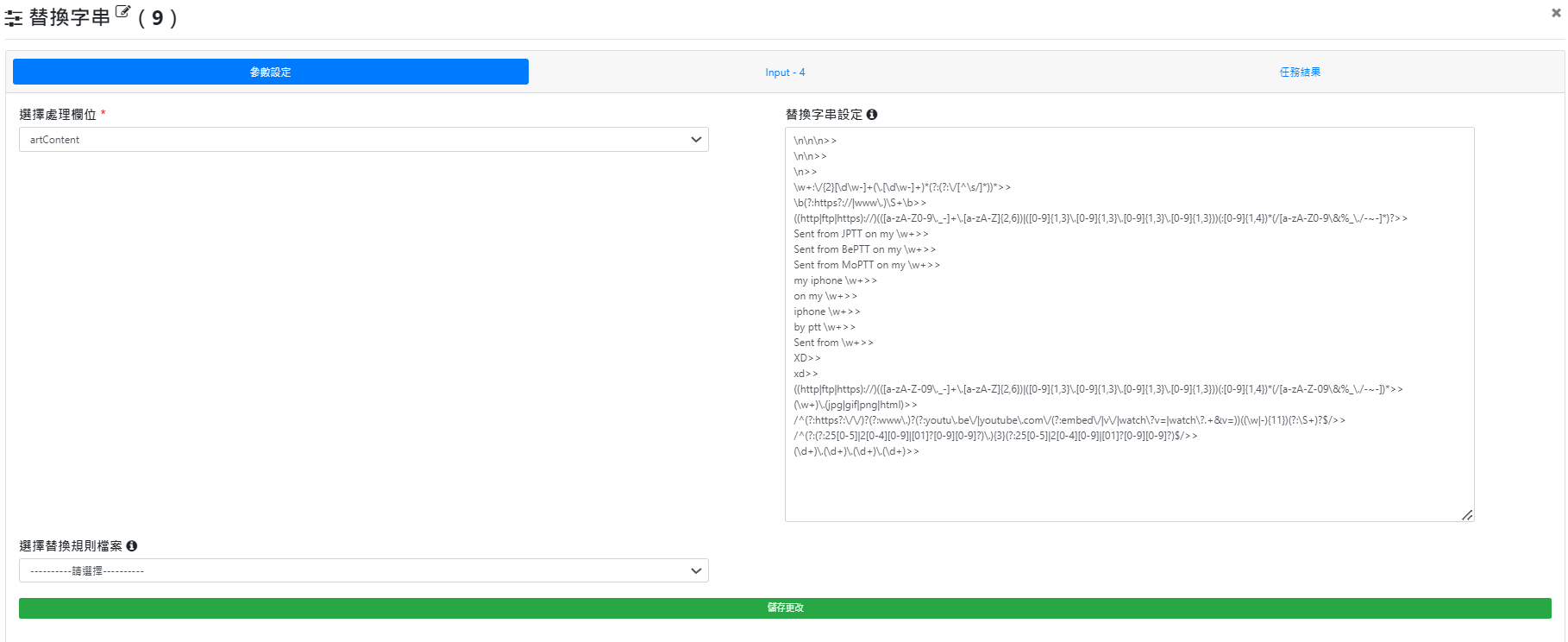
二、工作流程

圖一、本次project工作流程圖

三、資料清理

(一) 替換字串

以「artContent」欄位為標的進行替換字串，主要清除文章內容裡面的顯示的發文裝置類型、網址、HTML符號等。

圖二、替換字串設定

一張含有 文字, 螢幕擷取畫面, 軟體, 字型 的圖片

自動產生的描述

圖三、替換字串結果

(二)中文斷詞

以「result」欄位為標的，將句子切割成適當詞語，主要設想以霸凌、學校、職場、新聞、校園霸凌等相關領域為主，定義霸凌相關重要詞彙(關鍵字)同時設定權重，如網路霸凌1000、言語霸凌1000、體罰500等。

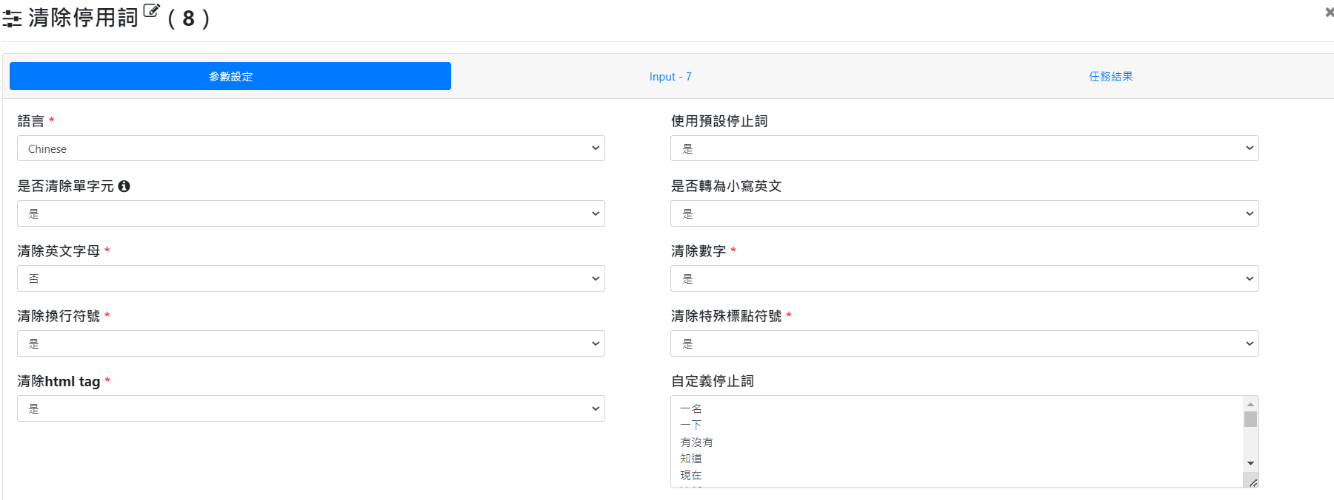
圖四、中文斷詞設定

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述圖五、中文斷詞結果

(三)清除停用詞

使用清除停用字詞設定，將不具分析意義的詞彙列出清除，過濾掉無意義或非主要分析對象詞語的詞彙，且清除單字元、換行符號、html tag、數字、特殊標點符號，並將英文大寫轉為小寫。



圖六、清除停用詞設定

一張含有 文字, 螢幕擷取畫面, 數字, 字型 的圖片

自動產生的描述

圖七、清除停用詞結果

(四)詞頻分析

討論度較高的字詞，即為相對重要的參數。同時也對應並改善我們定義的字典詞彙，以下我們選取出現頻率最高的前200個詞組來產生文字雲。依據文字雲結果，關於霸凌議題的題及詞語中，以「霸凌」最多，其次為「學生」，再者為「新聞」，學校、老師、媒體等字詞的頻率也不相上下，表示大家對於霸凌議題的可能較常關注於校園相關(校園霸凌、學生、學校)、訊息來源(新聞、媒體)等方面的資訊。

一張含有 文字, 螢幕擷取畫面, 數字, 收據 的圖片

自動產生的描述圖八、詞頻分析結果

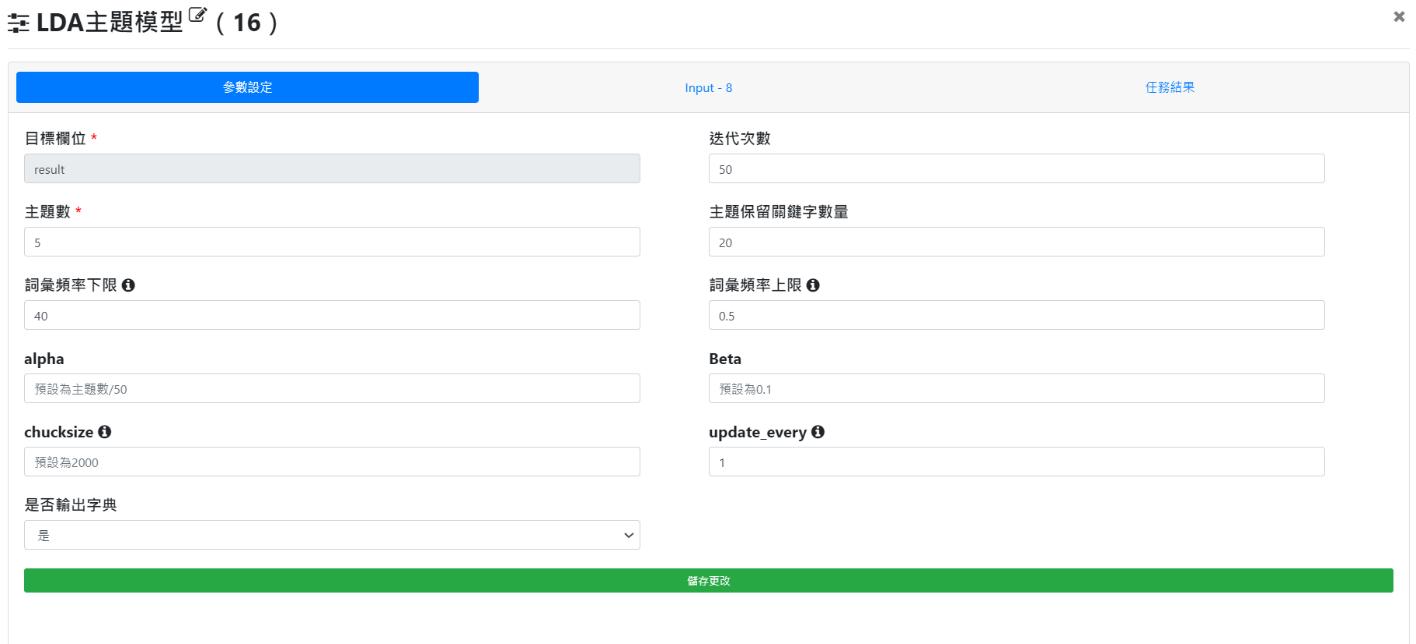
一張含有 文字, 字型, 圖表, 設計 的圖片

自動產生的描述

圖九、出現頻率最高的前200個字詞之文字雲

四、文章主題分析

運用LDA主題模型我們將主題數的參數設定為5，透過發文者的發問來了解所討論的前5大主題是什麼?



圖十、LDA主題模型參數設定



圖十一、LDA主題模型執行結果

(一) 第1主題:為「與台中高中校園霸凌輕生案相關議題」

一張含有 文字, 螢幕擷取畫面, 圖表, 圓形 的圖片

自動產生的描述

(二) 第2主題:為「各國家間相關議題的討論」

一張含有 文字, 螢幕擷取畫面, 圖表, 圓形 的圖片

自動產生的描述

(三) 第3主題；為「霸凌案於民代、執政黨等相關報導」

一張含有 文字, 螢幕擷取畫面, 圖表, 圓形 的圖片

自動產生的描述

(四) 第4主題；為「與校園霸凌相關各關係人之議題」

一張含有 文字, 螢幕擷取畫面, 圖表, 圓形 的圖片

自動產生的描述

(五) 第5主題；為「媒體對於相關案件的後續相關報導」

一張含有 文字, 螢幕擷取畫面, 圖表, 圓形 的圖片

自動產生的描述

五、社會網路圖

利用ptt爬蟲功能找出發文者、發文內容，再用留言萃取找出留言者、留言內容後合併資料並分群匯總，保留發文者、留言者，以及留言者共回覆幾則留言，再利用shiny社會網路圖，看出發文者、文章留言者及留言數量關係。

一張含有 文字, 字型, 行, 數字 的圖片

自動產生的描述

有多位特定發文者擁有自己的討論圈。

一張含有 文字, 螢幕擷取畫面, 圖表, 設計 的圖片

自動產生的描述

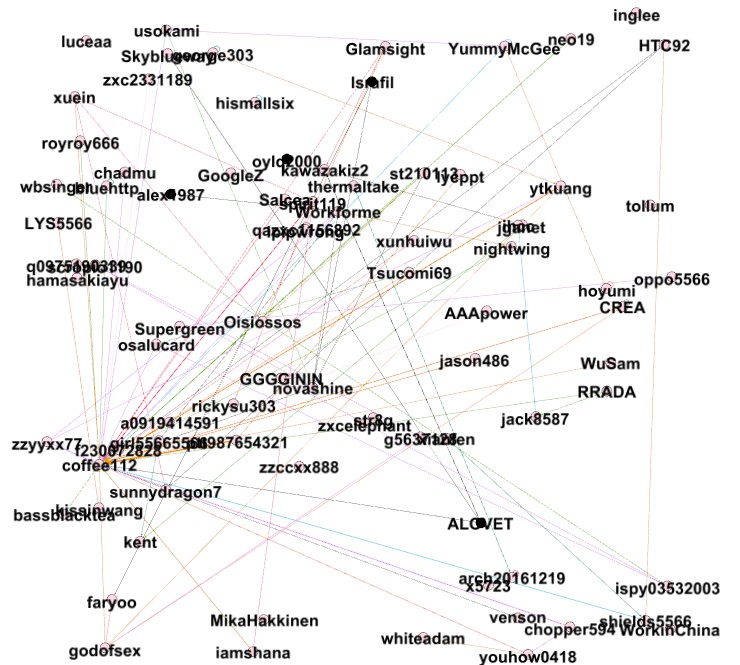
anly888這位回文者，除了踴躍回覆 spirit119的文章外，也參與好幾個討論度高的發文者。

一張含有 文字, 螢幕擷取畫面, 圖表 的圖片

自動產生的描述

六、視覺化工具Gephi應用

透過LDA主題模型分析後進行分群彙總，針對PO文者和回文者，在各主題之間的回文數量，透過Gephi的繪圖軟體進行繪製。



七、結論

(一)、 在初步探索文章的主題分類上，LDA是一個良好的工具，可以幫助我們尋找可分析的主題與詞彙。本次的資料來源由原本的八卦版，再增加了國中版、軍旅版、憂鬱版、精神疾病版、職場版、高中版這六個看板，欲從這些可能會與霸凌相關的看板做更深入的了解。

(二)、本次以五個主題作探討，從這些主題中可見，霸凌這個社會議題充斥於校園、職場及軍中，在進行霸凌防制的相關宣導時，因可先行了解以上場域的霸凌行為成因並提出解決策略，這可以提供教育主管機關、勞工、國軍相關申訴單位霸凌防治處理之參考依據。