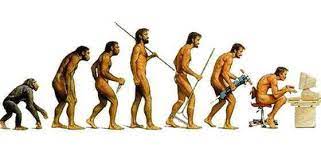
**社群媒體分析**

**讀書會第二次報告**

**第八組**

**分析主題：**

**沸沸揚揚的狒狒出逃事件**

**指導老師：黃三益教授**

組長：N104020023林義行

組員：N104020018謝明和

N104020020李宜臻

N104020021馮慧嬌

N104020022凌嫈琪

N104020024林品均

N104020025陳姿樺

目錄

[一、議題簡介 3](#_Toc131972584)

[（一）狒狒逃脫事件介紹 3](#_Toc131972585)

[（二）分析目的： 4](#_Toc131972586)

[二、工作流程 5](#_Toc131972587)

[三、資料清理 7](#_Toc131972588)

[（一）PTT爬蟲 7](#_Toc131972589)

[（二）替換字串： 9](#_Toc131972590)

[（三）中文斷詞 11](#_Toc131972591)

[（四）清除停用詞 12](#_Toc131972592)

[（五）詞頻計算 14](#_Toc131972593)

[（六）計算ngrams 15](#_Toc131972594)

[四、進階情緒分析 16](#_Toc131972595)

[（一）Corpus Based情緒分析 16](#_Toc131972596)

[（二）NTUSD情緒分析 17](#_Toc131972597)

[（三）Chinese Default情析分析 19](#_Toc131972598)

[五、視覺化圖表分析 21](#_Toc131972599)

[（一）視覺化儀表板-議題討論數量折線圖 21](#_Toc131972600)

[（二） 視覺化儀表板-詞頻計算 22](#_Toc131972601)

[（三） 視覺化儀表板-NTUSD正向、負向分群匯總 24](#_Toc131972602)

[（四） 視覺化儀表板-Chinese Default正向、負向分群匯總 27](#_Toc131972603)

[（五） 視覺化儀表板-Corpus Based分群匯總 28](#_Toc131972604)

[六、結論 30](#_Toc131972605)

# 一、議題簡介

## （一）狒狒逃脫事件介紹

1. 事件始末：

首先是桃園的一名里長3/22於臉書上宣稱於3月10日在路上看見東非狒狒，經當地居民圍捕兩次，皆以失敗做收，並該狒狒已離開轄區，故無繼續追捕的行動。東非狒狒生性兇猛，熱心民眾們希望各地能提供線索，協助迅速將脫逃的狒狒帶回（該狒狒為六福村動物園飼育），也因此將整個事件熱度推上高峰，並獲得各家新聞爭相報導，最後於3/27在桃園地區受麻醉後完成圍捕，在帶回後，醫療團隊發現狒狒極度虛弱，並在不久後死亡，相驗死因發現並非麻醉槍致死，狒狒於被捕前就已被獵槍射傷，並因未及時救治傷重身亡。

1. 案件關係人：
2. 狒狒： 於科學分類裡面屬於動物界-脊索動物門-哺乳綱-靈長目-猴科-狒狒屬-東非狒狒種，根據本草綱目紀載，中國古代曾稱之為梟羊、野人、人熊等等。
3. 林姓獵人：為林務局外來種移除的獵人，看到新聞前往，表示當時聽到以為是農業局的人說開槍才開槍，後來才知道該名人員是六福村的獸醫。
4. 縣政府：由農業處技士與林姓獵人聯繫，然而縣府未派遣林姓獵人。
5. 農業局：說明圍捕使用的是麻醉槍，但在圍捕的過程中似乎並沒有很好的流程及規範，最終導致了狒狒的身故。
6. 六福村：前後說法不一致，狒狒為該園區所飼養，在圍捕過程均攜帶麻醉槍並配合農業局指揮，最後接受違反野生動物法裁罰。

## （二）分析目的：

由於新聞延燒了好幾周，因此我們主要是想了解民眾是否真的關心狒狒出逃事件，抑或僅只是新聞譁眾取寵，由於希望目標是民眾，因此資料來源選擇從PTT八卦版中擷取，日期從事件發生03/22至4/4找出關於狒狒事件的文章，總共2070筆資料，希望能透過文字分析流程系統來分析鄉民對整個事件的參與討論狀況，來進一步探討民眾討論的情緒值，以及熱度分布狀況。

# 二、工作流程

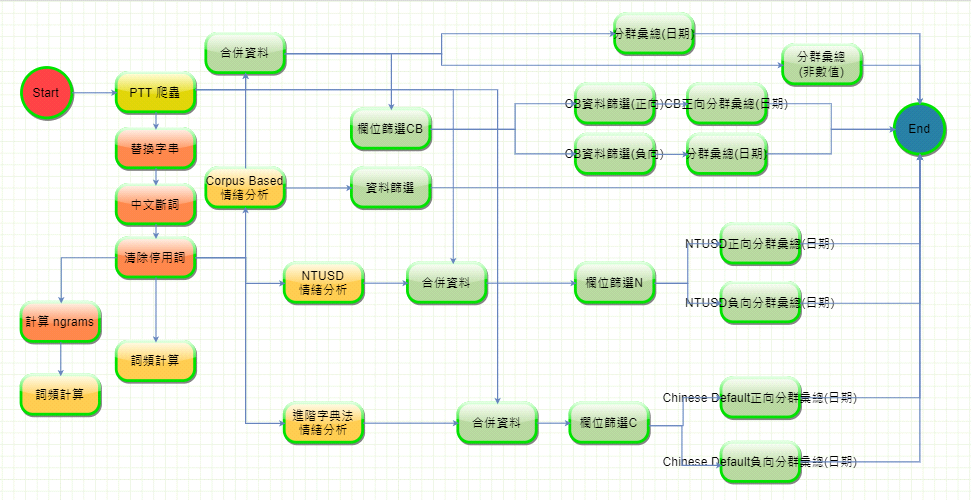
（一）使用平台：文字分析工作流程系統

（二）流程名稱：第二次讀書會

（三）資料來源：PTT八卦版

（四）時間：2023/03/22至4/4

（五）工作流程：



1. 從工作平台中的「PTT（八卦版）爬蟲」讀入資料，選定主題、設定搜尋關鍵字、排除關鍵字。
2. 使用「內容替換字串」將不要的字串替換為空白。
3. 使用「中文斷詞」加大特定詞彙權重。
4. 使用「清除停用字詞」設定，將不具分析意義的詞彙列出清除。
5. 使用「計算ngrams」設定來統計出現機率比較大的字詞組合，並用「詞頻計算」選取出現頻率最高的前200個詞組來產生文字雲。
6. 使用「詞頻計算」， 並用選取出現頻率最高的前200個詞組來產生文字雲。
7. 使用Corpus Based、NTUSD、Chinese Default三個情緒分析字典，瞭解情緒詞彙、正面與負面詞彙的發生情形。
8. 使用「合併資料」合併「PTT（八卦版）爬蟲」與 Corpus Based 情緒分析結果，用「分群彙總（日期）」找出每日的正負情緒值。
9. 使用「合併資料」合併「PTT（八卦版）爬蟲」與Corpus Based 情緒分析結果，用「分群彙總（非數值）」找出每日的每則文章的情緒數值。
10. 使用「合併資料」合併「PTT（八卦版）爬蟲」與Corpus Based情緒分析結果，用「新增欄位」篩選情緒層級為正負數的資料，並使用「分群彙總（日期）」，來彙總每日的正負情緒值。
11. 使用「合併資料」合併「PTT（八卦版）爬蟲」與NTUSD情緒分析結果，用「欄位篩選」分別篩選情緒值為負數及正數的資料，並使用 「分群彙總（日期）」，依日期來彙總每日的正負向情緒值（即每日情緒總和）。
12. 使用「合併資料」合併「PTT（八卦版）爬蟲」與進階情緒字典（Chinese Default）之分析 ，用「欄位篩選」分別篩選情緒值為負數及正數的資料，並使用 「分群彙總（日期）」，依日期來彙總每日的正負向情緒值（即每日情緒總和）。
13. 將工作平台分析資料結果製作成視覺化圖表：議題熱度討論圖、詞頻計算文字雲圖與直條圖、NTUSD正向/負向詞彙分群彙總圖、Chinese Default 正向/負向詞彙分群彙總圖、Corpus Based 正向/負向詞彙分群彙總圖。
14. 最後各段落備註上本組的看法，並將此次分析結果及讀書會心得彙整成結論。

# 三、資料清理

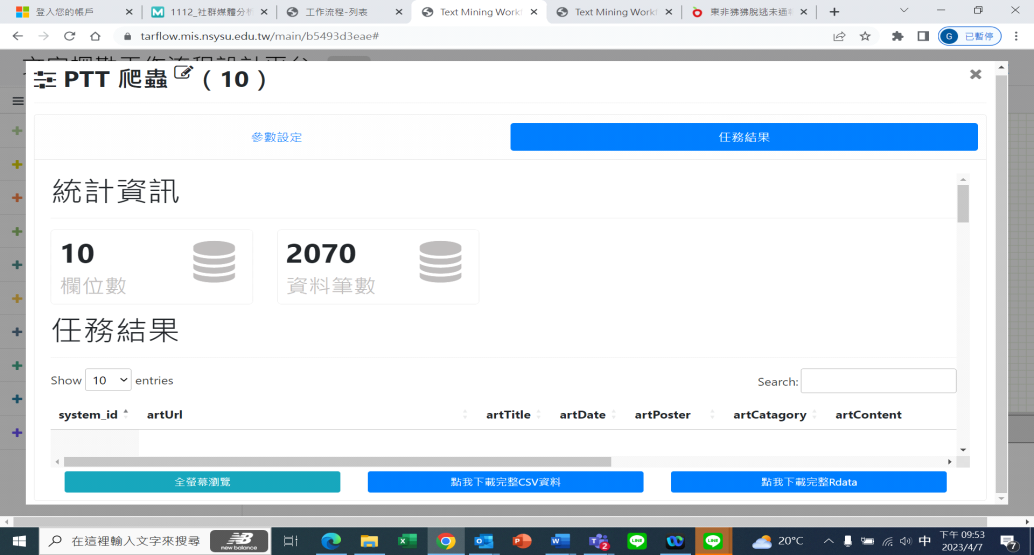
## （一）PTT爬蟲

本組自工作流程設計平台中「資料輸入與輸出」，選擇PPT爬蟲的套裝物件，參數設定選擇gossiping（八卦板），並設定日期為事件開始的第一天2023年3月22日，到2023年4月4日狒狒中槍不治當天，針對狒狒逃亡事件所有鄉民留言內容來做分析。

1. 設定的關鍵字：狒狒
2. 排除關鍵字：由於相關資料較多，為了讓資料更準確地搜尋到東非狒狒，因此設定了排除關鍵字，例如猴子、鄭州富士康、公視、斑馬、馬英九...等。



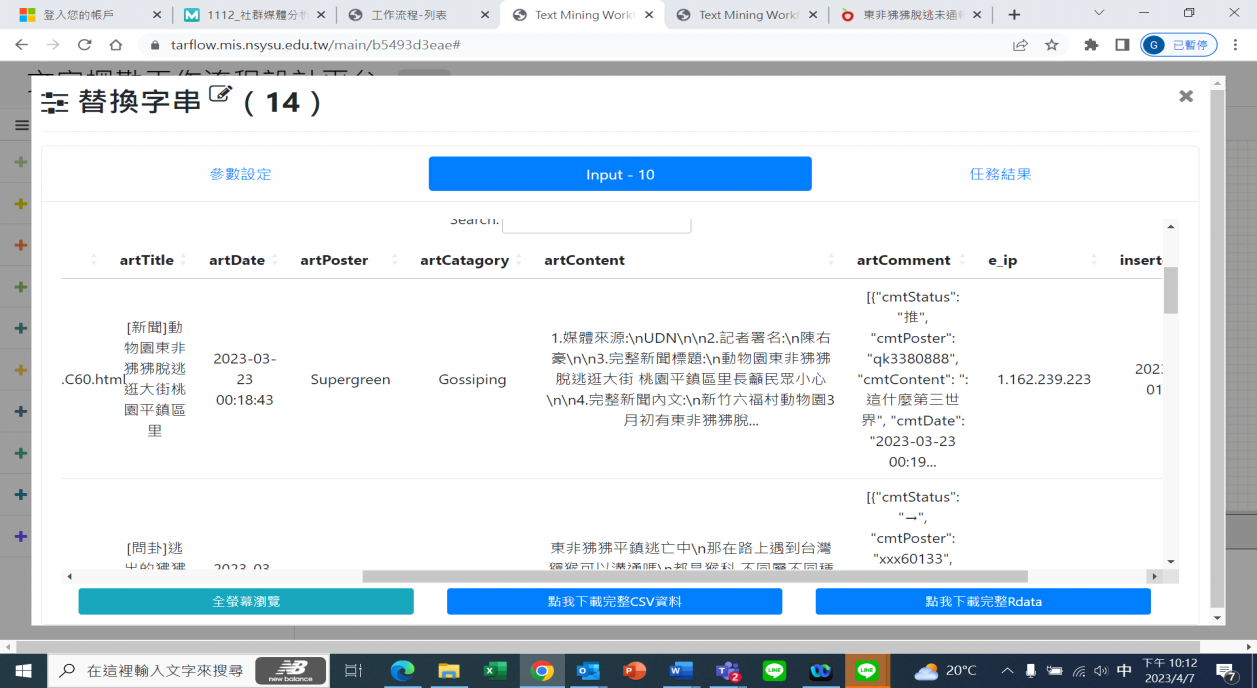
1. 任務結果為2070筆



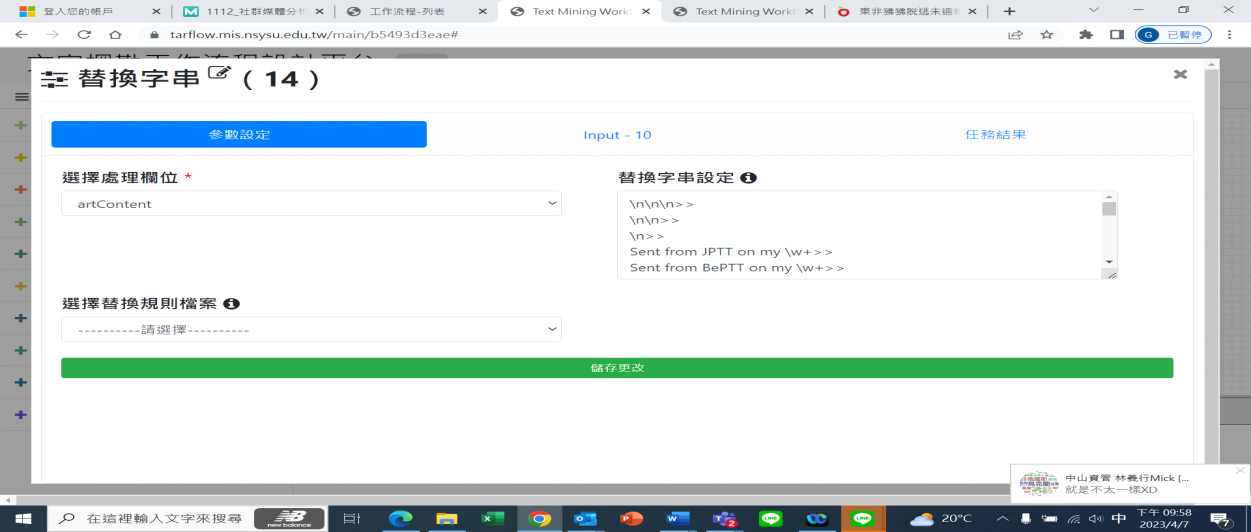
## （二）替換字串：

為了清理鄉民留言資料中無意義或是雜訊，讓搜尋出來的資料比較容易閱讀，我們做了替換字串。 清理前文章內容中有許多奇怪或是沒有意義的符號或是字體。

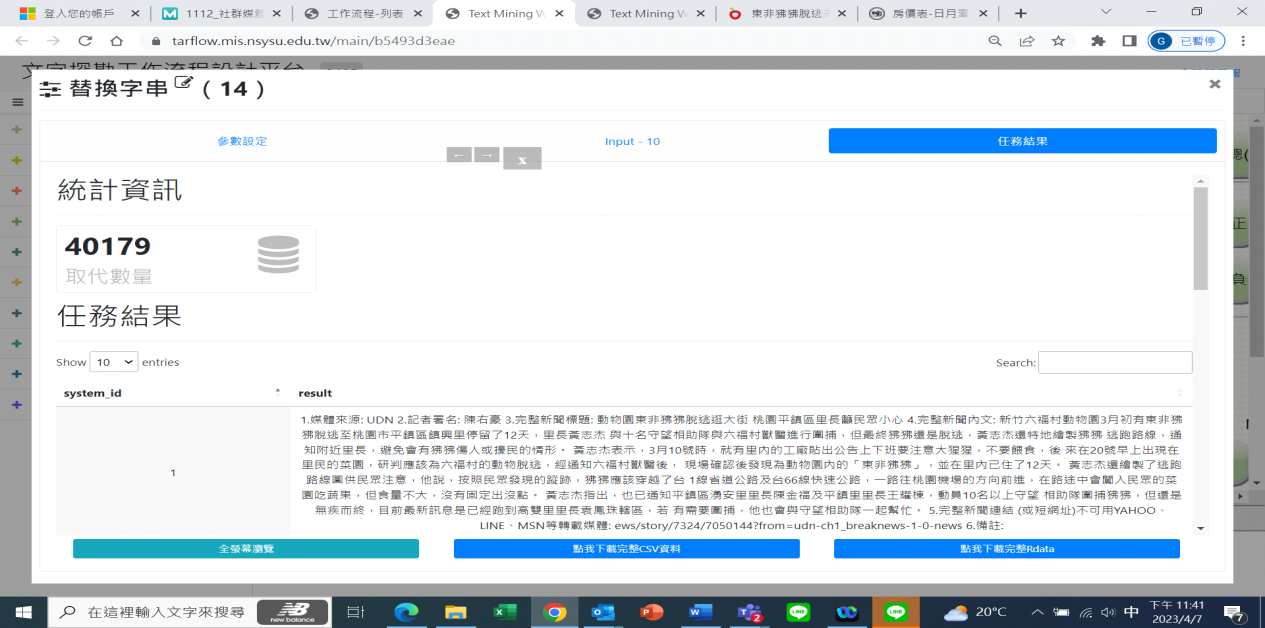
1. 清理前：



1. 替換字串的設定



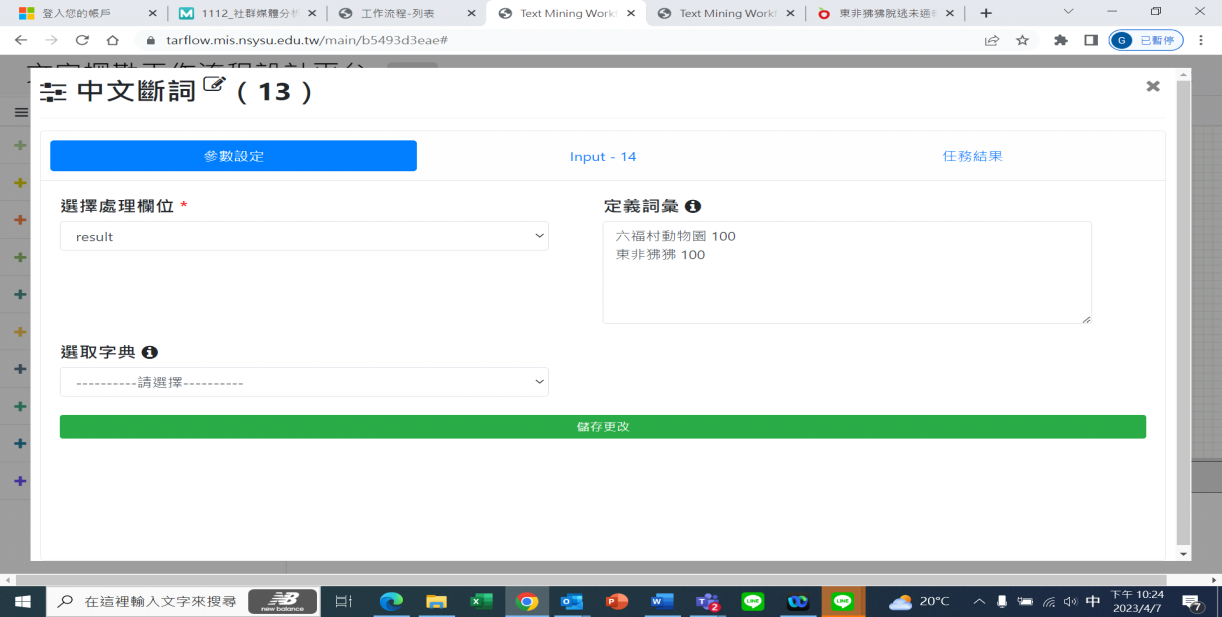
1. 替換後結果



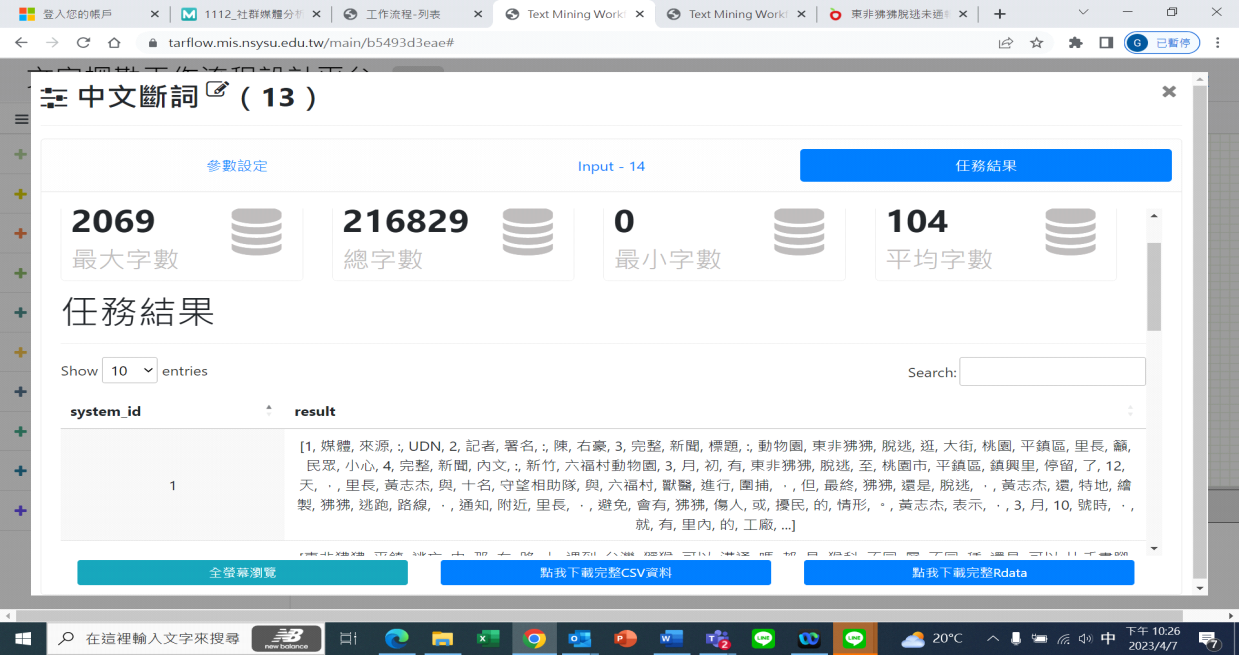
## （三）中文斷詞

依據斷詞所產出的結果，反覆清洗斷詞並修正，將會用在內容的特殊用語做為斷詞，讓分析可以更為準確，例如六福村動物園 100，東非狒狒 100。

1. 定義詞彙：



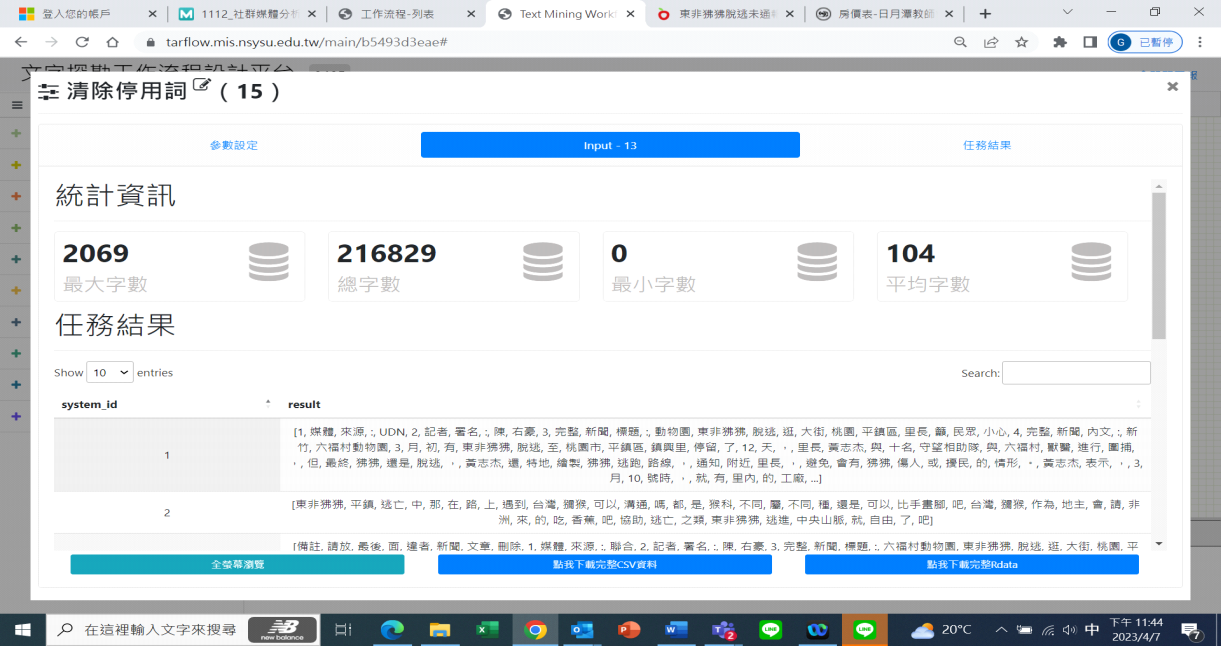
1. 斷詞後：



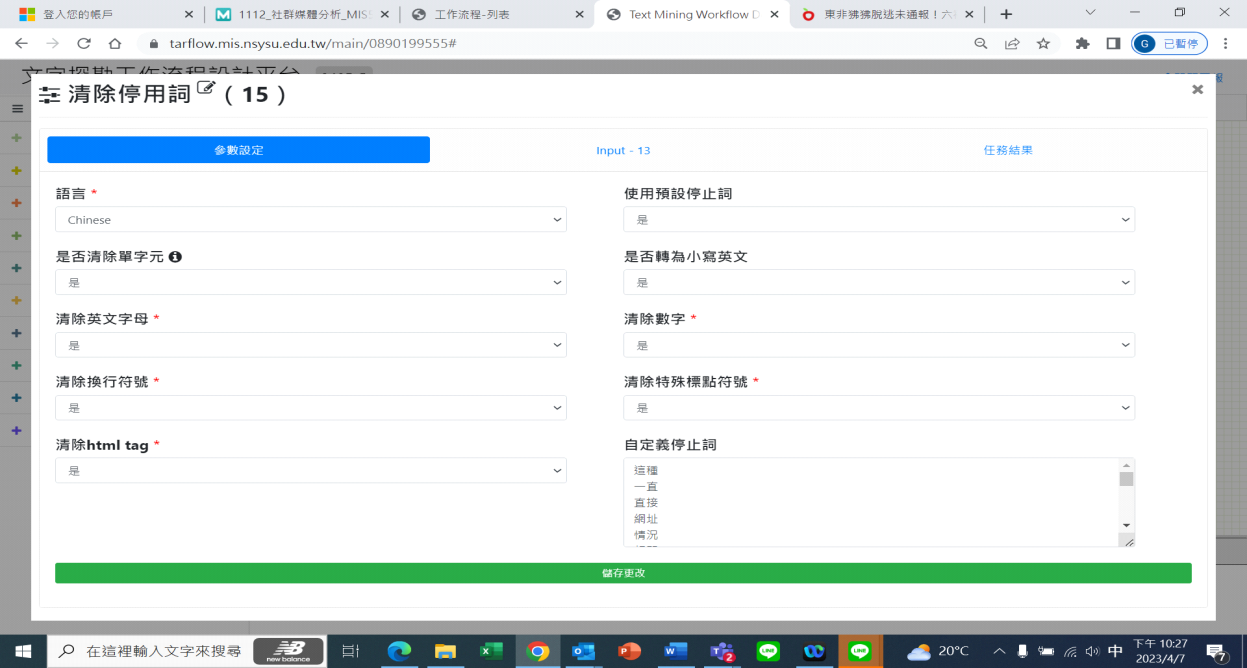
## （四）清除停用詞

本組將單字元、英文字母、換行符號、html tag、數字、特殊符號可能與內容無關的皆去除，並針對文字雲結果分析內容，對不具意義的字詞去除，包含：這種、一直、直接、網址、情況、相關、林姓、標題....等。

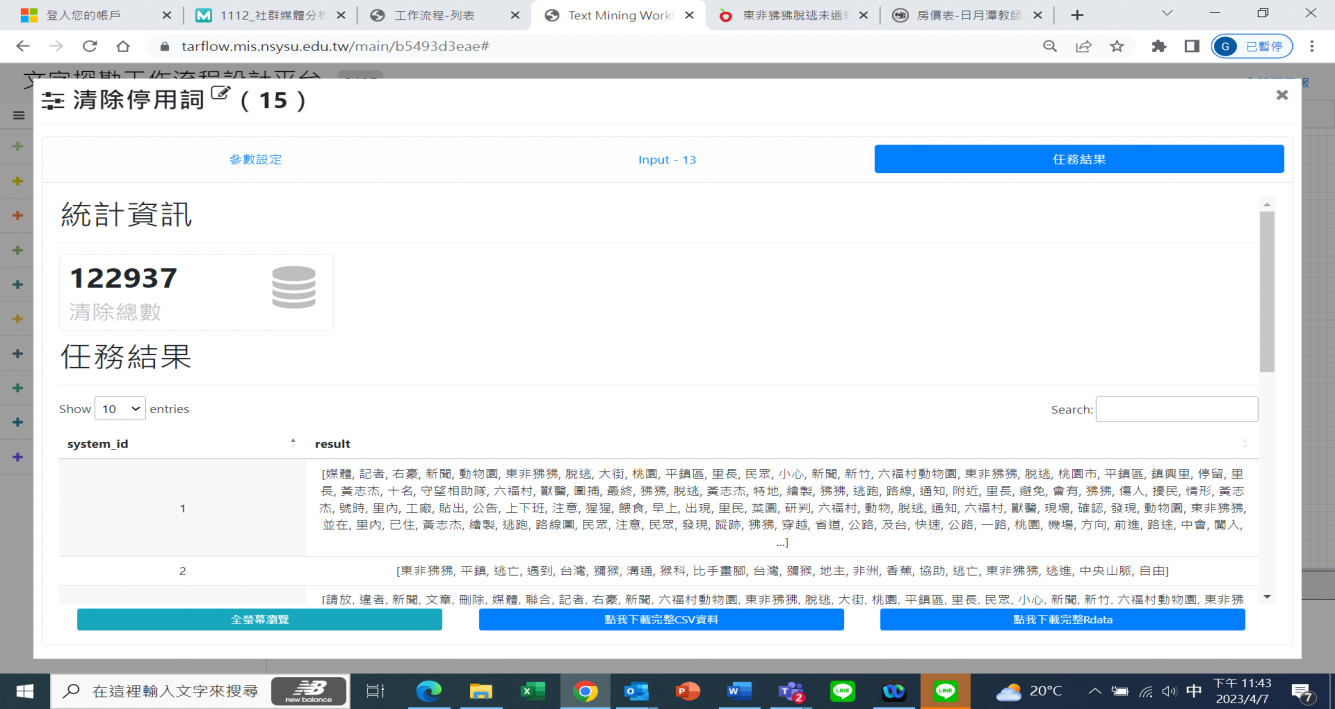
1. 清理前：



1. 清除設定：



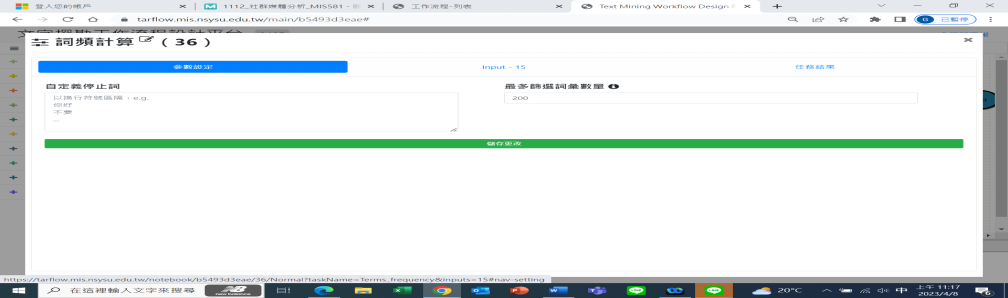
1. 任務結果：



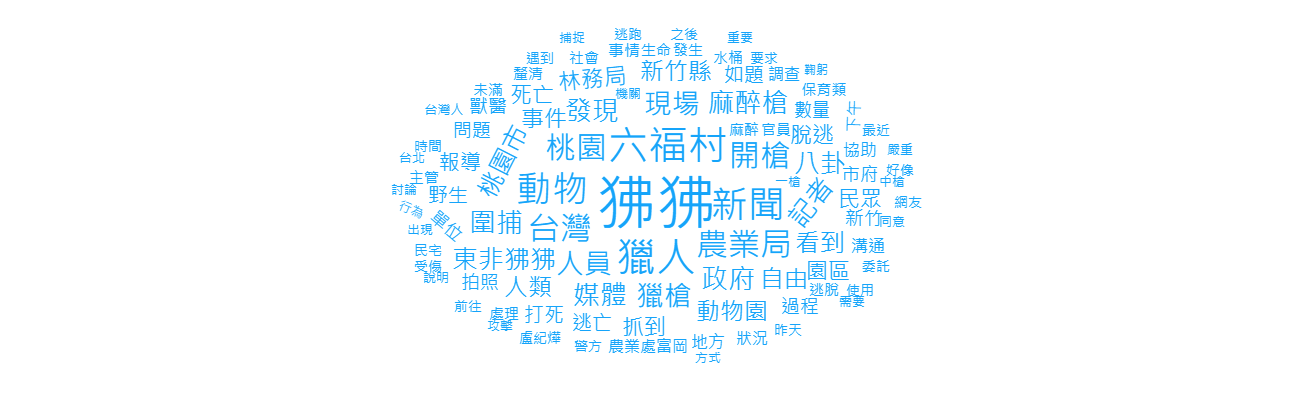
## （五）詞頻計算

為了了解這些鄉民留言或是討論文章所報導的內容與重點，我們設定詞彙數量200，找出討論度最高的200個詞頻。

1. 詞頻設定：



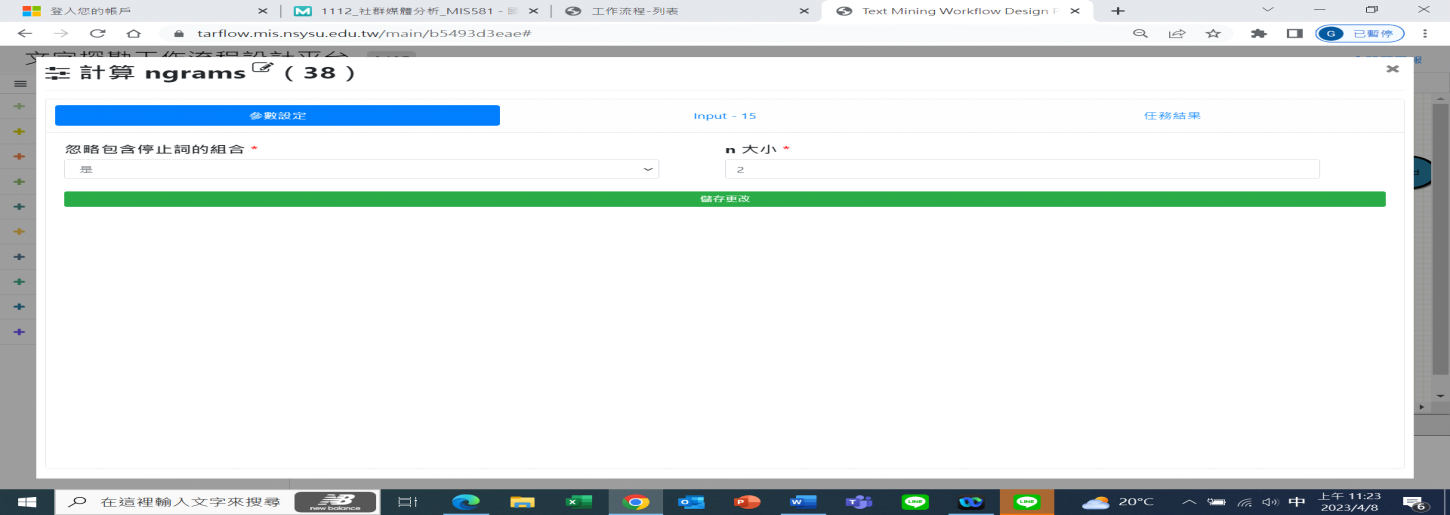
1. 詞頻任務結果如下：



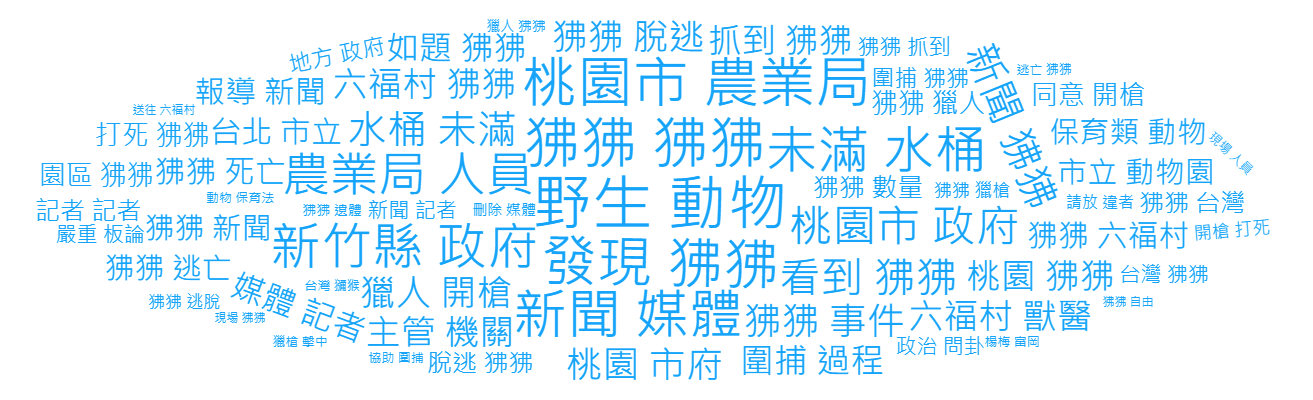
## （六）計算ngrams

透過ngrams，來觀察連接前後的詞，句子出現的機率，藉以看出事件的相關人或是負責單位，作為關鍵字預測。因此我們將設定n=2，是以兩個字（詞）為單位計算的詞頻 。

1. 設定：



1. 設定詞彙數量200，任務結果如下：

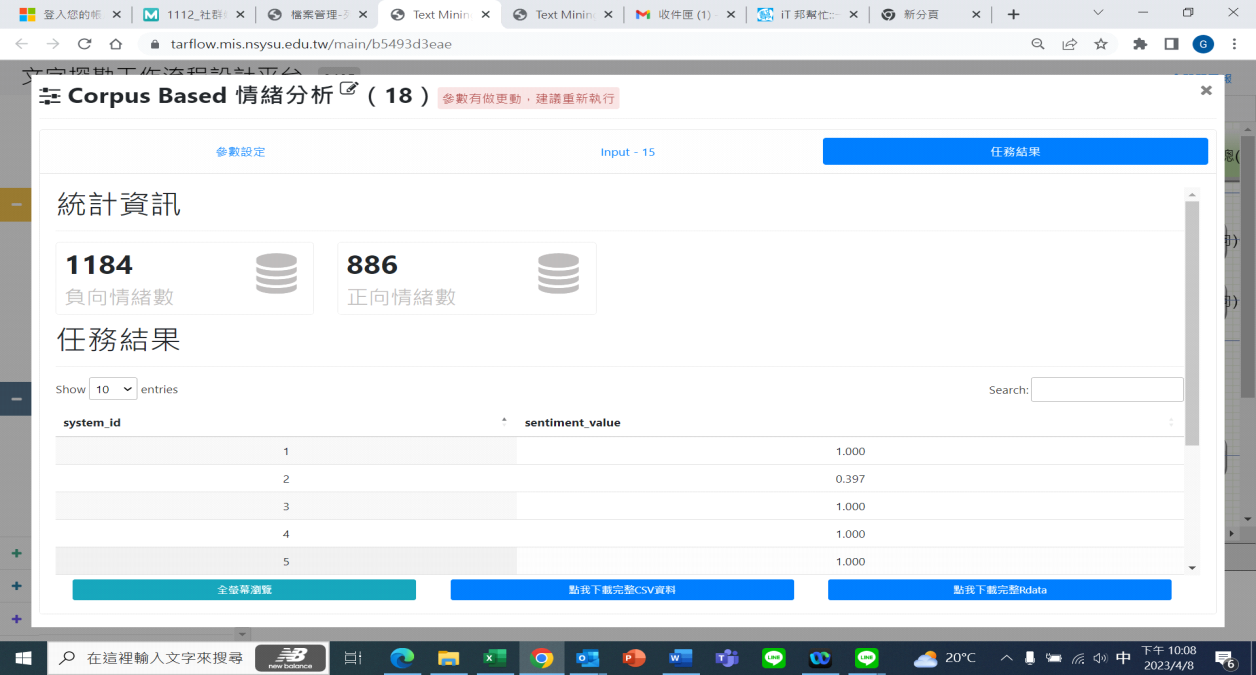


# 四、進階情緒分析

本組依據Corpus Based情緒分析、NTUSD情緒分析與進階字典法（Chinese Default），共三種情緒字典的來分析事件期間PTT八卦版上相關貼文，藉以瞭解東非狒狒逃脫期間PTT鄉民對於本事件的情緒風向。

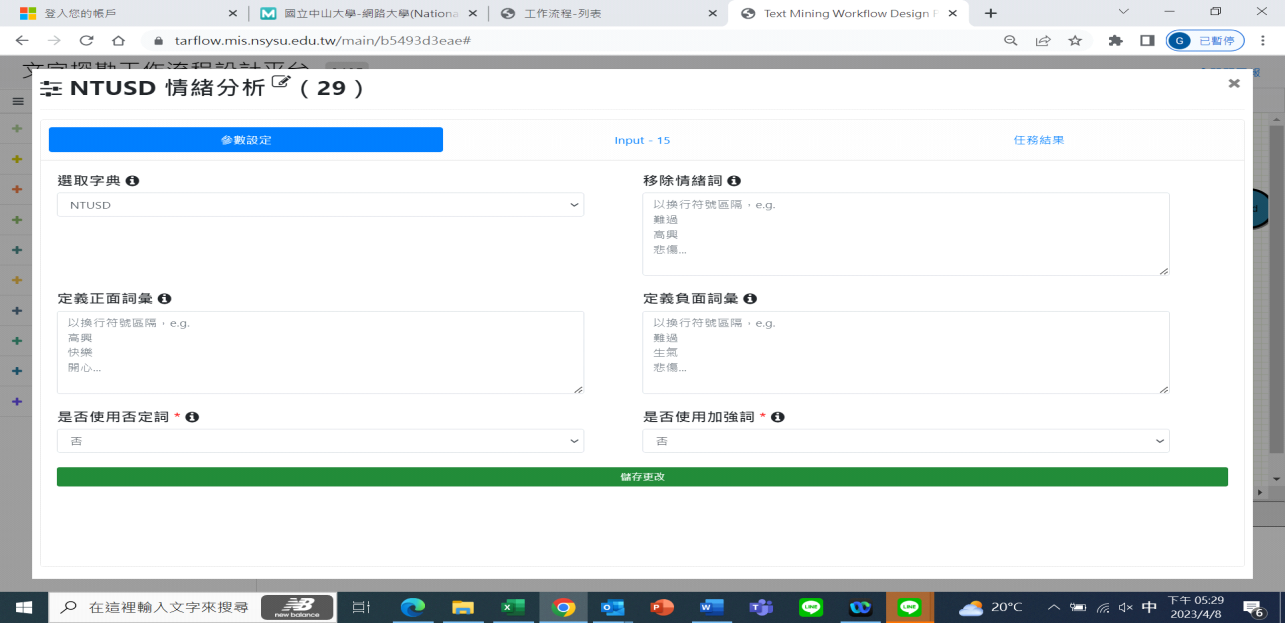
## （一）Corpus Based情緒分析

以模型預設值0.5作為基準，文章內容情緒分析值大於0.5，判定為正向情緒，若低於0.5，則判定為負向情緒。

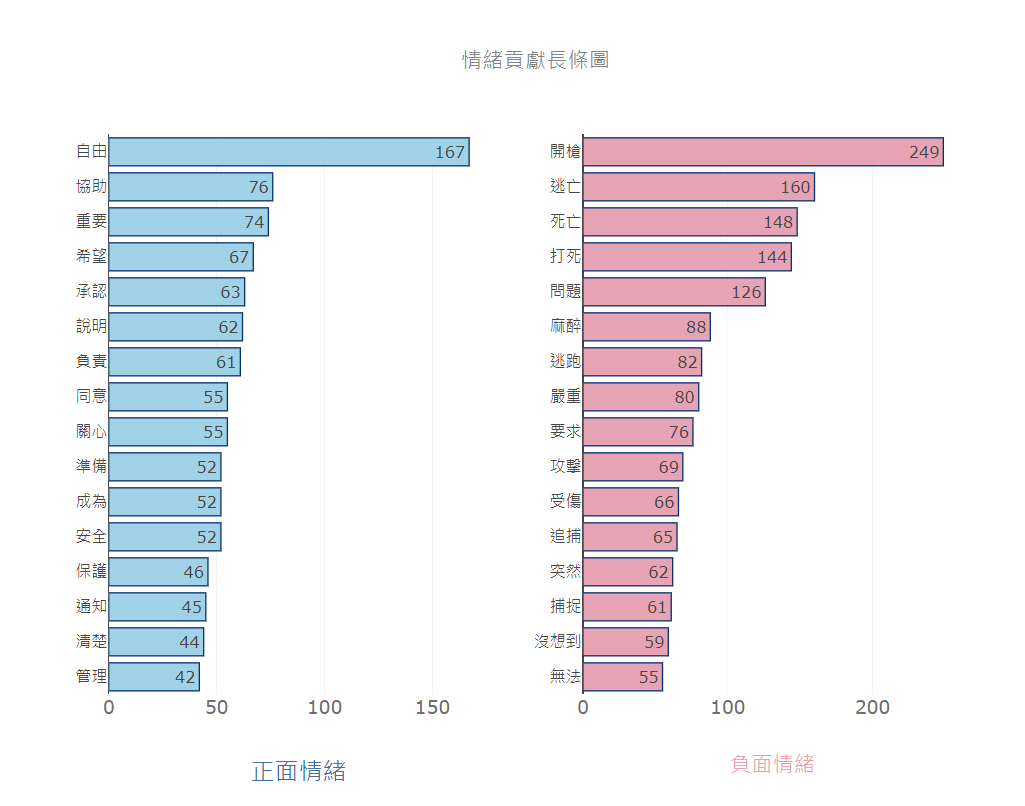


## （二）NTUSD情緒分析

1. 參數設定（依據分析結果發現，採用預設即達到效果）：



1. 長條圖結果如下：



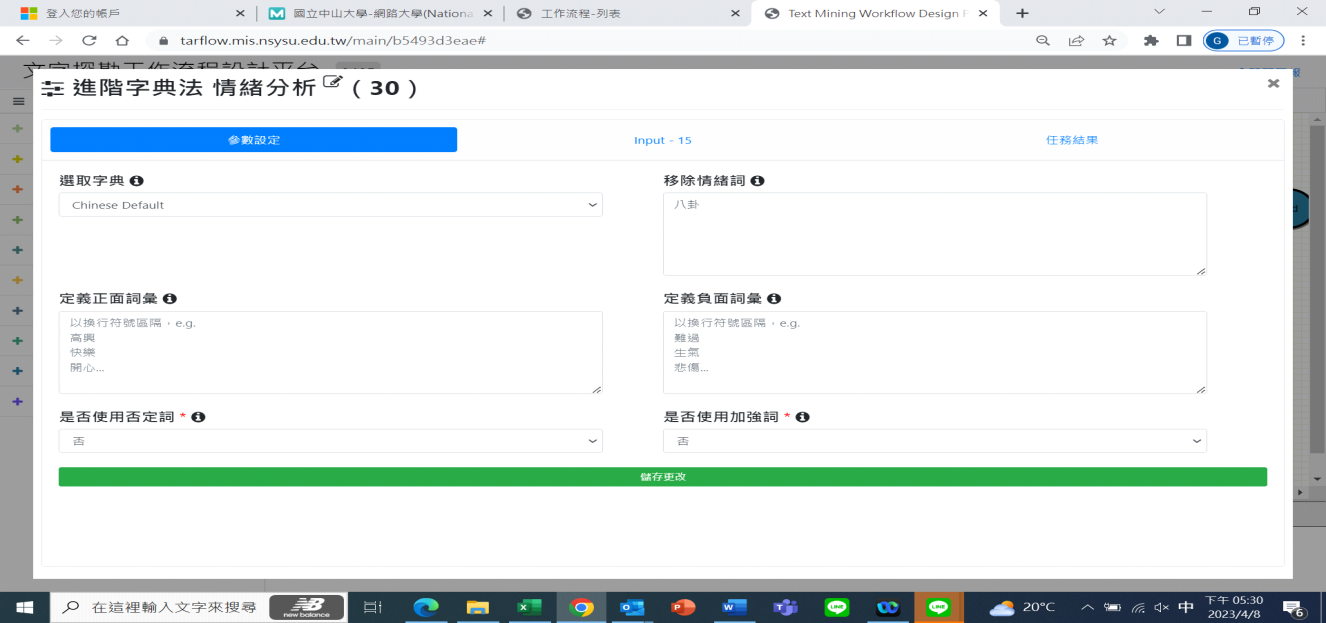
1. 文字雲上方藍色字體本分析中的正面詞彙，粉紅色則表示負面詞彙：



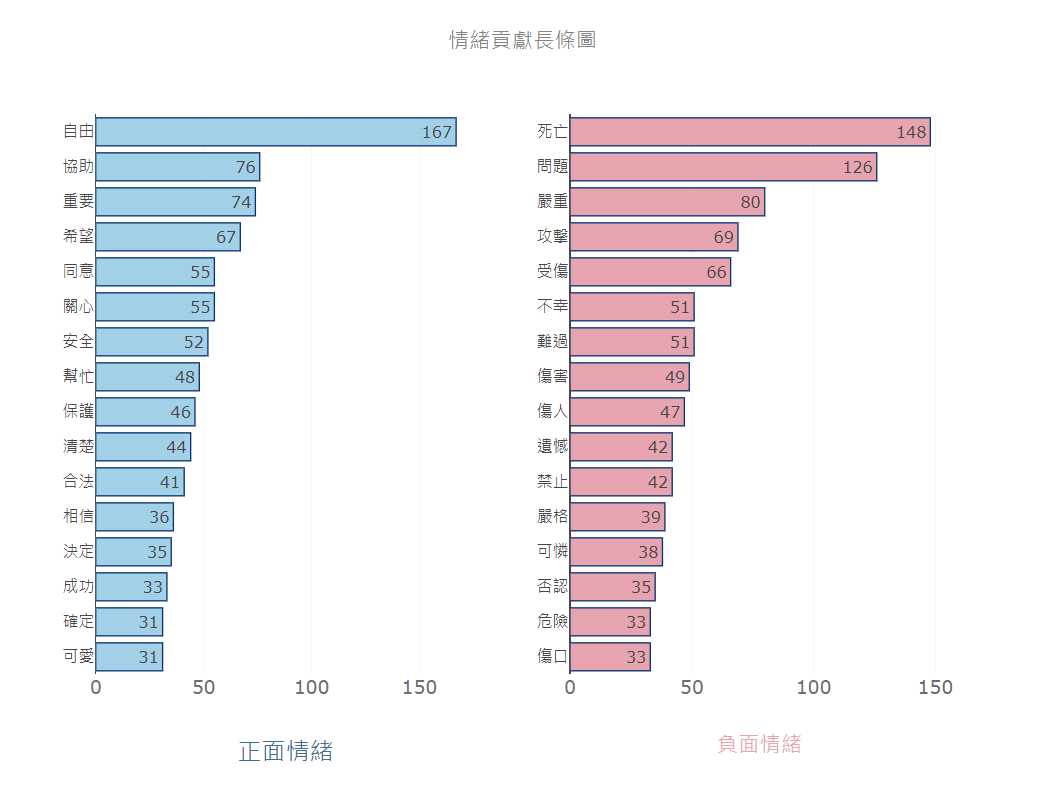
## （三）Chinese Default情析分析

在參數設定區中，本進階字典法特別將「八卦」列為情緒詞，但由於本分析爬蟲資料來源為 PTT 「八卦版」，詞彙「八卦」本身並非日常所指，所以我們將之列入移除情緒詞中，得到以下任務結果，並可以看出各正負面詞彙對於分析的貢獻程度。

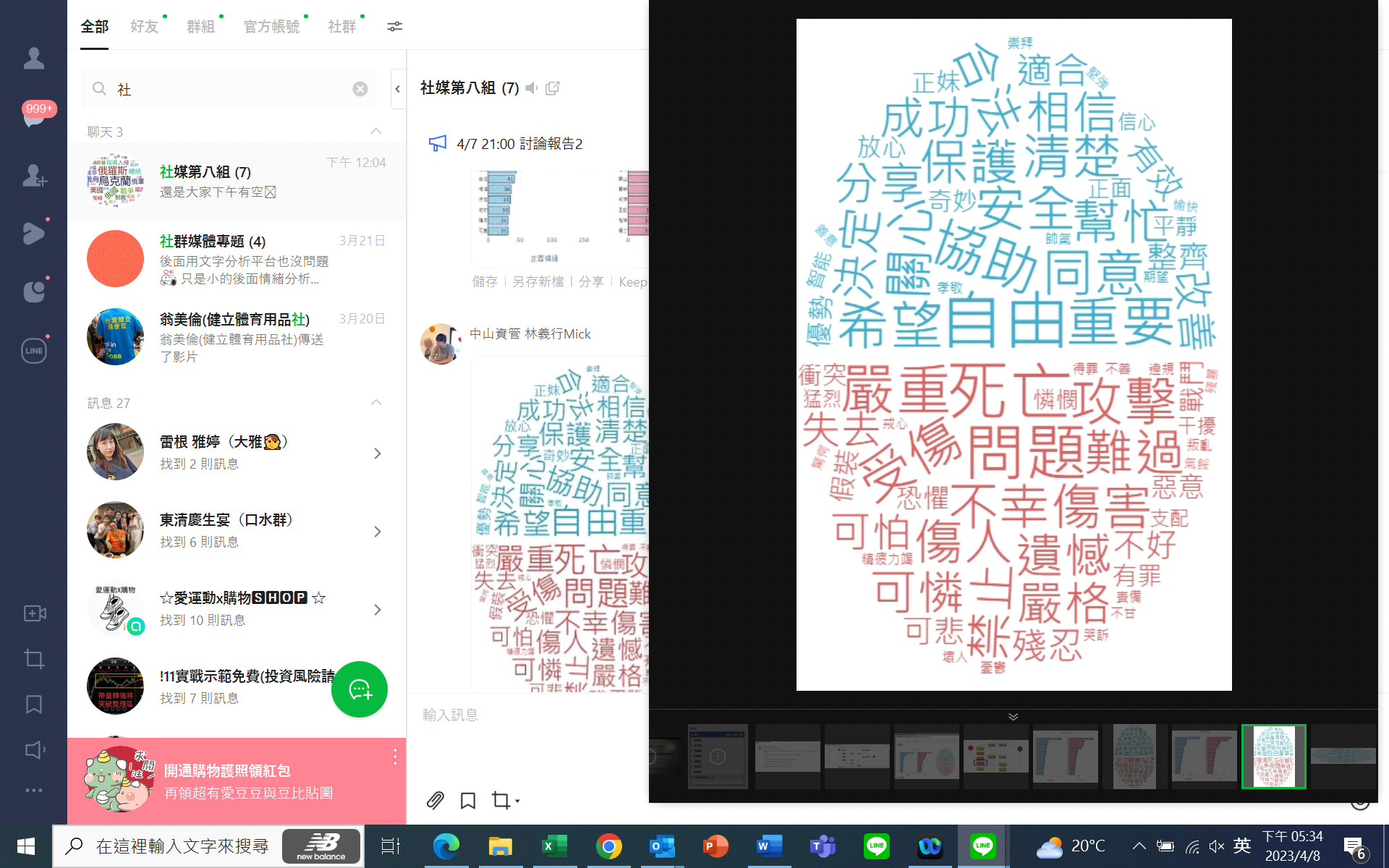
1. 參數設定：



1. 長條圖結果如下：



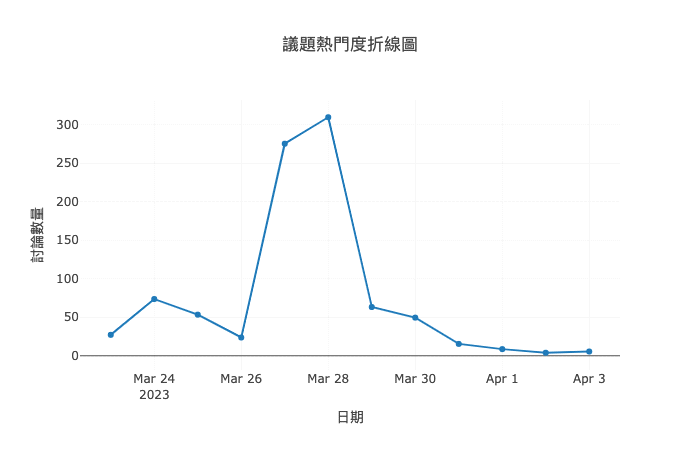
1. 文字雲上方藍色字體本分析中的正面詞彙，粉紅色則表示負面詞彙：



# 五、視覺化圖表分析

## （一）視覺化儀表板-議題討論數量折線圖

根據「分群彙總（日期）」加總3/23-4/3期間PTT對事件正負情緒意見討論數量，產生每日議題討論熱門度變化趨勢。此事件始於3/22桃園市平鎮區鎮興里里長黃志杰在臉書發文表示曾目擊狒狒現身桃園街頭，根據資料爬蟲結果發現，3/23即開始在PTT出現討論聲量，並在3/27狒狒遭槍擊死亡當天到隔天3/28，網路討論聲量達到最高，隨後慢慢趨緩。

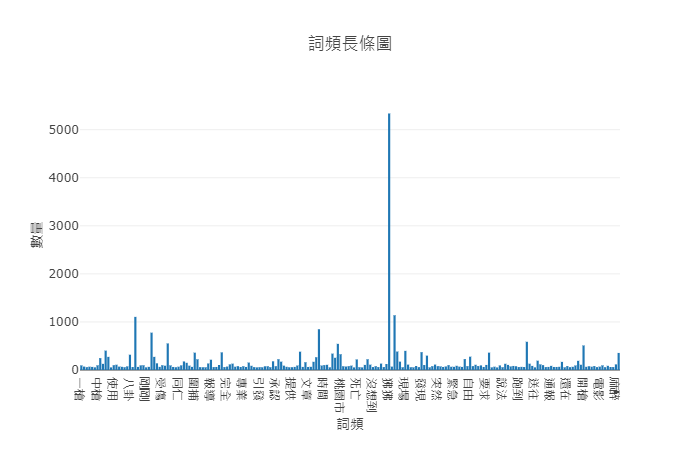


## （二） 視覺化儀表板-詞頻計算

根據事件關鍵字爬蟲資料，並進行資料處理及停用字清理後，將出現較多的100個詞彙出現的頻率產生文字雲，從單詞大小可看出民眾提及相關議題的頻率。從文字雲可看出針對狒狒脫逃事件，除了狒狒本身及出現地點外，民眾對相關單位如農業局、六福村、政府的討論度也相當高，由此可知，民眾認為狒狒脫逃事件，這幾個相關單位的責任也應該被檢視，還有獵人、獵槍、開槍等的字詞，也反映出民眾對於獵人與開槍權限的制度爭議也相當熱烈。

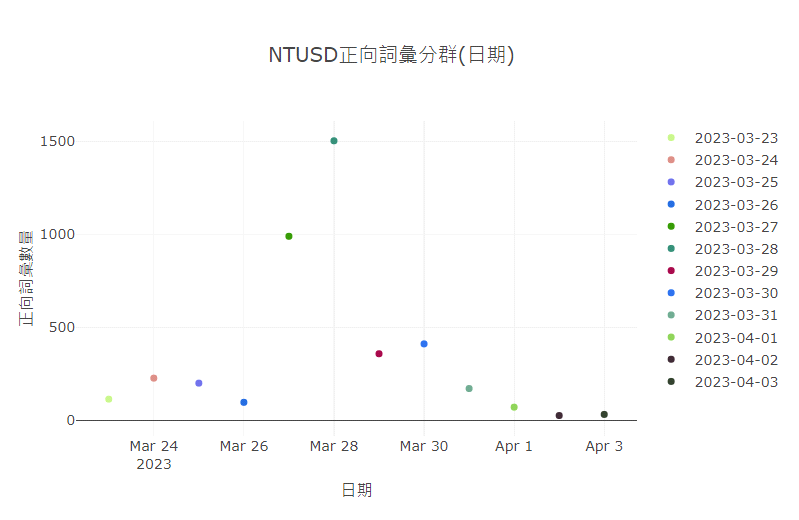


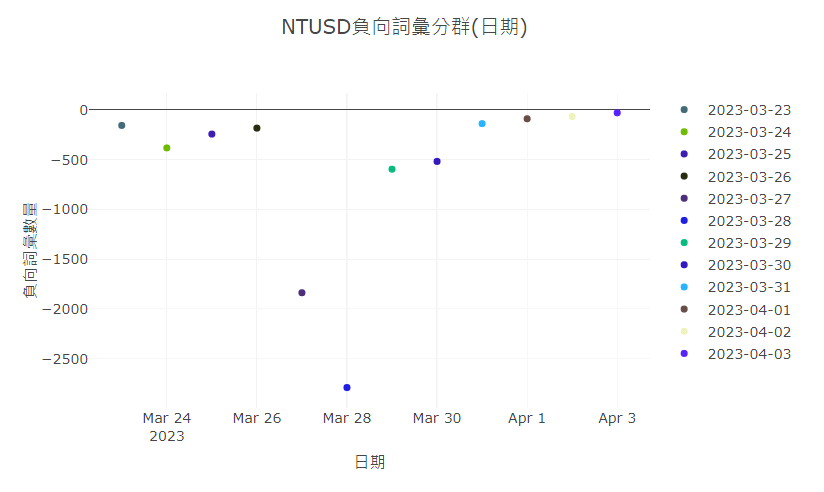
文字雲可看出每個詞彙出現的相對次數多寡，詞頻長條圖則是顯示討論度較高的200個詞彙次數。



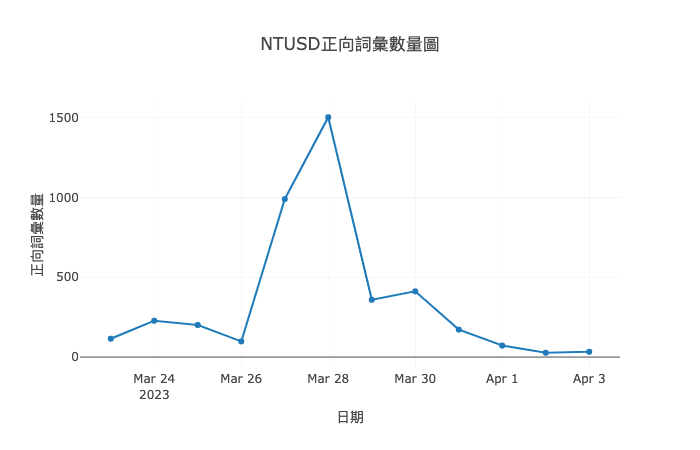
## （三） 視覺化儀表板-NTUSD正向、負向分群匯總

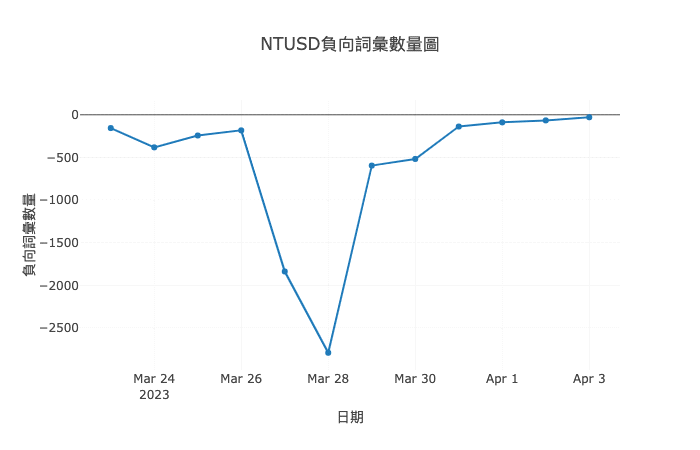
根據NTUSD情緒字典分析狒狒脫逃事件評論，產生正負向情緒詞彙數量分群散佈圖，由此圖可看出事件一開始，正負向詞彙數量尚無明顯差距，然而當狒狒遭槍擊死亡次日（3/28） 無論正向或負向情緒字彙皆於達到高峰，但可明顯看出負向詞彙數量較多，顯示一般民眾對於狒狒死亡一事持負面看法居多。而後隨著新聞的熱度及關注度下降，報導數量趨於平緩，正負向詞彙的數量波動也趨於平緩。





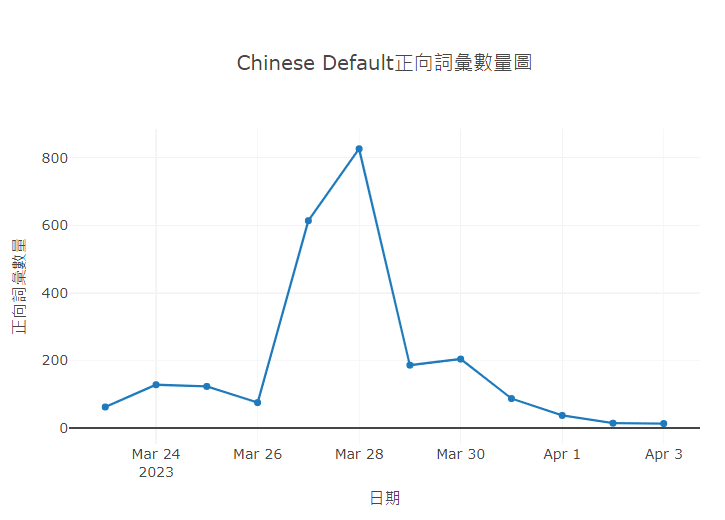
將散佈圖進一步產生折線圖則可更清楚表示數量變化趨勢。

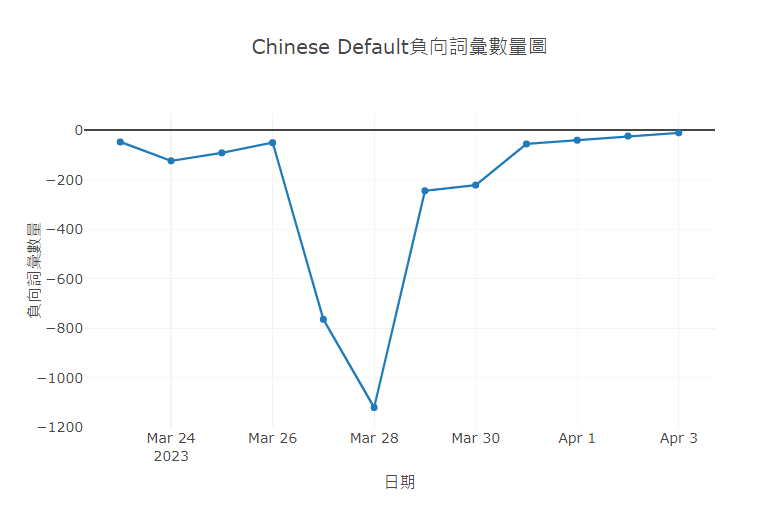




## （四） 視覺化儀表板-Chinese Default正向、負向分群匯總

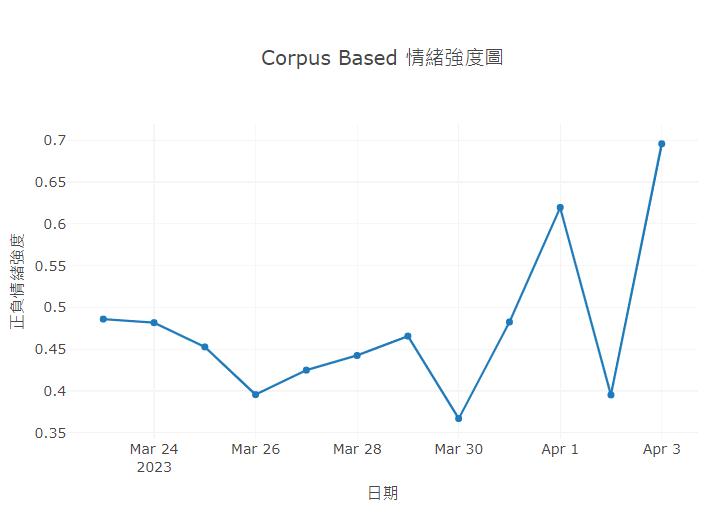
根據Chinese Default情緒字典分析狒狒脫逃事件評論，可看出結果與NTUSD情緒字典的分析結果趨勢相同，顯見狒狒脫逃事件的情緒分析無明顯爭議之處，分析結果還是以負向情緒居多。



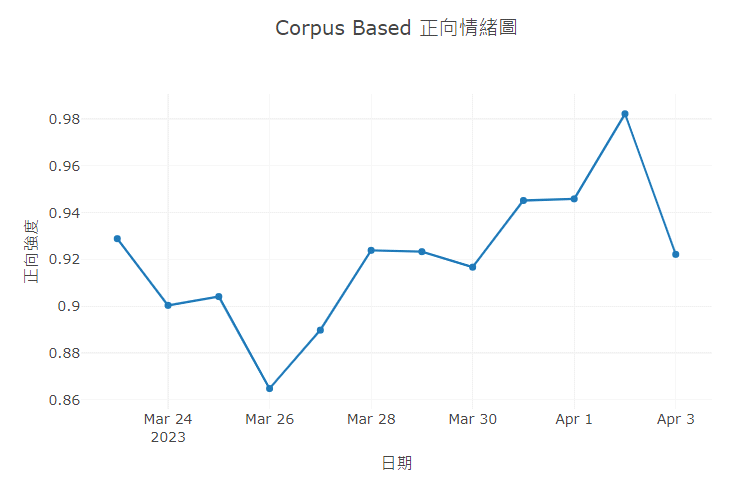


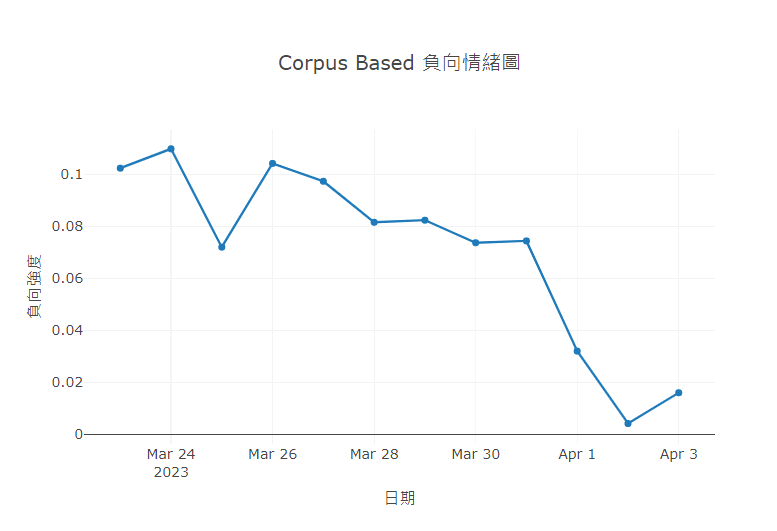
## （五） 視覺化儀表板-Corpus Based分群匯總

根據Corpus Based分群匯總，將所有評論情緒分數平均計算，依照事件發生日期觀察每日差異，大於0.5為正向，低於0.5為負向。從分析結果來看，3/31號事件漸漸沉寂後，評論數量下降，情緒分數才逐漸轉向正向。

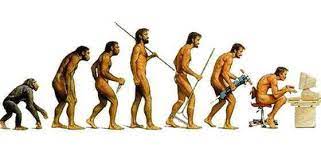


對照正向、負向情緒分群匯總，也可看出類似的結果，在狒狒死亡之時評論聲量最高，同時也是正負聲量起伏最高的時間，且與前述情緒分析一致，負面強度明顯高於正向強度。





# 六、結論

1. 事件一開始，眾人對於事件發生一無所悉，面對未知的事件造成眾人的恐懼。初期所引發的社會問題甚大，因此在PTT討論版上的討論容易引發迴響。
2. 從日期的推移觀察，可看出事件討論聲量大量增加是因為特定事件如：3/27狒狒死亡、3/28桃園市農業局局長帶隊向狒狒遺體鞠躬致意，雖然狒狒事件引起全民甚至國際關注，然而隨著事件落幕，討論聲量也逐漸下降，可見短暫的事件熱潮可能並不會因此增加人民對於動物權利或保護的注意。然而，此觀點還需要進一步觀察長期的趨勢來分析。
3. 在完成進階字典法分析後，本組將幾個關鍵字放入網路搜尋，讓本組對於事件有更不同角度的了解：在幾個較為強烈的詞彙，背後蘊含著民眾的想法、期待，其中業者及地方政府危機處理尤其重要，輿情隨事件發展，會不會造成更大騷動等，是這次觀察、思考到的地方。
4. 在狒狒事件發展後正好是4月清明連假，根據新聞報導，六福村表示連假人潮眾多，且每日遊客數量屢屢增加，在本情緒分析中，我們發現鄉民普遍對於六福村有大量不滿情緒，也對於事件多負面發文，本組對於此現象有以下推論：
   1. PTT鄉民與六福村到訪遊客非重疊之受眾，六福村動物園之負面事件並不影響樂園營運，甚至業績更好。
   2. 民眾有口嫌體正直的可能性，雖然揚言抵制，但一當商家祭出優惠，會馬上排隊入園，一起喊真香啊真香(據六福村說明連假期間各設施平均排隊時間超過30分鐘)。