# 社群媒體分析 第一次讀書會報告

以外某骨豐龍黑黑龍泉縣

指導老師:黃三益教授

組長: N104020023 林義行

組員: N104020018 謝明和

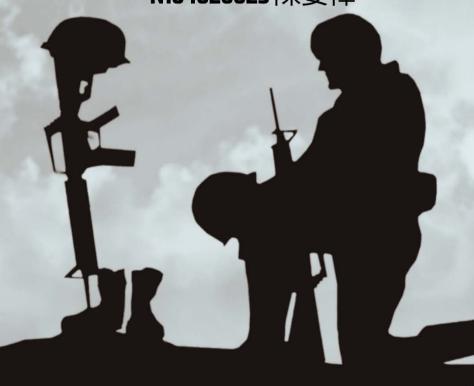
N104020020李宜臻

N104020021 馮慧嬌

N104020022 凌嫈琪

N104020024 林品均

N104020025 陳姿樺



讀書會GROUP 8 20 MARCH, 2023



# Agenda:

<b>-</b> \	議題簡介2
二、	工作流程4
三、	資料流程說明5
四、	情緒分析11
五、	視覺化圖表分析13
六、	結論20

# 一、議題簡介

#### 時間:

2022年迄今(近期之俄烏軍事衝突點為2022年2月24日開始)。

#### 背景:

1991 年蘇聯解體後的烏克蘭境內產生了「烏克蘭化、去俄羅斯化」、俄裔俄羅斯 民族主義造成的國族認同分裂、「北約東擴」引發俄羅斯的不安全感等事件,埋下二 國嫌隙的種子。

2014 年 2 月爆發東烏克蘭戰爭,俄烏同年 9 月,簽訂《明斯克協議》,2015 年簽署《新明斯克協議》,以化解烏克蘭武裝部隊和烏東親俄武裝之間的衝突,防止 在烏克蘭東部衝突進一步升溫。但後續仍有零星衝突發生,俄烏雙方多次指責對方違 反停火協議。

2022年2月24日俄羅斯入侵烏克蘭,是自2014年俄烏戰爭後的重大升級。起因為烏克蘭前總統維克多·亞努科維奇拒絕簽署與歐盟的貿易協定,轉而與俄羅斯簽署經濟合作協定,引發了烏克蘭國內一些親歐派的抗議活動。隨後,烏克蘭政局動盪,烏克蘭親歐盟勢力掌權,引發了俄羅斯的不滿和介入,並且佔領了烏克蘭克里米亞半島。此後,烏克蘭與俄羅斯之間的緊張關係不斷升級,雙方在克里米亞半島、東部地區和頓涅茨克地區展開了武裝衝突。

戰爭持續至今,已經造成數千人死亡,包括平民和軍事人員。國際社會對於此次 衝突持不同態度,一些國家支持烏克蘭,一些國家則支持俄羅斯。

#### 分析原因:

俄烏戰爭對於國際情勢具有高度的重要性,它牽涉到歐洲的安全、國際秩序和俄 羅斯在國際上的地位。以下是分析俄烏戰爭的一些原因:

- 地緣政治的影響:烏克蘭國土坐落於重要的地緣政治區域,位於俄羅斯和歐洲之間,因此成為了俄羅斯和西方國家之間的爭奪區域。
- 能源供應的危機:俄羅斯是歐洲最大的天然氣供應國之一,而烏克蘭是俄羅斯天 然氣向歐洲的重要過境國。因此,俄烏衝突對歐洲能源供應產生了重大影響。

- 3. 國際秩序的挑戰:俄羅斯佔領克里米亞半島和介入烏克蘭內政被認為是對國際秩序的嚴重挑戰。這導致西方國家對俄羅斯實行經濟和政治制裁,以維護國際秩序。
- 4. 國家主權和領土完整:烏克蘭是一個主權國家,其領土完整和國家主權得到國際 社會的承認。俄羅斯的佔領和介入被視為對烏克蘭主權和領土完整的侵犯,引起 了國際社會的譴責。而國際社會對於俄羅斯制裁的態度及實際效果,也成為其他 強權國家對於侵略後果的參考與評估。

總之,俄烏戰爭對於國際秩序、地緣政治和國家主權都產生了極其重大的影響。 對於台灣而言,戰爭爆發後各世界強權的態度也間接影響中國對於侵略周邊國家的看 法,更升高了海峽兩岸政治、軍事的緊張,甚至預測中國侵略台灣的各式分析也甚囂 塵上。

#### 分析期間:

2022/1/1~2022/12/31 (依近期俄烏戰爭為分析主軸)

### 資料來源:

聯合新聞網爬蟲

#### 使用工具:

工作流程系統

## 參考來源:

https://zh.wikipedia.org/zh-tw/%E4%BF%84%E4%B9%8C%E6%88%98%E4%BA%89

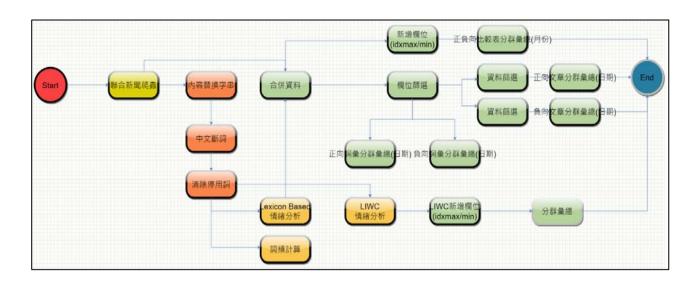
# 二、工作流程

工作流程名稱:第一次讀書會\_第八組

#### 設計步驟:

- 1. 從工作平台中的「聯合新聞爬蟲」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。
- 6. 使用「詞頻計算」, 並用選取出現頻率最高的前 100 個詞組來產生文字雲。
- 7. 使用 Lexicon Based 和 LIWC 兩個情緒分析字典,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「聯合新聞爬蟲」與 Lexicon 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(月期)」,來彙 總每月的正負情緒值。
- 9. 使用「合併資料」合併「聯合新聞爬蟲」與 Lexicon Based 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)。
- 10. 使用「合併資料」合併「聯合新聞爬蟲」與 Lexicon Based 情緒分析結果,用 「欄位篩選」+「資料篩選」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)」,依日期來彙總每日正面文章與負面文章趨勢。
- 11. 在LIWC 情緒分析字典後使用「新增欄位」計算正負向字詞哪一個出現的多,如果正向出現多次就會標記 positive;如果負向出現多次就會標記 negative。
- 12. 將工作平台分析資料結果製作成視覺化圖表:詞頻計算文字雲、正向/負向詞彙分群彙總圖、每月報導數量折線圖、正負向討論聲量直條圖與正負向彙整堆疊圖。
- 13. 最後在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、資料流程說明



# ❖ 聯合新聞爬蟲

俄烏戰爭始於 2022 年 2 月 24 日清晨,俄羅斯兵分四路從東、南、北、東北四個方向入侵烏克蘭。為了瞭解戰爭情勢的發展,本組選定聯合新聞網,作為本次分析的資料來源。

本組自工作流程設計平台中"資料輸入與輸出"元件庫,選擇聯合新聞爬蟲的套裝物件,參數設定選擇 global(全球)新聞,並設定日期自 2022 年 1 月到 2022 年 12 月這一整年期間,針對俄烏戰爭所發生的所有新聞內容來做分析。

#### ➡ 聯合新聞爬蟲 (59)



#### 搜尋關鍵字:

主要為俄烏戰爭、經濟制裁、天然資源、核能武器、人道救援、傭兵、軍事援助等。

#### 設定搜尋關鍵字的主要原因:

欲了解雙方發動戰爭以外,還有哪些議題是會被討論,或是為了那些利益衝突、 或是國際間的觀點。

#### 排除關鍵字:

疫情、美食、熱浪、天候、高溫、度假、火山、海嘯、諾貝爾、哈薩克、哈薩克 政變、內戰、維吾爾、拜習會、彩券、保全、剛果、豪宅、哈利梅根、習近平、種 樹、兒子、日本、耶誕、跨年、升息、同床異夢、在內。

#### 設定排除關鍵字的主要原因:

由於戰爭持續了一年多,雙方仍未有停戰的跡象,搜尋出來的新聞資料量很多。為了讓新聞資料的更集中在俄烏的議題與關係,因此設定了一些排除關鍵字,也讓資料先做第一步的篩檢。

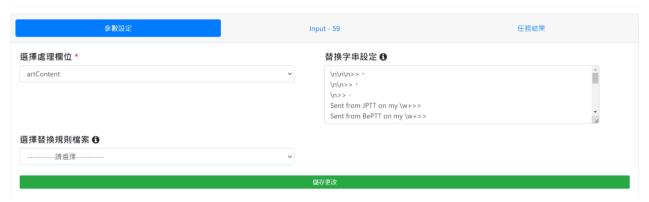
#### 任務結果:

➡ 聯合新聞爬蟲 Ѿ (59)

參數	設定	任務結果	
統計資訊			Â
7 欄位數	2313		
任務結果			
Show 10 ventries  system_id f artTitle artDate	artCatagory artSecondCatagory artUrl	Search: \$\display \text{ artContent} \display\$	dataSource ‡
全螢幕瀏覽	點我下藏完整CSV資料	美國總統拜登今天與歐洲領袖 召開危機會談後市,西方強 <sup>1888年</sup> 「興化」。 <sup>1880年 在東州</sup> <u> </u>	

# ❖ 內容替換字串

幸內容替換字串 ②(2)



1. 將單空白格替換成逗號

n >>

2. 兩個及三個空白格替換成句號

n n >

 $n\rangle \circ$ 

3. 常見字串替換成空白

Sent from JPTT on my  $\w+>>$ 

Sent from BePTT on my \w+>>

Sent from MoPTT on my \w+>>

my iphone \w+>>

on my  $\w+>>$ 

iphone \w+>>

by ptt \w+>>

Sent from \w+>>

XD>>

xd>>

4. 去除網址及影片

 $((http|ftp|https): //)(([a-zA-Z-09\._-]+\.[a-zA-Z]\{2,6\})|([0-9]\{1,3\}\.[0-9]\{1,3\}\.[0-9]\{1,3\}))(:[0-9]\{1,4\})*(/[a-zA-Z-09\._-])*>> (\w+)\.(jpg|gif|png|html)>>$ 

 $/^(?: https?: \/\/)?(?: www\.)?(?: youtu\. be\/|youtube\. com\/(?: embed\/|v\/|watch\?v=|watch\?.+&v=))((\w|-){11})(?: \S+)?$/>>$ 

#### 替換字串設定 ₫



#### 任務結果:

#### 華內容替換字串 ②(2)



## ❖ 中文斷詞

依據斷詞所產出的結果,反覆清洗斷詞並修正,將會用在新聞內容的特殊用語做為斷詞,讓分析可以更為準確,例如,從資料中可以發現澤倫和斯基,馬立波的馬立會被獨立出來,因此對該詞會設定 100 之權重。

### 參數設定:

澤倫斯基 100

北大西洋公約100

頓涅茨克 100

盧甘斯克 100

馬立波 100

### **幸**中文斷詞<sup>☑</sup>(21)

參數設定	Input - 2	任務結果
選擇處理欄位 *	定義詞彙 ❸	
result	深倫斯基 100 北大西洋公約 100 頓涅茨克 100 盧甘斯克 100 馬立波 100	A.
選取字典 ➊		
請選擇	•	
	儲存更改	

#### 任務結果:



9

## ❖ 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:

終於、角色、介紹、待補、附上、維基、百科、五大、整整、一名、無法、現在、目前、認為、包括、指出、進行、懶人、今天、表示、指出、各國、時刻、提供、 真情、宣布、手法、採取,以期出現更佳的視覺化效果。

幸清除停用詞 Ѿ (24)

Input - 21	任務結果	
使用預設停止詞		
是	v	
是否轉為小寫英文	是否轉為小寫英文	
是	•	
清除數字 *	清除數字*	
是	~	
清除特殊標點符號 *		
是	~	
自定義停止詞		
》 終於 角紹 符附上		
	使用預設停止詞  是  是  是  表  表  表  表  表  表  表  表  表  表	

#### 任務結果:

葦 清除停用詞 Ѿ (24)



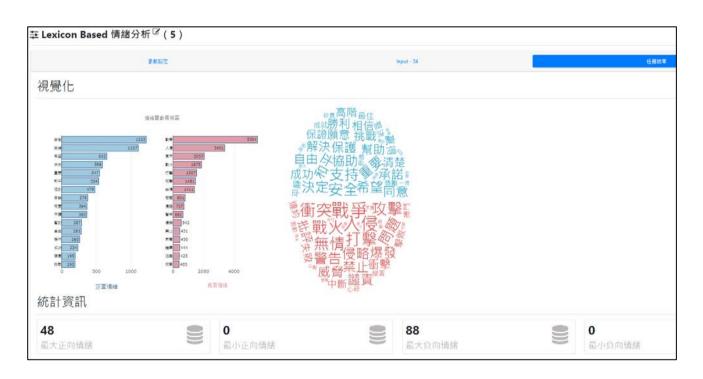
在清理階段,本組也嘗試將 ngram 函數放入,去分析出現頻率較高的詞句,在運行之後產出:「攻打烏克蘭」、「烏克蘭總統」等詞句,可以理解函數認為這些詞句為獨立詞句,但本組希望得到單純是單字的出現頻率,因此本組判斷在本階段不適合放入工作流程,也因此在下個階段得到較中性的文字雲產出。

# 四、情緒分析

本組依據 Lexicon based 與 LIWC 二種情緒字典的來分析新聞內容中的情緒文字,藉以瞭解俄烏戰爭期間新聞報導中的正負面詞彙。

首先運用 Lexicon based 情緒分析,載入預設的情緒字典,得到的結果正負面詞彙均為可接受範圍,由此推測對於 Lexicon based 來說,這類型戰爭報導的用詞較為典型,因此不另定義詞彙,出現最多的負面詞彙為戰爭,正面詞彙為安全,由於為重大國際安全事件,分析結果負面詞彙較多。

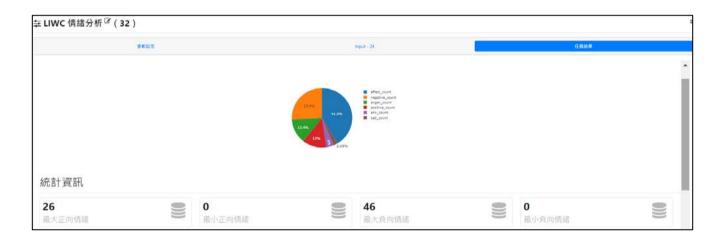




LIWC 情緒分析對情緒的區分更細緻,不只正負向情緒,亦包含 affect、anger、anx 和 sad 等,因此反而不容易針對每日彙總的正負向情緒做分析比較。

### 幸 LIWC 情緒分析 <sup>☑</sup> (32)





# 五、視覺化圖表分析

將資料與前面的日期作合併,並將正負向標籤化,將詞彙與文章分類,以利後續 視覺化圖表分析。

視覺化圖表的呈現,可以取代純文字、數字的敘述,用一張圖表來快速、清楚地 彙整並表達出分析的結果,具有優秀的動態效果和強大的交互體驗,有助於讀者閱讀 或是更有意願閱讀,理解數據,統合視覺解讀的需求,增加溝通效率。

#### 1. 視覺化儀表板-詞頻計算文字雲

使用"詞頻計算"資料產生文字雲,將前50名的字彙做成文字雲,文字雲的字體 大小會以詞頻出現的次數多寡,越多報導或是評論的文字,文字雲的字體就會越大。

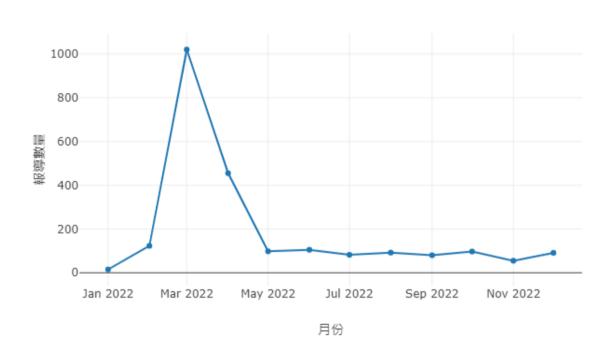
在俄烏戰爭的報導中,除了俄羅斯烏克蘭兩個國家報導主體外,還可以看到像是 民主與共產兩種政體,像是世界兩大強權國家中國與美國,由此可知烏俄戰爭,不僅 僅只是烏俄兩國交戰,亦對其他國家帶來深遠的影響,圖的部分其視覺化介面上的文 字雲可呈現較豐富的顏色變化,但可惜解析度不佳。故本組認為相較之下,原詞頻計 算模組內建所產生文字雲,更易於觀者解讀詞組資訊。



### 2. 視覺化儀表板-每月報導數量折線圖

使用"正負向比較表分群彙總(月份)"資料產生文章報導數量折線圖,是將正 向與負向的報導做加總。戰爭始於由 2022 年 2 月,故由圖表中可看出文章報導數量於 3 月最高,於同年 5 月隨著新聞的熱度及關注度下降,報導數量趨於平緩。

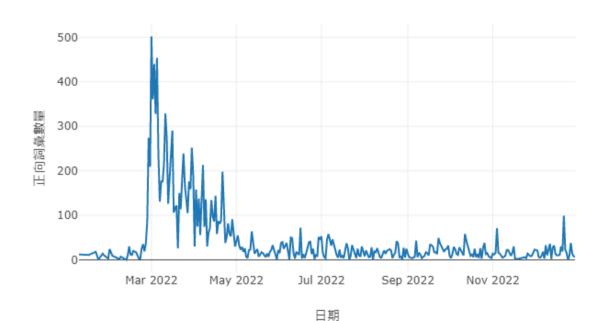




#### 3. 視覺化儀表板-正向詞彙分群彙總圖

使用"正向詞彙分群彙總(日期)"資料產生文章報導正向詞彙數量彙總,呈現 2022年一整年度戰爭新聞文章報導的正向詞彙數量彙總,圖表中的正向詞彙趨勢同於 每月報導數量折線圖,於3月最高,於同年5月隨著新聞的熱度及關注度下降,報導 數量趨於平緩,正向詞彙趨勢也趨於平緩,同時可以發現,5月隨著新聞的熱度及關注 度下降後,新聞討論內容中,正面詞彙的數量仍然存在輕微的波動。

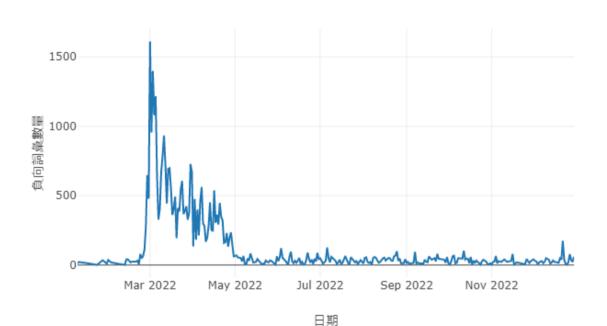
下向詞彙分群彙總圖



#### 4. 視覺化儀表板-負向詞彙分群彙總圖

使用"正向詞彙分群彙總(日期)"資料產生文章報導負向詞彙數量彙總,呈現 2022年一整年度戰爭新聞文章報導的負向詞彙數量彙總,圖表中的負向詞彙趨勢同於 每月報導數量折線圖,於3月最高,於同年5月隨著新聞的熱度及關注度下降,報導 數量趨於平緩,負向詞彙的數量波動也趨於平緩,兩張圖表乍看之下有些雷同,但仔 細看會觀察到本圖縱軸的數值大得多(相較於正向詞彙分群彙總圖),負向詞彙的數 量大幅多於正向詞彙的數量。

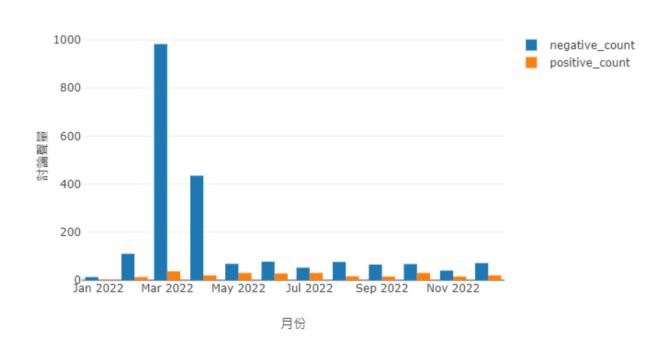




#### 5. 視覺化儀表板-正負向討論聲量直條圖

使用"正負向比較表分群彙總(月份)"資料產生直條圖,本組所做的分析是了解每個月的正、負向詞彙的數量比較。藉以了解是否每個月有所差異。戰爭始發於2022年2月份,開打之初正是全球矚目的焦點,所受到的新聞關注極高,在起初2月到3月的新聞報導非常熱絡,可以發現新聞報導量與正、負向的詞彙是呈正比,但戰爭開打後的二個月,由於新聞熱度與關注度逐漸下降,正、負向詞彙數量也隨之減少。因此可以發現正、負向的詞彙數量就逐漸減少。由直條圖每個月所呈現的正、負向詞彙數量,也可以反應出新聞報導的熱度。



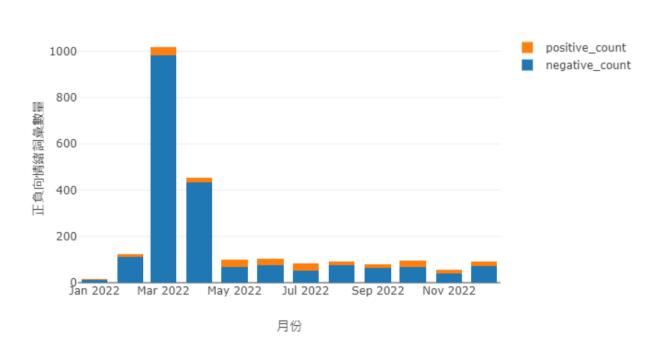


#### 6. 視覺化儀表板-正負向彙整堆疊圖

使用"正負向比較表分群彙總(月份)"資料產生堆積長條圖,透過堆積長條圖,可以清楚看見各個月份正負向的詞彙數量彙總,以及正向和負向的詞彙數量佔比分佈,2022年在5月、7月份,正向詞彙的數量佔比較多,其中7月份正向詞彙的數量幾乎與負向詞彙一樣多。

整體來說圖表中可看出新聞內容中正負向情緒差異,當中負向情緒多於正向情緒,多數媒體多於報導中呈現出戰爭所帶來的負面效應。

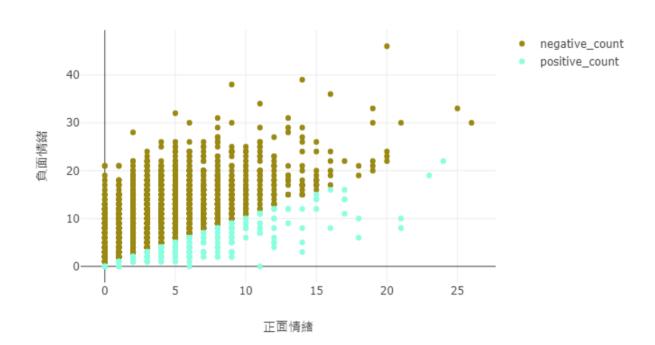




#### 7. 視覺化儀表板-LIWC 情緒分布圖

為了進一步了解正負向情緒詞彙的數量,本組利用 LIWC 字典法分析結果,製作了情緒分布圖,由圖中詞彙出現的數量,發現棕色負面詞彙出現的數量,遠大於湖水綠色的正面詞彙數量。代表這段期間的新聞報導,多屬於負面的情緒。顯示戰爭時間越長,負面的情緒也隨之越多,故媒體報導反映出民眾不願戰爭所帶來的負面效應,渴望正面的情緒會越來越強烈。

#### LIWC情緒分布圖



# 六、結論

- 1. 戰爭始發於 2022 年 2 月份,開打之初正是全球矚目的焦點,所受到的新聞關注極高,2 月到 3 月的新聞報導非常熱絡,但到了 4 月份之後媒體報導的篇幅大量下降,可能間接顯示出國內關注度的下降。
- 2. 情緒詞彙及文章均於戰爭開打時為最顛峰,吸引到最多人的關注以及報導,但其實烏克蘭戰後面臨的生活問題及後續該如何復原原本的生活也是很重要的議題,但目前情緒分析未涉及到此部分。
- 3. 依 Lexicon based 情緒分析圖、LIWC 情緒分析圖及正負向討論聲量直條圖觀察結果,整體而言負向詞彙比例高於正向詞彙,推測其原因在於戰爭所帶來大量負面效應,如經濟、能源、安全及穩定的威脅,媒體對於和平呼籲越發強烈,則對於戰爭產生負向詞彙的使用也會越多。
- 4. 工作流程系統的操作及使用,對於不會寫程式語言的同學,入門門檻較低,讀書會制度的設立,也能藉由彼此討論對操作更為了解整體流程設計,惟替換字串仍需孰悉程式語言的協助,以及後續視覺化欄位資料的篩選亦須互相協助討論,才能有更好的呈現。