

請擷取一則至少2000字的台南旅遊相關議題的專題報導或報告，彙整成一個文本，利用**主題探勘**技術將這些新聞區分成若干主題，並根據這些主題的關鍵詞，個別畫出文字雲，說明所分析呈現的資訊為何？

文章連結:

<https://www.twtainan.net/zh-tw/event/newsdetail/5350>

<https://www.twtainan.net/zh-tw/event/newsdetail/5345>

<https://www.twtainan.net/zh-tw/event/newsdetail/5342>

<https://www.twtainan.net/zh-tw/event/newsdetail/5328>

文章發布時間:111/11/7~111/11/18

使用**jieba**套件進行**斷詞**

停用詞除了預設的標點符號外有增加"台南"和"台南市"跟"臺南"和"臺南市"以及其他字詞如下表，因為這篇文章在找尋時本身就是以台南為中心，所以將他們設為停用詞來看觀察其他關鍵詞，否則出來的前10個關鍵字有兩個一定是"台南"和"台南市"跟"臺南"和"臺南市"。其他的字詞在呈現主題分析時認為非關鍵，所以將他們列為停用詞，使結果更為精準。最後出來後儲存關鍵詞長度大於1的詞彙做主題分析。

停用詞					
台南	台南市	臺南市	臺南	10	11
12	2022	表示	大家	市長	今年
局長	議員	單位	一起	今日	結束
只要	還有	不同	感謝	其中	前來
現場	日至	一同	歡迎	此外	參加

新增自定義詞庫，讓斷詞可以更加完善。

自定義詞庫				
台南	參與	臺南市	大魚的祝福	德陽艦園區
臺南	臺灣	台南市	臺南購物節	泰迪熊in新營
市長	圓滿	趙卿惠	馬來西亞	港濱歷史公園
次長	市價	陳宗彥	馬里亞納	新營文化中心
活潑	可愛	林佩君	共襄盛舉	1661臺灣船園區
貼心	局長	郭貞慧	佳馳國際	大臺南國際旅展
國家	區長	蔡承鴻	億載金城	大臺南會展中心
來自	在地	神農街	安平樹屋	貴人散步音樂節
舞臺	古蹟	伴手禮	美美的	臺南城市音樂節
商機	底蘊	服務臺	金曲獎	佳馳國際有限公司

邀請	音樂	關子嶺	管理處	德元埤荷蘭村風車節
終於	店家	雲嘉南	蟋蟀君	新營南瀛綠都心公園
視角	驗證	優美	魚頭君	

使用gensim套件裡的LdaModel製作主題探勘。

最終設定num\_topics=4:呈現4個主題，因為設定四個主題出來沒重複性且種類豐富，當設置五個主題時雷同性高，設定三個主題的時候雖然沒有重複性但樣式有點少，如下方三張表。

lda = LdaModel(corpus=word_num, id2word=word_id, num_topics=3) lda_list=lda.print_topics(3) num_topics=3:呈現3個主題											
詞彙	主題一	泰迪熊	活動	觀光	園區	音樂	民眾	旅展	泰迪熊in新營	旅遊局	旅遊
權重		0.016	0.010	0.009	0.007	0.007	0.006	0.005	0.005	0.005	0.005
詞彙	主題二	觀光	消費	音樂	活動	政府	旅遊局	登錄	發票	旅遊	設計
權重		0.010	0.010	0.009	0.009	0.008	0.006	0.006	0.006	0.006	0.005
詞彙	主題三	活動	泰迪熊	音樂	泰迪熊in新營	可愛	旅遊局	觀光	大臺南國際旅展	戶外	郭貞慧
權重		0.011	0.010	0.009	0.007	0.006	0.005	0.005	0.005	0.005	0.005

lda = LdaModel(corpus=word_num, id2word=word_id, num_topics=4) lda_list=lda.print_topics(4) num_topics=4:呈現4個主題											
詞彙	主題一	音樂	活動	觀光	旅遊局	戶外	郭貞慧	貴人散步音樂節	美食	台南城市音樂節	泰迪熊
權重		0.016	0.013	0.012	0.011	0.008	0.007	0.007	0.006	0.006	0.006
詞彙	主題二	活動	園區	觀光	黃偉哲	大臺南國際旅展	旅行	旅遊	消費	旅遊局	特色
權重		0.009	0.009	0.008	0.008	0.007	0.007	0.006	0.006	0.005	0.005
詞彙	主題三	泰迪熊	泰迪熊in新營	可愛	活動	民眾	新營	熊熊	展期	旅展	新營文化中心
權重		0.023	0.008	0.007	0.007	0.006	0.006	0.005	0.004	0.004	0.004

詞彙	主題四	消費	泰迪熊	活動	旅遊	發票	觀光	音樂	登錄	政府	民眾
權重		0.011	0.009	0.008	0.008	0.008	0.008	0.008	0.007	0.007	0.006

```
lda = LdaModel(corpus=word_num, id2word=word_id, num_topics=5)
```

```
lda_list=lda.print_topics(5)
```

num\_topics=5:呈現5個主題

詞彙	主題一	活動	泰迪熊	消費	民眾	發票	新營	特色	古蹟	音樂	登錄
權重		0.016	0.010	0.009	0.008	0.007	0.007	0.007	0.007	0.006	0.006
詞彙	主題二	活動	園區	泰迪熊	觀光	風車	戶外	旅遊局	泰迪熊 in新營	音樂	政府
權重		0.015	0.012	0.011	0.009	0.006	0.006	0.006	0.006	0.006	0.006
詞彙	主題三	泰迪熊	可愛	大臺南 國際旅展	泰迪熊 in新營	觀光	新營	使用	新化	旅行	熊熊
權重		0.012	0.008	0.007	0.007	0.006	0.006	0.006	0.005	0.005	0.005
詞彙	主題四	觀光	消費	旅展	活動	旅遊局	政府	登錄	旅遊	發票	參展
權重		0.013	0.011	0.009	0.009	0.008	0.008	0.006	0.006	0.006	0.006
詞彙	主題五	音樂	泰迪熊	觀光	旅遊局	貴人散步 音樂節	邀請	古蹟	民眾	台南城市 音樂節	神農街
權重		0.021	0.014	0.009	0.008	0.008	0.008	0.007	0.007	0.007	0.006

最後利用num\_topics=4製作出四個主題，並分別畫出文字雲

從下表和文字雲可以看出：

跟著觀光旅遊局局長郭貞慧一起暢遊貴人散步音樂節和臺南城市音樂節吧!在戶外享受台南太陽的熱情，一邊觀光，一邊享受美食之都的各式美食，並聽著美妙的音樂旋律放鬆自我。另外還有期限限定的泰迪熊等著你來合照留念!

主題一:台南音樂節，等你一起來聽音樂吃美食		
詞彙	權重	文字雲

音樂	0.016	
活動	0.013	
觀光	0.012	
旅遊局	0.011	
戶外	0.008	
郭貞慧	0.007	
貴人散步音樂節	0.007	
臺南城市音樂節	0.006	
泰迪熊	0.006	
美食	0.006	

從下表和文字雲可以看出：


台南市長黃偉哲在大台南會展中心的園區內和旅遊局推動了大臺南國際旅展，這個活動是為了推動旅遊和觀光，進而讓消費者可以去一些有特色的觀光景點消費，享受不一樣的台南旅遊體驗。透過這個活動使各方業者的業績都有上升。

主題二：來大臺南國際旅展，帶你遊台南		
詞彙	權重	文字雲
活動	0.009	
園區	0.009	
觀光	0.008	
黃偉哲	0.008	
大臺南國際旅展	0.007	
旅行	0.007	
旅遊	0.006	
消費	0.006	
旅遊局	0.005	
特色	0.005	

從下表和文字雲可以看出：

「2022泰迪熊in新營」活動在新營文化中心附近正式開跑，這裡有很多可愛的泰迪熊可以觀賞，這個展覽在展期間也間接推帶動了旅展的人潮，來到這裡的民眾皆表示泰迪熊十分可愛，很適合在紀念品店買一隻小熊熊回家收藏起來！

主題三：可愛泰迪熊在新營，等你來玩！		
詞彙	權重	文字雲

泰迪熊	0.023	
泰迪熊in新營	0.008	
可愛	0.007	
活動	0.007	
民眾	0.006	
新營	0.006	
熊熊	0.005	
展期	0.004	
旅展	0.004	
新營文化中心	0.004	

從下表和文字雲可以看出：  
 政府鼓勵民眾在旅遊時若是參與泰迪熊活動消費時記得登錄發票，一邊聽音樂，一邊看泰迪熊，還有機會可以中大獎!這麼好康的消費活動，你還不參加嗎?登錄發票是台南市政府推動觀光消費的一種活動，間接刺激了民眾的消費，帶動經濟發展。

主題四:逛泰迪熊，還可以登錄發票抽大獎，不參加的話機會就給別人囉!		
詞彙	權重	文字雲
消費	0.011	
泰迪熊	0.009	
活動	0.008	
旅遊	0.008	
發票	0.008	
觀光	0.008	
音樂	0.008	
登錄	0.007	
政府	0.007	
民眾	0.006	