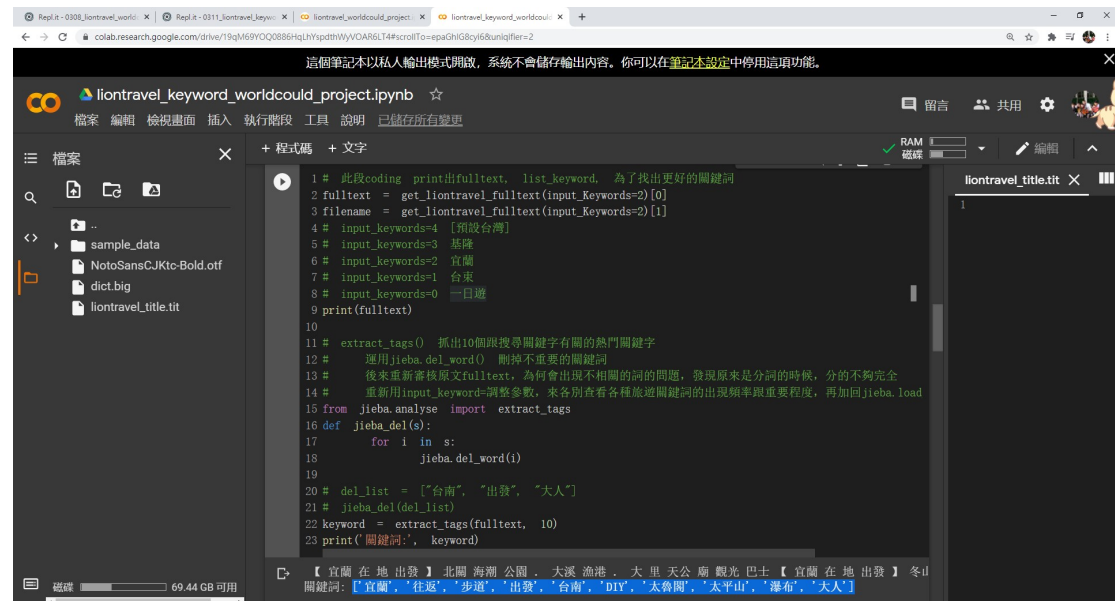


為了找更好的旅遊關鍵詞：10 個

先將 fulltext, input_keywords=2 設定:宜蘭

先看原始印出的 list_keyword = extract_tags(fulltext, 10)



```
1 # 此段coding print出fulltext, list_keyword, 為了找出更好的關鍵詞
2 fulltext = get_liontravel_fulltext(input_Keywords=2)[0]
3 filename = get_liontravel_fulltext(input_Keywords=2)[1]
4 # input_keywords=4  [預設台灣]
5 # input_keywords=3 基隆
6 # input_keywords=2 宜蘭
7 # input_keywords=1 台東
8 # input_keywords=0 一日遊
9 print(fulltext)
10
11 # extract_tags() 抓出10個跟搜尋關鍵字有關的熱門關鍵字
12 # 運用jieba.del_word() 刪掉不重要的關鍵詞
13 # 後來重新審核原文fulltext, 為何會出現不相關的詞的問題, 發現原來是分詞的時候, 分的不夠完全
14 # 重新用input_keyword=調整參數, 來各別查看各種旅遊關鍵詞的出現頻率跟重要程度, 再加回jieba.load
15 from jieba.analyse import extract_tags
16 def jieba_del(s):
17     for i in s:
18         jieba.del_word(i)
19
20 # del_list = ["台南", "出發", "大人"]
21 # jieba.del(del_list)
22 keyword = extract_tags(fulltext, 10)
23 print('關鍵詞:', keyword)
```

【宜蘭 在地出發】北關 海潮 公園, 大溪 漁港, 大里 天公 廟 觀光 巴士【宜蘭 在地出發】冬山
關鍵詞: ['宜蘭', '往返', '步道', '出發', '台南', 'DIY', '太魯閣', '太平山', '瀑布', '大人']

找一日遊

印出關鍵詞: ['宜蘭', '往返', '步道', '出發', '台南', 'DIY', '太魯閣', '太平山', '瀑布', '大人']

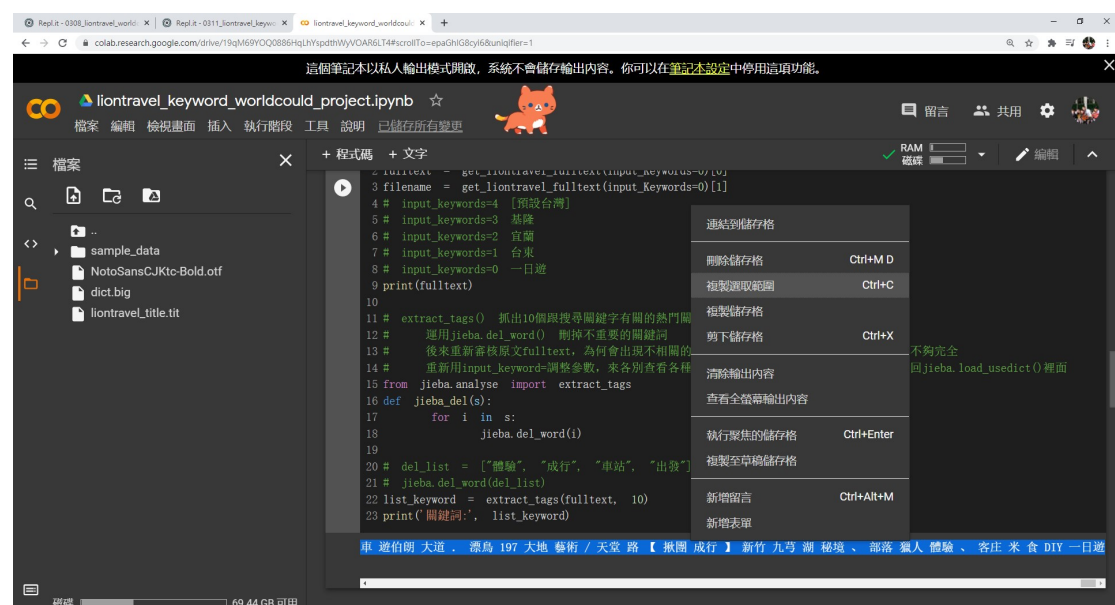
發現 '出發', '台南', '大人' 在宜蘭遊裡感覺突兀, 查查看其前後文



```
23 print('關鍵詞:', keyword)
```

【宜蘭 在地出發】北關 海潮 公園, 大溪 漁港, 大里 天公 廟 觀光 巴士【宜蘭 在地出發】冬山
關鍵詞: ['宜蘭', '往返', '步道', '出發', '台南', 'DIY', '太魯閣', '太平山', '瀑布', '大人']

複製 fulltext



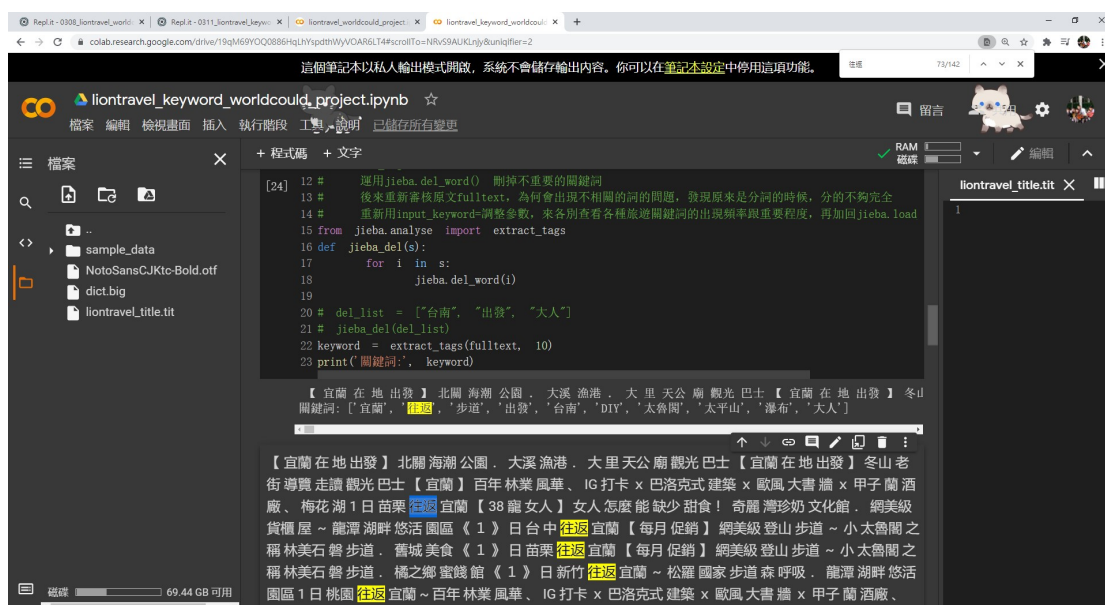
```
2 fulltext = get_liontravel_fulltext(input_Keywords=0)[0]
3 filename = get_liontravel_fulltext(input_Keywords=0)[1]
4 # input_keywords=4  [預設台灣]
5 # input_keywords=3 基隆
6 # input_keywords=2 宜蘭
7 # input_keywords=1 台東
8 # input_keywords=0 一日遊
9 print(fulltext)
10
11 # extract_tags() 抓出10個跟搜尋關鍵字有關的熱門關鍵字
12 # 運用jieba.del_word() 刪掉不重要的關鍵詞
13 # 後來重新審核原文fulltext, 為何會出現不相關的詞的問題, 發現原來是分詞的時候, 分的不夠完全
14 # 重新用input_keyword=調整參數, 來各別查看各種旅遊關鍵詞的出現頻率跟重要程度, 再加回jieba.load
15 from jieba.analyse import extract_tags
16 def jieba_del(s):
17     for i in s:
18         jieba.del_word(i)
19
20 # del_list = ["體恤", "成行", "車站", "出發"]
21 # jieba.del_word(del_list)
22 list_keyword = extract_tags(fulltext, 10)
23 print('關鍵詞:', list_keyword)
```

車 遊伯朗 大道, 漂鳥 197 大地 藝術 / 天堂 路【揪團 成行】新竹 九弓 湖 秘境、部落 獵人 體驗、客庄 米 食 DIY 一日遊

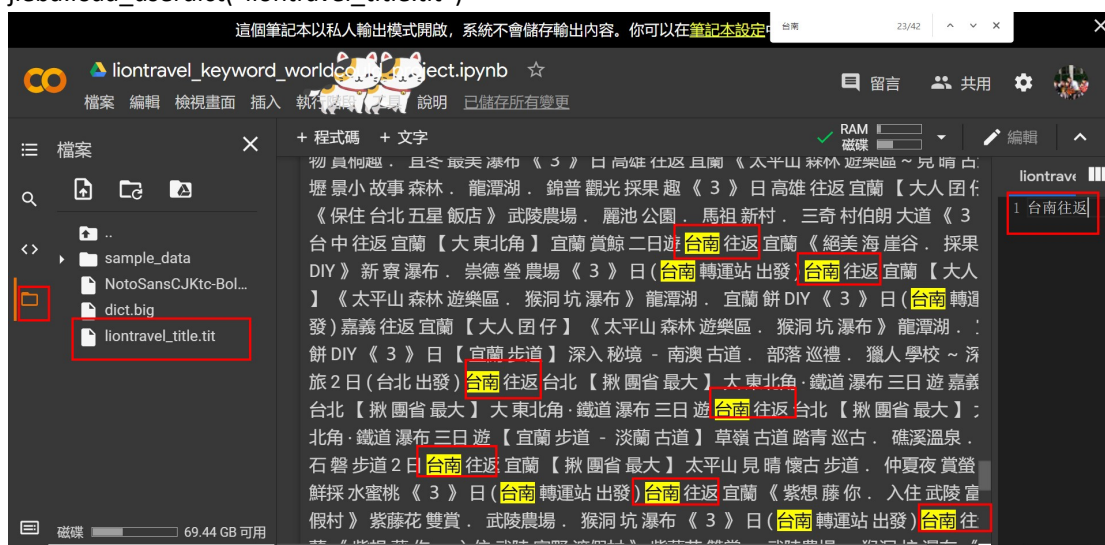
用 colab 新增文字



貼上之後 Ctrl + F 搜尋到，142 筆(這邊要 $\frac{1}{2}$ 大概 71 筆，因為剛剛複製至下面來查閱)“往返”的字詞，這個詞彙對於使用者來說應該是滿重要的詞彙，從哪邊搭車回程到哪邊，先不做刪除，接著看下一個關鍵詞“台南”



“台南”有 $\frac{42}{2}$ 大概 21 筆，其中發現上下文，大多是“台南往返”，“往返”在關鍵詞也有出現，代表宜蘭遊，大多行程是台南往返的行程，把它加入到先前建立好的 jieba.load_userdict(“liontravel_title.tit”)



接著是“出發”(52/2)約 26 筆，但看前後文覺得不重要，把它排除關鍵詞，jieba.del_word()

```
# extract_tags() 抓出10個跟搜尋關鍵字有關的熱門關鍵字
# 運用jieba.del_word() 刪掉不重要的關鍵詞
# 後來重新審核原文fulltext，為何會出現不相關的詞的問題，發現原來是分詞的時候，分的不夠完全
# 重新用input_keyword=調整參數，來分別查看各種旅遊關鍵詞的出現頻率跟重要程度，再加回jieba.load_us

from jieba.analyse import extract_tags
def jieba_del(s):
    for i in s:
        jieba.del_word(i)

# del_list = ["台南", "出發", "大人"]
# jieba_del(del_list)
del_list = ["出發"]
jieba_del(del_list)
keyword = extract_tags(fulltext, 10)
print('關鍵詞:', keyword)
```

【宜蘭在地出發】北關海潮公園．大溪漁港．大里天公廟觀光巴士【宜蘭在地出發】冬山老
關鍵詞: ['宜蘭', '往返', '步道', '出發', '台南', 'DIY', '太魯閣', '太平山', '瀑布', '大人']

在地出發】北關海潮公園．大溪漁港．大里天公廟觀光巴士【宜蘭在地出發】冬山老街導
觀光巴士【宜蘭】百年林業風華、IG打卡 x 巴洛克式建築 x 歐風大書牆 x 甲子蘭酒廠、梅
日苗栗往返宜蘭【38 寵女人】女人怎麼能缺少甜食！奇麗灣珍奶文化館．網美級貨櫃屋～
畔悠活園區《1》日台中往返宜蘭【每月促銷】網美級登山步道～小太魯閣之稱林美石磐

接著“大人”(50/2)約 25 筆，皆是“大人囡仔”，代表宜蘭遊裡面滿多大人囡仔的可以去的行程，將它放入自訂義辭典

```
18 jieba.del_word(i)
19
20 # del_list = ["出發", "成行", "兩人", "高雄", "台南", "車站"]
21 # jieba_del(del_list)
22 keyword = extract_tags(fulltext, 10)
23 print('關鍵詞:', keyword)
```

【台東出發．兩人成行】金樽眺望台、東河橋、泰源幽谷、水往上流、東海岸半日【台東出
關鍵詞: ['台東', '往返', '成行', '兩人', '車站', '單車', '伯朗', '綠島', '高雄', '大道']

【台東出發．兩人成行】金樽眺望台、東河橋、泰源幽谷、水往上流、東海岸半日【台東出發
兩人成行．縱谷線半日遊】單車遊伯朗大道．漂鳥197大地藝術．天堂路【台東出發．兩人成行
】小野柳、杉原護漁區、金樽眺望台、加路蘭遊憩區半日【台東出發．兩人成行】台東市區、
海濱公園戶外藝術裝置．國際地標．森林公園騎鐵馬．藏花村半日遊【台東出發兩人成行．南迴
線半日遊】多良車站．向陽薪傳木工坊．華源海灣．台東大學圖書館【台東出發．兩人成行】
山里車站．山里教堂．原生植物園．斑鳩冰品半日【台東出發兩人成行．海岸線半日遊】成功小
鎮巡禮【台東出發．兩人成行】單車遊龍田村．高台茶園．斑鳩冰品半日【台東出發．兩人成
行】多良車站．櫻木平交道．華源海灣半日【台東出發兩人成行．南迴線半日遊】太麻里金針山
青山農場【台東出發兩人成行．海岸線半日遊】馬武窟溪划竹筏．泰源幽谷【台東出發．兩人
成行】金崙溫泉．櫻木平交道．華源海灣半日【兩人成行／台灣觀巴】台東～玉長公路、百年教
堂、加路蘭遊憩區、泰源幽谷、水往上流、鹿野高台1日(台東出發)【台東出發．兩人成行】
加路蘭遊憩區．比西里岸幾米園區．三仙台．東河橋1日【台東出發兩人成行．縱谷線半日遊

做完兩個合併關鍵詞後再來看新的關鍵詞

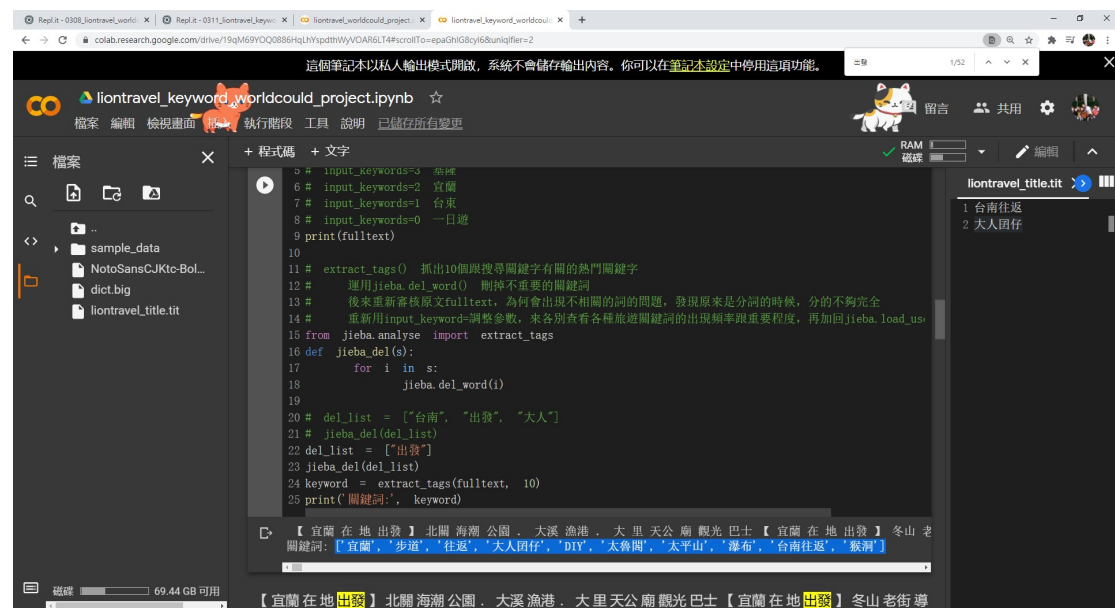
#要記得先把上一步的 coding 將辭典載入進去，再執行下一段 coding 再來看新的關鍵詞

```
1 # 下載jieba 大字典
2 import jieba
3 from urllib.request import urlopen
4 import requests
5 from bs4 import BeautifulSoup
6
7 url = "https://github.com/fxsjy/jieba/raw/master/extra_dict/dict.txt.big"
8 urlopen(url, "dict.big")
9
10 jieba.set_dictionary("dict.big") # 設定好原始字典
11 # 看情況加入自定義字典
12 jieba.load_userdict("liontravel_title.txt") # 載入自定義字典
13
14 data_godateststart = get_liontravel_datas()[0]
15 data_godateend = get_liontravel_datas()[1]
16
```

liontravel_title.txt

1 台南往返
2 大人囡仔

排除後確實有得到更好的關鍵詞



```
3 # input_keywords=3 基隆
6 # input_keywords=2 宜蘭
7 # input_keywords=1 台東
8 # input_keywords=0 一日遊
9 print(fulltext)
10
11 # extract_tags() 抓出10個跟搜尋關鍵字有關的熱門關鍵字
12 # 運用jieba.del_word() 刪掉不重要的關鍵字
13 # 後來重新審核原文fulltext，為何會出現不相關的詞的問題，發現原來是分詞的時候，分的不夠完全
14 # 重新用input_keywords調整參數，來各別查看各種旅遊相關關鍵詞的出現頻率跟重要程度，再加回jieba.load_us
15 from jieba.analyse import extract_tags
16 def jieba_del(s):
17     for i in s:
18         jieba.del_word(i)
19
20 # del_list = ["台南", "出發", "大人"]
21 # jieba_del(del_list)
22 del_list = ["出發"]
23 jieba_del(del_list)
24 keyword = extract_tags(fulltext, 10)
25 print('關鍵詞:', keyword)
```

【宜蘭在地出發】北關海潮公園，大溪漁港，大里天公廟觀光巴士【宜蘭在地出發】冬山老
關鍵詞: ['宜蘭', '步道', '往返', '大人團仔', 'DIY', '太魯閣', '太平山', '瀑布', '台南往返', '猴洞']

【宜蘭在地出發】北關海潮公園，大溪漁港，大里天公廟觀光巴士【宜蘭在地出發】冬山老街導

發現原本沒做自訂辭典的時候要刪除很多覺得不相關的詞，經過前後文查閱，並建入更適合的關鍵詞，最後只要刪掉“出發”，即可得到更好的 top 10 跟宜蘭相關的關鍵詞！

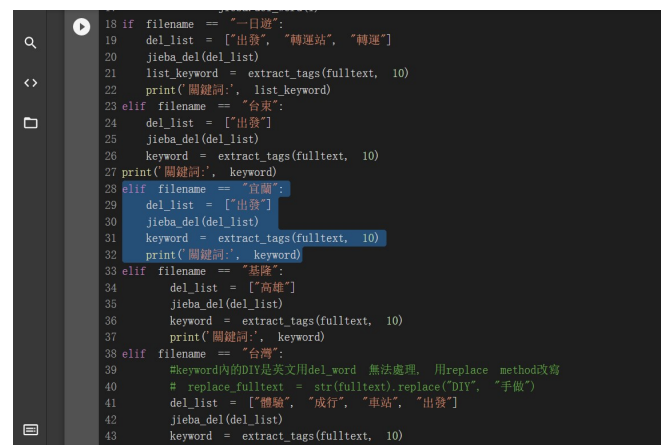


```
from jieba.analyse import extract_tags
def jieba_del(s):
    for i in s:
        jieba.del_word(i)

# del_list = ["台南", "出發", "大人"]
# jieba_del(del_list)
del_list = ["出發"]
jieba_del(del_list)
keyword = extract_tags(fulltext, 10)
print('關鍵詞:', keyword)
```

【宜蘭在地出發】北關海潮公園，大溪漁港，大里天公廟觀光巴士【宜蘭在地出發】冬
關鍵詞: ['宜蘭', '步道', '往返', '大人團仔', 'DIY', '太魯閣', '太平山', '瀑布', '台南往返', '猴洞']

把做好的 del_list 放回原文，繼續做下一個



```
18 if filename == "一日遊":
19     del_list = ["出發", "轉運站", "轉運"]
20     jieba.del(del_list)
21     list_keyword = extract_tags(fulltext, 10)
22     print('關鍵詞:', list_keyword)
23 elif filename == "台東":
24     del_list = ["出發"]
25     jieba.del(del_list)
26     keyword = extract_tags(fulltext, 10)
27     print('關鍵詞:', keyword)
28 elif filename == "宜蘭":
29     del_list = ["出發"]
30     jieba.del(del_list)
31     keyword = extract_tags(fulltext, 10)
32     print('關鍵詞:', keyword)
33 elif filename == "基隆":
34     del_list = ["高雄"]
35     jieba.del(del_list)
36     keyword = extract_tags(fulltext, 10)
37     print('關鍵詞:', keyword)
38 elif filename == "台灣":
39     # keyword內的DIY是英文用del_word 無法處理，用replace method改善
40     # replace_fulltext = str(fulltext).replace("DIY", "手做")
41     del_list = ["體驗", "成行", "車站", "出發"]
42     jieba.del(del_list)
43     keyword = extract_tags(fulltext, 10)
44     print('關鍵詞:', keyword)
```