

軟體工作

第四組

M134020043 張恩繁

M134020002 陳亭瑄

M134020018 吳信輝

B104020027 蕭宇廷

B104020028 林晏甄

M13B020007 陳學舊



目錄

1

動機與資料說明

2

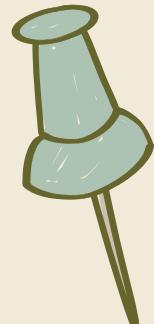
程式流程說明

3

遇到的困難&解決方式

4

結論



動機與 資料說明

01



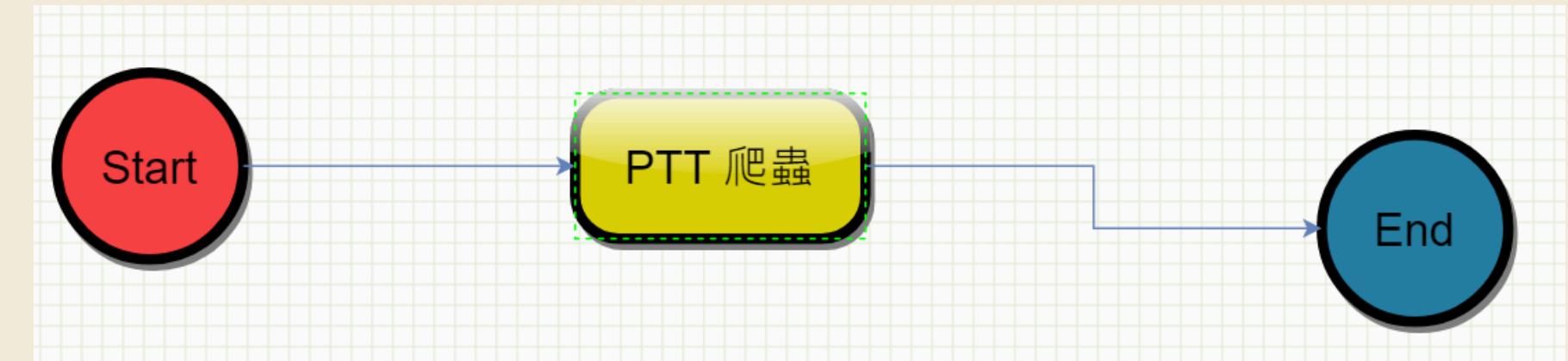
動機

畢業之後若想進入軟體領域工作有哪些該注意的地方，以及現階段可以努力和考慮的方向。



資料說明

Tarflow爬取PPT Soft_Job板資料



關鍵字

搜尋關鍵字

公司
面試
工作
能力
經驗
時間
程式

時間

搜尋起始日期

2022/01/01

搜尋結束日期

2024/12/31

資料筆數

共 2269 篇文章

2022.csv

2023.csv

2024.csv

clean_merge.csv



程式流程 說明

02



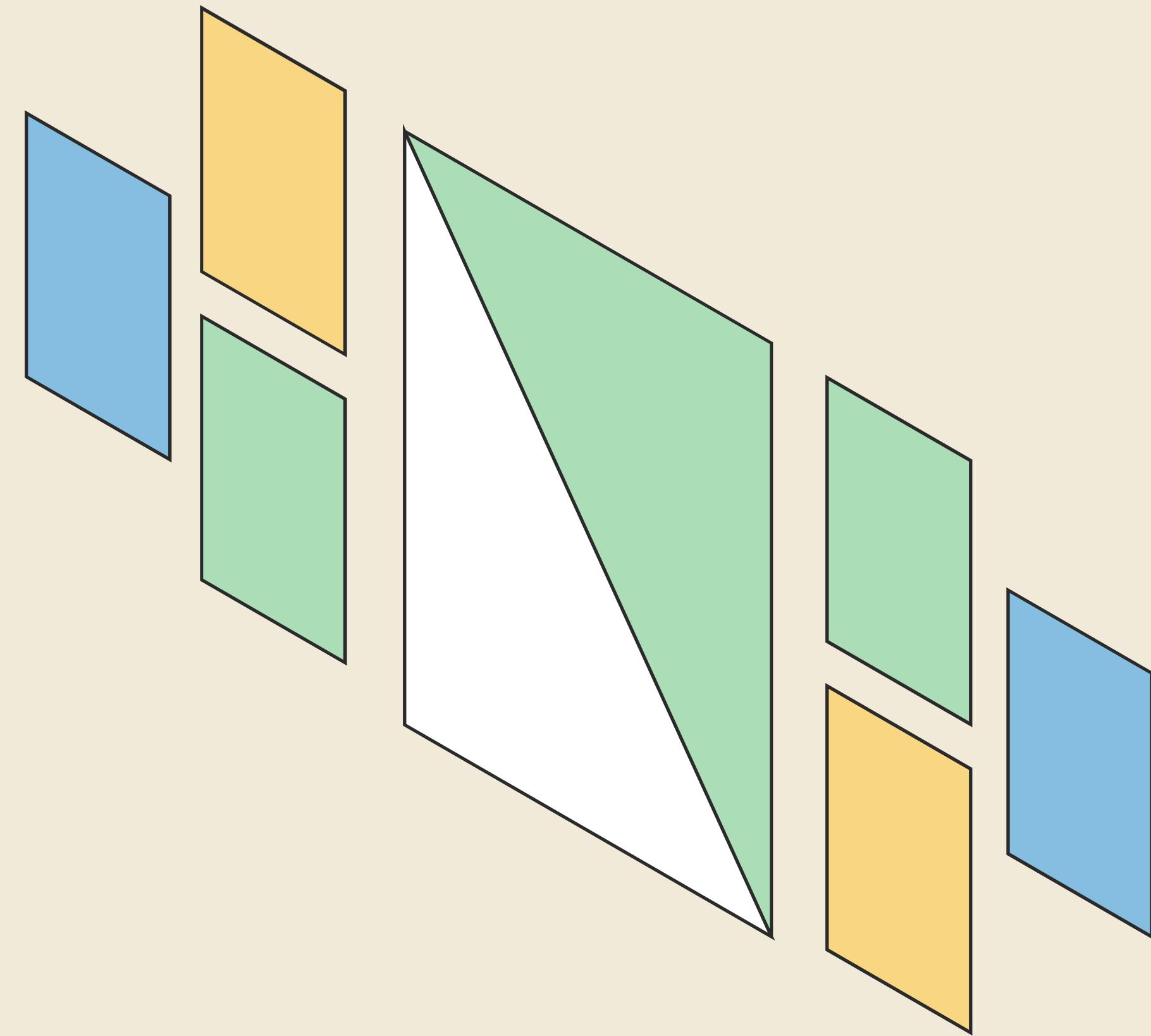
大綱

(1) 軟體工作文字雲

(2) lexicon-base情緒分析

(3) CKIP自然語言處理

(4) SnowNLP分析

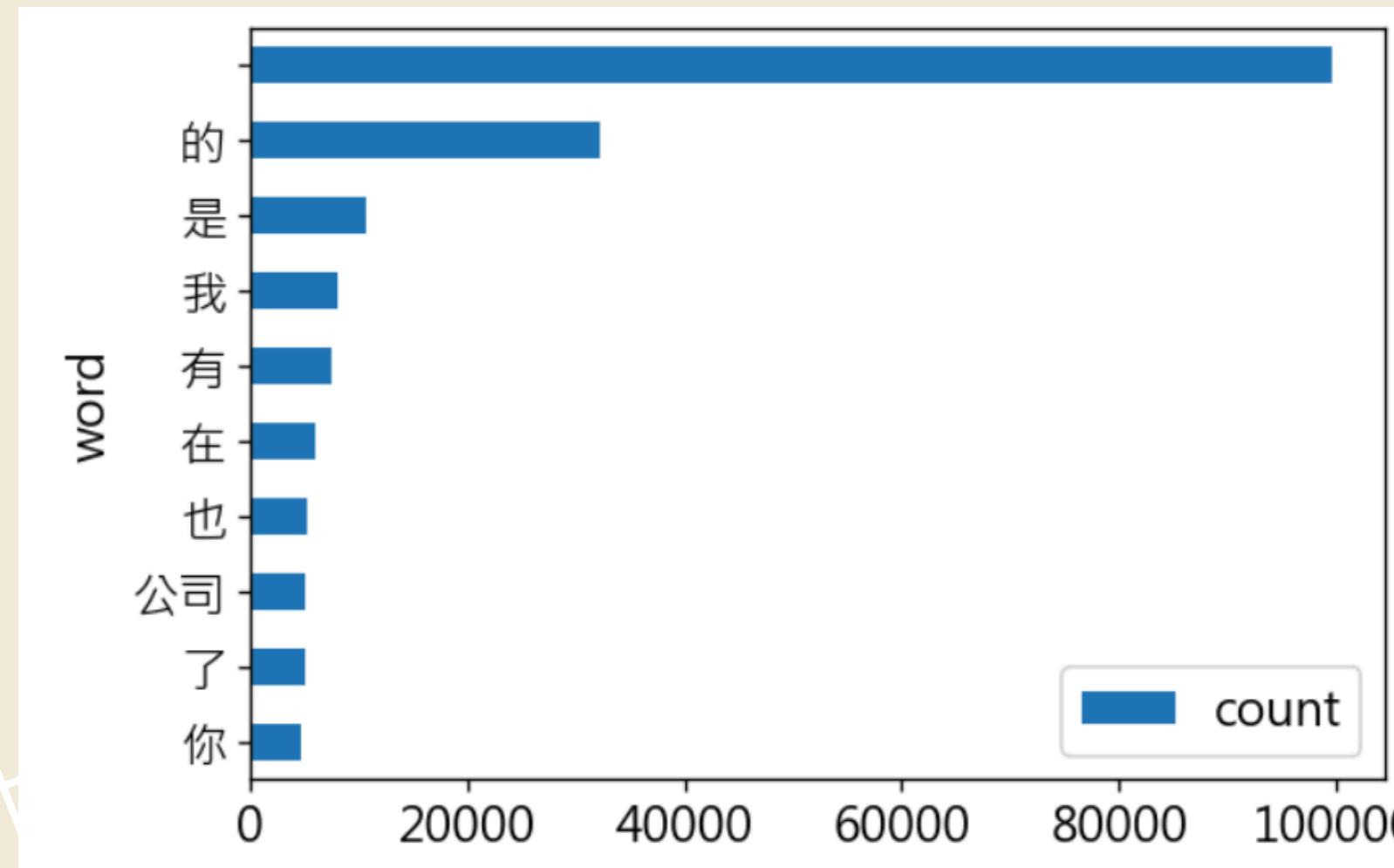
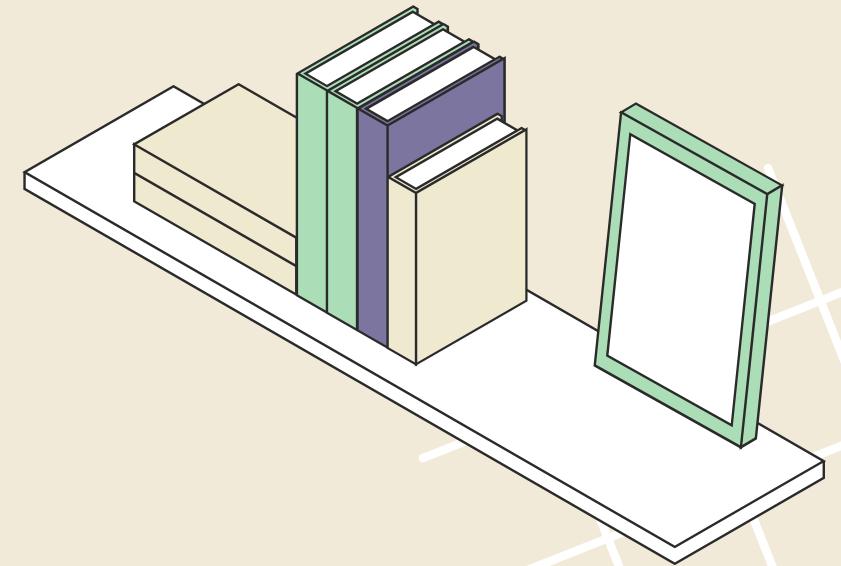




(1) 軟體工作文字雲

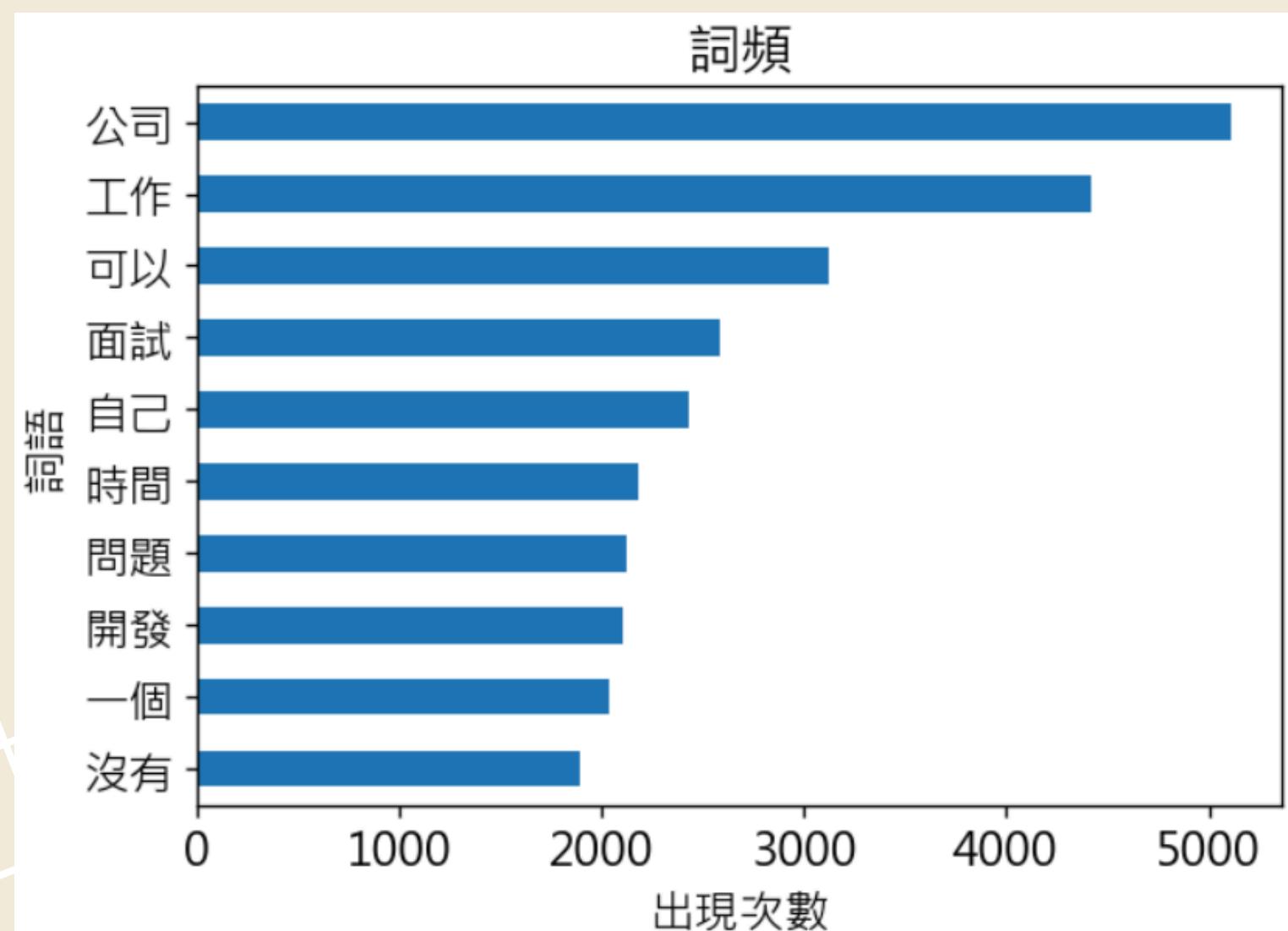
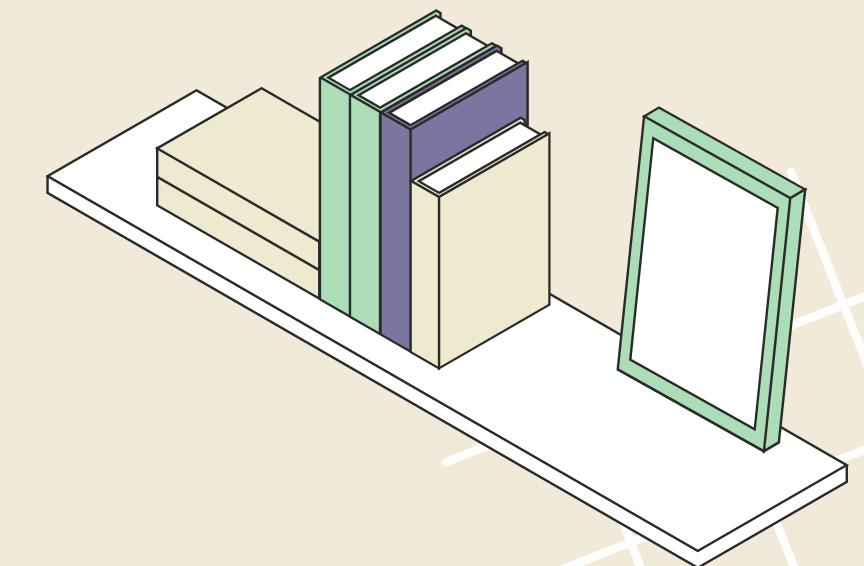
- 資料前處理
- 斷句
- 斷詞
- 計算去除停用字及更新字典
前後的詞頻並視覺化結果

初步斷句與斷詞後（去除停用字前）



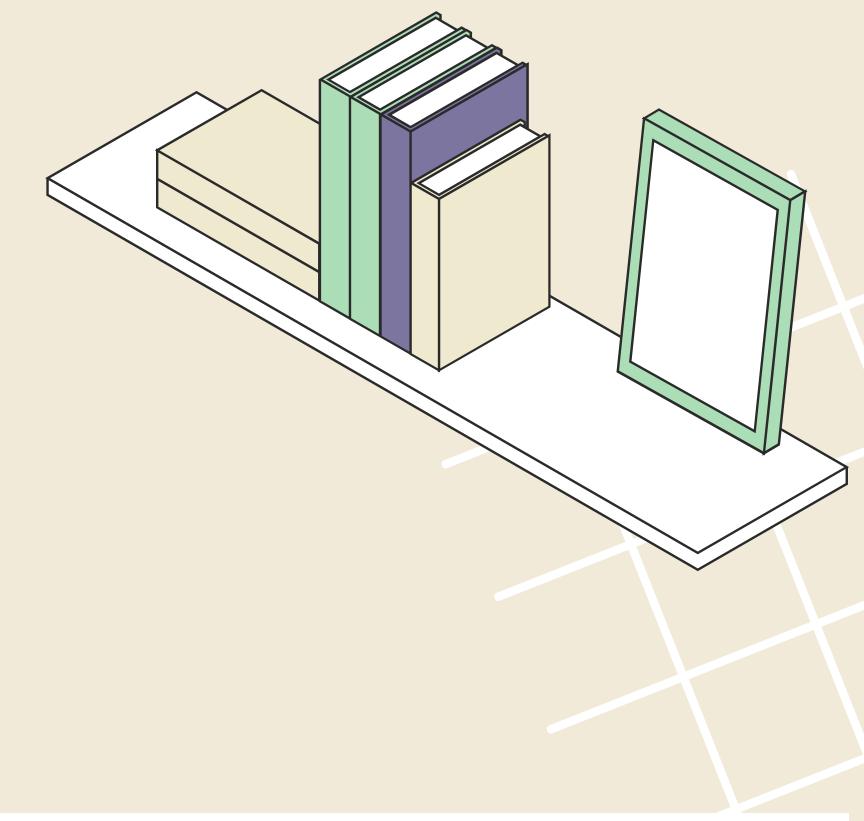
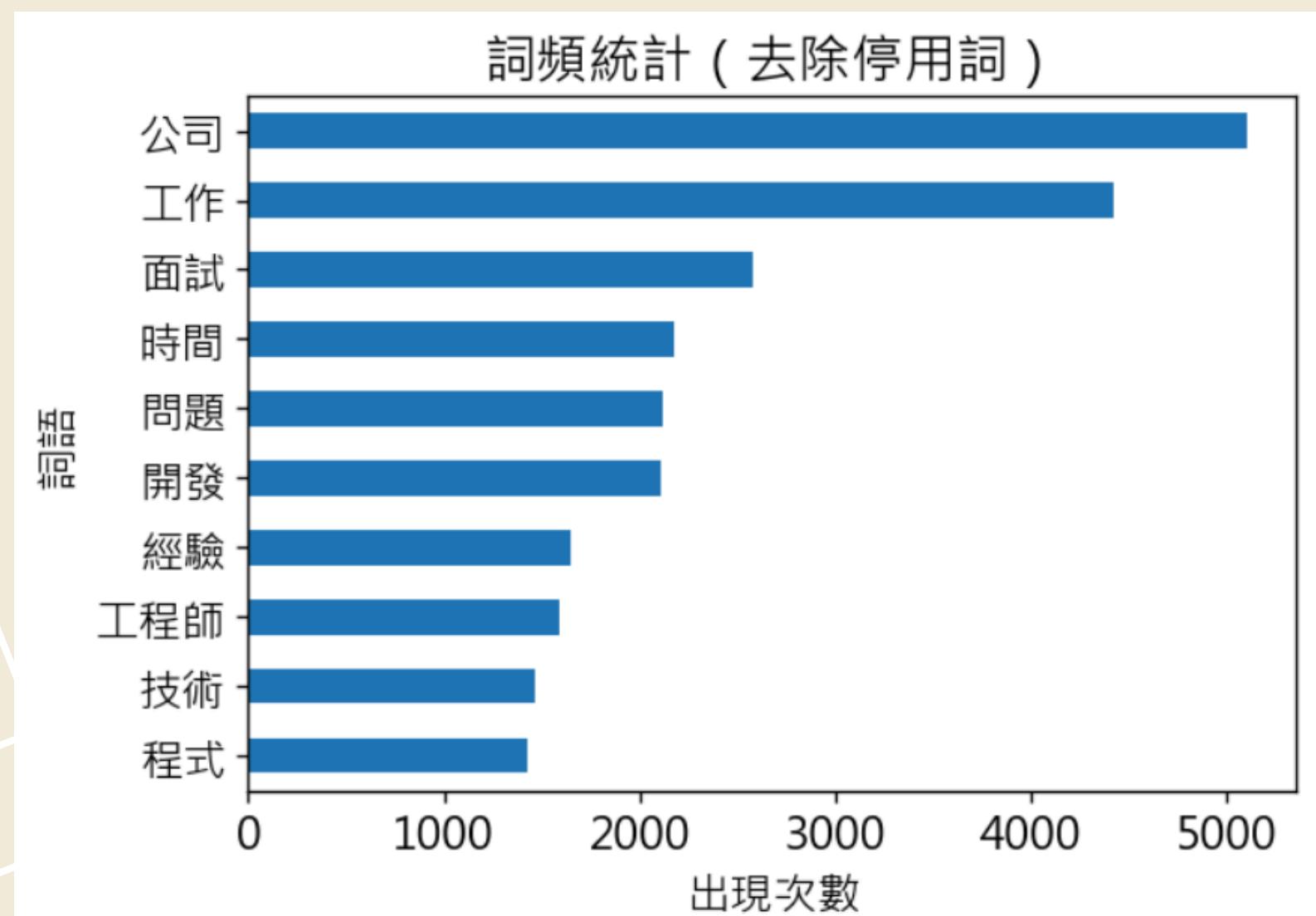
工作對時間自己工程師經驗上
寫就是公司都就會¹你公司
的去說如果跟人了想試
好個很來個他沒有因為所以²
但開發在也問題做不面後

去除少於兩個字的詞語

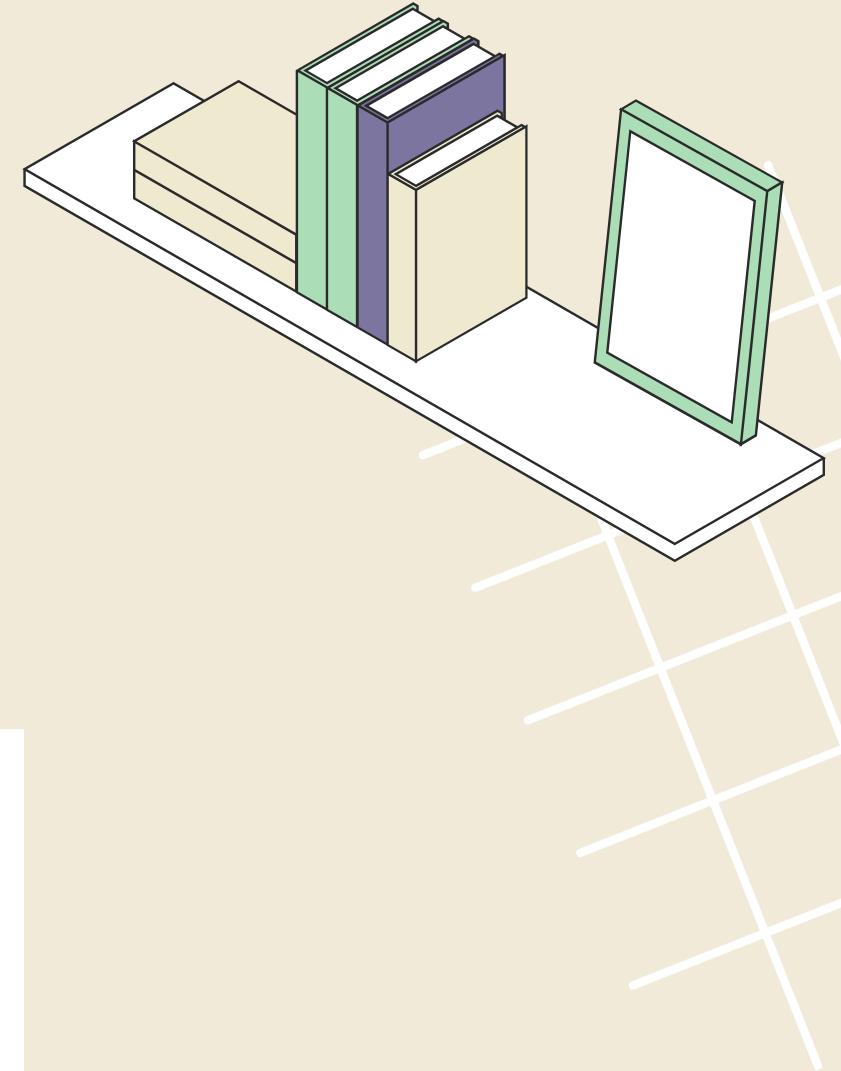


如果上工我們一時的能力很多使用這個大為的技術開發出來的系統真的現在已經開始覺得內容可謂是相當的複雜，因為我們在這幾年來的經驗累積，已經能夠很輕易地解決問題了。當然，這並不是說我們沒有問題，而是說我們已經有了很多的經驗，能夠很輕易地解決問題了。當然，這並不是說我們沒有問題，而是說我們已經有了很多的經驗，能夠很輕易地解決問題了。

移除通用停用字



更新停用字後



結果比較

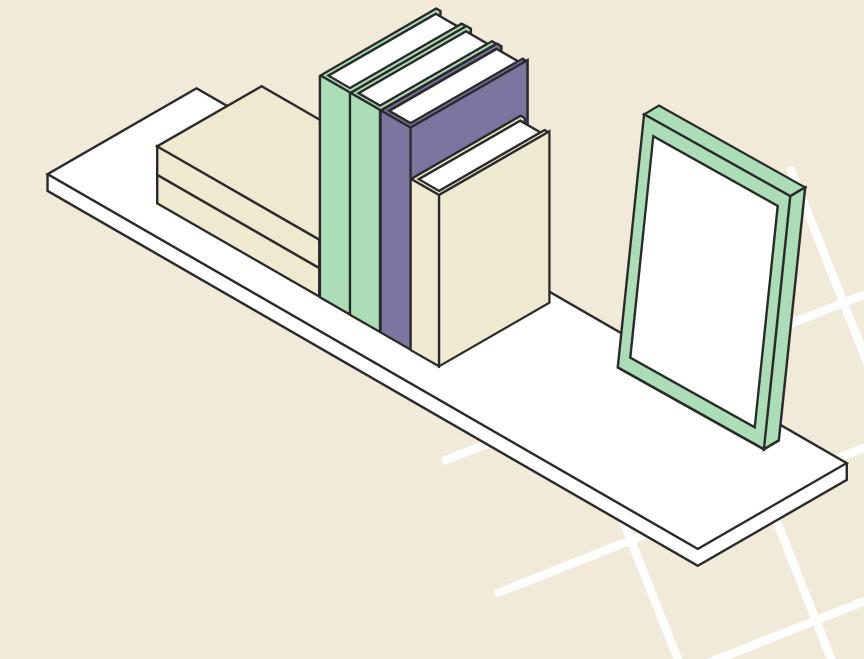


A photograph of a white laptop with its screen off, resting on a light-colored wooden desk. In front of the laptop, there is a closed notebook and a black pen. The background is slightly blurred, showing more of the desk area.

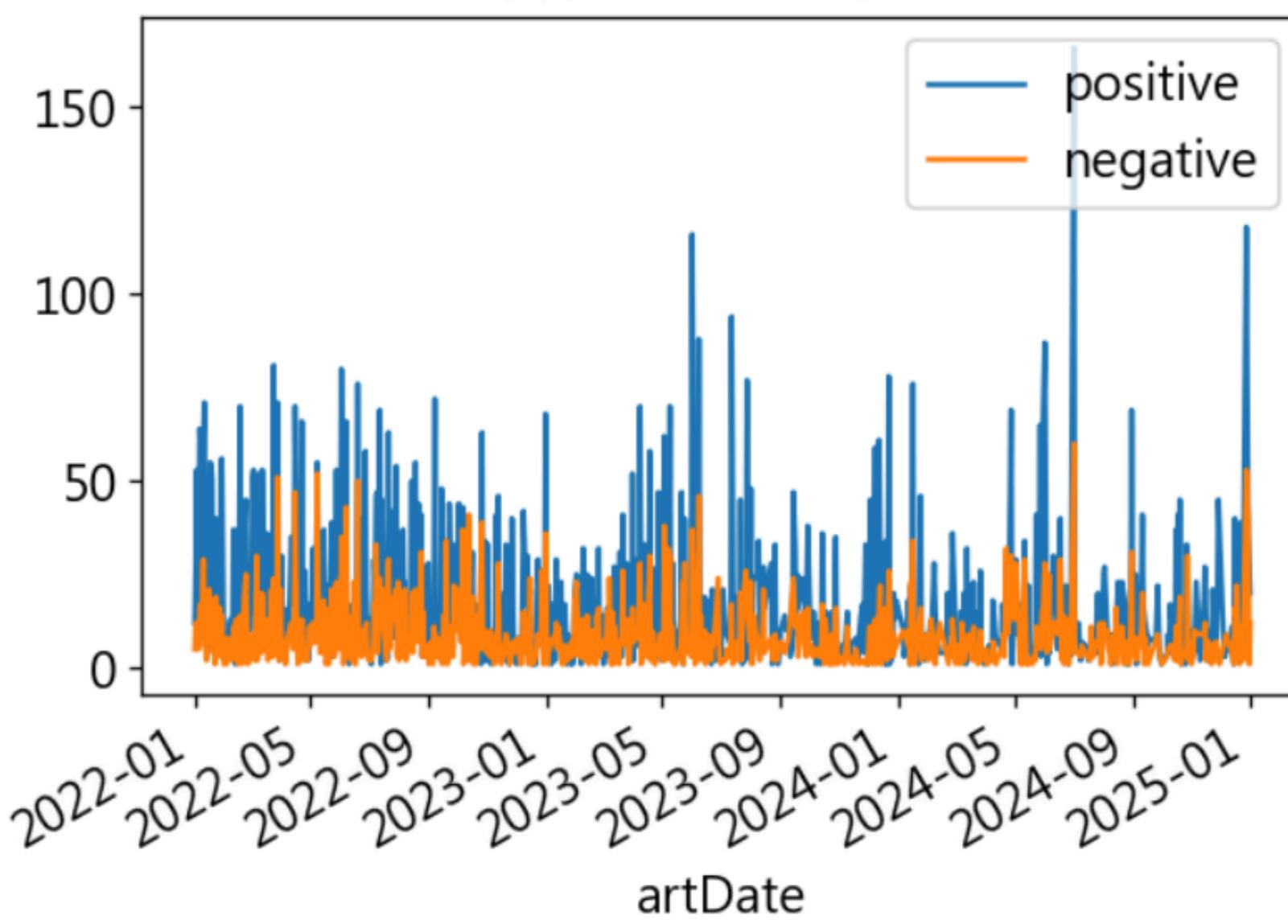
(2) lexicon-base情緒分析

- 準備LIWC字典
- 情緒字典與斷詞結果進行合併
- 正負向情緒詞彙頻率折線圖
- 找出情緒代表字
- 檢查情緒詞是否符合本文本

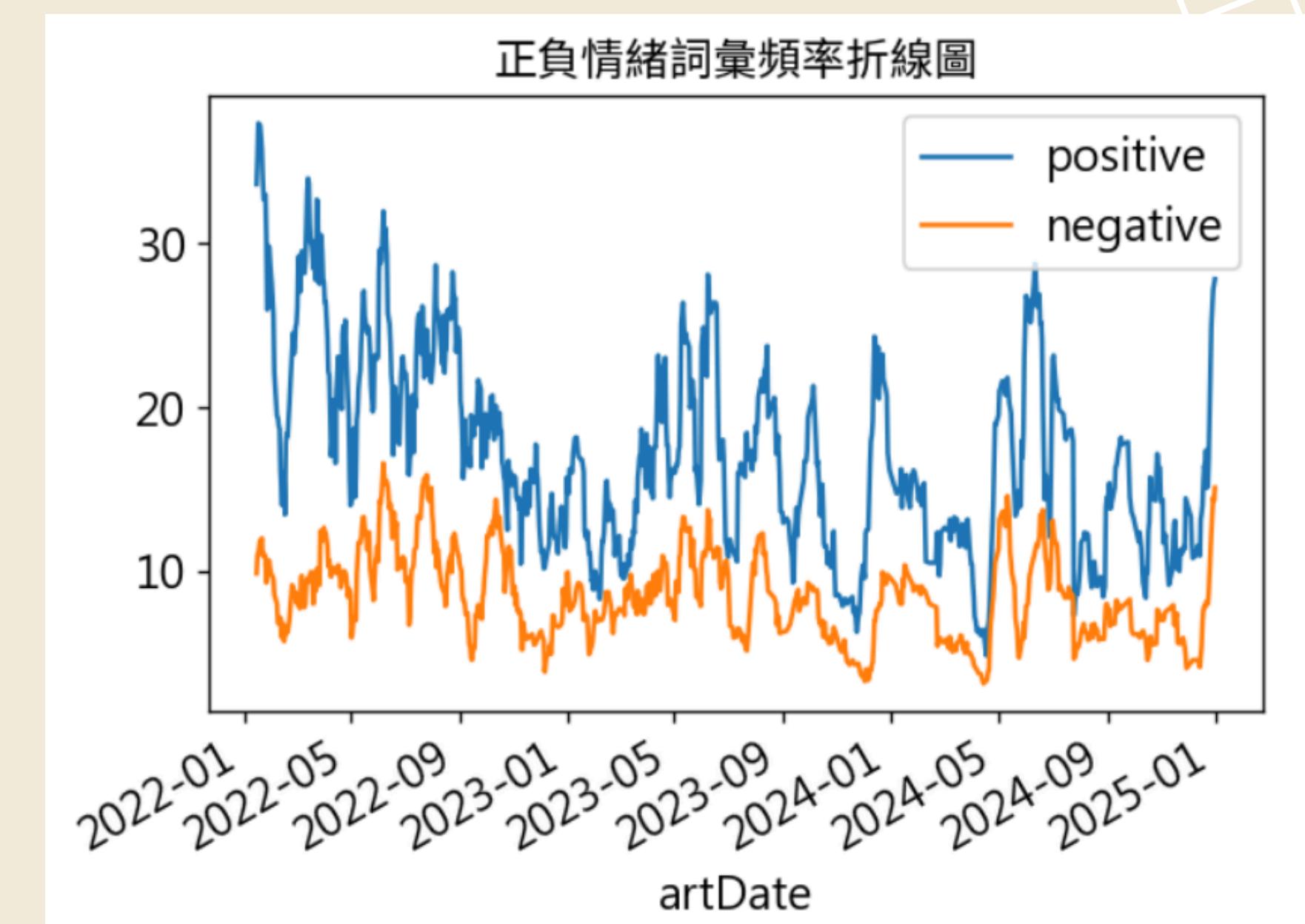
正負情緒詞彙頻率折線圖



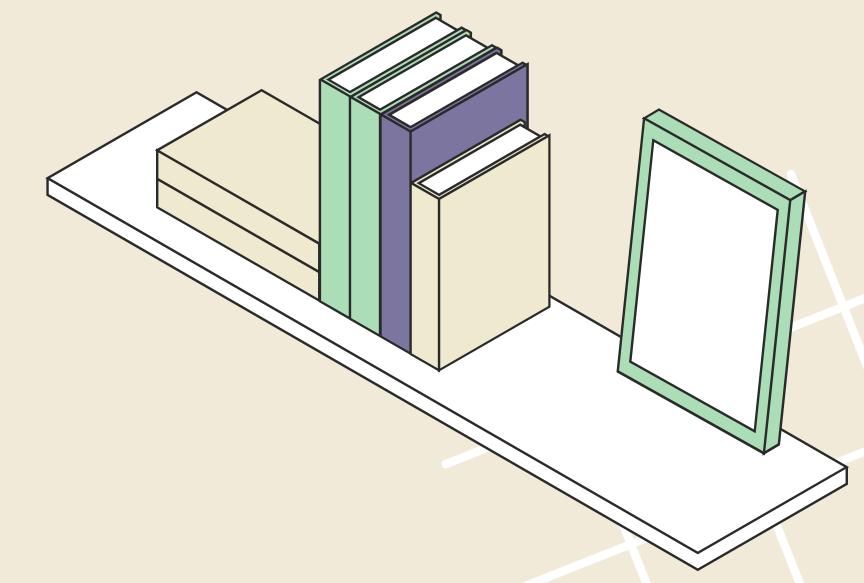
正負情緒詞彙頻率折線圖



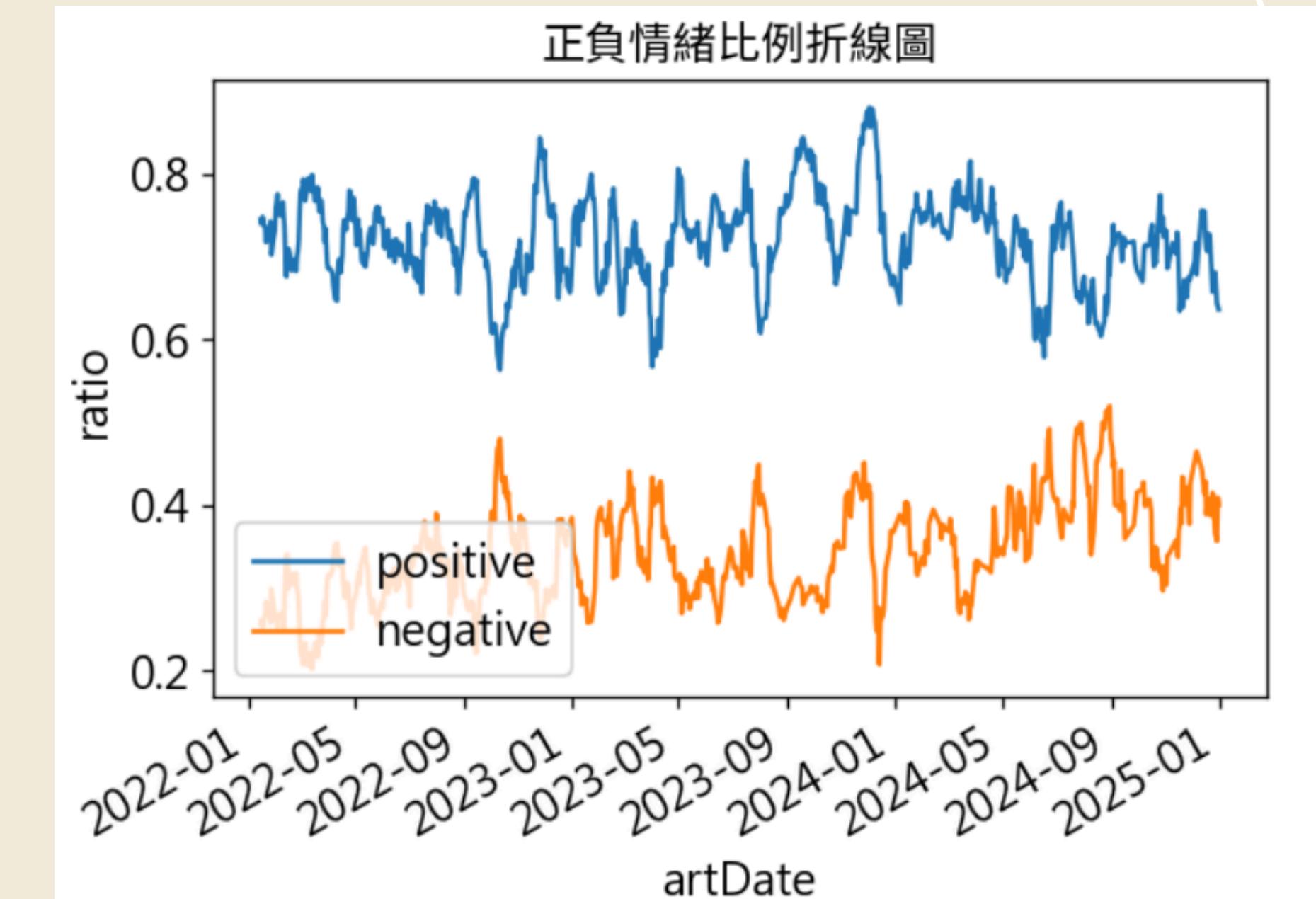
正負情緒詞彙頻率折線圖



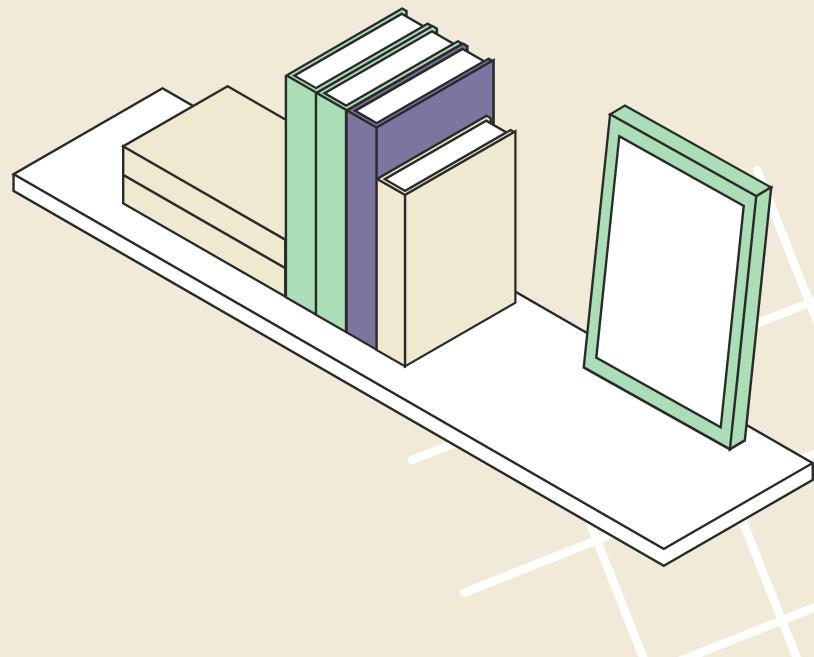
正負情緒的比例



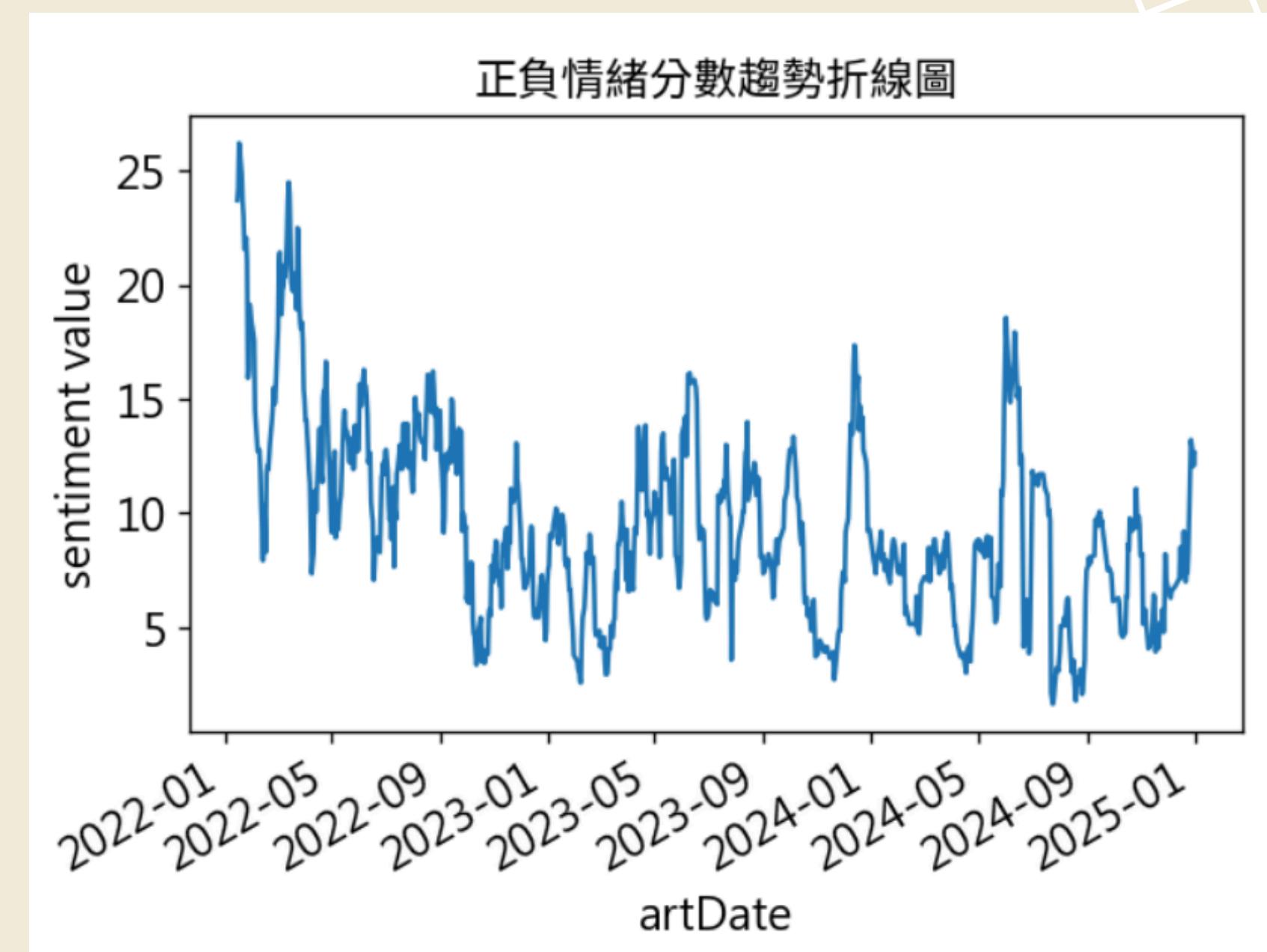
	artDate	sentiments	size	ratio
1	2022-01-01	negative	5	0.294118
2	2022-01-01	positive	12	0.705882
6	2022-01-02	negative	12	0.315789
7	2022-01-02	positive	26	0.684211
11	2022-01-03	negative	7	0.116667
12	2022-01-03	positive	53	0.883333
15	2022-01-04	negative	7	0.280000
16	2022-01-04	positive	18	0.720000
19	2022-01-05	negative	5	0.161290
20	2022-01-05	positive	26	0.838710

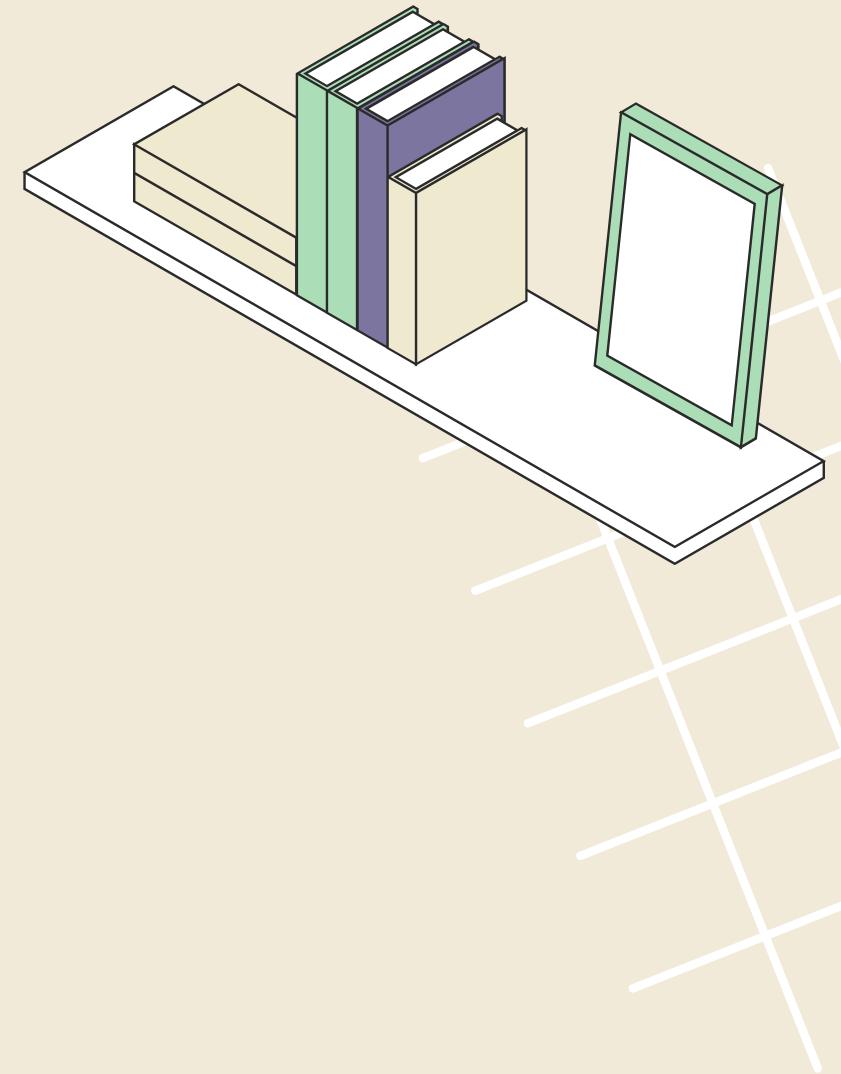


正負向情緒分數



	artDate	negative	positive	sentiment_value
0	2022-01-01	5.0	12.0	7.0
1	2022-01-02	12.0	26.0	14.0
2	2022-01-03	7.0	53.0	46.0
3	2022-01-04	7.0	18.0	11.0
4	2022-01-05	5.0	26.0	21.0

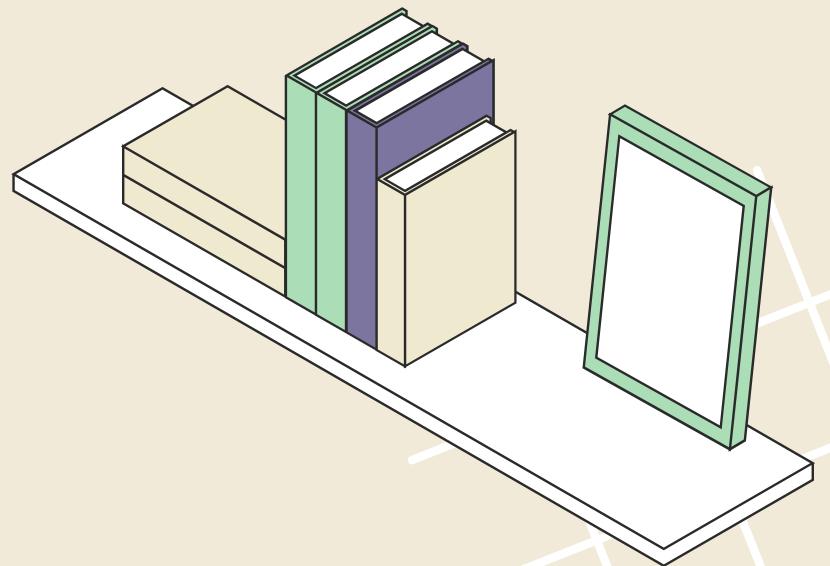




文章為單位的情緒分析

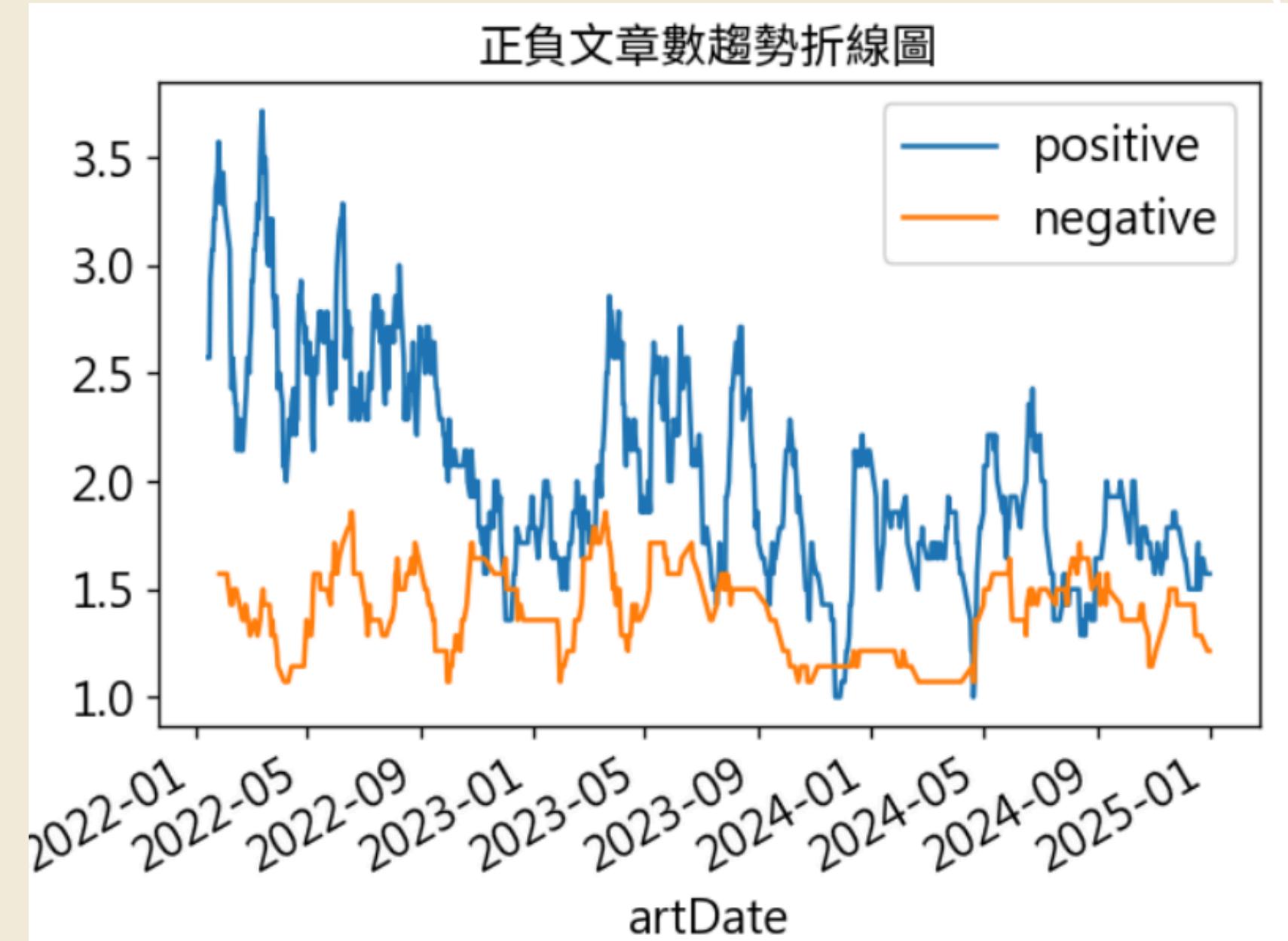
- 計算每篇文章的正負情緒詞數
- 將文章分類為「正向 or 負向」
- 計算每日正負向文章數量

文章為單位的情緒分析

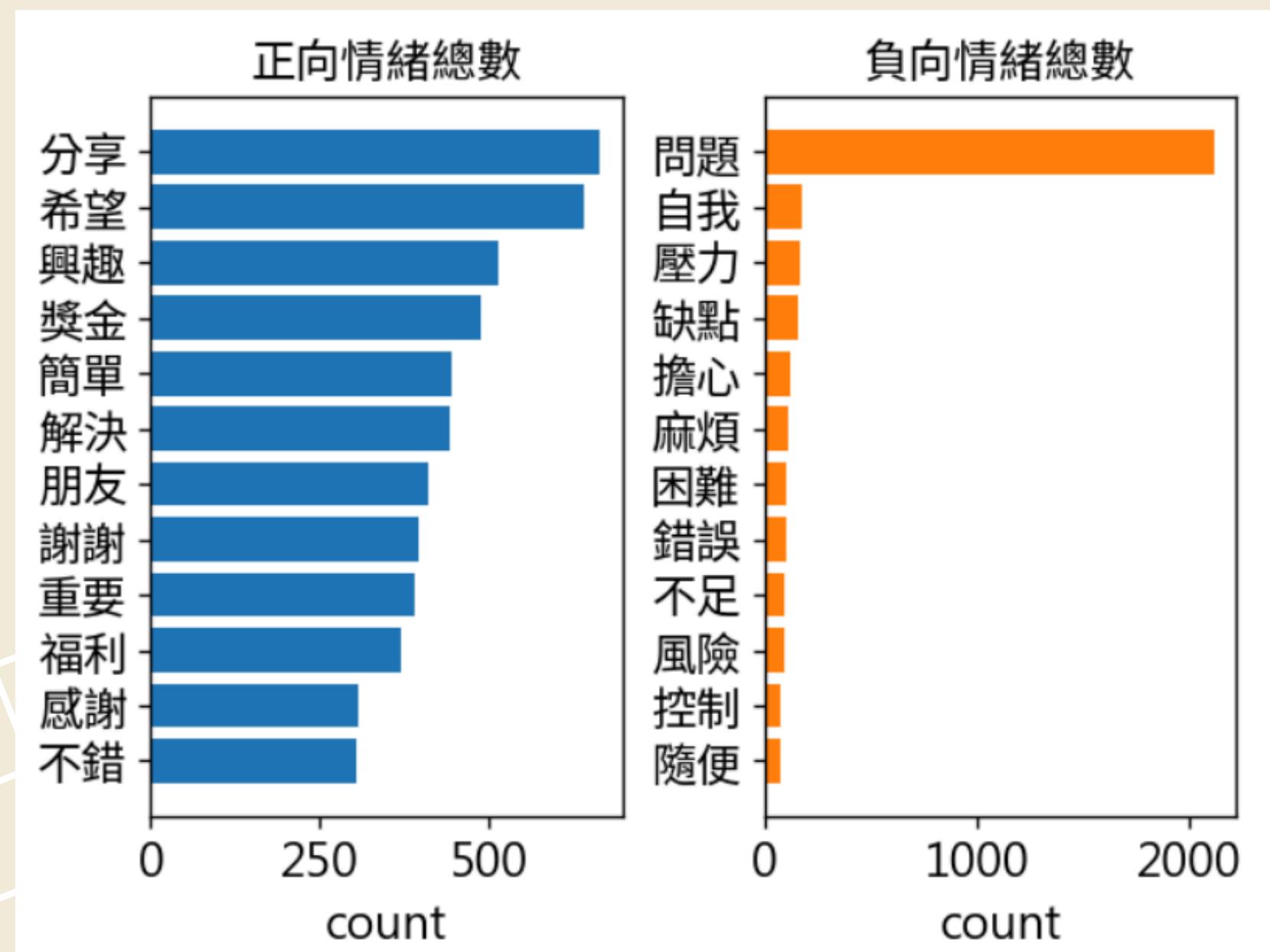


	artDate	sentiment_class	size
0	2022-01-01	正向	3
1	2022-01-02	正向	2
2	2022-01-03	正向	3
3	2022-01-04	正向	2
4	2022-01-05	正向	2
...
1157	2024-12-29	正向	1
1158	2024-12-29	負向	1
1159	2024-12-30	負向	1
1160	2024-12-31	正向	1
1161	2024-12-31	負向	1

1162 rows × 3 columns



情緒代表字



她先自我介紹了一下 並說明該部門PM要做的事情

=====

輪到我則是自我介紹了一下 並說明過去PM的經歷跟做事情的方式

=====

自我學習的動力也很重要

=====

有時會自我懷疑是否需辭掉工作去進修

=====

一開始自我介紹

=====

自我介紹

=====

一開始先自我介紹

=====

在只準備了自我介紹的情況下

=====

也會在後續提薪資的時候產生自我懷疑

=====

我自我介紹工作內容做過的專案會哪些技術

=====

自我感覺非常良好的公司

=====

但我們知道提升自我的這條路上充滿了許

=====

陪伴人們在自我成長這條路上前進

...

之後請我用英文自我介

=====

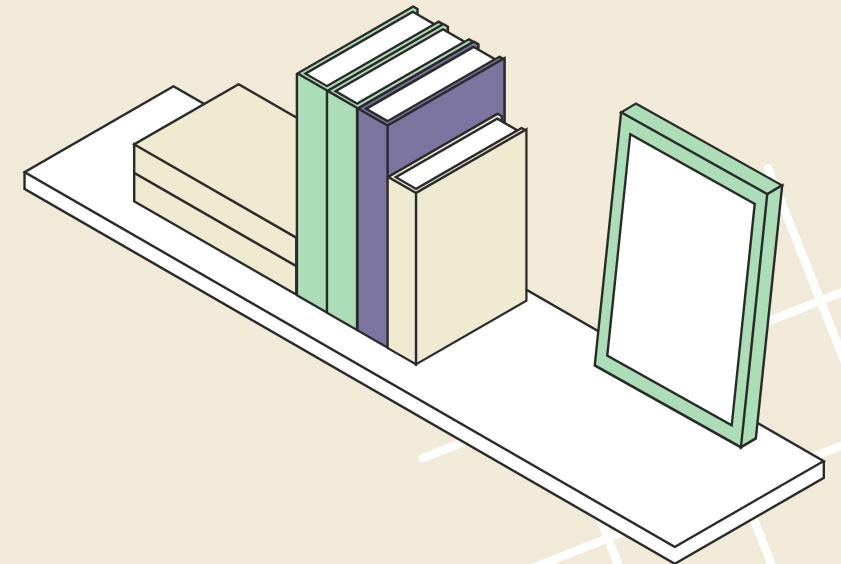
先自我介紹



(2) lexicon-base情緒分析(續)

- 設定要移除的字詞列表
- 繪製情緒分數趨勢圖
- 文字雲(2022-01 ~2022-06 在情緒分數趨勢圖上顯示較高的分數)

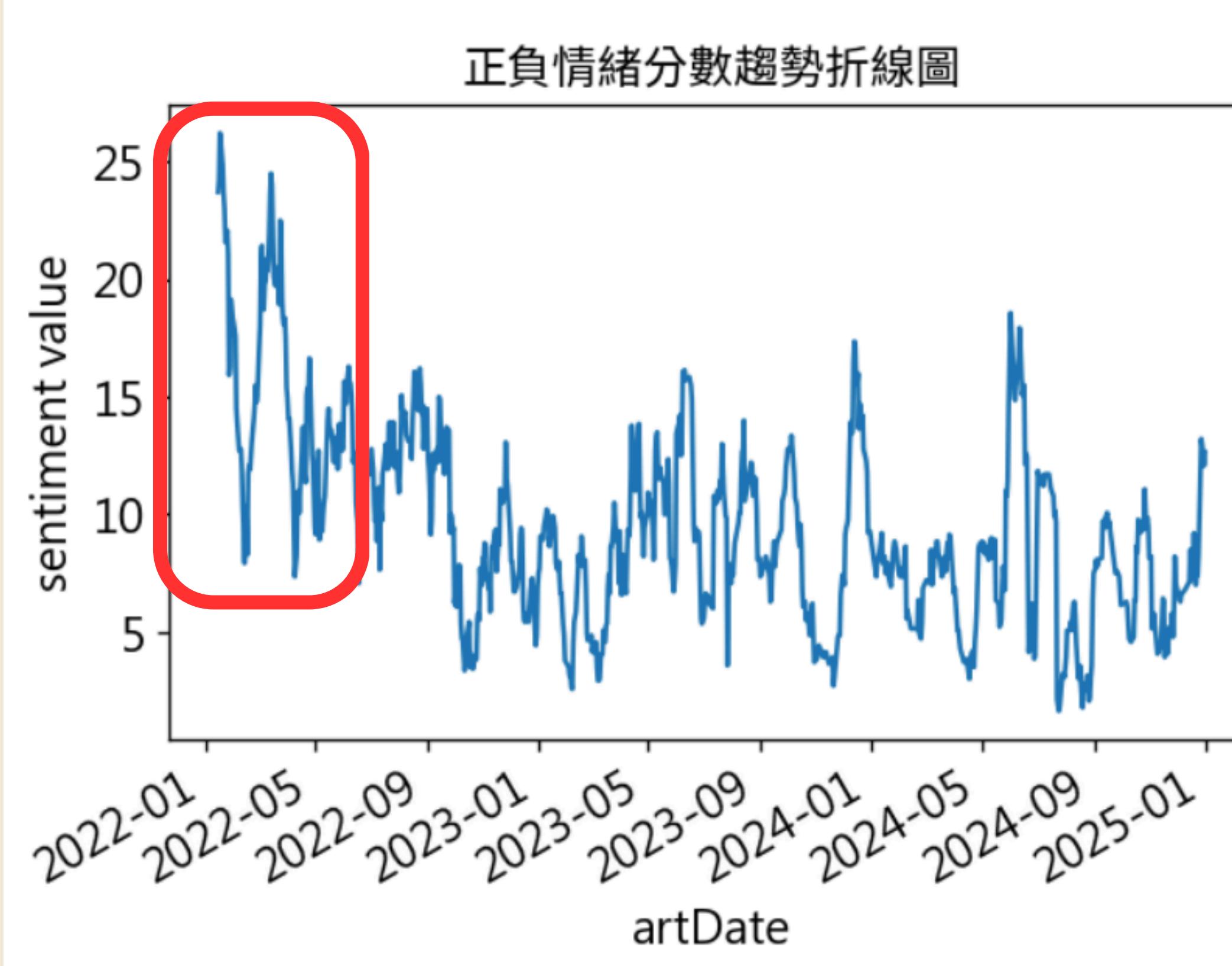




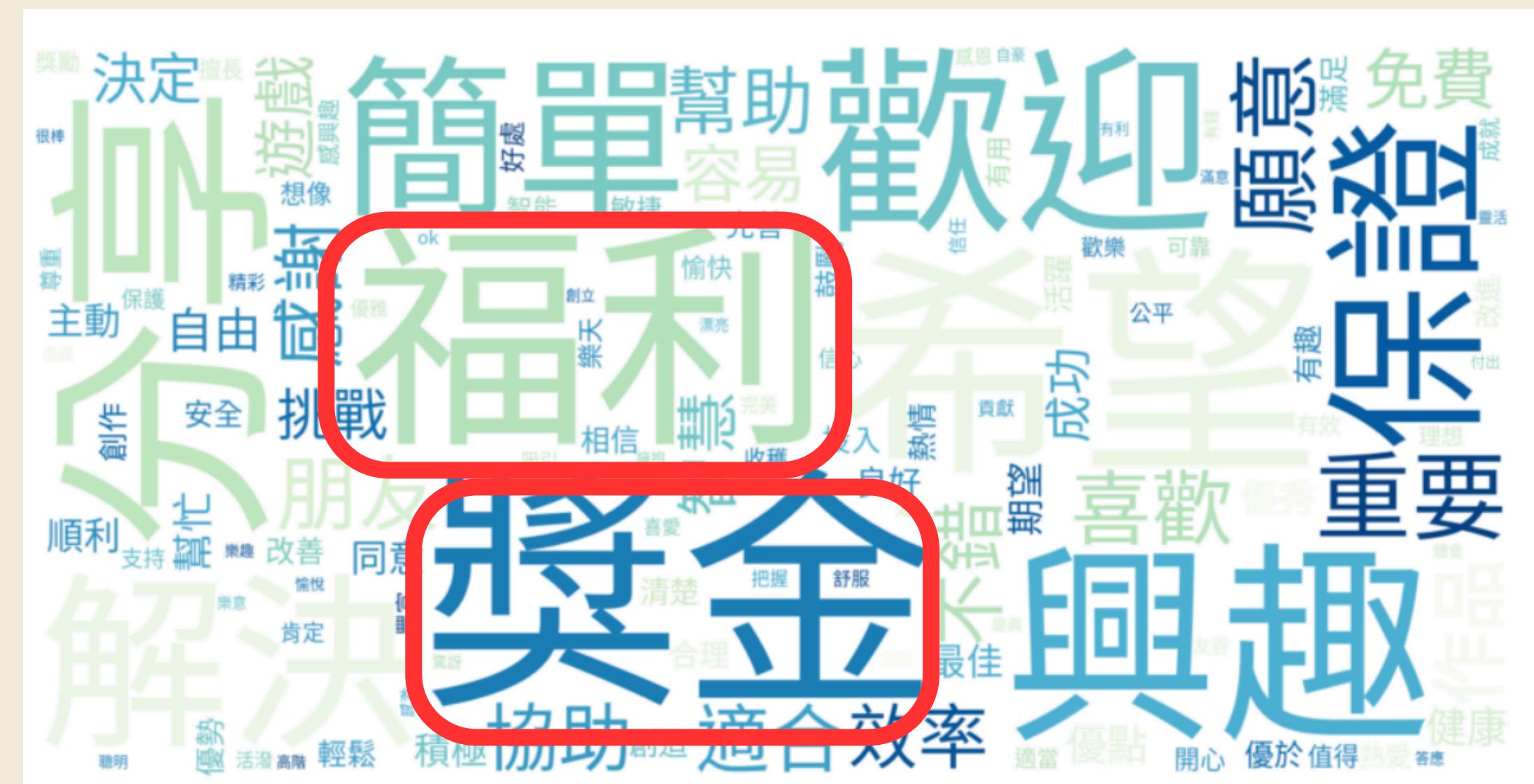
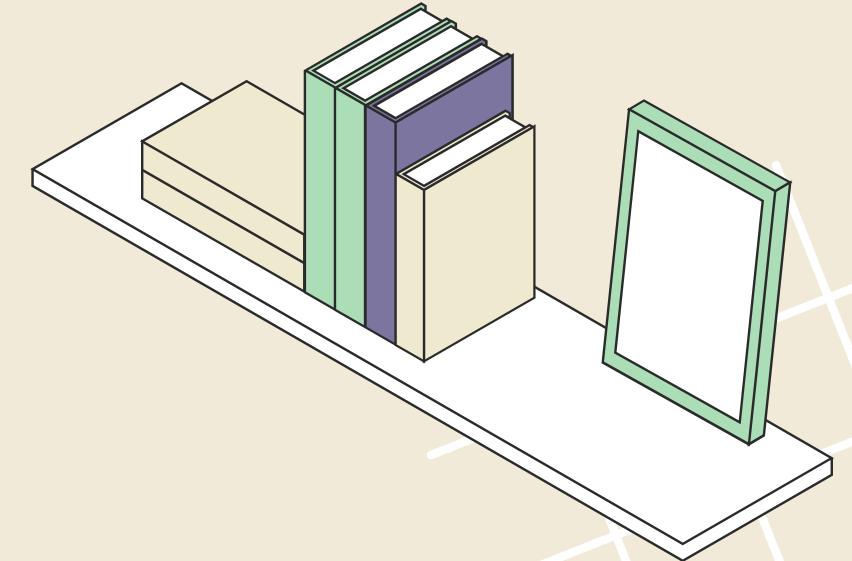
刪除詞彙

```
# 設定要移除的字詞列表  
words_to_remove = ['隨便', '自我', '謝謝', '確定']
```

分析在情緒分數趨勢圖上顯示較高分數的資料



正向文字雲



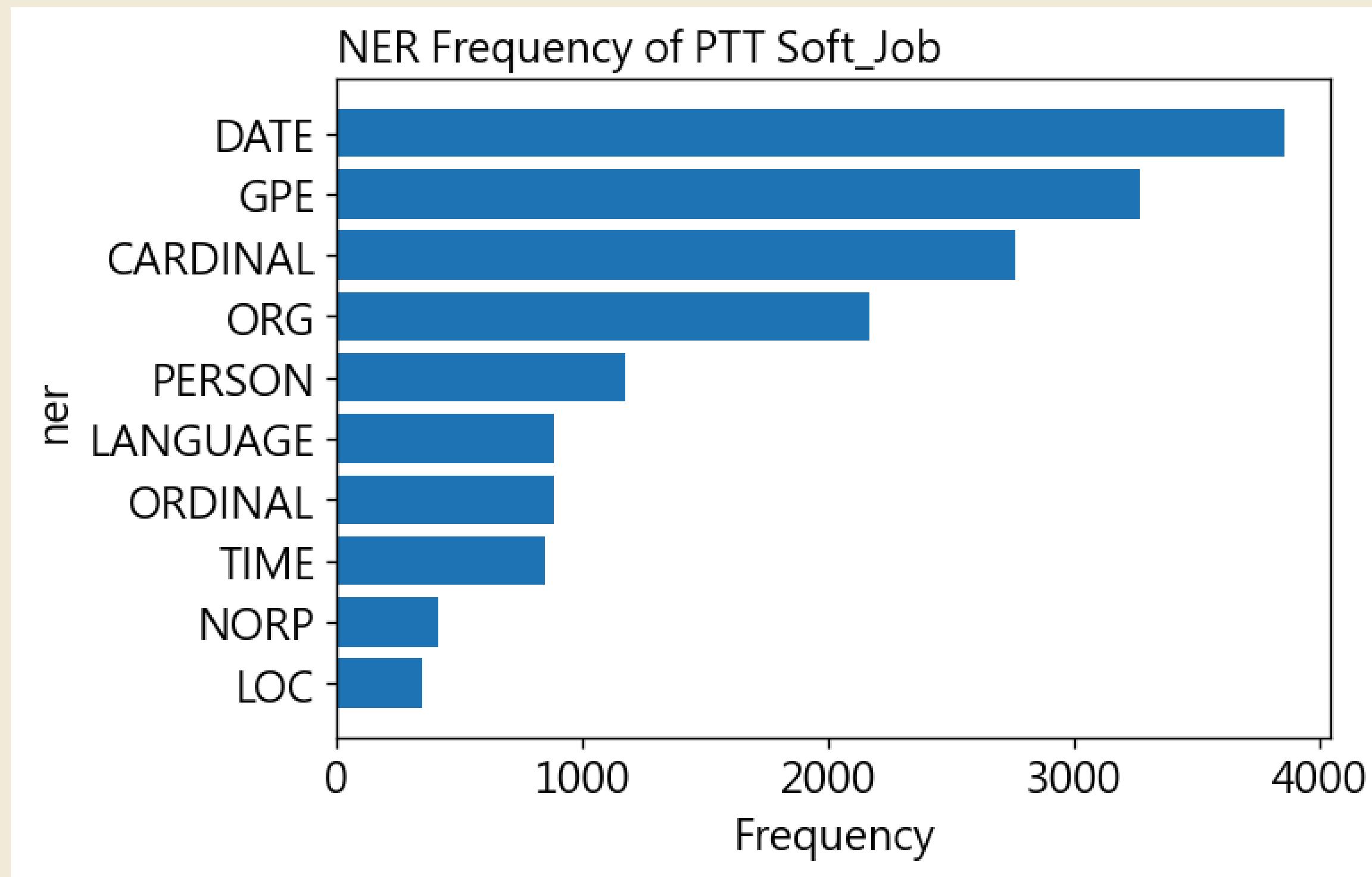


(3) CKIP自然語言處理

- 探索分析 - NER
- 統計整體 NER 類別次數
- 分析特定 NER 類別：ORG
- 分析特定 NER 類別：LANGUAGE
- 搜尋文章內含「英文」的貼文

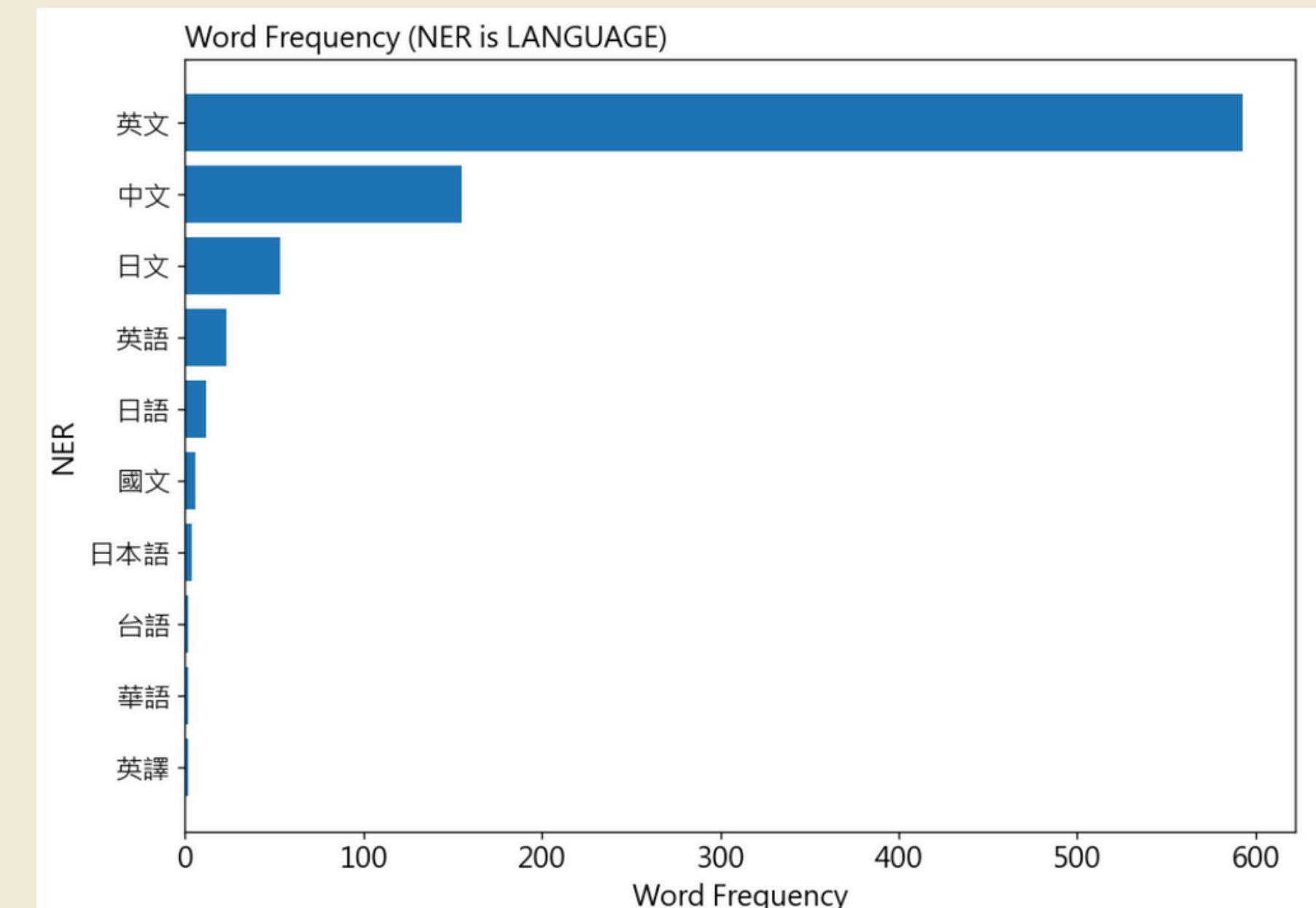
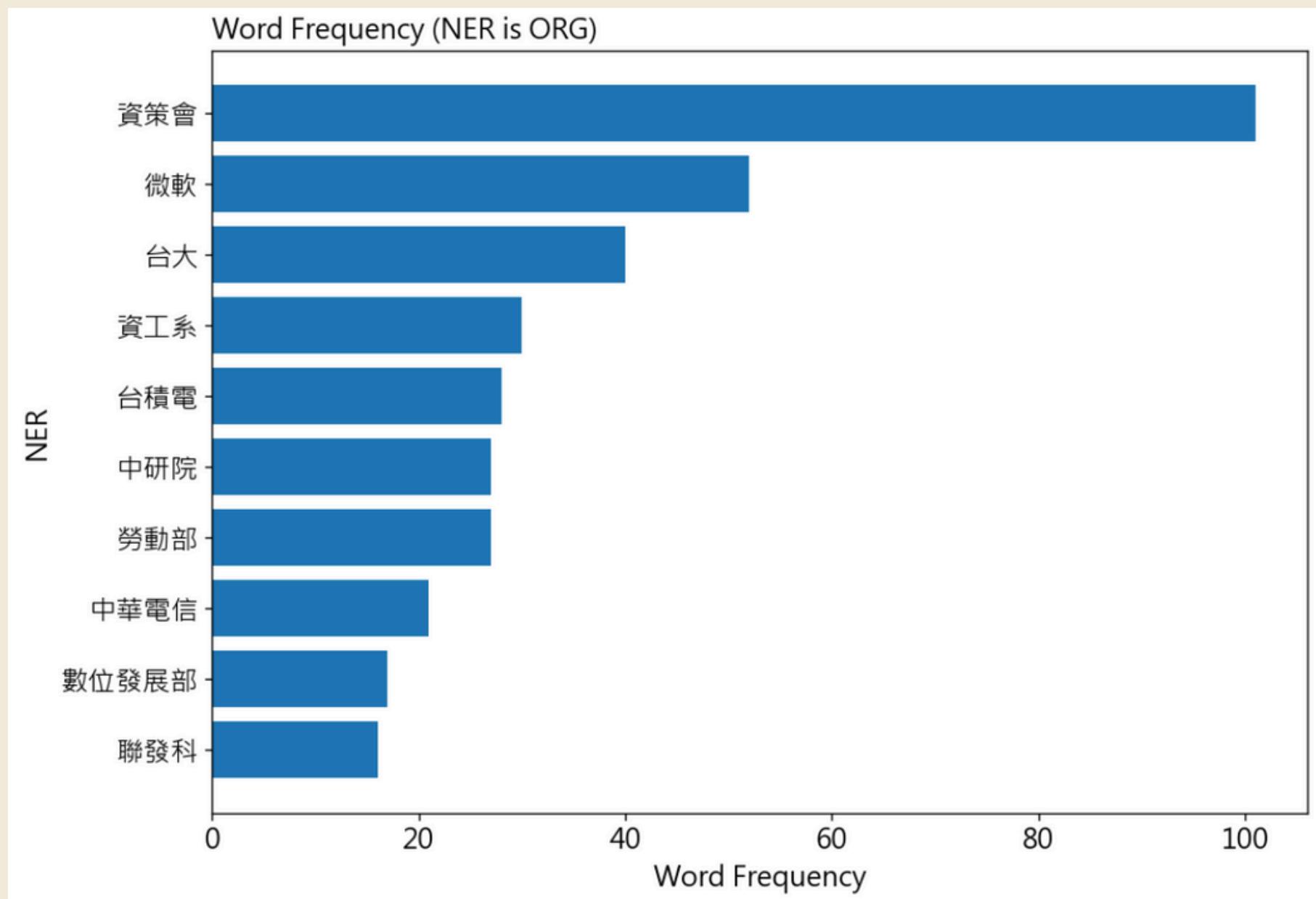


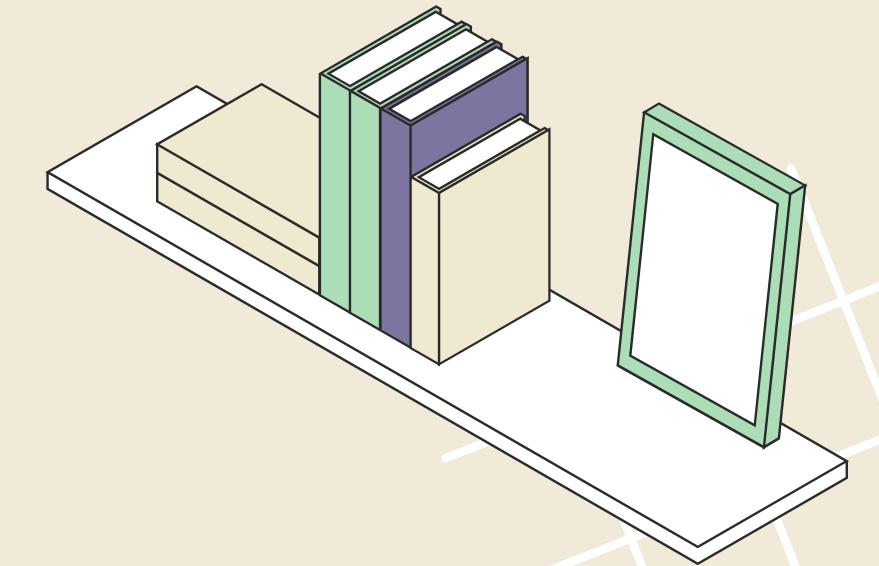
NER分析



ORG

LANGUAGE



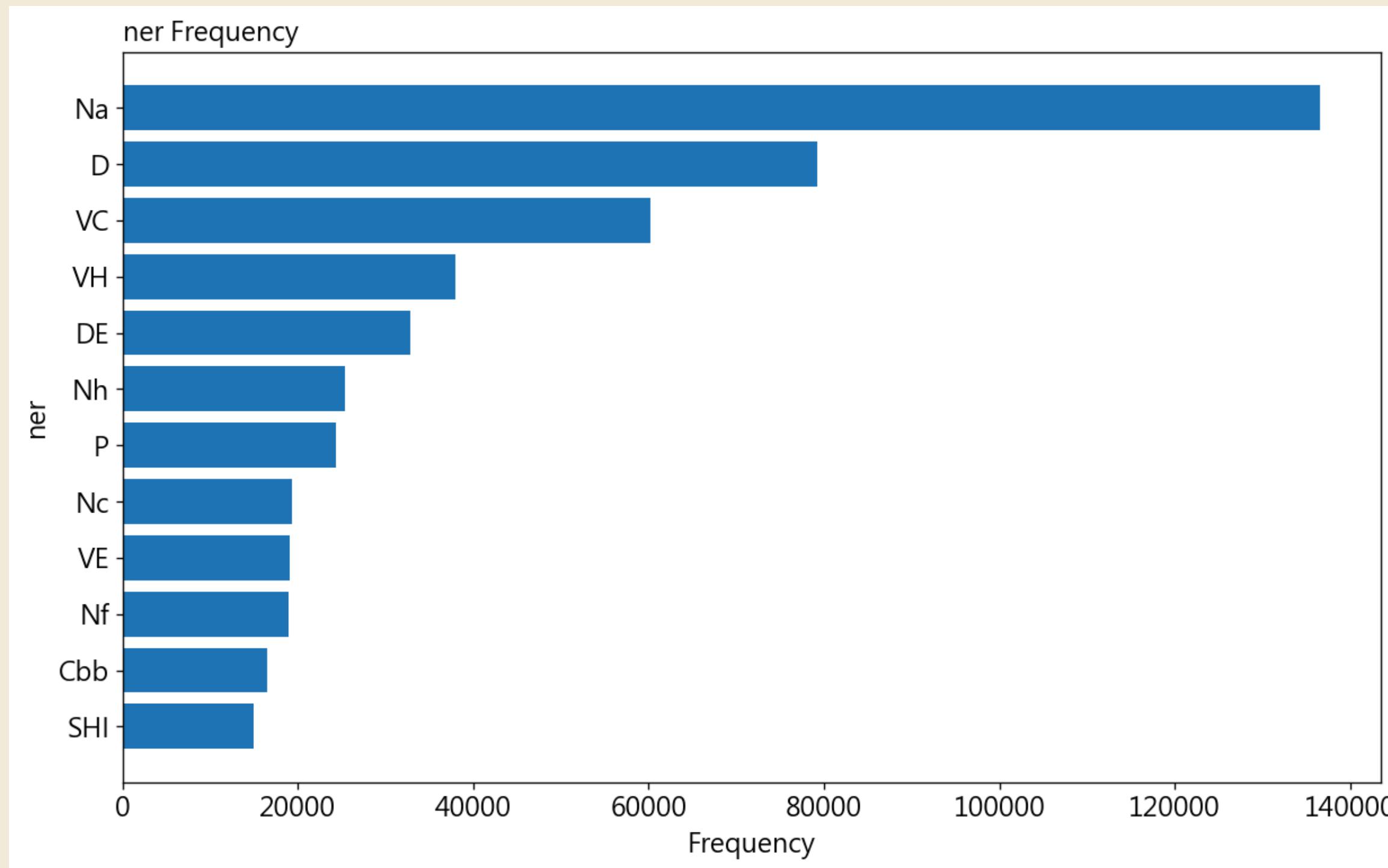


(3) CKIP自然語言處理(續)

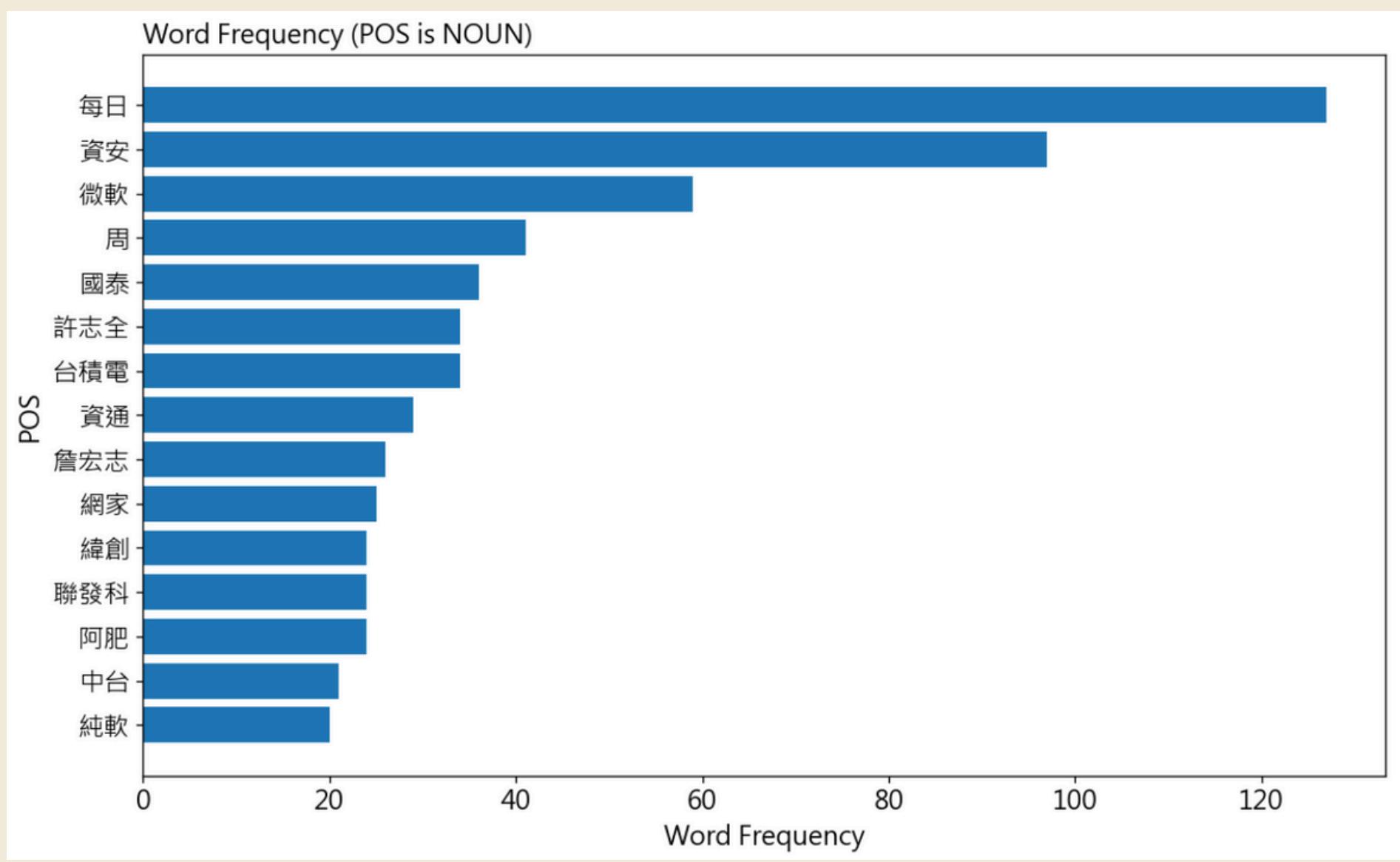
- 探索分析 - POS
- 統計 POS 種類出現次數
- 「專有名詞 Nb」進行詞頻分析
- 「動詞 VC」進行詞頻分析



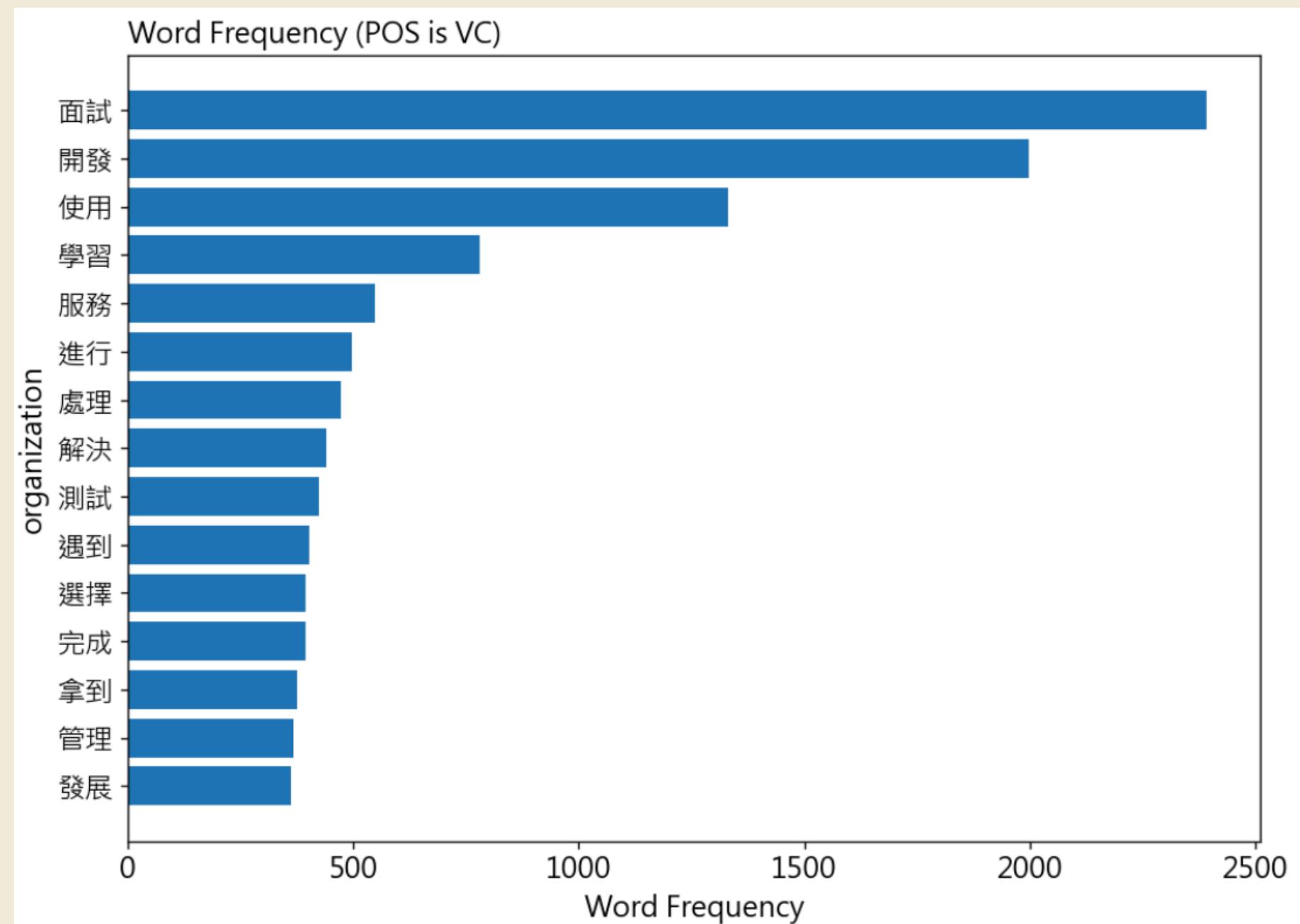
POS分析



專有名詞 Nb



動詞 VC



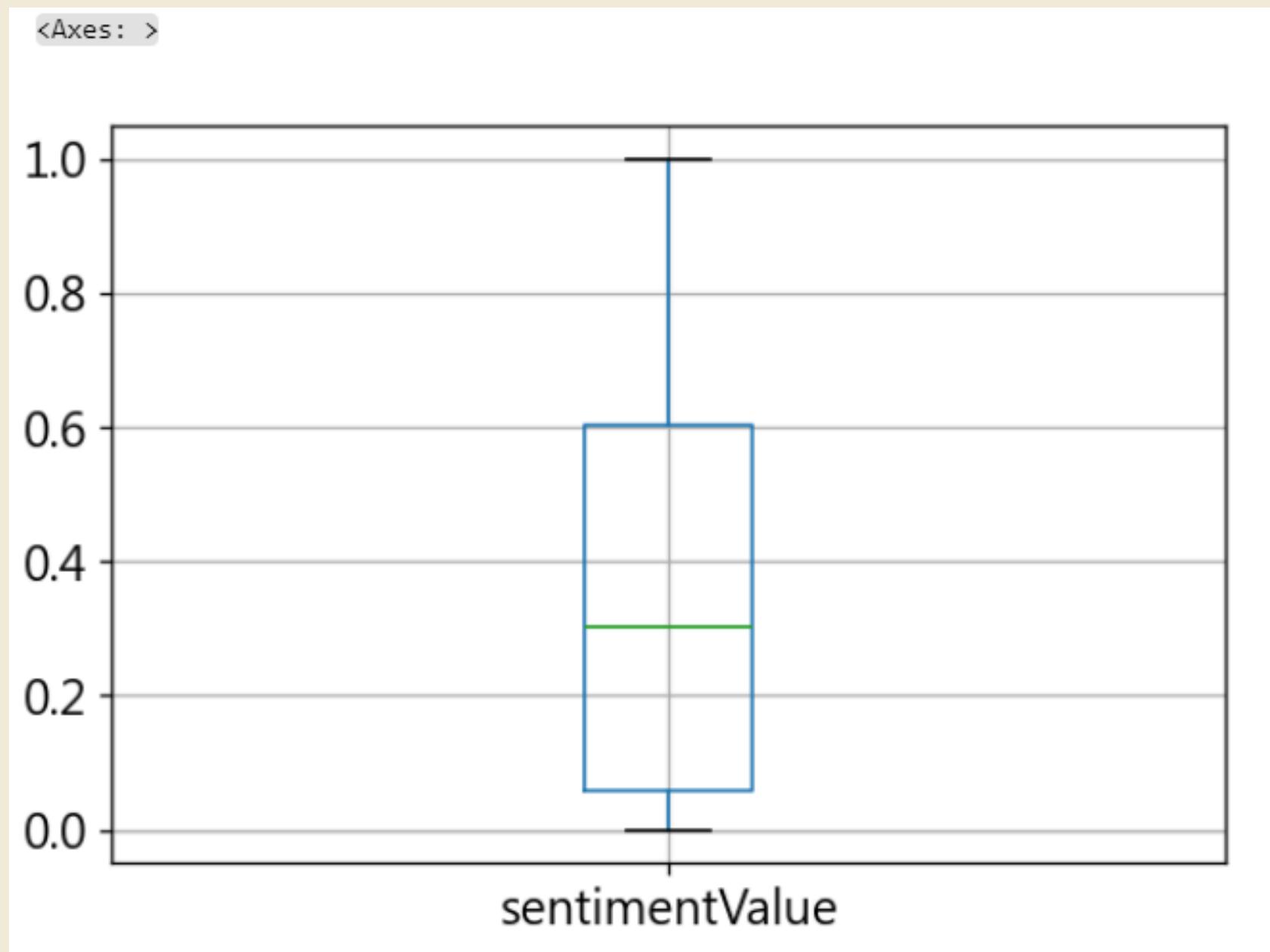


(4) SnowNLP分析

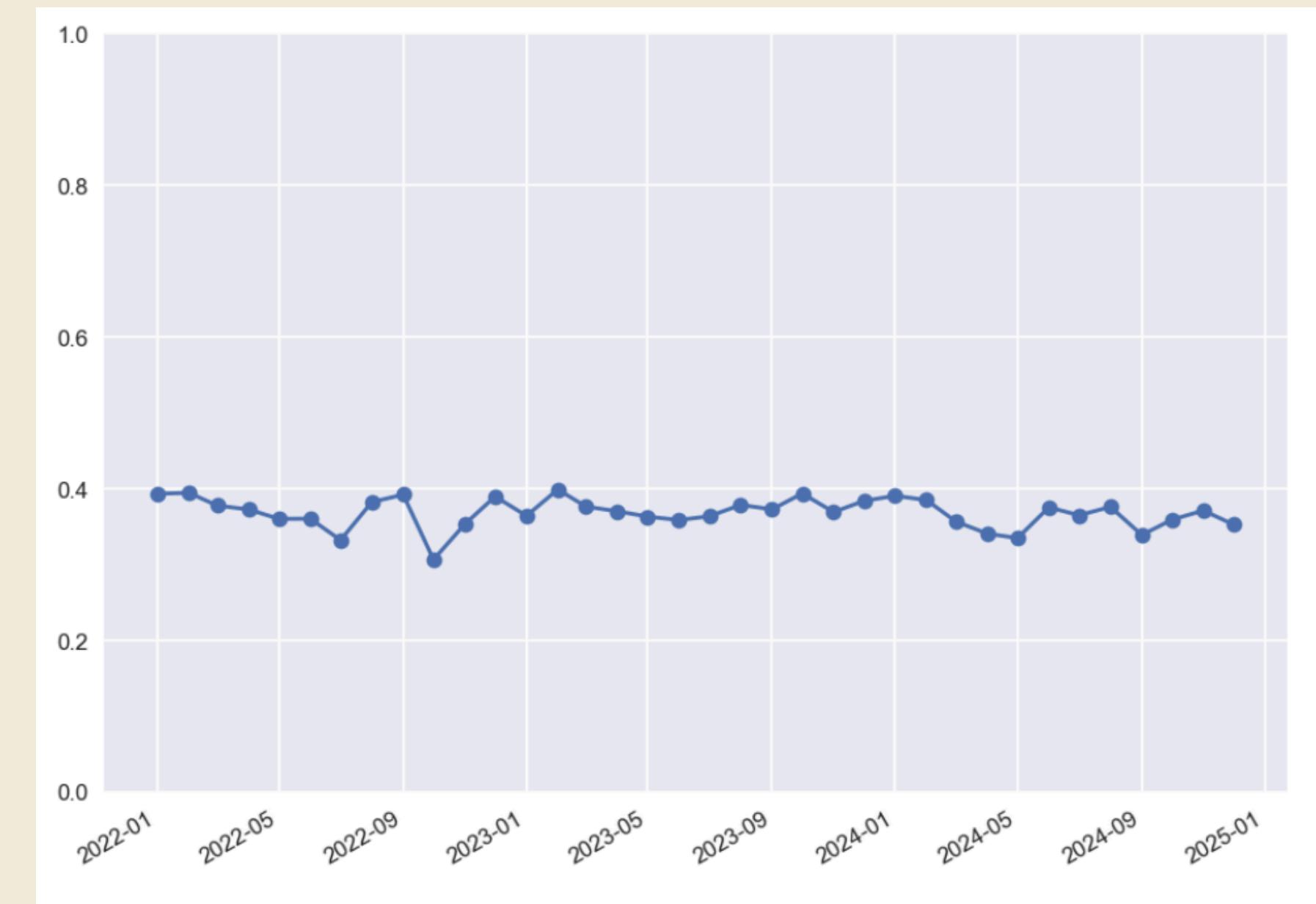
- 利用SnowNLP進行情緒分析
- 文章斷句
- 每句情緒分析
- 情緒分布
- 每月趨勢
- 繪製文字雲



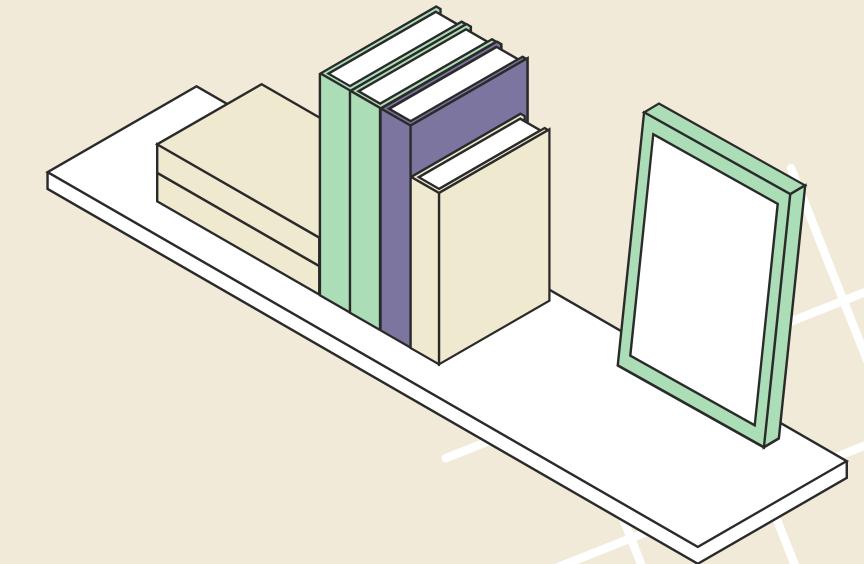
情緒分布的盒狀圖



情緒分數時間趨勢



特定時段文字雲

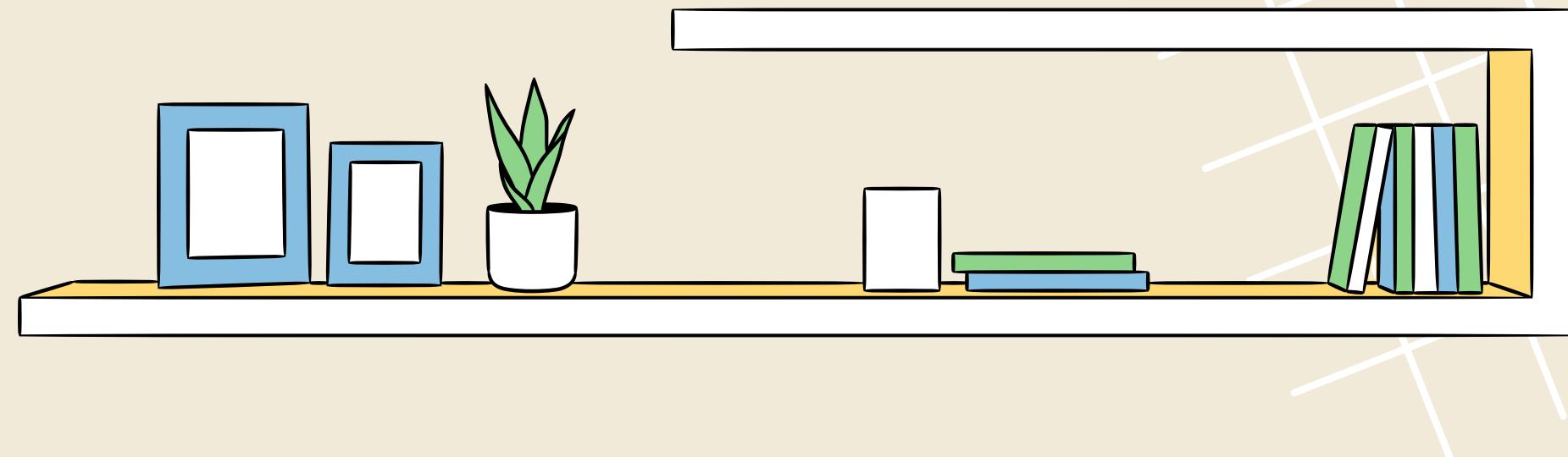


遇到的困難

03



遇到的困難



1

斷詞錯誤率高
手動更新耗時

2

有些詞可能不是
代表某一種情緒
需要額外印出並
分析文本

3

Tarflow資料
抓取不順利

結論

04



結論

文章大多偏正向

大家對獎金、福利超有興趣

資策會、微軟

英文、日文

時間令人焦慮



A stylized illustration of a woman with dark hair, wearing a white blazer over a green top, holding a large yellow circle. The background features a light beige color with a faint white grid pattern.

Thank you