

作業繳交方式：

點選下方連結

https://docs.google.com/forms/d/e/1FAIpQLSeof76JeBCESlws5043BYquFpTAuROH_0E34dSkmhJHuzvYgg/viewform?usp=sf_link

會看到一個Google表單，如圖所示：

第七次作業繳交

繳交作業前請先詳閱以下注意事項！

1. 請務必繳交.py檔案，若是以colab撰寫的程式碼，請先以.py形式下載後再繳交。
2. 檔案名稱請依照指定格式命名，相關格式範例以及注意事項都有寫在作業上傳區，上傳前請務必仔細檢查！
3. 上傳後如果發現自己不小心交錯，直接以相同的檔案名稱重新繳交即可。

當你上傳檔案並提交這份表單時，系統會記錄與你 Google 帳戶相關聯的名稱和相片。表單回覆不會包含你的電子郵件地址。

* 表示必填問題

請選擇你的學校 *

選擇

檔名格式檢查器 (點擊藍字即可開啟檢查器)

請同學們在上傳作業前務必先使用檢查器檢查過自己的檔名格式，確認無誤後再上傳

作業上傳區 *

繳交檔案名稱格式：學校名稱_學號_姓名

範例：國立基隆高級中學_1130921_易佰分

※ 中間不可加上任何空白或其它符號及文字，否則系統會無法判斷內容
※ 未依規定檔案名稱格式繳交一律視為缺交

可上傳 1 個支援的檔案，大小上限為 10 MB。

↑ 新增檔案

作業繳交說明請務必先閱覽

學校名稱記得要選對

上傳前記得先使用檢查器
檢查一下自己的檔名格式

上傳作業前請一定要看清楚
檔案名稱格式以及是否
為.py檔

作業內容

- 目標: 請根據給定的 **text**, 運用**jieba**、**Counter**來斷句並統計詞頻以後, 並根據以下要求來回答問題
 1. 將文本統計詞頻後, 停用給定的詞以及用字典自訂義增加新的詞。最後過濾出現次數大於1的詞彙, 並按詞頻次數從高到低排序, 且不出現詞頻只有1的詞彙(過濾出現次數大於1的詞彙)
 2. 利用**jieba.analyse**來將斷句過後的**text**來作摘要
 3. 利用**jieba.analyse**來找到關鍵詞**keywords**

輸入格式

本題無文字輸入

輸出格式

李樹桐認為高祖未稱臣的論點有四, 其一為高祖不稱天子表示其未受突厥封號, 也未稱臣, 且當時情勢並不危急, 無稱臣之必要; 其二高祖致書突厥用啓字為政治操作; 其三記載高祖稱臣之紀錄為史官欲討太宗喜所寫, 因修史者缺乏史德, 素有「誣高祖之惡, 溢太宗之美」之舉, 且高祖與太宗間有嫌隙, 其晚景淒涼, 故當朝史官多抹黑高祖以悅太宗, 而李樹桐以記載高祖稱臣與太宗渭水之恥的時間點推論史官欲以高祖誠稱臣掩太宗的渭水之恥, 此亦符合史官一貫貶高祖的敘事筆法, 若真稱臣當為國恥, 豈會不察而錄; 其四因未有稱臣與結束稱臣的紀錄, 故未稱臣。陳寅恪先以太宗欲改白旗、符讖和興國寺兵脅高祖證之, 言起兵時的絳白旗一方面欲以絳色示初時擁立隋幼主的立場, 另方面表示其不純服於突厥, 並以「童子木上懸白旗, 胡兵紛紛滿前後」, 證白旗為突厥之旗色, 再論太宗與突厥有和突利結香火之盟的特殊關係, 故突厥視太宗為其部落之人。而陳寅恪曾在《論唐高祖稱臣於突厥事》中提出唐高祖李淵曾向突厥稱臣的說法, 而李樹桐撰寫《唐高祖稱臣於突厥考辯》以後, 根據封號使用問題、對突厥使用「啓」字、史官記載之過以及稱臣紀錄不存在等方面, 反駁了陳寅恪的觀點

突厥稱臣高祖李淵太宗史官劉文靜記載陳寅恪白旗

[('突厥', 20), ('稱臣', 11), ('高祖', 11), ('太宗', 8), ('李淵', 6), ('史官', 5), ('劉文靜', 4), ('記載', 4), ('後', 4), ('其', 4), ('陳寅恪', 4), ('臣', 4), ('白旗', 4), ('起兵', 3), ('並', 3), ('中', 3), ('可汗', 3), ('亦', 3), ('之', 3), ('唐高祖', 3), ('李樹桐', 3), ('封號', 3), ('唐朝', 3), ('且', 3), ('隋末', 2), ('實力', 2), ('中原', 2), ('強大', 2), ('勢力', 2), ('壓力', 2), ('反', 2), ('決定', 2), ('財帛', 2), ('借兵', 2), ('從', 2), ('隋', 2), ('根據', 2), ('可以', 2), ('時', 2), ('因此', 2), ('等', 2), ('策略', 2), ('曾', 2), ('使用', 2), ('啓', 2), ('字', 2), ('紀錄', 2), ('方面', 2), ('不同', 2), ('初期', 2), ('及其', 2), ('研究', 2), ('係', 2), ('認為', 2), ('或', 2), ('立場', 2), ('絳', 2), ('表示', 2), ('因', 2), ('渭水之恥', 2)]

text內容

text = ""

隋末民間動亂削弱了隋朝的軍事實力, 而導致突厥崛起, 成為中原一帶強大的勢力, 藉此對中原形成巨大壓力。而李淵在起兵反隋時, 由於意識自身力量不足, 依劉文靜建議, 決定借助突厥的支持, 以財帛金寶作為借兵的交換條件, 從而增強反隋的實力。而若根據《舊唐書》記載可以發現到, 李淵起兵時兵源有限, 因此決定

「財帛金寶入突厥」以求借兵，並借用胡騎作為援兵。此外，在突厥的統治者中像是始畢可汗、處羅可汗等相繼去世後，李淵也罷朝致哀，並「詔百官就館吊其使」。而從《資治通鑑》中亦記載：「劉文靜勸李淵與突厥相結，資其士馬以益兵勢。淵從之，自以卑辭厚禮，遺始畢可汗。」。

所以，可以反映出唐初為應對內外壓力而採取的妥協策略。而陳寅恪曾在《論唐高祖稱臣於突厥事》中提出唐高祖李淵曾向突厥稱臣的說法，而李樹桐撰寫《唐高祖稱臣於突厥考辯》以後，根據封號使用問題、對突厥使用「啓」字、史官紀載之過以及稱臣紀錄不存在等方面，反駁了陳寅恪的觀點。兩者基於對史料的不同解讀，揭示出學者對同一歷史事件的不同詮釋。因此，針對唐朝建國初期面對強大突厥勢力的複雜抉擇及其學術爭論，本研究將深入探討李淵與突厥之間的關係及其對唐朝建立的影響，並分析學術界長期辯論的原因及意義。這不僅有助於理解唐朝初期的外交策略，亦為唐史研究提供新視角。陳寅恪認為隋末中國北方群雄多臣於突厥，或受其封號，或得具突厥圖騰的狼頭纛，唐高祖亦不例外，且稱臣突厥之主謀實為唐太宗與劉文靜，而後李唐欲改稱臣突厥的立場，故殺劉文靜。陳寅恪先以太宗欲改白旗、符讖和興國寺兵脅高祖證之，言起兵時的絳白旗一方面欲以絳色示初時擁立隋幼主的立場，另方面表示其不純服於突厥，並以「童子木上懸白旗，胡兵紛紛滿前後」，證白旗為突厥之旗色，再論太宗與突厥有和突利結香火之盟的特殊關係，故突厥視太宗為其部落之人。李樹桐認為高祖未稱臣的論點有四，其一為高祖不稱天子表示其未受突厥封號，也未稱臣，且當時情勢並不危急，無稱臣之必要；其二高祖致書突厥用啓字為政治操作；其三記載高祖稱臣之紀錄為史官欲討太宗喜所寫，因修史者缺乏史德，素有「誣高祖之惡，溢太宗之美」之舉，且高祖與太宗間有嫌隙，其晚景淒涼，故當朝史官多抹黑高祖以悅太宗，而李樹桐以記載高祖稱臣與太宗渭水之恥的時間點推論史官欲以高祖誠稱臣掩太宗的渭水之恥，此亦符合史官一貫貶高祖的敘事筆法，若真稱臣當為國恥，豈會不察而錄；其四因未有稱臣與結束稱臣的紀錄，故未稱臣。

.....

模組內容

```
import jieba
from collections import Counter
import jieba.analyse
```

#參考資料:助教許家瑋的史學方法分組報告

#引用書目：
任士英,〈說李淵稱臣突厥事—兼述劉文靜被殺原因〉,《煙台師範學院學報》, 1991年第4期, 山東, 1991, 頁37-38。
陳寅恪,〈論唐高祖稱臣於突厥事〉,《嶺南學報》, 11(2), 廣州, 嶺南大學, 1951, 頁1-10。
李樹桐,〈唐高祖稱臣於突厥考辨〉,收入氏著,《唐史考辨》,臺北,中華書局, 1965, 頁214-246。

Hint

##提示-詞頻

```
#避免輸出有\n跟空格出現
text = text.replace("\n", "")
```



```
# 排除停用詞
stop_words = set(['所以', '好', '因為', '；', '的', '是', '了', '欲', '也', '在', '和', '就', '不', '有',
                  '以', '與', '、', '為', '這個', '而', '「', '」', '，', '。', '《', '》', '對', '於', '\n'])
```



```
# jieba中添加自定義詞彙
custom_dict = [
    '陳寅恪',
    "濁水之恥",
    "論唐高祖稱臣於突厥事",
    "資治通鑑",
    "舊唐書"# 直接添加词汇
]

for word in custom_dict:
    jieba.add_word(word)
```



```
## 過濾停用詞
filtered_words = [word for word in words if word not in stop_words]

# 統計詞頻
word_count = Counter(filtered_words)
# 過濾出現次數大於1的詞彙
filtered_word_count = {word: count for word, count in word_count.items() if count > 1}
```



```
# 排序句子，選擇句子作為摘要
top_sentences = sorted(sentences, key=lambda x: sum(1 for word in jieba.cut(x) if word in keywords), reverse=True)[:3]

# 生成摘要並去除換行符
summary = '。'.join(top_sentences).replace("\n", "")
```



```
print(summary)
print("-----")
print(''.join(keywords))
print("-----")
print(sorted_filtered_word_count)
```



Hint

- 本次作業是要讓同學們對於**JIEBA**跟**Counter**有一些初步的認識，可以幫助同學們對於一篇長文當中可以快速地找到重點

- 本次作業無輸入，只要輸出所需的答案即可，所以不須額外建立輸出的語法

- 最後還是要再度提醒，在繳交作業前一定要再三檢查，檔案名稱的格式有沒有打對，程式也要測試輸出格式是否與範例完全一致喔！
