



大數據與商業分析

HOMEWORK2

Group 12 徐熾鎔 廖文豪 鄭皓 陳宜君 陳牧忠 藤田教譽

PRESENTATION DETAILS

TOPICS TO DISCUSS

前處理
篩選方法
計算相似度
GUI
Demo

○○○○○●

資料前處理

PREPROCESSING

READ FILES

將“hw1_text.xlsx”與“bda2019_hw2_table.xlsx”
讀入

CONVERT TO LIST

分成doc_list(新聞內容)以及term_list(關鍵字)
其中doc_list組成有將新聞標題2倍加權後放
入

FUNCTION

get_df(term_list, doc_list), get_tf(term_list, doc_list)
計算各個term的DF及於各篇新聞中的TF，得到df_list
與tf_list



篩選方法

FILTER

METHOD1 TF-IDF

Function: `get_tfidf(tf, df)`

找出245個關鍵字對所有文章的tf-idf
值

METHOD2 CHI-SQUARE

Function: `get_chi_square(tf, df)`

找出245個關鍵字對所有文章的Chi-Square
值



計算相似度 **Similarity**

1

FUNCTIONGET_NORM_WEIGHT(WEIGHT)

將2081篇新聞的向量進行normalized weighting

2

FUNCTIONGET_VSM(WEIGHT,N_QUERY)

將normalized的值兩兩dot，得到相似度



GUI

PACKAGE: PYQT5

大數據與產業分析 HW2

請輸入鴻海新聞編號 (1~2081):

45

Method

☐ TF-IDF

☒ Chi-Square

參考文章

編號	類別	時間	標題	內容
45	鴻海新聞	2016-01-12 13:54:30	大單拉尾盤 台股收盤小跌19點	台股早盤雖開高上漲約60點，而後盤中買氣不濟，蘋果股成過街老鼠股持續挨打，先前漲多的生技股也加入補跌行列，但尾盤包括台積電、台塑集團等出現買盤拉抬指數，終場台股僅小跌19.97點，加權指數收盤為7768.45點，成交值新台幣861億元。今天盤面表現上，科技股、生技股尤為弱勢，蘋果股在蘋果手機首季銷售看跌下，包括可成、鴻準持續破底，大立光收跌95元；生技股的健亞跌 6.4元、F-麗豐跌停為301.5元、智擎跌11.5元、中裕跌22元、浩鼎重挫52元等，走勢疲弱。不過，被市場視為政府基金護盤標的的台積電、鴻海、台塑集團等權值股，尾盤出現大單買盤進場，讓指數收盤由先前約下跌50點，終場僅下跌19.97點，台股得以縮小跌幅收收，1050119

推薦排序

	編號	類別	時間	標題	內容
第一名	1456	鴻海新聞	2016-05-06 22:21:56	夏普股價重挫 傳淨損高達3000億日圓	由於市場擔心夏普公司上個會計年度虧損會比預期大很多，夏普今天股價跳水，盤中最多挫跌11%，根據彭博，有消息人士表示，夏普淨損可能達3000億日圓（28億美元），幾乎是分析師預估均值1610億日圓的兩倍。台北時間早上9時4分，夏普下跌8.5%，報129日圓，創1個多月來最大跌幅。夏普多年來因消費電子產品和提供給蘋果公司（Apple Inc.）等手機製造商的顯示器需求驟降，陷入困境，夏普先前預測營業虧損為1700億日圓。鴻海斥資35億美元收購夏普，夏普發言人透過電話表示：「我們正在調查重組相關的虧損，所以現階段無法揭露淨損或淨資產。」（譯者：中央社陳昱婷）
第二名	1270	鴻海新聞	2016-04-03 01:32:27	iSE新機鋪貨 貨運迎小旺季	蘋果推出全新入門款手機iPhone SE，在美國等地已於3月31日搶先上市，台灣列為第二波，本周四（4月7日）開賣，新機鋪貨效應帶動3月底上海飛美國的航空運價大漲一倍，國內「航空雙雄」華航、長榮航，以及中菲行、捷迅、台聯等貨運承攬商喜迎營收小旺季。捷迅總經理孫錫銀透露，年初以來一直呈現疲弱的航空貨運運價最近出現小旺季，由於蘋果iPhone SE開始出貨，3月底上海飛美國航空運價由原本的每公斤約人民幣12-13元（約新台幣60-65元），一下子飆漲到人民幣20多元，漲幅高達一倍。iPhone SE主要在鴻海河南鄭州廠生產，上海是出貨主要機場，所以運價漲幅最大；台灣、香港的運價變動沒有那麼明顯，但因為訂單外溢效果，多少還是可為運價提供一些支撐。華航、長榮航2月因為碰上農曆年長假，許多工廠休假，貨運需求減少；加上去年初因為美西塞港事件，許多貨品從海運轉空運，推升運價大漲，...
第三名	287	鴻海新聞	2016-09-20 11:45:33	iPhone 7爆噪音門 小摩潑冷水	iPhone 7上市後供不應求，激勵蘋果供應鏈股價昨天大漲表態，國外陸續傳出「噪音門」及飛航模式設定等瑕疵，加上摩根大通證券指出iPhone 7實際銷量沒有看起來那麼好，衝擊蘋果股今天休息。國外科技媒體「9to5Mac」、「512 Pixels」報導，有用戶反映，使用iPhone 7時機身會發出類似中央處理器運作的「嘶嘶聲」，錄音、錄影時雜音還會被收錄進去。此外，外電也報導，有iPhone 7用戶抱怨，將手機設定為飛航模式後，之後即使解除該模式，iPhone 7仍持續顯示「沒有服務」，無法進行通訊。iPhone 7上市後造成熱潮，16日正式開賣後，全球各地都傳出缺貨的消息，儘管蘋果表示不公佈iPhone 7和iPhone 7 Plus的首個周末銷量，但上週卻宣佈首批iPhone 7 Plus已銷售一空，最搶手的新色鑽石黑也已賣光。對於iPhone 7熱銷，分析師各有解讀，摩根大通證券指出，銷售一空是供應不足的結果，認為市場看到的只...

Word Count

15

Matched Word Count

3

4

5

RESULT & DEMO