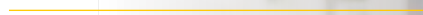


# 大數據與商業分析

台灣2019電子媒體關鍵字探索





# Outline

## 01 文字前處理

- 流程
- Demo

## 02 關鍵詞篩選

- 篩選標準
- 六大主題結果

## 03 總結



# 文字處理

*Text Processing*

01

# Text Preprocessing 程序





# 關鍵詞篩選

*Screening of Kewwords*

02

# 篩選標準說明



keyword

1

成詞達義

2

主題相關性

3

精確：排除空泛常用的詞

4

連接資訊充足：排除特定的詞



# 各主題篩選標準

## DF卡方

1. 銀行
2. 信用卡
3. 台灣
4. 台積電

## TF卡方

5. 匯率
6. 日本

## Topic 1

# 銀行

排名	keyword	排名	keyword
1	銀行	11	兆豐銀
2	金融	12	金融機構
3	分行	13	貸款
4	中央銀行	14	信用
5	銀行業	15	兆豐銀行
6	國銀	16	人民銀行
7	金管會	17	商業銀行
8	央行	18	信用卡
9	金控	19	中國人民銀行
10	金融業	20	貨幣

### ▶ 篩選標準

DF卡方值

### ▶ 原因

不會太空泛，也不會太特殊

### ▶ 關鍵詞議題

- 高頻提及銀行
- 功能業務
- 金融監督



## Topic 2

# 信用卡

排名	keyword	排名	keyword
1	信用卡	11	持卡
2	發卡	12	支付
3	刷卡	13	支付工具
4	行動支付	14	卡友
5	金融卡	15	持卡人
6	手機信用卡	16	付款
7	現金回饋	17	台灣行動支付
8	銀行信用卡	18	聯名卡
9	使用信用卡	19	聯名卡
10	銀行信用	20	簽帳

### ▶ 篩選標準

DF卡方值

### ▶ 原因

除TF-IDF外，其餘結果相似

### ▶ 關鍵詞議題

- 基本內容
- 未來發展方向
- 銀行相關

### Topic 3

## 台灣

排名	keyword	排名	keyword
1	台灣	11	台灣人
2	候選人蔡英文	12	莫蘭蒂
3	候選人朱立倫	13	梅姬
4	國民黨總統	14	候選人
5	總統候選人	15	颱風
6	民進黨總統	16	周子瑜
7	朱立倫	17	暴風圈
8	國際民航組織	18	造勢晚會
9	兩岸	19	掃街
10	候選人宋楚瑜	20	立委候選人

▶ 篩選標準

DF卡方值

▶ 原因

能了解關鍵事件及高頻討論內容

▶ 關鍵詞議題

- 政治
- 經濟發展

## Topic 4

# 台積電

排名	keyword	排名	keyword
1	大立光	11	美股
2	權值股	12	半導體
3	台股	13	晶圓
4	聯發科	14	電子股
5	個股	15	晶圓代工
6	董事長張忠謀	16	蘋概股
7	金融股	17	鴻海
8	聯電	18	日月光
9	外資	19	成交值
10	加權指數	20	友達

### ▶ 篩選標準

DF卡方值

### ▶ 原因

能了解關鍵事件及高頻討論內容

### ▶ 關鍵詞議題

- 投資
- 台積電的營運
- 其他產業龍頭



## Topic 5

# 匯率

排名	keyword	排名	keyword
1	匯率	11	升值
2	人民幣	12	新台幣兌美元
3	美元	13	外匯
4	美元匯率	14	台幣
5	介在	15	中間價
6	人民	16	高低
7	貶值	17	貨幣
8	人民幣匯率	18	間價
9	人民幣兌美元	19	中間
10	台幣兌美元	20	新台幣匯率

### ▶ 篩選標準

TF卡方值

### ▶ 原因

較其他指標有意義、有關聯

### ▶ 關鍵詞議題

- 基本內容
- 經濟發展

## Topic 6

# 日本

排名	keyword	排名	keyword
1	日本	11	日本的
2	蓮舫	12	電玩展
3	慰安婦	13	試爆
4	小桃阿嬤	14	日本央行
5	北韓	15	日本人
6	東京	16	氫彈
7	日本政府	17	安倍
8	核試	18	產業革新
9	夏普	19	溫泉
10	日圓	20	郵票

### ▶ 篩選標準

TF卡方值

### ▶ 原因

較其他指標成詞、準確

### ▶ 關鍵詞議題

- 日本政治
- 東北亞
- 重大事件

03

# 總結

*Conclusion*