

巨量資料探勘與應用

HW2

一、組員名單與分工

資工碩一 112598017 楊淨雯 Q2、Q3

資工碩一 112598028 余珮綺 Q1、Q4

二、環境設定

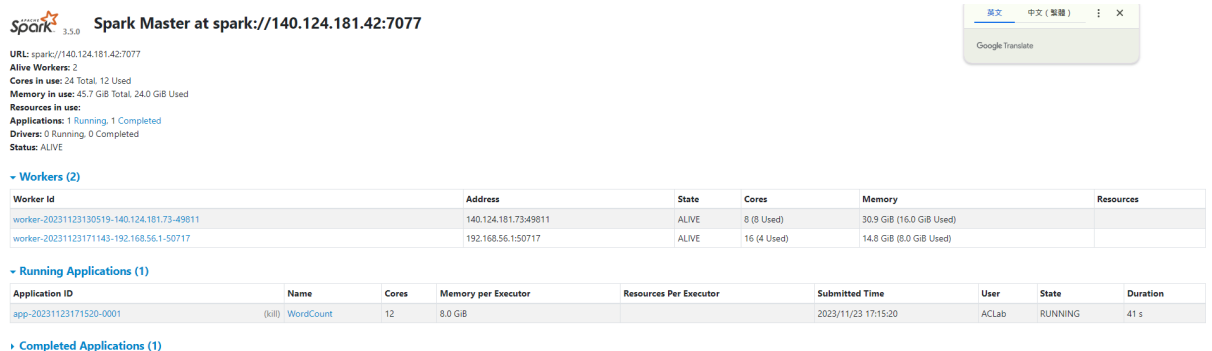
Multi-node on spark

master 和 worker 皆為 windows, 一樣都要先照之前的作業把 python、spark、hadoop 安裝好

master: 140.124.181.42

worker 1: 192.168.56.1, 4 cores, 8GB Used

worker 2: 140.124.181.73, 8 cores, 16GB Used



The screenshot shows the Spark Master web interface at `spark://140.124.181.42:7077`. It displays the following information:

- URL:** `spark://140.124.181.42:7077`
- Alive Workers:** 2
- Cores in use:** 24 Total, 12 Used
- Memory in use:** 45.7 GiB Total, 24.0 GiB Used
- Resources in use:**
- Applications:** 1 Running, 1 Completed
- Drivers:** 0 Running, 0 Completed
- Status:** ALIVE

Workers (2)

Worker Id	Address	State	Cores	Memory	Resources
worker-20231123130519-140.124.181.73-49811	140.124.181.73:49811	ALIVE	8 (8 Used)	30.9 GiB (16.0 GiB Used)	
worker-20231123171143-192.168.56.1-50717	192.168.56.1:50717	ALIVE	16 (4 Used)	14.8 GiB (8.0 GiB Used)	

Running Applications (1)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
app-20231123171520-0001	(kill) WordCount	12	8.0 GiB		2023/11/23 17:15:20	ACLab	RUNNING	41 s

Completed Applications (1)

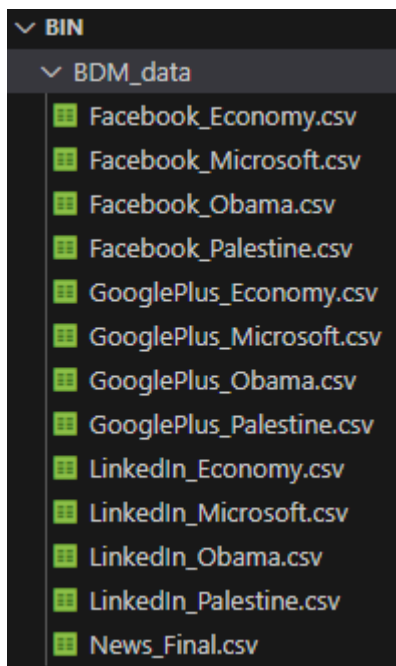
三、實驗執行說明

Step1. 連線

- Master:
以 Admin 開啟 cmd, 到 `spark \ bin` 資料夾, 輸入指令:
`spark-class org.apache.spark.deploy.master.Master`
直到出現 Alive 即可
- worker:
以 Admin 開啟 cmd, 到 `spark \ bin` 資料夾, 輸入指令:
`spark-class org.apache.spark.deploy.worker.Worker spark://<master_ip>:7077`

Step2. 執行

- 確保指定路徑中的資料夾裡有以下這些檔案



- 打開 cmd，進入 spark \ bin 資料夾
- 輸入 `spark-submit --master spark://http://<master_ip>:7077 --executor-cores 4 --executor-memory 8g <code_file_name>`

附註：

在我們的電腦上

<master_ip> 為 140.124.181.42

<code_file_name> 為 BDM_hw2_final.py

畫底線部分為分配 worker 資源

要下載作業有用到的額外套件：Pandas、nltk

四、實驗結果

1.

word	total
economy	26198
obama	22576
microsoft	17570
obamas	5203
us	4713
palestine	3812
new	3740
says	3518
president	3039
economic	2951
microsofts	2939
windows	2734
2016	2428
global	2197
10	2150
growth	1936
trump	1721
palestinian	1599
china	1543
could	1361

only showing top 20 rows

word	total
obama	26438
economy	25510
president	22863
microsoft	20694
barack	13257
us	10049
said	9793
new	9385
economic	9076
year	5734
first	5159
one	4996
palestine	4780
windows	4749
growth	4686
obamas	4261
last	4152
2016	4035
washington	3949
global	3878

only showing top 20 rows

Topic	Total_Title_Word	count
economy	economy	26079
economy	us	2880
economy	economic	2791
economy	global	1946
economy	growth	1812
economy	says	1727
economy	china	1241
economy	2016	1155
economy	chinas	1135
economy	new	1075
economy	brexit	985
economy	uk	889
economy	boost	819
economy	could	819
economy	world	752
economy	quarter	723
economy	oil	715
economy	bank	638
economy	jobs	631
economy	minister	620

only showing top 20 rows

Topic	Total_Title_Word	count
microsoft	microsoft	17559
microsoft	microsofts	2937
microsoft	windows	2731
microsoft	10	1968
microsoft	new	1568
microsoft	surface	1250
microsoft	xbox	1248
microsoft	cloud	971
microsoft	one	802
microsoft	office	727
microsoft	app	703
microsoft	google	664
microsoft	apple	648
microsoft	2016	625
microsoft	pro	620
microsoft	lumia	575
microsoft	data	566
microsoft	update	564
microsoft	azure	554
microsoft	hololens	537

only showing top 20 rows

Topic	Total_Title_Word	count
obama	obama	22231
obama	obamas	5112
obama	president	2655
obama	us	1388
obama	trump	1321
obama	says	1299
obama	barack	1243
obama	visit	1010
obama	court	929
obama	clinton	888
obama	house	871
obama	white	813
obama	gun	798
obama	new	795
obama	administration	791
obama	supreme	758
obama	state	656
obama	michelle	628
obama	gop	580
obama	isis	574

only showing top 20 rows

Topic	Total_Title_Word	count
palestine	palestine	3810
palestine	palestinian	1573
palestine	israel	967
palestine	israeli	560
palestine	2016	330
palestine	new	302
palestine	israelpalestine	285
palestine	peace	264
palestine	un	264
palestine	palestinians	251
palestine	state	216
palestine	east	213
palestine	palestines	191
palestine	us	169
palestine	gaza	165
palestine	says	162
palestine	conflict	159
palestine	bank	157
palestine	support	154
palestine	west	142

only showing top 20 rows

Topic	Total_Headline_Word	count
economy	economy	25228
economy	economic	8654
economy	growth	4509
economy	said	4407
economy	us	3739
economy	year	3527
economy	global	3196
economy	new	2939
economy	percent	2621
economy	quarter	2246
economy	minister	1852
economy	government	1786
economy	first	1742
economy	bank	1737
economy	last	1687
economy	years	1664
economy	world	1624
economy	chinas	1574
economy	2015	1559
economy	2016	1558

only showing top 20 rows

Topic	Total_Headline_Word	count
microsoft	microsoft	20679
microsoft	windows	4740
microsoft	new	4044
microsoft	microsofts	3823
microsoft	10	2733
microsoft	company	2107
microsoft	announced	2076
microsoft	one	2041
microsoft	xbox	1636
microsoft	surface	1589
microsoft	users	1486
microsoft	cloud	1463
microsoft	today	1350
microsoft	software	1339
microsoft	app	1154
microsoft	2016	1148
microsoft	office	1148
microsoft	said	1080
microsoft	year	1078
microsoft	last	1054

only showing top 20 rows

Topic	Total_Headline_Word	count
obama	obama	26052
obama	president	20742
obama	barack	13039
obama	us	5314
obama	obamas	4173
obama	said	3615
obama	washington	3195
obama	house	3034
obama	white	2885
obama	administration	2348
obama	first	2057
obama	state	1924
obama	new	1903
obama	tuesday	1621
obama	visit	1488
obama	wednesday	1458
obama	friday	1442
obama	would	1384
obama	thursday	1350
obama	one	1327

only showing top 20 rows

Topic	Total_Headline_Word	count
palestine	palestine	4778
palestine	palestinian	3551
palestine	israeli	1211
palestine	israel	1079
palestine	said	691
palestine	east	666
palestine	2016	592
palestine	state	524
palestine	president	521
palestine	international	516
palestine	new	499
palestine	west	497
palestine	minister	490
palestine	bank	467
palestine	ramallah	438
palestine	palestinians	393
palestine	wafa	377
palestine	city	367
palestine	people	367
palestine	one	365

only showing top 20 rows

PublishDate	Total_Title_day	count
2002-04-02	obama	1
2002-04-02	national	1
2002-04-02	wreath	1
2002-04-02	cemetery	1
2002-04-02	arlington	1
2002-04-02	lays	1
2008-09-20	health	1
2008-09-20	chinese	1
2008-09-20	look	1
2008-09-20	economy	1
2012-01-28	roubini	1
2012-01-28	nouriel	1
2012-01-28	economy	1
2012-01-28	back	1
2012-01-28	global	1
2012-01-28	2008	1
2015-03-01	microsoft	65
2015-03-01	hololens	42
2015-03-01	economy	26
2015-03-01	microsofts	17

only showing top 20 rows

PublishDate	Total_Headline_day	count
2002-04-02	obama	2
2002-04-02	wreath	2
2002-04-02	president	1
2002-04-02	national	1
2002-04-02	unknowns	1
2002-04-02	lays	1
2002-04-02	honor	1
2002-04-02	tomb	1
2002-04-02	arlington	1
2002-04-02	barack	1
2002-04-02	laid	1
2002-04-02	cemetery	1
2008-09-20	state	1
2008-09-20	haywood	1
2008-09-20	economy	1
2008-09-20	discusses	1
2008-09-20	china	1
2008-09-20	director	1
2008-09-20	book	1
2008-09-20	investment	1

only showing top 20 rows

2. 會根據每個不同平台的結果，另存在csv檔。

Facebook_popular_day.csv
Facebook_popular_hour.csv
GooglePlus_popular_day.csv
GooglePlus_popular_hour.csv
LinkedIn_popular_day.csv
LinkedIn_popular_hour.csv

Facebook the average popularity by hour

IDLink	hour1	hour2	hour3	hour4	hour5	hour6	hour7	hour8	hour9	hour10
1.0	-1.0	-1.0	1.6666666666666667	7.0	7.666666666666667	8.0	8.0	8.0	8.0	8.0
4.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
7.0	-1.0	-0.6666666666666666	0.0	2.6666666666666665	4.0	4.0	7.0	7.0	7.0	8.0
8.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-0.6666666666666666	0.0
18.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.6666666666666665	10.0	10.0
25.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
29.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-0.6666666666666666	0.0	0.0
34.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
35.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	13.333333333333334	42.0
36.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
37.0	-1.0	1.0	5.0	29.666666666666668	48.0	48.0	59.0	59.0	59.0	65.0
39.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
41.0	-1.0	0.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
42.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
44.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
45.0	-1.0	-0.6666666666666666	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
47.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
49.0	-1.0	-1.0	-1.0	-1.0	0.0	2.0	2.0	2.0	2.0	2.0
50.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
56.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0

only showing top 20 rows

Facebook the average popularity by day

IDLink	day1	day2
1.0	8.527777777777779	12.25
2.0	16.333333333333332	41.15277777777778
3.0	24.875	92.43055555555556
4.0	1.5833333333333333	5.555555555555555
5.0	7.833333333333333	32.166666666666664
6.0	66.40277777777777	129.66666666666666
7.0	6.944444444444445	12.680555555555555
8.0	-0.3611111111111111	0.0
9.0	4.569444444444445	12.166666666666666
10.0	16.486111111111111	129.08333333333334
11.0	0.7916666666666666	2.0
12.0	1.7361111111111112	10.916666666666666
13.0	8.875	23.125
14.0	-0.1527777777777778	0.0
15.0	-1.0	113.875
16.0	-1.0	7.5
17.0	0.2777777777777778	1.0
18.0	6.625	11.847222222222221
19.0	448.8888888888889	650.9444444444445
20.0	14.666666666666666	18.5

only showing top 20 rows

GooglePlus the average popularity by hour

IDLink	hour1	hour2	hour3	hour4	hour5	hour6	hour7	hour8	hour9	hour10	hour11	hour12	hour13	hour14
1.0	-1.0	-1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.0
7.0	-1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.0	0.0	0.0	0.0	0.0	0.0
18.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
25.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
29.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
34.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.0	0.0
35.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	13.0	13.666666666666666	14.0	15.333333333333334	17.0	17.0
36.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	1.0	1.0	1.0	1.0
37.0	-1.0	2.6666666666666665	5.333333333333333	6.666666666666667	7.0	7.0	7.0	9.0	9.0	9.0	9.0	9.0	9.0	9.6
39.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
41.0	-1.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
42.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
44.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
45.0	-1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
47.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.0	0.0	0.0
49.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
50.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
56.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	1.0

only showing top 20 rows

GooglePlus the average popularity by day

IDLink	day1	day2
1.0	-0.08333333333333333	0.8055555555555556
2.0	-0.08333333333333333	0.0
3.0	7.472222222222222	22.930555555555557
4.0	-0.5416666666666666	0.6388888888888888
5.0	-0.625	0.0
6.0	25.51388888888889	47.44444444444444
7.0	-0.04166666666666666...	0.0
8.0	-0.3333333333333333	0.0
9.0	0.0	1.0
10.0	0.3333333333333333	2.0
11.0	-0.375	0.0
12.0	0.6527777777777778	2.888888888888889
13.0	0.0	1.0
14.0	-0.125	0.0
15.0	-1.0	1.9166666666666667
16.0	-1.0	2.888888888888889
17.0	-0.3333333333333333	0.0
18.0	0.4166666666666667	1.0
19.0	0.5833333333333334	2.236111111111111
20.0	1.625	2.0

only showing top 20 rows

LinkedIn the average popularity by hour

IDLink	hour1	hour2	hour3	hour4	hour5	hour6	hour7	hour8	hour9	hour10	hour11	hour12	hour13	hour14
1.0	-1.0	-1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
4.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	6.0
7.0	-1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.0	0.0	0.0	0.0	0.0	0.0
18.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	2.0	3.0	3.0	3.0	3.0
25.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
29.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
34.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	1.0	1.0
35.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	37.33333333333333	39.0	40.0	40.33333333333333	42.0	42.0
36.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.6666666666666665	4.0	4.0	4.0
37.0	-1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
39.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
41.0	-1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
42.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
44.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
45.0	-1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
47.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.0	0.0	0.0
49.0	-1.0	-1.0	-1.0	-1.0	17.666666666666668	20.0	20.0	20.0	20.0	20.0	21.0	21.0	21.0	21.0
50.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
56.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.0

only showing top 20 rows

LinkedIn the average popularity by day

IDLink	day1	day2
1.0	0.9305555555555556	2.0
2.0	-0.08333333333333333	0.0
3.0	-0.625	0.0
4.0	3.4166666666666665	10.0
5.0	-0.625	0.0
6.0	34.31944444444444	104.36111111111111
7.0	-0.04166666666666666	0.0
8.0	-0.3333333333333333	0.0
9.0	4.027777777777778	11.111111111111111
10.0	2.3472222222222223	23.166666666666668
11.0	3.138888888888889	9.444444444444445
12.0	2.7777777777777777	27.97222222222222
13.0	2.5	10.986111111111111
14.0	-0.125	0.0
15.0	-1.0	48.77777777777778
16.0	-1.0	5.8055555555555555
17.0	-0.3333333333333333	0.0
18.0	1.8472222222222223	4.0
19.0	2.9444444444444446	5.0
20.0	20.444444444444443	38.166666666666664

only showing top 20 rows

3.

SentimentTitle			
topic	sum	average	
economy	-350.89723953455314	-0.01034240861632142	
microsoft	51.56650733416542	0.002359159453479981	
obama	-27.60443536577117	-9.64852686674979...	
palestine	-177.61514431068477	-0.02008539458449449	

SentimentHeadline			
topic	sum	average	
economy	-1340.13881721557	-0.0394994935515082	
microsoft	-322.2826752054231	-0.01474438078531536	
obama	-507.62019742134794	-0.01774275419158...	
palestine	-393.38238008336015	-0.04448517246221...	

4.

economy Matrix: economy Title																																		
col_col	2015	2016	ahead	amid	back	bad	bank	better	big	boost	brexit	budget	business	central	china	chinas	chinese	could	crisis	data	despite	digital	economic	economy	economys	eu	e							
still	3	0	2	1	1	1	8	13	4	1	2	0	6	6	26	16	7	3	2	8	23	1	38	310	4	6								
japans	4	2	0	0	9	0	3	1	0	5	3	1	3	1	0	6	0	3	0	16	6	0	23	233	0	1								
growth	78	107	10	39	11	3	64	8	8	21	17	4	25	31	139	76	31	38	10	25	39	12	497	826	25	8								
grow	12	96	1	2	1	0	24	0	2	1	3	2	1	7	24	30	20	17	0	0	17	2	7	517	0	2								
boost	6	8	1	1	1	1	11	5	15	819	3	11	3	4	37	9	7	50	4	4	1	9	23	727	0	12								
stocks	0	2	5	33	1	2	2	1	1	2	6	0	0	0	49	13	14	4	1	21	5	0	31	251	6	0								
ahead	0	18	253	3	4	0	3	1	2	1	8	2	0	3	6	11	4	5	0	11	3	0	50	167	0	9								
us	28	73	26	19	13	14	24	42	15	19	66	3	23	7	70	24	12	63	3	68	38	11	197	2254	35	7								
markets	1	12	4	0	0	2	1	3	2	3	7	1	2	3	17	10	10	3	0	8	5	1	27	172	2	0								
2016	16	1159	18	7	9	3	36	15	6	8	8	76	12	13	39	47	11	27	6	2	10	2	121	895	5	4								
local	4	4	2	3	2	0	2	1	8	25	3	0	10	0	2	0	0	7	0	0	2	0	2	237	0	2								
hit	2	1	1	3	0	1	8	0	14	0	75	2	6	17	7	18	7	24	2	1	1	0	19	242	0	14								
slows	16	2	2	11	1	2	3	0	0	1	5	0	4	0	45	44	9	2	1	0	1	0	38	207	0	12								
could	3	27	5	1	2	16	14	9	22	50	113	4	5	3	26	21	10	823	22	5	3	7	69	611	1	40								
prices	5	5	0	7	0	6	5	3	2	5	5	1	0	6	8	6	6	13	2	2	8	0	15	234	4	2								
low	4	19	0	8	0	6	3	0	1	5	8	1	6	2	5	12	3	9	1	1	5	0	23	216	2	0								
russian	12	30	3	2	0	0	20	0	3	3	7	6	2	17	1	0	4	2	10	1	4	0	37	239	3	1								
sharing	4	7	1	2	4	0	1	0	7	5	1	1	9	0	7	5	1	6	0	5	0	6	3	426	11	12								
grew	63	8	0	0	1	0	5	8	0	0	0	0	1	1	4	7	1	0	0	2	2	0	2	325	0	6								
recession	2	12	3	1	15	1	5	0	1	2	17	1	7	2	5	3	2	13	3	7	3	0	17	257	0	1								

only showing top 20 rows

Matrix: economy Headline																												
	col_col	2015	2016	according	also	bank	business	cent	central	china	chinas	chinese	could	country	countryrs	data	despite	development	domestic	economic	economies	economy	european	e				
	still	32	40		30	24	39		18	21		22	41	87	14	28		15	36	49		6	15	148	19	621	33	
	growth	541	340		230	133	394		151	442		182	248	423	133	193		151	153	299	248		96	279	2088	158	3353	111
	spending	73	46		38	26	34		74	33		13	27	25	3	39		19	12	84	27		10	68	183	6	722	9
	grow	62	137		74	18	81		33	131		37	31	43	29	35		23	17	31	29		22	44	118	16	819	12
	2016	120	1660		114	25	106		47	95		52	80	107	47	51		30	55	56	51		37	62	450	28	1308	15
	financial	45	36		56	23	96		29	17		59	71	95	47	92		31	35	53	57		6	26	287	47	904	50
	two	41	66		45	19	45		30	30		17	30	44	17	45		30	20	31	26		8	20	206	28	673	28
	us	177	193		208	94	152		117	48		99	119	63	45	192		45	41	293	122		12	112	769	70	3439	65
	markets	38	63		22	21	65		21	5		44	114	136	78	84		14	15	65	54		3	20	209	48	727	57
	could	38	51		87	70	73		64	25		49	59	33	45	1483		26	65	39	19		15	33	305	45	1216	152
	prices	65	73		34	54	101		20	25		47	52	51	34	55		46	47	61	34		10	49	234	42	965	22
	finance	23	33		19	20	38		23	8		41	14	27	5	25		33	68	5	11		15	15	168	49	643	48
	low	48	48		28	18	48		16	23		21	13	37	4	32		26	28	20	26		7	35	114	25	544	11
	month	22	22		49	11	39		38	47		24	28	34	18	18		10	17	67	25		4	27	121	6	541	15
	grew	345	93		125	27	91		46	222		68	86	162	28	38		42	45	234	32		9	186	150	11	1576	12
	new	133	169		279	52	111		157	108		64	111	97	75	122		73	78	139	65		76	48	630	54	2314	87
	recession	55	51		52	17	47		22	19		35	16	19	9	71		39	40	75	14		6	36	160	13	803	25
	growing	54	55		45	24	27		33	60		23	70	46	23	25		32	21	44	41		14	25	133	35	846	10
	economy	1480	1308		1292	689	1429		898	902		806	1054	1468	699	1216		718	871	1138	655		340	837	2377	188	30460	860
	investment	58	26		35	19	39		99	40		17	35	92	28	24		10	32	55	28		19	49	159	17	595	23

only showing top 20 rows

microsoft																															
Matrix: microsoft Title																															
	col_col	10	2	2016	3	365	4	5	950	adds	ai	amazon	android	announces	app	apple	apps	available	azure	band	big	book	build	business	buy	ceo	cloud	corporation	cortana	could	
microsoft	iphone	4	0	0	1	0	0	0	4	2	0	0	8	2	44	24	7	5	0	0	1	0	3	6	0	2	0	2	11	1	
	us	7	3	1	4	2	4	0	6	0	1	0	2	3	2	19	1	6	2	2	0	2	3	5	0	7	9	1	2	2	
	2016	21	6	644	11	5	10	9	0	10	3	4	2	23	2	9	5	15	13	2	8	10	62	2	5	2	13	10	0	11	
	stock	1497	308	502	164	243	333	156	232	223	112	249	285	279	520	540	264	152	460	381	200	188	262	303	200	268	800	324	186	174	
	azure	3	1	10	1	0	1	3	1	0	0	3	0	0	0	7	1	0	0	1	2	0	0	0	6	1	10	40	0	11	
	server	4	0	89	0	0	1	0	0	4	0	1	0	5	0	0	0	7	17	0	3	0	2	1	1	0	15	0	0	5	
	android	24	10	2	1	4	0	0	0	14	0	0	398	1	73	4	86	6	2	12	3	0	1	9	4	1	0	0	46	0	
	buy	4	2	5	6	0	6	3	4	1	1	4	4	0	2	9	5	2	3	2	4	1	0	1	228	2	8	17	0	4	
	could	19	0	11	3	0	1	3	0	0	1	2	0	1	4	1	3	0	1	4	7	0	2	5	4	3	4	15	1	259	
	office	29	1	56	3	290	1	2	7	28	0	0	28	15	23	7	42	10	28	0	5	0	12	23	1	2	20	1	1	2	
	apps	42	2	5	0	8	1	2	0	5	2	1	86	2	17	1	338	2	3	2	0	0	12	10	5	3	3	1	2	3	
	dynamics	0	1	31	0	7	0	4	0	9	0	0	0	0	0	0	1	6	14	0	1	0	0	7	0	1	16	0	1	1	
	3	12	4	11	217	3	13	5	0	4	1	2	1	0	0	4	0	2	1	6	0	3	0	1	6	0	4	13	0	3	
	launches	21	2	5	1	5	8	0	10	0	5	0	16	2	31	8	5	1	29	0	4	0	4	12	0	0	9	4	12	0	
	365	7	0	5	3	293	0	1	5	8	0	0	4	8	13	2	8	6	16	0	0	0	2	15	0	2	12	1	0	0	
	phone	43	0	11	4	2	1	1	13	0	0	0	5	2	18	1	4	2	0	1	1	0	3	35	1	0	5	5	0	10	
	corporation	12	3	10	13	1	0	1	0	2	5	1	0	2	5	14	1	0	0	0	3	6	0	0	5	17	1	10	330	0	15
	new	180	13	45	20	41	27	5	4	43	23	2	29	45	116	24	40	9	35	20	13	23	46	32	0	11	62	18	18	23	
	power	5	4	3	0	0	3	2	0	10	0	0	0	9	4	1	5	0	6	6	2	2	2	6	0	0	4	0	7	2	
only showing top 20 rows																															

Matrix: microsoft Headline																															
	col	col	10	2015	2016	365	4	according	also	android	announced	app	apple	apps	available	azure	back	big	billion	build	business	ceo	cloud	companies	company	companys	conferen				
	two	74	22	31	19	16		22	36		28	86	27	42	28		32	42	32	40	19	48		43	28	72	76	78	25		
	us	55	32	31	13	17		18	27		19	47	18	46	12		24	9	12	24	14	26		38	14	57	82	121	30		
	2016	116	32	1340	40	48		25	25		12	275	9	11	20		70	62	18	35	32	174		58	26	86	22	112	31		
microsoft	2737	520	1185	514	494		690	675	637		2427	1203	720	767		629	943	527	450	571	845	999	729	1609	621	2161	685				
	azure	14	20	62	74	2		15	31		0	211	12	0	16		66	1061	11	12	10	39		54	22	501	35	67	35		
	android	70	5	12	10	1		24	25	765		72	244	20	287		57	0	25	12	2	23		17	6	3	7	78	14		
	could	70	7	18	2	4		27	17	6		23	33	28	22		9	6	11	12	30	15		35	16	27	19	55	14		
	apps	167	21	20	28	3		27	41	287		77	167	21	955		43	16	15	18	4	76		71	17	45	25	89	23		
	office	61	27	139	656	2		28	56	79		180	95	13	114		77	85	17	31	10	49		114	25	143	32	101	30		
	companies	27	16	22	16	2		23	13	7		55	10	80	25		15	35	11	26	23	23		47	16	87	685	62	11		
	365	11	9	40	584	0		10	14	10		90	25	2	28		47	74	8	14	2	10		68	12	113	16	39	12		
	phone	139	4	14	2	5		42	29	48		35	50	12	41		26	2	19	8	14	21		85	9	1	4	71	31		
	using	74	4	18	27	7		12	13	27		53	49	8	32		17	31	12	9	6	32		28	18	62	26	44	29		
	new	837	125	239	198	104		102	160	136		544	440	135	175		158	158	79	96	30	317		216	85	253	130	414	126		
	back	77	23	18	8	12		17	14	25		55	21	27	15		15	11	466	11	10	20		13	17	21	11	47	10		
	apple	34	28	11	2	43		28	24	20		13	35	862	21		7	0	27	26	11	9		33	52	8	80	59	23		
	launched	109	26	20	21	32		13	32	37		33	79	24	25		30	18	7	5	4	23		43	4	39	16	76	21		
	recently	74	6	20	9	8		7	14	17		95	26	10	19		13	21	7	9	6	31		13	4	26	7	61	19		
	security	86	12	41	51	1		37	23	1		78	23	9	9		13	51	8	8	3	2		10	25	93	26	69	30		
	technology	33	23	54	8	9		30	20	7		106	36	30	16		6	40	11	42	10	24		62	44	81	94	124	49		
only showing top 20 rows																															

obama																								
Matrix: obama Title																								
	col_col	2016	action	address	administration	america	americans	attack	attacks	back	barack	bill	brexit	budget	calls	campaign	change	climate	clinton	congress	control	could	court	
obama	watch	7	0	6	1	1	1	2	1	0	32	1	0	1	1	0	0	2	8	2	4	0	3	
	donald	5	0	1	0	2	1	7	2	5	64	3	4	0	7	12	1	2	26	0	1	3	0	
	us	5	7	17	19	14	2	23	20	8	89	10	13	3	17	15	17	20	10	13	8	15	39	
	2016	318	0	14	2	1	1	0	1	0	13	0	0	0	4	21	2	5	7	2	15	1	1	
	trumps	0	0	0	1	2	0	4	6	0	4	1	0	0	10	9	0	2	5	0	0	0	0	
	islamic	0	0	9	2	6	1	2	12	2	4	0	0	0	10	10	1	0	2	2	1	0	0	
	could	1	1	0	3	1	1	0	7	4	15	3	11	0	0	0	4	8	11	0	3	265	26	
	hillary	4	3	0	1	1	0	6	2	1	48	10	3	0	2	17	0	0	279	3	4	11	5	
	americans	1	2	3	4	1	215	8	1	5	3	0	0	0	12	1	4	3	0	0	2	1	4	
	nominee	2	0	0	0	0	0	0	0	4	0	0	0	0	3	2	0	1	4	1	0	5	156	
	show	1	2	0	2	0	0	0	0	1	0	9	2	0	0	3	3	2	1	1	1	2	3	0
	gop	5	4	3	3	2	1	2	2	7	6	7	2	4	10	12	4	25	18	6	6	10	81	
	america	1	0	0	1	219	1	1	0	7	11	0	1	0	1	0	0	0	1	0	1	1	0	
	isis	3	2	6	5	3	15	10	27	1	22	2	0	0	14	3	3	3	20	6	1	2	1	
	shooting	0	5	4	1	0	3	5	0	1	5	0	0	0	27	1	0	0	3	0	15	0	0	
new	2	5	8	73	8	5	4	6	2	22	4	1	15	13	10	1	6	7	7	21	12	12		
summit	4	0	2	0	1	0	0	1	0	4	0	3	0	5	0	7	38	0	0	0	0	0		
paris	0	2	0	2	0	0	21	59	1	6	0	0	0	16	0	12	107	0	1	1	4	0		
nuclear	0	10	0	3	0	2	11	1	1	6	0	0	0	6	0	7	1	1	0	0	5	1		
trump	13	3	7	4	14	7	12	17	16	72	6	18	0	18	33	6	12	103	1	5	11	14		
only showing top 20 rows																								

Matrix: obama Headline

	col	col	2016	address	administration	also	american	americans	announced	ap	barack	called	campaign	change	climate	clinton	cnn	conference	congress	could	country	court	cuba	da	
	donald	44	25		24	20	29	11	3	33	510	33	124	6	10		96	53		18	4	19	15	16	1
	day	19	28		21	13	10	31	13	28	293	18	16	14	11		7	18		12	17	10	27	20	16
	us	146	165		368	112	216	160	161	185	3988	161	153	129	137		52	57	128	199	134	170	275	301	13
	two	20	24		65	14	27	19	29	42	414	13	23	22	21		20	13	10	37	11	17	21	39	2
	2016	759	51		42	12	14	14	22	29	426	6	80	4	2		42	10	25	24	21	18	22	12	1
	could	21	15		61	9	15	12	8	30	282	9	16	24	22		29	13	7	18	566	9	63	5	1
	islamic	3	41		70	30	42	53	21	54	446	19	55	10	11		7	2	12	10	9	8	0	1	
	office	47	115		45	15	40	37	9	42	570	15	19	15	8		15	9	6	50	28	19	19	36	2
	hillary	37	1		22	14	11	10	7	23	426	14	171	4	1		747	38	2	3	16	4	18	1	
	americans	14	76		52	18	21	713	5	56	345	32	11	9	9		11	10	4	15	12	43	8	20	3
	cnn	10	20		36	5	16	10	11	1	493	18	35	14	13		45	670	5	7	13	6	38	18	1
	nominee	8	1		20	9	6	2	8	51	338	7	46	1	4		72	16	6	11	19	2	349	2	
	new	49	48		296	31	55	41	109	121	803	47	54	37	33		54	34	26	82	43	26	49	57	5
	summit	16	12		9	28	26	3	4	19	351	19	3	32	71		0	8	33	1	9	5	0	0	1
	paris	1	3		12	4	16	11	4	38	320	14	23	106	262		0	21	86	9	10	3	0	3	1
	nuclear	6	6		43	15	13	9	2	37	328	15	0	11	10		2	12	12	7	9	6	1	2	1
	trump	45	31		27	36	36	14	2	18	448	39	118	9	15		116	52	22	4	18	15	8	2	2
	press	22	7		26	12	10	6	6	14	254	14	10	18	21		11	15	206	13	7	5	15	12	1
	news	28	11		29	9	10	10	7	9	249	18	5	7	8		26	12	179	20	7	13	22	26	1
	minister	20	9		8	11	19	0	13	20	435	10	8	35	43		1	2	37	0	1	14	0	1	1

only showing top 20 rows

palestine

Matrix: palestine Title

	col	col	2015	2016	4	abbas	arab	army	arrested	bank	boys	call	calls	children	city	conference	conflict	day	de	east	eu	first	flood	flooding	forces	france	french	gaza	gre	
	recognize	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	1	0	0	0	0	9	2	0	
	israel	5	1	0	5	5	2	4	27	0	5	19	10	4	11	0	0	0	14	7	6	12	3	0	0	1	17	9	24	
	day	1	0	0	0	1	0	0	0	0	0	0	1	1	1	0	0	0	96	0	3	0	3	0	0	0	0	0	2	
	army	0	1	0	0	0	61	0	4	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	0	1	0	0	2	
	swaraj	0	0	0	1	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	two	0	0	0	0	0	2	3	4	1	3	0	1	2	0	0	8	0	0	0	0	0	0	0	1	0	0	0	3	
	us	0	0	0	2	2	0	0	2	0	2	3	2	0	0	0	4	1	0	2	1	2	0	0	0	0	0	0	0	
	2016	0	330	9	5	0	1	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	0	0	0	0	0	0	0	
	university	0	1	0	0	2	3	0	0	0	0	2	0	0	1	0	2	0	0	0	0	2	0	0	0	1	0	0	0	
	israelipalestinian	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	24	1	0	0	2	0	0	0	0	0	8	0	0	
	rights	2	0	0	1	0	0	0	1	0	0	2	1	0	0	2	0	1	0	0	0	0	0	0	0	0	0	1	0	
	greek	0	0	0	2	1	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	palestinians	1	2	1	5	3	10	0	23	0	2	4	1	0	1	0	1	1	0	0	2	2	1	0	19	0	3	6		
	shooting	0	0	0	1	0	0	19	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	new	2	1	1	5	1	1	3	10	1	3	3	3	2	2	4	0	0	6	4	2	0	0	0	0	3	1	4		
	palestinian	1	5	0	28	18	24	6	34	0	4	19	31	10	6	8	24	2	12	10	5	0	0	0	28	16	12	32		
	trump	0	0	0	0	0	0	0	3	0	2	2	0	1	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	
	man	0	0	0	0	0	0	21	1	0	0	0	3	0	0	0	0	1	0	10	0	0	1	0	4	0	0	0		
	report	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	3	0	0	0	2	0	0	0	1	
	israels	1	0	0	1	1	0	0	2	0	0	1	3	0	0	0	0	0	2	0	1	3	0	0	0	0	0	4	3	

only showing top 20 rows

Matrix: palestine Headline

	col	col	2015	2016	abbas	according	affairs	agency	arab	authority	bank	called	center	city	conflict	day	de	december	department	east	first	forces	foreign	friday	gaza	genera
israel	10	41	14	27	85	7	27	57	53	38	7	19	79	27	0	3			5	47	51	19	47	31	68	2
day	13	19	3	3	8	2	6	1	11	4	2	10	2	201	0	0			1	4	16	4	1	10	12	1
two	10	18	9	7	10	3	10	4	24	10	6	20	18	7	0	8			16	37	21	23	8	19	15	
2016	10	654	20	44	6	31	15	16	26	16	98	21	13	19	27	1			47	111	16	18	3	30	27	5
us	3	15	7	6	3	3	2	16	6	6	3	9	22	3	0	3			11	6	3	5	20	11	2	
university	2	18	0	2	9	1	4	1	8	1	5	8	5	2	1	1			2	5	3	7	0	0	5	
islamic	1	7	3	0	1	8	4	2	10	4	3	18	5	3	0	1			1	0	6	26	5	18	10	
israelipalestinian	0	3	1	1	2	0	1	3	3	8	4	1	115	1	4	0			0	9	11	1	19	11	2	
rights	3	6	2	4	2	0	0	1	5	10	8	2	3	1	1	1			1	6	0	4	3	3	3	
wafa	184	198	10	58	2	27	9	4	14	14	1	9	7	19	0	118			8	16	6	43	0	15	32	4
palestinians	24	33	9	20	4	5	7	11	48	6	1	17	5	15	0	10			1	15	31	28	7	8	28	1
new	11	40	5	12	29	16	5	14	18	11	10	30	23	13	1	3			3	33	50	7	12	23	8	
palestinian	191	170	261	113	48	66	90	304	301	70	21	152	32	105	6	103			31	105	124	162	87	83	195	9
agency	17	31	4	7	4	180	7	2	17	3	1	2	5	2	0	12			1	48	3	13	7	13	9	1
man	13	11	3	15	0	3	1	2	14	2	1	15	0	1	0	9			15	33	10	16	0	6	4	
news	3	13	7	7	2	56	9	1	3	1	0	6	2	4	0	5			2	4	8	5	7	9	2	
israels	5	5	6	6	1	8	0	4	19	6	0	2	1	5	0	1			2	7	3	9	4	10	8	
minister	7	28	37	8	138	13	7	38	4	36	17	2	13	12	0	6			1	19	40	1	155	14	6	
security	5	9	4	2	18	7	7	12	44	2	0	13	6	1	0	2			6	11	5	52	7	19	10	1
sunday	22	7	8	7	27	5	3	3	8	2	2	8	3	9	0	14			8	5	10	6	7	2	5	

only showing top 20 rows