BA1: TOP 5

86609 followermonitor
43836 delicious50
37459 porngus
36394 tweet_the_time
34357 drunk_bot

time:
delicious50: 07 ~ 13   15058
delicious50: 14 ~ 21   10704
delicious50: 22 ~ 06   18027
drunk_bot: 07 ~ 13   9559
drunk_bot: 14 ~ 21   9837
drunk_bot: 22 ~ 06   14892
followermonitor: 07 ~ 13   36697
followermonitor: 14 ~ 21   19705
followermonitor: 22 ~ 06   30103
porngus: 07 ~ 13   12869
porngus: 14 ~ 21   9760
porngus: 22 ~ 06   14807
tweet_the_time: 07 ~ 13   14822
tweet_the_time: 14 ~ 21   8541
tweet_the_time: 22 ~ 06   12992

What:

We use
ensiz_news
rockonmobi
dwzmedia
realitysocial
Ganjaboy74

Real-real people
delicious50
thinkingstiff
dominiquerdr
mariolavandeira
thegamingscoop

92855	you
91140	that
86599	stopped
86579	following
86418	Ive
86148	noticed
66738	httpcligsgV4jg
58643	the
46494	of
43290	is
38109	now
37409	#porn #nsfw
36926	It
34293	RT
26458	drunk
24999	and
21220	am
18456	a
18071	to
16649	pm
16363	I
15854	August
14667	httpbitly14MylO
11524	in
11465	httptwitdrunkcom
11396	September
8818	on
8531	for
6813	my
6402	Sunday
6310	with
6127	Saturday
6045	The
6020	at
5964	i
5763	was
5334	October
5279	Monday
5242	it
5220	get
5174	Wednesday
4875	Thursday
4862	Tuesday
4842	me

4827   her
4674   Friday

wordCount:yarn jar ./wordcount-1.0.jar BA1.WordCountJob /p2/*.txt /outAA02

yarn jar ./wordcount-1.0.jar BA1.whatAndWhen.Top5Only /p2/*.txt /outP215
yarn jar ./wordcount-1.0.jar BA1.whatAndWhen.Top5Only /p2/*.txt /outPBA101

When: yarn jar ./wordcount-1.0.jar BA1.whatAndWhen.when.When /outP216/part-r-00000 /outP217

What: yarn jar ./wordcount-1.0.jar BA1.whatAndWhen.what/What /outP216/part-r-00000 /outP218
yarn jar ./wordcount-1.0.jar BA1.whatAndWhen.what/What /outPBA101/part-r-00000 /outPBA102

What Sort: yarn jar ./wordcount-1.0.jar BA1.UserSort /outP221/part-r-00000 /outP222
yarn jar ./wordcount-1.0.jar BA1.UserSort /outPBA102/part-r-00000 /outPBA103

BA2: HashTags

//step1: get hashtag - time
Step1: week:hashtag
Step2: week:hashtag - number
Step2: hashtag - month - frequency

Jobs:
1 yarn jar ./wordcount-1.0.jar BA2.GetWeekTagJob /p2/*.txt /outP2BA21
2 sort tag by number: yarn jar ./wordcount-1.0.jar util.sortByIntValue.UserSort /outP2BA24 /outP2BA25
3: use localTasks.BA2 to compute top5 of each week and trend

BA3: Sentiment analyze

Results:
User:
delicious50: -4
mariolavandeira: -3
dominiquerdr: -5
thegamingscoop: -5
thinkingstiff: -5

Hashtags
iranelection: -3
jobs: 1
followfriday: 0
iremember: 1
spymaster: 0

Jobs:
Step1 get metadata(name - tweet) of chosen 5 users and top5 hashtags of week2

yarn jar ./wordcount-1.0.jar BA3.TOP5User.Top5Only /p2/*.txt /outPBA301
yarn jar ./wordcount-1.0.jar BA3.TOP5HashTagWeek2.Top5Only /p2/*.txt /outPBA302

After got these two data, run localTasksBA3(Azure APIKey needed), the program will print the result on console

D

Advanced Analyze

Time - mood analyze

Normal active user(2-4 tweets per day), get: user - number, choose 100 user
User - twitch -> time(hour)-tweet
分析

Step1: step1_candidates: totally 120000
wordCount:yarn jar ./wordcount-1.0.jar AA.Step1.NormalActiveUser /outAA02/part-r-00000 /outAA03
So we choose 120 of them
Used localTask, choose one every 1000 lines!

Step2: find all tweets that belong to lucky120
Job: yarn jar ./wordcount-1.0.jar AA.Step2.GetLucky120Tweets /p2/*.txt /outAA04

Step3: every hour, choose 100, call api, See localTask.AA2
Result:
Hour: 0 Score: -7
Hour: 1 Score: -6
Hour: 2 Score: 6
Hour: 3 Score: 10
Hour: 4 Score: 9
Hour: 5 Score: -7
Hour: 6 Score: 9
Hour: 7 Score: 21
Hour: 8 Score: 10
Hour: 9 Score: 14
Hour: 10 Score: 13
Hour: 11 Score: 20
Hour: 12 Score: 16
Hour: 13 Score: 9
Hour: 14 Score: 6
Hour: 15 Score: -2
Hour: 16 Score: 17
Hour: 17 Score: 5

Hour: 18 Score: 20
Hour: 19 Score: -8
Hour: 20 Score: 15
Hour: 21 Score: 8
Hour: 22 Score: 10
Hour: 23 Score: 14

Jobs:
Step1: find out active users
wordCount:yarn jar ./wordcount-1.0.jar AA.NormalActiveUser /outAA02/part-r-00000 /outAA03

```
86609     followermonitor
43836     delicious50
37459     porngus
36394     tweet_the_time
34357     drunk_bot
32828     otvrecorder
31695     tea_robot
30573     bildarchiv
30040     nieuwslogr
29821     eproducts24
27947     dominiquerdr
27404     redscarebot
27258     lingonews
25915     dudebrochill
25334     tweet_words
25296     pflive_announce
25108     transalchemy2
24676     hashjobs
23641     buzzurls
23267     localmosac
23231     znatrainer
23056     bibliotaph_eyes
22899     ezf_toppolitics
22818     mariolavandeira
22055     dwzmedia
21997     bezplatno_net
21788     blogtoblog
21549     travelnewsfeeds
21394     pulsesearch
21364     jobhits
21211     santa_claus_np
20839     thegamingscoop
20837     faithclubdotnet
20797     lukstruk
20700     sportlogr
20481     vaguetweets
20250     wootboot
20131     craigslistjobs
19534     luvorhate
19318     jamokie
19036     richardhejah
```