

BIA-660 Midterm Project Report

Analytics & Comparison of Apple Products with Twitter

Group Members:

Colin Mills: cmills@stevens.edu
Qiang Tong: qtong1@stevens.edu
Xiangda Shi: xshi5@stevens.edu
Xin Gao: xgao12@stevens.edu

The Questions:

- 1) Which Apple mobile OS is more popular? iOS 4 ~ iOS 7.
- 2) Which version of iPhone is more popular? iPhone 3GS ~ iPhone 5S.
- 3) iPhone 5S vs. iPhone 5C, which is more popular?

How We Plan to Answer the Questions:

We planed to get the a large number of tweets about certain Apple product or software around their release date, store them into a database. Then tried to figure out which of these tweets are positive and which of them are negative. Then calculate the percentage of positive/negative tweets and the radio of the positive to the negative. Then display them using Matplotlib. We also planed to find out the “key-word” of each Apple Product

Data Collection:

We decided to get data from the website topsy.com by using their searching url.

For example, the following URL: <http://topsy.com/s?type=tweet&q=iOS%204&language=en&offset=00&mintime=1277078425&maxtime=1308700814>

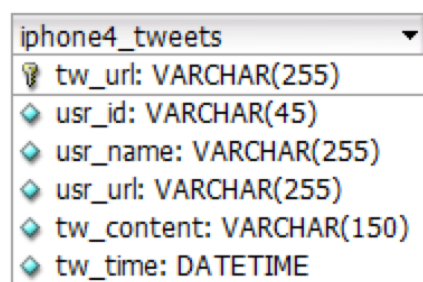
means the searching content is “iOS 4”, the language is set to English, it’s the first page of the results, and the searching is limit between 2010-06-21 and 2010-06-22. We changed the “offset” variable in the url to get to the different pages of the results, and and change the time to get more tweets.

The reason why we did not use Twitter API is because we can only get tweets only in recent 8 to 9 days. But we’d like to get the tweets when the Apple software or products were first released and see what was people’s response at that time in twitter.

The reason why we did not use Twitter searching URL is because the web pages of Twitter searching URL auto-load next a few tweets when you scroll down the page. So when we use urllib to get HTML source code, there are only 20 tweets we can get in that page. There is no “page” variable in Twitter search url. So it’s very inconvenient for us to get data.

There is another problem. If we use urllib to open the above topsy.com searching URL, there is not any tweets in the HTML, the contents of searching results are generated by Javascript. So we decided to use the tool “Selenium”. What “Selenium” do is to open up browser and get the Javascript generated HTML code. We grab the code, then use BeautifulSoup to parse it, get contents we need and then save them into our database.

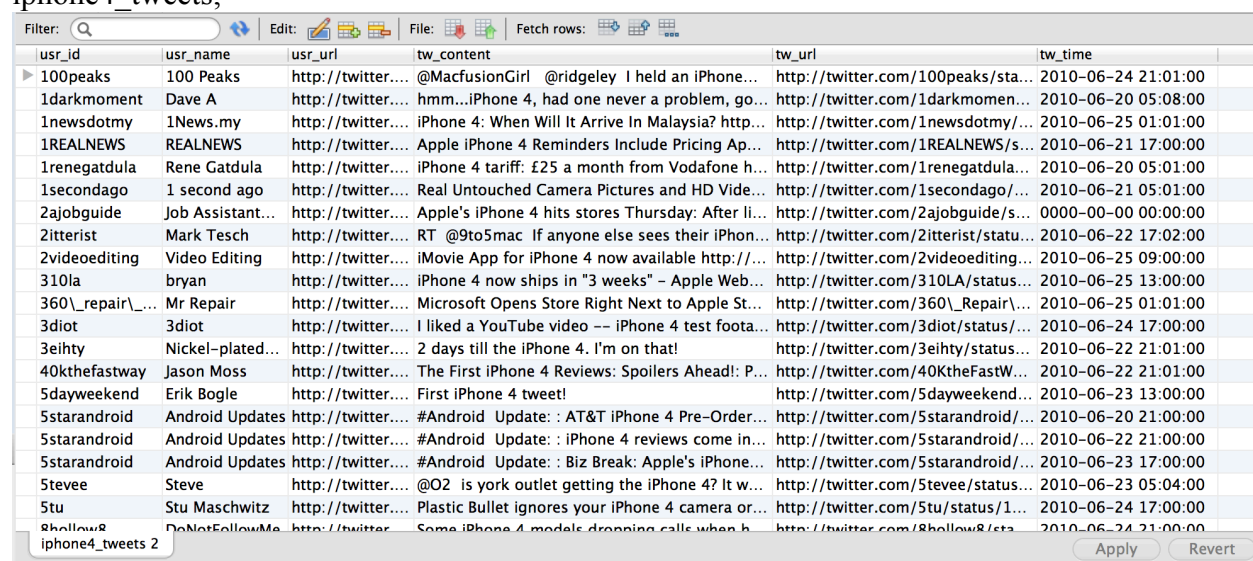
We use mysql database to store all the information of tweets we get, we create several tables, each one of these tables saves all the tweets we gathered about iOS/iPhone at its release time. Here is a graph of one of the tables we created, the variables are tweet url (which is the primary key), user id, user name, user url, tweet content, tweet time.



iphone4_tweets	
tw_url	VARCHAR(255)
usr_id	VARCHAR(45)
usr_name	VARCHAR(255)
usr_url	VARCHAR(255)
tw_content	VARCHAR(150)
tw_time	DATETIME

We managed to get more than 2200 rows in each table, which means there are more than 2200 tweets for each Apple product/software at its release time. We created 11 tables, they are ios_four_tweets, ios_five_tweets, ios_six_tweets, ios_seven_tweets, iphone3gs_tweets, iphone4_tweets, iphone4s_tweets, iphone5_tweets, iphone5s_tweets, iphone5s_only_tweets, iphone5c_only_tweets.

Here is the screen shot of results Mysql Workbench executing sql query “select * from iphone4_tweets;”



usr_id	usr_name	usr_url	tw_content	tw_url	tw_time
100peaks	100 Peaks	http://twitter....	@MacfusionGirl @ridgeley I held an iPhone...	http://twitter.com/100peaks/sta...	2010-06-24 21:01:00
1darkmoment	Dave A	http://twitter....	hmm...iPhone 4, had one never a problem, go...	http://twitter.com/1darkmomen...	2010-06-20 05:08:00
1newsdotmy	1News.my	http://twitter....	iPhone 4: When Will It Arrive In Malaysia? http...	http://twitter.com/1newsdotmy/...	2010-06-25 01:01:00
1REALNEWS	REALNEWS	http://twitter....	Apple iPhone 4 Reminders Include Pricing Ap...	http://twitter.com/1REALNEWS/s...	2010-06-21 17:00:00
1renegatdula	Rene Gatdula	http://twitter....	iPhone 4 tariff: £25 a month from Vodafone h...	http://twitter.com/1renegatdula...	2010-06-20 05:01:00
1secondago	1 second ago	http://twitter....	Real Untouched Camera Pictures and HD Vide...	http://twitter.com/1secondago/...	2010-06-21 05:01:00
2ajobguide	Job Assistant...	http://twitter....	Apple's iPhone 4 hits stores Thursday: After li...	http://twitter.com/2ajobguide/s...	0000-00-00 00:00:00
2litterist	Mark Tesch	http://twitter....	RT @9to5mac If anyone else sees their iPhon...	http://twitter.com/2litterist/statu...	2010-06-22 17:02:00
2videoediting	Video Editing	http://twitter....	iMovie App for iPhone 4 now available http://...	http://twitter.com/2videoediting...	2010-06-25 09:00:00
310la	bryan	http://twitter....	iPhone 4 now ships in "3 weeks" - Apple Web...	http://twitter.com/310LA/status...	2010-06-25 13:00:00
360_repair\...	Mr Repair	http://twitter....	Microsoft Opens Store Right Next to Apple St...	http://twitter.com/360_Repair\...	2010-06-25 01:01:00
3diot	3diot	http://twitter....	I liked a YouTube video -- iPhone 4 test foota...	http://twitter.com/3diot/status/...	2010-06-24 17:00:00
3eihty	Nickel-plated...	http://twitter....	2 days till the iPhone 4. I'm on that!	http://twitter.com/3eihty/status...	2010-06-22 21:01:00
40kthefastway	Jason Moss	http://twitter....	The First iPhone 4 Reviews: Spoilers Ahead! P...	http://twitter.com/40KtheFastW...	2010-06-22 21:01:00
5dayweekend	Erik Bogle	http://twitter....	First iPhone 4 tweet!	http://twitter.com/5dayweekend...	2010-06-23 13:00:00
5starandroid	Android Updates	http://twitter....	#Android Update: : AT&T iPhone 4 Pre-Order...	http://twitter.com/5starandroid/...	2010-06-20 21:00:00
5starandroid	Android Updates	http://twitter....	#Android Update: : iPhone 4 reviews come in...	http://twitter.com/5starandroid/...	2010-06-22 21:00:00
5starandroid	Android Updates	http://twitter....	#Android Update: : Biz Break: Apple's iPhone...	http://twitter.com/5starandroid/...	2010-06-23 17:00:00
5teevee	Steve	http://twitter....	@O2 is york outlet getting the iPhone 4? It w...	http://twitter.com/5teevee/status...	2010-06-23 05:04:00
5tu	Stu Maschwitz	http://twitter....	Plastic Bullet ignores your iPhone 4 camera or...	http://twitter.com/5tu/status/1...	2010-06-24 17:00:00
8hollow8	DoNotFollowMe	http://twitter....	Some iPhone 4 models dropping calls when b...	http://twitter.com/8hollow8/sta...	2010-06-24 21:00:00

Data Analysis:

After gathering the data, we need to analysis the data we got and try to solve the problem. What we did is to first retrieve the tweets from database, then try to figure out whether the tweet is positive or negative. To achieve that, we made two small “dictionaries”, one to save all the positive words, the other to save all negative words. Then we check the tweet content to see if it contains those positive words or negative words to determine whether this particular tweet is positive or negative. Then we counted the positive and negative numbers and make comparison of them with the total number and calculate the ratio and the percentage.

The following are the numbers we get about each topic:

iOS 4/5/6/7

iOS 4:		
Positive:	745 / 2552	29.19279 %
Negative:	530 / 2552	20.76803 %
Positive vs. Negative:	1.40566	
=====		
iOS 5:		
Positive:	787 / 2491	31.59374 %
Negative:	451 / 2491	18.10518 %
Positive vs. Negative:	1.74501	
=====		
iOS 6:		
Positive:	843 / 2611	32.28648 %
Negative:	492 / 2611	18.84336 %
Positive vs. Negative:	1.71341	
=====		
iOS 7:		
Positive:	729 / 2374	30.70767 %
Negative:	595 / 2374	25.06318 %
Positive vs. Negative:	1.22521	

iPhone 5S/5C

iPhone 5S Only:		
Positive:	565 / 2207	25.60036 %
Negative:	478 / 2207	21.65836 %
Positive vs. Negative:	1.18201	
=====		
iPhone 5C Only:		
Positive:	619 / 2234	25.29096 %
Negative:	352 / 2234	15.75649 %
Positive vs. Negative:	1.75852	

iPhone 3GS/4/4S/5/5S

iPhone 3GS		
Positive:	1048 / 3091	33.90489 %
Negative:	554 / 3091	17.92300 %
Positive/Negative:	1.89170	
=====		
iPhone 4:		
Positive:	707 / 3027	23.35646 %
Negative:	571 / 3027	18.86356 %
Positive/Negative:	1.23818	
=====		
iPhone 4S:		
Positive:	688 / 2758	24.94561 %
Negative:	554 / 2758	20.08702 %
Positive/Negative:	1.24188	
=====		
iPhone 5:		
Positive:	633 / 2492	25.40128 %
Negative:	552 / 2492	22.15088 %
Positive/Negative:	1.14674	
=====		
iPhone 5S:		
Positive:	615 / 2225	27.64045 %
Negative:	422 / 2225	18.96629 %
Positive/Negative:	1.45735	

We also count frequency of each word in the data base, and managed to find out the “key-words” of each product. First, we remove all the punctuations, then use string.split() to put words into a list, then we use collections.Counter to count words in the tweet content, then get rid of the “useless words” like ‘iphone’, ‘apple’, ‘5s’, ‘an’, ‘of’, ‘and’...

Here are the key-words with number over 100 of each Apple product/software:

iOS 4: multitasking(280), jailbreak(141), facebook(134), 3g(127), ‘92’(121), navigon(116)

iOS 5: wp7(295), windows(268), jailbreak(197), mango(195), android(151), leak(131), icloud(126)

iOS 6: maps(285)

iOS(7): battery(173)

iPhone 3GS: squarespace(305), att(259), win(246), 3g(225), faster(173), twitition(151)

iPhone 4: att(128), best(111), case(104), shipping(104)

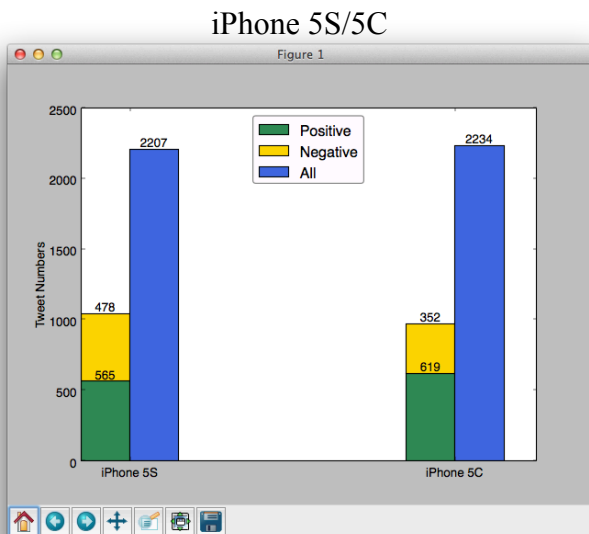
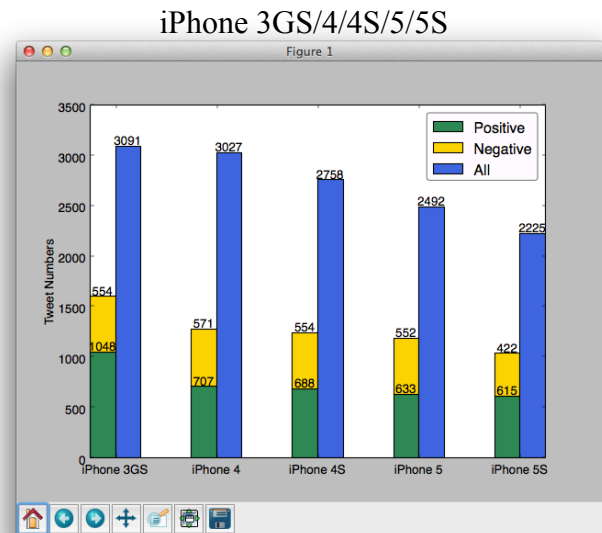
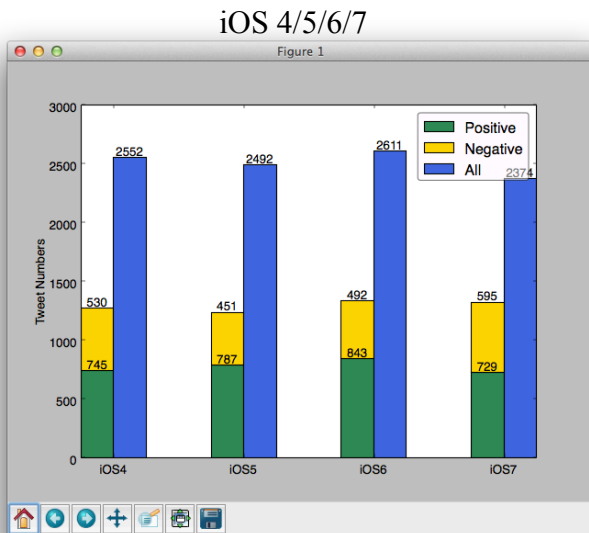
iPhone 4S: siri(256), samsung(115)
iPhone 5S: gold(183), fingerprint(131)

iPhone 5: samsung(96)
iPhone 5C: plastic(137), cheap(102)

Data Visualization:

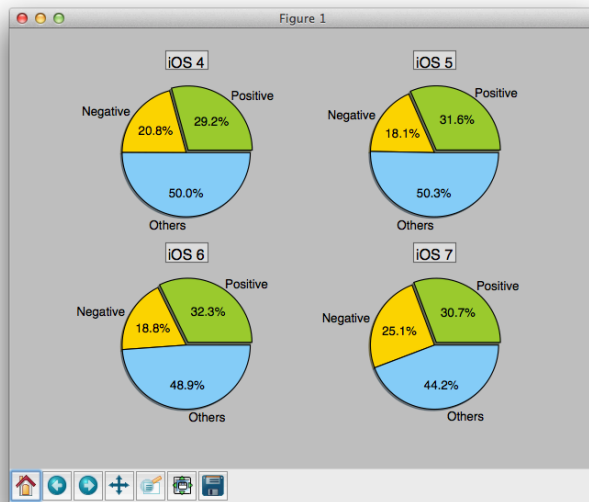
We use Matplotlib library to draw graphs about the statics above. We draw a bar-chart and a pie-chart for each question, the following are the screenshots.

Bar-charts:

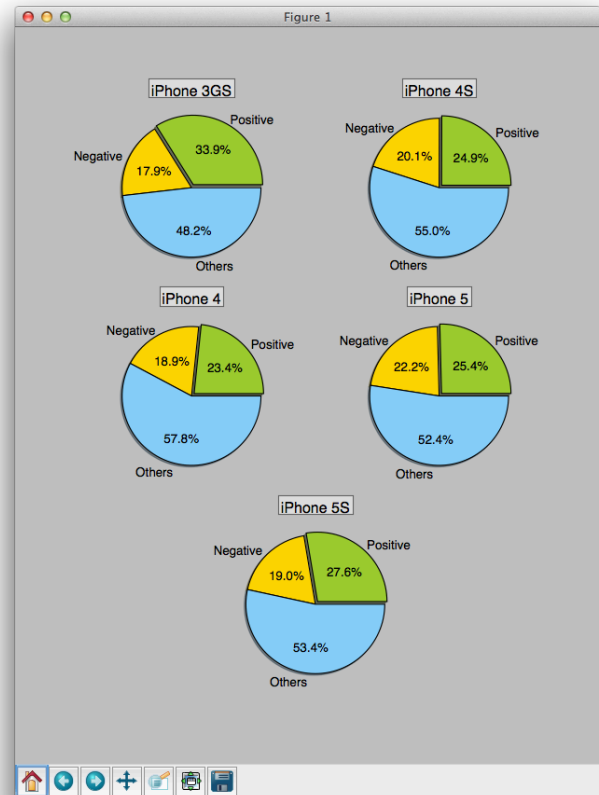


Pie-charts:

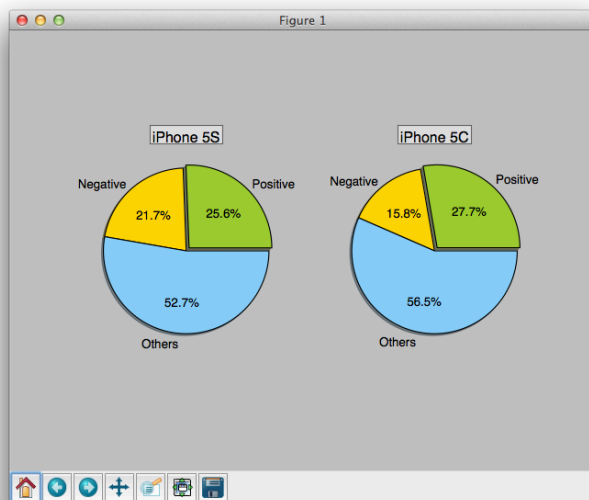
iOS 4/5/6/7



iPhone 3GS/4/4S/5/5S



iPhone 5S/5C



Conclusion:

Among all versions of iOS, iOS 6 has the highest positive tweets percentage, and iOS 5 has the greatest ratio of positive to negative.

Among all iPhones, iPhone 3GS has the the highest positive tweets percentage, and iPhone 3GS also has the greatest ratio of positive to negative.

Between iPhone 5S and iPhone 5C, iPhone 5C has higher percentage of positive tweets, it also has larger ratio of positive to negative.