

Twitter Sentiment Analysis

by

Chayanit Nadam

Dissertation submitted in partial fulfilment of
The requirements for the
Bachelor of Technology (Hons)
(Business Information System)

JANUARY 2014

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Twitter Sentiment Analysis

by

Chayanit Nadam

A project dissertation submitted to the
Business Information System Programme
Universiti Teknologi PETRONAS
in partial fulfillment of the requirement for the
BACHELOR OF TECHNOLOGY (Hons)
(BUSINESS INFORMATION SYSTEM)

Approved by,

ALIZA BINTI SARLAN

UNIVERSITI TEKNOLOGI PETRONAS

TRONOH, PERAK

January 2014

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

CHAYANIT NADAM

ABSTRACT

Social media continues to gain increased presence and importance in society. Public and private opinion about a wide variety of subjects are expressed and spread continually via numerous social media. Twitter is one of the social media that is gaining increased popular. Twitter offers organizations a fast and effective way to analyze customers" perspectives toward the critical to success in the marketplace. Developing a program for sentiment analysis is an approach to be used to computationally measure customers" perceptions. This paper reports on the design of a sentiment analysis extracting a vast amount of tweets. Prototyping is used in this development. Results classify customers" perspective via tweets into positive and negative which is represented in pie chart and html page. However, the program has planned to develop on web application system but due to limitation of Django which can be worked on Linux server or LAMP, for further this approach need to be done.

ACKNOWLEDGEMENT

First and foremost, it is an honor to express my deeply sincere gratitude to my supervisor, Miss Aliza Binti Sarlan for her kind help, support and encouragement during final year in order to complete this project. Certainly my project has been successful. She has not only served as a supervisor but also patiently guided me throughout the final year project process, never accepted less than my best efforts.

Finally, I would like to extend my appreciation and faithful thanks to family and to all around me, whose name have not been mentioned, who either directly or indirectly have encouraged me to successfully complete this final year project and stand all challenges.

TABLE OF CONTENTS

CERTIFICATION		i
ABSTRACT		iii
ACKNOWLEDGEMENT		iv
CHAPTER 1:	INTRODUCTION	1
	1.1 Background Study	1
	1.2 Problem Statement	5
	1.3 Objectives	5
	1.4 Scope of Study	6
CHAPTER 2:	LITERATURE REVIEW	7
	2.1 Social Media	7
	2.2 Twitter	8
	2.3 Sentiment Analysis	8
	2.4 Technique of Sentiment Analysis	11
	2.5 Python	13
	2.6 Related Work	14
CHAPTER 3:	METHODOLOGY	17
	3.1 Methodology	17
	3.2 Milestones	19
	3.3 GANTT Chart	20
CHAPTER 4:	RESULTS AND DISCUSSION	21
	4.1 Methods	21
	4.2 Techniques	22
	4.3 Networking Issues	23
	4.4 Create an Application	24
	4.5 Modeling Functions	27
	4.6 System Architecture	30
	4.7 Results	31
CHAPTER 5:	CONCLUSION AND RECOMMENDATION	34
REFERENCES		36
APPENDIX		41

LIST OF FIGURES

Figure 1	Top ten reasons why people use Twitter	3
Figure 2	The Sentiment Analysis Application User Interface	13
Figure 3	Hypothetical Results of Analyzing Tweets	14
Figure 4	TweetFeel Application	14
Figure 5	Twitter Sentiment Results Page	16
Figure 6	Methodology – 2 Phases	17
Figure 7	Sign In Twitter Developer	24
Figure 8	Create an Application	25
Figure 9	OAuth Settings for Coding Process	26
Figure 10	Application Created	26
Figure 11	Activity Diagram	27
Figure 12	Use Case Diagram	29
Figure 13	System Architecture	32
Figure 14	Pie Chart	33

LIST OF TABLES

Table 1	Milestones	19
Table 2	GANTT Chart	20
Table 3	Comparison Between Machine-based Learning Approach and Lexicon-based Approach	21
Table 4	Parse API Response Keys	25
Table 5	Description of Activity Processes	28

CHAPTER 1

INTRODUCTION

1.1 Background Study

Rambocas and Gama (2013) has mentioned that a millions of people are using social network site to express their emotion, opinion and disclose about their daily lives. However, people write anything such as social activities or any comment on products. Through the online communities provide an interactive forum where consumers inform and influence others.

Moreover, social media provides an opportunity for business that giving a platform to connect with their customers such as social media to advertise or speak directly to customers for connecting with customer's perspective of products and services.

In contrast, consumers have all the power when it comes to what consumers want to see and how consumers respond. From this, company's success, failure is publically to be shared and end up with word of mouth. However, social network can be change the behavior and decision making of consumers, Example from Jose et al., (2010) mentions that 87% from internet users are influenced their purchase and decision by customer's review. So that, if organization can catch up faster on what their customer's think, it would be more benefit to organization to react on time and come up with a good strategy to compete their competitors.

1.1.1 Sentiment Analysis

The sentiment can be found in comments or tweet to provide useful indicators for many different purposes (Prabowo and Thelwall, 2009). Prabowo and Thelwall (2009) together with Saif et al., (2011) stated that a sentiment can be categorized into two groups, which is negative and positive words. Sentiment analysis is a natural language processing techniques to quantify an expressed opinion or

sentiment within a selection of tweets (Carpenter and Way, 2010).

1.1.2 Opinion Mining

Opinion mining refers to the broad area of natural language processing, text mining, computational linguistics which, involves the computational study of sentiments, opinions and emotions expressed in text (Carpenter and Way, 2010). A thought, view or attitude based on emotion instead of reason is often colloquially referred to as a sentiment (Carpenter and Way, 2010). Hence, lending to an equivalent for opinion mining or sentiment analysis.

Osimo and Mureddu (2010) stated that opinion mining has many application domains including accounting, law, research, entertainment, education, technology, politics, and marketing. In earlier days many social media have given web users a venue for opening up to express and share their thoughts and opinions (Pak and Paroubek, 2010).

1.1.3 Twitter

Twitter is a popular real time microblogging service that allows users to share short information known as tweets which are limited to 140 characters (Jose et al., 2010; Lai, 2012; Lohmann et al., 2012). Users write tweets to express their opinion about various topics relating to their daily lives. Twitter is an ideal platform for the extraction of general public opinion on specific issues (Pak and Paroubek, 2010; Osimo and Mureddu, 2010). A collection of tweets is used as the primary corpus for sentiment analysis which refers to the use of opinion mining or natural language processing (Rambocas and Gama, 2013).

Twitter, with 500 million users and million messages per day, has quickly become a value asset for organizations to invigilate their reputation and brands by extracting and analyzing the sentiment of the tweets by the public about their products, services market and even about competitors (Saif et al., 2012). Jose et al., (2010) mentions that, from the social media generated opinions with the mammoth growth of the world wide web, super volumes of opinion texts in the form of tweets,

reviews, blogs or any discussion groups and forums are available for analysis, thus making the world wide web the fastest, most comprising and easily accessible medium for sentiment analysis. However, from figure 1, is showing the top reasons why people are using Twitter. All the reasons are involving with business area. From this point, Twitter will be selected to be platform that will do the sentiment analysis.

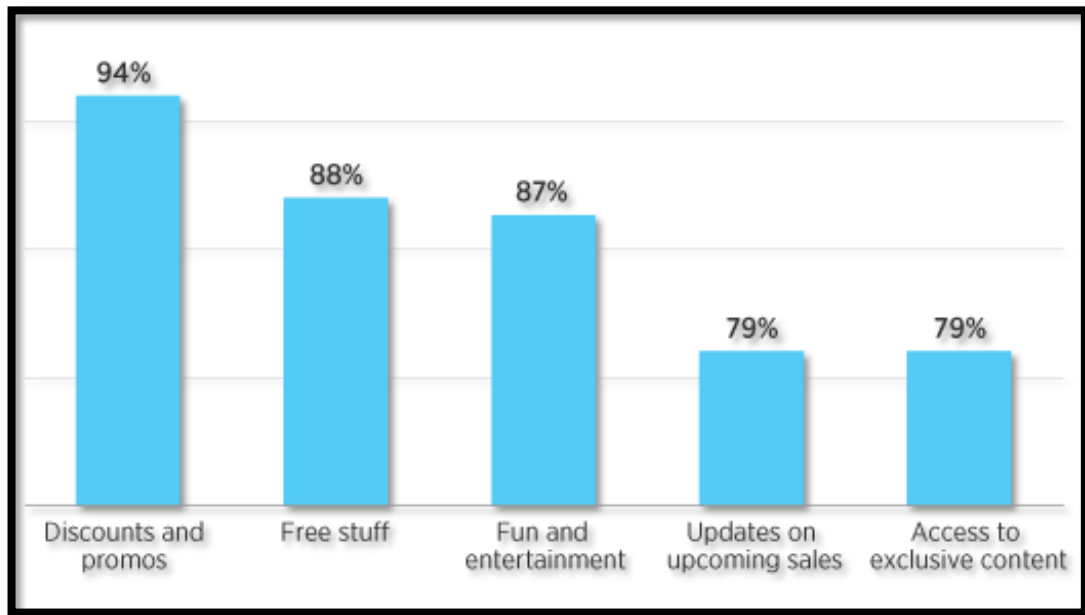


FIGURE 1: Top reasons why people use Twitter

1.1.4 Microblogging with E-commerce

A microblogging platform such as Twitter is alike to a conventional blogging platform just single posts are shorter (Agarwal et al., 2012). Twitter has limited for a small number of words which are designed for the quick transmission of information or exchange of opinion (Wikipedia, 2013a). However, small business or large organizations are initiation to the potential of microblogging as an e-commerce marketing tool (Lai, 2012). Though, microblogging platform has been developed a few years time for promoting foreign trade website by using a foreign microblogging platform as Twitter marketing (Lai, 2012).

The instant of sharing, interactive, community-oriented features are opening an e-commerce launched a new bright spot which it can be showed that microblogging platform has become many companies do brand image, product important sales

channel, improve product sales, talk to consumer for a good interaction and other business activates involved (Jose et al., 2010; Lai, 2012; Zhang et al., 2010).

Agarwal et al., (2012) said in fact, companies manufacturing such products have started to poll theses microblogs to get a sense of general sentiment for a product. Many times these companies study user reactions and reply to users on microblogs (Zhang et al., 2010).

1.1.5 Why Twitter is selected for Sentiment Analysis

A vast majority of people around the world is using Twitter (Pak and Paroubek, 2010). Through this project has been selected the Twitter to be a platform for sentiment analysis. The reasons are: sentiment analysis over Twitter data faces variety of new challenges due to the short length that each tweet up to 140 characters only and sporadic structure of such content (Jose et al., 2010; Lai, 2012; Lohmann et al., 2012). Moreover, Twitter presents an opportunity to learn about customer perceptions, feelings in real time and provide full of meaning insights of customer behavioral tendencies as customer occur without interruption or frustration (Prabowo and Thelwall, 2009; Rambocas and Gama, 2013). From these point, marketer can be learned about customer feelings, attitudes and perspectives on the products, services or company itself.

As more and more users tweet about products and services they use, or express their perspective, microblogging websites become valuable sources that can be used in opinion mining and sentiment analysis tasks. Such data can be efficiently used for variety purposes especially for business purpose such as marketing.

Despite that, from the currently available system, the system can retrieve only one tweet at a time for sentiment analysis (Lai, 2012). Organizations use sentiment analysis for business purposes, at the same time organizations still have a problem when it comes to analyzing heavy reviews tweet by tweet. Hence sentiment analysis for tweet will be created to increase interest due to its promising and potential application (Jose et al., 2010).

1.2 Problem Statement

Despite the availability of software to extract data regarding a person's sentiment on a specific product or service, organizations and other data workers still face issues regarding the data extraction. These issues are highlighted below.

1.2.1 Sentiment Analysis on Web Based Applications Focus on Single Tweet Only.

With the rapid growth of the World Wide Web, people are using social media such as Twitter which generates big volumes of opinion texts in the form of tweets which is available for the sentiment analysis (Lai, 2012). This translates to a huge volume of information from a human's viewpoint which make it difficult to extract a sentences, read them, analyze tweet by tweet, summarize them and organize them into understandable format in a timely manner.

1.2.2 Difficulty of Sentiment Analysis with Inappropriate English

The usage of inappropriate English and slangs that are used on social media such as Twitter has rendered making everyday decisions harder (Annett and Kondrak, 2009). Systems currently in place do not possess the ability to extract a person's sentiment if these have been conveyed using inappropriate English or slang words (Annett and Kondrak, 2009). Therefore there is a need for a system that is able to detect subjective data from social media and categories it to improve decision making

a. Informal Language

Informal language refers to the use of colloquialisms and slang in communication, employing the conventions of spoken language (Wikipedia, 2013c) such as „would not“ and „wouldn't“. Not all systems are able to detect sentiment from use of informal language and this could hanger the analysis and decision-making process.

b. Emoticons

Emoticons are a pictorial representation of human facial expressions, which in the absence of body language and prosody serve to draw a receiver's attention to the tenor or temper of a sender's nominal verbal communications, improving and changing its interpretation (Wikipedia, 2013b). For example ☺ indicates a happy state of mind. Systems currently in place do not have sufficient data to allow them to draw feelings out of the emoticons. As humans often turn to emoticons to properly express what they cannot put into words. Not being able to analyze this puts the organization at a loss.

c. Short-form

Short-form is widely used even with short message service (SMS). The usage of short-form will be used more frequently on Twitter so as to help to minimize the characters used. This is because Twitter has put a limit on its characters to 140 (Wikipedia, 2013a). Example „Tba“ refers to be announced.

1.3 Objective

1.3.1 To Study the sentiment analysis in microblogging which in view to analyze feedback from customer of organization's product.

1.3.2 To develop a program for customers' review on a product which allows an organization or individual to sentiment and analyze a vast amount of tweet into a useful format.

1.4 Scope of Study

- Sentiment Analysis using NLP technique under machine-learning based approach

This method need to be used to computationally measure customers' perceptions which is designed for a sentiment analysis to extract a vast

amount of tweets and inappropriate English.

- To focus on sentiment analysis on Twitter

The program will be developed by focusing on one of the most leading social networking which is Twitter because of a free format of messages and an easy accessibility of microblogging platforms, moreover internet users tend to shift from traditional communication tools to microblogging services (Pak and Paroubek, 2010).

- Using Python for system development

The program will be developed by Python which is an open source software (Nareyko, 2013). “Pattern” is a Python package for machine learning, natural language processing, web mining and network analysis with a focus on ease of use (Wiki, 2012). Moreover it can be learned through variety source online. As follow, this project is not required any costs. Besides, this project can be only one person developed the system. Through 8 months, the system can be completed.

CHAPTER 2

LITERATURE REVIEW

2.1 Social Media

Kalia (2013) and Wikipedia (2013c) defined a social media as a group of Internet-based applications that create on the ideological and technological foundations of Web 2.0 which is allowed to build and exchange of user generated contents. In a discussion of Internet World Start (2013) identified that a trend of internet users is increasing and continuing to spend more time with social media by the total time spent on mobile devices and social media in the U.S. across PC increased by 37 percent to 121 billion minutes in 2012 compared to 88 billion minutes in 2011. In the other hand, businesses uses social networking sites to find and communicate with clients, business can be demonstrated damage to productivity caused by social networking (Jung and Media, 2013; Kaplan and Haenlein, 2010). As social media can be posted so easily on the public which can be harm private information to spread out in the social world (Lohmann et al., 2012).

On the contrary, Tang et al., (2012) discussed that a benefits of participating in social media have gone beyond simply social sharing to build reputation and bring in career opportunities and monetary income. In addition, Hawkins (2013) and Kalia (2013) mentioned that the social media is also being used for advertisement and companies for promotions, professionals for searching, recruiting, social learning online and electronic commerce. Electronic commerce or E-commerce refers to the purchase and sale of goods or services online which can via by social media, such as Twitter which is convenient due to its 24-hours availability, ease of customer service and global reach (Hawkins, 2013).

These is the reason of why business tends to use more social media for insight into consumer behavioral tendencies, marketing intelligence and present an opportunity to learn about customer review and perceptions.

2.2 Twitter

Twitter is an online social networking and microblogging service that enables users to send and read "tweets", which are text messages limited to 140 characters (Pak and Paroubek, 2010; Jose et al., 2010; Lohmann et al., 2012). Micro-blogging is a term described by Wikipedia (2013b) as "a form of blogging that allow users to exchange small elements of content such as short sentences, individual images, or video links".

Twitter is the most popular social media, from Pak and Paroubek (2010) gave a reason that because of a free format of messages and an easy accessibility. Moreover Carpenter and Way (2010) also stated that because people always interested in what others think and comment. As a result of a large with 500 million registered users Twitter in 2012, who posted 340 million tweets per day and the service also handled 1.6 billion search queries per day (Holter, 2012; Weber, 2012; Farber, 2012). Therefore Twitter is contains a large valuable source of people"s perspective. In that, Twitter is the best platform to be used for a sentiment analysis.

2.3 Sentiment Analysis

Sentiment analysis refers to the general method to extract polarity and subjectivity from semantic orientation which refers to the strength of words and polarity text or phrases (Taboada et al., 2011). There has two main approaches for extracting sentiment automatically which are the lexicon-based approach and machine-learning-based approach (Annett and Kondrak, 2009; Goncalves et al., 2013; Kouloumpis et al, 2011; Sharma, 2008; Taboada et al., 2011).

2.3.1 Lexicon-based Approach

Lexicon-based methods make use of predefined list of words where each word is associated with a specific sentiment (Goncalves et al., 2013). The lexicon methods vary according to the context in which they were created and involve calculating orientation for a document from the semantic orientation of texts or phrases in the

documents (Taboada et al., 2011). Besides, Sharma and Dey (2012) also states that a lexicon sentiment is to detect word-carrying opinion in the corpus and then to predict opinion expressed in the text. Annett and Kondrak (2009) has showed the lexicon methods which has basic paradigm which are:

- a. Preprocess each tweet post by remove punctuation
- b. Initialize a total polarity score (s) equal 0 $\rightarrow s=0$
- c. Check if token is present in a dictionary, then

If token is positive, s will be positive (+)

If token is negative, s will be negative (-)

- d. Look at total polarity score of tweet post

If $s >$ threshold, tweet post as positive

If $s <$ threshold, tweet post as negative

However, Goncalves et al., (2013) mentions one advantage of leaning-based methods has ability to adapt and create trained models for specific purposes and contexts. In contrast, an availability of labeled data and hence the low applicability of the method on new data which is cause labeling data might be costly or even prohibitive for some tasks (Goncalves el al., 2013).

2.3.2 Machine-learning-based Approach

Machine learning methods often rely on supervised classification approaches where sentiment detection is framed as a binary which are positive and negative (Sharma and Dey, 2012). This approach requires labeled data to train classifiers (Goncalves el al., 2013). This approach, it become apparent that aspects of the local context of a word need to be taken in to account such as negative (e.g. Not beautiful) and intensification (e.g. Very beautiful) (Taboada et al., 2011). However, Annett and Kondrak (2009) have showed a basic paradigm for create a feature vector is:

- a. Apply a part of speech tagger to each tweet post
- b. Collect all the adjective for entire tweet posts
- c. Make a popular word set composed of the top N adjectives
- a. Navigate all of the tweets in the experimental set to create the following:
 - i. Number of positive words
 - ii. Number of negative words
 - iii. Presence, absence or frequency of each word

Taboada et al., (2011) showed some example of switch negation, negation simply to reverse the polarity of lexicon: changing beautiful (+3) into not beautiful (-3). More examples:

She is not terrific ($6-5=1$) but not terrible ($-6+5=-1$) either.

In this case, the negation of a strongly negative or positive value reflects a mixed perspective which is correctly captured in the shifted value. However, Goncalves et al., (2013) has mentioned the limitation of machine-learning-based approach to be more suitable for Twitter than the lexical based method. Furthermore, Annett and Kondrak (2009) stated that machine learning methods can generate a fixed number of the most regularly happening popular words which assigned an integer value on behalf of the frequency of the word in the Twitter.

2.4 Techniques of Sentiment Analysis

The semantic concepts of entities extracted from tweets can be used to measure the overall correlation of a group of entities with a given sentiment polarity (Saif et al., 2011). Polarity refer to the perhaps most basic form, which is if a text or sentence is positive or negative (Spencer and Uchyigit, 2008). However, sentiment analysis has techniques in assigning polarity such as:

2.4.1 Natural Language Processing (NLP)

NLP techniques are based on machine learning and especially statistical learning which uses a general learning algorithm combined with a large sample, a corpus, of data to learn the rules (Blom and Thorsen, 2012). Sentiment analysis has been handled as a Natural Language Processing denoted NLP, at many levels of granularity. Starting from being a document level classification task (Pang and Lee, 2008), it has been handled at the sentence level (Hu and Liu, 2004; Kim and Hovy, 2004) and more recently at the phrase level (Agarwal et al., 2012). NLP is a field in computer science which involves making computers derive meaning from human language and input as a way of interacting with the real world.

2.4.2 Case-Based Reasoning (CBR)

Case-Based Reasoning (CBR) is one of the techniques available to implement sentiment analysis. CBR is known by recalling the past successfully solved problems and use the same solutions to solve the current closely related problems (Nakov et al., 2013). Spencer and Uchyigit (2008) identified some of the advantages of using CBR that CBR does not require an explicit domain model and so elicitation becomes a task of gathering case histories and CBR system can learn by acquiring new knowledge as cases. This and the application of database techniques make the maintenance of large columns of information easier (Spencer and Uchyigit, 2008).

2.4.3 Artificial Neural Network (ANN)

Agarwal et al., (2012) mentioned that Artificial Neural Network (ANN) or known as neural network is a mathematical technique that interconnects group of artificial neurons. It will process information using the connections approach to computation. ANN is used in to find relationship between input and output or to find patterns in data (Spencer and Uchyigit, 2008).

2.4.4 Support Vector Machine (SVM)

Support Vector Machine is to detect the sentiments of tweets (Sharma, 2008). Pak and Paroubek (2010) together with Saif et al., (2012) stated SVM is able to extract and analyze to obtain up to 70% - 81.3% of accuracy on the test set. Wu et al., (2006) collected training data from three different Twitter sentiment detection websites which mainly use some pre-built sentiment lexicons to label each tweets as positive or negative. Using SVM trained from these noisy labeled data, they obtained 81.3% in sentiment classification accuracy.

2.4.5 Application Programming Interface (API)

Alchemy API performs better than the others in terms of the quality and the quantity of the extracted entities (Zhang et al., 2010). As time passed the Python Twitter Application Programming Interface (API) is created by collected tweets (Smedt and Daelemans, 2012). Python can automatically calculated frequency of messages being re-tweeted every 100 seconds, sorted the top 200 messages based on the re-tweeting frequency, and stored them in the designated database (Saif et al., 2011.). As the Python Twitter API only included Twitter messages for the most recent six days, collected data needed to be stored in a different database (Zhang et al., 2010).

2.5 Python

Python was found by Guido Van Rossum in Natherland, 1989 which has been public in 1991 (Wiki, 2012). Python is a programming language that available and solves a computer problem which is providing a simple way to write out a solution (Wiki, 2012). Seberino (2012) mentioned that Python can be called as a scripting language. Moreover, Lukaszewski (2010) and Seberino (2012) also supported that actually Python is a just description of language because it can be one written and run on many platforms. In addition, Nareyko (2013) mentioned that Python is a language that is great for writing a prototypes because Python is less time consuming and working prototype provided, contrast with others programming languages.

Many researchers have been said that Python is efficient especially to a complex

project, as Lukaszewski (2010) has mentioned that Python is suitable to start up social networks or media steaming projects which most always are a web-based which is driven a big data. Nareyko (2013) gave the reason that because Python can handle and manage the memory used. Besides Python creates a generator that allow an iterative processing of things, one item at a time and allow program to grab source data one item at a time to pass each through the full processing chain (Wiki, 2012).

2.6 Related Works

Twitter is available for the sentiment analysis since people is using Twitter to generate a big volume of opinion texts in form of tweets (Lai, 2012). However, a currently available system does not invalid to retrieve and analyze whole tweets in the same time (Lai, 2012; Rich, 2008). Moreover, the currently system does not have function to extract and analyze inappropriate English tweets in terms of informal language, emoticons and short-form (Annett and Kondrak, 2009). Thus the emergence of the need for a system that is able to remove all these drawbacks and improve the way data is analyzed and decisions are made.

2.6.1 “Build a Sentiment Analysis Application with JavaScript”

This application is created by Scott Rich in 2008. The end user needs to enter a single tweet for evaluation score (Rich, 2008).

- a. Start an application and enter a browser to <http://localhost:3000/testSentiment>.

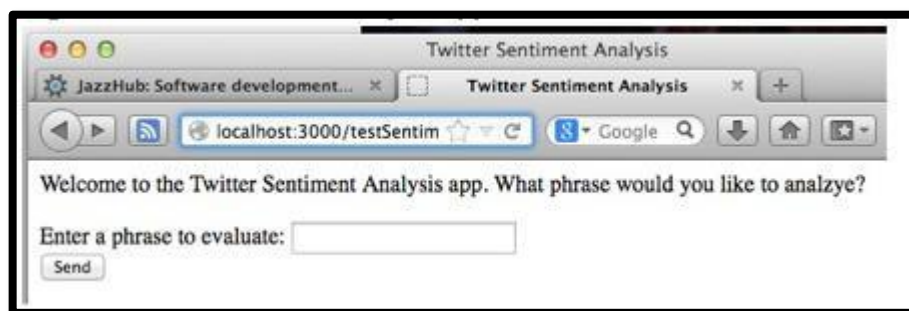


FIGURE 2. The Sentiment Analysis Application User Interface

- b. Click send, Application analyze and score a tweets. Here are some hypothetical results from sentiment analysis application.

Sentiment	Score	Interpretation
Node.js is cool, I love it	4	Very positive sentiment
Node.js is uncool, I hate it	-3	Very negative sentiment
Mi piace Node.js. Node.js é bellissima	0	No sentiment detected. Sentiment has only an English dictionary.
Node.js is not cool, I do not love it	4	Oops! Very positive, same score as the first example. Sentiment is not aware of grammar or negation.

FIGURE 3. Hypothetical Results of Analyzing Tweets

2.6.2 “TweetFeel”

This application named TweetFeel which is working well if user wants to look at the sentiment of a trend on Twitter (Widrich, 2010). Tweetfeel will check latest hundred tweets matched a search keyword by classify them in negative or positive within different colors. Other than end user cannot notify the positive and negative on their Twitter page, together with no neural tweets.



FIGURE 4. TweetFeel Application

2.6.3 “Twitter Sentiment”

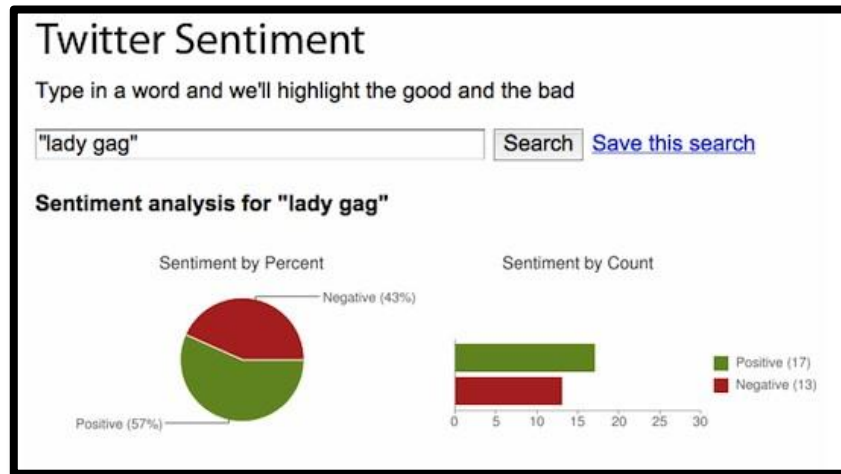


FIGURE 5. Twitter Sentiment Results Page

This program, user needs to enter a keyword and click search, then a program will analyze the latest 30 tweets which is containing the keyword and show a percentage of positive or negatives in different colors without neural tweets (Young, 2011). However, this program is providing a feedback page for users to comment whether results are correct or not at the end of page. Moreover, this program is allowed user to save a result page after signing in with Google account (Young, 2011).

CHAPTER 3

METHODOLOGY

3.1 Methodology

The methodology used in this project has been divided into 2 phases which are preliminary methodology and program developments.

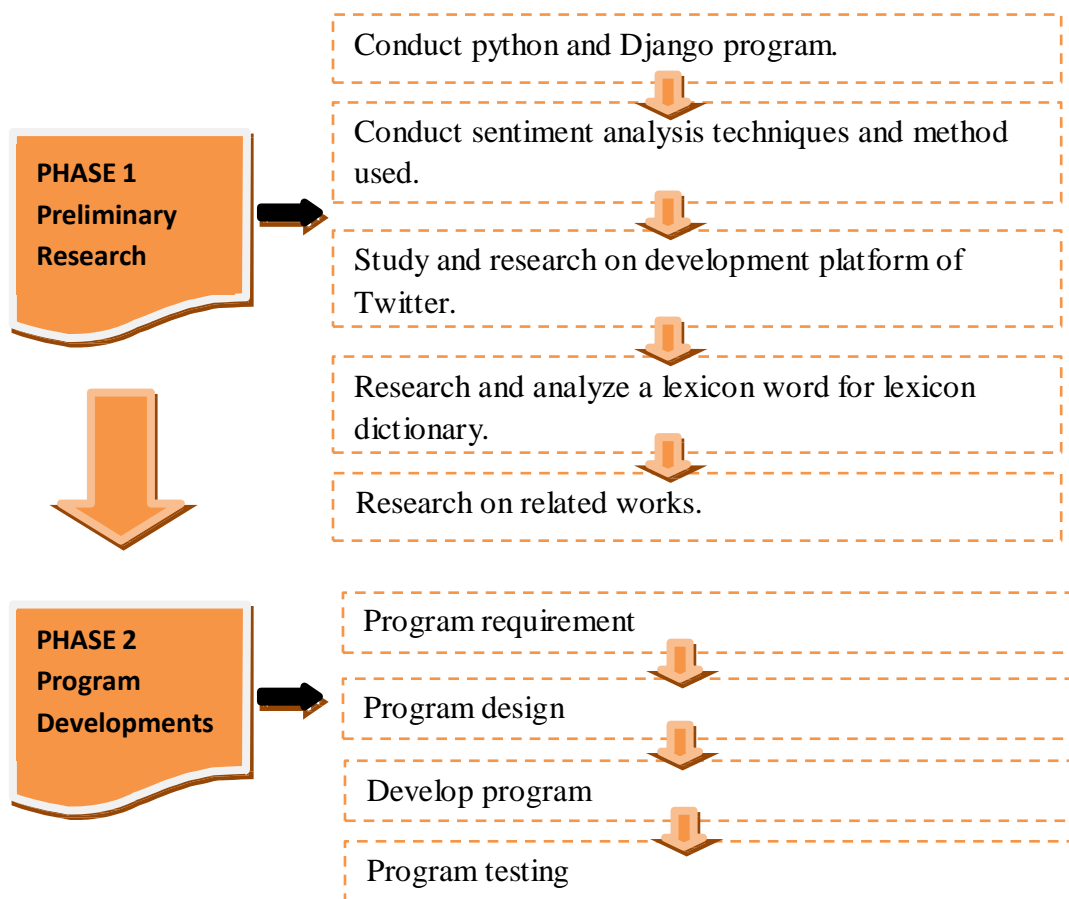


FIGURE 6. Methodology – 2 Phases

In the first phase is preliminary research which involve conducting research and gathering needed such as study on tools, development twitter platform, related works, and sentiment analysis techniques and method used by literature review.

The second phase is program development which is focused on defining the program requires need and functionality in relationship with the needs of the business. The desired output for this phase is a list of specific requirements to be designed and implemented during project. Besides, study on benefit gained out of project will also been studied. Next, program design is mainly focus on architectural, output designed and how its interaction by diagrams and interfaces. Then, program development with program testing will be at the end of this phase.

3.2 Milestones

TABLE 1. Milestones

Activities	Final Year Project I																Final Year Project II																
	October'13				November'13				December'13				January'14				February'14				March'14				April'14								
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4					
FYP Deliverable																																	
Title Proposal Submission		◆		17/10/2013																													
Extended Proposal Submission			◆	30/10/2013																													
Proposal Defence											◆	4/12/2013																					
Interim Report												◆	18/12/2013																				
Progress Report																																	
Pre-Sedex																																	
Disertation																																	
Online Technical Report and Desertation																																	
Viva																																	
Final desertation (Hard Bound)																																	

CHAPTER 4

RESULT AND DISCUSSION

4.1 Method Selection

Machine learning based approach will be used in Twitter sentiment analysis. The reasons are; machine learning based approach provides high accuracy of classification (mentioned in literature review section 2.3.2) and solves problem of computational linguistics (Blinov et al., 2012). However, machine learning based approach the task of sentiment analysis is regarded as a simple problem of text classification and it can be solved by training the classifier on a labeled text collection (Blinov et al., 2012). In contrast, lexicon based approach requires powerful linguistic resources which is not always available, from that it is difficult to take the context into account (Blinov et al., 2012). From this view point, Twitter sentiment analysis is using machine-based learning approach. As the following reasons:

TABLE 3. Comparison Between Machine-based Learning Approach and
Lexicon-based Approach

Machine-based learning approach	Lexicon-based approach
Solved by training the classifier on a labeled text collection	No need for labeled data and the procedure of learning
High accuracy of classification	Require powerful linguistic resources which it not always available

Machine-based learning approach	Lexicon-based approach
Solve various problem of computational linguistics	Calculate by method of relevance frequency (RF)

4.2 Techniques

Techniques that has been used in Twitter sentiment analysis is a Natural Language Processing (NLP) which incorporates processing algorithms to provide certain rules and guidelines of speech and language in order for examination and analysis. Blinov et al. (2012) have come up with some techniques that are useful for Twitter sentiment analysis are:

4.2.1 Tokenization

To consider punctuation sign that system need to pay attention, for example “e.g.” is a one token, two token or four token. This step is usually fairly straightforward and rule-based.

4.2.2 Structure and Sentence Detection

NLP in an auto-coding application needs to identify the section of clinical narrative. Moreover, the system need to figure it out when it end of an abbreviation and when it end of a sentence such as “Mr.” which is does not mean the end of sentence. For human, it is very easy to identify but for it difficult for computer to consistently get right.

4.2.3 Part of Speech (POS)

A part of speech for a word is complicated like “there are cats and dogs” or “Having rain like cats and dogs.” NLP needs to understand a world like “cats”, “dogs” and “cats and dogs” which is must be able to understand the context in which the word is used and interpret the surrounding terms.

4.3 Networking Connection Issues

Twitter sentiment analysis cannot retrieve a tweet through Internet in Universiti Teknologi PETRONAS (UTP) accrediting to security issue, following these reason:

- UTP Internet is having some issues using the library from behind a proxy because UTP internet only allows HTTP request, no other protocol is opened by ITMS.
- UTP Internet does not allow multiple connections when the program is running to get tweets from twitter.
- UTP internet has user name and password policy problem.

Example username : student
 password : UTP@2013.

However, from this issue is creating a problem to run python program on line because proxy: (160.0.226.8:808) has been setting like proxy: <http://student:UTP@2013@160.0.226.8:808/> which is making student as username, UTP as password and "2013@160.0.226.6" as proxy server are not valid. Together with 808 as port for incoming and outgoing. The reason is, UTP@2013 password is including '@' symbol.

- To create an instance of the Twitter API with login credentials, many API calls required the client to be authenticated which is blocked by UTP Wi-Fi and LAN because of proxy connection.
- Twitter sentiment analysis does not need to run twitter API from a proxy, in order to multiply our request limit. All we have to do is to create an application with the respective keys from different twitter accounts.

- UTP internet does not allow traceback URL availability for fetching tweets in different way.

From these issues, twitter sentiment analysis need an Internet from data plan to retrieve tweets at the first process.

4.4 Project Development Procedures

4.4.1 Register and Sign in

Register and sign in on <https://dev.twitter.com/>.

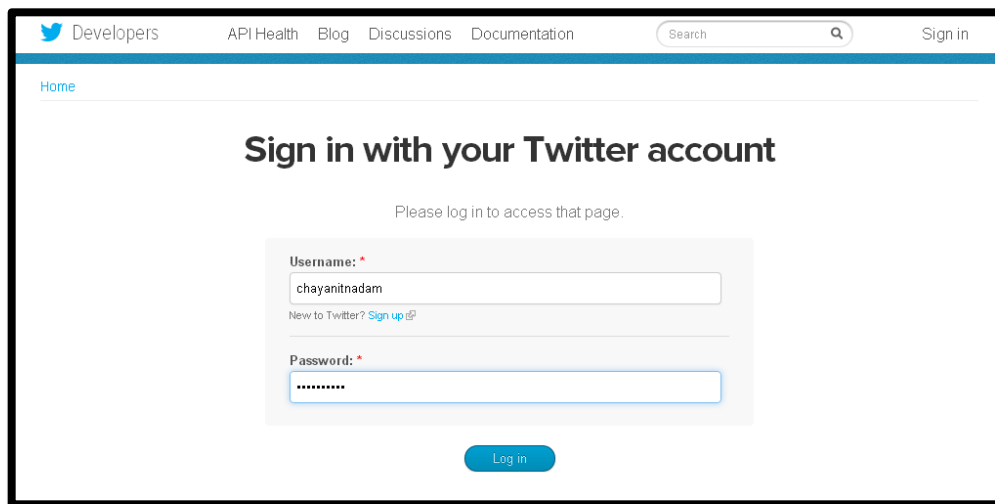


FIGURE 7. Sign In Twitter Developer

4.4.2 Create an Application

Create an application on development twitter platform that has been provided. A developer need to agree on terms and conditions.

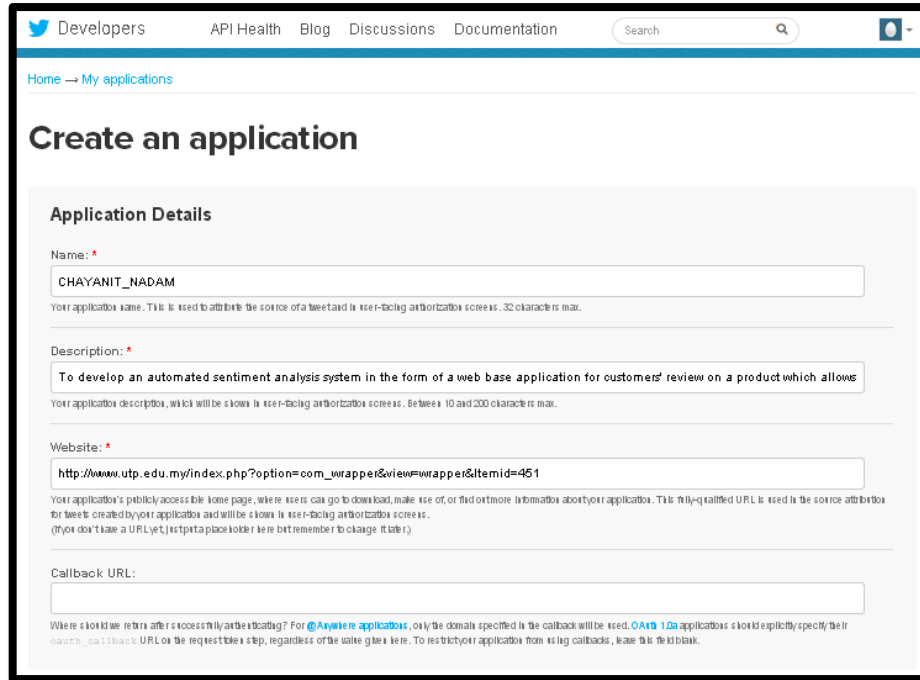


FIGURE 8. Create an Application

4.4.3 Parse API Response

Developer will get important keys. Then, use consumer key and secret key to request barer token for using Python library such as Oauth for an authorization.

TABLE 4. Parse API Response Keys

Consumer key	UMyUKVMcGwAmjqsTaC0xw
Consumer secret	NyFsbcmloLKKGs89jrwVwSjgGh9sNQVIirTw2hzBZo
Access token	2241761480- 9E0v1IYJNarBuNVPZigtDqYP9POhXz9O1XtJahS
Access token secret	Y1IuiUqrkCwt6KgawMeZ7WlpWPoptlFkzYwk0q6UhEXBh

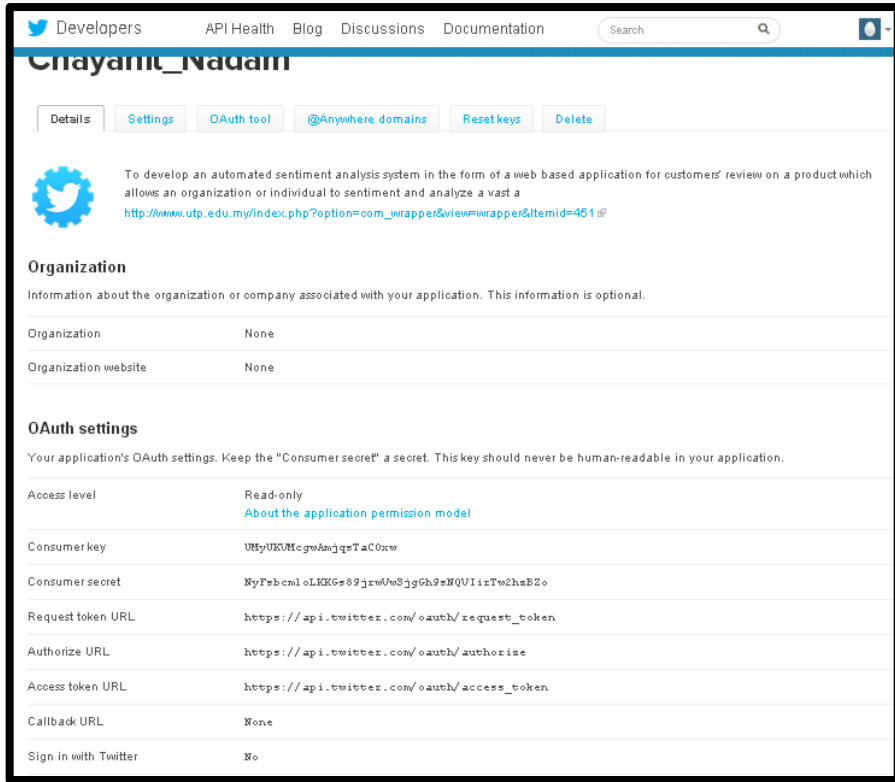


FIGURE 9. OAuth Settings for Coding Process

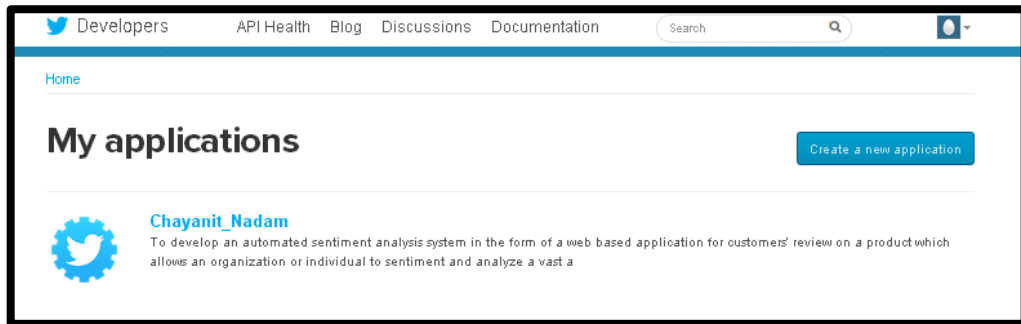


FIGURE 10. Application Created

4.5 Functional Modeling

4.5.1 Activity Diagram of Sentiment Analysis

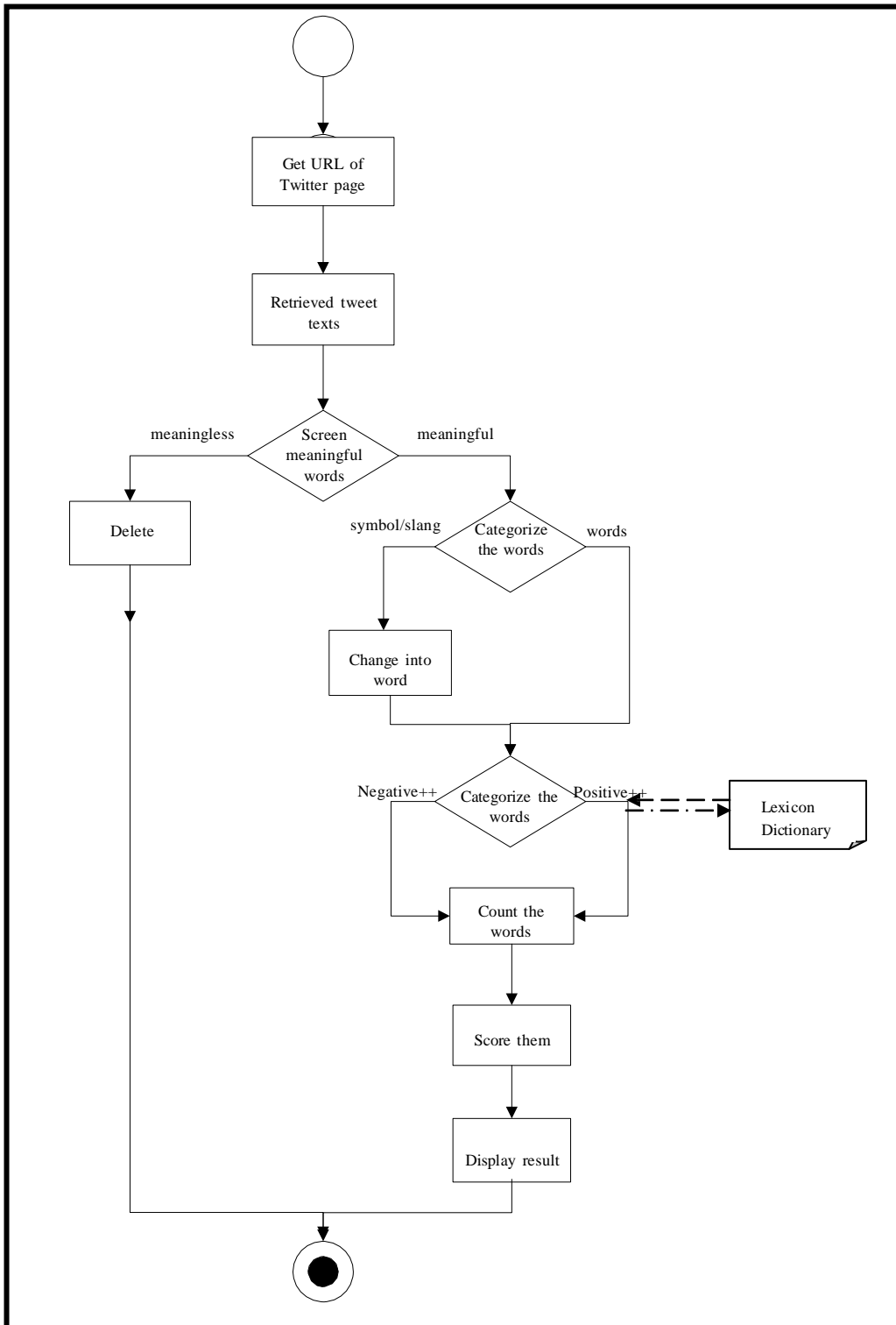


FIGURE 11. Activity Diagram

4.5.2 Description of activity processes

TABLE 5. Description of Activity Processes

Process/Activity	Description
Get URL of Twitter page	Get URL of Twitter page that user want to sentiment analysis on.
Retrieved tweet texts	The program will be automated retrieve entire tweet texts from URL page that user has selected.
Screen meaningful words	<p>The program will determine the meaningful and meaningless word for categorize onward.</p> <p>If meaningless word, it will be deleted</p> <p>If meaningful word, it will be categorize in positive or negative word.</p>
Delete	The program will delete the meaningless words.
Categorize the words	The program will categorize the meaningful words into symbol/slang and words.
Check the words	The program will categorize the words into positive or negative by following the lexicon dictionary.
Count the words	The program will count how many frequency of each positive and negative words have been use on Twitter selected.

Process/Activity	Description
Score them	Assign the value of each word for sentiment analysis to calculate positive or negative Twitter page.
Get result	The result will be shown in a form of graph, top 10 lists of frequency each positive and negative words, together with percentage of negative and positive from customer's perspective.

4.5.3 Use Case Diagram

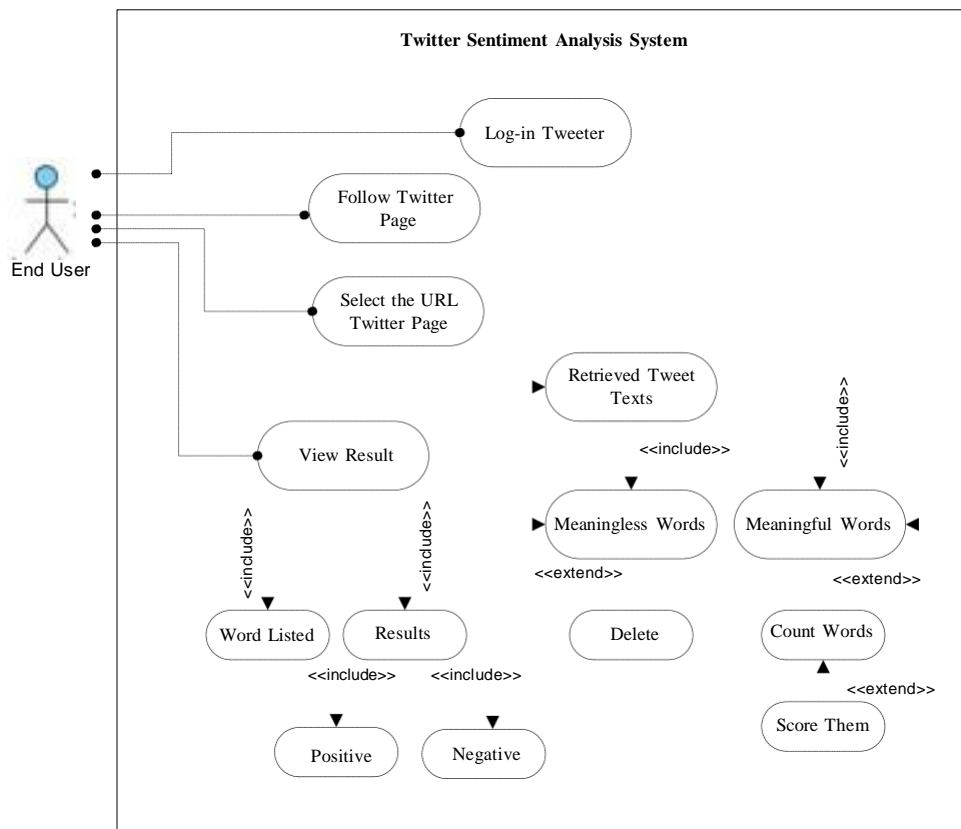


FIGURE 12. Use Case Diagram

4.6 System Architecture

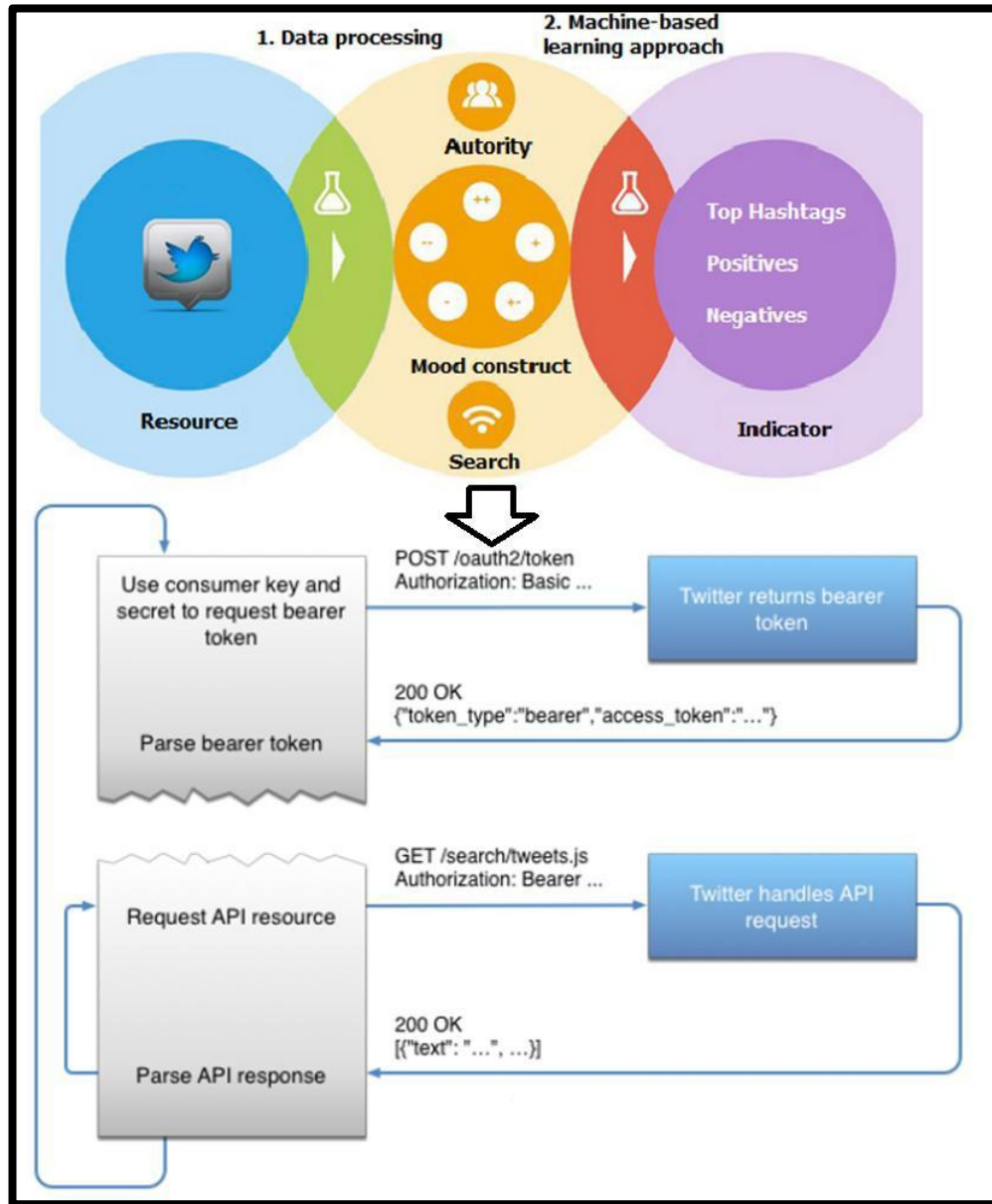


FIGURE 13. System Architecture

4.6.1 Source

Development Twitter platform and twitter pages which is providing a tweet for analyzing a sentiment.

4.6.2 Authority

Program needs to search for authority before accessing a source to do a sentiment analysis into positive and negative. However, Twitter development platform will provide access token key and token key to allow a program to access their data. After that, program will match a word in each sentence of tweet with lexicon dictionary (mood construct) for assigning the value of each word and analyze a sense of each tweets.

4.6.3 Indicator

Once, program has accessed data and categorizes a sentiment word. Program will represent a result in a form of pie chart and html page.

4.7 Results

Results have been divided into 3 parts, which are;

4.7.1 Twitter Retrieved

To associate with Twitter API, developer need to agree in terms and conditions of development Twitter platform which has been provided to get an authorization to access a data.

Output from this process will be saved in JSON file. The reason is, JSON (JavaScript Object Notation) is a lightweight data-interchange format which is easy for humans to write and read (Droettboom, 2014). Moreover, Droettboom (2014) stated that, JSON is simple for machines to generate and parse. JSON is a text format that is totally language independent but uses convention that is known to programmers of the C-family of

languages, including Python and many others (Droettboom, 2014). However, output's size depends on the time for retrieving tweets from Twitter.

Nevertheless, the output will be categorized into 2 forms, which are encoded and un-encoded. According to security issue for accessing a data, some of output will be showed in an ID form such as string ID.

4.7.2 Sentiment Analysis

Tweets from JSON file will be assigned the value of each word by matching with the lexicon dictionary. As a limitation of words in the lexicon dictionary which is not able to assign a value to every single word from tweets. However, as a scientist language of python which is able to analyze a sense of each tweet into positive or negative for getting a result.

4.7.3 Information Presented

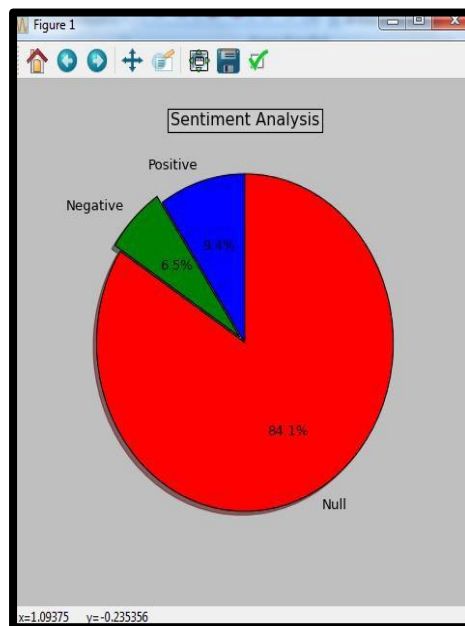


FIGURE 14: Pie Chart

The result will be shown in a pie chart which is representing a percentage of positive, negative and null sentiment hash tags. For null hash tag is representing the hash tags that were assigned zero value. However, this program is able to list a top ten positive and negative hash tags.

Refer to figure 14, the pie chart is representing of each percentage positive, negative and null sentiment hash tags in different color.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

In this project, twitter sentiment analysis is developed to analyze customers' perspective which crucial towards organization's critical success in the marketplace. The program is using a machine-based learning approach which found to be more accurate for analyzing a sentiment; together with natural language processing techniques be used.

The result, program has categorized sentiment into positive and negative which is represented in pie chart and html page, although, the program has been planned to be developed as a web application, due to limitation of Django which can only work on Linux server or LAMP. However, it cannot be realized. Therefore, further enhancement of this element is recommended in future study.

Moreover, this project is only focusing on one available URL platform that has provided by Twitter for code testing purpose. Therefore, this program can be enhanced to focus on single keyword on multiple twitter pages for various products.

REFERENCES

- Agarwal, A., Xie, B., Vovsha, I., Rambow, O., and Passonneau, R. (2012). *Sentiment Analysis of Twitter Data*. Annul International Conferences. New York: Columbia University.
- Annett, M., and Kondrak, G. (2009). *A Comparison of Sentiment Analysis Techniques: Polarizing Movie Blogs*. Conference on web search and web data mining (WSDM). University of Alberia: *Department of Computing Science*.
- Blinov, P.D. (2012). *Research of lexicon approach and machine learning method for sentiment analysis*. Proceeding of the Conference on Empirical mehotds in natural language processing (EMNLP). Russia: *Vyatka State Humanities University*.
- Blom, A., and Thorsen, S. (2012). *Automatic Twitter replies with Python*. International conference “Dialog 2012”.
- Carpenter, T., and Way, T. (2010). *Tracking Sentiment Analysis through Twitter*. ACM computer survey. Villanova: Villanova University.
- Droettboom, M. (2014). *Understanding JSON Schema Release 1.0*. Space Telescope Science Institute.
- Farber, D. (2012). *Twitter hits 400 million tweets per day, mostly mobile*. Retrieved 10 18, 2013 from: http://news.cnet.com/8301-1023_3-57448388-93/twitter-hits-400-million-tweets-per-day-mostly-mobile/
- Goncalves, P., Benevenuto, F., Araujo, M., and Cha, M. (2013). *Comparing and Combining Sentiment Analysis Methods*.

- Hawkins, A. (2013). There is more to becoming a thought leader than giving yourself the title. Retrieved 10 18, 2013 from: <http://www.thesocialmediashow.co.uk/author/admin/>
- Holter, K. (2012). Twitter: 340 million tweets, 140 million daily active users. Retrieved 10 21, 2013 from: <http://www.dailydot.com/business/twitter-340-million-tweets-140-million-daily-users/>
- Hu, M and Liu, B. (2004). Mining and summarizing customer reviews. ACM computer survey
- Internet World Start (2013). Usage and Population Statistic. Retrieved 10 15, 2013 from: <http://www.internetworldstats.com/stats.htm>
- Jose, A.K., Bhatia, N., and Krishna, S. (2010). Twitter Sentiment Analysis. National Institute of Technology Calicut.
- Jung, B., and Media, D. (2013). The Negative Effect of Social Media on Society and Individuals. Retrieved 10 23, 2013 from: <http://smallbusiness.chron.com/negative-effect-social-media-society-individuals-27617.html>
- Kalia, G. (2013). A Research Paper on Social Media: *An Innovative Educational Tool*. (Vol. 1, pp. 43-50). Chitkara University.
- Kaplan, A.M., and Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. France: *Paris*.
- Kim, S and Hovy, E. (2004). Determining the Sentiment of Opinions.
- Kouloumpis, E., Wilson, T., and Moore, J. (2011). Twitter Sentiment Analysis: *The Good the Bad and the OMG!* (Vol. 5). International AAAI.

- Lai, P. (2012). Extracting Strong Sentiment Trend from Twitter. Stanford University.
- Lohmann, S., Burch, M., Schmauder, H., and Weiskopf, D. (2012). Visual Analysis of Microblog Content Using Time-Varying Co-occurrence Highlighting in Tag Clouds. Annual conference of VISVISUS. Germany: *University of Stuttgart*
- Lukaszewski, A. (2010). MySQL for Python. *Integrate the flexibility of Python and the power of MySQL to boost the productivity of your applications*. UK: *Birmingham*. Packt Publishing Ltd.
- Nareyko, V. (2013). Why python is perfect for startups. Retrieved 01 10, 2014 from: <http://opensource.com/business/13/12/why-python-perfect-startups>
- Nakov, P., Kozareva, Z., Ritter, A., Rosenthal, S., Stoyanov, V., and Wilson, T. (2013). SemEval-2013 Task 2: *Sentiment Analysis in Twitter* (Vol.2, pp. 312-320). Georgia.
- Osimo, D., and Mureddu, F. (2010). Research Challenge on Opinion Mining and Sentiment Analysis. Proceeding of the 12th conference of Fruct association. United Kingdom.
- Pak, A., and Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. Special Issue of International Journal of Computer Application. France: Universite de Paris-Sud.
- Pang, B., and Lee, L. (2008). Opinion mining and sentiment analysis. 2nd workshop on making sense of Microposts. Ithaca: Cornell University. Vol.2(1).
- Prabowo, R., and Thelwall, M. (2009). *Sentiment Analysis: A Combined Approach*.

International World Wide Web Conference Committee (IW3C2). United Kingdom: University of Wolverhampton.

- Rambocas, M., and Gama., J. (2013). Marketing Research: *The Role of Sentiment Analysis*. The 5th SNA-KDD Workshop“11. University of Porto.
- Rich, S. (2008). *Build a sentiment analysis application with Node.js, express, sentiment and ntwitter*. IBM Corporation 2013.
- Saif, H., He, Y., and Alani, H. (2011). Semantic Sentiment Analysis of Twitter. *Proceeding of the Workshop on Information Extraction and Entity Analytics on Social Media Data*. United Kingdom: Knowledge Media Institute.
- Saif, H., He, Y., and Alani, H. (2012). Alleviating Data Scarcity for Twitter Sentiment Analysis. Association for Computational Linguistics.
- Seberino, C. (2012). Python. *Faster and easier software development*. Annual Conference. California: *San Diego*.
- Sharma, S. (2008). Application of Support Vector Machines for Damage detection in Structure. *Journal of Machine Learning Research*.
- Sharma, A., and Dey, S. (2012). Performance Investigation of Feature Selection Methods and Sentiment Lexicons for Sentiment Analysis. Association for the advancement of Artificial Intelligence.
- Smedt, T.D., and Daelemans, W. (2012). Pattern for Python. *Proceeding of COLING 2012*. Belgium: University of Antwerp.
- Spencer, J., and Uchyigit, G. (2008). Sentimentor: *Sentiment Analysis of Twitter Data*. Second joint conference on lexicon and computational semantics.

Brighton: University of Brighton.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. Association for computational Linguistics.

Tang, Q., Gu, B., and Whinston, A.B. (2012). Content Contribution in Social Media: *The case of YouTube*. 2nd conference of social media. Hawaii: *Maui*.

Weber, H., (2012). With 140 million active users & 340 million tweets per day, Twitter is officially mainstream. Retrieved 10 18, 2013 from: <http://thenextweb.com/socialmedia/2012/03/21/twitter-has-over-140-million-active-users-sending-over-340-million-tweets-a-day/>

Widrich, L. (2010). 6 Tools For Twitter Sentiment Tracking. Retrieved 03 10, 2014 from: <http://socialmouths.com/blog/2010/03/31/6-tools-for-twitter-sentiment-tracking/>

Wikipedia (2012). Python (programming language). Retrieved 1 10, 2014 from: http://en.wikipedia.org/wiki/Python_%28programming_language%29

Wikipedia (2013a). Microblogging. Retrieved 10 14, 2013 from: <http://en.wikipedia.org/wiki/Microblogging>

Wikipedia (2013b). Spoken Language. Retrieved 10 14, 2013 from: http://en.wikipedia.org/wiki/Spoken_language

Wikipedia (2013c). Social Media. Retrieved 10 12, 2013 from: http://en.wikipedia.org/wiki/Social_media

Wu, J., Wang, J., & Liu, L., (2006). Kernel-Based Method for Automated Walking Patterns Recognition Using Klnematics Data. 5th Workshop on Natural Language Processing. China: *Xi'an Jiaotong University*.

Young. (2011). Best 5 Sentiment Analysis Tools For Twitter. Retrieved 04 01, 2014
from: <http://freenuts.com/best-5-sentiment-analysis-tools-for-twitter/>

Zhang, J., Qu, Y., Cody, J., & Wu, Y., (2010). A case study of Microblogging in the Enterprise: *Use Value, and Related Issues*. Proceeding of the workshop on Web 2.0.

APPENDIX

1. Retrieve a

```
Tweet import
oauth2 as oauth import
urllib2 as urllib

# put these credentials
access_token_key = "2342537311-
OnOQxsSWixeWGbyN4fm65PTwsORsLQFT2FHn56Z" access_token_secret =
"mWT70cMxmSjuMUFE4LisH7gli4sHmFpMUT2GKRflJjYiX"

consumer_key = "qMPDRMT0FO7yOA4wp1X2TQ"
consumer_secret = "nAxH5rVpIdSMTe6aSmP220RMfYpB9mJHnI6E73bqI"

_debug = 0

oauth_token = oauth.Token(key=access_token_key, secret=access_token_secret)
oauth_consumer = oauth.Consumer(key=consumer_key,
secret=consumer_secret) signature_method_hmac_sha1 =
oauth.SignatureMethod_HMAC_SHA1() http_method = "GET"

http_handler = urllib.HTTPHandler(debuglevel=_debug)
https_handler = urllib.HTTPSHandler(debuglevel=_debug)

def twitterreq(url, method, parameters):
    req =
        oauth.Request.from_consumer_and
        _token(oauth_consumer,
        token=oauth_token,
        http_method=http_method,
        http_url=url,
        parameters=parameters)

    req.sign_request(signature_method_hmac_sha1, oauth_consumer,
oauth_token) headers = req.to_header()

    if http_method == "POST":
```



```

    encoded_post_data =
req.to_postdata() else:
    encoded_post_data =
None url = req.to_url()

opener =
urllib.OpenerDirector()
opener.add_handler(http_handl
er)
opener.add_handler(https_hand
ler)

response = opener.open(url, encoded_post_data)

return

response def

fetchsamples(

):
    url =
"https://stream.twitter.com/1/statuses/sample.json"
    parameters = []
    response = twitterreq(url, "GET", parameters)
    for line in response:
        print line.strip()

if __name__ == '__main__':
    fetchsamples()

```

2. Un-encoded

Program will direct match with lexicon dictionary. Then, program will assign a value according to lexicon dictionary.

```

import sys
import json
import re

def main():

    # load a tab delimited dict of sentiment
    scores a finnfile = open(sys.argv[1])
    scores = {}

```

```

for line in afinnfile:
    term, score = line.split("\t")
    scores[term] = int(score)

# load each tweet as json
for line in open(sys.argv[2]):
    score = 0
    tweet_json = json.loads(line)

    # only accept records with a 'text'
    field if tweet_json.get('text'):
        tweet_text = tweet_json['text'].encode('utf8').split()
        for word in tweet_text:

            if re.match("[A-Za-z0-9_-]*$", word):
                score +=
                    scores.get(word, 0)
            print word, float(score)
            #print float(score)
        if __name__ == '__main__':
            main()

```

3. Encoded

Program is encoding output.json format to human readable format and assigning sentiment value according to lexicon dictionary.

```

import
sys
import
json
import
re

def main():
    # load a tab delimited dict of sentiment
    scores afinnfile = open(sys.argv[1])
    scores = {}
    for line in afinnfile:
        term, score = line.split("\t")
        scores[term] = int(score)

    tweet_sentiments = {}
    word_sentiments = {}

    # load each tweet as json

```

```

for line in open(sys.argv[2]):
score = 0
tweet_json = json.loads(line)

# only accept records with a 'text'
field if tweet_json.get('text'):
tweet_text = tweet_json['text'].encode('utf8').split()
for word in tweet_text:
# only read alphanumeric words and mentions (e.g., "@username")
if re.match("^[@|A-Za-z0-9_-]*$", word):
score += scores.get(
word, 0) tweet_senti
ments[line] = score for
everyword in tweet_text:
# only read alphanumeric words and mentions (e.g., "@username")
if re.match("^[@|A-Za-z0-9_-]*$", everyword):
word_sentiments[everyword] = tweet_sentiments[line]

# empty dict for storing non-AFINN word
sentiments_unscored_words = {}

# cycle through each tweet, keep running tweet score for each word not in AFINN
for line in open(sys.argv[2]):
tweet_json = json.loads(line)

# only accept records with a 'text'
field if tweet_json.get('text'):
tweet_text = tweet_json['text'].encode('utf8').split()
for word in tweet_text:
# only read alphanumeric words and mentions (e.g., "@username")
if re.match("^[@|A-Za-z0-9_-]*$",
word):
if not
scores.get(word):
unscored_words[word] = word_sentiments[word]

# print full dict <term:string> <sentiment:float>
for key,value in unscored_words.items():
print key,
float(value)

if __name__ == '__main__':
main()

```

4. Frequency Calculation

Program will calculate of frequency all the tweet terms based on output.json file from

Twitter.

```
import
sys
import
json
import
re
from collections import Counter

def main():
    all_words = []
    # load each tweet as json
    for line in open(sys.argv[1]):
        tweet_json = json.loads(line)
        # only accept records with a 'text'
        field if tweet_json.get('text'):
            tweet_text = tweet_json['text'].encode('utf8').split()
            for word in tweet_text:
                # only read alphanumeric words and mentions (e.g.,
                "@username") if re.match("^[| @ A-Za-z0-9_]*$", word):
                    all_words.append(word)

    words_hash = Counter(all_words)
    denominator = float(sum(words_hash.values()))
    frequency_dict = {}
    for (key, value) in words_hash.items():
        frequency_dict[key] = float(value/denominator)
```

```

# print term frequencies <term:string>
<frequency:float>
for (key, value) in
frequency_dict.items():
    print key,
    value

if __name__ == '__main__':
    main()

```

5. Top Ten Hashtags

```

import
sys
import
json
from collections import Counter

def main():

    hashtags = []

    # load each tweet as json
    for line in open(sys.argv[1]):
        tweet_json = json.loads(line)

        # if a tweet has a
        hashtag...
        if "entities" in tweet_json.keys() and "hashtags" in
        tweet_json["entities"]:
            # and isn't
            blank ...
            if tweet_json["entities"]["hashtags"] !=
            []:
                # append each hashtag (in
                unicode)
                for hashtag in
                tweet_json["entities"]["hashtags"]:
                    unicode_hashtag =
                    hashtag["text"].encode('utf-8')
                    hashtags.append(unicode_hashtag)

    # print top ten hashtag counts to
    stdout
    top_ten =
    Counter(hashtags).most_common(10)
    for key, value in
    top_ten:

```

```

    print      key,
    float(value)

if __name__ == '__main__':
    main()

```

6. Positive Listed

```

import
csv
import
sys

sentimentData = open(sys.argv[1] + '#abcd1.csv sorted excel with decending
order')
csvReader1 = csv.reader(sentimentData)
for i in range(10):
    print csvReader1.next()

```

7. Negative Listed

```

import csv
import sys
import
HTML

sentimentData = open
#sentimentData = sys.argv[1] + '#abcd1.csv sorted excel with Ascending
order')
csvReader1 = csv.reader(sentimentData)
for i in range(10):
    print csvReader1.next()

```

8. Result Displayed

```

import csv
import sys
import
HTML
from pylab import *

sentimentData_pos = open
#sentimentData = sys.argv[1] + '#abcd1.csv sorted excel with decending
order')
csvReader1 = csv.reader(sentimentData_pos)

```

```

HTMLFILE = HTML_output.html'
f= open(HTMLFILE, 'w')
f.write('Top ten positive tags , value <tr>')
for i in range(10):
    #print csvReader1.next() table_data_pos =
    csvReader1.next() htmlcode_pos =
    HTML.table(table_data_pos) print
    htmlcode_pos
    f.write(htmlcode_pos)
f.write('<tr
>')
print '!'*79

# make a square figure and
axes figure(1, figsize=(6,6))
ax = axes([0.1, 0.1, 0.8, 0.8])

# The slices will be ordered and plotted counter-
clock wise. labels = 'Positive', 'Negative', 'Null'#, 'Logs'
fracs = [x, y, z]#, 10]
explode=(0, 0.05, 0)#, 0)

piecode = pie(fracs, explode=explode, labels=labels,
    autopct='%1.1f%%', shadow=True,
    startangle=90)
# The default startangle is 0, which would start
# the Frogs slice on the x-axis. With startangle=90,
# everything is rotated counter-clock wise by 90 degrees,
# so the plotting starts on the positive y-axis.

title('Sentiment Analysis', bbox={'facecolor':'0.8',

'pad':5})

#f.write(piecode)
show()
f.close()
print '\nOpen the file %s in a browser to see the result.'% HTMLFILE

```