



SENTIMENT ANALYSIS BASED ON EMOTICONS

CSE 3021: SOCIAL AND INFORMATION NETWORKS

A2+TA2

SUBMITTED TO: PROF.MEENAKSHI S P

submitted by

VINEET NALAWADE(15BCE0284)

SATYAK BABAR (15BCE0150)

RAJAT MUKATI (15BCE0538)

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

NOVEMBER 2017

CERTIFICATE

This is to certify that the project work entitled “Collection of tweets and performing sentiment analysis based on emoticons and NLP (Text Blob) ” that is being submitted by “VINEET NALAWADE,SATYAK BABAR,RAJAT MUKATI” for SOCIAL AND INFORMATION NETWORKS(CSE3021) is a record of bonafide work done under my supervision. The contents of this Project work, in full or in parts, have neither been taken from any other source nor have been submitted for any other CAL course.

Place: Vellore

Date- 07/11/2017

ACKNOWLEDGEMENTS

We feel immense pleasure while expressing our deep sense of gratitude to each and every individual co-operating with this work of ours. First of all, we are very much thankful to our Course Faculty Prof. MEENAKSHI S P for providing us with the great opportunity to accomplish this work as our project. The timely guidance and high-end encouragement we received from him is the mightiest arrow in our quiver. In the realization of this project, there is also a great helping hand of our colleagues whose valuable support is making us to chase down the track.

VINEET NALAWADE(15BCE0284)

SATYAK BABAR (15BCE0150)

RAJAT MUKATI (15BCE0538)

1. INTRODUCTION

With the advancement of web technology and its growth, there is a huge volume of data present in the web for internet users and a lot of data is generated too. Internet has become a platform for online learning, exchanging ideas and sharing opinions.

Social networking sites like Twitter, Facebook, Google+ are rapidly gaining popularity as they allow people to share and express their views about topics, have discussion with different communities, or post messages across the world. There has been lot of work in the field of sentiment analysis of twitter data.

As people increasingly use emoticons in text in order to express, stress, or disambiguate their sentiment, it is crucial for automated sentiment analysis tools to correctly account for such graphical cues for sentiment. We analyze how emoticons typically convey sentiment and demonstrate how we can exploit this by using a novel, manually created emoticon sentiment lexicon in order to improve a state-of-the-art lexicon-based sentiment classification method. We evaluated our approach on 2,080 Dutch tweets and forum messages, which all contain emoticons and have been manually annotated for sentiment. On this corpus, paragraph-level accounting for sentiment implied by emoticons significantly improves sentiment classification accuracy. This indicates that whenever emoticons are used, their associated sentiment dominates the sentiment conveyed by textual cues and forms a good proxy for intended sentiment.

2. PROBLEM STATEMENT

- ☐ Extraction of tweets from Twitter based on a topic.
- ☐ Sentimental analysis using emoticons.
- ☐ Classifying the sentiments into three subcategories:-
 1. Positive
 2. Negative
 3. Neutral
 4. Confused

3. LITERATURE SURVEY

Twitter Sentiment Analysis with Emoticons WAGHODE POONAM B,PROF. MAYURA KINIKAR MIT ACADEMY OF ENGINEERING, ALANDI, PUNE, MAHARASHTRA, INDIA . According to them Sentiment analysis, which is also called opinion mining is the field of study which analyzes people's opinions, sentiments, evaluations, appraisals, attributes and emotions towards entities such as products services, organizations, individuals, issues, events, topics, and their attributes through twitter. People use microblogging (twitter) to talk about their daily activities and to seek or share information. It is an online social networking and micro-blogging service that enables users to send and read "tweets", which are text messages limited to 140 characters. In this paper we propose a model that can spot the public opinion with their emotions.

Exploiting Emoticons in Sentiment Analysis by ALEXANDER HOGENBOOM. According to them as people increasingly use emoticons in text in order to ex-press, stress, or disambiguate their sentiment, it is crucial for automated sentiment analysis tools to correctly account for such graphical cues for sentiment. We analyze how emoti-cons typically convey sentiment and demonstrate how we can exploit this by using a novel, manually created emoti-con sentiment lexicon in order to improve a state-of-the-art lexicon-based sentiment classi cation method. We evalu-ate our approach on 2,080 Dutch tweets and forum mes-sages, which all contain emoticons and have been manually annotated for sentiment. On this corpus, paragraph-level accounting for sentiment implied by emoticons signi cantly improves sentiment classi cation accuracy. This indicates that whenever emoticons are used, their associated senti-ment dominates the sentiment conveyed by textual cues and forms a good proxy for intended sentiment.

4. WORK DONE AND IMPLEMENTATION:

☐ **LANGUAGES AND TOOLS USED**

- ☐ Python
- ☐ M.S. excel
- ☐ Javascript

☐ **DATA SET**

The data set consists of the set of the tweets posted on twitter obtained with the help of Tweepy.

☐ **TOOLS AND LIBRARIES USED**


Tweet listener


1. Tweepy
2. Json
3. Textblob


Sentiment analysis


1. Re(regular expression)
2. Codecs
3. Random
4. Pickle
5. Maths
6. Numpy

5. METHODOLOGY

 **Collection of Data:** Twitter provides REST APIs that can be used to interact with their service. There is also a bunch of Python-based clients that we could help in writing the queries. In particular, Tweepy is one of the most interesting and straightforward to use.

 **Text Pre-processing:** After collecting a number of tweets we can store them in JSON. **JSON** (JavaScript Object Notation) is a lightweight data-interchange format that makes the parsed text easy for humans to read. It is based on a subset of the JavaScript Programming Language.

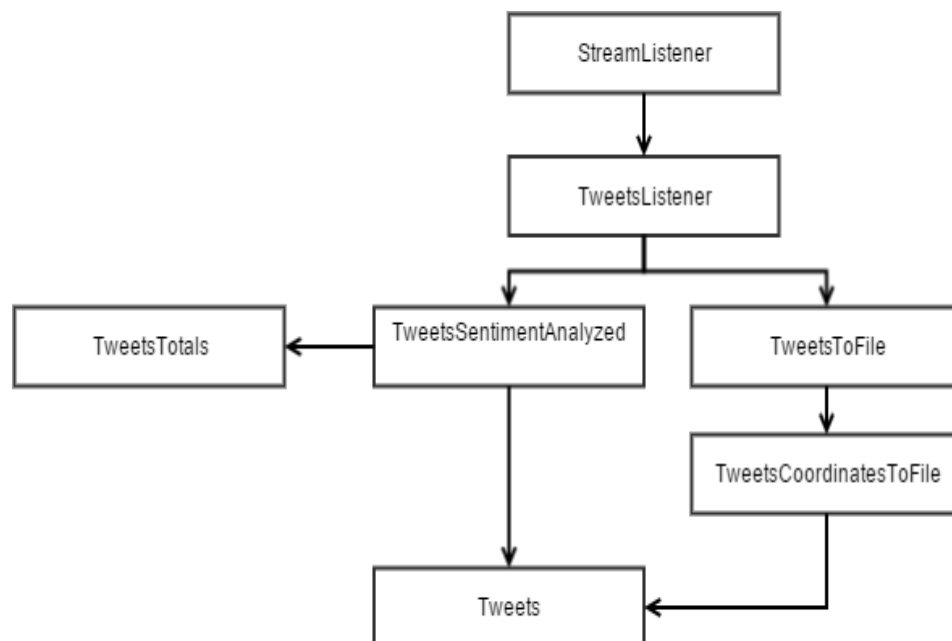
 **Application Domain:** The application domain will be any current topic. For example: It could be a football match or important treaty signed by the government.

 **Data Visualisation :** Generating a bar graph for the comparison between the various sentimental attitudes.

6. WORKING

- ☐ **TweetsListener.py** provides classes to perform tweets collection and sentiment analysis
- ☐ There are several classes based on **StreamListener** from **tweepy**

- ❑ **TweetsListener** is a base class to listen tweets only and stop listening when a specific number of tweets collected or time listening is out. If both parameters are 0 then it can run forever.
- ❑ **TweetSentimentAnalyzed** analyzes the tweet emotions based on emoticons in the tweet if any or applies NLP procedures from **TextBlob** package to perform sentiment analysis
- ❑ **TweetsToFile** saves collected tweets in a csv file. In addition to the limit number of tweets or maximum time collection reached, the collection of tweets can be stopped when the file reaches a limit. The class can split the data in several files, based on the tweets number or an individual file size.
- ❑ In addition to the tweet itself, **TweetsCoordinatesToFile** saves GEO info in a separate file. "_geo_" is added to the file name The GEO file is in csv format, many-to-one relations to the tweets file All files total size is taken into account to stop tweets collection at file_size_limit Only tweets file is splitted in multiply files
- ❑ **Tweets** class inherits **TweetsCoordinatesToFile** and **TweetSentimentAnalyzed** methods. It saves the tweet, it's sentiments in a csv file and the coordinates in a separate file
- ❑ **TweetsTotals** saves only total sentiments daily in a csv file. There is also hourly data but they are overwritten



7. CODES

TWITTERLISTENER.PY

This module provides classes to perform tweets collection and sentiment analysis.

FUNCTIONALITIES IN THE CODE

```
class TweetsListener(StreamListener):
    """
        This is a base class to listen tweets only
        and stop listening when a specific number of tweets collected
        or time listening is out
    """

    def on_data(self, data):
        # The presence of 'in_reply_to_status' indicates a "normal" tweet.
        # The presence of 'delete' indicates a tweet that was deleted after posting.

    def on_status(self, status):
        """
            Collects tweet in an inner data structure,
            checks if there is a time stop tweet collection

    def on_tweets_limit(self):
        """
            When the total tweets number reached the limit

    def on_time_limit(self):
        """
            When the time to collect tweets is out
        """

    def get_tweet_data(self, status):
        """
            Reads the tweet from status XML and
            adds some info in tweet{} dictionary
```

```

def save_tweet(self,tweet):
    """
        Should be used to save the tweet from the tweet dictionary
        It just prints dot in the base class

class TweetSentimentAnalyzed(TweetsListener):
    """
        The class analyzes the tweet emotion
        based on emoticons in the tweet if any
        or or applies NLP procedures from textblob package
        to perform sentiment analysis

def sentiment_analysis(self,tweet):
    """
        Runs the sentiments analysis based on emoticons in the tweet
        If there are no emoticons or the sentiment is NEUTRAL or CONFUSED
        runs a sentiment analysis from TextBlob package

def get_tweet_data(self,status):
    """
        The same as in the base class plus
        sentiment analysis of the tweet

def sentiment_analysis_by_emoticons(self,tweet):
    """
        Based on the emoticons set the sentiment for the tweet
    """

def sentiment_analysis_by_text(self,tweet):
    """
        Set the sentiment of the tweet based on the polarity
        from TextBlob sentiment analysis

class TweetsToFile(TweetsListener):
    """
        The class saves collected tweets in a csv file

def on_status(self, status):
    """
        saves the tweet and
        compares collected files size to the file_size_limit

```



```

def on_file_size_limit(self):
    """
        When the total File Size limit was reached
    """

def on_save_tweet(self):
    """
        When the tweet is saved
        It opens a new file if limits are reached

def save_tweet(self,tweet):
    """
        Saves the tweet in a csv file

class TweetsCoordinatesToFile(TweetsToFile):
    """
        In addition to the tweet itself, the class saves GEO info
        in a separate file. "_geo_" is added to the file name
        The GEO file is csv, many-to-one relations to the tweets file
        All files total size is taken into account to stop tweets collection at file_size_limit
        Only tweets file are splitted in multiply files

        def save_coordinates(self,tweet):
            """
                Saves coordinates in a file
            """

class Tweets(TweetsCoordinatesToFile,TweetSentimentAnalyzed):
    """
        The class saves the tweet, it's sentiments in a csv file and
        the coordinates in a separate file

class TweetsTotals(TweetSentimentAnalyzed):
    """
        The class saves only total sentiments daily in a csv file
        There is also hourly data but they are overwritten
    """

```

8. OUTPUT

```
saransh@saransh-pc: ~/TweetsListener
saransh@saransh-pc:~/TweetsListener$ python TweetsListener.py "IPHONE X" Project
Resources.yml
Collecting tweets regarding IPHONE X Click Ctrl+C to stop the programm
█
```

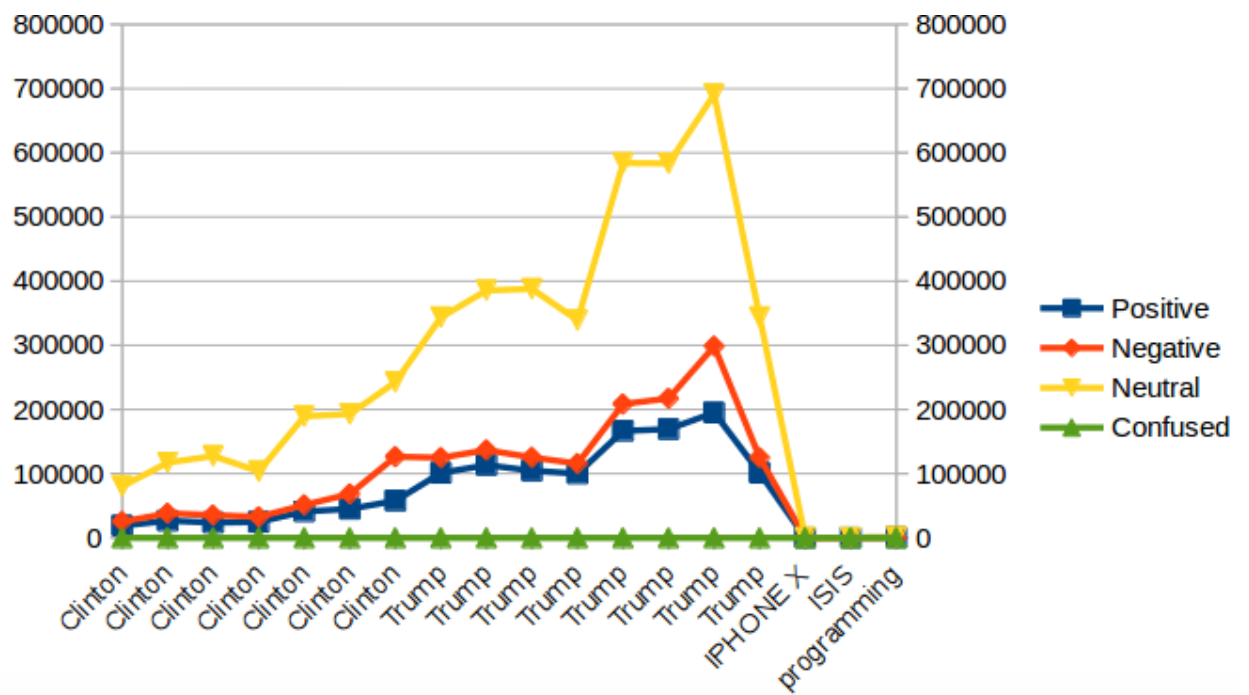
	Standard	Standard	Standard	Standard	Standard	Standard
1	SearchTerm	EndDateTime	Positive	Negative	Neutral	Confused
2	Clinton	28-05-2016 00:00	18707	25053	80258	0
3	Clinton	29-05-2016 00:00	26527	38092	117176	0
4	Clinton	30-05-2016 00:00	23683	35247	127383	0
5	Clinton	31-05-2016 00:00	24771	32553	103485	0
6	Clinton	01-06-2016 00:00	40881	51066	189286	0
7	Clinton	02-06-2016 00:00	44848	68257	192782	0
8	Clinton	03-06-2016 00:00	57715	126342	242797	0

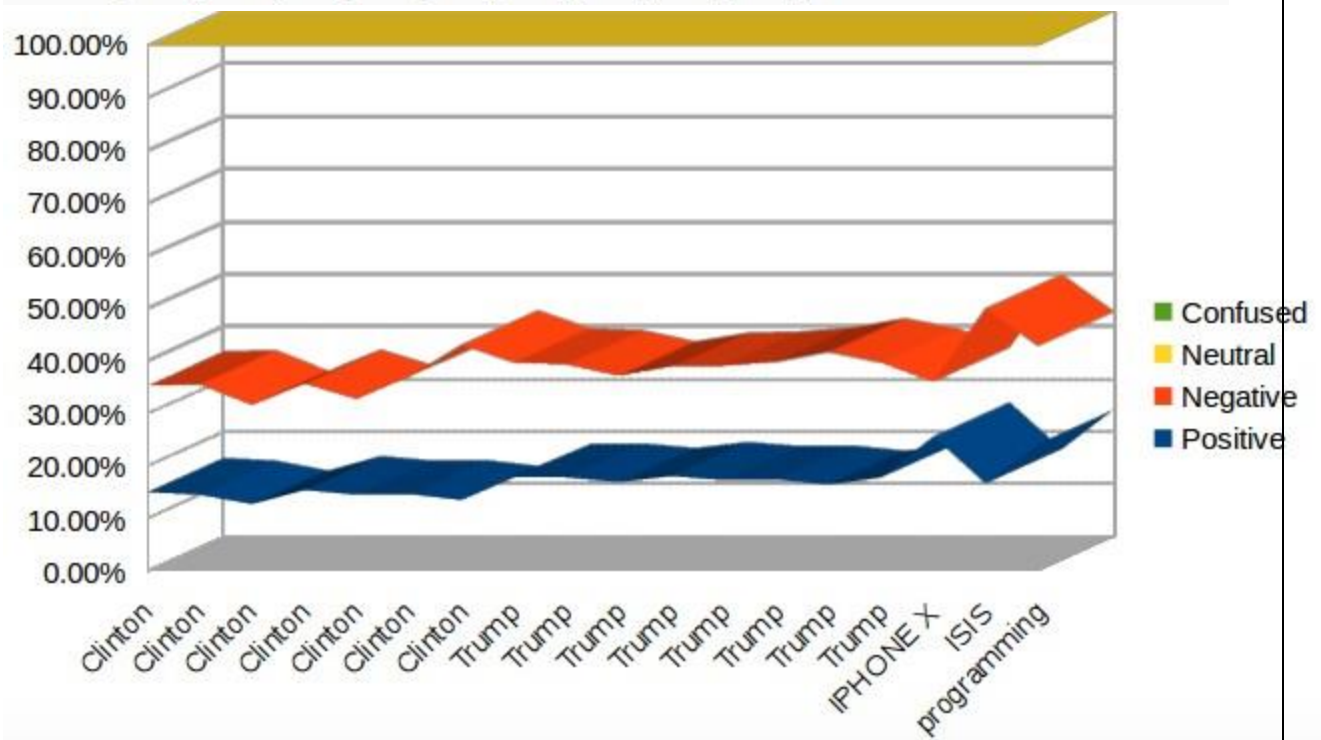
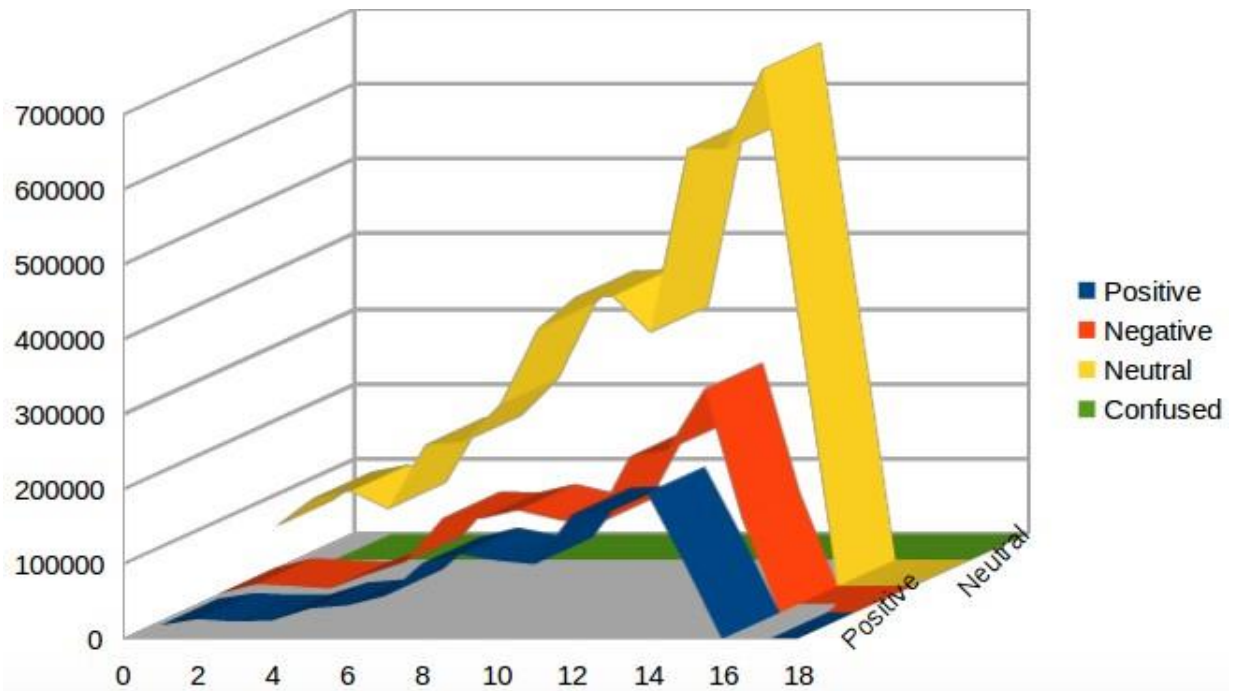
	Standard	Standard	Standard	Standard	Standard	Standard
1	SearchTerm	EndDateTime	Positive	Negative	Neutral	Confused
2	Trump	28-05-2016 00:00	101201	124845	343689	0
3	Trump	29-05-2016 00:00	113340	136688	385403	0
4	Trump	30-05-2016 00:00	104475	125239	387828	0
5	Trump	31-05-2016 00:00	100083	115670	338967	0
6	Trump	01-06-2016 00:00	166539	208409	584058	0
7	Trump	02-06-2016 00:00	169233	217233	583357	0
8	Trump	03-06-2016 00:00	195010	298512	691062	0

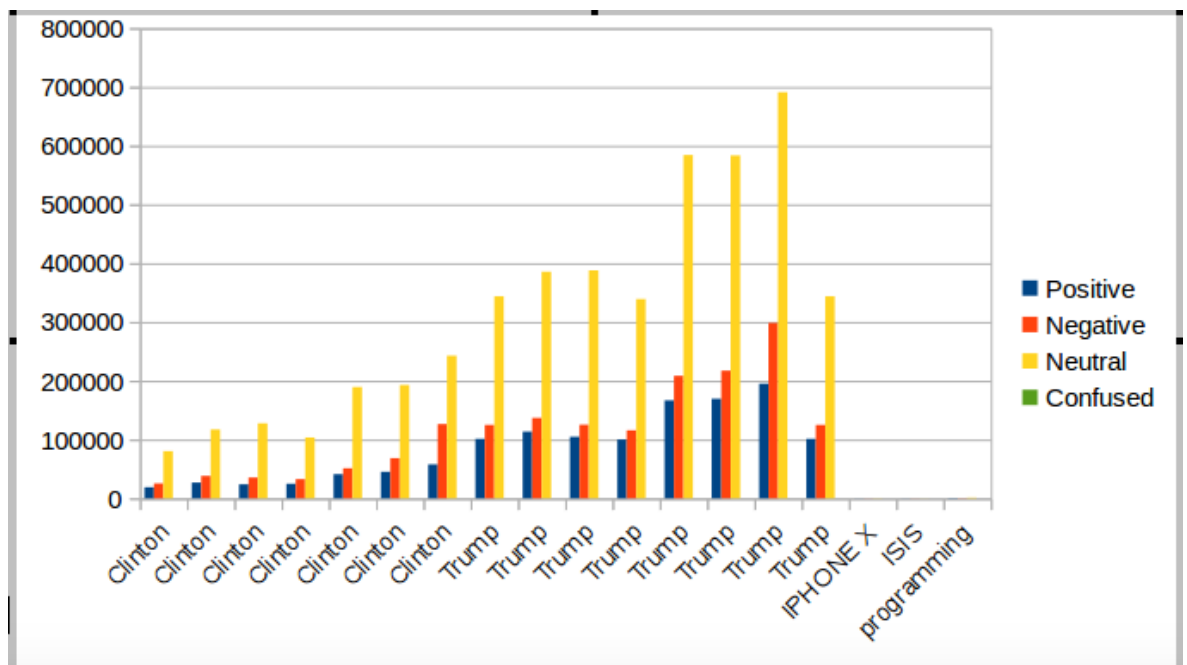
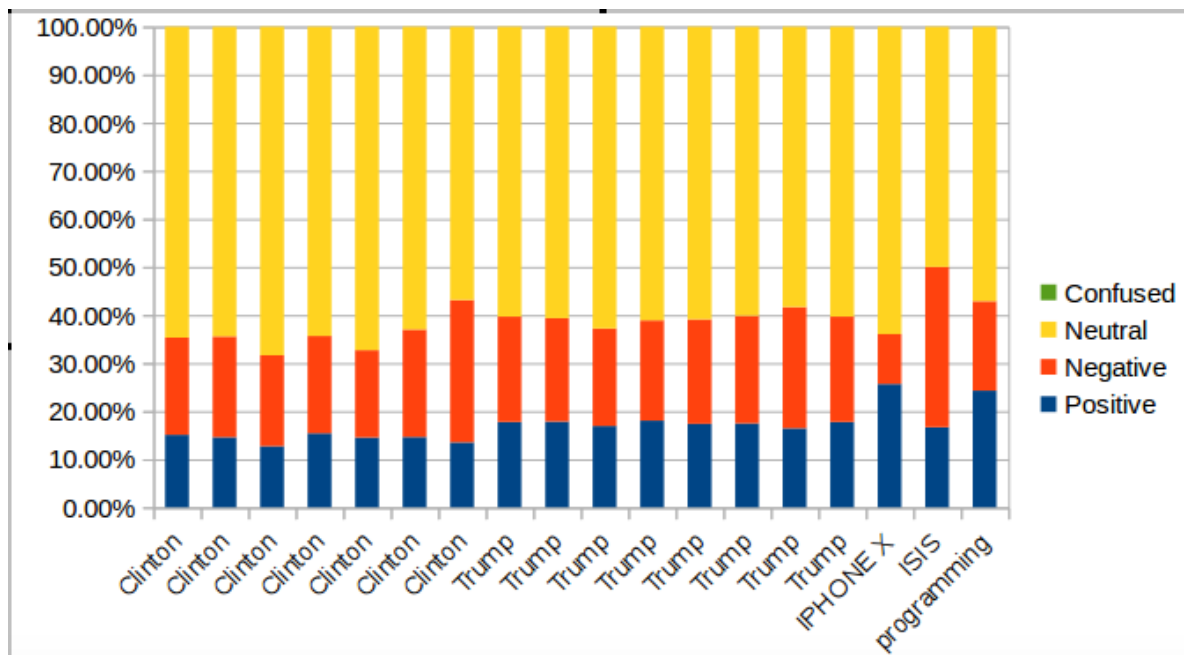
	Standard	Standard	Standard	Standard	Standard
1	SearchTerm	Positive	Negative	Neutral	Confused
2	IPHONE X	20	5	58	0

	Standard	Standard	Standard	Standard	Standard
1	SearchTerm	Positive	Negative	Neutral	Confused
2	ISIS	1	2	3	0

	Standard	Standard	Standard	Standard	Standard
1	SearchTerm	Positive	Negative	Neutral	Confused
2	programming	390	299	918	0







9. RESULT

By running sentimental analysis on Hilary Clinton, Donald triumph, isis and iphone x, we obtained the following inferences:-

☐ HILLARY CLINTON:

We observe that neutral sentiments are more prominent in this regard followed by negative and positive sentiments

.

☐ DONALD TRUMPH

We observe that neutral sentiments are more prominent in this regard followed by negative and positive sentiments.

☐ IPHONE X

We observe that neutral sentiments are more prominent in this regard followed by positive and negative sentiments.

☐ ISIS

We observe that neutral sentiments are more prominent in this regard followed by negative and positive sentiments.

10. REFERENCES

 <http://www.ijecs.in/issue/v4-i4/39%20ijecs.pdf>

 <http://ieeexplore.ieee.org/document/7975659/>

 https://link.springer.com/chapter/10.1007/978-3-319-11119-3_12