Sentiment Analysis of Swachh Bharat Abhiyan

Sahil Raj*, Tanveer Kajla**

Abstract

The present paper is about the social media analytics. It is a new tool to analyse the behaviour of the users who use social networking sites and other social sites like blogs, forums etc. Every organisation uses this tool to analyse their customers. Even the government agencies are using these analytical tools to get the feedback of their newly launched missions and their policies. In this paper the sentiment analysis of Swachh Bharat Abhiyan is done with the help of tweets extracted from twitter. Tweets regarding Swachh Bharat Abhiyan are extracted with the help of an open source software R-studio. The geo-locations of tweets are also extracted in the software and the results are plotted on the map of India. The pattern of tweets is analysed and the popularity of the mission is evaluated. The word cloud of the popular and the most used words is also formed in the R-studio software. With the overall analysis, the popularity of the mission is perceived according the regions on the map of India, and the strategies can be applied to popularize the campaign in the lesser known regions of India.

Keywords: Swachh Bharat, Word Cloud, Geo-Location, Campaign

Introduction

Swachh Bharat Abhiyan is a special campaign by the BJP government to clean the roads, streets and infrastructure of the country. It is the visionary mission launched by our honourable Prime Minister Shri Narendra Modi. It was launched on 2nd October, 2014. This campaign is one of the India's biggest campaign, covering around 3 million government employees. This mission is widely popular among the citizens of the country as it directly gives them the responsibility to clean up their own country. The cleanliness campaign is also covering the schools, colleges and universities. The Prime Minister also nominated the nine big personalities of the country and also gave them the responsibility to nominate nine more people to join the campaign, making a chain to increase the participants of the mission. The aim of the campaign is to achieve the vision of cleanIndia by the year 2019. The main objectives of the campaign are to finish up the manual scavenging and to eliminate the open defecation which is the main cause of the tuberculosis in India. The construction of individual, community and cluster toilets were also included. The villages should be cleaned and to lay water pipelines in the villages to ensure 24 hour water supply to all the households by 2019.

Review of Literature

Social media produces massive amount of data (Ediger *et al.*, 2010).Twitter is a micro blogging service where users create messages called tweets. These tweets sometimes express opinions about different topics (Go, Huang & Bhayani, 2009). Social media is used as the official media platform by the celebrities, politicians but the research can be centred around the events also (Tumasjan *et al.*, 2010). Millions and trillions of users share their opinions on social media sites. Athletes use tweets to interact with their fans. Twitter feeds can also be used for emergency events like natural disaster and crisis management during or after the time of disaster(Zielinski *et al.*, 2012). The research can be done in English or any other language, where relevant tweets can be classified and extracted. Sentimental classifier is able to determine the neutral, negative and positive tweets. Algorithm can be made accurately to classify Twitter messages as positive or

* Assistant professor, School of Management Studies, Punjabi University, Patiala, Punjab, India.
E-mail: er_sahil@yahoo.com
** Research Scholar,School of Management Studies, Punjabi University, Patiala, Punjab, India.
E-mail: er.tanveer47@gmail.com

negative, with respect to a query term (Go *et al.*, 2009). Searching the tweets can be more easily done by using the hashtags ahead of the subject to be searched. Hashtags are used to categorise the messages in the twitter. With the use of the hashtags, twitter users can propagate the ideas and promote specific topics and people. Hashtags are used to streamline the search and at the same time, to increase the effectiveness of the research (Wang *et al.*, 2011). But the biggest problem is the size of an unstructured data, which is in chunks, and there can be lot of repetition in the data, so this unstructured data should be taken in large volumes and therefore more complex algorithms are used to classify very high number of tweets. Linguistic features can also be used to identify the language used in the tweets. The research is also done on how many retweets came and what are the factors contributing to the retweets (Naveed *et al.*, 2011). The researchers brought much attention to the data from the tweets as the data is very irregular due the 140 character limits put on the tweets (Saif, He & Alani, 2012). The authors show how to automatically collect a corpus for sentiment analysis and opinion mining purposes. Using the corpus, the authors build a sentiment classifier that is able to determine positive, negative and neutral sentiments for a document (Pak & Paroubek, 2010).

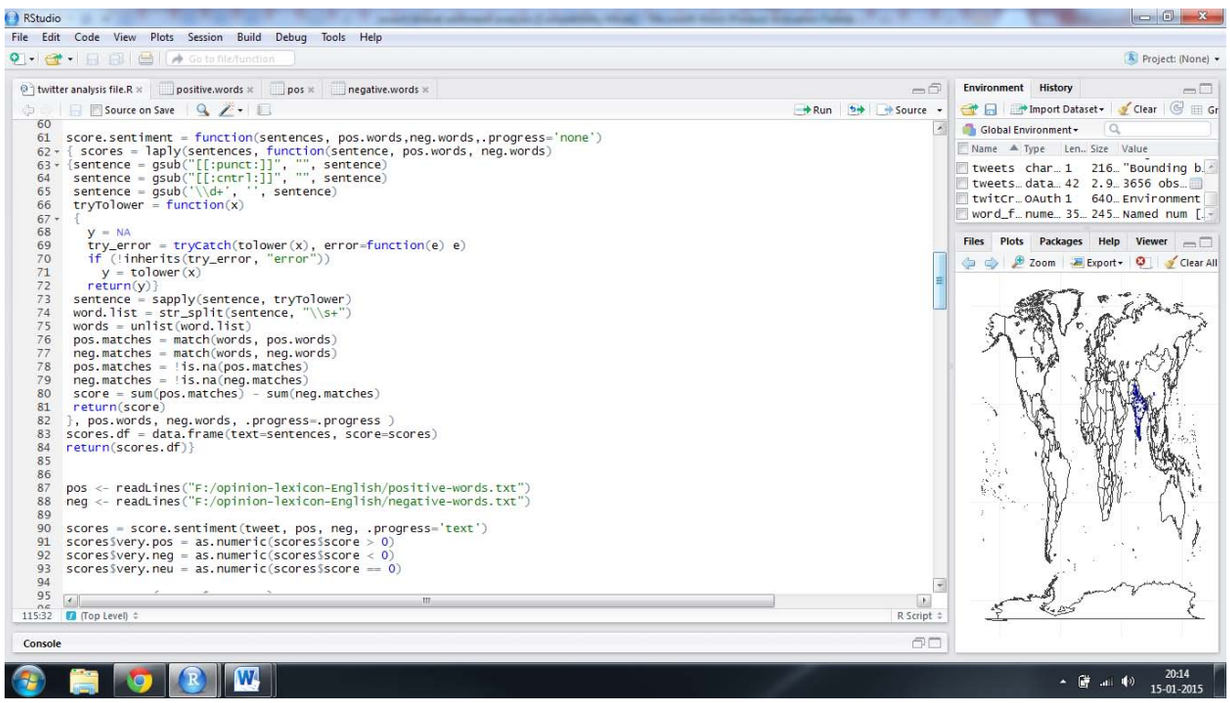
Objectives

The objective of the research paper is to do the sentimental analysis on the cleanliness campaign launched by the present government. The analysis will give the perception of the citizens regarding this new campaign. The second objective of the paper is to find the location of the tweets regarding the Swachh Bharat, and then plot them on the map of India to check the popularity of the mission. The third objective of the paper is to make a word cloud of the tweets which will prompt the most used words in the tweets.

Research Methodology

The data related to the research are the unstructured data. The unstructured data are extracted from twitter in the form of tweets. For extracting the data from the twitter we used open-source R-Studio software which is based on windows platform. First of all the developer account has to be created in the twitter, and then the application to mine the text is created in the application account of the twitter. In this application, keys and tokens were generated. With these keys and tokens, authorisation is provided by twitter to extract the tweets. In R-studio software, these tweets were extracted with the help of

Figure 1: Function for Sentiment Analysis



some text mining packages. When the extraction of tweets is done, the analysis is done on the set of tweets. In the analysis part, we are categorising the tweets into the positive tweets, negative tweets and neutral tweets. It is done with the assistance of text files having the list of positive and negative words. With these lists tweets are compared and the positive and negative ratings are given to the tweets. The rated tweets are plotted on the pie chart with the help of which analysis is done.

Figure 1: shows the function for sentiment analysis. With this function the tweets are rated to positive, negative and neutral.

The word cloud of the tweets is the pictorial representation of the words in the software. It will highlight the most used or talked about words in the middle of the cloud. With this cloud some prominent parts of the mission can be detected. The cloud can also prompt us about the important subjects of the mission. The word cloud is created with the help of the word cloud package and then the word cloud function is created to display the word cloud.

The Highlighted Portion of the Screenshot in Figure 2 Shows the Word Cloud Function

The geolocation of tweets is done with the assistance of ggplot package and maps package.The tweets are extracted according to the location. According to the research, tweets should be extracted from the Indian region. So the longitude and latitude of the area is given from the south-west corner to the north-east corner in the command interface and the tweets will be extracted from this bounded region. Maps package will print the map in the software and ggplot package will plot the extracted tweets from the twitter on the map. The geo-located tweets will give the pattern on the map, and will give the region from where more number of tweets are coming.

The tweets which are fetched according to the Indian region will be saved in bharat2.json file. Then these tweets be cleaned for the relevant text, and then plotted on the world map.

Analysis And Discussion

When the tweets are extracted from twitter on swachhbharat, then the sentiment analysis is done on the tweets. Analysis is done through pie chart, as shown in Figure 4. Pie chart clearly explains that twitter users are having a liking towards the swachhbharat mission as there are very less negative responses towards the campaign.

Figure 2: Word Cloud Function

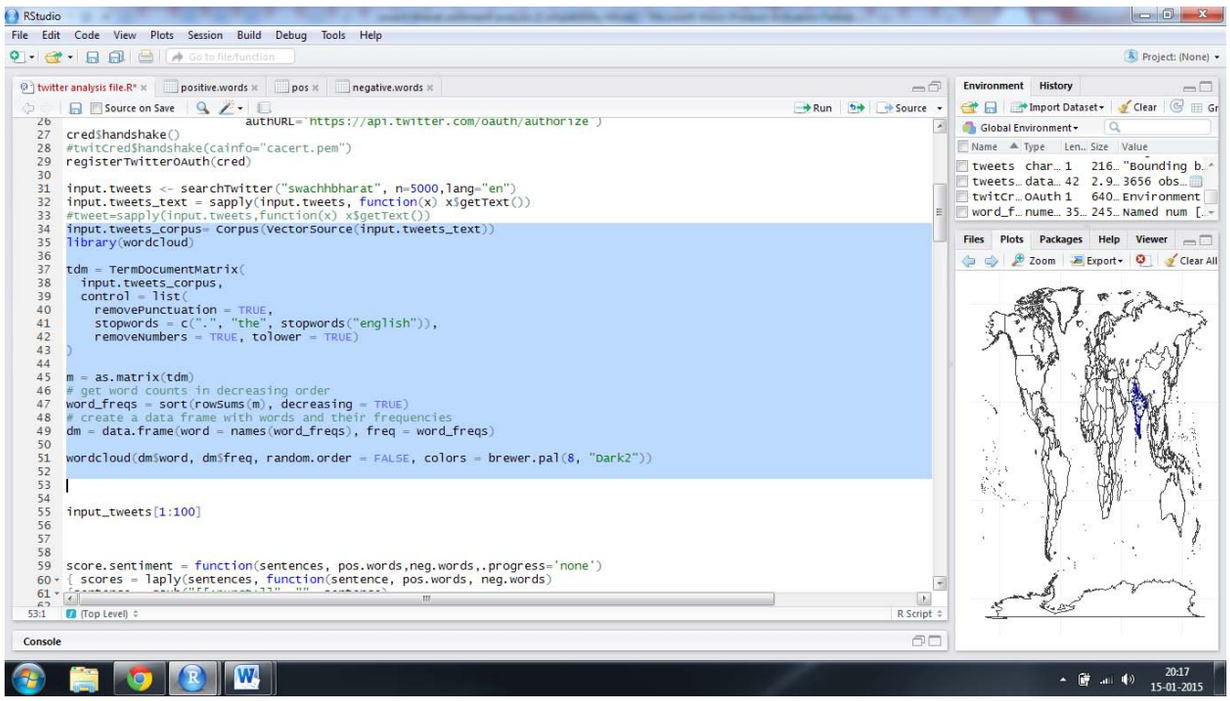


Figure 3: Steps to Save Tweets in bharat2.json file

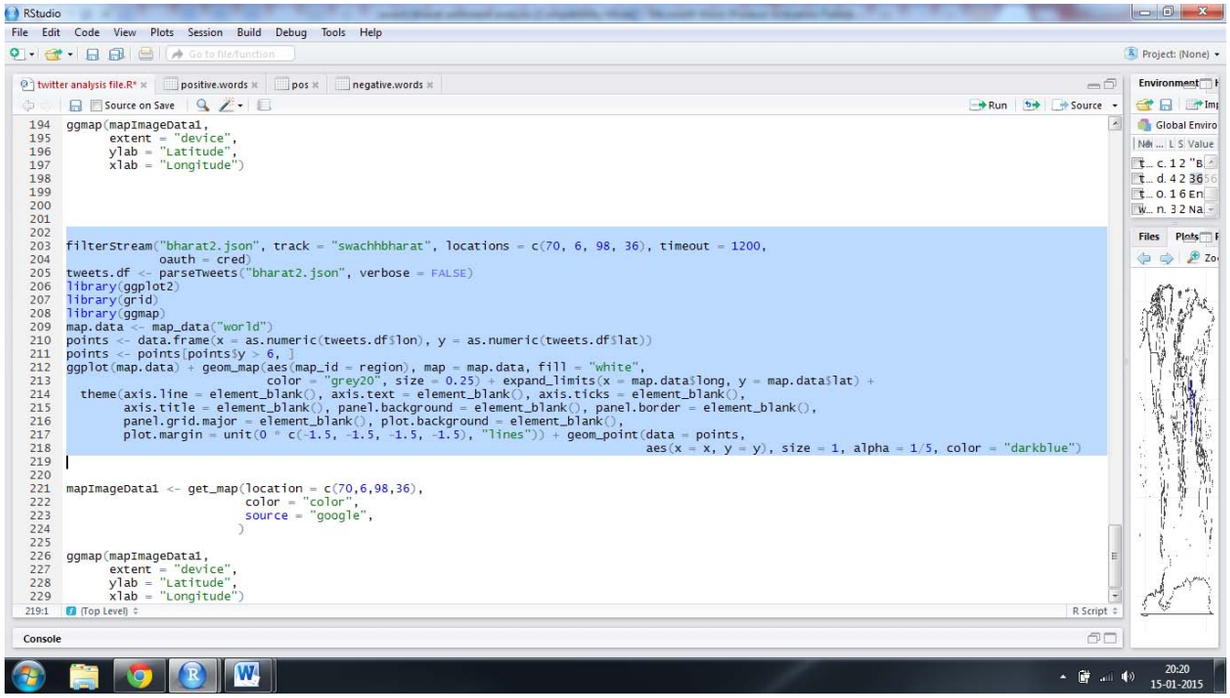


Figure 4: Analysis Through Pie Chart

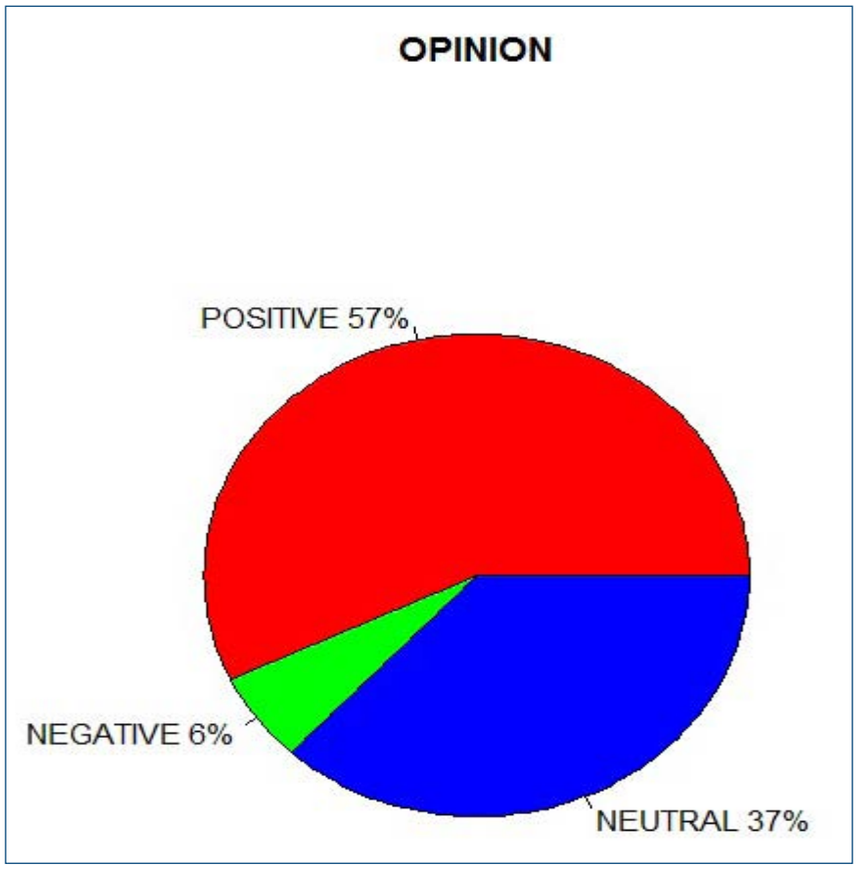
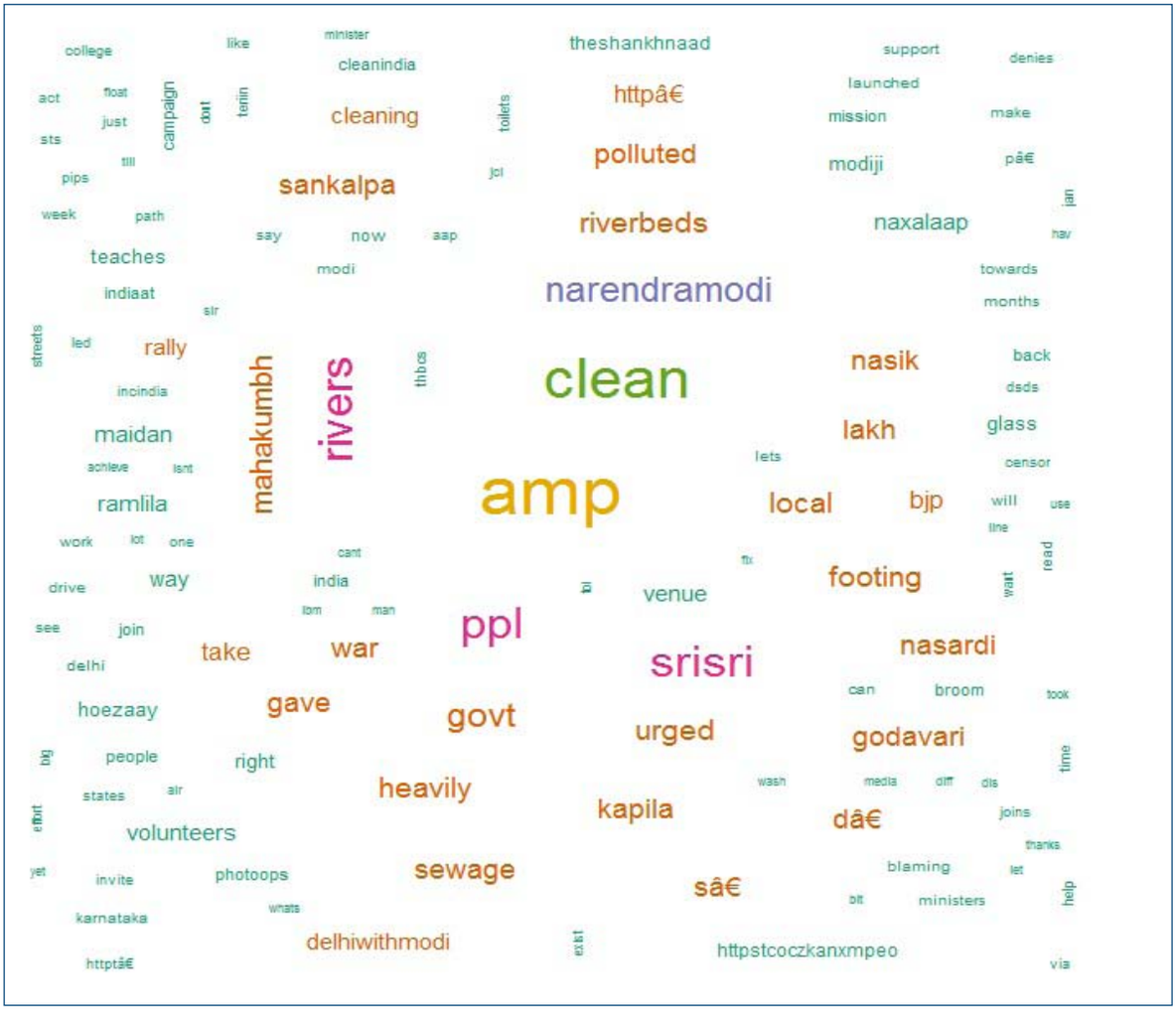


Figure 5: Word Cloud of the Tweets

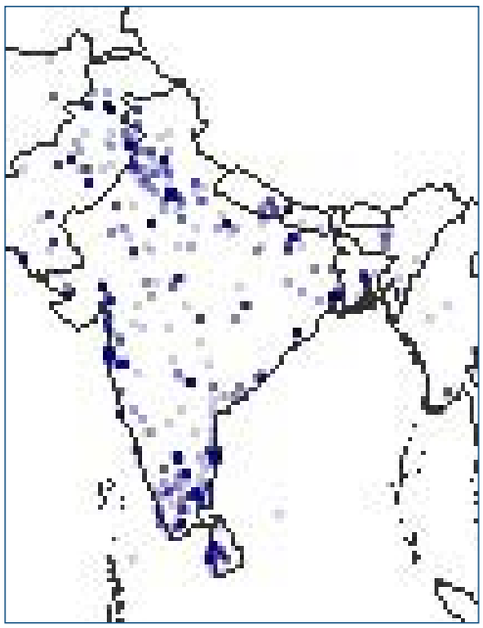


There is more number of neutral and positive responses on this campaign. This clearly explains that the people of India are adapting to this kind of campaign.

Word cloud of the tweets is also made in the software. In the word cloud it is clearly seen that the Narendra Modi is a heavily used word, as it is seen in the middle of the cloud. This shows that he is an integral part of the mission. It is a well-known fact that he is the creative mind behind this whole campaign. The word "clean" is the most highlighted word in the word cloud. This indicates that the twitter users, who are sending tweets on the twitter about the Swachh Bharat were more concerned about cleanliness, which is the main objective of this campaign.

Tweets are extracted from the Indian region to see the popularity of the region, from where the people are more concerned about the Swachh Bharat. When the tweets are plotted on the map of India then the tweets are easily distinguished on the basis of their recurrence. More number of tweets are coming from northern and southern region of India as it is clearly seen on the map. That means the campaign is more popular in these regions. The campaign has started from New Delhi, and that is the only reason it is very popular among them. In the southern region, the campaign is popular in the state of Tamil Nadu. The campaign is not so popular in the Gujarat region and in other states of India. Many tweets have been received from the neighbouring countries also. This shows that campaign got worldwide popularity and the new government is connecting with the neighbouring countries in a positive way.

Figure 6: Origin of Tweets



Conclusion

Twitter is the emerging social media tool, and every organisation is using this tool for marketing of their product. Internet users from different background and countries come under a single medium to share their views and opinions about any new happenings, and can also recommend to other users or some experts, who can even give their advice to internet users, who are new to the world of gadgets. The sentiment analysis clearly shows that this campaign is a success among the people of India. People have given a very positive response on this initiative by the Indian Prime Minister.

In this analysis many loopholes are also found for the campaign. Though the campaign is popular and is very much appreciated, but still it is not popular in the central region of India. Very less number of tweets has been received from these parts of India. Northern part of the India is also less involved in this campaign. The main aim of this campaign was cleanliness and the tweets show that 'clean' is a highly used word in the tweets as shown in the word cloud.

Recommendations

The government should not focus on the urban areas, as this campaign is popular in urban areas only. The campaign should be encouraged in the central parts of India also. The campaign has been accepted by the people, so this initiative can be easily extended to the lesser known regions also.

References

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011, June). Sentiment analysis of twitter data. In *Proceedings of the Workshop on Languages in Social Media* (pp. 30-38). Association for Computational Linguistics.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market.*Journal of Computational Science*, 2(1), 1-8.

Bruns, A., & Stieglitz, S. (2013). Towards more systematic Twitter analysis: Metrics for tweeting activities. *International Journal of Social Research Methodology*, 16(2), 91-108.

Ebner, M., Altmann, T., & Softic, S. (2011). @ twitter analysis of# edmedia10–is the# informationstream usable for the# mass. *Form@ re-Open Journal per la formazione in rete*, 11(74), 36-45.

Ediger, D., Jiang, K., Riedy, J., Bader, D. A., Corley, C., Farber, R., & Reynolds, W. N. (2010, September). Massive social network analysis: Mining twitter for social good. In *Parallel Processing (ICPP), 2010 39th International Conference on* (pp. 583-593). IEEE.

Go, A., Bhayani, R., & Huang, L. (2009a). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1-12.

Go, A., Huang, L.,&Bhayani, R. (2009b). Twitter sentiment analysis. *Entropy*, 17.

Hao, M., Rohrdantz, C., Janetzko, H., Dayal, U., Keim, D. A., Haug, L., & Hsu, M. C. (2011, October). Visual sentiment analysis on twitter data streams. In Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on (pp. 277-278). IEEE.

Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the omg!. *ICWSM*, 11, 538-541.

Li, R., Lei, K. H., Khadiwala, R., & Chang, K. C. (2012, April). Tedas: A twitter-based event detection and analysis system. In *Data Engineering (ICDE), 2012 IEEE 28th International Conference on* (pp. 1273-1276). IEEE.

Mathioudakis, M., & Koudas, N. (2010, June).Twitter monitor: Trend detection over the twitter stream. In *Proceedings of the 2010 ACM SIGMOD*

International Conference on Management of data (pp. 1155-1158).ACM.

Naveed, N., Gottron, T., Kunegis, J., &Alhadi, A. C. (2011, June). Bad news travel fast: A content-based analysis of interestingness on twitter. In*Proceedings of the 3rd International Web Science Conference* (p. 8).ACM.

Pak, A., &Paroubek, P. (2010, May).Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *LREC*.

Saif, H., He, Y., &Alani, H. (2012).Alleviating data sparsity for twitter sentiment analysis. CEUR Workshop Proceedings (CEUR-WS. org).

Thomas, K., Grier, C., Song, D., & Paxson, V. (2011, November). Suspended accounts in retrospect: An analysis of twitter spam. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference* (pp. 243-258).ACM.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. *ICWSM*, 10, 178-185.

Wang, X., Wei, F., Liu, X., Zhou, M., & Zhang, M. (2011, October). Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In *Proceedings of the 20th ACM international conference on Information and knowledge management* (pp. 1031-1040). ACM.

Yardi, S., & Boyd, D. (2010). Dynamic debates: an analysis of group polarization over time on Twitter. *Bulletin of Science, Technology & Society*, 30(5), 316-327.

Zielinski, A., Bügel, U., Middleton, L., Middleton, S. E., Tokarchuk, L., Watson, K., & Chaves, F. (2012, April). Multilingual analysis of twitter news in support of mass emergency events. In *EGU General Assembly Conference Abstracts* (Vol. 14, p. 8085).