

extraction of the real time tweets to be collected from the twitter website, news websites on the windows based operating system and executed the program in JAVA for the same. Also we have used the JSOUP libraries to obtain the tweet from the news websites. To retrieve live tweet from twitter streaming API is used and for retrieving old tweet from twitter GET search API method is used.

3. RESULTS AND DISCUSSIONS

The tweets were obtained on Friday Nov 21 11:16:25 IST 2014 of the keyword search Narendra Modi using twitter API's, which contain so many things such as URL's, slangs, hash, abbreviations, user ID's, stopwords which is not required for the further analysis of the tweet. The sample of the data is given in Table 1. Thus recent and old tweets have been extracted on the particular on the social topic is discussed in news media and news papers and tweeter website within a day or other day such as Narendra Modi on which peoples have given their opinion and related events happened. The cleaning of the tweets has been done by different methods. Initially we have done the preprocessing of the raw tweets, only useful sentences are sorted out from the garbage. To get the meaning of each word we used in three English dictionaries WordNet. Also we identified presence of URL using a regular expression and remove all the URLs from the tweet. Remove all the private usernames identified by @user and the hashtags identified by the # symbol. Thus we have tried to normalize complete tweets by preprocess raw tweets Further pre-processed the tweets data will be used for further complete processing

4. CONCLUSION

Thus we have tried to extract the social network related tweets from the twitter web site using the twitter 4j Libraries and their API's. Further these tweets were preprocessed to for the removal of Hash, URL and user ID.

REFERENCES

- [1] Bifet, E. Frank, Sentiment Knowledge Discovery in Twitter Streaming Data,, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 1-15.
- [2] Bifet, G. Holmes, B. Pfahringer, MOA-Tweet Reader: real-time analysis in twitter, streaming data, in: T. Elomaa, J. Hollm'en, H. Mannila (Eds.), DS 2011, LNCS 69AUTHORS
- [3] 26, Springer-Verlag, Berlin Heidelberg, 2011, pp. 46-60.
- [4] Lambodar Jena, Narendra Kumar Kamila, Data Extraction and Web page Categorization using Text Mining, IJAIEEM, ISSN 2319 - 4847, Volume 2, Issue 6, June 2013
- [5] Farhan Hassan Khan et. al. , TOM: Twitter opinion mining framework using hybrid classification scheme, Decision Support Systems, 2013, <http://dx.doi.org/>
- [6] 10.1016/j.dss.2013.09.004
- [7] <https://api.twitter.com/1.1/geo/search.json>
- [8] <https://dev.twitter.com/>
- [9] <http://search.twitter.com/search.format>

AUTHORS



Bapurao M Bandgar received his MCA Degree in 2010, M.Sc. (Physics) degree in 1999 and B.Sc. (physics) in 1995 from Pune University. Perusing Ph.D(Comp. Sci.) from Karpagam University. His research interest includes Social Network Analysis and Data Mining.



Dr. Binod Kumar is Director & Professor at JSPM's Jayawant Technical Campus, affiliated to University of Pune , India . He has been in the field of teaching since more than 16 years. At present six Ph.D. candidates are registered under his supervision. He has published nearly 30 papers in International and National Journals. His areas of interest are Data Mining, Bioinformatics and Software Engineering