Abstract

Various posts on social media contain number of spam comments which are counterproductive. Such comments may include advertisements, abusing words or pointless arguments. Current paper presents the findings on such comments and provides the frequency of unique words after stop words filtration. Total 10 Facebook links were evaluated for spam words identification. These 10 links contained a total 13,100 comments among which 2,404 spam comments were detected. Total Spam words from Spam comments were 18,505 including Stop words and Duplicates while total unique spam words were 12,266. It has been found that the words ‘recharge’, ‘best’, ‘like’, ‘free’ and ‘god’ are the most probable spam words in the spam comments in profiles of celebrities on Facebook.

References

1. Link1, Available Online: https://www.facebook.com/AmitabhBachchan/photos/a.452056338161494.108384.4490828417
Identification and Analysis of Spam Words from Facebook Spam Comments

92177/1010186939015095/?type=1
2. Link2, Available Online:
https://www.facebook.com/JhoothiBewaFa/photos/a.648422241855030.1073741826.53438309925612/94697851532733/?type=1
3. Link3, Available Online:
https://www.facebook.com/CircleofCricket.MSDhoni/photos/a.1183268344883194.9394.118306468218564/861540440561826/?type=1
4. Link4, Available Online: https://www.facebook.com/ehsaas.sharma.904?fref=ufi
5. Link5, Available Online:
https://www.facebook.com/SainaNehwal.OGQ/photos/a.692139554160390.1073741829.10078783962234/990319264342733/?type=1
6. Link6, Available Online:
https://www.facebook.com/shreyaghoshal/photos/a.10152074044476484.1073741825.11541726483/10153194737396484/?type=1
7. Link7, Available Online:
https://www.facebook.com/Discovery/posts/10153273842238586?comment_tracking=%7B%22tn%22%3A%22O%22%22%7D
8. Link8, Available Online:
https://www.facebook.com/TimesofIndia/photos/a.10150135992497139.327198.26781952138/10153229934242139/?type=1&comment_tracking=%7B%22tn%22%3A%22O%22%7D
9. Link9, Available Online:
https://www.facebook.com/History/photos/a.480647736183.258349.8429246183/10152615184926184/?type=1&comment_tracking=%7B%22tn%22%3A%22O%22%7D
10. Link10, Available Online:
https://www.facebook.com/UNBfacts/photos/a.241405075971682.47251.241392719306251/750756091703242/?type=1
http://research.ijcaonline.org/volume114/number9/pxc3901957.pdf
13. Facebook, Available Online: https://www.facebook.com/
14. LinkedIn, Available Online: https://in.linkedin.com/
15. Reddit, Available Online: https://www.reddit.com/
17. Twitter, Available Online: https://twitter.com/
19. Tumblr, Available Online: https://www.tumblr.com/
20. Pinterest, Available Online: https://au.pinterest.com/
21. Del.icio.us, Available Online: http://del.icio.us/
22. YouTube, Available Online: https://www.youtube.com/?gl=IN

Index Terms

Computer Science | Information Sciences
Keywords

Bag of Words (BOW), Facebook, Spam, Tokenization, Vector Space Document Model (VSDM)