An Analysis on Top Commented Posts in Reddit Social Network about COVID-19

At the moment, people have been quarantined for COVID-19 management. In such a situation, people share their concerns, ideas, questions, etc., through social media more than before. Face-to-face relationships have been replaced with virtual relationships through the Internet. COVID-19 will not be controlled without people's collaboration with medical personnel, managers, and policy-makers (source: authors' observation). Besides working on the treatment of patients, researching to find the COVID-19 vaccine, etc. It is essential to consider people's concerns and try to address them for gaining more co-operation from their sides. In this regard, social media mining can help to understand what is on the point of discussion about COVID-19 by the public. By gaining a list of topics, we can plan better for dealing with COVID-19 as the public have a vital role in the prevention of virus dissemination.

For that purpose, we follow some sources and tutorials on social network mining and textual data analyzing.[1-4] Based on the mentioned sources, the methodology of our research is presented below. We selected Reddit as the social media for study. Based on Statista, Reddit is in the list of popular social networks globally and has about 430 million active users.[5] Reddit visitors are mainly from the USA, UK, Canada, Australia, and Germany.[6] Most of Reddit users are male and young.[7] To understand what people say about COVID-19, we searched the terms "COVID-19" and "corona virus" in the title of Reddit's posts (search date was March 31, 2020). Then, the top commented posts for each searched keywords were extracted. The total number of extracted posts was about 460. This process was carried out semi-automatically by the R tool and its packages. Then, to understand what people said about COVID-19, we utilized topic modeling and the Latent Dirichlet Allocation (LDA) algorithm. LDA identify topics in the textual data and present a list of keywords for each topic. In our research, LDA detected topics in the posts in which authors had left more comments on them. It is necessary to state that before using LDA, we should clean and prepare the extracted data to make it suitable for LDA analysis (preprocessing technique such as stop words removing, stemming, punctuation removing, etc.).

Using the mentioned method, we presented the keywords related to each topic in Figure 1 as the words cloud (WC). A total of 12 topics was identified in the data and shown as the WC.

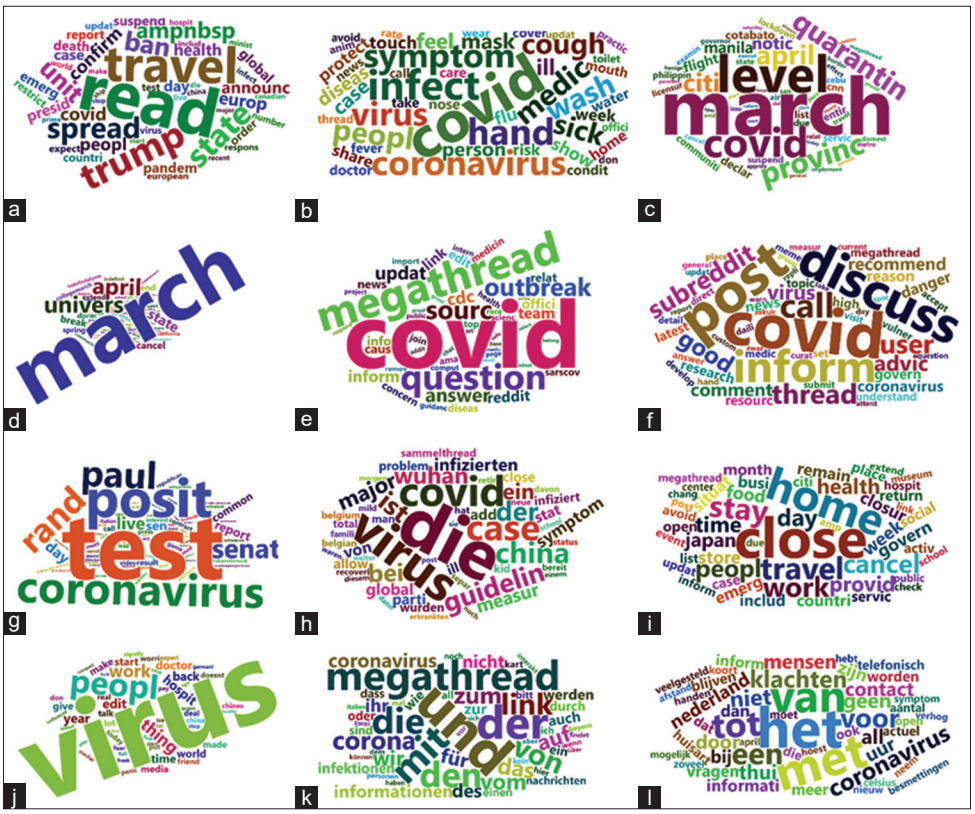

Figure 1: Words cloud of topics related to COVID-19 posts in Reddit

WC (a) reflects COVID-19 in the USA and the effect of this virus on this country. People discussed the spread of COVID-19 in the USA and made decisions to confront and limit the burden of this virus. WC (b) shows discussions about the symptoms of COVID-19, the ways this virus spreads, and its confrontation methods. This topic highlights attraction to the information needs of public people. People share and discuss the symptoms of COVID-19 and want to understand whether they are infected or not. In such a condition, people need to have access to valid and fruitful information. This will also reduce unnecessary referrals to hospitals. In some countries, the Ministry of Health has developed mobile/web-based applications to check people health statuses by answering some self-assessment questions about COVID-19. WC (c) emerged due to discussions about quarantine in different states, provinces, cities, etc. This topic indicates the number of detected COVID-19 cases after the quarantine and its effect on the prevention of COVID-19. By considering this topic, we need to provide suitable plans for leisure, support, education, and persuasion to comply with quarantine conditions for the public. WC (d) is about the cancellation of universities and colleges due to COVID-19. Some people share posts that show which universities and colleges suspended their activity and note the date of ending this suspension. These may be critical concerns for students, lecturers, academic staff, etc. They must meet some deadlines for research works or passing courses. There is a need to identify the academic population's concern as a subset of the public and address their concerns. WC (e) presents topics about questioning and answering in the social media (Reddit) or projects which encourage people to collaborate with sharing computing resources (sharing their computers' central processing units and random access memorys to conduct computations related to COVID-19 researches faster than ever). In this regard, it is necessary to provide a list of confirmed physicians in the social media for people to ask their questions. It is essential to detect spammers or fake accounts to prevent fake news. Policy-makers and health managers should introduce valid accounts and webpages to people for receiving valid information and news. WC (f) contains topics related to the management of information, conduction of discussions, removal of possible fake posts, etc., in Reddit. Bloggers and social media users should take care of the information they receive from social media and check the source of information. In the case of fake information, they should report posts since some users new to social media may interpret any information as valid. WC (g) is about positive COVID-19 tests of famous people such as senators, players, gamers, etc. It also contains information about the cancellation of some meetings, as well as self-quarantine. People follow news related to celebrities or famous people in the COVID-19 days! WC (h) emerged due to discussions about the reported cases of COVID-19 and discussions about this virus, mainly in China. The public is interested in such statistics as they hope to come back to regular living days. Hopefully, there are online websites such as Worldometer (https://www.worldometers.info/coronavirus/), which reports COVID-19 cases for each country. WC (i) reflects that stay at home to be safe from COVID-19. This topic contains posts that discuss the cancelation of travel, works, etc., to stay at home for preventing the spread of COVID-19. However, they should be intensives and laws so that people respect self-quarantine. Based on the WC (j), the risk of COVID-19 is another discussed topic. Some believe that COVID-19 is more dangerous than influenza, and some discuss the economic devastation of COVID-19. The public has concerns about what will happen after COVID-19 harness. These concerns should be addressed so that people collaborate with harnessing COVID-19. WC (j and k) appeared due to the posts about COVID-19 in Dutch.

Jelodar *et al*. used Reddit and analyzed available comments in subreddits. Our findings are somewhat similar to their results. The difference is the source of data, we analyzed posts, but they focused on users' comments. We focused on top commented posts, but they considered all available comments in selected subreddits. Hence, our research finding can be considered as the subset of their finding, because we focused on the most discussed content, not all them.[8] In the research by Stokes *et al*., they analyzed all available posts between March 3 and March 31, (94,467 posts) from r/Coronavirus threads. They also share similar results.[9] There are other research on the COVID-19 comments and posts analysis in Reddit or other social media such as tweeter.[10-14] An interesting point is here, many researchers used social media analyzing technique to understand COVID-19 in the perception of people. The difference between these works is in the number of posts/comments/tweets, search date, data selection strategy, or/and method. Ordun *et al*. presented a list of 14 papers that analyzed tweets to understand COVID-19.[10] The usage of social media is popular to understand the perception of people.

By considering H1N1, the recent pandemic in the world, there are similarities and differences between COVID-19 and H1N1 in public concerns. By reviewing available researches,[15-17] we understand that during the H1N1 pandemic, the public considered doing some activities (such as washing hands) or refraining some others (avoiding overseas travel, avoiding crowded places, not eating pork, etc.) to cope with H1N1. They also have concerns about H1N1 infections and symptoms. This is somewhat similar to what people shared in their posts in WC (a), (b), and (i) for COVID-19. However, it seems that the severity of actions for coping with COVID-19 is more intense than that for H1N1. Furthermore, the economic devastation of COVID-19 is more critical. The total infected countries are also more than H1N1.

Conclusion

This piece highlights the discussed topics in the top commented posts in Reddit by users. Based on the findings, it is essential to understand the public concerns and addressed them. Different researchers used Reddit or other available social media to provide analysis about what people shared. These studies used different strategy to find related content. We recommend policy-makers to use all them to gain more comprehensive picture about phenomenon.

The presented information is based on the text mining of Reddit's posts (as the source of data) besides authors' interpretations and recommendations for addressing related concerns. As mining textual data is a somewhat challenging task, there may be some tolerances in the result.

Used packages and tool:

- R tool[18]
- RedditExtractoR[19]
- dplyr[20]
- ldatuning[21]
- topicmodels[22]
- wordcloud2[23]
- tm.[24,25]

Khalil Kimiafar[1], Mehdi Dadkhah[2], Masoumeh Sarbaz[1], Mohammad Mehraeen[2]

[1]Department of Medical Records and Health Information Technology, School of Paramedical Sciences, Mashhad University of Medical Sciences, [2]Department of Management, Faculty of Economics and Administrative Sciences, Ferdowsi University of Mashhad, Mashhad, Iran

Address for correspondence: *Dr. Mehdi Dadkhah, Department of Management, Faculty of Economics and Administrative Sciences, Ferdowsi University of Mashhad, Mashhad, Iran. E-mail: mehdidadkhah@mail.um.ac.ir*

Submitted: 27-May-2020 **Revised:** 03-Aug-2020
Accepted: 15-Aug-2020 **Published:** 30-Jan-2021

References

1. Silge J, Robinson D. Text Mining with R: A Tidy Approach. Printed in the United States of America: O'Reilly Media, Inc.; 2017.
2. Kumar A, Paul A. Mastering Text Mining with R. Livery Place35 Livery StreetBirmingham B3 2PB, UK: Packt Publishing Ltd.; 2016.
3. Keys T. Using Topic Models Package and Latent Dirichl Allocation to Location to Identify Topics in Texts; 2017. Available from: https://rstudio-pubs-static.s3.amazonaws.com/266565_171416f6c4be464fb11f7d8200c0b8f7.html. [Last accessed on 2020 Aug 12].
4. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003;3:993-1022.
5. Statista. Most Popular Social Networks Worldwide as of July 2020, Ranked by Number of Active Users. Available from: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/. [Last accessed on 2020 Aug 04].
6. Statista. Regional Distribution of Desktop Traffic to Reddit.com as of May; 2020. Available from: https://www.statista.com/statistics/325144/reddit-global-active-user-distribution/. [Last accessed on 2020 Aug 04].
7. Duggan M, Smith A. 6% of Online Adults are Reddit Users; 2013. Available from: https://www.pewresearch.org/internet/2013/07/03/6-of-online-adults-are-reddit-users/. [Last accessed on 2020 Aug 04].
8. Jelodar H, Wang Y, Orji R, Huang H. Deep sentiment classification and topic discovery on novel coronavirus or covid-19 online discussions: Nlp using lstm recurrent neural network approach. arXiv preprint arXiv 2020. doi: 10.1109/JBHI.2020.3001216.
9. Stokes DC, Andy A, Guntuku SC, Ungar LH, Merchant RM. Public priorities and concerns regarding COVID-19 in an online discussion forum: Longitudinal topic modeling. J Gen Intern Med 2020;35:2244-7.
10. Ordun C, Purushotham S, Raff E. Exploratory analysis of covid-19 tweets using topic modeling, umap, and digraphs. arXiv preprint arXiv 2020.
11. Gozzi N, Tizzani M, Starnini M, Ciulla F, Paolotti D, Panisson A, *et al*. Collective response to the media coverage of COVID-19 Pandemic on Reddit and Wikipedia. arXiv preprint arXiv 2020.
12. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, *et al*. The covid-19 social media infodemic. arXiv preprint arXiv 2020.
13. Murray C, Mitchell L, Tuke J, Mackay M. Symptom extraction from the narratives of personal experiences with COVID-19 on Reddit. arXiv preprint arXiv 2020;10454.
14. Zhang JS, Keegan BC, Lv Q, Tan C. A tale of two communities: Characterizing reddit response to COVID-19 through/r/China_Flu and/r/Coronavirus. arXiv preprint arXiv 2020.
15. Chew C, Eysenbach G. Pandemics in the age of twitter: Content analysis of tweets during the 2009 H1N1 outbreak. PLoS One 2010;5:e14118.
16. Dhand NK, Hernandez-Jover M, Taylor M, Holyoake P. Public perceptions of the transmission of pandemic influenza A/H1N1 2009 from pigs and pork products in Australia. Prev Vet Med 2011;98:165-75.
17. Seale H, McLaws M, Heywood AE, Ward KF, Lowbridge CP, van D, *et al*. The community's attitude towards swine flu and pandemic influenza. Med J Australia Wiley Online Lib 2009;191:267-9.
18. R Core Team. R: A Language and Environment for Statistical Computing Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/. [Last accessed on 2020 Aug 12].
19. Rivera I. RedditExtractoR: Reddit Data Extraction Toolkit; 2019. Available from: https://CRAN.R-project.org/package=RedditExtractoR. [Last accessed on 2020 Aug 12].
20. Wickham H, François R, Henry L, Müller K. Dplyr: A Grammar of Data Manipulation; 2020. Available from: https://CRAN.R-project.org/package=dplyr. [Last accessed on 2020 Aug 12].
21. Nikita M. ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters; 2019. Available from: https://CRAN.R-

project.org/package=ldatuning. [Last accessed on 2020 Aug 12].
22. Grün B, Hornik K. Topicmodels: An R package for fitting topic models. J Statistical Software 2011;40:1-30.
23. Lang D, Chien G. Wordcloud2: Create Word Cloud by "htmlwidget; 2018. Available from: https://CRAN.R-project.org/package=wordcloud2. [Last accessed on 2020 Aug 12].
24. Feinerer I, Hornik KT. Text Mining Package; 2019. Available from: https://CRAN.R-project.org/package=tm. [Last accessed on 2020 Aug 12].
25. Feinerer I, Hornik K, Meyer D. Text mining infrastructure in R. J Statist Software 2008;25:1-54.

Access this article online

Quick Response Code:

Website: www.jmssjournal.net

DOI: 10.4103/jmss.JMSS_36_20

How to cite this article: Kimiafar K, Dadkhah M, Sarbaz M, Mehraeen M. An analysis on top commented posts in reddit social network about COVID-19. J Med Sign Sens 2021;11:62-5.