

EPIDEMIOLOGICAL STUDY

Evaluating YouTube as a source of information on COVID-19: analysis with latent Dirichlet allocation method

Doganer A¹, Zhang ZJ²

Department of Biostatistics and Medical Informatics, Faculty of Medicine, Kahramanmaraş Sutcu Imam University, Kahramanmaraş, Turkey. adem_doganer@hotmail.com

ABSTRACT

OBJECTIVE: The global impact of COVID-19 pandemic has gained momentum rapidly. People have little information about SARS-CoV-2 (Coronavirus). Internet has become a frequently used tool to obtain information in recent years, while YouTube is one of the popular sources of information with many videos on its platform. This study aims to identify the topics regarding Coronavirus that people learned about on YouTube. The videos about Coronavirus were also evaluated in terms of the reliability of their source of information.

METHODS: In total, 160 videos on Coronavirus that had 500,000 or more views were analysed. The latent Dirichlet allocation method was used in the process of identifying the topics that were then compared in terms of video parameters. The reliability of the source of information provided by videos was assessed with a modified DISCERN tool.

RESULTS: A proportion of 15.6 % of these videos had a scientific content, while 45 % of these videos were about the process entailed by the COVID-19 pandemic. In terms of video reliability, the difference between video types was found to be significant; videos with scientific content had more reliable sources of information ($p < 0.001$).

CONCLUSION: It has been determined that the videos about the symptoms, diagnosis and treatment of COVID-19, and those with scientific content have the most reliable source of information on Coronavirus (Tab. 5, Fig. 1, Ref. 35). Text in PDF www.elis.sk

KEY WORDS: coronavirus, SARS-CoV-2, COVID-19 pandemic, coronavirus pandemic, latent Dirichlet allocation, information, YouTube.

Introduction

The disease caused by virus SARS-CoV-2 (COVID-19) was first diagnosed in a human being in China's Wuhan province in late December 2019 (1). The virus SARS-CoV-2 (Coronavirus) was quickly transmitted from person to person. As a result, many people in Wuhan got quickly infected and infection then spread into the whole of China. After the Coronavirus had spread to many other countries over the course of January 2020, the pandemic was declared (2). The COVID-19 pandemic has been observed in Europe, America, and many countries in Africa. A large number of deaths related to Coronavirus were reported in a short period of time. The pandemic continued to disperse to many countries during February and March. As of April 25, approximately 3 million cases and 210,000 deaths were documented worldwide (3).

Patients infected with Coronavirus may show different symptoms. A certain number of infected patients have recovered asymptotically. Some of the patients may recover with only mild symptoms (4). Approximately 15% of patients have severe symptoms. The most common symptoms of COVID-19 are high fever, dry cough, difficulty in breathing, problems related to taste, and a ground-glass opacity on CT scan (5). The groups at risk of severe symptoms of the disease have been found to be those of the elderly, individuals with chronic conditions, and individuals with immune disorders.

Coronavirus has caused great anxiety worldwide. Most citizens have complied with the measures taken by governments. Curfew, quarantine, social distancing practices, and use of disinfectants are among the most applied measures (6). People are now facing a pandemic that they do not know much about. People had limited information on the COVID-19 pandemic. They needed to learn more about the Coronavirus in order to protect themselves from the pandemic. The fact that people had to distance themselves from the social environment limited their information sources. The most effective information sources in this process have been TV programs, radio programs, and internet. TV and radio programs provide information only at certain points of time of the day, whereas internet enables us to obtain information instantly whenever we want. The widely used tools such as smartphones and computers

¹Department of Biostatistics and Medical Informatics, Faculty of Medicine, Kahramanmaraş Sutcu Imam University, Kahramanmaraş, Turkey, and ²Department of Management, Coggin College of Business, University of North Florida, Jacksonville, FL, USA

Address for correspondence: A. Doganer, PhD, Department of Biostatistics and Medical Informatics, Faculty of Medicine, Kahramanmaraş Sutcu Imam University, Bahcelievler Yerleskesi, Dulkadiroglu, Kahramanmaraş, Turkey, 46050.

Phone: +90.344.3002656, Fax: +90.344.30034 09

play the most important role in obtaining information on internet. Internet can be used for a wide range of reasons. Internet is often preferred because it provides easy access to information (7). One of the websites that people prefer to obtain information from is YouTube (<http://www.youtube.com>). YouTube is a video-sharing and video-watching platform followed by many people (8). As one of the most important information sources for the humankind, YouTube contains many videos that vary in genre, form and content. Video parameters such as view count, number of comments, and number of likes and dislikes are also available on YouTube (9).

Text-mining methods have made great progress in recent years, which plays a major role in text-based analyses on websites. Depending on their usage area, there are different text-mining methods that can be used in many fields such as pattern recognition, classification, topic extraction, sentiment analysis, information retrieval, information extraction, summarization and modelling (10). The latent Dirichlet allocation (LDA) model conducts topic prediction from text, which then can be applied to identify hidden issues out of words in text documents (11). LDA is an unsupervised classification method that considers the probability distribution of words. The LDA method allows for topic predictions that can be made in many documents and in large text chunks. In the theoretical structure of the LDA method, the Dirichlet distribution, Gibbs sampling, and expectation-maximization processes are applied (12). LDA is very successful in identifying hidden issues in internet-based texts and posts on social media (13).

YouTube is one of the most popular websites containing many videos on various areas. One of the purposes of using YouTube is to obtain information. Researchers have conducted many studies on the use of YouTube as an information source. For instance, Barry et al (2016) conducted a study on YouTube's contribution to anatomy training (14). YouTube is often preferred as an information on health and science. Tackett et al conducted research in order to identify YouTube's contribution to medical education (15). YouTube provides detailed information on protection, infection prevention, and early diagnosis for many diseases. For instance, Lim et al (2018) conducted research on identifying the contribution of videos containing information on "infection prevention and control" to science (16).

YouTube was considered an important and effective source for people to access information during the epidemic of the H1N1 flu and the Ebola virus (17–20). Since the COVID-19 pandemic is a rather current issue, very few publications have been found on this subject. Several recent studies have explored the role of YouTube as a source for people to obtain information about Coronavirus. For instance, Khatri et al (2020) analysed the reliability of YouTube as an information source in the COVID-19 pandemic (21). In the study, it was reported that YouTube was preferred more in this pandemic compared to previous pandemics in terms of obtaining information. It was also stated that the content of some videos was medically insufficient. D'Souza et al (2020) studied 113 YouTube videos related to Coronavirus and attempted to identify useful versus misleading videos (22). However, these studies have not specifically identified people's preferences when watching YouTube videos, which is the focus of our research. We studied

the topics of YouTube videos related to Coronavirus and found that they prefer choosing scientific videos to access information.

This study aims to identify topics of YouTube videos on Coronavirus from their titles using the LDA method. In particular, we identify the topics that people chose on YouTube while trying to learn more about Coronavirus during the pandemic process. Topics related to Coronavirus that people were curious about, video contents that they watched, and their reactions to these videos were analysed.

The remainder of the paper proceeds as follows. The next section briefly summarizes related works. Section 3 outlines the material and method of this research. Section 4 presents the results and discusses the statistical analysis. Section 5 concludes the paper by highlighting contributions and limitations of the study as well as future research directions.

Material and methods

Latent Dirichlet allocation

Various machine-learning methods have been used to identify the content of videos on YouTube. LDA is one of the most used unsupervised classification methods for topic modelling. In the LDA method, it is aimed to model documents and texts among topics defined according to the probability distribution of words (23). LDA produces graphs used for modelling discrete data such as text files, while revealing hidden issues within the text (12). LDA conducts 3 productive processes for each text document (24).

1. A random topic should be chosen from the topic distributions in each text document.
2. Word sampling should be chosen from the word distribution related to the selected topic.
3. Operations should be repeated for all words in the text (25).

LDA identifies hidden issues comprised in the text file. LDA conducts these operations using the Dirichlet distribution, a multivariate form of beta distribution (26). LDA does not take into account for modelling the word order in the text. LDA evaluates the possibility of words being together (27). The graphical model of the LDA is given in Figure 1.

LDA is one of the methods used in identifying the topic of contents from video titles and video comments. Several studies have been conducted on identifying the topic of content on YouTube videos using the LDA method (28–30). Our research adopts LDA to predict titles according to the titles of YouTube videos.

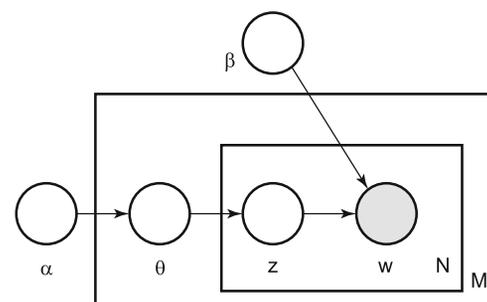


Fig. 1. Graphical model of latent Dirichlet allocation (12).

Methodology

In the study, the word “Coronavirus” was searched on YouTube. The research was conducted between 20–25 April 2020 (31). As a result, a ranking list was done for videos with the highest number of views. Videos with 500,000 or more views were included in the study. Videos that appeared on the list more than once were excluded from the study. Finally, 160 videos containing information on Coronavirus were obtained. In the study, the information on these videos such as the number of views, number of comments, likes, dislikes, and duration were recorded. We did not evaluate any human participants or animals in this study. Videos that anyone can access were evaluated. Therefore, there was no need for approval of the ethics committee for this study.

LDA method was used to predict topics from video titles. As a result of the analysis, 4 topics related to COVID-19 were identified, including (i) symptoms, diagnosis, and treatment; (ii) case statistics, (iii) measures against the spread of Coronavirus, and (iv) pandemic process. These videos were then classified as news videos, informational videos, or scientific videos. In addition, according to the speakers providing the information, the videos were categorized as containing information provided by health professionals (physician, academic, etc.) or other (news anchor, program presenter, etc.). The content of videos was also taken into account in the process of classifying these videos. The number of views, number of comments, number of likes and dislikes, and duration were compared according to their topics and classification. Table 1 shows the topics determined using the LDA method.

Tab. 1. Topics determined with LDA method.

Topics	Words in Titles
Symptoms, diagnosis and treatment of COVID-19	Disease, Symptom, Information, Patient, Body, Infection, Harm, Virus, Treatment, Vaccine, Lab, Pneumonia, Facts, Experience, Skin, Recovery, Similar, Doctor, Infected, Flu, Allergy, Resistance, SARS
Measures against the spread of Coronavirus	Country, Management, Music, Disinfectant, Struggle, Prevention, Manager, Warning, Precaution, defence, Emergency, Hydroxychloroquine, Panic, Restriction, Announcement, Quarantine, Health, Expert, Mask
Case statistics on COVID-19	Case, Number, Information, Death, Mortality, Young, American, WHO, Peak, Level, epidemiology, Modelling
COVID-19 pandemic process	Wild, Animal, Propagation, Society, Good, Hospital, Task, Force, Future, Satellite, Media, Brave, Italy, Bangladesh, War, USA, Process, City, Complo, Epidemic, President, China, Bad, Will End, Protest, Criticism, Why, Equator, Hospital, Second, Wave, News, Economy, India, Wuhan, Attitude, Ecuador, Bioswar, Death, Martial Law, Economy, Colombia, Danger, Manufacturer, Bat, Spain, France, New York, Pandemic, Taiwan, Nurse, Grave

Tab. 2. Modified DISCERN tool questions.

Questions
Is the video clear, concise, and understandable?
Are reliable sources of information used?
Is the presented information balanced and unbiased?
Are additional sources of information listed for reference?
Are areas of uncertainty mentioned?

Assessment of reliability

A modified DISCERN tool was used to evaluate the videos in terms of the reliability of the source of provided information. The modified DISCERN tool consists of 5 questions. The reliability score of each video is in the range of 0–5 points, while a 5-point score represents the highest level of reliability. The questions were adopted from the DISCERN tool (32–33) and modified. The videos were evaluated by the authors of the study. Support was received from physicians in terms of the assessment of videos that required medical knowledge. The agreement between the two authors was assessed using the Kappa coefficient. The modified DISCERN tool questions are presented in Table 2 (33).

Statistical analysis

The fit of variables to normal distribution was analysed using the Kolmogorov–Smirnov test in the dataset. The Mann-Whitney U test was used in two-group comparison of variables that did not have normal distribution. The comparison of three or more groups was performed using the Kruskal-Wallis H test. The Dunn-Sidak test was used for post-hoc analysis. Statistical parameters were shown as median (Q1–Q3). The statistical significance was accepted at $p < 0.05$. The data analysis was performed using IBM SPSS Statistics for Windows Version 22 (IBM SPSS for Windows version 22, IBM Corporation, Armonk, New York, United States). The R-3.3.2 software was used in the process of identifying the topics by means of latent Dirichlet allocation method.

Results

In the study, 160 videos about Coronavirus most frequently watched on YouTube were analysed. A proportion of 56.9 % of these videos are news videos while 15.6 % of them contain scientific information and 45 % of these videos contain information on events and news during the COVID-19 pandemic. A proportion of 10% of these videos contain information presented by doctors or academics. The findings on the distribution of video contents are given in Table 3.

Table 4 shows the findings on comparisons according to the types of videos. The difference between types of videos was found to be statistically significant in terms of the numbers of views ($p = 0.016$). Scientific videos were viewed more frequently than the news videos. The highest number of views was achieved in the category of scientific videos. The difference between the types of videos was found to be significant in terms of the number of likes given to videos ($p = 0.001$). It was observed that the highest number of likes was achieved in the category of scientific videos. The difference between the types of videos was found to be statisti-

Tab. 3. Distribution of videos.

		n	%
Video types	Informational	44	27.5
	Scientific	25	15.6
	News	91	56.9
Video topics	Measures against the spread of Coronavirus	34	21.3
	COVID-19 pandemic process	72	45.0
	Symptoms, diagnosis and treatment of COVID-19	37	23.1
	Case statistics on COVID-19	17	10.6
Video speakers	Healthcare professional (physician, academician, etc.)	16	10.0
	Others (program speaker, news anchor, etc.)	144	90.0

cally significant in terms of video duration ($p = 0.006$). The duration of informational videos is longer as compared to other types of videos. There was no difference between the types of videos in terms of the number of comments and dislikes. Table 4 demonstrates the findings on comparisons according to video topics. There was no statistically significant difference between the topics of videos on Coronavirus regarding the number of views, number of comments, number of likes and dislikes, or video duration ($p > 0.05$). Table 4 displays the findings on comparisons according to narrators. No statistically significant difference was found between videos narrated by healthcare professionals and others in terms of the number of views, number of comments, number of likes and dislikes or video duration ($p > 0.05$).

Tab. 4. Comparison of video parameters.

Types [†]		Number of views	Number of comments	Number of likes	Number of dislikes	Duration of videos (min.)
Informational	Median	2247500	4535	27000	2050	10.11
	(Q1–Q3)	(1237500–5158500)	(2290–11111)	(16500–99000) ^c	(1050–3900)	(6.15–21.04) ^{b,c}
Scientific	Median	3934000	5139	61000	2100	6.08
	(Q1–Q3)	(2440000–8854000) ^c	(2012–13113)	(22000–158000) ^c	(1300–5200)	(4.49–8.45) ^a
News	Median	1921000	7102	17000	2400	7.32
	(Q1–Q3)	(1212000–3365000) ^b	(4024–12888)	(9700–50000) ^{a,b}	(1400–6000)	(4.22–12.45) ^a
p		0.016*	0.146	0.001*	0.508	0.006*
Topics [†]		Number of views	Number of comments	Number of likes	Number of dislikes	Duration of videos (min.)
Measures against the spread of Coronavirus	Median	2388000	9818	40000	2800	8.23
	(Q1–Q3)	(1167000–5997000)	(4024–22612)	(15000–112000)	(1300–9400)	(5.45–19.28)
COVID-19 pandemic process	Median	1915500	5987	20000	2500	7.42
	(Q1–Q3)	(1211000–3236000)	(3162–12888)	(10000–58000)	(1300–5300)	(5.19–13.23)
Symptoms, diagnosis treatment of COVID-19	Median	2680000	4578	38000	1600	5.58
	(Q1–Q3)	(1390000–4134000)	(2290–8922)	(20000–103000)	(1200–3700)	(4.52–9.02)
Case statistics on COVID-19	Median	2445000	7023	13000	2500	8.01
	(Q1–Q3)	(1782000–4980000)	(4787–11859)	(9800–52000)	(1500–5700)	(5.29–19.35)
p		0.224	0.113	0.055	0.284	0.236
Speakers ^{††}		Number of views	Number of comments	Number of likes	Number of dislikes	Duration of videos (min.)
Healthcare Professional	Median	3283500	7952	33000	2750	5.78
	(Q1–Q3)	(1935000–5800500)	(3893–19067)	(22000–131500)	(1300–5000)	(4.40–11.30)
Others	Median	2187000	6218	21000	2250	7.49
	(Q1–Q3)	(1256000–4074500)	(3048–12218)	(11000–71000)	(1300–5200)	(5.19–13.23)
p		0.137	0.670	0.138	0.945	0.380

[†]Kruskal–Wallis H test; ^{††}Mann–Whitney U test; $\alpha: 0.05$; Post-hoc: Dunn–Sidak test; Median (%25 Quartile–%75 Quartile) *Difference between the groups is statistically significant; ^aDifference with informational videos is statistically significant; Difference with scientific videos is statistically significant; Difference with news videos is statistically significant

Table 5 shows the findings on comparisons for the use of videos as an information source. The difference between the types of videos was found to be statistically significant in terms of the DISCERN score ($p < 0.001$). The highest modified DISCERN tool score was achieved in the category of scientific videos. The difference between the types of topics was found to be statistically significant in terms of the DISCERN score ($p < 0.001$). The highest modified DISCERN tool score was achieved by vid-

eo topics informative of symptoms, diagnosis and treatment of COVID-19.

Discussion

The COVID-19 pandemic has changed the global agenda. This pandemic has caused anxiety for the whole of humanity. People did not have enough information on the situation that they were about to face. Internet has been one of the information sources on Coronavirus. The primary information sources in the pre-internet era were television, radio, and printed publications. Websites have become popular information sources with the advent of internet (34). Internet provides quick access to information. One of the

Tab. 5. Comparison of modified DISCERN tool score by topics and types of video.

Modified DISCERN Tool Score			p
Types	Informational, Median(Q1–Q3)	3.00(2.00–4.00) ^{y,z}	p<0.001*
	Scientific, Median(Q1–Q3)	4.00(3.00–4.00) ^{x,z}	
	News, Median(Q1–Q3)	2.00(1.00–3.00) ^{x,y}	
Topics	Measures against the spread of Coronavirus, Median(Q1–Q3)	2.00(1.00–3.00) ^c	p<0.001*
	COVID-19 pandemic process, Median(Q1–Q3)	2.00(1.00–3.00) ^c	
	Symptoms, diagnosis and treatment of COVID-19, Median(Q1–Q3)	4.00(4.00–4.00) ^{a,b,d}	
	Case statistics on COVID-19, Median(Q1–Q3)	2.00(2.00–3.00) ^c	

Kruskal–Wallis H test; α :0.05; Post-hoc: Dunn-Sidak test; Median(%25 Quartile-%75Quartile);*Statistical significance; ^a difference with group Informational is statistically significant; ^y difference with group Scientific is statistically significant; ^z difference with group News is statistically significant; ^b difference with group Measures against the spread of Coronavirus is statistically significant; ^c difference with group COVID-19 pandemic process is statistically significant; ^d difference with group Symptoms, diagnosis and treatment of COVID-19 is statistically significant; ^e difference with group Case statistics on COVID-19 is statistically significant

popular sources of information on health is YouTube (35). The data obtained from internet are an important indicator in the process of identifying the topics, on which people want to obtain information.

When analysing 160 videos on Coronavirus, that were viewed most frequently, our research identified 4 topics from video titles using the LDA method. These topics are diagnosis and treatment, case statistics, measures against the spread of Coronavirus, and the pandemic process. A proportion of 45 % of these videos were related to the pandemic process, of which 56.9 % were news videos. Health professionals were narrators in 10 % of the videos. The number of uploaded videos that were related to the COVID-19 pandemic process was higher than the number of videos on other topics. The number of news videos on Coronavirus uploaded on YouTube is higher compared to other types of videos. The number of videos, in which health professionals were speakers was the lowest among them. It is observed that these videos are generally related to the television news programs. News program videos have mostly been aimed at describing the events during the pandemic. Information has generally been provided by the news announcers on these programs.

Although the number of uploaded news program videos was higher, people preferred to view scientific videos. News program videos were the least viewed video type. Similarly, the scientific videos were most frequently liked by viewers, while the news program videos were liked least frequently. The highest numbers of dislikes were given to news program videos. The highest number of comments were made in response to news program videos. Most frequently, the viewers preferred to obtain information on Coronavirus from scientific videos.

When people's preferences on obtaining information were analysed according to video topics, it was observed that the most frequently viewed video topics were those dealing with symptoms, diagnosis, and treatment of COVID-19. The topic with the highest number of comments was found to be that dealing with the measures against the spread of Coronavirus. People preferred to obtain information mostly on symptoms, diagnosis, and treatment of COVID-19, as well as on measures taken against the spread of Coronavirus.

When people's preferences on obtaining information were analysed according to speakers, the videos with information provided by health professionals were viewed more frequently than other videos. Similarly, it was observed that the number of com-

ments and likes was the highest in response to videos with health professionals as speakers. People preferred to obtain information from videos, in which health professionals were the narrators.

There are certain limitations to this study. Since COVID-19 pandemic is a current issue, the information on Coronavirus may be changing constantly. Although there are many videos on the subject, only those narrated in English and with 500,000 or more views have been included in the study. YouTube is a dynamic video website. The number of videos, views, comments, likes, and dislikes keeps changing. Future studies should take into account the dynamic changes in the pandemic and YouTube content to predict the topics preferred by people.

Conclusion

This study uses the numbers of views, comments, likes and dislikes on YouTube videos to identify topics, on which people want to obtain information. With the LDA method, we capture and classify hidden topics from YouTube video titles about Coronavirus.

Our findings show that people often use YouTube in order to learn more about Coronavirus. People mostly prefer to watch scientific videos in order to obtain information on Coronavirus. Videos on symptoms, diagnosis, and treatment of COVID-19 as well as on measures against the spread of Coronavirus have been viewed more frequently as compared to videos dealing with other topics. Videos containing information provided by health professionals have been viewed more frequently than other videos. YouTube has become a quick and easily accessible information source during the COVID-19 pandemic.

The videos were assessed in terms of the reliability of the source of provided information. A modified DISCERN tool was applied to evaluate the videos. The highest reliability of the source of information was observed in videos with the topic of symptoms, diagnosis and treatment of COVID-19. Similarly, the highest reliability of the source of information was observed in videos with a scientific content.

References

- Jiang F, Deng L, Zhang L, Cai Y, Cheung CW, Xia Z. Review of the Clinical Characteristics of Coronavirus Disease 2019 (COVID-19). *Journal of General Internal Medicine* 2020; 4: 1–5.

2. **Riou J, Althaus CL.** Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance* 2020; 25 (4), 2000058: 1–5.
3. **Worldometer** (2020), “COVID-19 Coronavirus Pandemic”. available at <https://www.worldometers.info/coronavirus/>. (accessed 25 April 2020).
4. **Adhikari SP, Meng S, Wu YJ, Mao YP, Ye RX, Wang QZ, et al.** Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: A scoping review. *Infect Dis Poverty* 2020; 9 (1): 1–2.
5. **Qin C, Liu F, Yen TC, Lan X.** 18F-FDG PET/CT findings of COVID-19: a series of four highly suspected cases. *Eur J Nucl Med Mol Imaging* 2020; 22: 1–6.
6. **Studdert DM, Hall MA.** Disease Control, Civil Liberties, and Mass Testing – Calibrating Restrictions during the COVID-19 Pandemic. *New Engl J Med* 2020; 383 (2): 102–104.
7. **Majid S, Tan AT.** Usage of information resources by computer engineering students: A case study of Nanyang technological university, Singapore. *Online Information Rev* 2002; 26 (5): 318–325.
8. **Lange PG.** Publicly private and privately public: Social networking on YouTube. *J Comp Comm* 2007; 13 (1): 361–380.
9. **Khan ML.** Social media engagement: What motivates user participation and consumption on YouTube? *Comp Human Behav* 2017; 66: 236–247.
10. **Vidhya K. A, Aghila G.** Text Mining Process , Techniques and Tools : an Overview. *Internat J Inform Technol Knowledge Manag* 2010; 2 (2): 613–622.
11. **Maskeri G, Sarkar S, Heafield K.** Mining business topics in source code using Latent Dirichlet Allocation. In: *Proceedings of the 2008 1st India Software Engineering Conference, ISEC'08.* 2008: 113–120.
12. **Blei DM, Ng AY, Jordan MI.** Latent Dirichlet allocation. *J Machine Learning Res* 2003; 3: 993–1022.
13. **Ostrowski DA.** Using latent dirichlet allocation for topic modelling in twitter. In: *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing, IEEE ICSC 2015.* 2015: 493–497.
14. **Barry DS, Marzouk F, Chulak-Oglu K, Bennett D, Tierney P, O’Keefe GW.** Anatomy education for the YouTube generation. *Anat Sci Educ* 2016; 9 (1): 90–96.
15. **Tackett S, Slinn K, Marshall T, Gaglani S, Waldman V, Desai R.** Medical education videos for the world: An analysis of viewing patterns for a youtube channel. *Acad Med* 2018; 93 (8): 1150–1156.
16. **Lim K, Kilpatrick C, Storr J, Seale H.** Exploring the use of entertainment-education YouTube videos focused on infection prevention and control. *Amer J Infect Control* 2018; 46 (11): 1218–1223.
17. **Pandey A, Patni N, Singh M, Sood A, Singh G.** YouTube As a Source of Information on the H1N1 Influenza Pandemic. *Amer J Prevent Med* 2010; 38 (3): 1–3.
18. **Pathak R, Poudel DR, Karmacharya P, Pathak A, Aryal MR, Mahmood M, et al.** YouTube as a source of information on ebola virus disease. *North Amer J Med Sci* 2015; 7 (7): 306–309.
19. **Basch CH, Basch CE, Ruggles K V., Hammond R.** Coverage of the Ebola Virus Disease Epidemic on YouTube. *Disaster Med Publ Health Preparedness* 2015; 9 (5): 531–535.
20. **Nagpal SJS, Karimianpour A, Mukhija D, Mohan D, Brateanu A.** YouTube videos as a source of medical information during the Ebola hemorrhagic fever epidemic. *SpringerPlus* 2015; 4 (1): 1–5.
21. **Khatrri P, Singh SR, Belani NK, Yeong YL, Lohan R, Lim YW, et al.** YouTube as source of information on 2019 novel coronavirus outbreak: a cross sectional study of English and Mandarin content. *Travel Med Infect Dis* 2020; 35 (2020): 1–6.
22. **D’Souza RS, D’Souza S, Strand N, Anderson A, Vogt MNP, Olatoye O.** YouTube as a source of medical information on the novel coronavirus 2019 disease (COVID-19) pandemic. *Global Publ Health* 2020; 15 (7): 935–942.
23. **Blei D, Lafferty J.** “Topic Models”, In “Text mining”, Chapman and Hall/CRC, 2009: 101–124.
24. **Hidayatullah AF, Ma’Arif MR.** Road traffic topic modeling on Twitter using latent dirichlet allocation. In: *Proceedings - 2017 International Conference on Sustainable Information Engineering and Technology, SIET 2017.* 2018: 47–52.
25. **Hong L, Davison BD.** Empirical study of topic modeling in Twitter. In: *SOMA 2010 – Proceedings of the 1st Workshop on Social Media Analytics 2010:* 80–88.
26. **Ng KW, Tian GL, Tang ML.** Dirichlet and Related Distributions: Theory, Methods and Applications. *John Wiley&Sons* 2011.
27. **Ekinci E, Omurca Sİ, KIRIK E, TAŞÇI Ş.** Tıp Veri Kümesi için Gizli Dirichlet Ayrımı. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi* 2020; 22 (64): 67–80.
28. **Sureka A, Kumaraguru P, Goyal A, Chhabra S.** Mining YouTube to discover extremist videos, users and hidden communities. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* 2010.
29. **Daniel C, Dutta K.** Automated Generation of Latent Topics on Emerging Technologies from YouTube Video Content. In: *Proceedings of the 51st Hawaii International Conference on System Sciences* 2018.
30. **Morchid M, Linares G.** A LDA-based method for automatic tagging of Youtube videos. In: *International Workshop on Image Analysis for Multimedia Interactive Services.* IEEE 2013: 1–4.
31. **YouTube** (2020). “Coronavirus”, available at https://www.youtube.com/results?search_query=coronavirus.
32. **Charnock D, Shepperd S, Needham G, Gann R.** DISCERN: An instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Comm Health* 1999; 53 (2): 105–111.
33. **Singh AG, Singh S, Singh PP.** YouTube for information on rheumatoid arthritis – A wakeup call? *J Rheum* 2012; 39 (5): 899–903.
34. **Fernandez-Luque L, Elahi N, Grajales FJ.** An analysis of personal medical information disclosed in youtube videos created by patients with multiple sclerosis. *Studies in Health Technology and Informatics* 2009; 150: 292–296.
35. **Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK.** Healthcare information on YouTube: A systematic review. *Health Informatics J* 2015; 21 (3): 173–194.

Received November 19, 2020.
Accepted December 15, 2020.