

소셜미디어에 나타난 코로나 바이러스(COVID-19) 인식 분석

윤상후¹, 정상윤¹, 김영아^{2*}

¹대구대학교 수리빅데이터학부, ²제주대학교 간호대학 · 건강과간호연구소

Trend Analysis of Corona Virus(COVID-19) based on Social Media

Sanghoo Yoon¹, Sangyun Jung¹, Young A Kim^{2*}

¹Division of Mathematics and Big data science, Daegu University

²College of Nursing, Jeju National University · Health & Nursing Research Institute

요약 본 연구는 국내 소셜미디어를 기반으로 코로나 확산 시기에 따른 코로나19 관심사 변화를 텍스트 기반으로 살펴 보았다. 연구자료는 2020년 1월 20일부터 8월 15일까지 네이버와 다음의 블로그와 카페에 올라온 글이다. 코로나 확산 시기는 총 3단계로 분류하였다. 중국에서 발견된 코로나19가 한국에 확산되기 시작한 1월 20일부터 2월 17일을 '전조기', 대구를 중심으로 본격적 확산을 진행한 2월 18일부터 4월 20일을 '심각기', 그리고 일 확진자 수가 안정화되는 4월 21일부터 8월 15일을 '안정기'로 명명하였다. 코로나19와 연관된 상위 50개 단어를 추출하여 TF-IDF를 이용하여 군집 분석 하였다. 분석결과 전조기는 코로나 '상황'에 관련된 텍스트가 많았고, 심각기에는 '국가'와 '감염경로'에 관련된 텍스트가 많았다. 안정기에는 '치료'가 주로 언급되었다. 시기와 무관하게 공통적으로 언급이 많이 된 단어는 '감염', '마스크', '사람', '발생', '확진', '정보'이다. 시기별 감정의 변화를 살펴보면 시간이 지남에 따라 긍정의 비율이 높아지고 있다. 카페와 블로그는 글쓴이의 생각과 주관이 담긴 글을 인터넷을 통해 공유하므로 코로나19로 인한 비대면 시대의 주요 정보 공유 공간이다. 그러나 정보전달의 선택성과 임의성이 존재하므로 소셜미디어에서 생산되는 정보를 비판적으로 바라보는 시각이 필요하다.

Abstract This study deals with keywords from social media on domestic portal sites related to COVID-19, which is spreading widely. The data were collected between January 20 and August 15, 2020, and were divided into three stages. The precursor period is before COVID-19 started spreading widely between January 20 and February 17, the serious period denotes the spread in Daegu between February 18 and April 20, and the stable period is the decrease in numbers of confirmed infections up to August 15. The top 50 words were extracted and clustered based on TF-IDF. As a result of the analysis, the precursor period keywords corresponded to congestion of the Situation. The frequent keywords in the serious period were Nation and Infection Route, along with instability surrounding the Treatment of COVID-19. The most common keywords in all periods were infection, mask, person, occurrence, confirmation, and information. People's emotions are becoming more positive as time goes by. Cafes and blogs share text containing writers' thoughts and subjectivity via the internet, so they are the main information-sharing spaces in the non-face-to-face era caused by COVID-19. However, since selectivity and randomness in information delivery exists, a critical view of the information produced on social media is necessary.

Keywords : Blog, Clustering, COVID-19, Sentimental analysis, Social media

이 논문은 2020학년도 제주대학교 교원성과지원사업에 의하여 연구되었음.

*Corresponding Author : Young A Kim(Jeju National Univ.)

email: yakim@jeju.ac.kr

Received January 27, 2021

Accepted May 7, 2021

Revised February 26, 2021

Published May 31, 2021

1. 서론

코로나 바이러스 감염증-19(Corona virus disease 19, COVID-19, 이하 코로나19)는 2019년 12월 중국 후베이성 우한시에서 발생하여 2020년 1월 중순 한국으로 전염되기 시작했다. 한국은 2월 중순 대구와 경북 중심으로 확진자 수가 급증하여 국가적으로 감염병 위기단계를 ‘심각’ 수준으로 격상하여 관리하였다. 세계보건기구(WHO)도 세계적으로 코로나19가 확산되자 2020년 3월 11일 팬데믹(pandemic)을 선포하였다. 하지만 코로나19는 기존 감염병에 비해 전염성이 높아 사태가 장기화되면서 경제적, 사회적, 문화적으로 큰 변화를 맞고 있다[1].

바이러스 확산을 막기 위해 2020년 9월 12일 정부는 질병관리본부를 질병관리청으로 격상하여 일상생활의 사회적 거리두기, 기업의 재택근무, 개인의 외출 자제를 권고하고 있다. 물리적 활동의 위축은 기업실적 부진, 고용 불안 등을 야기할 수 있으며 바이러스 확산에 버티기 어려운 산업과 기업은 사라지고 생활 필수적인 산업과 바이러스 영향이 적은 산업 위주로 산업구조를 변화시킨다[2]. 즉, 코로나19와 함께 사는 언컨택트 사회가 형성되는 것이다. 이러한 변화는 사람 간 물리적 교류를 감소시켜 ‘코로나 블루’라는 우울증을 유발시킨다[3-5].

키워드를 이용한 빅데이터 분석은 키워드와 연관된 이슈를 이해하는데 도움이 된다. 코로나19에 관한 빅데이터 분석의 예로 언론보도를 토픽모델링 한 국가 위기 커뮤니케이션 정책 개발[6]과 유아교육기관 의 교사와 부모의 대응 및 대처방안을 들 수 있다[7]. 이 외에도 코로나19와 관련된 언론 기사 및 댓글을 분석하여 시기별 이슈를 살펴보기도 하였다[8]. 하지만 선행연구는 뉴스를 중심으로 연구되어 코로나19를 바라보는 사람들의 감정이나 관심사를 능동적으로 제시하지 못하는 한계점이 있다.

소셜네트워크서비스(Social network service, SNS)는 인터넷을 통해 지인과의 인간관계를 강화하거나 새로운 인맥을 형성하는 서비스로 프로필기반, 블로그 기반, 버티컬 기반, 마이크로 블로그, 커뮤니티 기반으로 구분된다[9]. 여기서 블로그는 웹과 로그의 줄인 말로 사용자가 글을 올리면 관심 있는 사람들이 찾아와 자유롭게 글을 볼 수 있고, 카페는 분야별 관심사를 공유하는 커뮤니티므로 마케팅의 효과적 수단으로 사용된다. 언론 보도는 중립보도 원칙에 따라 긍정과 부정이 균형을 맞추려는 경향이 있으나 블로그는 개인의 주관에 반영되어 편향될 수 있다[10]. 또한 블로그에서 다루는 범위는 언론

보도에 비해 제한적이지만 개인의 생각과 감정을 주관적으로 표현하므로 음식, 숙박, 마케팅 등의 이슈 분석에 많이 활용된다[11-13]. 따라서 본 연구는 코로나19에 대한 이슈를 살펴보기 위해 언론보도가 아닌 블로그나 카페에서 수집하여 빅데이터 분석을 수행하였다.

코로나19의 확산 시기에 따라 정부의 정책 및 언론 보도의 관심사가 상이하므로 SNS 활동도 확산 시기에 영향을 받는다[6]. 본 연구는 국내에서 1번 코로나19 확진자가 발견된 2020년 1월 20일부터 2020년 8월 중순 집회 이후 2차 확산이 진행되기 전까지 텍스트 자료를 수집하여 분석하였다. 네이버와 다음의 블로그와 카페 글은 ‘텍스툼(Textom)’을 이용하여 시기별로 수집하였고 형태소 및 매트릭스 분석을 통해 단어별 인식변화 및 단어 간의 연결성을 살펴보았다. 추가로 감정사전을 이용한 감정 분석도 수행하였다. 코로나19 확산에 따른 SNS의 반응 해석과 시사점 도출은 향후 팬데믹을 효율적으로 대책을 수립하는데 도움이 될 것이다.

2. 본론

2.1 연구범위 및 방법

2.1.1 연구대상 및 범위

연구자료 수집채널은 국내에서 접근이 용이한 포털사이트 중 네티즌의 사용빈도가 높은 네이버와 다음으로 한정하였다. 코로나19 관련 키워드는 ‘코로나’, ‘COVID-19’, ‘코로나 바이러스’로 지정하여 자료를 수집하였다. 연구기간은 국내에서 처음 코로나19 확진자가 발견된 2020년 1월 20일부터 급격히 증가하기 전인 2월 18일까지를 ‘전조기’로 명하였다. ‘심각기’는 대구와 경북을 중심으로 많은 수의 확진자가 발생한 2020년 2월 19일부터 2020년 4월 20일이고, ‘안정기’는 확진자 수가 안정적으로 감소하기 시작한 2020년 4월 21일부터 2차 확산이 발생하기 전인 2020년 8월 15일까지이다.

2.1.2 텍스트 마이닝

텍스트 마이닝은 비정형인 텍스트 형태의 데이터를 수집, 처리, 분석하는 연구기법이다. 텍스트 마이닝을 위해 자연어 처리에 용이한 텍스툼을 선정하여 형태소 분석을 수행하였다[14]. 형태소란 언어의 최소 의미단위로 어휘적 의미와 문법적 의미를 포함한다. 형태소 분석이란 문장을 최소 의미 단위로 분절하여 어휘 형태소

와 문법 형태소로 구분한다. 본 연구는 ‘코로나19’와 가장 연관된 어휘 형태소 중 빈도가 높은 50개를 선정하였다. 추가로 확진자 수 시기별로 어휘 형태소의 빈도를 제시하여 시기별 인식변화를 비교하였다. 마지막으로 UCINET을 사용하여 텍스트의 관계를 살펴보았다.

2.1.3 군집분석

텍스트 마이닝에서 군집분석이란 형태소 분석을 통해 생성된 키워드를 TF-IDF(term frequency- inverse document frequency) 기반으로 벡터화하여 키워드 간 유사성이나 관련성이 높은 키워드끼리 묶어준다. TF는 문서가 주어졌을 때 키워드의 출현 정도를 정량화 한 수치이다. TF-IDF는 연관성이 없는 단어에 제약조건을 주어 연관성이 높은 단어만 얻을 수 있어 검색엔진에서 주로 사용된다[15]. TF-IDF는 단어의 빈도와 역문서 빈도의 곱으로 특정 문서 안의 키워드의 중요한 정도를 나타내는 수치이다. 본 연구는 TF와TF-IDF 가중치를 사용하여 K 평균 군집분석을 실시하였다. 연구에 고려된 군집수는 6개이다.

2.1.4 감정분석

감정분석(sentimental analysis)은 오피니언 마이닝 (opinion mining)이라고도 불리며 텍스트로 나타난 사람들의 태도, 의견, 성향을 감정사전을 통해 분석한다. 네티즌들이 남긴 텍스트에 언급된 동사와 형용사의 긍정과 부정의 빈도로 감정을 파악한다. 감정은 사람들의 주관적인 의견이 반영되어 표현되므로 키워드에 대한 감정을 평가하면 이를 활용한 마케팅이 가능하다. 본 연구는 코로나19에 대한 긍정과 부정의 빈도를 통해 시기별 감정변화를 살펴보았다.

2.2 연구결과

2.2.1 코로나19 연관 키워드

코로나 19에 관한 뉴스 기사를 토픽 분석하면 정부의 대응조치, 국내 확진자 발생 상황, 그리고 감염경로/사망자 관련으로 구분할 수 있다[6]. 하지만 중단 자료를 시기별로 군집을 분석하면 시기별 차이를 동일한 기준으로 비교하는데 어려움이 있다. 이 문제를 해결하기 위해 연구기간의 모든 자료를 통합하여 텍스트 마이닝을 수행하였다. 포털사이트에서 수집된 블로그와 카페 문서를 형태소 분석을 통해 주요 명사의 빈도로 정리하였다(Table 1). 코로나19와 관련된 단어를 1위부터 50위까지 선별하

였으며, 관련성이 떨어지는 조사, 형용사, 부사, 등은 제외하였다. 단어빈도 옆에 있는 숫자는 군집번호이며 각 군집의 토픽은 ① 코로나19 상황, ② 국가, ③ 정부, ④ 예방, ⑤ 치료, ⑥ 감염경로이다.

Table 1. The top 50 keywords related to COVID-19

No.	Keyword	Freq.	No.	Keyword	Freq.
1	Confirmation①	10,536	26	Visiting②	3,436
2	People⑥	10,110	27	Area②	3,227
3	China②	9,588	28	Pneumonia①	3,223
4	Mask④	9,440	29	Start	3,201
5	Infection①	8,566	30	Government③	3,138
6	Occurrence①	7,906	31	Response③	3,105
7	Case①	7,883	32	Treatment⑤	3,044
8	Nose⑥	7,405	33	Lately	2,904
9	Symptom①	6,341	34	Domestic②	2,833
10	Circumstance①	5,523	35	Situation③	2,728
11	Information	5,350	36	Hand⑥	2,693
12	Work	5,232	37	Support③	2,594
13	Patient	5,089	38	Society③	2,591
14	Present	5,042	39	Method	2,573
15	USA②	5,029	40	Prevention④	2,542
16	Relation	4,596	41	Management③	2,472
17	Diffusion①	4,562	42	Confirmed case	2,280
18	Vaccine⑤	4,550	43	Dead	2,162
19	Today	4,056	44	Hospital④	2,083
20	Korea②	4,049	45	Status①	2,065
21	World②	4,034	46	Remedy⑤	2,054
22	Japan②	3,912	47	Travel	1,978
23	Health④	3,560	48	Wearing④	1,874
24	Development⑤	3,519	49	Declaration③	1,848
25	Inspection①	3,510	50	Dying	1,798

코로나 확산 시기로 구분하여 단어의 빈도를 살핀 것은 Table 2와 같다. 전조기는 중국에서 발생한 감염병이 국내에 확산된 시기이고 심각기는 대구와 경북을 중심으로 1차 지역 확산이 진행된 시기이며 안정기는 기업 및 소상공인들이 정상적인 업무를 위해 준비하며 일반인들도 조금씩 원래 일상으로 돌아가려는 시기이다. 키워드를 중심으로 사람들의 사고변화를 살펴보려 했으나 중복되는 단어가 많지 않아 TF-IDF 기반의 군집분석 결과를 토대로 시기별 특성을 보다 구체적으로 살펴보았다.

Table 2. The top 15 keywords by period

No	Phase 1		Phase 2		Phase 3	
	Keyword	(%)	Keyword	(%)	Keyword	(%)
1	China ^②	13.7	Confirmation ^①	11.0	People ^⑥	15.1
2	Infection ^①	9.5	Nose ^⑥	9.3	Confirmation ^①	10.9
3	Mask ^④	9.3	People ^⑥	8.2	Vaccine ^⑤	7.9
4	People ^⑥	7.5	Work	7.9	Occurrence ^①	7.5
5	Symptom ^①	7.2	Mask ^④	7.4	Mask ^④	7.3
6	Occurrence ^①	7.0	Support ^③	6.5	USA ^②	6.0
7	Confirmation ^①	6.7	Infection ^①	6.5	Development ^⑤	5.9
8	Pneumonia ^①	6.5	USA ^②	6.3	Today	5.7
9	Nose ^⑥	6.5	Occurrence ^①	6.2	Circumstance ^①	5.6
10	Patient	5.4	Information	5.9	Infection ^①	5.2
11	Prevention ^④	5.2	Circumstance ^①	5.8	Japan ^②	5.1
12	Present	4.3	Diffusion ^①	5.0	Remedy ^⑤	4.7
13	Information	3.8	China ^②	5.0	Information	4.6
14	Infection disease	3.7	World ^②	4.7	Korea ^②	4.3
15	Diffusion ^①	3.7	Area ^②	4.4	Symptom ^①	4.3

2.2.2 군집분석 결과

코로나19와 연관된 중요 키워드와 군집은 Table 3이다. 군집을 구성하는 키워드가 1개 또는 2개인 경우를 제외하면 총 6개의 군집으로 나타난다.

Table 3. The result of clustering analysis about COVID-19

Cluster	Keywords
Situation	Confirmation, Infection, Occurrence, Case, Symptom, Circumstance, Diffusion, Pneumonia, Inspection, Status
Nation	China, USA, Korea, World, Japan, Area, Domestic, Visiting
Government	Government, Response, Support, Management, Declaration, Situation, Society
Prevention	Mask, Health, Wearing, Prevention, Hospital
Treatment	Vaccine, Development, Treatment, Remedy
Infection Route	People, Nose, Hand

첫 번째 군집을 구성하는 키워드는 확진, 감염, 발생, 경우, 증상, 상황, 확산, 폐렴, 검사 그리고 현황이다. 코로나19는 평균 5-6일 잠복기 후 발열과 함께 상기도 감염 증상을 보이며 최대 잠복기는 14일이다. 대부분의 환자는 경증의 상기도 감염이나 중등증 이하의 폐렴 후 회복된다. 호흡관이나 저산소증은 중증 사례이며 호흡부전, 쇼크는 심각 사례이다[16]. 코로나19의 초기 증상과

징후 그리고 현황은 감염병으로부터 자신을 보호하기 위한 가장 기본적인 행동으로 볼 수 있다. 즉, 첫 번째 군집은 코로나19 바이러스에 대한 지식과 현재 현황을 대표하는 명칭은 ‘Situation’이다.

두 번째 군집을 구성하는 키워드는 중국, 미국, 한국, 세계, 일본, 지역, 국내 그리고 방문이다. 초기 코로나19는 중국 우한시 야생동물 판매 시장을 중심으로 사람과 사람 사이의 전파로 확산되었다[17]. 2020년 1월 30일을 기점으로 코로나19가 중국을 넘어 전 세계로 확산되었다. WHO는 국제공중보건위기상황을 선포하였으나 전 세계 확진자 수가 급증하여 3월 11일 팬데믹을 선포하였다. 미국은 코로나19의 높은 전염성을 간과하여 초기 진압에 실패하여 확산세가 급속히 진행되었고, 일본은 도쿄 올림픽이라는 국가적 행사와 소극적 대응으로 대규모 확산을 방지하였다. 중국은 봉쇄정책을 통해 확산세를 완화시키는 가시적인 성과를 보였다. 한국은 중국과의 왕래가 많은 국가임에도 불구하고 발 빠른 초기 대처로 방역 관리를 하였으나 지역감염이 급증하는 위기를 겪었다. 두 번째 군집은 코로나19가 발생된 공간을 설명하므로 대표하는 명칭은 ‘Nation’이다.

세 번째 군집을 구성하는 키워드는 정부, 대응, 지원, 관리, 신고, 사태 그리고 사회이다. 질병관리본부는 과거 메르스 대응의 경험을 바탕으로 코로나19도 초기부터 적극적으로 방역 조치하였다. 확진자의 역학조사 결과를 상세히 공개하였고, 보건의로 종사자들도 치료와 선별진료 등의 업무에 헌신적으로 협조하였다. 공공기관이나 대형 건물에서는 손 소독제를 비치하고 열감지기를 설치해 운영하고 있다. 코로나19의 국가적 대응체계는 과거에 비해 진일보하였으나 부처 간 협업, 소통, 지역 재난 지원 역량 강화는 아직 필요하다[18]. 감염병은 추가 감염 및 전파 등에 대한 불확실성이 높아 일상생활 및 사회경제 전반에 영향을 미치며, 지역사회사회 단위의 피해가 파생되므로 피해를 최소화하기 위해선 정부 주도의 재난대응이 필요하다. 세 번째 군집은 코로나19로 인한 피해를 줄이기 위한 정부 단위의 대응정책에 연관된 키워드이므로 대표하는 명칭은 ‘Government’이다.

네 번째 군집은 마스크, 건강, 착용, 예방 그리고 병원으로 구성되었다. 코로나19는 감염 초기에 전파되므로 개인이 주도하는 방역이 필요하다. 핵심적인 방역 행동은 사람들 간의 접촉을 최소화하는 것이다. 마스크는 침방울을 통한 감염 전파 차단에 도움이 되어 개인 방역의 필수 품이다. 코로나19 초기 마스크 수요가 공급보다 높아 공적 마스크를 제한적으로 구입하기도 했다. 이와 같은 개

인방역 대응에 관한 키워드로 구성된 네 번째 군집을 대표하는 명칭은 'Prevention'이다.

다섯 번째 군집은 백신, 개발, 치료 그리고 치료제 키워드로 구성되어 있다. 코로나19는 RNA바이러스로 인간에게 감염된 일곱 번째 코로나 바이러스이다. 코로나 바이러스의 변이는 스파이크 당단백질(spike glycoprotein)에 영향을 미쳐 변이를 일으키는 것으로 알려지고 있다. 바이러스의 변이는 백신과 치료제의 개발이 쉽지 않음을 의미한다. 그럼에도 불구하고 전 세계는 코로나19 사태를 해결하기 위한 백신 개발에 집중하고 있다. 국내 제약 회사는 코로나19 백신보다는 치료제 개발에 중심을 두고 연구 활동이 진행되고 있다[19]. 전 세계가 코로나19 연구에 집중된 결과 2020년 12월 중순 개발 중인 백신 후보는 162개로 이 중 52개가 임상시험 단계이다. 다섯 번째 군집은 코로나19 극복을 위한 치료에 연관된 키워드로 구성되어 있어 대표 명칭은 'Treatment'이다.

여섯 번째 군집을 구성하는 키워드는 사람, 코, 손이다. 코로나19는 비말감염과 접촉감염을 통해 전파되므로 사람의 감염경로에 해당하는 코, 손과 사람이 군집으로 파악되었다. 비말감염은 감염자의 체액이 기침 등으로 튀어나와 다른 사람의 입 또는 코로 들어가 감염된다. 일반적으로 기침 한 번에 약 3,000개의 비말이 2m 내로 분산되므로 마스크를 끼는 것이 좋다. 접촉감염은 감염환자의 호흡기에서 방출된 바이러스가 물체의 표면에 있다 다른 사람이 해당 물체를 만진 후 코, 입 또는 눈을 통해 몸에 스며들어 감염된다. 이 때문에 손 씻기가 중요하다. 여섯 번째 군집은 감염경로에 관련된 키워드로 구성되어 'Infection Route'가 대표 명칭이다.

2.2.3 코로나19 확산시기별 특징

확산시기별 군집에 포함된 단어의 빈도를 시각화하면 Fig 1이다. 모든 시기에서 공통적으로 높은 빈도는 'Situation'으로 네티즌들은 코로나19 관련 상황에 관심이 많다. 전조기(Phase 1)에서는 'Prevention'이 두 번째로 높았으며, 'Infection Route'와 'Nation'이 그 다음이었다. 이 시기에는 예방에 대한 관심이 조금 더 많다. 심각기(Phase 2)는 전조기에 비해 'Nation'의 빈도가 급격하게 높아졌으며, 'Infection Route'의 빈도도 다소 높아졌다. 전 세계적으로 코로나19가 확산되면서 국가 관련 단어의 빈도가 높아지면서 감염경로에 대한 빈도도 같이 높아지게 되었다. 그리고 빈도에 나타나지 않았던 'Government'의 빈도도 발생하였다. 안정기(Phase 3)는 'Nation'의 빈도는 조금 낮아졌지만, 'Treatment'의

빈도가 급격하게 높아졌다.

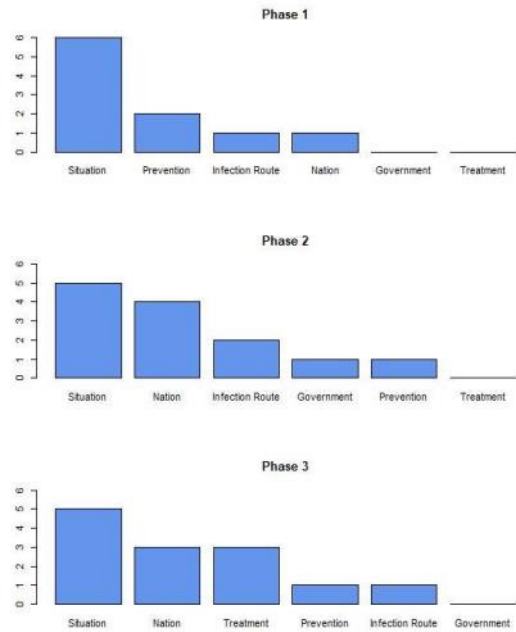


Fig. 1. Barplot of cluster by COVID-19 period

1) 코로나19 전조기(2020.01.20.~2020.02.17.)

중국 여행객으로부터 코로나19가 한국으로 유입되어 코로나19의 직접적인 영향을 받기 시작한 시기다. 1월 20일부터 2월 17일까지 코로나19 확진자 수는 소수로 서울, 수도권, 광주에서 간헐적으로 나왔다. 1월 27일 위기경보가 '주의'에서 '경계'로 격상되면서 중앙사고수습본부 출범하였고, 2월 18일 대구에서 확진자가 발생하기 전까지 보건용 마스크를 기부하는 등 대응에 여유가 있던 시기이다. 중국에서 처음으로 발생했기 때문에 '중국'이라는 단어의 빈도가 눈에 띄게 높았다. 상황 군집인 감염, 확진, 폐렴, 발생 등의 빈도가 높았고, 자신을 보호하기 위한 마스크도 많이 언급되었다. 뉴스의 토픽모델링에서도 국내 확진자 발생, 중국 전염병 확산, 정부 대응조치 강화, 신종 감염병 증상 및 예방, 세계 각국 확진자 발생으로 본 연구와 비슷하다[6].

2) 코로나19 심각기(2020.02.18.~2020.04.20.)

대구와 경북을 중심으로 확산이 시작된 심각기는 한국은 물론 세계 언론에서 언급될 정도로 코로나19의 위험성이 높았던 시기다. 매일 수백 명의 확진자가 발생하였고 사망자가 속출하여 대구 도시 전체에 공포의 그림자

가 드리워졌지만 전국에서 몰려든 의료종사자의 분투로 희망도 볼 수 있었다[20]. 2월 23일에는 위기경보가 ‘경제’에서 ‘심각’ 단계로 상향되면서 중앙재난안전대책본부의 본부장으로 국무총리가 임명되어 모든 중앙부처의 자원을 활용하여 코로나19의 확산을 막기 위한 방역, 치료, 지원 등의 효율을 높였다. 심각기 역시 ‘Situation’ 군집을 구성하는 키워드가 순위권을 차지하고 있다. 지역적 집단감염이 발생하여 ‘지역’이 언급되었다. 본격적인 확산이 지역에서 발생하여 감염경로에 높은 관심을 보였다. 정부에서 사회적 거리두기를 시행하면서 경제활동이 일시정지가 됨으로써 ‘일(work)’이라는 키워드의 사용 빈도가 높다. 이 시기에는 사회적 거리두기가 시행되었지만, 정부에서 개인의 1차 방역으로 권고하고 체계적으로 정립되어 관리되지는 않았다. 3월 9일에는 정부가 보건용 마스크의 원활한 공급을 위해 공적 마스크 제도를 시행하여 ‘마스크’의 빈도도 높다. 베트남, 타이완, 뉴질랜드와 같이 초기 봉쇄와 방역을 시행한 소수의 국가를 제외하고 전 세계적으로 코로나19가 확산되면서 ‘미국(USA)’, ‘중국(China)’, ‘세계(World)’가 언급되었다.

3) 코로나19 안정기(2020.04.21.~2020.08.15.)

일 확진자 수가 점차 감소추세를 보인 시기로 국민들이 일상생활로 복귀하며 코로나19의 극복을 위한 치료제 개발을 원하는 시기이다. 사회적 거리두기로 인한 영업제한, 재택근무 등이 완화되면서 5월 이후에는 ‘생활방역’에 들어가게 된다. ‘백신’, ‘치료제’, ‘개발’이 단어 빈도 상위에 들면서 코로나19의 위기의 완벽한 극복을 희망하고 있으나, 전염성이 높아 장기적으로 대응해야하므로 ‘마스크’, ‘확진’, ‘발생’, ‘감염’, ‘정보’가 여전히 높은 순위를 보인다. ‘드라이브 스루’의 빠른 검사, IT 인프라를 활용한 확진자 동선 추적 등이 효율적으로 수행되면서 한국의 코로나19 현황은 확진자 수, 사망자 수 등 대부분의 기준에서 최상위권으로 나타났다.

2.2.4 감정분석

SNS에 표출된 감정을 분석하면 Table 4이다. 감정 판별에 사용된 단어는 총 90,935개이고 시간이 지날수록 감정 단어 빈도수가 증가하는 추세를 보인다. 전조기에는 긍정과 부정의 비율이 각각 52.33%와 47.67%였다. 블로그와 카페에 표출된 SNS 감정은 심각기를 지나 안정기에는 긍정의 비율이 62.31%로 상승하였다. 이는 지역 확산으로 인해 아픔을 겪었지만 의료/공공기관의 빠른 대응과 공격적인 코로나19검사로 다른 나라에 비해

성공적으로 코로나19 확산을 막았기 때문으로 판단된다.

Table 4. The result of sentimental analysis

Period	Emotion	Frequency	Percent(%)
First half	Positive	14,488	52.33
	Negative	13,209	47.67
Serious	Positive	17,691	57.32
	Negative	13,175	42.68
Normalization	Positive	20,172	62.31
	Negative	12,200	37.69
Total	Positive	52,351	57.57
	Negative	38,584	42.43

2.2.5 광복절 이후 코로나19

코로나19로 인하여 전 국민이 힘든 시기를 거쳐 확진자 발생 수가 안정기로 접어들었으나 8월 15일 대구모집회를 통해 지역감염이 확산되면서 정부는 사회적 거리두기를 2단계로 상향하였다. 집회 이후 전국적으로 2주간 매일 200명의 확진자가 나왔다. 2단계로 격상된 사회적 거리두기는 국민의 노력으로 일일 확진자 수가 점차 감소하면서 10월 12일부로 1단계로 조정되었다. Fig 2에서 2월 18일 1차 집단감염 그리고 8월 15일 집회로 인한 2차 집단감염의 일 확진자 수 급증을 시각적으로 확인할 수 있다.

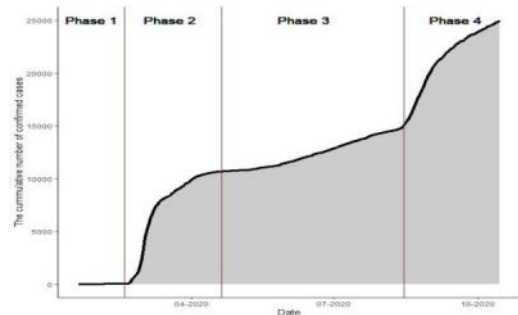


Fig. 2. The plot of daily confirmed cases in COVID-19

3. 결론

본 논문은 텍스트 마이닝을 이용하여 코로나19 확산 시기에 따른 SNS 키워드와 군집을 살펴보았다. 전조기에는 중국 단어와 상황 군집의 키워드가 많았고, 심각기는 지역감염에 따른 감염경로와 전 세계 대처방안에 관한 키워드가 많았고, 안정기에는 코로나19 극복을 위한 치료에 대한 관심이 많았다. 시기와 무관하게 SNS에서는

‘감염’, ‘마스크’, ‘사람’, ‘발생’, ‘확진’, ‘정보’가 언급되고 있었다.

최근 많은 연구에서 코로나 블루와 코로나 레드와 같은 우울감과 불안감 등을 제시하고 있다. 그러나, 본 연구에서는 코로나 팬데믹의 장기화로 인한 불안과 우울에 대한 키워드 검색 빈도는 높지 않았다. 대인관계에서 인간은 타인에게 자신을 포장하여 보이고자 하고[21], SNS는 개인의 생각, 감정, 일상을 과시적 또는 모범적으로 관리하여 대인관계를 형성한다. 카페와 블로그는 대인 관계를 안정적으로 유지하기 위한 도구이므로 코로나19 감염에 대한 두려움, 불안, 우울 등의 개인 감정보다 코로나19에 관한 객관적인 정보가 언급되고 있었다.

본 연구는 개인이 자유롭게 글을 쓰고 볼 수 있는 블로그와 분야별 관심사가 공유되는 커뮤니티인 카페에서 코로나19가 어떻게 표현되는지 확인하였지만, 이를 위해 연구범위를 제한한 한계점이 있다. 그러나 코로나19 확산 시기에 따라 사람들이 왜 이러한 키워드들에 관심을 가지게 되었는지에 대한 함의를 제공한 본 연구결과는 감염병 관련 교육자료의 구성 및 관련 보건의료 정책 마련을 위한 기초자료로 활용될 수 있다.

References

- [1] H. J. Lee, H. Y. Lee, "COVID-19 stress: Is the level of COVID-19 stress same for everybody? - Segmentation approach based on COVID-19 stress level", *KOREA LOGISTICS REVIEW*, Vol.30, No.4, pp.75-87, 2020. DOI: <https://dx.doi.org/10.17825/klr.2020.30.4.75>
- [2] B. Y. Lee, "Developments and future challenges of COVID-19", *Weekly Financial Brief*, Vol.29, No.11, pp.3-9, 2020.
- [3] Y. H. Kim, "From contact to uncontact: Marketing, beat the COVID-19!", *Excellence Marketing for Customer*, Vol.54, No.4, pp.9-17, 2020.
- [4] E. W. Lee, "Generation of COVID-19, Does mental health remain healthy?", *Issue & Analysis*, Vol.414, pp.1-25, 2020.
- [5] J. S. No, "How to overcome Corona blue?", *Local Finance Association*, Vol.2020, No.4, pp.124-127, 2020.
- [6] T. J. Kim, "COVID-19 News analysis using news big data : Focusing on topic modeling analysis", *The Journal of the Korea Contents Association*, Vol.20, No.5, pp.457-466, 2020. DOI: <https://doi.org/10.5392/JKCA.2020.20.05.457>
- [7] H. J. Ju, "Early childhood institute-related issues in the COVID-19 Pandemic using big data", *Journal of Learner-Centered Curriculum and Instruction*, Vol.20, No.22, pp.1085-1111, 2020. DOI: <http://dx.doi.org/10.22251/jlcci.2020.20.22.1085>
- [8] M. A. Oh, J. A. Jun, "Analysis of COVID-19's major issues based on social big data", *Issue & Focus*, Vol.376, pp.1-12, 2020.
- [9] D. Boyd, N. B. Ellison, "Social network sites: Definition, history, and scholarship", *Journal of Computer-Mediated Communication*, Vol.13, No.1, pp.210-230, 2007.
- [10] J. H. Ki, S. H. Ahn, "Application of sentiment analysis and topic modeling on rural solar PV issues : Comparison of news articles and blog posts", *Journal of Digital Convergence*, Vol.18, No.9, pp.17-27, 2020. DOI: <http://dx.doi.org/10.14400/JDC.2020.18.9.017>
- [11] B. K. Kim, M. K. Kim, "Keywords network analysis of temple stay using big-data", *Journal of Tourism Sciences*, Vol.39, No.4, pp.27-40, 2015. DOI: <http://dx.doi.org/10.17086/JTS.2015.39.5.27.40>
- [12] S. H. Hwang, D. Y. Ku, J. K. Kim, "Application of social big data analysis for cosmedical cosmetics marketing: H company case study", *Journal of Digital Convergence*, Vol.17, No.7, pp.35-41, 2019. DOI: <https://doi.org/10.14400/JDC.2019.17.7.035>
- [13] J. H. Koh, Y. K. Chong, "Analysis of Taiwanese food trends in Korea using social big data: Focusing on blog text analysis", *Journal of Tourism Management Research*, Vol.24, No.3, pp.71-91, 2020. DOI: <http://dx.doi.org/10.18604/tmro.2020.24.3.4>
- [14] Textom, Bigdata solution Textom manual version 3.0. Seoul: The IMC, 2018.
- [15] H. Christian, M. P. Agus, D. Suhartono, "Single document automatic text summarization using term frequency-inverse document frequency (TF-IDF)", *ComTech: Computer, Mathematics and Engineering Applications*, Vol.7, No.4, pp.285-294, 2016. DOI: <https://doi.org/10.21512/comtech.v7i4.3746>
- [16] J. Y. Heo, "Clinical and epidemiological characteristics of coronavirus disease 2019 in the early stage of outbreak", *Korean Journal of Medicine*, Vol.95, No.2, pp.67-73, 2020. DOI: <https://doi.org/10.3904/kim.2020.95.2.67>
- [17] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, B. Cao, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China", *The Lancet*, Vol.395, No.10223, pp.497-506, 2020. DOI: [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)
- [18] Y. K. Oh, H. S. Ru, J. Y. Heo, H. Y. Kim, D. H. Kim, "Infectious disease disaster response issues and policy implications based on COVID-19 infections cases", *Issue paper*, Vol.87, 2020.
- [19] H. J. Park, J. Y. Im, "Vaccine development trends and vaccine reserves for COVID-19", *Environmental and Resource Economics Review*, Vol.29, No.2, pp.273-292, 2020. DOI: <https://doi.org/10.15266/KEREA.2020.29.2.273>

- [20] S. C. Lee, "Analyzing early responses to COVID-19: Daegu case", *The Korean Journal of Public Administration*, Vol.29, No.3, pp.1-42, 2020.
DOI: <https://doi.org/10.22897/kipain.2020.29.3.001>
- [21] B. S. Chon, "Perceived homophily by the Degree of self disclosure among SNS users", *The Journal of the Korea Contents Association*, Vol.12, No.12, pp.170-178, 2012.
DOI: <http://dx.doi.org/10.5392/JKCA.2012.12.12.170>

윤 상 후(Sanghoo Yoon)

[정회원]



- 2011년 8월 : 전남대학교 통계학과 (이학박사)
- 2012년 12월 ~ 2013년 11월 : The university of Southampton 박사후 연구원
- 2016년 3월 ~ 현재 : 대구대학교 조교수

<관심분야>

시공간모형, 극단치 분포, 통계학습, 인공지능

정 상 윤(Sangyun Jung)

[준회원]



- 2016년 3월 ~ 현재 : 대구대학교 수리빅데이터학부 학생

<관심분야>

인공지능, 기계학습, 텍스트마이닝

김 영 아(Young A Kim)

[정회원]



- 2003년 2월 : 전남대학교 간호학과 (간호학사)
- 2013년 2월 : 전남대학교 간호대학 (간호학석사)
- 2016년 2월 : 전남대학교 간호대학 (간호학박사)
- 2016년 3월 ~ 현재 : 제주대학교 간호대학 부교수

<관심분야>

여성건강간호, 간호정책, 질적연구