

비정형 빅데이터를 이용한 COVID-19 주요 이슈 분석¹⁾

Analysis of Major COVID-19 Issues Using Unstructured Big Data

김진솔 (Jinsol Kim) 삼성서울병원²⁾
신동훈 (Donghoon Shin) 연세대학교 정보대학원³⁾
김희웅 (Heewoong Kim) 연세대학교 정보대학원⁴⁾

< 국문초록 >

2019년 12월 말, 전 세계를 혼란에 빠트린 코로나바이러스감염증-19(COVID-19)의 팬데믹이 시작되었다. 이러한 위기를 극복하고 피해를 최소화하기 위해 정부와 기관에서는 기존의 정책지원 효과를 극대화하고 변화하는 사회를 반영한 전방위적 대응책 마련이 필요하다. 사회적으로 부각되는 논제와 관심사항의 주제를 파악하기 위해, 본 연구는 소셜미디어의 빅데이터를 통해 코로나19와 관련된 주요 생각, 태도, 감정 등을 파악한다. 특히 정부의 대응에 관한 대중의 견해를 알기 위해 ‘정부 대응방향’을 기준으로 시기를 나누어 분석을 진행했다. 분석에 활용한 데이터는 네이버를 통해 2019년 12월 31일부터 2020년 12월 12일까지 수집되었다. 또한, 분석을 위해 텍스트마이닝 기법 중 TF-IDF 키워드 추출과 LDA 토픽모델링을 활용하였다. 그 결과, 8개의 코로나19 관련 주요 이슈가 도출되었으며, 이러한 이슈 사항과 주요 키워드를 기반으로 해당 분야에서의 코로나19 및 감염병 대응 정책 전략을 제시하였다. 본 연구는 코로나19 팬데믹과 같은 위기상황에 정부와 관련 기관이 국민의 필요와 요구에 따른 정확한 대응책을 마련하는 데 기초자료를 제공했다는 점에서 의의를 가진다.

주제어: 코로나19, 텍스트마이닝, 블로그 빅데이터, TF-IDF 키워드 추출, LDA 토픽모델링

1) 이 논문 또는 저서는 2018년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2018S1A3A2075114)

2) 제 1저자, jinsolkim719@gmail.com

3) 제 2저자, dh_shin@yonsei.ac.kr

4) 교신저자, kimhw@yonsei.ac.kr

1. 서론

2019년 12월 31일, 중국 후베이성 우한시에서 발생한 원인 미상의 폐렴이 세계보건기구(WHO)에 보고되면서 코로나바이러스감염증-19(이하 ‘코로나19’) 대유행이 시작되었다. 2020년 12월 31일을 기준으로 전 세계적으로 8,147만 명 이상의 코로나19 환자와 약 180만 명의 사망자가 나타나고, 지금도 그 수는 가파르게 증가하고 있다. 코로나19로 인한 경제타격, 달라진 일상 등으로 인한 피해를 최소화하고 팬데믹 위기를 극복하기 위해 정부와 기관에서는 기존의 정책지원 효과를 극대화하고 변화하는 사회를 반영한 전방위적이고 촘촘한 대응책 마련이 필요하다. 특히 코로나19 사태가 더욱 심화되거나 장기화될 경우를 대비하고, 추후 새로운 신종 감염병으로 인한 팬데믹 상황에 대응하기 위한 정부와 각계의 큰 관심이 촉구되는 바이다(옥철, 2020).

시의적절한 정부정책과 위기 대응 방안 및 지원책을 마련하기 위해서는 사회적으로 어떠한 논제가 부각되고 있으며, 사람들이 중요시하는 의제와 관심사항을 정확하게 파악할 필요가 있다. 이러한 감염병 확산 시기에 소셜미디어는 사회의 현상과 흐름을 파악할 수 있는 자료의 원천으로서 중요한 역할을 할 수 있다(홍순구 외, 2020). 특히, 이새미 외(2020)는 소셜미디어 플랫폼에 게시된 공개적으로 접근할 수 있는 빅데이터를 대중들의 주요 생각, 태도, 감정 등을 파악할 수 있는 주요 자료로 설명한다. 따라서, 사회적 이슈를 파악하기 위하여 온라인 상 비정형 빅데이터 분석을 기반으로 정책의제설정에 활용해야 함을 강조하고 있다.

현재까지의 코로나19 관련 국내문헌으로, 김기탁(2020)의 프로야구 관람문화 및 대중의 인식 변화를 파악하기 위한 키워드 분석 연구, 김상미(2020)의 온

라인 교육에 대한 이슈 및 주요 동향을 파악하기 위한 토픽 모델링 연구 등이 있다. 하지만, 정책적 시사점 도출을 위한 사회 전반적 범위 내에서의 국민들의 실질적인 의견을 분석한 연구가 부재한 상황이다. 또한, 정부의 대응에 대한 국민의 반응과 요구사항을 파악하기 위해 대응정책 변화에 따른 시기별 분석이 필요하다. 하지만 국내 선행 연구들은 소셜미디어의 검색량 또는 월별 기준으로 시기를 구분하여 분석이 진행되었다는 점에서 정책적 활용을 위한 시사점 도출에는 한계점이 존재한다.

따라서 본 연구의 목적은 소셜미디어의 비정형 빅데이터를 텍스트마이닝 분석방법을 통해 코로나19 관련 사회적으로 형성되고 있는 주요 이슈를 분석하여 사회적 및 경제적 정책개발을 위한 정책적 전략 방안을 제안하는 데에 있다. 이를 위해 본 연구는 코로나19 상황에서의 사회적 거리두기와 같은 정책 변화에 따라 6개의 시기로 구분하여 시기별 주요 이슈를 파악하고자 한다. 또한, 한국의 사회와 문화를 반영한 해석을 추가하여 한국 사회에 걸맞는 팬데믹 관련 대응방안의 기초자료를 제공한다.

2. 개념적 배경

2.1. 코로나바이러스감염증-19

전 세계 코로나19 확진자가 12만 명에 달한 2020년 3월 11일, 세계보건기구(WHO)는 코로나19에 대해 전염병 최고단계인 ‘팬데믹(Pandemic)’을 선포했다. 한국은 2월 23일 코로나19 사태와 관련해 정부의 위기경보 단계를 최고단계인 ‘심각’으로 격상하고 이에 대한 대응체계를 대폭 강화했다. 심각 단계에서는 정부에서 보다 적극적인 코로나19 확산 방지를 위해 해외

유입 차단 및 사회적 거리와 같은 강력한 대응조치를 추진한다(중앙사고수습본부, 2020).

코로나19의 세계적 대유행과 지역사회로의 확산은 사회적 수준에서 보건, 복지, 경제, 교육, 외교 등 사회 전반에 큰 영향을 미치고 있을 뿐만 아니라 개인의 일상에도 큰 변화를 불러오고 있다. 특히 코로나19의 확산 방지와 감염예방을 위해 전 세계에서 ‘사회적 거리두기’가 시행되고 있다. 한국에서는 코로나19의 확산 정도에 따라 사회적 거리두기를 생활방역(1단계), 지역유행(1.5단계, 2단계), 전국유행(2.5단계, 3단계)으로 크게 구분한다. 또한, 지역유행과 전국유행 단계를 보다 세분화해 총 5개 단계로 구분하여 시행하고 있으며 2020년 12월 31일 기준 사회적 거리두기 2.5단계를 시행 중에 있다. 이에 따라 불필요한 외출과 모임 자체가 권고되고 다중이용시설의 영업시간이 제한되었으며, 일상 및 사회·경제적 활동의 많은 부분에 제약이 생긴 상황이다. 또한, 사회적 거리두기가 일상화되며 비대면·비접촉의 언택트(untact) 시대가 열렸다. 이러한 언택트 시대에 사람들은 타인 그리고 사회와 소통하기 위해 온라인 매체를 활용하고 있다. 이에 따라 사람들이 타인과 소셜미디어를 통해 소통하는 양이 크게 늘었으며, 코로나19 발생 이후 SNS를 이용하는 비율이 45.7% 증가하고, 온라인 커뮤니티와 카페 등에서 활동하는 비율은 24.2%가 증가했다(KBS, 2020; 임현아 외, 2019).

이처럼 사람들은 언택트 시대에서의 커뮤니케이션을 위해 온라인 매체를 적극적으로 활용하고 있다. 온라인 매체를 통해 사회의 변화를 확인하고, 이에 대한 자신의 심경과 코로나19로 영향을 받은 개인의 일상을 일기형식으로 소셜미디어 블로그에 기록한다. 특히 한국의 소셜미디어 이용률은 87%로 세계 평균(49%)의 약 1.8배를 기록하고 있는 것으로 나타났다(DMC미디어, 2020). 따라서, 본 연구에서는 소셜미디

어 블로그의 비정형 빅데이터를 통해 코로나19와 관련된 사회적 주요 이슈를 분석하여 재난위기에 대응하기 위한 정책개발에 도움이 되는 기초자료를 제공하고자 한다.

2.2. 문헌 연구

코로나19의 첫 발생 직후부터 현재까지 코로나19 관련 논제 파악에 관한 많은 연구가 진행되고 있다. 이러한 연구는 연구초점에 따라 1) 소셜미디어를 이용해 코로나19와 관련한 대중의 견해와 의견을 분석한 연구, 2) 코로나19에 관한 언론보도 내용과 보도프레임을 분석한 연구, 3) 코로나19가 진행되며 시기별로 주요하게 형성되는 이슈를 분석한 연구로 나뉜다.

첫 번째 범주에 속한 연구는 코로나19가 여가생활에 미친 영향을 중심으로 분석하였다. Yu et al.(2020)은 코로나19 위기상황에 온라인 여행 플랫폼인 트립 어드바이저(TripAdvisor)에서 나타난 관광 커뮤니케이션의 주요 주제를 파악해 관광 위기관리를 위한 기초 자료를 제공하였다. 또한 김기탁(2020)은 포스트코로나 시대의 프로야구 관람문화를 예측하고 대비방안을 제시하기 위해 코로나19 이후 프로야구 팬들이 프로야구에 관해 공통적으로 인지하고 있는 사항을 조사하고, 팬들의 행동, 인식, 패턴 그리고 요구사항을 분석하였다.

두 번째 범주에 속한 연구는 뉴스미디어를 이용해 코로나19에 관해 보도된 뉴스내용과 보도프레임을 분석하고 있다. 예를 들어, Aslam et al.(2020)은 코로나19 관련 뉴스 헤드라인의 긍·부정 또는 중립의 감성결과를 도출하여 코로나19 상황에서 정보와 정서적 감정 사이의 상호작용을 분석하였다. 또한, 코로나19와 관련한 중국의 매체 보도 방향과 내용을 탐색해 코로나19 위기상황에서 미디어의 역할을 분석한 연구가 있

다(Liu et al., 2020).

세 번째 범주에 속한 연구로는 코로나19 사태의 진행상황을 시기별로 나누어 각 시기별 특징을 분석하는 것에 초점을 맞추고 있다. Damiano et al.(2020)은 위험인식이론에 기반하여 미국의 코로나19 발생 초기의 트위터 대화를 분석하여, 2월과 3월 시기에 따라 위협, 비난, 분노의 정도가 다르게 나타나는 것을 확인했다. Zhao et al.(2020)은 소셜미디어의 검색량에 따라 시기를 구분하여 중국에서 코로나19와 관련하여 사회적으로 나타난 주요 사건을 파악하는 연구를 진행했다. 특히 코로나19와 관련하여 대중의 주요 관심 변화를 파악하고 코로나19 사태가 진행됨에 따라 감정 수준이 부정에서 중립 그리고 긍정으로 변화하는 양상을 파악했다. 김상미(2020)는 코로나19 시대의 온라인 교육에 관해 보도된 국내 언론기사 키워드를 대상으로 3월부터 5월까지의 월별 교육현장의 주요 이슈와 시기별 키워드 변화 동향을 분석하였다.

기존의 코로나19 관련 이슈 분석에 관한 해외 연구를 살펴보면 다양한 소셜미디어 분석을 기반으로 하고 있다. 하지만, 국내문헌 중 코로나19에 관한 한국인의 실질적인 의견을 분석한 연구가 부재한 상황이다. 한국에서 발생하고 있는 코로나19 사태와 관련하여 한국의 상황에 적합한 정책과 대응책 개발을 위해서는 국민의 실질적인 의견을 분석한 연구가 필수적이다. 또한, 기존의 국내 연구에서는 시기의 구분에 있어 소셜미디어의 검색량 또는 일별, 월별 기준으로 시기를 나누어 분석을 진행했다. 그러나 정부의 코로나19의 대응에 대한 국민의 정확한 반응을 확인하고 구체적인 필요 및 요구사항을 파악하기 위해서는 정부의 대응정책 변화에 따른 시기로 구분하여 분석하는 것이 필요하다.

따라서, 본 연구에서는 코로나19 대응에 있어 정부의 정책변화에 따라 시기를 구분하여 시기별로 국민의

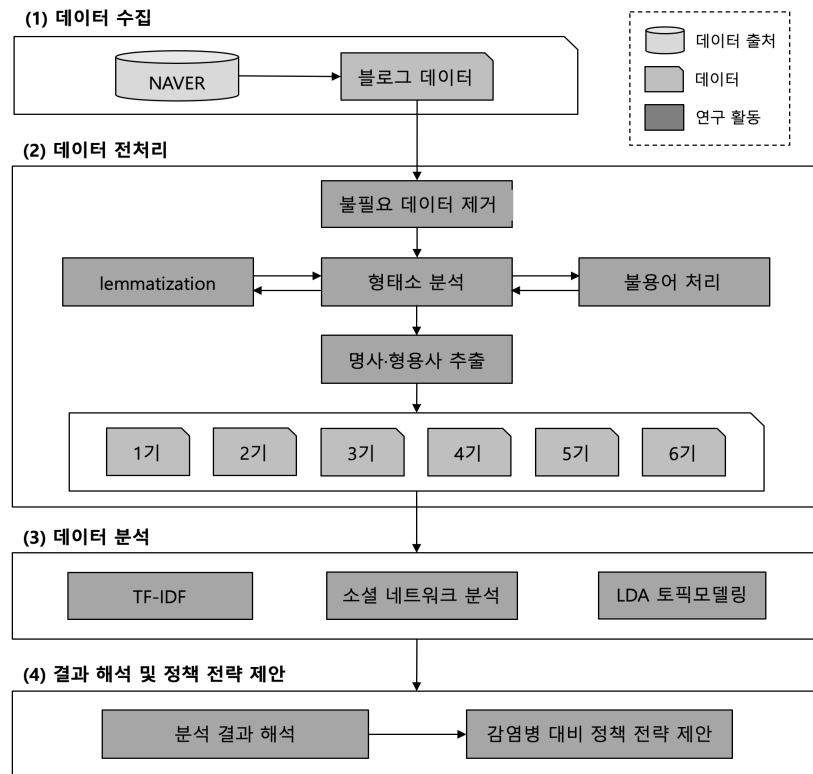
반응과 의견 및 생각을 분석한다. 이를 기반으로 코로나19 위기상황에서 정부와 관련 기관이 국민의 요구를 반영한 정확한 정책 및 대응책을 마련하고, 추후 감염병 위기상황을 체계적으로 대비하기 위한 정책개발에 실질적인 도움을 주는 기초자료를 제공한다.

3. 연구방법론

3.1. 연구절차

본 연구는 네이버 블로그 게시글을 대상으로 텍스트마이닝 분석을 통해 코로나19와 관련하여 사람들의 시기별 관심사항과 주요 토픽을 파악한다. 윤명희(2007)는 블로그와 같은 1인 커뮤니티를 개인중심의 네트워킹에 기반한 새로운 공동체의 구성으로 정의한 바 있다. 즉, 블로그는 개인에 기반한 새로운 사회관계를 형성하며 개인과 공동체의 모든 성격을 띠게 된다. 따라서 블로그는 개인의 지극히 주관적인 생각과 더불어 사회 속에서의 커뮤니케이션의 주제 및 방향을 알 수 있는 가장 적합한 수단이라고 할 수 있다. 본 연구에서 선정한 데이터 수집 대상인 포털사이트 네이버 블로그는 소셜미디어 점유율 중 25.7%의 비율을 차지하며 33.6%의 비율을 차지한 인스타그램 다음으로 2위를 기록하였다(오픈서베이, 2020). 인스타그램은 사진 업로드 중심 소셜미디어이며, 네이버 블로그는 글 작성 중심의 소셜미디어인 특성에 따라 코로나19와 관련된 언어적 정보를 활용하고자 네이버 블로그를 수집 대상으로 선정하였다.

본 연구의 연구절차는 <그림 1>과 같으며, 크게 (1) 데이터 수집, (2) 데이터 전처리, (3) 데이터 분석, (4) 결과 해석 및 정책 전략 제안 단계로 구분된다. 구체적으로, 코로나19에 관해 작성된 네이버 블로그 게시



〈그림 1〉 연구절차

물을 수집하여 전처리 작업을 진행하였다. 정제된 데이터는 기간에 따라 6개의 시기로 분류한 후, 시기별로 TF-IDF(Term Frequency-Inverse Document Frequency) 키워드 추출, 그리고 LDA(Latent Dirichlet Allocation) 토픽모델링 분석이 진행됐다. 이러한 분석 결과를 통합하여 해석 및 토의를 진행하고, 이를 바탕으로 코로나19와 관련하여 정부와 관련 기관을 위한 정책적 제안을 하였다.

3.2. 데이터 수집과 전처리

본 연구를 위한 분석 대상 데이터는 코로나19 관련 블로그 게시글 텍스트 데이터다. 분석기간은 코로나19가 처음 WHO에 보고된 2019년 12월 31일부터 2020년 12월 12일까지 총 348일이다. 전체 분석기간은 정부가 코로나19 확산정도에 따라 시행하는 정책인 ‘사회적 거리두기’단계와 실질적인 코로나19 확진자 발

〈표 1〉 분석 대상 데이터 규모

시기	기간(일수)	구분	데이터 규모(개)
제1기	2019/12/31-2020/1/19 (20)	발생 초기	282
제2기	2020/1/20-2020/5/5 (107)	1차 대유행	59,450
제3기	2020/5/6-2020/8/15 (102)	안정기	60,337
제4기	2020/8/16-2020/10/11 (57)	2차 대유행	60,908
제5기	2020/10/12-2020/11/18 (38)	감소기	60,445
제6기	2020/11/19-2020/12/12 (24)	3차 대유행	60,777
합계	2019/12/31-2020/12/12 (348)	전체 기간	302,199

생 수를 기준으로 여섯 개의 시기로 나누어 분석을 진행했다(<표 1> 참조).

제1기는 ‘코로나19 발생 초기’로 중국 우한에서 첫 코로나19 환자가 보고된 시점부터 국내 코로나19 발생 직전까지의 기간이다. 제2기는 ‘1차 대유행’이 나타난 시기로, 국내 첫 코로나19 환자가 발생한 시점부터 전국적으로 사회적 거리두기가 처음 시행된 2020년 5월 5일까지의 기간이다. 제3기는 확진자 수가 안정된 ‘안정기’로서 사회적 거리두기 1단계가 시행된 기간이다. 제4기는 ‘2차 대유행’이 발생한 시기로, 사회적 거리두기 2단계와 2.5단계가 시행된 기간이다. 제5기는 확진자 수가 감소한 ‘감소기’로, 사회적 거리두기의 단계가 1단계로 완화된 기간이다. 마지막 제6기는 확진자 수가 급격하게 증가한 ‘3차 대유행’이 나타난 시기로, 사회적 거리두기 단계가 다시 2단계와 2.5단계로 강화된 기간이다.

데이터 수집에 사용된 검색어는 ‘코로나19’, ‘신종 코로나바이러스’, ‘우한폐렴’으로 선정하였다. 정부가 신종 코로나바이러스의 한글 공식명칭을 ‘코로나19’로 명명한 2월 12일을 기준으로, 그 전 시기에 소셜미디어에서 코로나바이러스의 명칭으로 가장 많이 사용된 단어인 ‘신종 코로나바이러스’와 ‘우한폐렴’을 추가 검색어로 선정하였다. 이와 같은 기준을 기반으로 포털사이트 네이버 블로그의 게시글을 파이썬(Python)을 이용한 웹 크롤링(web crawling)을 활용하여 수집하였다. 수집한 데이터는 분석에 사용하기 전에 전처리 과정을 거쳤다. 전처리 과정은 다음과 같이 5가지 단계, 즉 (1) 코로나19와 관련 없는 데이터 제거, (2) 형태소 분석, (3) lemmatization, (4) 불용어 처리, (5) 명사 및 형용사 추출 순으로 진행됐다. 전처리를 거쳐 총 302,199개의 네이버 블로그 게시글이 분석에 활용되었다.

전처리 과정 중 첫째로, 분석에 사용될 데이터를 선

정하고 불필요한 데이터를 제거하였으며, 그 기준은 코로나19에 관해 개인의 의견이 작성된 블로그 데이터다. 즉, 수집한 데이터에서 코로나19와 관련이 없는 내용, 온라인 기사를 그대로 옮겨 놓은 게시글, 광고 및 홍보 또는 공지 성격을 띠는 데이터를 제거하였다. 둘째, 형태소 분석을 위해 KoNLPy 패키지의 Mecab 한글 형태소 분석기를 이용하였다. 셋째, 다른 형태를 가지고 있지만 같은 의미를 가진 단어들을 사전에 기반하여 기본형 단어로 복원시켜 단어의 어근을 찾는 lemmatization 작업을 진행했다. 예를 들어, ‘힘들어’와 ‘힘들다’를 ‘힘든’으로 통일시켜주는 작업이다. 넷째, 빈번하게 나타나지만 구체적인 의미를 갖지 않는 ‘것’, ‘이때’, ‘관련’, ‘정도’ 등의 단어를 불용어 처리하였다. 형태소 분석과 표제어 추출 그리고 불용어 처리는 반복해서 진행하며 데이터 전처리의 정확도를 높이고자 하였다. 마지막으로, 동사, 대명사, 수사, 관형사, 부사, 조사, 감탄사와 기타 문장부호를 제외하고, 유의미한 해석이 가능한 명사와 형용사를 추출하였다.

3.3. TF-IDF 키워드 추출

TF-IDF(term frequency-inverse document frequency)는 문서를 대표하는 주요 키워드를 추출하는 대표적인 기법이다. 이는 단순히 문서 내 단어의 빈도를 기반으로 키워드를 도출하는 단어빈도분석보다 더 정교한 계산방식을 사용한다. 즉, 단어빈도분석은 문서 내에 많이 등장하는 단어일수록 중요한 단어로 계산하는 방식이라면, TF-IDF 기법은 다른 문서에서는 많이 등장하지 않고 특정 문서 내에서 높은 비율로 등장할 때 그 단어를 해당 문서의 주요 키워드로 도출하는 방식이다. 특정 단어가 특정 문서와 강한 상관관계를 보일 때 그 단어의 TF-IDF 값이 높아지게 되고, 해당 문서의 주요 키워드라고 할 수 있다.

TF-IDF 값은 각 단어의 가중치를 의미하며, TF(term frequency)와 IDF(inverse document frequency) 수치의 곱을 통해 계산된다. 또한, 계산된 TF-IDF 값의 순위로 키워드의 중요도를 파악하여 키워드를 추출하게 된다. TF는 단어빈도분석과 같이 특정 문서에서 특정 단어의 등장빈도를 나타내며, IDF는 전체 문서의 수를 특정 단어가 담긴 문서의 수로 나눈 후 로그를 취한 값이다. 본 연구에서는 TF-IDF 가중치를 통해 코로나19 관련 주요 키워드를 파악하는 데 이용한다.

3.4. LDA 토픽모델링

토픽모델링(topic modeling)은 구조화되지 않은 문서의 텍스트에서 단어들의 패턴을 찾는 비지도학습 기반 계층적 확률모델 알고리즘으로, 단어들의 패턴에 따라 문서의 주제, 곧 토픽을 추론한다(Blei & Lafferty, 2006). 즉, 토픽모델링은 텍스트에 내재되어 있는 토픽들을 찾기 위해 문서를 토픽의 확률적 집합으로 표현하는 통계적 추론 기법이다(Blei et al., 2012; 이정현 외 2020).

토픽모델링 알고리즘 중 가장 단순한 기법으로 널리 사용되는 LDA(latent Dirichlet allocation)는 문서가 여러 개의 토픽으로 혼합되어 표현되고, 각 토픽이 단어의 분포로 구성된다는 개념에 기초한다. 구체적으로, LDA는 문서를 표현하는 잠재적인 토픽들과 단어들이 토픽에 포함될 가능성을 디히클레 분포(Dirichlet distribution)에 기반하여 추정하고, 문서 내에 잠재된 토픽의 단어들을 집합으로 추출하는 알고리즘이다(Blei et al., 2003).

LDA 토픽모델링에서 문서에 내재된 최적의 토픽 수를 결정할 때는 일반적으로 혼잡도 점수(perplexity score)와 일관성 점수(coherence score)를 활용한다. 혼잡도 점수는 특정 확률모델이 실제로 관측되는 값을

얼마나 잘 예측하는지를 나타낸다. 즉, 잠재된 토픽이 문서 내용을 얼마나 잘 예측하는가를 평가하는 지표로 그 값이 낮을수록 토픽이 문서를 잘 반영하고 있다고 할 수 있다. 또한, 일관성 점수는 주제의 일관된 정도를 측정한다. 즉, 토픽모델링이 잘 되어 있을수록 해당 토픽 내에는 의미론적으로 유사한 단어가 많이 모이게 되어 주제의 일관성을 높게 된다.

4. 데이터 분석 및 결과

4.1. TF-IDF 키워드 추출 결과

본 연구에서는 TF-IDF 키워드 추출을 통해 문서 내에서 핵심적인 비중을 차지하는 키워드를 분석하였다. TF-IDF 값이 가장 큰 상위 30개의 핵심 키워드들과 가중치 값을 시기별로 구분하여 <표 2>에 나타냈다.

각 시기별 주요 키워드 분석 결과, 제1기에는 전반적으로 중국·폐렴·감염·우한·발생과 같이 코로나19 발생에 관한 키워드가 많이 도출되었다. 이는 코로나19가 중국 우한에서 처음 발생한 후, 코로나19에 대한 사람들의 관심을 보여준다. 또한, 다른 시기에는 도출되지 않았지만 제1기에만 나타난 키워드로는 폐렴·신종·우한·발생·증상·검사·나라·병원·정보·활동·예방·확산·겨울·국내·모임이 있다. 이를 통해 제1기에는 코로나19의 국제적 발생 상황과 코로나19의 증상 및 예방 등 코로나19가 무엇인지에 대해 이해하고자 하는 현황이 나타나는 것을 볼 수 있다.

제2기에는 지역·대구·감염과 같은 국내 감염에 관련한 키워드가 다수 도출되었다. 이는 국내에서 코로나19 확진자가 처음 발생하고 1차 대유행이 진행되면서, 전국적으로 확산하고 있는 코로나19에 대한 상황에 대한 관심으로 해석할 수 있다. 또한, 제2기에 유일

하게 나타난 키워드로는 소독·대구·감사·정부가 있다. 이는 대구에서 발생한 신천지 교회 관련 대규모 집단감염과 연관된 키워드로 분석된다. 특히 ‘감사’ 키워드는 1차 대유행이 발생함에 따라 의료진의 헌신과 노고에 감사를 표하는 캠페인인 ‘덕분에 챌린지’가 대국민 응원 캠페인으로 발전하면서 처음 등장한 것으로 해석된다.

제3기에는 ‘교육’, ‘학교’, ‘온라인’과 같이 교육에

관한 키워드가 많이 나타났다. 이는 1차 대유행이 끝나고 안정기에 접어들면서 개학이 미뤄졌던 초·중·고등학교가 순차적 개학을 시작하면서 학생들의 교육 상황에 높은 관심이 생긴 것으로 해석된다. 특히 제3기에 유일하게 나타난 키워드로는 생활·학교·신청으로 나타났다. 이를 통해 학생들에게는 개학 연기 및 온라인수업과 같이 변화한 학교생활과 직장인들에게는 재택근무가 증가하며 바뀐 생활환경 등 많은 국민

〈표 2〉 TF-IDF 주요 키워드 추출 결과

제1기		제2기		제3기		제4기		제5기		제6기	
키워드	TF-IDF	키워드	TF-IDF	키워드	TF-IDF	키워드	TF-IDF	키워드	TF-IDF	키워드	TF-IDF
여행	11.31	마스크	1197.96	마스크	881.57	마스크	945.67	마스크	853.11	확진	1179.40
중국	11.21	확진	1093.28	아이	830.94	확진	932.39	확진	826.06	투자	913.57
페렴	7.92	아이	897.13	확진	691.10	아이	817.69	아이	738.09	단계	866.33
건강	7.30	중국	694.62	교육	675.23	카페	755.98	교육	720.76	가족	778.09
감염	7.02	소독	662.72	여행	665.33	추석	717.72	시장	689.06	백신	764.44
환자	6.71	카페	638.88	시장	654.81	사회	708.64	백신	659.50	아이	727.71
미국	6.68	마음	636.37	카페	635.61	방역	681.12	투자	654.56	방역	710.25
서울	6.51	건강	624.22	지원	628.28	거리	658.02	미국	648.41	사회	707.02
한국	6.47	지역	624.19	수업	619.49	마음	642.95	사회	632.40	마스크	701.46
신증	6.32	상황	592.05	지역	611.15	수업	625.30	기업	629.07	시장	691.94
세계	5.79	운동	581.15	마음	593.16	지원	608.95	여행	616.52	거리	689.93
우한	5.69	한국	580.43	투자	572.77	시장	602.70	지역	611.15	주가	671.82
발생	5.64	사회	574.10	기업	567.03	건강	598.24	카페	608.92	입장	670.24
증상	5.34	대구	567.24	건강	562.44	여행	581.71	방역	604.89	카페	658.77
검사	5.22	방역	542.37	제품	561.16	교육	581.19	지원	593.88	마음	637.30
나라	4.87	지원	542.32	미국	558.90	운동	570.49	제품	580.64	교육	621.84
병원	4.75	감사	534.44	사회	557.76	상황	568.57	건강	579.59	상황	613.48
마스크	4.52	수업	530.70	한국	554.12	단계	559.94	한국	577.95	기업	607.04
정보	4.24	환자	530.69	상황	552.87	서울	555.52	관리	575.46	지역	596.51
활동	4.23	세계	524.89	경제	552.02	온라인	552.79	사업	574.07	건강	594.06
예방	4.18	시장	519.81	운동	540.48	친구	544.35	마음	566.89	제품	572.66
경제	4.05	감염	518.31	친구	532.40	지역	540.57	온라인	556.46	수업	571.49
확산	4.01	미국	515.53	생활	530.89	제품	531.40	단계	554.66	거리 두기	560.33
겨울	3.98	여행	509.89	방역	527.16	기업	512.01	거리	546.68	한국	559.23
거리	3.93	거리	505.58	사업	519.68	거리 두기	509.46	상황	543.45	미국	542.94
국내	3.86	친구	505.08	학교	519.44	한국	509.13	경제	532.06	친구	536.23
상황	3.83	정부	491.23	온라인	516.05	관리	504.35	서울	522.54	관리	532.21
모임	3.73	제품	488.70	관리	514.46	투자	499.31	수업	511.03	서울	531.00
해외	3.71	경제	485.05	신청	502.95	가족	496.26	상승	510.97	사업	525.37
확진	3.41	신증	468.12	세계	496.09	확산	492.41	활동	506.05	여행	523.11

에게 나타난 생활방식의 변화가 사회적 이슈가 되고 있는 것으로 해석할 수 있다.

제4기에는 단계·거리·사회적 거리두기와 같이 사회적 거리두기 시행에 관련한 키워드가 도출되었다. 이로부터 2차 대유행이 진행되며 격상된 사회적 거리두기 2단계 및 2.5단계 조치와 강화되는 방역수칙에 대한 큰 관심을 알 수 있다. 특히 제4기에 유일하게 나타난 키워드는 추석으로 나타났으며, 이는 정부가 9월 29일부터 10월 11일까지의 기간을 ‘추석 특별방역 기간’으로 지정해 거리두기 등 방역을 강화함에 따라 추석 기간의 사회적 거리두기가 사회적 이슈로 부각된 것으로 해석된다.

제5기에는 백신·투자·상승과 같이 백신 관련 주식에 대한 투자와 연관된 키워드가 다수 나타났으며, 제5기에 유일하게 나타난 키워드는 상승으로 분석됐다. 이는 글로벌 제약사들의 코로나19 백신 개발에 속도가 붙으면서 코로나19 백신과 관련한 주식이 투자 종목으로 인기가 크게 상승한 사회적 현상과 연관이 있는 것으로 해석할 수 있다.

마지막으로 제6기에는 확진·투자·단계·백신·방역·주가·상황·사회적 거리두기 등 사회적 거리두기와 백신 관련 주식에 대한 투자를 나타내는 키워드가 도출되었다. 제6기는 확진자 수가 급격하게 증가한 시기로 3차 대유행이 나타나 사회적 거리두기 2단계와 2.5단계가 시행되었다. 한편, 여러 제약사의 백신 개발이 성공하면서 세계적으로 백신 접종이 시작된 시기이기도 하다. 이에 따라 사회적 거리두기 방침과 코로나19 백신 관련 주식에 대한 투자에 이목이 집중된 것으로 해석할 수 있다.

4.2. LDA 토픽모델링 결과

제1기부터 제6기까지 시기별로 LDA 토픽모델링을

진행하여 코로나19와 관련하여 시간의 흐름에 따라 나타나는 토픽의 변화 양상을 살펴보았다. 토픽수를 객관적으로 선정하기 위해 토픽의 개수를 2개에서 10개까지 모형화하여 나타나는 혼잡도 및 일관성 점수를 기준으로, 제1기의 토픽수는 2개, 제2기부터 제6기까지의 토픽수는 8개가 가장 적합할 것으로 판단하였다. 각 토픽의 토픽명과 토픽에 대한 해석은 추출된 키워드들의 의미와 키워드 간의 관계 및 키워드가 포함된 원본 데이터 내용을 고려하여 3명의 연구자가 독립적으로 분석한 후 최종합의를 통해 결정되었다.

제1기 LDA 토픽모델링 분석으로 도출된 두 가지의 토픽 중 첫째는 ‘코로나19 개념 이해’는 핵심 키워드로 중국·폐렴·감염·증상 등으로, 코로나19가 무엇인지를 이해하려는 모습을 나타낸다. 둘째, ‘코로나19 상황 파악’은 핵심 키워드로 여행·중국·환자·감염 등이 도출됨으로써 코로나19에 관한 상황을 파악하고 코로나 사태를 파악하고자 하는 것을 나타낸다. 제2기부터 제6기까지의 LDA 토픽모델링 분석 결과로 여덟개의 토픽을 도출하였으며, 2기의 토픽들은 다음과 같다. 첫째, ‘코로나19 상황 파악’은 핵심 키워드로 확진·중국·감염 등이 도출됨으로써 코로나19가 세계적으로 확산하는 현황과 국내 코로나19 확산 상황에 대한 관심으로 해석할 수 있다. 둘째, ‘정신건강’은 핵심 키워드로 마음·이야기·친구·걱정 등이 도출됨으로써 코로나블루와 같은 정신건강에 대한 우려를 의미한다. 셋째, ‘식사문화’는 핵심 키워드로 카페·주문·식당·방문 등이 도출됨으로써 코로나19 감염 위험인식이 높아짐에 따라 외식에 대한 우려와 변화하는 식사문화를 나타낸다. 여기서 변화하는 식사문화는 코로나19 발생 이후, 사회적 거리두기 단계가 격상될수록 카페와 식당의 영업시간이 제한되고 배달음식의 주문 빈도가 크게 증가하는 현상을 의미한다. 넷째, ‘제한된 야외 및 스포츠 활동’은 핵심 키워드로 운동·거리·

공원·마스크 등이 도출됨으로써 코로나19 감염 위험 인식이 높아짐에 따라 외식에 대한 우려가 커진 현상으로 해석된다. 다섯째, ‘마스크 대란’은 핵심 키워드로 마스크·구매·가격·사이즈 등이 도출됨으로써 코로나19 감염 예방의 필수 품으로 인식된 마스크 구매의 어려움을 의미한다. 마스크 수요를 공급이 따라가지 못해 생긴 마스크 대란 상황 마스크를 구매하는 데 큰 어려움이 발생하면서 마스크에 대한 이슈가 크게 나타나게 된 것으로 해석된다. 여섯째, ‘긴급재난지원금’은 핵심 키워드로 지원·신청·지역·재난 등이 도출됨으로써 코로나19로 어려움을 겪는 국민을 위해 정부가 지급하기 시작한 긴급재난지원금에 대한 높은 관심을 나타낸다. 일곱째, ‘코로나19 예방’은 핵심 키워드로 건강·소독·마스크·면역력 등이 도출됨으로써 코로나19를 예방하기 위한 노력을 의미한다. 예를 들어, 코로나19 감염에 대비해 몸의 면역력을 증진시키기 위해 영양제를 복용하고 손씻기와 손소독제 사용 등 위생을 준수하는 행동이다. 여덟째, ‘교육’은 핵심 키워드로 수업·학원·학교·시험 등이 도출됨으로써 코로나19로 변화한 교육환경을 의미한다.

제3기의 분석 결과, ‘여름휴가’, ‘글로벌 경제상황’, ‘긴급재난지원금 사용’ 등의 새로운 토픽들이 등장하였다. ‘여름휴가’는 핵심 키워드로 여행·영화·제주·가족 등이 도출됨으로써 코로나19로 인해 변화한 여름휴가 풍경을 나타낸다. 또한, ‘글로벌 경제상황’은 핵심 키워드로 기업·경제·미국·정책 등이 도출됨으로써 코로나19로 인해 발생한 전 세계적 경제 침체를 의미한다. 이에 따라, ‘주식시장’ 또한 새로운 주제로 등장하며 코로나19로 인해 변화한 주식시장 및 코로나19 치료제와 백신 개발의 영향으로 바이오와 제약 산업의 주식이 폭등한 상황을 반영한다. ‘긴급재난지원금 사용(외식)’은 핵심 키워드로 카드·지원금·재난·식당 등이 도출되었으며 재난지원정책이 시행됨에 따

라 사용처로 외식업계가 높은 비율로 나타났음을 의미한다.

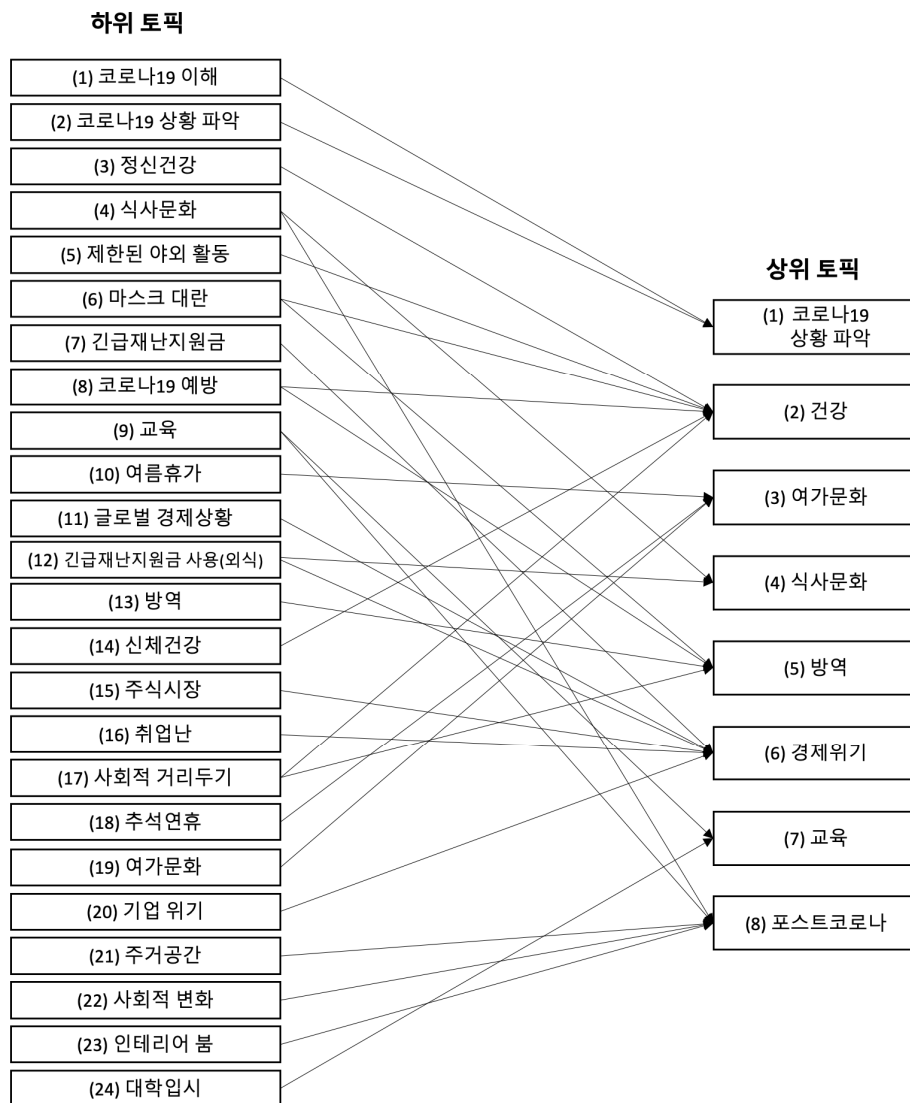
제4기의 분석 결과, 가장 높은 비중의 주제는 ‘여름휴가’로 제3기에 이어 지속적으로 관심을 갖고 있음을 파악할 수 있다. 이어서 새롭게 등장한 토픽인 ‘취업난’이 두번째로 높은 비중을 보였으며 핵심 키워드로 온라인·면접·청년·취업 등이 도출되었다. 지속적으로 악화되는 국내 경제 속에서 좁아진 채용시장과 이로 인한 스트레스를 의미한다. 특히 제4기에는 2차 대유행이 발생하며 사회적 거리두기가 2단계 및 2.5단계로 격상되고 사회의 많은 부분이 달라지면서 이에 대해 이목이 집중된 것으로 볼 수 있다. 대표적으로 ‘식사문화’, ‘사회적 거리두기’ 토픽이 새롭게 등장한 것을 예로 들 수 있다. 특히, 주문·포장·배달·소독 등의 키워드를 통해 외식에 대한 우려 및 변화하는 식사문화에 대한 관심이 높아진 것을 파악할 수 있다.

제5기의 분석 결과, ‘교육’ 토픽이 높은 비중을 차지하며 교육·학교·온라인·수업 등의 키워드를 바탕으로 학교의 등교 여부와 수능과 같은 시험에 대한 우려가 증가함을 의미한다. 5기에는 건강, 교육 등 다양한 분야에서의 걱정과 우려와 동시에 ‘여가문화’, ‘주거공간’과 같은 토픽의 등장을 통해 코로나19 사태 속에서 새로운 문화가 발생되는 것을 파악할 수 있다. 또한, 넷플릭스·인테리어 등의 키워드가 도출되었으며 집에 머무는 시간이 길어지는 등 사람들의 행동양식이 변화하며 실내 여가 및 주거공간에 대해 높아진 관심을 나타낸다. 제6기의 분석 결과를 통해 ‘식사문화’와 ‘여가문화’와 같이 문화에 대한 관심과 ‘사회적 변화’와 ‘주식시장’과 같이 급변하는 경제에 대한 관심이 지속적으로 높다는 것을 파악할 수 있다. 특히 6기는 코로나19 3차 대유행이 발생하고 사회적 거리두기가 강력하게 제재되면서 ‘사회적 거리두기’ 토픽이 4기에 이어 다시 높은 비중을 차지함을 알 수 있다.

코로나19 관련 블로그 텍스트 데이터를 시기별로 LDA 토픽모델링하여 분석한 결과를 종합적으로 해석하기 위해 시기별로 중복된 토픽을 제외하고 총 24개의 토픽을 도출하였다. 이후, 24개의 하위토픽들은 (1) 코로나19 상황 파악, (2) 건강, (3) 여가문화, (4) 식사문화, (5) 방역, (6) 경제위기, (7) 교육, (8) 포스트코로나, 총 8개의 상위토픽으로 분류되었으며, 상위토픽 분류에 대한 연구자 3명의 평가자 간 신뢰도(inter-rater reliability)는 0.91로 나타났다. 또한, ‘미국 대선’과 ‘부

동산시장’의 경우 코로나19 관련짓기 어려운 독립적인 토픽으로 해석되어 특정 영역으로 분류하지 않았으며 구체적인 분류 결과는 <그림 2>와 같다.

이를 기반으로 상위토픽에 해당하는 주제가 언급된 시기를 분석하여 <표 3>에 표기하였다. 상위토픽에 해당하는 하위토픽들의 비중을 합산한 수치가 높을수록 음영이 짙어지도록 표시하였다. 이를 통해 시기별로 상위토픽의 비중을 비교하며 주요 논의사항을 파악하였으며 해석은 다음과 같다. 첫째, ‘코로나19 발



<그림 2> LDA 토픽모델링 결과 토픽 분류

〈표 3〉 LDA 토픽모델링 결과 종합 요약

No.	토픽명	시기						해석
		1기	2기	3기	4기	5기	6기	
1	코로나19 발생 현황							코로나19 발생 현황 정보 파악
2	건강							감염 예방을 위한 신체건강 및 코로나19의 영향으로 나타나는 정신건강 스트레스 관리
3	여가문화							코로나19 이후 휴가 및 여가생활의 변화
4	식사문화							코로나19 이후 식사문화의 변화
5	방역							코로나19 예방을 위한 정부 차원의 방역조치 및 개인 차원의 생활방역
6	경제위기							코로나19가 변화시킨 경제상황
7	교육							코로나19 이후 교육현장의 변화
8	포스트코로나							코로나19가 변화시킨 사회모습, 그 이후의 삶

생 현황’은 코로나19의 발생 현황에 대한 정보를 파악하려는 모습을 의미한다. 해당 토픽은 제1기, 제2기, 제3기, 즉 코로나19가 발생한 초기 기간에 집중적으로 나타나고 있다. 이는 위협의 지각(risk perception)에 있어 다양한 위협 특성 중 두려움(dread)와 낯섦(newness)이 가장 두드러진다는 사실(Slovic, 1987)로 해석할 수 있다. 즉, 코로나19라는 신종 바이러스가 세상에 모습을 드러내면서 사람들이 이러한 새로운 위협에 민감하게 반응하고 제4기 이후로 시간이 흐르면서 익숙한 위협에 보다 둔감해지는 것이다.

둘째, ‘건강’은 코로나19 바이러스로부터의 감염 예방을 위한 신체건강 및 코로나19의 영향으로 나타나는 정신건강 스트레스를 관리하고자 하는 높은 관심을 의미한다. ‘건강’ 토픽은 제2기에서 특히 높은 비중을 보이며 제6기까지 꾸준히 나타나고 있는 주제로 신체 및 정신건강 측면에서 많은 논의가 이루어지고 있는 분야다. 코로나19가 장기화되면서 방역 및 의료적 대응과 함께 개인의 수준에서 건강을 지키기 위한 노력이 다방면에서 나타나고 있다. 특히 사회적 거리두기가 지속되면서 신체활동이 감소해 비만발생의 우려가 높아지고, 사회활동 감소와 감염에 따른 불안감으로 우울감과 무기력감이 증가하고 있다(중앙재난안

전대책본부, 2020). 또한, 서울시(2020)의 조사에 의하면 코로나19로 인해 신체건강이 나빠졌다고 응답한 비율은 25%, 정신건강이 나빠졌다고 응답한 비율은 40%로 신체건강보다 정신건강 문제가 더욱 심각한 것으로 나타났다. 이처럼 코로나19가 사람들의 신체 및 정신건강에 지속적인 악영향을 미치고 있으며, 이에 따라 코로나19로부터 건강을 지키고자 하는 관심이 높게 나타나고 있는 것으로 분석된다.

셋째, ‘여가문화’는 코로나19 발생 이후 변화한 휴가와 여가생활을 의미하며, 여름 휴가철인 제3기와 제4기에 특히 높은 비중을 보인다. 서울시(2020)의 조사에 의하면 응답자의 32%가 코로나19로 인해 느끼는 어려움 중 가장 힘든 점으로 ‘여가활동이나 여행’이라고 응답한 바 있다. 더불어 코로나19 사태로 개인의 여가시간은 늘었지만, 여가공간 범위는 집 주변으로 제한되었으며, 비대면으로 혼자 보내는 여가활동이 대폭 증가했다(문화체육관광부, 2020). 즉, 여가활동과 여행의 제약이 정신건강 악화와 직결되어 나타나고 있는 것이다. 결론적으로, 코로나19가 장기화되면서 여가활동과 여행 그리고 사회적 교류의 제한으로 사람들이 고립감과 단절 등 어려움을 겪고 있으며 이에 대한 논의가 꾸준히 이루어지고 있다고 할 수 있다.

넷째, ‘식사문화’는 코로나19 이후 변화한 식사문화를 의미한다. 코로나19가 장기화됨에 따라 식사문화에 변화가 나타나고 있으며, 시기가 지날수록 해당 토픽의 비중이 증가하는 추세를 보인다. 여럿이 모여 함께 먹는 모습은 많이 사라지고, 개인별 음식 먹기 문화로 바뀌고 있으며 배달 활용이 급증하고 있다. 통계청(2020)에 따르면 2020년 8월 온라인 음식배달 서비스의 거래액이 전년 동기 대비 83% 높아졌다. 방역수칙에 따라 음식점과 카페 등의 이용이 제한되면서 음식을 포장하고 배달시켜 먹는 인구가 급증하게 된 것이다. 이처럼 위생과 방역이 중요한 코로나19 시대인 만큼 변화하는 식사문화에 많은 관심이 나타나고 있다.

다섯째, ‘방역’은 코로나19 예방을 위한 정부 차원의 방역조치와 개인 차원의 생활방역을 의미한다. 코로나19 확산 정도에 따라 사회적 거리두기 단계가 조정되고, 이에 따라 방역조치가 결정된다. 현재 정부에서는 감염 확산 상황에 따라 한 주에서 두 주 정도의 기간을 기준으로 사회적 거리두기 단계를 조정하고 있다. 이렇게 사회적 거리두기 단계가 조정되면서 자주 변화하는 방역조치에 대해 높은 관심이 나타나고 있다. 특히 일부 영역의 방역조치에 있어 세부적인 기준이 정확히 확립되지 않은 경우에는 혼란이 나타나며 이에 대한 논란과 논의가 집중되는 현상이 나타나고 있다. 또한, 개인 수준에서 방역수칙을 준수하여 코로나19 감염으로부터 본인과 가족을 보호하고자 하는 노력과 관심이 크게 나타나고 있다. 국민건강보험(2020)에 의하면 코로나19 발생 이후 감기, 폐렴 등 호흡기 감염 환자 수가 전년 대비 51.9% 감소했다. 이는 마스크 착용과 손씻기 실천 등 개인 수준에서 생활방역이 잘 지켜진 결과로 해석되고 있다. 이처럼 정부의 방역조치와 개인의 생활방역과 관련한 이슈가 많이 논의되고 있다.

여섯째, ‘경제위기’는 코로나19가 변화시킨 경제상

황을 의미한다. ‘경제위기’ 토픽은 제2기부터 제6기까지 공통적으로 도출되면서 경제에 대한 사람들의 높은 관심을 보여주고 있다. 코로나19의 확산은 공급과 수요의 양 측면에서 경제 성장을 위축시키고 있다(박성욱, 2020; 도보람 외, 2021). 공급 측면에서는 감염병 발생 국가의 생산활동이 차질을 겪고 있으며, 이러한 생산차질은 글로벌 공급망을 훼손시켜 감염병 확산이 심하지 않은 교역상대국의 공급마저 위축시킨다(박성욱, 2020). 이러한 상황에서 기업과 자영업자는 공급망의 훼손으로 위기를 겪게 되며, 이는 고스란히 개인의 경제활동에 큰 타격으로 이어진다. 더불어 코로나19 사태가 장기화되면서 많은 기업에서 고용 유지에 어려움을 겪고 있으며, 이에 따라 실제 인원을 감축하거나 채용을 취소하는 기업이 증가하고 있다. 결론적으로, 코로나19로 인한 사회의 경제 위축과 고용충격 등으로 개인의 경제적 삶에 큰 불확실성을 불러일으키고 있는 만큼, 경제에 대한 관심이 집중되고 있다.

일곱째, ‘교육’은 코로나19 발생 이후 변화하는 교육환경을 의미한다. 전반적으로 높지 않은 토픽 비중을 차지하고 있지만, 대부분의 시기에 걸쳐 꾸준히 이슈가 되고 있는 토픽으로 볼 수 있다. 코로나19가 발생하면서 초·중·고등학교 및 대학교의 개학이 연기되고 사상 최초로 온라인 개학이 실시됐다. 공교육뿐 아니라 사교육시장도 큰 영향을 받고 있는데, 경기도교육연구원(2020)의 조사에 따르면, 코로나19 이후 사교육 시간이 27.1% 증가한 것으로 나타났다. 이는 온라인 등교에 한계를 느끼며 사교육에 의존하는 비율이 증가한 것으로 해석된다.

여덟째, ‘포스트코로나’는 코로나19가 변화시킨 사회의 모습과 코로나19 이후의 삶을 의미한다. 특히 제5기와 제6기에서 도출된 토픽인 ‘인테리어 붐’과 ‘주거공간’이 이러한 포스트코로나의 대표적인 사례로, 해당 시기에 많은 관심을 보이고 있음을 표를 통해 확

인하 수 있다. 코로나19의 확산으로 많은 사람들이 재택근무와 온라인 등교 등으로 집에서 생활하는 시간이 길어지면서, 집이 단순히 주거공간을 넘어 휴식·문화·레저를 즐기는 공간으로 확대되고 있다. 이에 따라 집 내부에 일상공간을 꾸미는 인테리어 붐이 나타나고 있다. 예를 들어, 재택근무를 편하게 하기 위해 홈오피스를 구현하고, 방역조치로 카페를 가지 못하는 상황인 만큼 집에 홈카페 인테리어를 하는 것이다. 더불어 코로나19로 인해 재택근무를 도입한 많은 기업들이 이에 대한 효율성을 경험하면서, 코로나19 이후에도 재택근무를 채택하고자 하는 기업이 증가하고 있다. 이에 따라 사람들은 주거공간에 대한 중요성을 더욱 크게 인식하게 된 것으로 해석된다.

5. 토의 및 시사점

5.1. 연구 결과 토의 및 제안

본 연구에서는 소셜미디어에서의 코로나19 팬데믹 관련 이슈들을 도출하고 분석하여 종합적으로 해석한 후 코로나19와 같은 감염병 위기상황에서 실질적으로 활용될 수 있는 정책 전략을 제안한다. 첫째, 코로나19로 인해 변화한 교육현장에 대한 지속적인 높은 관심이 나타나고 있다. 특히 최근 코로나19 관련 해외 연구에서 또한 불안한 온라인 접속과 수업 내용 부실에 대한 불만 등 공교육의 온라인 수업 문제점이 제기되고 있다(M. P. Murphy, 2020). 또한, 온라인 등교 및 수업에 대한 한계를 느끼며 온라인 사교육에 의존하는 비율이 증가하는 결과가 나타나고 있다. 이러한 온라인 사교육 의존율의 증가는 사회적으로 교육격차로 이어져 교육양극화가 심해지는 현상이 나타날 수 있다(L. W. Van, Z. Parolin., 2020). 따라서 정부에서는 공

교육의 온라인 수업 인프라를 확충하고, 콘텐츠를 충실하게 준비할 수 있도록 정부의 지원을 강화해야 할 것이다.

둘째, 감염예방과 신체건강에 대한 관심이 높게 나타나고 있다. 코로나19로 인해 전국적으로 사회적 거리두기가 시행되면서 외부활동이 현저히 줄고 있다. 이는 신체활동 부족을 일으켜 결국에는 면역력 감소를 나타나게 한다. 코로나19와 같은 감염병을 예방하는 데 면역력의 중요성이 매우 높으며 대중 또한 이를 강하게 인식하고 있는 것으로 나타났다. 특히, ‘면역력’과 관련하여 영양제·식품·섭취가 연관 키워드로 나타나면서 면역력을 키우고자 하는 다각도의 관심이 나타났다. 이와 동시에 전반적인 운동 문화에도 큰 변화가 이루어지고 있기 때문에 정부와 관련 기관에서는 국민의 감염 예방과 신체건강에 대해 관심을 집중할 필요가 있다(권오정, 2020). 구체적으로, 개인의 차원에서 면역력을 키울 수 있는 여러 방안들을 구상하여 홍보할 필요가 있다. 예를 들어, 사회적 거리를 유지하고 방역수칙을 준수하면서도 실천할 수 있는 ‘걷기’ 운동을 장려하거나, 자가격리 중인 집에서도 쉽게 할 수 있는 홈트레이닝과 같은 운동에 대한 가이드라인을 마련하고 홍보할 수 있을 것이다. 결론적으로 정부는 국민의 신체 활동량 증진과 더불어 면역력을 높이는 데 초점을 맞추어 정책을 수립할 것이 권장된다.

셋째, 감염병에 대한 신체건강에 관한 우려뿐 아니라 정신건강에 관한 관심이 고조되고 있다. 코로나19가 불러온 재택근무 및 온라인 개학, 사회활동 감소, 고용불안정, 경제적 어려움, 부정확한 정보로 인한 혼란으로 정신적 스트레스가 가중되고 있다. 또한 코로나19로 인해 제한되는 여가생활로 스트레스 해소의 어려움이 나타나고 있다(중앙재난안전대책본부, 2020). 이에 대응하여 정부에서는 코로나 우울 극복과 국민정신건강 증진을 위한 심리방역 관리 체계를 강

화할 필요가 있다. 예를 들어, 상담전화와 휴식 및 치유 프로그램, 우울 자가진단 앱, 검사 및 치료 대책 등 정신건강 증진을 위한 인프라 확장이 필수적이다. 또한, 국민의 ‘정신건강복지’ 계획을 수립하고 추진하여 민간과 정부 간 협력을 강화할 필요가 있다.

넷째, 사회적 거리두기로 인해 여가생활에 많은 제약이 생겼으며, 이는 큰 정신적 스트레스로 이어지고 있다. 특히 여가문화와 관련해서 여행과 영화 키워드가 도출됨으로써 대중이 여행을 가지 못하고, 영화관 입장 제한에 따른 정신적 피로도를 보여주고 있다. 이에 따라 여행과 영화 소비의 제약을 해결할 수 있는 대책을 강구할 필요가 있다. 배영임, 신혜리(2020)는 교육을 포함한 다양한 범위에서의 AR 및 VR 기술 활용도가 지속적으로 높아질 것으로 보고 있으며, 이를 통해 한정된 여가문화의 범주가 더욱 넓어질 것으로 예상된다. 대표적으로, 한 유적지와 여행 관광지 방문 서비스를 확대하거나 영화 스트리밍 서비스에 대한 지원을 확대하는 방안이 있다.

마지막으로, 코로나19 이후의 삶인 포스트코로나에 대해 관심이 높아지고 있다. 코로나19가 변화시킨 사회의 다양한 모습들이 코로나 이후의 삶에 크게 영향을 미칠 것으로 예상된다. 특히 코로나19 위기는 디지털 기반 사회로의 가속화 모멘텀으로 생각할 수 있다. 따라서, 정부는 비대면 신산업을 적극 육성하기 위한 대폭적인 지원과 활성화를 위한 기반을 구축할 필요가 있다. 또한, 한국의 우수한 정보통신 기술 인프라를 활용해 글로벌 디지털 경제를 리드하는 기회를 확보해 한국의 경제 성장을 위한 기회를 잡아야 할 것이다. 또한, 증가하는 재택근무와 온라인 등교로 인해 주거공간과 인테리어에 대한 관심이 높게 나타났다. 특히 많은 기업에서 재택근무의 효율성을 경험하면서 포스트코로나 시대에도 재택근무를 채택할 것으로 예상되며, 주거공간의 중요성이 커지고 있다. 이에 따라

오피스 기능과 더불어 휴식 및 문화를 즐길 수 있는 주거공간의 수요가 커질 것으로 보인다. 따라서, 이와 같은 주거시설 시스템의 확대를 고려해 부동산 제도를 정비하고 정책을 개발해야 할 것이다.

5.2. 연구 한계 및 향후 연구 방향

본 연구는 다음과 같은 한계점이 나타나며 이를 보완한 향후 연구 방향은 다음과 같다. 첫째, 본 연구에서는 소셜미디어 게시글의 작성자 인구통계학적 특성을 확인하지 못했다. 또한, 소셜미디어의 특성상 20, 30대의 소셜미디어 활동이 왕성한 연령층은 분석이 용이하나 상대적으로 소셜미디어 활동량이 적은 노년층 및 비참여자들의 행태는 확인하지 못했다. 데이터 작성자의 인구통계학적 특성을 추가로 조사하여 연령별 및 성별에 따른 추가 분석을 진행한다면 코로나19 관련 주요 이슈에 관해 연령대와 성별에 따른 더 구체적인 결과를 제시할 수 있을 것이다. 이러한 결과는 정부의 감염병 대응 정책수립에 있어 더 세밀하고 체계적인 전략을 수립하는 데 도움을 줄 것이다.

둘째, 코로나19와 관련한 논의 주제는 감염병 발생의 특성상 지역별로 다르게 나타날 수 있으므로, 지역을 구분하여 연구를 진행한다면 더욱 흥미로운 결과가 도출될 것으로 예상된다. 특히 지역을 구분한 공간적 특성과 시기를 구분한 시간적 특성을 동시에 고려하여 분석을 진행한다면, 시공간의 변화에 따라 나타나는 코로나19 관련 주요 이슈를 파악할 수 있을 것이다. 이는 감염병이 시공간으로 확산하는 정도에 따라 나타날 수 있는 사회적 어려움과 위기를 미리 예측해 이를 대비한 대응책을 구상하는 데 큰 도움을 줄 것이라 기대된다.

마지막으로, 텍스트마이닝 분석에 이어 감성분석을 추가로 실시한다면 사람들이 논의하는 주요 토픽에

대한 감성수준을 파악하여 해당 토픽에 대한 관심정도를 객관적으로 확인할 수 있을 것이다. 이러한 결과는 기존의 정부정책에 대한 사람들의 생각과 더불어 감성수준을 확인하여 기존 정책을 보완하는 데 적절하게 활용될 것으로 보인다. 또한, 정책 개발 시 분야별 중요도를 파악할 수 있어 정책 개발의 우선순위를 결정하는 데 큰 도움을 줄 것으로 예상된다.

5.3. 학술적·실무적 시사점

본 연구의 학술적 시사점은 다음과 같다. 첫째, 본 연구는 코로나19에 관한 한국인의 관심사항과 의견을 분석한 실증적 연구이다. 기존의 연구는 시기별 코로나19 주요 이슈 분석과 관련하여 언론보도만을 분석하였으며(김상미, 2020), 언론매체는 국민의 실질적인 의견을 담고 있다고 보기 어렵다. 즉, 코로나19와 같은 위기상황에서 국민에게 실질적으로 필요한 대응책을 제시하기 위해 국민의 의견을 파악하는 것이 매우 중요함에도 불구하고 관련 연구가 부재한 상황이다. 본 연구는 소셜미디어 데이터를 TF-IDF 키워드 추출과 LDA 토픽모델링과 같은 텍스트마이닝 기법으로 분석하고 복합적으로 해석하여 코로나19 상황 속에서 한국인의 실질적인 의견과 반응을 분석했다는 점에서 학술적 시사점이 있다.

둘째, 본 연구는 기존의 선행연구(김상미, 2020; 김기탁, 2020)와 달리 코로나19에 관해 특정 분야에 한정하지 않고 사회 전반에 걸쳐 나타나는 다양한 주요 토픽을 도출한 것에 그 학술적 의의가 있다. 즉, 감염병과 같은 위기상황에 사회의 다양한 영역을 아울러 사람들의 논의사항을 파악해 분야별로 나타나는 주요 주제를 비교할 수 있는 기초자료를 제공하였다.

셋째, 본 연구는 코로나19에 대한 정부의 대응에 관한 사람들의 직접적인 견해를 파악하기 위해 ‘정부의

대응방향’을 기준으로 시기를 나누어 분석을 진행했다. 기존 연구에서는 소셜미디어의 검색량 또는 일별, 월별 기준으로 시기를 구분해 연구를 진행했다(박주현, 2020; 이종수, 2020). 하지만, 본 연구는 정부의 감염병 대응정책 변화에 따라 시기를 구분하여, 정부의 코로나19 대응에 대한 국민의 실질적인 반응과 견해를 파악하였다는 점에서 학술적 의의를 가진다.

본 연구의 실무적 시사점은 다음과 같다. 첫째, 본 연구는 코로나19 장기화에 대비하고, 포스트코로나 시대를 준비하기 위한 구체적 방안을 제시하였다. 코로나19가 발생하고 약 1년 간 형성된 사회적 여론을 분석함으로써 현재 주요하게 나타나고 있는 사회적 시스템의 문제를 지적하고 이를 개선하기 위한 방안을 제안했다. 즉, 코로나19 장기화를 대비해 코로나 시대 속에 살아가는 국민에게 실질적이고 효율적인 도움을 주는 정책 개선을 위한 자료를 제공한 점에 실무적 의의가 있다.

둘째, 본 연구의 결과 분석과 해석을 통해 제안한 정책 전략은 추후 감염병 발생과 같은 사회재난 등의 위험 발생에 직면하여 사회와 국민을 지키기 위한 정부와 기관의 정책 수립과 이행에 구체적인 방향을 제시한다. 특히 국가적 차원에서 다양한 방면에서 국민의 필요와 요구에 따른 정확한 대응책을 개발하기 위한 기초자료를 제공했다는 점에서 실무적 시사점을 가진다.

6. 결론

본 연구는 코로나19 사태 속에서 한국인이 갖고 있는 주요 의견, 태도, 감정 등을 파악하고, 이를 통해 코로나19 팬데믹과 같은 위기상황에 한국 정부와 관련 기관이 국가적 차원에서 국민의 필요와 요구에 따

른 정확한 대응책을 마련하는 데 전략방안을 제시하는 데 목적이 있다. 특히 정부의 대응정책 변화에 따른 국민의 실질적인 의견을 분석하고자 시기의 구분에 있어 정부의 정책방향을 기준으로 시기를 분류하여 연구를 진행했다.

한국인의 견해를 분석하기 위해 한국의 최대 포털 사이트인 네이버의 블로그 데이터를 수집하여 텍스트 마이닝 분석을 진행했다. 텍스트마이닝 기법으로는 TF-IDF 키워드 추출과 LDA 토픽모델링을 활용했으며 각각의 분석결과를 종합하여 해석하였다. 특히 TF-IDF 키워드 추출을 통해 도출한 주요 키워드를 기반으로 코로나19 관련 시기별 반응에 대하여 파악하였다. 최종적으로 LDA 토픽모델링 결과를 통해 시기로 나뉜 주요 이슈를 분석하였고, 토픽모델링 결과의 해석에는 TF-IDF 키워드 추출 결과를 반영하여 종합적으로 해석하였다.

LDA 토픽모델링 분석 결과, 최종 스물네 개의 하위 토픽과 여덟 개의 상위토픽이 도출되었으며, 상위토픽은 (1) 코로나19 상황 파악, (2) 건강, (3) 여가문화, (4) 식사문화, (5) 방역, (6) 경제위기, (7) 교육, (8) 포스트코로나와 같다. 분석 결과와 해석을 기반으로 코로나19 사태와 관련하여 한국의 상황에 적합한 정책과 대응책을 개발하기 위한 구체적인 전략을 제시했다. 결론적으로 본 연구는 감염병과 같은 위기상황에 대비해 사회와 국민을 지키기 위한 정책을 개발하는 데 이바지한다.

<참고문헌>

[국내 문헌]

1. DMC미디어 (2020). **소셜 미디어 현황 및 전망**.
2. KBS (2020). **코로나19: 2차 국민인식 조사**.
3. 국민건강보험 보도자료 (2020). **코로나19로 인한 국민의 의료 이용행태 변화**.
4. 권오정 (2020). 코로나 19에 의한 노인 운동행동 변화 사례 연구. **한국스포츠심리학회지**, 31(2), 123-134.
5. 김기탁 (2020). 소셜네트워크 빅데이터를 활용한 코로나 19에 따른 프로야구 관람문화조사. **한국엔터테인먼트산업학회논문지**, 14(6), 139-150.
6. 김상미 (2020). 코로나 19 관련 온라인 교육에 관한 국내 언론보도기사 분석. **한국디지털콘텐츠학회 논문지**, 21(6), 1091-1100.
7. 도보람, 윤여림, 김태년, 최정혜 (2021). 코로나 19 상황에서 직무만족도와 모바일 생산활동: 결정요인 연구. **지식경영연구**, 22(1), 39-56.
8. 문화체육관광부 보도자료 (2020). **생활권 내 여가생활과 비대면 문화 활동 증가**.
9. 박성욱 (2020). 코로나 19의 경제적 영향 및 정책대응 방향. **주간금융브리프**, 29(6), 3-10.
10. 박주현 (2020). 언론의 이념성향에 따른 '코로나 19' 보도 프레임 비교 연구. **한국언론학보**, 64(4), 40-85.
11. 배영임, 신혜리 (2020). 코로나 19, 언택트 사회를 가속화하다. **이슈 & 진단**, 416, 1-26.
12. 서울시 보도자료 (2020). **(석간) 서울시민 10명중 4명 코로나 19로 '정신건강 나빠졌다'...여가여행 제약 제일 힘들어**.
13. 오픈서베이 (2020). **소셜미디어와 검색 포털에 관한 리포트 2020**.
14. 옥철 (2020). WHO, 코로나19 팬데믹 선언. **연합뉴스**.
15. 윤명희 (2007). 블로그의 사회적 유형분석: 1인 커뮤니티의 다층화. **한국사회학**, 41(1), 156-193.
16. 윤정현 (2020). COVID-19 이후, 뉴노멀과 미래 사회 변화. **FUTURE HORIZON**, 4-7.
17. 이새미, 유승의, 안순재 (2020). 텍스트 마이닝을 활용한 매스 미디어와 소셜 미디어 의제 분석: '마스크 5부제'를 중심으로. **한국콘텐츠학회논문지**, 20(6), 460-469.
18. 이정연, 박미희, 소미영, 안수현 (2020). **코로나19와 교육:**

학교구성원의 생활과 인식을 중심으로. 경기도교육연구원 이슈페이퍼 2020-08.

19. 이정현, 선형주, 이흥주 (2020). 텍스트 마이닝을 활용한 스마트 스피커 제품의 포지셔닝: 인공지능 속성을 중심으로. **지식경영연구**, 21(1), 197-210.
20. 이종수 (2020). 코로나 19 관련 CNN 뉴스 영상분석: '타자'의 질병에서 '우리'의 질병으로. **미디어, 젠더 & 문화**, 35(2), 245-298.
21. 임현아, 최재원, 이흥주 (2019). 텍스트 분석을 통한 제품 분류 체계 수립방안: 관광분야 App을 중심으로. **지식경영연구**, 20(3), 139-154.
22. 중앙사고수습본부 보도자료 (2020). **코로나바이러스감염증-19. 중앙사고수습본부 정례브리핑**.
23. 중앙재난안전대책본부 보도자료 (2020). **과일채소 섭취·잠잠이 운동... '코로나 시대 건강 생활수칙'**.
24. 통계청 보도자료 (2020). **2020년 8월 온라인쇼핑 동향**.
25. 홍순구, 유승의, 김나랑, 이태현, 이새미, 안순재 (2020). **스마트 거버넌스 정책과정의 혁신**. 유원북스.

[국외 문헌]

26. Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study. *Journal of Medical Internet Research*, 22(4), e19016.
27. Aggarwal, J., Rabinovich, E., & Stevenson, S. (2020). *Exploration of gender differences in COVID-19 discourse on reddit*. arXiv preprint arXiv: 2008.05713.
28. Aslam, F., Awan, T. M., Syed, J. H., Kashif, A., & Parveen, M. (2020). Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak. *Humanities and Social Sciences Communications*, 7(1), 1-9.
29. Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.
30. Blei, D. M., & Lafferty, J. D. (2006, June). Dynamic topic models. *In Proceedings of the 23rd International Conference on Machine Learning*, 113-120.
31. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning*

- Research*, **3**, 993–1022.
32. Damiano, A. D., & Allen Catellier, J. R. (2020). A content analysis of coronavirus tweets in the United States just prior to the pandemic declaration. *Cyberpsychology, Behavior, and Social Networking*, **23**(12), 889–893.
 33. Freeman, L. C. (2005). Graphical techniques for exploring social network data. *Models and Methods in Social Network Analysis*, 248–269.
 34. Liu, Q., Zheng, Z., Zheng, J., Chen, Q., Liu, G., Chen, S., ... & Ming, W. K. (2020). Health communication through news media during the early stage of the COVID-19 outbreak in China: Digital topic modeling approach. *Journal of Medical Internet Research*, **22**(4), e19118.
 35. Murphy, M. P. (2020). COVID-19 and emergency eLearning: Consequences of the securitization of higher education for post-pandemic pedagogy. *Contemporary Security Policy*, **41**(3), 492–505.
 36. Park, Y. J., Choe, Y. J., Park, O., Park, S. Y., Kim, Y. M., Kim, J., ... & COVID-19 National Emergency Response Center, Epidemiology and Case Management Team (2020). Contact tracing during coronavirus disease outbreak, South Korea, 2020. *Emerging Infectious Diseases*, **26**(10), 2465–2468.
 37. Slovic, P. (1987). Perception of risk. *Science*, **236**(4799), 280–285.
 38. Van Lancker, W., & Parolin, Z. (2020). COVID-19, school closures, and child poverty: A social crisis in the making. *The Lancet Public Health*, **5**(5), e243–e244.
 39. Yu, M., Li, Z., Yu, Z., He, J., & Zhou, J. (2020). Communication related health crisis on social media: A case of COVID-19 outbreak. *Current Issues in Tourism*, 1–7.
 40. Zhao, Y., Cheng, S., Yu, X., & Xu, H. (2020). Chinese public's attention to the COVID-19 epidemic on social media: Observational descriptive study. *Journal of Medical Internet Research*, **22**(5), e18825.

● 저 자 소 개 ●



김진솔 (Jinsol Kim)

연세대학교 정보대학원에서 비즈니스 빅데이터 분석을 전공하고, 정보시스템학 석사 학위를 취득하였다. 주요 연구분야는 빅데이터 분석, 머신러닝, 딥러닝 등이다.



신동훈 (Donghoon Shin)

연세대학교 정보대학원 비즈니스 빅데이터 분석 트랙 석사과정이다. 현재 디지털서비스연구실에서 데이터분석 관련 연구들을 진행 중이다. 주요 연구 분야는 빅데이터분석, 머신러닝, 딥러닝 등이다.



김희웅 (Heewoong Kim)

National University of Singapore 정보시스템학과에서 근무한 후, 현재 연세대학교 정보대학원 교수로 재직 중이다. 주요 연구분야는 디지털 비즈니스, 정보시스템 관리 및 활용 등이다. 관련 연구들은 MIS Quarterly, Information Systems Research, Journal of Management Information Systems 등에 50여 편의 논문이 게재되었다. JAIS, IEEE TEM의 편집위원으로 활동했고, KrAIS 회장을 역임했다.

〈 Abstract 〉

Analysis of Major COVID-19 Issues Using Unstructured Big Data

Jinsol Kim^{*}, Donghoon Shin^{**}, Heewoong Kim^{***}

As of late December 2019, the spread of COVID-19 pandemic began which put the entire world in panic. In order to overcome the crisis and minimize any subsequent damage, the government as well as its affiliated institutions must maximize effects of pre-existing policy support and introduce a holistic response plan that can reflect this changing situation- which is why it is crucial to analyze social topics and people's interests. This study investigates people's major thoughts, attitudes and topics surrounding COVID-19 pandemic through the use of social media and big data. In order to collect public opinion, this study segmented time period according to government countermeasures. All data were collected through NAVER blog from 31 December 2019 to 12 December 2020. This research applied TF-IDF keyword extraction and LDA topic modeling as text-mining techniques. As a result, eight major issues related to COVID-19 have been derived, and based on these keywords, this research presented policy strategies. The significance of this study is that it provides a baseline data for Korean government authorities in providing appropriate countermeasures that can satisfy needs of people in the midst of COVID-19 pandemic.

Key Words: COVID-19, Text-mining, Blog bigdata, TF-IDF Keyword Extraction, LDA Topic Modeling

* Samsung Medical Center

** Graduate School of Information, Yonsei University

*** Graduate School of Information, Yonsei University