

트위터 게시물 분석을 통한 코로나바이러스감염증-19 백신에 대한 의견 탐색

Exploring Opinions on COVID-19 Vaccines through Analyzing Twitter Posts

정우진 (Woojin Jung)*, 김규리 (Kyuli Kim)**
유승희 (Seunghee Yoo)***, 주영준 (Yongjun Zhu)****

초 록

본 연구는 코로나바이러스감염증-19(이하 코로나바이러스) 백신에 대한 사회적 의견을 파악하기 위해 트위터에서 작성된 백신 관련 게시물들을 분석하였다. 2020년 3월 16일부터 2021 3월 15일까지 1년간 트위터에서 작성된 코로나바이러스 백신 이름을 키워드로 포함한 45,413개의 게시물을 수집하여 분석하였다. 데이터 수집을 위해 활용된 코로나바이러스 백신 키워드는 총 12개이며, 수집된 게시물 수준으로 '화이자', '아스트라제네카', '모더나', '얀센', '노바백스', '시노팜', '시노백', '스푸트니크', '바라트', '칸시노', '추마코프', '벡토르'이다. 수집된 게시물들은 수기와 자동화된 방법을 동시에 활용하여 키워드 분석, 감성 분석, 및 토픽모델링을 통하여 백신들에 대한 의견을 탐색하였다. 연구결과에 따르면 전반적으로 백신에 대한 부정적인 반응이 많았으며, 백신 접종 후유증에 대한 불안 및 백신의 효능에 대한 불신이 백신들에 대한 부정적인 주요 요소로 파악되었다. 이와는 반대로, 백신 접종에 따른 코로나바이러스 확산 억제에 대한 기대감이 백신에 대한 긍정적인 사회적 요소인 것을 확인할 수 있었다. 본 연구는 기존의 선행연구들이 뉴스 등 대중매체 데이터를 통해 코로나바이러스 백신에 대한 사회적 분위기를 파악하고자 했던 것과 달리, 소셜 미디어 데이터 수집 및 이를 활용한 키워드 분석, 감성 분석, 토픽 모델링 등의 여러 분석방법들을 사용하여 대중들의 의견을 파악하는 것으로 학술적 의의를 지닌다. 또한, 본 연구의 결과는 백신에 대한 사회적 분위기를 반영한 백신 접종 권장 정책 수립 기여라는 실질적 함의를 시사한다.

ABSTRACT

In this study, we aimed to understand the public opinion on COVID-19 vaccine. To achieve the goal, we analyzed COVID-19 vaccine-related Twitter posts. 45,413 tweets posted from March 16, 2020 to March 15, 2021 including COVID-19 vaccine names as keywords were collected. The 12 vaccine names used for data collection included 'Pfizer', 'AstraZeneca', 'Moderna', 'Jansen', 'NovaVax', 'Sinopharm', 'SinoVac', 'Sputnik V', 'Bharat', 'KhanSino', 'Chumakov', and 'VECTOR' in the order of the number of collected posts. The collected posts were analyzed manually and automatedly through keyword analysis, sentiment analysis, and topic modeling to understand the opinions for the investigated vaccines. According to the results, there were generally more negative posts about vaccines than positive posts. Anxiety about the aftereffects of vaccination and distrust in the efficacy of vaccines were identified as major negative factors for vaccines. On the contrary, the anticipation for the suppression of the spread of coronavirus following vaccination was identified as a positive social factor for vaccines. Different from previous studies that investigated opinions about COVID-19 vaccines through mass media data such as news articles, this study explores opinions of social media users using keyword analysis, sentiment analysis, and topic modeling. In addition, the results of this study can be used by governmental institutions for making policies to promote vaccination reflecting the social atmosphere.

키워드: 코로나바이러스감염증-19, 백신, 소셜 미디어, 트위터, 감성 분석, 토픽 모델링
COVID-19, vaccine, social media, Twitter, sentiment analysis, topic modeling

* 성균관대학교 문헌정보학과 박사과정(dnwls54@g.skku.edu) (제1저자)

** 성균관대학교 문헌정보학과 석사과정(kyulikim@g.skku.edu) (공동저자)

*** 성균관대학교 데이터사이언스융합전공 학사과정(yoush20@naver.com) (공동저자)

**** 성균관대학교 문헌정보학과 조교수(yzhu@skku.edu) (교신저자)

■ 논문접수일자: 2021년 11월 15일 ■ 최초심사일자: 2021년 12월 7일 ■ 게재확정일자: 2021년 12월 20일

■ 정보관리학회지, 38(4), 113-128, 2021. <http://dx.doi.org/10.3743/KOSIM.2021.38.4.113>

© Copyright © 2021 Korean Society for Information Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

1.1 연구의 필요성

2019년 12월 처음 발견된 코로나바이러스감염증-19(이하 코로나바이러스)는 빠른 전염률과 강력한 치사율로 인해, 최초 발견 이후 3개월도 되지 않아 세계보건기구(WHO)에 의해 팬데믹으로 지정되었다(Domenico & Maurizio, 2020). 현재까지도 코로나바이러스의 끊임없는 변이와 방역 및 통제의 어려움으로 인해 코로나바이러스의 확산 세는 계속되고 있다(최성민, 2021). 2021년 11월 6일 기준으로 전 세계에서 코로나바이러스에 약 2억 4천 8백만명이 감염되었고, 코로나바이러스로 인해 약 5백만 명이 사망한 것으로 집계되었다(Worldometers.info, 2021). 국내의 경우에는 2021년 11월 6일까지 코로나바이러스에 377,712명이 감염되었고, 2,956명의 사람들이 코로나바이러스로 인해 사망하였다(중앙방역대책본부, 2021). 코로나바이러스로 인한 전 세계적 위기를 타개하기 위해서는 백신의 접종이 필수적이며(Samira et al., 2021), 백신 접종은 자기 보호 효과와 더불어 바이러스 배출까지 억제할 수 있기 때문에 바이러스 전염 가능성을 낮출 수 있다(성백린, 2021). 국내의 경우 2021년 2월 26일부터 화이자, 아스트라제네카, 모더나, 얀센 등의 백신들을 수급 및 접종을 진행하고 있다. 2021년 11월 6일 자 대한민국의 백신 완전 접종률은 약 76.5%로(김장현, 2021), 이는 2021년 11월 5일 기준으로 OECD 국가 38개국 중 8위에 해당하는 접종률이다(Mathieu et al., 2021). 전 세계적으로 대한민국의 접종률은 높은 수준이지만, 임상실험에

서 제시된 백신의 예방효과 수준의 집단면역 효과를 얻기 위해서는 예방효과보다 높은 접종률이 필요하다(김서영, 2021). 김서영(2021)의 기사에 따르면 대한민국 국민의 약 90%가 접종을 완료해야 70%의 집단면역 효과를 기대할 수 있다고 밝혔으며, 목표 수준의 접종률에 최대한 빨리 도달하기 위한 노력이 필요하다고 언급하였다.

목표 접종률에 도달하기 위해서는 백신에 대한 대중들의 부정적인 인식 원인을 파악하고, 부정적인 인식을 완화할 수 있는 접종 독려 방침을 구상할 필요가 있다(Samira et al., 2021). 이에 본 연구는 소셜 미디어 데이터를 활용하여 코로나바이러스 백신에 대한 대중들의 의견을 알아보려고 하였다.

1.2 선행연구

소셜 미디어 상에서 이용자들은 솔직하게 자신의 의견을 피력하며(서하림, 송민, 2019), 이러한 소셜 미디어의 특성을 활용해 소셜 미디어 상에서 논해지는 다양한 주제들에 대한 여러 연구들이 진행되었다(서하림, 송민, 2019; 주영준 외, 2019; 진설아 외, 2013). 소셜 미디어의 방대한 양의 텍스트 데이터를 분석하는 기법들이 존재하며(주영준 외, 2019), 그중 감성 분석과 토픽 모델링은 여러 연구에서 활용되고 있다. 감성 분석은 텍스트에 포함된 작성자의 감성을 파악하는 방법으로, 다양한 주제에서 소셜 미디어 게시물 작성자들의 감정을 파악하는데 활용되고 있다(서하림, 송민, 2019). 토픽 모델링은 Latent Dirichlet Allocation(LDA) 기법을 활용하여 분석하고자 하는 말뭉치에서 주제 및 각 주제를 구성하는 주제어들의 분포를

나타내며(Blei, Ng, & Jordan, 2003), 여러 연구들에서 소셜 미디어 상에서 논해지는 의견들의 탐색을 위해 사용된다.

소셜 미디어는 백신과 관련한 사회적 분위기를 파악하는데도 유용할 수 있다. 해외에서 소셜 미디어는 백신에 대한 일반 대중의 의견에 대한 정보를 검색하는데 유용한 정보원으로써(Lucia et al., 2021), 코로나바이러스 발생 이전에도 소셜 미디어 상에서의 전염병 백신 접종에 대한 의견 탐색을 진행하는 연구들이 진행되어왔다. Lara et al.(2020)은 2016년부터 2017년 사이 이탈리아에서 실시된 예방접종에 관한 트윗들을 수집 및 분석하였으며, 백신과 관련된 부정적인 사건들은 접종 관련 부정적인 트윗의 수에 영향을 미친다고 분석하였다. 소셜 미디어상에서 백신에 대한 부정적인 정서 및 정보에 노출된 사람은 백신 접종을 주저하거나 거부까지 이어질 수 있다(Samira et al., 2021).

국내에서는 코로나바이러스에 대한 사회적 분위기를 분석하는 연구들이 진행되었으며, 해당 연구들은 코로나바이러스 관련 보도자료를 활용하였다(김태중, 2020; 유소연, 임규건, 2021; 최원주, 홍장선, 2021). 코로나바이러스 관련 뉴스에서 불안, 공포, 우려 등의 코로나바이러스에 대한 사회적 분위기를 반영한 단어들이 다수 발견되었으며(김태중, 2020; 최원주, 홍장선, 2021), 사회적 거리, 자가 격리 등과 같은 코로나바이러스로 인한 사회적 변화 또한 자주 언급되었다(유소연, 임규건, 2021). 하지만 해당 연구들은 모두 분석 대상이 뉴스 등 대중매체에 집중되어 있으며, 이는 코로나바이러스 및 백신과 관련한 다양한 대중들의 의견을 수집할 수 없다는 한계점을 지닌다(유소연, 임규건, 2021).

따라서, 본 연구는 소셜 미디어 플랫폼 중 하나인 트위터에서 세계 최초로 코로나바이러스 백신 임상실험이 시작된 2020년 3월 16일부터 2021년 3월 15일까지의 1년간 코로나바이러스 백신들과 관련된 게시물들을 수집 및 분석을 통해 코로나바이러스 백신에 대한 대중들의 생각을 파악하고자 한다. 이후, 분석 결과를 통해 향후 백신 접종 권장을 위한 정책 수립에 기여하고자 한다.

2. 데이터

본 연구를 위해 트위터에서 2020년 3월 16일부터 2021년 3월 15일까지의 1년간 작성된 코로나바이러스 백신 관련 게시물들을 수집하였다. 2021년 3월 15일 기준 최소 한 국가에서 배포 승인이 허가된 12개의 백신들에 대하여 국내에서 대중적으로 통용되는 명칭들을 키워드로 사용하였다. 예를 들어, '아스트라제네카'사에서 개발한 'AZD1222' 백신은 공식 명칭보다는 흔히 '아스트라제네카'라는 이름으로 통용된다. 파이썬 라이브러리인 snsrape(JustAnotherArchivist, 2020)를 활용해 수집된 게시물은 총 45,413개이고, 키워드 별 수집 게시물 수는 <표 1>과 같다. <표 1>에 따르면, 관련 게시물 수 기준으로 상위 네 개의 백신(화이자, 아스트라제네카, 모더나, 안센)과 관련된 게시물의 수(42,261개)와 하위 여덟 개 백신(노바백스, 시노팜, 시노백, 스포트닉, 바라트, 켄시노, 추마코프, 벡토르)들의 게시물의 수(3,152개)에 큰 차이가 있음을 발견할 수 있었다. 게시물 수 기준 상위 백신 관련 게시물과 하위 백신 관련 게시물로 나누어 따로 분

석을 진행하였다. 하위 백신 관련 게시물은 연구자들이 수기로 전처리 및 분석 작업을 진행하였으며, 상위 백신 관련 게시물은 자동화된 방식을 활용하여 전처리 및 분석 작업을 진행하였다.

〈표 1〉 백신 별 수집 게시물

백신 명	수집 게시물 수
화이자	20,133
아스트라제네카	11,720
모더나	8,521
얀센	1,887
노바백스	1,062
시노팜	965
시노백	643
스푸트니크	264
바라트	190
칸시노	14
추마코프	7
벡토르	7

3. 연구방법

3.1 게시물 수 기준 하위 백신 관련 게시물 전처리 및 분석

본 연구는 게시물 수 기준 하위 백신 관련 게시물 3,152개에 대하여 세 명의 연구자들이 총 2단계의 전처리 작업을 진행하였다. 1단계 전처리에 서 연구자들은 수기로 백신과 관련 없는 1,118개의 게시물을 제거하였다. 이후 2단계 전처리 과정에서는 본 연구의 목적에 부합하지 않는 1,471개의 게시물을 제거하였다. 제거 기준으로는 백신에 대한 소셜 미디어 이용자들의 의견이 반영되지 않은 게시물들(예: 신문기사, 단순 정보 전

달을 위한 게시물)을 제거하였다. 위 과정에서 네 개의 백신(스푸트니크, 바라트, 칸시노, 추마코프) 관련 게시물들은 본 연구에 적합한 데이터가 존재하지 않아 분석 대상에서 제외되었다.

2단계 작업을 통해 남은 279개의 게시물들에 대하여 세 명의 연구자들이 개별적으로 각 백신들에 대한 긍정/부정 레이블링을 진행하였다. 레이블링이 일치하지 않는 게시물들은 상호 합의를 통해 최종 레이블링을 진행하였으며, 이후 트윗 별 긍정/부정 레이블링 원인들을 텍스트에서 추출하였다. 레이블링 원인이 명확히 드러나지 않은 게시물들은 세 명의 연구자가 작성일자 기준 1일 이내에 작성된 해당 백신 관련 기사 및 수집된 게시물들의 내용을 토대로 레이블링 원인들을 유추하였다. 예를 들어 ‘xx백신은 안 되...’라는 게시물의 경우, 해당 백신과 관련된 1일 이내의 기사들 및 같은 일자에 수집된 해당 백신 관련 게시물들로부터 백신에 부정적인 원인이 부작용인 경우, 레이블링 원인을 부작용으로 추출하였다. 추출된 원인들은 이후 연구자들의 상의를 통해 표준화 작업을 진행하였다. ‘비싸다’, ‘가격오바야’, ‘가격실화냐’ 등의 키워드들이 특정 백신에 대한 부정적인 요인들로 추출된 경우, 해당 키워드들은 부정적인 의미의 ‘가격’으로 표준화 되었다. 위 과정을 거친 게시물들의 분석을 통해, 각 게시물 수 기준 하위 백신 별 긍정/부정 게시물들에 대한 소셜 미디어상에서의 의견을 탐색하고자 하였다.

3.2 게시물 수 기준 상위 백신 관련 게시물 전처리 및 분석

게시물 수 기준 상위 백신 관련 게시물 42,261

개 중 백신 이름이 두 개 이상 언급된 게시물들은 제거하였다. 이는 텍스트에 복수의 백신들이 포함돼 있을 때, 감성 분석을 통한 평가 대상 백신을 파악할 수 없기 때문이다. 위 과정에서 안센의 경우, 전처리 후 남은 게시물이 백신과 관련된 게시물이 존재하지 않아 분석 대상에서 제외되었다. 안센 관련 게시물들은 만화, 애니메이션, 게임 관련한 내용들로써, 본 연구에 필요한 코로나바이러스 백신으로써의 안센 관련 게시물은 없는 것으로 파악되었다. 남은 34,357개의 게시물들에서 한국어 텍스트 분석에 불필요한 특수기호, 외국어, 숫자, URL 링크들을 제거하였다. 이후 한국어 형태소 분석을 위한 파이썬 라이브러리인 코엔엘파이(KoNLPy)(박은정, 조성준, 2014)를 활용하여 게시물의 토큰화를 시행하였다. 형태소들 중 분석에 불필요한 조사와 어미들을 제외하였으며, 분석에 불필요하다고 파악되는 단어들을 추가로 제거하였다.

다음으로 해당 백신들에 대한 소셜 미디어 이용자들의 평가를 긍정/부정으로 나누어 파악하고자 감성 분석을 진행했으며 이를 위해 KNU 한국어 감성사전을 사용하였다. KNU 한국어 감성사전은 14,843개의 단어들로 이루어져 있으며 표준어뿐만이 아닌 온라인상에서 사용되는 신조어나 축약어들도 포함되어 있어 소셜

미디어 텍스트 감성 분석에 있어 효과적이다(박상민 외, 2018). KNU 한국어 감성사전에 등록된 단어들은 매우 부정, 부정, 중립, 긍정, 매우 긍정으로 레이블링 되었으며, 레이블링에 따라 -2점부터 2점까지 감성 점수를 가진다. 각 게시물들의 형태소들 중 KNU 한국어 감성 사전에 포함된 단어들에 부여된 감성 점수의 합산을 통해 합산 점수가 0보다 크면 긍정, 작으면 부정으로 각 게시물들에 나타난 백신에 대한 의견을 계산하였다.

추가로 소셜 미디어상에서 코로나바이러스 백신들에 대한 평가 및 논의되는 주제를 파악하기 위해 토픽 모델링을 사용하였다. 토픽 모델링을 진행하기에 앞서, 토픽 수에 따른 주제 일관성 측정을 통해 분석하고자 하는 텍스트의 최적 토픽 수를 선정하였다(Linton et al., 2017). 분석을 위해 각각의 백신들에 대한 게시물 중, 앞선 감성 분석을 통해 긍정 혹은 부정으로 레이블링된 13,507개의 게시물들만을 대상으로 파이썬 라이브러리인 gensim(Rehurek & Sojka, 2011)을 활용하여 2부터 20까지의 주제 일관성을 계산하였으며, 각각의 백신 별 긍정/부정 텍스트에 대한 주제 일관성 결과를 토대로 선정된 백신 별 긍정/부정 게시물들에 대한 최적 토픽 수는 <표 2>와 같다. <표 2>의 최적 토픽 수

<표 2> 게시물 수 기준 상위 백신 별 긍정/부정 게시물 최적 토픽 수

백신 명	긍정/부정	최적 토픽 수
화이자	긍정	6
	부정	5
아스트라제네카	긍정	3
	부정	6
모더나	긍정	6
	부정	3

에 따라 각각의 게시물 집합들의 토픽 모델링을 gensim(Rehurek & Sojka, 2011)을 활용하여 진행하였으며, 총 여섯 개의 토픽 모델링 결과가 추출되었다.

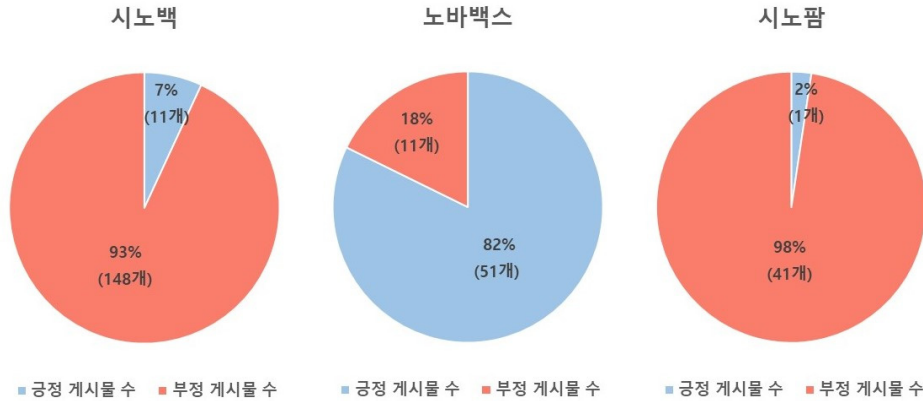
각각의 토픽 모델링 결과는 토픽과 토픽을 구성하는 주제어들로 이루어져 있으며, 주제어들이 각각의 토픽에 포함될 확률은 다르게 존재한다. 본 연구는 주영준 외(2019)의 연구에서 제시한 토픽 별 주제어 처리 방법을 사용하여 주제어들이 가장 높은 확률 값을 가지는 토픽에만 주제어들을 배정하였다. 토픽들 간 주제어 중복 출현을 제거함으로써 각 토픽들을 구성하는 주제어들의 대표성이 보장된다. 또한, 단어의 중복 출현으로부터 발생하는 토픽 레이블링의 모호함을 방지할 수 있다. 고유한 주제어들만을 포함한 토픽들에 대해 세 명의 연구자들이 개별적으로 레이블링을 진행하였다. 이후 토픽 레이블링 통일을 위해 협의를 진행하였으며, 두 명 이상의 연구자가 토픽에 대해 동일하게 레이블링을 진행한 경우에만 토픽의 레이블을 확정하였다. 그렇지 못한 토픽의 경우 세 명 외의 연구자가 중재자로 참여, 해당 토픽에 대한 레이블을 네 명의 연구자가 협의하여 최종 확정하였다. 레이블링된 토픽 별 확률 분포 파악을 통해, 가장 높은 확률 값을 가지는 주제가 해당 게시물의 주요 주제인 것으로 할당하였다. 예를 들어 토픽 수를 세 개로 설정한 모델에서 게시물의 토픽 별 분포 값이 토픽 1(70%), 토픽 2(12%), 토픽 3(18%)인 경우, 토픽 1을 해당 게시물의 중심 주제로 할당하였다. gensim을 활용하여 게시물의 대표 토픽을 계산하였으며, 이를 통해 백신들에 대한 토픽 별 게시물 분포를 분석하였다.

4. 연구 결과

4.1 시노백, 노바백스, 시노팜 관련 게시물 분석

게시물 수 기준 하위 백신들과 관련한 279개의 트위터 게시물들에 대하여 감성 분석을 실시하였으며, 백토르의 경우 전처리 과정 이후의 관련 게시물의 수(1개)가 충분하지 않아 분석에서 제외하였다. 시노백 관련 게시물이 164개, 노바백스 73개, 시노팜 42개의 게시물에 대하여 감성 분석을 실시하였고 백신 별 긍정/부정 게시물 수는 <그림 1>과 같다. 시노백의 경우, 부정적인 게시물(153개)이 긍정적인 게시물(11개) 보다 많은 것을 확인할 수 있었고 시노팜 역시 부정적인 게시물(41개)과 긍정적인 게시물(1개) 개수에서 큰 차이를 보였다. 이를 통해, 시노백과 시노팜에 대해 트위터에서 부정적인 의견이 대다수를 차지한다는 것을 파악할 수 있었다. 반면 노바백스의 경우, 긍정적인 게시물(61개)의 수가 부정적인 게시물(12개) 보다 많았으며, 트위터상에서 노바백스에 대해 긍정적인 반응이라는 것을 유추할 수 있었다.

백신 별 긍정/부정적 의견을 자세히 살펴보기 위해 키워드 추출 및 표준화 작업을 통한 키워드 분석을 실시하였다. 시노백의 경우, 긍정적인 반응을 보이는 게시물들에서 '생산방식', '백신효능' 등의 키워드를 추출할 수 있었다. '생산방식'의 경우, 시노백의 생산방식이 백신 개발에서 과거부터 사용된 전통적인 방식으로 안전성에 신뢰가 가기 때문에 긍정적이라는 것을 파악할 수 있었다. 또한, 브라질에서 실시한 임상실험을 통해 백신의 유효성이 입증됨으로



〈그림 1〉 시노백, 노바백스, 시노팜 긍정/부정 게시물 분포

써 '백신효능'에 대해 긍정적인 반응을 보이고 있음을 알 수 있었다. 부정적인 게시물의 경우 '생산방식', '백신효능', '가격' 등의 키워드를 추출하였다. 전통적인 생산방식이 긍정적인 게시물에서는 안전성에 대한 신뢰 요인으로 나타난 반면, 부정적인 게시물에서는 전통적인 생산방식에 따른 백신 효능의 한계점이 백신의 '생산방식'에 대한 부정적인 요인으로 작용하는 것을 볼 수 있었다. '백신효능'의 경우, 임상실험으로 나타난 저조한 백신의 예방효과와 부작용 및 임상실험 결과 발표에 대한 불신으로 인해 부정적인 반응을 나타내는 것을 알 수 있었다. 다른 백신들에 비해 높은 '가격' 또한 시노백에 대한 부정적인 평가요인으로 작용한다고 유추할 수 있었다.

노바백스에 대해 긍정적인 게시물들에서 '국내생산', '백신효능', '보관 및 관리', '가격' 등의 키워드를 추출하였다. 노바백스가 국내 기업의 위탁 생산을 통해 공급된다는 측면에서 긍정적인 반응을 파악할 수 있었으며, 노바백스의 임상실험 결과가 '백신효능'에 대한 긍정적인 평가로 작용한 것을 알 수 있었다. '보관 및 관리'

의 용이함 및 '가격'이 다른 백신들에 비해 저렴한 점 또한 노바백스에 대한 긍정적 평가요인인 것을 유추할 수 있었다. 부정적인 반응으로는 노바백스의 '생산방식'이 기존의 전통적인 백신 개발 방식과는 다른 점에서 발생하는 불안함과 관리 및 용이의 어려움이 언급되었으며, 불만족스러운 임상실험 결과에 따른 저조한 '백신효능'이 키워드로 추출되었다.

시노팜의 경우 임상결과가 '백신효능'에 긍정적인 평가의 원인으로 파악할 수 있었다. 시노팜에 대한 부정적인 키워드로는 신빙성이 낮은 임상실험 결과로 인한 '백신효능'과 다른 백신들에 비해 높은 '가격'이 부정적인 키워드로써 추출되었다.

〈그림 2〉는 백신 별 긍정/부정 게시물들을 대상으로 추출된 키워드들의 분포를 나타낸 것으로, 한 게시물에서 여러 개의 키워드가 나온 경우 중복으로 집계하였다. 〈그림 2〉에 따르면, 시노백에 대한 긍정 게시물들 중 '백신효능', '생산방식' 순으로 백신에 대한 긍정적인 요소로 언급되었으며, 부정적인 게시물들 중에는 '백신효능', '가격', '생산방식' 순으로 부정적인 이유



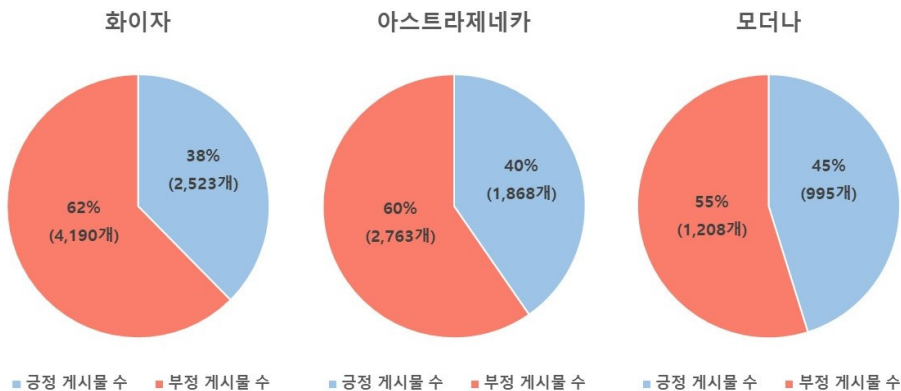
〈그림 2〉 시노백, 노바박스, 시노팜 긍정/부정 게시물 키워드 분포

로 언급되었다. 특히, ‘백신효능’(141개)에 대한 불신이 다른 요소들에 비해 해당 백신에 대한 부정적인 의견의 대다수를 차지하는 것을 알 수 있었다. 노바박스의 경우 긍정적인 요소는 ‘국내생산’, ‘백신효능’, ‘보관 및 관리’, ‘가격’의 순서로, 부정적인 요소로는 ‘백신효능’과 ‘생산방식’의 비중이 동일하였다. ‘국내생산’(30개)과 ‘백신효능’(27개)이 다른 요소들에 비해 소셜 미디어상에서 노바박스를 긍정적으로 평가하는 주요 요인이라고 유추할 수 있었다. 시노팜의 경우, 긍정적인 의견을 내포한 게시물이 한 개로 ‘백신효능’에 관한 것이었으며, 부정적인 게시물들은 ‘백신효능’, ‘가격’ 등의 원인을 포함하였다. ‘백신효능’에 대해 부정적인 반응을 보이는 게시

물이 40개로, ‘백신효능’이 시노팜의 부정적인 평가에 대한 가장 큰 원인을 알 수 있었다.

4.2 화이자, 아스트라제네카, 모더나 관련 게시물 분석

상위 세 개 백신 관련 게시물은 감성 분석을 통해 긍정 혹은 부정으로 레이블링된 총 13,507 개의 게시물을 대상으로 감성 분석을 진행하였다. 백신 별 분석 대상 게시물은 화이자 6,713 개, 아스트라제네카 4,631개, 모더나 2,163개이며 감성 분석 결과는 〈그림 3〉과 같다. 화이자의 경우 분석 게시물 중 2,523개(38%)가 긍정, 4,190개(62%)가 부정적인 것을 알 수 있었다.



〈그림 3〉 화이자, 아스트라제네카, 모더나 긍정/부정 게시물 분포

이를 통해 트위터상에서 화이자에 대해 부정적인 반응이 더 많다는 것을 확인할 수 있었다. 아스트라제네카와 모더나 또한 분석 결과로 아스트라제네카는 긍정 게시물 1,868개(40%), 부정 게시물 2,763개(60%), 모더나의 경우 긍정 955개(44%), 부정 1,208개(56%)로 파악되었다. 이를 통해, 화이자뿐만이 아닌 상위 세 개 백신들 모두 트위터에서 긍정적인 의견보다는 부정적인 반응이 많은 것을 확인할 수 있었다.

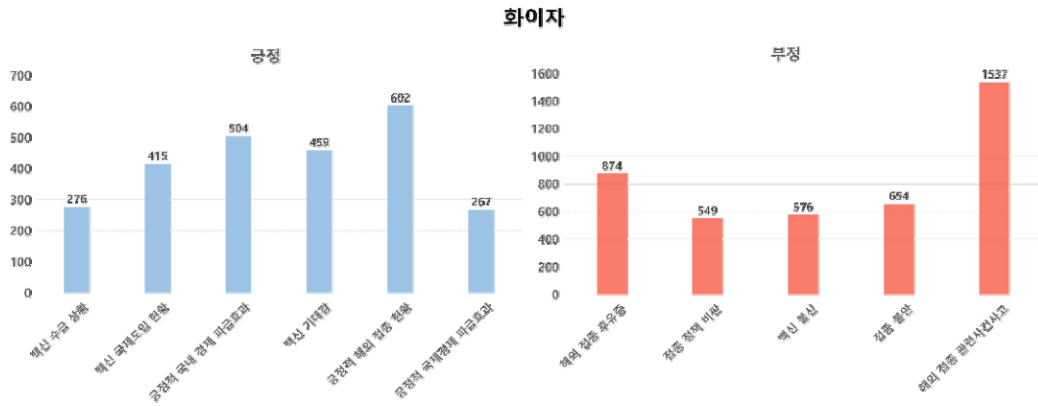
백신들에 대한 트위터상에서 논의되는 의견들의 기간별 변화를 살펴보기 위해 토픽 모델링을 진행하였다. 감성 분석을 통해 계산된 각 백신들에 대한 긍정/부정 게시물 별로 토픽 모델링을 실시하였으며, 이를 통해 백신들에 대한 긍정 혹은 부정적인 의견들을 심층적으로 살펴보고자 하였다. 백신 별 긍정/부정 게시물들에 대하여 주제 일관성 검사로 추출된 최적 토픽수로 생성된 토픽 모델링의 토픽들과 각 토픽들을 대표하는 주요 키워드들을 추출하였으며, <표 3>은 화이자 게시물에 대한 토픽 모델링 분

석에 따른 토픽들과 각 토픽들을 대표하는 주요 키워드들을 나타낸 것이다. 화이자에 대한 긍정 게시물들의 경우, 총 여섯 개의 토픽이 추출되었으며 각각의 토픽들은 토픽 번호순으로 '백신 수급 상황', '백신 국제도입 현황', '긍정적 국내 경제파급효과', '백신 기대감', '긍정적 해외 접종 현황', '긍정적 국제경제 파급효과'로 레이블링 되었다. 부정적 게시물은 다섯 개의 토픽은 각각 '해외 접종 후유증', '접종 정책 비판', '백신 불신', '접종 불안', '해외 접종 관련사건사고'로 레이블링 되었다.

<그림 4>는 토픽 모델링 결과를 토대로 화이자 긍정/부정 게시물의 토픽 분포를 시각화한 것이다. 화이자 긍정 게시물에서 토픽 관련 게시물 순으로 '긍정적 해외 접종 현황'(602개), '긍정적 국내 경제 파급효과'(504개), '백신 기대감'(459개), '긍정적 국제경제 파급효과'(267개) 등이 화이자에 대한 긍정적인 의견을 내포한 토픽들로 판단하였다. 해외에서 화이자 백신의 임상실험 및 접종을 통해 나타난 백신의

<표 3> 화이자 긍정/부정 게시물 토픽, 레이블 및 키워드

긍정/부정	토픽	레이블	키워드
긍정	1	백신 수급 상황	도착, 유통, 수송, 물량
	2	백신 국제도입 현황	확보, 유럽, 러시아, 싱가포르
	3	긍정적 국내 경제 파급효과	좋은, 투자, 증가, 강세, 증시, 급등
	4	백신 기대감	승인, 안정성, 임상실험, 안정, 검증, 안전, 기대
	5	긍정적 해외 접종 현황	효과, 이스라엘, 노르웨이, 빛, 면역, 가능성, 긍정, 남아공, 스코틀랜드
	6	긍정적 국제경제 파급효과	미국, 독일, 중국, 나스닥, 시장, 다우,
부정	1	해외 접종 후유증	독일, 영국, 이스라엘, 포르투갈, 싱가포르, 심장마비, 후유증, 불안, 멕시코
	2	접종 정책 비판	재앙, 난리, 정권, 유통, 농락, 차별, 불안감
	3	백신 불신	사망, 사망자, 불신
	4	접종 불안	부작용, 알레르기, 걱정, 우려
	5	해외 접종 관련사건사고	주사기, 사건, 사고, 세계, 감염



〈그림 4〉 주제 별 화이자 긍정/부정 게시물 분포

효용성이 입증됨에 따라, 트위터에서 백신에 대한 기대감 증가 및 백신 보급에 따른 코로나 바이러스로 인한 경기 침체 완화를 예상하는 것으로 유추할 수 있었다. 부정 게시물의 경우, '해외 접종 관련사건사고'(1,537개)와 '해외 접종 후유증'(874개) 등이 화이자 백신에 대해 트위터 이용자들이 부정적인 의견을 가지는 주원인인 것으로 유추할 수 있었다.

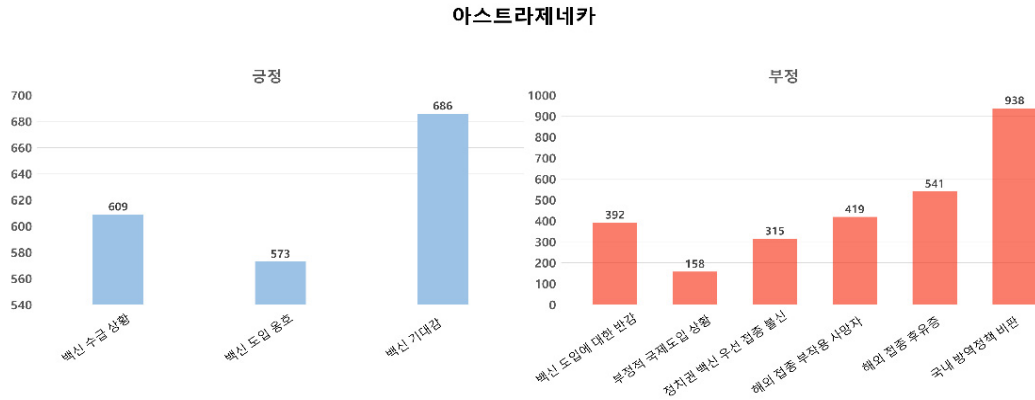
아스트라제네카의 경우 긍정 게시물로부터 세 개, 부정 게시물로부터 여섯 개의 토픽이 추출되었으며, 토픽 레이블 및 토픽 별 주요 키워

드는 〈표 4〉와 같다. '백신 수급 상황', '백신 도입 옹호', '백신 기대감'이 긍정 게시물 토픽의 키워드로, '백신 도입에 대한 반감', '부정적 국제도입 상황', '정치권 백신 우선 접종 불신', '해외 접종 부작용 사망자', '해외 접종 후유증', '국내 방역정책 비판'이 부정 게시물 토픽으로 레이블링 되었다.

〈그림 5〉는 아스트라제네카 관련 긍정/부정 게시물의 토픽 분포를 시각화한 것으로, '백신 기대감'(686개), '백신 도입 옹호'(609개) 등의 아스트라제네카에 대한 긍정적인 토픽들이 파

〈표 4〉 아스트라제네카 긍정/부정 게시물 토픽, 레이블 및 키워드

긍정/부정	토픽	레이블	키워드
긍정	1	백신 수급 상황	생산, 예방접종, 확보, 유통
	2	백신 도입 옹호	정부, 좋은, 방역, 신뢰, 안정, 우리나라, 희망
	3	백신 기대감	접종, 예방효과, 성공, 효능, 승리, 안전
부정	1	백신 도입에 대한 반감	거짓, 기피, 문제, 재앙, 정부, 가짜, 도입,
	2	부정적 국제도입 상황	유럽, 찬밥신세, 오스트리아, 금지, 이탈리아, 쓰레기, 투약
	3	정치권 백신 우선 접종 불신	접종, 대통령, 먼저, 정치, 불신, 불안감, 거짓말, 민주당
	4	해외 접종 부작용 사망자	부작용, 사망, 질환, 혈전, 사망자
	5	해외 접종 후유증	근육통, 증상, 후기, 몸살, 두통, 열, 오한, 통증, 걱정, 이쁨, 관절통
	6	국내 방역정책 비판	무능한 정권, 거부, 풀찌, 논란, 예방접종



〈그림 5〉 주제 별 아스트라제네카 긍정/부정 게시물 분포

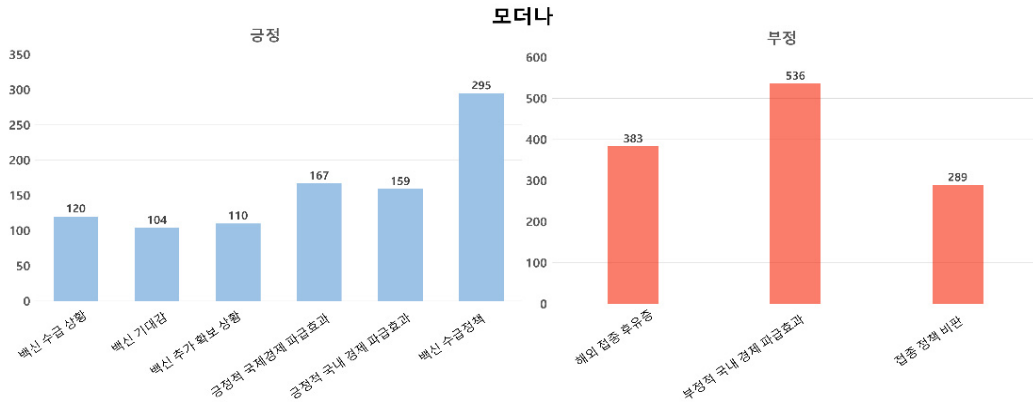
악되었다. 아스트라제네카의 효능에 대한 트위터에서의 기대감을 파악할 수 있었으며, 이러한 기대감으로부터 백신의 수급 상황에 대해 관심을 가지는 것으로 유추할 수 있었다. 부정적인 게시물로부터 '해외 접종 후유증'(541개), '해외 접종 부작용 사망자'(419개), '백신 도입에 대한 반감'(392개), '부정적 국제도입 상황'(158개) 등이 아스트라제네카와 직접적인 관련이 있는 부정적인 토픽들로 파악되었으며, 해외에서 발생된 접종 관련 부정적인 사건들이 트위터상에서 아스트라제네카에 대한 부정적인 의견 형성에

영향을 미쳤음을 알 수 있었다.

모더나는 긍정 게시물로부터 여섯 개, 부정 게시물로부터 세 개의 토픽이 추출되었으며, 토픽 레이블 및 토픽 별 주요 키워드는 〈표 5〉와 같다. 긍정 게시물의 여섯 개의 토픽은 '백신 수급 상황', '백신 기대감', '백신 추가 확보 상황', '긍정적 국제경제 파급효과', '긍정적 국내 경제 파급효과', '백신 수급정책'으로 레이블링되었다. 부정적인 게시물은 '해외 접종 후유증', '부정적 국내 경제 파급효과', '접종 정책 비판'으로 토픽 레이블링되었다.

〈표 5〉 모더나 긍정/부정 게시물 토픽, 레이블 및 키워드

긍정/부정	토픽	레이블	키워드
긍정	1	백신 수급 상황	생산, 접종, 확보, 위탁
	2	백신 기대감	효과, 항체, 예방, 결과, 기대, 성공, 임상실험
	3	백신 추가 확보 상황	나라, 협상, 가격, 대한민국, 추가, 요청, 도스
	4	긍정적 국제경제 파급효과	증시, 강세, 급등, 다우, 지구촌, 뉴욕증시
	5	긍정적 국내 경제 파급효과	주식, 추가, 코스피, 코스닥, 취직, 기회
	6	백신 수급정책	공급, 계약, 대통령, 정부
부정	1	해외 접종 후유증	접종, 부작용, 알레르기, 사망, 질환, 이상, 현기증, 고혈압, 식은땀, 오한
	2	부정적 국내 경제 파급효과	부작용, 추가, 급락, 증시, 곤두박질
	3	접종 정책 비판	방역, 재앙, 문제, 우리나라



〈그림 6〉 주제 별 모더나 긍정/부정 게시물 분포

모더나 긍정/부정 게시물 토픽 분포를 시각화한 것은 〈그림 6〉과 같다. 〈그림 6〉에 따르면, 게시물 순으로 ‘긍정적 국제경제 파급효과’(167개), ‘긍정적 국내 경제 파급효과’(159개), ‘백신 기대감’(104개) 등의 모더나에 대한 긍정적인 토픽들이 파악되었다. 이를 통해 모더나 접종으로부터 발생하는 코로나바이러스 확산 방지에 따른 긍정적인 효과들이 기대되고 있음을 유추할 수 있었다. 반면에 부정적인 게시물의 경우 ‘부정적 국내 경제 파급효과’(536개), ‘해외 접종 후유증’(383개) 등이 모더나와 직접적인 연관이 있는 부정적인 토픽들로 파악되었으며, 코로나바이러스의 장기화에 따른 경기 침체와 모더나 백신 접종 후유증에 대한 부정적인 의견이 트위터상에서 나타난 것을 파악할 수 있었다.

5. 결론 및 제언

본 연구는 코로나바이러스 백신에 대한 소셜 미디어 이용자들의 의견을 파악하기 위해 2020

년 3월 16일부터 2021년 3월 15일까지 트위터에서 작성된 42,261개의 백신 관련 게시물들을 수집하였다. 수집된 게시물들은 게시물 수 기준 상위 백신 관련 게시물(42,261개)과 하위 백신 관련 게시물(3,152개)로 나누어 전처리 및 분석을 진행하였다. 게시물 수 기준 하위 백신 관련 게시물의 전처리 및 분석은 연구자들의 수기를 통해 진행되었으며, 이를 통해 백신에 대한 트위터 이용자들의 감성 분석 및 긍정/부정 요인을 파악하고자 하였다. 분석에 사용된 게시물 수 기준으로 시노백(164개), 노바백스(73개), 시노팜(42개) 등의 백신들에 대한 감성 분석 및 키워드 분석을 진행하였으며, 노바백스를 제외한 시노백과 시노팜에 대해서는 트위터 상에서 부정적인 의견이 대다수임을 파악할 수 있었다. 시노백과 시노팜에 부정적인 요소로는 백신의 효능에 대한 불신인 것으로 파악되었다. 노바백스의 경우, 국내에서 생산된다는 점과 백신의 효능에 대한 기대감이 노바백스가 트위터상에서 긍정적인 평가를 받는 요소로 파악되었다.

게시물 수 기준 상위 백신 관련 게시물의 경

우 자동화된 방식을 활용하여 전처리 및 분석 작업을 진행하였다. 전처리 과정 중 안센이 분석 대상에서 제외되었으며, 화이자, 아스트라제네카, 모더나 백신과 관련한 13,507개의 게시물로 감성 분석 및 토픽 모델링을 실시하였다. 감성 분석 결과, 세 개의 백신 모두 트위터상에서 긍정적인 평가보다는 부정적인 평가가 많은 것을 파악할 수 있었다. 백신 별 긍정/부정 평가로 논의되는 주제들을 세부적으로 파악하기 위해 토픽 모델링을 실시하였으며, 경제적 파급효과, 백신 효능, 접종 후유증, 접종 관련 사건 사고와 같은 다양한 주제들이 논의됨을 알 수 있었다. 해외에서 검증된 백신의 효능에 따른 백신에 대한 기대감이 세 백신 모두에서 논의되고 있었으며, 화이자와 모더나에서는 코로나바이러스로 인한 경기 침체 완화가 긍정적으로 논의되고 있음을 알 수 있었다. 반면에, 부정적인 의견의 경우 해외에서 발견된 접종 후유증이 세 백신 모두에서 논의되고 있음을 알 수 있었다. 화이자의 경우 접종 후유증에 의한 백신에 대한 불신 및 불안감이 세분화된 주제로 논의되었으며, 아스트라제네카는 접종 후유증으로 인한 접종 규제가 아스트라제네카에 대한 부정적인 의견으로 논의되고 있음을 알 수 있었다. 마지막으로 모더나의 경우, 접종 후유증이라는 부정적인 영향이 경제 상황에 악영향을 끼친 것이 논의되었음을 알 수 있었다.

본 연구에는 다음과 같은 한계점이 존재한다. 본 연구는 코로나바이러스 백신 접종이 시행되기 전의 시점부터 백신 접종이 시작된 지 얼마 되지 않은 시기까지의 게시물을 대상으로

분석을 진행하였다. 이로 인해 이후 작성된 트위터상에서 백신과 관련된 더 많은 수의 게시물들을 분석 대상에 포함하지 못하였다. 그리하여 실제 대규모 백신접종시의 상황보다는 백신 접종 전 또는 초기의 의견을 탐색하였다. 따라서 추후 연구에서는 백신 접종이 활성화된 시기까지의 게시물을 토대로 백신에 대한 소셜 미디어 이용자들의 보다 더 심도 있는 의견을 탐색하고자 한다.

본 연구는 기존의 선행연구에서 다루지 않았던 소셜 미디어 데이터를 활용하여 국내에서 논의되는 코로나바이러스 백신과 관련된 사회적 의견을 분석하고자 하였다. 이를 위해 게시물 수 기준 상위, 하위 백신으로 나누어 각 데이터에 맞는 분석방법을 활용하였으며, 백신 별 감성 분석 및 토픽 모델링을 활용한 주제 추출 등을 통해 코로나바이러스 백신들에 대한 대중들의 부정적, 긍정적인 의견을 탐색하는 것에 학술적 함의를 가진다. 또한, 본 연구의 결과를 통해 백신 접종 후유증에 대한 불안함과 백신의 효능에 대한 불신과 같은 원인들이 백신 접종을 저해하는 원인이 될 수 있다는 것을 유추할 수 있었다. 반면에, 백신 효능에 대한 기대감이 백신에 대한 긍정적인 요소로 작용할 가능성을 유추할 수 있었다. 이를 통해 대중들의 백신에 대한 불안함 등 부정적인 요인들을 완화하고, 안전성 및 효능에 대해 대중들에게 인식시킬 수 있는 캠페인 등의 백신 접종 권장 정책 수립의 필요성을 제안하는 실질적 함의가 기대된다.

참 고 문 헌

- 김서영 (2021. 2. 21). '집단면역 70%' 달성하려면... "사실상 전국민 90% 백신 접종해야". 연합뉴스, 출처: <https://www.yna.co.kr/view/AKR20210219114600530>
- 김장현 (2021. 11. 6). 신규 확진 2,248명...나흘 연속 2천명대 확산. 연합뉴스TV, 출처: <https://www.yonhapnewstv.co.kr/news/MYH20211106003000641?did=1825m>
- 김태중 (2020). 뉴스 빅데이터를 활용한 코로나19 언론보도 분석: 토픽모델링 분석을 중심으로. 한국콘텐츠학회논문지, 20(5), 457-466. <http://doi.org/10.5392/JKCA.2020.20.05.457>
- 박상민, 나철원, 최민성, 이다희, 온병원 (2018). Bi-LSTM 기반의 한국어 감성사전 구축 방안. 지능정보연구, 24(4), 219-240. <http://doi.org/10.13088/jiis.2018.24.4.219>
- 박은정, 조성준 (2014). KoNLPy: 쉽고 간결한 한국어 정보처리 파이썬 패키지. 제26회 한글 및 한국어 정보처리 학술대회 논문집, 춘천, 133-136.
- 서하림, 송민 (2019). 소셜미디어를 통한 우울 경향 이용자 담론 주제 분석. 정보관리학회지, 36(4), 207-226. <http://doi.org/10.3743/KOSIM.2019.36.4.207>
- 성백린 (2021). 코로나19 백신 연구와 개발. 지식의 지평, 30, 117-127.
- 유소연, 임규건 (2021). 텍스트 마이닝과 의미 네트워크 분석을 활용한 뉴스 의제 분석: 코로나 19 관련 감정을 중심으로. 지능정보연구, 27(1), 47-64. <http://doi.org/10.13088/jiis.2021.27.1.047>
- 주영준, 김동훈, 이창호, 이용정 (2019). 페이스북 그룹 게시물 분석을 통한 우울증 관련 주제에 대한 고찰. 한국문헌정보학회지, 53(4), 171-187. <http://doi.org/10.4275/KSLIS.2019.53.4.171>
- 중앙방역대책본부 (2021). 코로나 19 국내 검사 및 확진자 현황. 출처: <http://ncov.mohw.go.kr/>
- 진설아, 허고은, 정유경, 송민 (2013). 트위터 데이터를 이용한 네트워크 기반 토픽 변화 추적 연구. 정보관리학회지, 30(1), 285-302. <http://doi.org/10.3743/KOSIM.2013.30.1.285>
- 최성민 (2021). 팬데믹의 현재와 백신의 미래. 오토피아, 36(1), 5-38. <http://doi.org/10.32355/oughtopia.2021.06.36.1.5>
- 최원주, 홍장선 (2021). 코로나19에 대한 국내 언론보도 검색키워드 패턴 연구. 커뮤니케이션학 연구, 29(2), 29-58. <http://doi.org/10.23875/kca.29.2.2>
- Blei, D. M., Ng, A. Y., & Jordan, M. (2003). "Latent Dirichlet Allocation." *Journal of Machine Learning Research*, 3(4/5), 993-1022.
- Domenico, C. & Maurizio, V. (2020). WHO declares COVID-19 a pandemic. *Acta Bio Medica: Atenei Parmensis*, 91(1), 157-160. PubMed. <http://doi.org/10.23750/abm.v91i1.9397>
- Härdle, W., Chen, C. H., & Overbeck L. eds. *Applied Quantitative Finance. Statistics and Computing*.

- JustAnotherArchivist (2020). snsrape 0.3.4. <https://github.com/JustAnotherArchivist/snsrape>
- Lara, T., Filippo, Q., Eleonora, D., Pietro, D., Marco, V., Francesco, M., & Luigi, L., Pier. (2020). Twitter as a sentinel tool to monitor public opinion on vaccination: an opinion mining analysis from September 2016 to August 2017 in Italy. *Human Vaccines & Immunotherapeutics*, 16(5), 1062-1069. PubMed. <http://doi.org/10.1080/21645515.2020.1714311>
- Linton, M., Teo, E., Bommers, E., Chen, C., & H. Wolfgang Karl. (2017). *Dynamic Topic Modelling for Cryptocurrency Community Forums*. Springer, Berlin, Heidelberg. http://doi.org/10.1007/978-3-662-54486-0_18
- Lucia, P. S., Manuel, T., Juan, Diego, F. P.-B., Almudena, J., Manuel, C., Ernestina, M., Antonio, C. F., Amalia, A., Angel, G. de M., & Alejandro, R. G. (2021). Influenza and Measles-MMR: two case study of the trend and impact of vaccine-related Twitter posts in Spanish during 2015-2018. *Human Vaccines & Immunotherapeutics*, 1-15. <http://doi.org/10.1080/21645515.2021.1877597>
- Mathieu, E., Ritchie, H., Ortiz-Ospina, E., Roser, M., Appel, C., Giattino, C., & Rodes-Huirao, L. (2021). A global database of COVID-19 vaccinations. *Nat Hum Behav*. <http://doi.org/10.1038/s41562-021-01122-8>
- Rehurek, R. & Sojka, P. (2011). Gensim--python framework for vector space modelling. NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic, 3(2).
- Samira, Y., Rozita, D., Samira, M., Andrew, P., & Shayan, S. (2021). An analysis of COVID-19 vaccine sentiments and opinions on Twitter. *International Journal of Infectious Diseases*, 108, 256-262. <http://doi.org/10.1016/j.ijid.2021.05.059>
- Worldometers.info. (2021). Coronavirus Cases - Worldometer. Available: <https://www.worldometers.info/>

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

- Central Quarantine Countermeasure Headquarters (2021). COVID-19 domestic testing and confirmed cases. <http://ncov.mohw.go.kr/>
- Choi, S. M. (2021). The present and future of vaccines for pandemic. *Oughtopia*, 36(1), 5-38. <http://doi.org/10.32355/oughtopia.2021.06.36.1.5>
- Choi, W. J. & Hong, J. S. (2021). A study on the search keyword pattern of COVID-19 in the domestic media. *Korean Journal of Communication Studies*, 29(2), 29-58.

<http://doi.org/10.23875/kca.29.2.2>

- Jin, S. A., Heo, G. E., Jeong, Y. K., & Song, M. (2013). Topic-Network based topic shift detection on twitter. *Journal of the Korean Society for Information Management*, 30(1), 285-302. <http://doi.org/10.3743/KOSIM.2013.30.1.285>
- Kim, J. H. (2021, November 6). 2,248 new confirmed cases... Spread of 2,000 people for four days in a row. *Yonhap News TV*. Available: <https://www.yonhapnewstv.co.kr/news/MYH20211106003000641?did=1825m>
- Kim, S. Y. (2021, February 21). To achieve '70% herd immunity'... "Actually, 90% of the nation should be vaccinated". *Yonhap News*, Available: <https://www.yna.co.kr/view/AKR20210219114600530>
- Kim, T. J. (2020). COVID-19 news analysis using news big data: focusing on topic modeling analysis. *The Journal of the Korea Contents Association*, 20(5), 457-466. <http://doi.org/10.5392/JKCA.2020.20.05.457>
- Park, E. J. & Cho, S. J. (2014). KoNLPy: Korean natural language processing in python. In *Proceedings of the 26th Annual Conference on Human and Cognitive Language Technology*, Chuncheon, 133-136.
- Park, S. M., Na, C. W., Choi, M. S., Lee, D. H., & On, B. W. (2018). KNU Korean sentiment lexicon - Bi-LSTM-based method for building a Korean sentiment lexicon. *Journal of Intelligence and Information Systems*, 24(4), 219-240. <http://doi.org/10.13088/jiis.2018.24.4.219>
- Seo, H. R. & Song, M. (2019). An analysis of the discourse topics of users who exhibit symptoms of depression on social media. *Journal of the Korean Society for Information Management*, 36(4), 207-226. <http://doi.org/10.3743/KOSIM.2019.36.4.207>
- Seong, B. L. (2021). COVID-19 vaccine research and development. *Orbis Sapientiae*, 30, 117-127.
- Yoo, S. Y. & Lim, G. G. (2021). Analysis of news agenda using text mining and semantic network analysis: focused on COVID-19 emotions. *Journal of Intelligence and Information Systems*, 27(1), 47-64. <http://doi.org/10.13088/jiis.2021.27.1.047>
- Zhu, Y. J., Kim, D. H., Lee, C. H., & Yi, Y. J. (2019). Investigating major topics through the analysis of depression-related facebook group posts. *Journal of the Korean Library and Information Science*, 53(4), 171-187. <http://doi.org/10.4275/KSLIS.2019.53.4.171>