

보건복지 ISSUE & FOCUS



제376호 (2020-07)
발행일 2020. 03. 17.
ISSN 2092-7117

발행인 조흥식 발행처 한국보건사회연구원 (30147) 세종시 시청대로 370 세종국책연구단지 사회정책동(1~5층) T 044)287-8000 F 044)287-8052

코로나바이러스감염증-19 특집호 ④

코로나바이러스감염증-19 소셜 빅데이터 기반 주요 이슈 분석



오미애
정보통계연구실 빅데이터·정보연구센터장
전진아
보건정책연구실 건강정책연구센터장

- 2020년 1월 20일 국내에서 첫 확진자가 발생한 후 코로나바이러스감염증-19(이하, 코로나19)가 확산되면서 국민의 일상생활부터 국가와 경제·사회 전반에 걸쳐 영향을 미치고 있음. 이 글에서는 코로나19 관련 언론 보도 및 댓글들(2020년 1월 20일~2020년 3월 8일)을 분석하여 시기별로 발생한 이슈들을 살펴보았음.
- 빈도 분석 및 네트워크 분석 결과, 중국과 관련된 키워드 중심에서 대구, 인천지 키워드 중심으로 변화하고 지역사회 2차 감염 관련 키워드로 확산되는 양상이 보임.
- 코로나19에 신속하게 대응하기 위해 현상을 전체적으로 모니터링하고 방향을 설정할 수 있는 데이터 기반 시스템 체계가 필요함.

01. 들어가며

◆ 2020년 1월 20일 국내에서 첫 확진자가 발생한 후 코로나19가 확산되면서 국민의 일상생활부터 국가와 경제·사회 전반에 걸쳐 영향을 미치고 있음. 이 글에서는 코로나19 관련 언론 보도 및 댓글들을 분석하여 시기별로 발생한 이슈들을 살펴보고 정리해 보고자 함.

- 코로나19로 인해 지역별 의료진, 의료시설 등의 의료자원 공급과 배분 문제, 마스크 사재기, 소비 위축에 따른 자영업 위기, 감염병 불안에 따른 휴원, 휴교 조치 등 다양한 문제들이 복합적으로 발생하고 있어 국민이 느끼는 불안감이 고조·확산되고 있음.

- 이러한 현상은 온라인상의 뉴스, 블로그, 카페, SNS, 커뮤니티 등 소셜 미디어를 통해 공유·확산되고 있음.
 - 코로나19 관련 소셜 빅데이터 분석을 통해 사람들 간에 주로 논의된 이슈를 시간 흐름에 따라 파악할 수 있고 국민들의 반응을 다각도로 분석할 수 있음.
- ◆ 소셜 미디어상의 코로나19에 대한 버즈양¹⁾은 증가하는 추세를 보이며, 확진자가 누적 증가하면서 코로나19에 대해 공유·확산되는 버즈양도 함께 증가하고 있음.
- 2020년 1월 20일 국내 첫 번째 확진자가 발생한 이후 코로나19에 대한 버즈양이 증가세를 보이며, 2020년 2월 18일 31번째 확진자가 발생했다는 정부의 발표 이후 매우 많은 버즈양 증가폭을 보임.
 - 확진자 수 발표, 신천지 집단 감염, 마스크 판매 이슈 등이 버즈양 증가를 견인한 것으로 보임. 이에 따라 다양한 이슈들이 공유되고, 외부 활동 자제, 마스크 판매 급증, 사재기, 개학 연기, 오프라인 매출 감소 등 다양한 사회·경제적 파급효과를 불러옴.

02. 분석 데이터 및 분석 방법

◆ 데이터 수집 및 분석 방법

- 데이터 수집 방법
 - 수집 키워드: ‘코로나’, ‘우한’
 - *코로나19 발생 초기, ‘우한 폐렴’이라는 표현을 사용하여 ‘우한’이라는 키워드를 포함함.
 - 2020년 1월 20일(국내 첫 확진자 발생일)부터 2020년 3월 8일까지 뉴스, 블로그, 카페, SNS, 커뮤니티에서 수집 키워드 관련 데이터(문서/댓글) 약 260만 건을 수집함.²⁾

〈표 1〉 코로나 관련 데이터의 주차별 문서, 댓글 버즈양

(단위: 천 건)

구분	2020. 1. 20. ~1. 26. (1주차)	2020. 1. 27. ~2. 2. (2주차)	2020. 2. 3. ~2. 9. (3주차)	2020. 2. 10. ~2. 16. (4주차)	2020. 2. 17. ~2. 23. (5주차)	2020. 2. 24. ~3. 1. (6주차)	2020. 3. 2. ~3. 8. (7주차)	합계
문서	26	149	210	171	236	421	412	1,625
댓글	34	171	117	68	264	209	177	1,040

1) 버즈양은 어떤 주제에 대한 온라인상에서의 언급 횟수를 의미함.

2) 네이버 뉴스, 다음 뉴스, 다음 카페, 다음 블로그, 트위터, 커뮤니티 등.

- ‘코로나’ 키워드로 수집된 빅데이터를 활용하여 시계열 버즈양 분석, 연관어 분석, 네트워크 분석 방법을 적용하였음.
 - 일별로 ‘코로나’ 단어를 포함하고 있는 버즈양을 통해 코로나19에 대한 관심 정도와 이슈 정보들을 파악하고, 버즈양이 많은 날의 이슈를 파악함.
 - 자연어 처리 방법을 적용하여 자주 등장한 연관어 빈도 분석과 키워드 간의 연결 관계를 파악하는 네트워크 분석을 통해 다각도로 분석함.
 - (데이터별) 분석 데이터에 대해 분석 키워드인 ‘코로나’를 포함하는 문서³⁾(본문)와 그 문서에 달린 댓글⁴⁾을 따로 분석하여, 정보 전달 위주의 문서와 이에 대한 반응으로 논의·공유되는 댓글 데이터를 분석함.
 - (시기별) 주차별로 상위(핵심) 빈도를 갖는 단어를 중심으로 비교하여 시간 흐름에 따라 어떻게 이슈들이 변화했는지 파악함.

03. 소셜 빅데이터 분석

가. 시계열 버즈양

◆ 코로나19 확진자 수 대비 코로나19 관련 소셜 미디어 버즈양 비교 분석

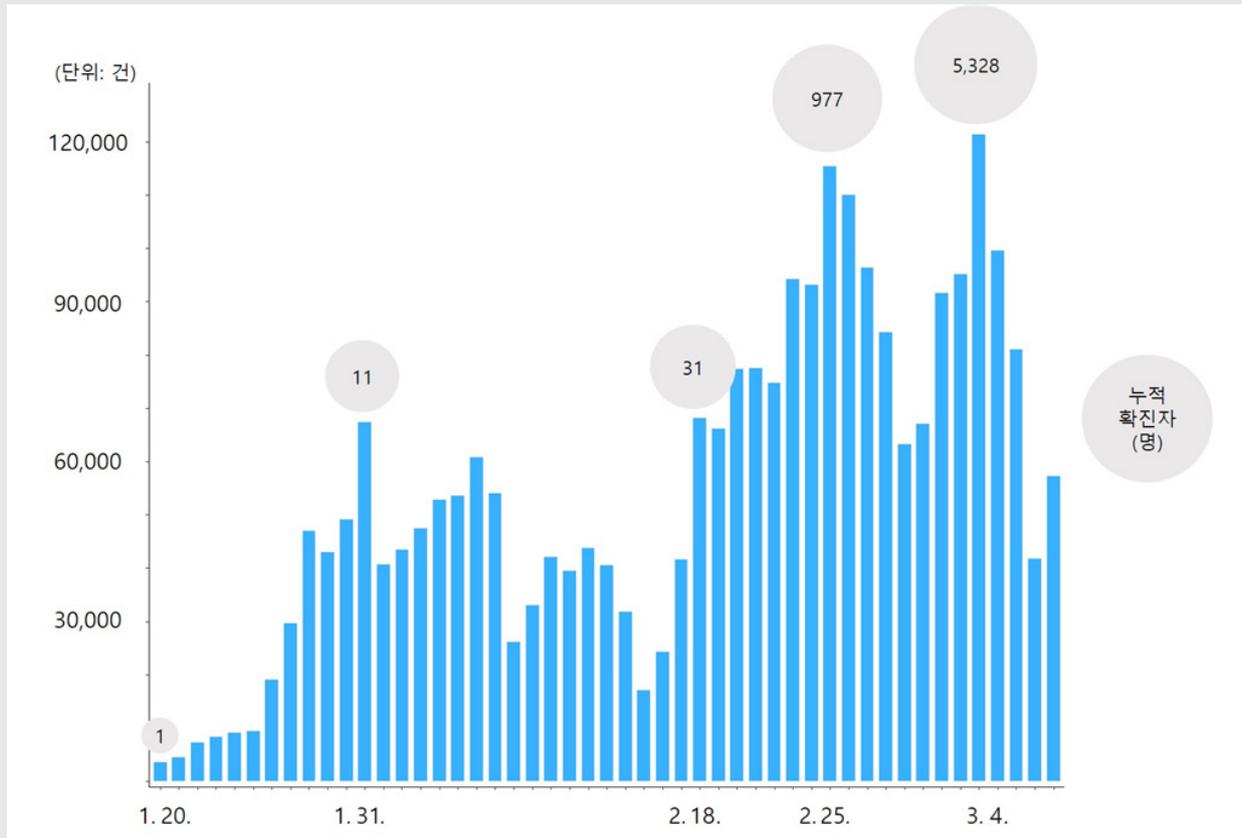
- 코로나19 확진자 수 추이와 코로나19 관련 데이터양 비교 분석 결과는 다음과 같음.
 - 2019. 12. 31. 국내에서 처음으로 ‘우한 폐렴’이라는 이름으로 코로나19 관련 보도가 이루어진 후, 2020. 1. 20. 국내 첫 확진자 발생 전까지 국내에서는 큰 반응이 없었음.
 - 2020. 1. 20. 국내 첫 확진자가 발생한 후, 과거 대비 버즈양이 급증함으로써 다수의 사람들이 이전보다 관심을 가지기 시작함.
 - 2020. 1. 31. 국내 확진자가 10명을 넘어서면서 버즈양이 증가하였고, 이후 국내 확진자 수가 큰 증가를 보이지 않으며 버즈양 또한 감소하는 양상을 보임.
 - 2020. 2. 18. 국내에서 31번째 확진자가 발생하면서 버즈양이 증가함.
 - 2020. 2. 20. 국내 첫 사망자 발생 및 국내 확진자 수가 100명 이상 증가한 시점에 다시 한번 버즈양이 급증하였고, 확진자 수와 버즈양 증가 추이가 궤를 같이하는 것을 확인할 수 있었음.

.....

3) 여기에서의 문서는 뉴스, 블로그, 카페, SNS, 커뮤니티의 본문을 의미함.

4) 여기에서의 댓글은 수집된 뉴스, 블로그, 카페, SNS, 커뮤니티의 본문에 달린 댓글을 의미함.

[그림 1] 코로나 관련 데이터 버즈양 및 누적 확진자 수(2020년)



자료: 저자 작성, 누적 확진자(명) 자료는 질병관리본부 보도자료 활용.

나. 코로나19 연관어 순위 분석⁵⁾

◆ 코로나 연관어 빈도 순위 분석

- 분석 기간 전체에 등장한 코로나19 연관어 빈도 순위 분석 결과는 다음과 같음.
 - 문서에서는 중국, 확진, 환자, 신천지, 마스크 순으로 높은 빈도를 보였으며, 댓글에서는 중국, 신천지, 중국인, 입국 순으로 빈도가 높음.
 - 문서와 댓글에서 나타난 키워드들을 유형별로 구분해 보면, 코로나19 발생, 유행 및 확산과 관련한 중국, 대구, 신천지가 많이 언급되었으며 코로나19 확진, 환자, 증상에 대한 관심도 높은 것으로 나타남.
 - 마스크 수급과 관련한 지속된 우려를 반영하듯 마스크가 주요한 키워드로 나타나며, 코로나19 유행 대응과 관련한 '정부', '대통령' 키워드 역시 많이 나타남.

5) 코로나19 관련 연관어 빈도는 TF-IDF(Term Frequency - Inverse Document Frequency)값으로 산출하였으며, 산술적 빈도수(Term Frequency)에 따라 선별된 단어가 아님.

〈표 2〉 분석 기간 전체 코로나19 연관어 빈도 순위

빈도 순위	문서	댓글
1	중국(인)	중국(인)
2	확진	신천지
3	환자	입국
4	신천지	정부
5	마스크	대구
6	교회	대통령
7	대구	마스크
8	대통령	확진
9	정부	환자
10	증상	격리
11	일본	검사
12	병원	일본
13	격리	치료
14	입국	병원
15	검사	경제

자료: 저자 작성.

◆ 주차별 코로나 연관어 빈도 순위

- 분석 기간 주차별 코로나 연관어 빈도 순위 분석 결과는 다음과 같음.
 - 문서와 댓글의 연관어 빈도 순위를 보면, 코로나19가 국내에서 대구와 경북 지역을 중심으로 확산되기 전에는 중국이 가장 높은 빈도를 보였으나, 코로나19가 신천지 신도 중심으로 확산됨에 따라 신천지가 가장 높은 빈도를 보임.
 - 5주차(2020. 2. 17.~2. 23.)에는 2월 18일 대구에서 31번 확진자가 발생(해외여행력 없음)함으로써 확진자가 2월 17일 30명에서 2월 23일 602명으로 급증하는 상황이 발생함. 또한 2월 20일 청도 대남병원에서 첫 사망자가 발생하였고, 2월 23일 청도 대남병원에서 6번째 사망자가 발생함.
 - 또한 국내에서 코로나19가 확산되기 전에는 중국, 일본, 교민, 크루즈 등의 이슈가 부각되었으나, 확산된 이후에는 개인 보건위생과 관련한 손, 물 등에 대한 키워드가 높은 순위를 보이고 있음. 이는 코로나19 유행 초기에는 코로나19를 ‘나’ 혹은 ‘우리’의 문제로 바라보기보다 ‘중국’ 혹은 ‘일본’, ‘크루즈’에서 유행하는 문제로 보아 객관적 입장을 견지하였다면, 국내 확산이 본격화되면서 코로나19를 ‘나’ 혹은 ‘우리’의 문제로 여기게 되었음을 의미함.

- 실제로 코로나19를 예방하고자 예방 소독 및 물청소,⁶⁾ 손 씻기⁷⁾ 등의 행동이 강조되었으며, 심지어 뜨거운 물을 마시면 코로나19가 예방 된다는 가짜 뉴스⁸⁾도 공유되었음.
- ‘마스크’는 분석 대상 기간에 지속적으로 높은 빈도를 보였는데, 이는 예방에 대한 국민의 관심 급증과 마스크 수급 부족 등을 의미하는 것일 수 있음.

〈표 3〉 문서 대상 주차별 연관어 빈도 순위

순위	2020. 1. 20. ~ 1. 26. (1주차)	2020. 1. 27. ~ 2. 2. (2주차)	2020. 2. 3. ~ 2. 9. (3주차)	2020. 2. 10. ~ 2. 16. (4주차)	2020. 2. 17. ~ 2. 23. (5주차)	2020. 2. 24. ~ 3. 1. (6주차)	2020. 3. 2 ~ 3. 8. (7주차)
1-5 순위	중국 환자 확진 증상 병원	중국 환자 확진 마스크 증상	중국 확진 환자 마스크 증상	중국 일본 확진 환자 대통령	신천지 확진 중국 교회 대구	신천지 중국 확진 대구 마스크	신천지 중국 교회 대구 확진
6-10 순위	미국 투자 증시 북한 질병관리본부	교민 정부 격리 전세기 접촉	정부 격리 일본 면역력 병원	마스크 정부 경제 교민 코로나	환자 대통령 마스크 병원 정부	교회 입국 대통령 정부 환자	마스크 대통령 예배 정부 지원
11-15 순위	확산 마스크 사망자 격리 대통령	박쥐 예방 병원 시설 손	대통령 검사 코로나 사망자 미국	격리 미국 사망자 병원 지역	손 일본 물 지역 예배	검사 병원 예배 금지 격리	환자 입국 일본 검사 미국

6) 홍해인. (2020. 2. 20.). 광화문광장 코로나19 예방 소독 및 물청소. 연합뉴스. <https://www.yna.co.kr/view/PYH20200220142900013?input=1196m>에서 2020. 3. 13. 인출.

7) 신호정. (2020. 2. 23.). [따람의 피부읽기] 신종 코로나에 대응하는 올바른 손 씻기. 스포츠경향 http://sports.khan.co.kr/bizlife/sk_index.html?art_id=202002231637003&sec_id=561801&pt=nv#csidx0b21da1a9fb62929d8d0030db5b7275에서 2020. 3. 13. 인출.

8) 임재희. (2020. 2. 21.). 뜨거운 물 마시면 코로나19 예방?...가짜 권고안 주의. 뉴시스. http://www.newsis.com/view/?id=NISX20200221_0000927254&clD=10201&plD=10200에서 2020. 3. 13. 인출.

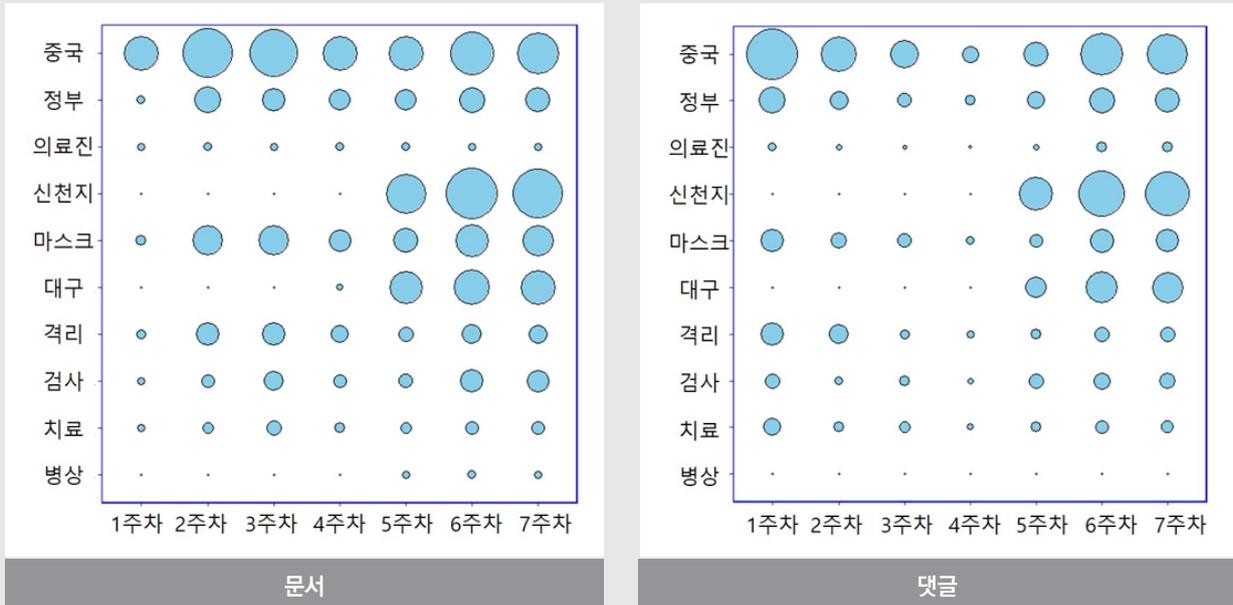
〈표 4〉 댓글 대상 주차별 연관어 빈도 순위

순위	2020. 1. 20. ~1. 26. (1주차)	2020. 1. 27. ~2. 2. (2주차)	2020. 2. 3. ~2. 9. (3주차)	2020. 2. 10. ~2. 16. (4주차)	2020. 2. 17. ~2. 23. (5주차)	2020. 2. 24. ~3. 1. (6주차)	2020. 3. 2 ~3. 8. (7주차)
1-5 순위	중국(인) 입국 정부 확진 마스크	중국(인) 입국 격리 독감 정부	중국(인) 마스크 입국 정부 대통령	중국(인) 일본 정부 경제 확진	신천지 중국(인) 대구 정부 입국	신천지 중국(인) 대구 입국 정부	신천지 마스크 중국(인) 대구 돈
6-10 순위	격리 독감 교민 환자 일본	교민 대통령 확진 마스크 미국	확진 사망자 치료 환자 검사	대통령 마스크 입국 교민 환자	대통령 확진 교회 검사 환자	대통령 마스크 확진 봉쇄 확산	정부 대통령 추경 확진 이만희
11-15 순위	메르스 미국 사망자 치료 지역	메르스 환자 지역 정권 금지	일본 독감 병원 메르스 격리	격리 사망자 메르스 크루즈 병원	이만희 마스크 경제 병원 종식	검사 교회 종식 환자 금지	입국 세금 검사 의사 일본

◆ 주차별 코로나 연관어 빈도 순위

- 문서, 댓글에서 지속적으로 등장하는 키워드인 중국, 정부, 마스크, 신천지, 대구와 더불어 빈도 순위는 높지 않지만 코로나19 유행 대응과 관련한 주요 키워드인 의료진, 격리, 검사, 치료, 병상 키워드에 대해 주차별로 빈도의 증감을 살펴봄.
 - 문서와 댓글 모두 해당 키워드의 증감 양상은 유사하며 신천지, 대구는 코로나19가 신천지 신도 중심으로 확산됨에 따라 이후 새롭게 높은 빈도의 키워드로 등장하였고, 코로나19 확산 이후, 정부, 마스크 키워드 또한 높은 빈도를 보임.

[그림 2] 주차별 선정 키워드 빈도 변화

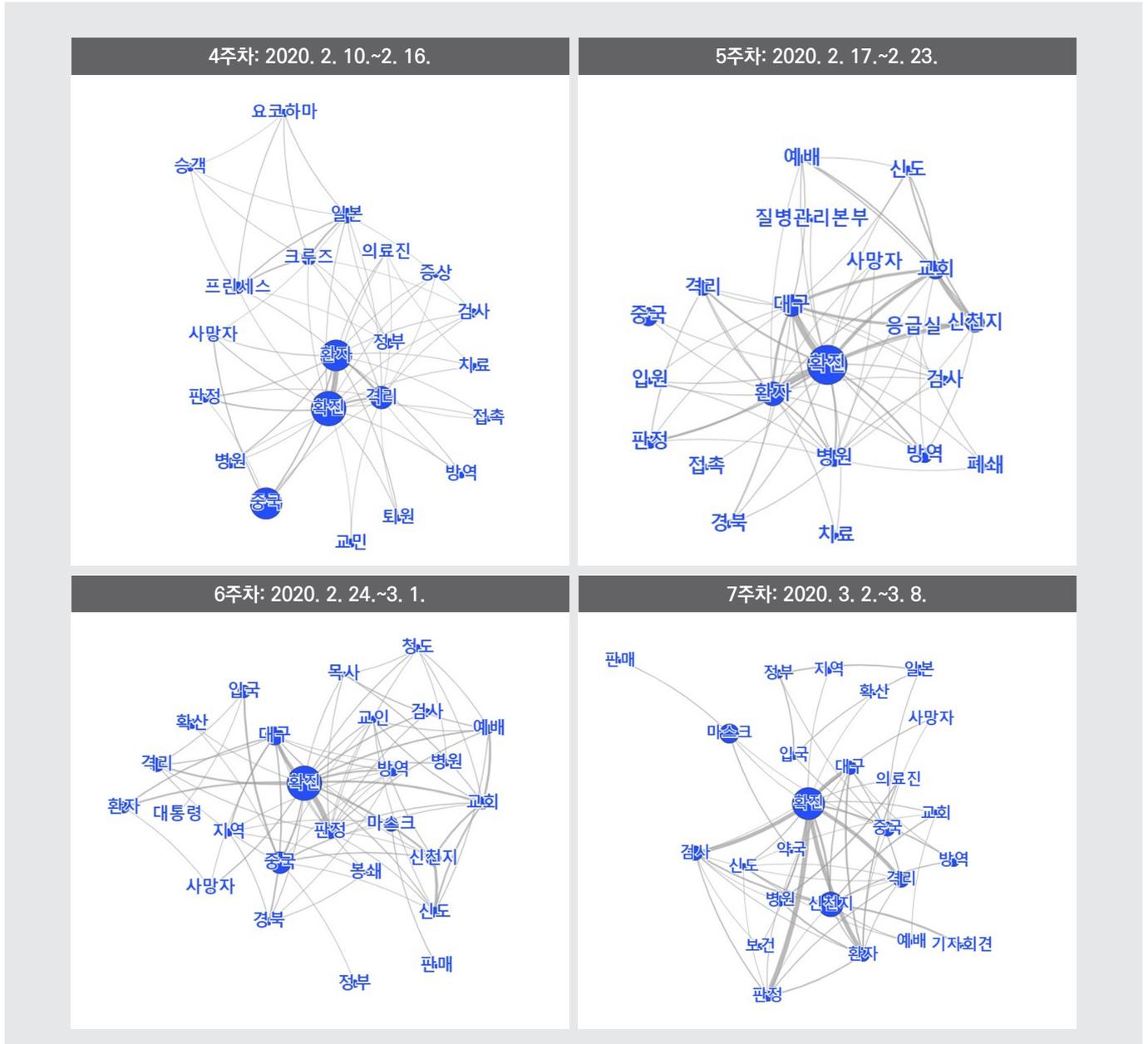


자료: 저자 작성.

다. 코로나19 연관어 네트워크 분석

◆ 연관어 네트워크 분석 결과

- 코로나19와 연관된 키워드들로 네트워크 분석을 실시한 결과는 다음과 같음.
 - 네트워크에서 각 노드(node)는 단어를 의미하고 에지(edge)는 단어 간의 관계를 나타냄. 단어 간의 관계는 개념적으로 두 단어가 같은 문서에서 많이 나타나면 두 단어 간의 관계가 강한 것으로 이해할 수 있음.
 - 문서 전체의 네트워크 결과를 보면 중국, 확진, 환자, 감염 키워드를 중심으로 방역, 의료진, 등 다양한 이슈와의 연관성이 높게 나타남.
 - 1주차(2020. 1. 20. ~ 1. 26.)에는 중국, 환자, 확진 키워드를 중심으로 코로나19의 증상, 발생 경로, 확산 등의 이슈와 연관성이 높았음.
 - 2주차(2020. 1. 27. ~ 2. 2.)에는 중국 우한에서 전세기로 귀국하는 교민들과 관련된 키워드가 주를 이루었고 정부의 초기 대응과 관련한 방역, 보건소 키워드의 연관성도 높게 나타남.
 - 3주차(2020. 2. 3. ~ 2. 9.)의 네트워크 결과에 보이는 태국, 일본, 크루즈 키워드는 태국을 여행한 한국 여성이 확진 환자로 판정받은 것과, 일본 대형 크루즈선에 한국인 9명이 탑승해 있었다는 사실과 관련이 있음. 중국, 공장, 부품, 가동 키워드는 코로나19 영향으로 중국 부품 업체가 가동을 중단함으로써 해당 부품 재고가 없어 한국의 공장 가동이 중단되는 상황과 관련이 있음.
 - 4주차(2020. 2. 10. ~ 2. 16.)의 네트워크 결과를 보면 여전히 중국, 확진, 환자, 격리 키워드가 중심을 이루면서 일본, 크루즈, 승객, 요코하마, 프린세스 키워드가 도출되었는데, 이는 일본 요코하마에 정박 중인 미국 크루즈선인 다이아몬드 프린세스호의 탑승자 중 355명이 감염된 상황과 관련이 있음.



- 6주차(2020. 2. 24.~3. 1.)는 지역사회 2차 감염이 시작된 때로, 지역 중심, 방역 중심으로 이슈가 변화되었음을 알 수 있음
- 7주차(2020. 3. 2~3. 8.) 네트워크 결과를 보면 마스크, 판매, 약국 키워드와 확진, 검사, 판정 키워드의 연관성이 높게 나타났는데, 이는 정부의 공적 마스크 공급, 한국의 코로나19 대처 및 대구모 검사 진단에 대한 외국의 보도와 관련이 있음.

04. 나가며

- ◆ 빅데이터를 활용하여 코로나19 유행 상황에 대한 키워드들을 분석한 결과, 코로나19의 발생 및 유행 단계에 따라 주요 키워드들이 변화하는 것을 확인할 수 있었음.

- 코로나19가 국내에서 발생한 이후 매일 ‘확진’ 환자에 대한 기사들이 쏟아졌기 때문에 ‘확진’은 주요한 키워드일 수밖에 없음. 하지만 유행 상황에 따라 키워드를 통한 국민 혹은 사회의 관심도가 달라지는 것을 확인할 수 있었음.
 - 발생 및 유행 초기에는 중국, 일본 등 해외에 대한 관심도가 높았던 반면, 코로나19 확산이 본격화됨에 따라 인천지, 대구에 대한 관심도가 높아졌음을 확인할 수 있었음. 또한 대구와 경북 지역 외 17개 시도에서 모두 확진자가 발생한 6주차부터는 ‘지역’이라는 키워드 역시 중요하게 부각되었음.
- 빅데이터 분석을 통해 코로나19가 국민 개개인의 일상에 영향을 미치고 있음을 알 수 있었으며, 정부 대응 과정에 대한 관심도가 높음을 알 수 있었음.
 - 코로나19가 확산됨에 따라 ‘마스크’ 수급 부족 문제가 본격화되었으며, 여전히 다수의 국민은 마스크 부족 문제를 경험하고 있음.
 - 정부 대응 과정에 대한 관심도는 ‘정부’, ‘대통령’이 주요한 키워드로 부각되고 있는 것에서 확인할 수 있음. ‘정부’, ‘대통령’ 키워드는 정부와 대통령에 대한 지지와 비판의 양가적인 입장을 보여 주는 것임. 하지만 코로나19가 국내에서 발생한 이후부터 꾸준히 높은 빈도를 차지하고 주요한 네트워크로 표현되고 있다는 점은 정부와 대통령에 대해 어떠한 입장을 보이든지 간에 우리 사회가 정부의 코로나19 유행 대응에 높은 관심을 가지고 지켜보고 있음을 의미하는 것임.
- ◆ 현재는 전염병 확산 방지를 위해 실시간 지역별 확진자 현황 및 동선 정보 제공 등 객관적인 정보 전달을 위해 노력하고 있지만, 이를 전체적으로 살펴볼 수 있는 데이터 기반 시스템 체계가 필요함.
 - 국토교통부와 과학기술정보통신부, 질병관리본부는 코로나19 확산 방지를 위해 국토부와 과기정통부에서 연구·개발 중인 스마트시티 기술을 활용해 역학조사를 지원하는 시스템을 시범 운영할 계획임.⁹⁾ 이 시스템을 활용하면 확진자 동선을 확인하는 데 걸리는 시간을 평균 5시간에서 10분가량으로 줄일 수 있으며, 신속하게 정보를 전달할 수 있게 됨.
 - 국민의 정서 관리가 필요한 지금과 같은 국가 재난 상황에서는 현재의 이슈를 실시간으로 파악하고 전문가들이 객관적인 데이터를 통해 국민의 반응을 실시간으로 모니터링할 수 있는 시스템이 필요함.

9) 강세훈. (2020. 3. 11.). 확진자 동선 확인, 5시간→10분으로 줄인다…정부, 연계 시스템 개발. 뉴스시. <https://m.news.nate.com/view/20200311n03232?sect=sisa&list=rank&cate=interest>에서 2020. 3. 13. 인출.

극심한 스트레스를 극복하는 것은
쉽지 않은 일이지만,
우리의 마음 속에는
스스로를 치유하는 힘이 있습니다.

- ✓ 생활패턴을 회복하시고
 - ✓ 충분한 수면과 건강한 식사,
 - ✓ 과도한 음주와 카페인 섭취를 제한하고,
 - ✓ 가벼운 운동을 하는 것은
- 스트레스 완화에** 도움이 됩니다.



여러분은 혼자가 아닙니다.

언제나 정신건강 전문가의 도움을
받을 수 있습니다!

확진자 및 가족

- ✓ 02-2204-0001~2 국가트라우마센터
- ✓ 055-520-2777 영남권국가트라우마센터

격리자 및 일반인

- ✓ 1577-0199 정신건강복지센터