ORIGINAL ARTICLE

J Korean Acad Community Health Nurs (지역사회간호학회지)

ISSN 1225-9594 (Print) / ISSN 2288-4203 (Online)
Vol. 32 No. 4, 467-476, December 2021
https://doi.org/10.12799/jkachn.2021.32.4.467

텍스트 마이닝과 토픽모델링 분석을 활용한 코로나19와 간호사에 대한 언론기사 분석

안지연[1] · 이윤정[2] · 이복임[3]

경인여자대학교 간호학과 부교수[1], 경인여자대학교 간호학과 조교수[2], 울산대학교 간호학과 교수[3]

Analysis of Media Articles on COVID-19 and Nurses Using Text Mining and Topic Modeling

An, Jiyeon[1] · Yi, Yunjeong[2] · Lee, Bokim[3]

[1]Associate Professor, Department of Nursing, Kyung-In Women's University, Incheon
[2]Assistant Professor, Department of Nursing, Kyung-In Women's University, Incheon
[3]Professor, Department of Nursing, University of Ulsan, Ulsan, Korea

Purpose: The purpose of this study is to understand the social perceptions of nurses in the context of the COVID-19 outbreak through analysis of media articles. **Methods:** Among the media articles reported from January 1st to September 30th, 2020, those containing the keywords '[corona or Wuhan pneumonia or covid] and [nurse or nursing]' are extracted. After the selection process, the text mining and topic modeling are performed on 454 media articles using textom version 4.5. **Results:** Frequency Top 30 keywords include 'Nurse', 'Corona', 'Isolation', 'Support', 'Shortage', 'Protective Clothing', and so on. Keywords that ranked high in Term Frequency-Inverse Document Frequency (TF-IDF) values are 'Daegu', 'President', 'Gwangju', 'manpower', and so on. As a result of the topic analysis, 10 topics are derived, such as 'Local infection', 'Dispatch of personnel', 'Message for thanks', and 'Delivery of one's heart'. **Conclusion:** Nurses are both the contributors and victims of COVID-19 prevention. The government and the nurses' community should make efforts to improve poor working conditions and manpower shortages.

Key Words: COVID-19; Nurses; Data mining

서 론

1. 연구의 필요성

전 세계적으로 처음 코로나19 환자가 발생한 것은 2019년 12월 31일 중국 후베이성 우한시에서부터였고, 우리나라에서 첫 환자가 나온 것은 2020년 1월 20일로 알려져 있다[1]. 전 세계적으로는 2021년 6월 9일 기준 현재 173,185,146명의 환자가 발생하여 3,733,604명의 사망자가 있었고[2], 우리나라에서는 145,692명(해외유입 9,220명)의 환자가 발생하고 1,977명이 사망하였다(21.06.09., 0시 기준)[3].

코로나19 팬데믹 상황에서 전 세계가 백신 개발에 총력을 기울여 화이자, 모더나, 아스트라제네카 등에서 백신이 개발되었고 우리나라도 2021년 6월 9일 현재 백신 접종자가 11,527,605명에 이르고 있다[4]. 그러나 최근 발생한 변이바이러스로 인해 백신접종률이 높아도 환자발생률이 감소되지 않아[5] 현재

주요어: 코로나19, 간호사, 데이터 마이닝

Corresponding author: Lee, Bokim
Department of Nursing, University of Ulsan, 93 Daehak-ro, Nam-gu, Ulsan 44610, Korea.
Tel: +82-52-259-1283, Fax: +82-52-259-1236, E-mail: bokimlee@ulsan.ac.kr

Received: Jun 11, 2021 / Revised: Oct 20, 2021 / Accepted: Oct 26, 2021

This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

의 팬데믹 상황이 언제 종결될지 가늠할 수 없는 상황이다.

이러한 코로나19 팬데믹 상황의 발생으로 전국의 보건의료기관과 의료진들은 휴일도 없이 환자 치료와 검진 업무에 투입되고 있으며 그 어느 때보다 간호사에 대한 업무요구가 많아지고 있다[6,7]. 코로나 환자를 돌보는 간호사들의 활약상은 여러 언론 매체를 통해 연일 보도되었다. 마스크와 방호복을 몇 겹씩 겹쳐 입고 긴 시간 화장실도 못가며 음압병상에서 감염 환자를 돌보는 간호사들, 초기에 대구 지역에서 폭발적인 환자가 발생하자 자발적으로 전국에서 환자치료를 위해 달려온 간호사들, 이를 만화로 그려내어 간호사들의 노고를 국민들에게 알려준 또 다른 간호사의 이야기, 그리고 이러한 간호사들의 고생과 헌신에 감동하고 감사해하는 국민들의 얘기가 연일 신문기사를 장식했다.

그러나 사실 지금까지는 간호사 역할의 중요성에 비해 일반 대중들은 간호사의 위상과 역할을 제대로 알지 못하고 간호사에 대한 시각도 상당히 부정적으로 왜곡되어 있었다[8]. 간호사 이미지는 시대에 따라 자비의 천사, 여비서, 영웅, 어머니, 성적대상으로 변해왔다[9]. 최근 인터넷 신문에 나타난 간호사 이미지를 분석한 한 연구에 의하면 '사회적 역할 확대'에 관한 기사가 가장 많고, '전문적인 모습'은 상대적으로 적게 보도되었으며, 호의적인 기사의 비율(44.3%)만큼이나 비호의적인 기사의 비율(20.5%)도 높았다[10].

그런데 코로나19라는 세계적인 재앙으로 인해 간호사의 역할에 대한 요구는 높아졌고, 간호사 업무의 중요성에 대한 국민의 관심과 호응도 같이 높아지고 있는 것으로 보인다. 코로나19 발생 이전과 이후의 인터넷 포털과 소셜미디어에 나타난 간호사 관련 토픽을 비교한 연구에서 코로나19 발생 후 간호사 관련 토픽이 15% 증가한 것을 보여주었다[11]. 또한 코로나19 발생 이전의 주요 키워드가 간호, 제왕절개, 물리치료, 청소, 영양이었던 것에서 간호, 응급, 관계, 가운, 대학으로 변화하였으며, 이는 대중의 인식이 코로나19 확산 위험을 감수하면서도 간호사에 대한 존경과 자부심을 가진다는 것이라고 해석하였다[11]. 그러나 기존 연구는 글의 제목과 2~3줄의 요약 글을 분석하였다는 점, 취업, 진로에 대한 단순 광고가 포함되었을 가능성이 있다는 점, 다양한 매체가 포함되어 같은 키워드라 하더라도 그 의미가 다를 수 있다는 점 등의 한계가 있다[11]. 이러한 문제 인식하에서 우리나라 언론사에서 생산하는 종합일간지, 경제지, 지역일간지, 방송 등의 뉴스에서 코로나19와 간호사가 실제로 어떻게 나타나고 있는지를 분석할 필요가 있다고 판단하였다. 특히 이에 대한 언론의 태도와 내용분석이 정밀하게 이루어진다면 코로나19 시대에 간호사에 대한 인식이 어떻게 형성되어 있는지 확인할 수 있을 것이다.

언론 보도 기사는 대표적인 소셜미디어 데이터로서 빅데이터 분석으로 가장 많이 활용되는 소스이다. 언론기사의 경우 빅데이터의 조건인 볼륨(volume), 속도(velocity), 다양성(variety)을 모두 충족하는 대표적인 비정형 데이터로써 연구목적에 맞게 수집된 자료는 아니지만 데이터가 가지는 잠재적 의미는 매우 크다고 할 수 있다. 언론기사와 같은 자연 언어로 작성된 비구조적(unstructured) 데이터를 대상으로 유용한 정보를 추출하는 기술을 텍스트 마이닝이라고 한다[12]. 키워드 검색의 수준을 넘어서 지능적 분석기술인 텍스트 마이닝을 통해 얻어진 문서 집합에서 추상적인 주제를 발견하기 위한 통계적 모델 중 하나가 바로 토픽 모델링이다. 토픽 모델링을 위해 통계 도구인 Python 패키지 또는 R 통계 프로그램를 활용할 수도 있으나 최근에는 텍스트 마이닝과 토픽 모델링을 함께 실시하고 토픽모델링의 시각화 결과까지 얻을 수 있는 프로그램이 개발되고 활용되고 있다. 본 연구에서는 텍스트 마이닝과 토픽 모델링 기법을 활용하여 코로나와 간호사가 함께 언급된 언론기사를 분석하여 코로나라는 사회적 이슈 속에서 간호사가 어떠한 맥락으로 기사화되었는지 확인하고자 하였다.

2. 연구목적

본 연구는 코로나19와 그 대응을 위해 최일선에서 일하는 간호사에 대한 언론기사를 분석하여 언론을 통해 형성되고 있는 간호사에 대한 사회적 인식을 파악하는 것을 목적으로 한다.

연구방법

1. 데이터 수집

코로나 및 간호사와 관련된 단어를 파악하고자 뉴스 빅데이터 분석 시스템인 빅카인즈(www.bigkinds.or.kr)를 사용하여 자료를 수집하였다. 한국언론진흥재단에서 1990년부터 운영하고 있는 빅카인즈는 2016년부터 일반인을 위해 키워드 기반으로 키워드 언급 추이, 뉴스 히스토리, 매체별 언급 빈도 등 입체적으로 뉴스를 파악할 수 있는 서비스를 제공하고 있다.

데이터 수집을 위한 키워드는 코로나와 간호사이지만 연관어까지 포함하기 위하여 검색어는 '[코로나 or 우한폐렴 or covid] and [간호 or 간호사]'의 조합에 따른 검색식을 모두 포함하였다. 언론사는 특정하지 않았고 빅카인즈에서 제공하는 모든 언론사를 포함하였다. 언론사 개수는 중앙지 총 11개, 경

제지 8개, 지역종합지 28개, 방송사 5개, 전문지 2개를 모두 포함하는 54개였다. 검색 기간은 코로나바이러스가 국내에 처음 유입된 2020년 1월 1일에서부터 2020년 9월 30일까지였다. 통합 검색에서 추출된 기사는 총 5,939개였으나, 중복기사와 URL을 통해 기사 원문이 확인되지 않는 기사를 제외한 총 5,523개의 기사를 1차로 선정하였다. 1차 선정된 기사 중에서 분석에 포함시킬 최종 기사를 선택하기 위해 배제기준을 마련하였고, 연구자 간의 반복적인 회의를 통해 연구자의 주관성을 배제하고 기사 선정에 신뢰성을 확보하였다. 배제기준 마련을 위한 연구자 간의 의견 일치가 쉽지 않았기에 1차 선정된 기사는 모든 연구자가 각자 원문을 확인하였고, 둘 이상의 연구자의 일치 의견이 없는 경우 기사 건별 선정 여부에 대한 회의를 추가하였다. 그 결과 총 454편의 기사가 최종 분석데이터로 선정되었다. 이때 배제기준에는 특정 병원 또는 특정 학교 홍보 기사, 특정 기관 또는 모임에 간호사의 단순 참여, 간호사 이력을 가진 특정인 언급, 간호조무사 관련 기사 등이 포함되었다.

2. 데이터 분석

1) 데이터 전처리

수집된 자료는 텍스트 마이닝 기술을 이용한 빅데이터 분석 프로그램인 텍스톰(textom 4.5 version)을 사용하여 분석하였다. 텍스트 마이닝을 위한 전 처리는 두 단계를 거쳤다. 정형화되지 않은 빅테이터로부터 도출된 결과의 정확성을 높이기 위한 데이터 전처리 과정이 필수적이다. 첫 번째 단계는 원문데이터에서 연구자가 직접 실시한 수동적 단어 정제과정이다. 단어빈도분석을 실시하여 띄어쓰기 수정, 불필요한 단어 삭제, 숫자 및 고유명사 제거, 단어 일치 등 원문데이터를 직접 수정하는 정제과정을 거쳤다. 수동적 단어 정제과정은 언론기사마다 동일한 단어일지라도 띄어쓰기와 단어조합이 다르기 때문에 반드시 선행해야 하는 과정이다. 예를 들어, '코로나19'는 '코로나'로 대체, '간호인력'는 '간호', '인력'으로 띄어쓰기, 특정 기업명은 삭제 등으로 빈도분석에서 분석 단어의 적절성을 평가하기 위해 빈도표에 대한 반복적인 확인과 수동 정제과정을 실시하였다, 특히 수동적 단어 정제과정은 연구결과에 중요한 영향을 미치는 전처리 과정이므로 모든 연구자의 합의에 의해서 진행되었다. 두 번째 단계는 텍스톰의 기본 전처리 과정인 형태소분석에 의한 정제과정이다. 단어의 형태소분석을 위해 분석언어는 한국어, 분석모듈은 Espresso K, 분석품사는 명사로 설정하여 실시하였다. 이때 분석모듈인 Espresso K는 고유명사 및 복합명사를 그대로 반영하는 기법으로 원문의 단어를 최대한 살리는 방법이다. 두 번의 단어 정제과정 후에도 정제되지 않는 일부 단어는 분석 직전에 웹상에서 바로 편집하였다.

2) 데이터 분석

본 연구에서는 텍스트 마이닝과 토픽모델링 분석을 활용하여 코로나 및 간호사와 관련된 단어 또는 이슈를 분석하였다. 텍스트 마이닝을 통해 단어분석, N-gram, TF-IDF, 연결중심성(degree centrality) 분석을 실시하였고, 이들 결과값은 시각화 결과와 매칭하여 제시하였다. 단어분석은 추출된 단어와 데이터 내 해당 단어의 빈도수를 나타내는 결과값으로 검색어인 코로나 및 간호사가 포함된 기사에서 어떠한 단어들이 언급되는지를 서술적으로 파악할 수 있는 결과이다. N-gram은 두 단어의 쌍으로 두 단어가 나란히 등장하는 빈도(동시출현빈도, Co-occurrence)를 순위화한 값으로 본 연구에서는 N-gram 네트워크 시각화 결과로 제시하였다. TF-IDF는 단어빈도(TF, Term Frequency)와 문서빈도의 역수(IDF, Inverse Document Frequency)를 곱한 값으로 문서 내에서 키워드가 실제로 핵심적인 의미를 갖는지 알아볼 수 있는 값이다[13]. TF-IDF 값이 높다는 의미는 다른 문서에는 많지 않고, 해당 문서에서 자주 등장하는 단어로 해석할 수 있으며 쉽게 말해, 가중치가 반영된 단어 빈도분석이라고 할 수 있다. 단어 개수가 증가할수록 값은 증가하므로 TF-IDF의 상한값은 정해져 있지 않다. 본 연구에서는 단어빈도분석과 TF-IDF를 함께 결과로 제시함으로써 코로나 및 간호사 관련 언론기사에서 주로 어떤 단어가 자주 등장하는지 그리고 주로 어떤 단어가 중요하게 언급되는지를 보여주고자 하였다. 연결중심성 분석은 단어가 얼마나 많은 연결관계를 가지고 있고 중심이 되는지를 보여주는 통계적 수치로 0에서 1의 값을 갖는다. 1에 가까울수록 키워드 네트워크 중심에 있어서 다른 단어에 미치는 영향 정도가 크고, 0에 가까울수록 키워드의 네트워크 내의 중심에서 떨어져 있다고 해석한다.

본 연구에서 토픽분석은 LDA (Latent Dirichlet Allocation) 토픽모델링 방법을 사용하였다. LDA는 2003년 처음 소개된 토픽분석방법으로 수집된 원문 내용에 담긴 다양한 키워드를 기반으로 내용을 유형화(집락화) 시켜주는 방법이다[13]. 이 방법은 인터넷 매체와 같이 대량의 데이터로부터 주제를 찾기 위한 알고리즘으로 유사한 의미를 가진 단어들을 집합화하는 방식이다. 즉, LDA에서 데이터는 특정 확률에 의해 선택된 단어들로 구성된 토픽들의 집합으로 표현된다. LDA의 핵심의 최적의 토픽 수를 결정하는 것인데 첫 번째는 집합화 분석을 반복하면서 적절한 토픽의 수를 연구자가 결정할 수 있고, 두 번째는 Coherence와 Perplexity의 모델평가에 따른 계산식에

의해서 최적의 토픽 수를 결정할 수도 있다[13-15]. 연구자가 직접 결정할 경우에는 LDA 분석의 시각화 결과를 확인하면서 토픽 수가 적절한지 평가할 수 있다. 시각화는 원 형태의 토픽이 표시되고,토픽 간의 거리가 공간적으로 구분되는 지도(Intertopic Distance Map, IDM) 형태이다. 토픽의 원들이 서로 겹치지 않고 독립적이며 토픽간의 거리가 충분히 떨어져있으면 토픽 분류가 잘되었다고 해석할 수 있다. 본 연구는 최적의 토픽 수를 산출하기 위하여 다양한 개수의 토픽을 설정해보았고, 원의 크기는 비록 작지만 다른 토픽과 뚜렷하게 구분되는 IDM 확인을 통해 최종 10개로 결정하였다. 또한 각 토픽에 포함되는 단어 수도 5~15개로 다양하게 설정해보았고, 각 토픽의 특징이 명확하게 드러나는 단어가 포함되었는지 확인을 반복한 결과, 최종적으로 포함한 단어의 수는 총 15개로 하였다. 본 연구는 빈도가 낮은 단어까지 모두 LDA 분석에 포함하였기에 λ(람다 값)은 1로 설정하여 토픽별로 가장 자주 등장하는 단어들을 우선적으로 키워드로 선택하였다. 텍스톰에서 LDA 토픽모델링은 토픽을 랜덤으로 할당한 후 토픽의 재할당을 반복수행하면서 단어의 토픽을 찾게 되는데 본 연구에서는 무작위 할당을 위해 랜덤값은 사용하지 않았다. 텍스톰에서는 토픽분석 후 시각화 결과도 함께 제시하므로 본 연구에서도 토픽별 거리를 가시적으로 표현하였다.

3. 윤리적 고려

본 연구가 활용한 분석데이터는 개인을 식별할 수 있는 정보를 가지고 있지 않기에 울산대학교의 생명윤리위원회로부터 IRB 심의면제(IRB No.: 2021R0019-001)를 득하고 연구를 진행하였다. 개인정보를 포함하지 않았다 할지라도 분석된 키워드를 통해 특정 개인의 정보가 노출될 우려가 있으므로 데이터 전처리 과정에서 해당 단어는 모두 가명처리 하였다.

연구결과

1. 단어 빈도분석

Table 1은 빈도수에 따른 상위 30개의 단어의 빈도 및 백분율을 나타내고 있다. 단어분석으로 확인된 총 단어수는 8,469개였고, 누적백분율은 27.0%로 나타났다. 1,000개 빈도 이상의 단어는 '간호사(Nurse)', '코로나(Corona)', '환자(Patient)', '병원(Hospital)', '대구(Daegu)', '의료진(Medical staff)'이었고, 이들 단어의 누적백분율은 13.9%였다. 코로나바이러스 감염증으로 발생한 여러 가지 의료적 상황과 관련된 단어(격리/Isolation, 입원/Admission, 상황/Circumstances, 지원/Support, 현장/Site, 병동/Ward, 병상/Hospital bed 등)와

Table 1. Frequency to Rank 30 of Keyword

Rank	Keyword	n	%	Cumulative %
1	Nurse	2,953	3.5	3.5
2	Corona	2,889	3.4	6.9
3	Patient	1,904	2.3	9.1
4	Hospital	1,656	2.0	11.1
5	Daegu	1,211	1.4	12.5
6	Medical staff	1,129	1.3	13.9
7	Confirmed case	779	0.9	14.8
8	Medical treatment	719	0.9	15.6
9	Infection	717	0.9	16.5
10	Manpower	613	0.7	17.2
11	Circumstances	592	0.7	17.9
12	Working	586	0.7	18.6
13	Treatment	526	0.6	19.2
14	Confirmed	522	0.6	19.8
15	Isolation	502	0.6	20.4
16	Doctor	491	0.6	21.0
17	Support	483	0.6	21.6
18	Nursing	446	0.5	22.1
19	Area	432	0.5	22.6
20	Site	418	0.5	23.1
21	Ward	407	0.5	23.6
22	Protective clothing	403	0.5	24.1
23	Thinking	354	0.4	24.5
24	Shortage	339	0.4	24.9
25	Mask	324	0.4	25.3
26	Admission	310	0.4	25.6
27	Tested (positive/negative)	309	0.4	26.0
28	Hospital bed	303	0.4	26.3
29	Quarantine	297	0.4	26.7
30	Emerging	296	0.4	27.0
Total keyword				8,469
Total frequency				84,713



Figure 1. Word cloud of the keywords.

의료인 또는 의료지원인력 등 인적자원과 관련된 단어(인력/Manpower, 의사/Doctor, 의료진/Medical staff 등)가 30위 안에 포함되어 있었다. 또한 24위인 '부족(Shortage)'이라는 단어는 총 339개의 빈도로 나타났는데 다양한 자원의 부족이라는 맥락 속에서 추출된 단어이다. Figure 1은 상위 30개 단어 빈도의 시각화 결과로 빈도에 따라 상, 중, 하로 구분되어 3가지 색상으로 나타나고 있다.

2. N-gram, TF-IDF 및 연결중심성 분석

Table 2는 N-gram, TF-IDF, 연결중심성 분석 결과를 나타내준다. Figure 2는 N-gram을 시각화한 결과이다. N-gram은 확률 모델을 통해 단어쌍을 선별해주는 기법이다. 분석 문서에서 2개 단어가 연쇄적으로 표현된 개수를 표현한 값으로 본 연구에서 나타난 가장 많은 단어쌍은 '코로나-환자(Corona-Patient)'로 전체 문서에서 총 223회였다. 그다음은 '신종-코로나(Emerging-Corona)(n=213)', '전담-병원(Dedicated-Hospital)(n= 194)', '코로나-바이러스(Corona-Virus)(n=191)' 순위로 나타났다.

TF-IDF는 단어가 특정 문서에서 얼마나 핵심적인 의미인지를 알 수 있는 지수인데 본 연구에서는 '대구(Daegu)'가 855.51로 가장 높은 수치로 나타났다. 단어 빈도분석에서 '대구(Daegu)'는 5위였으나 TF-IDF 값에서는 가장 높은 순위로 나타난 셈이다. 2위인 '대통령(President)'과 3위인 '광주(Gwangju)' 는 단어빈도분석에서 상위 30위에 포함되지도 않았던 단어로 확인된다. 단어와 단어의 연결 정도를 나타나는 연결중심성 분석 결과에서는 '간호사(Nurse)'가 .18로 나타났다. 그 다음 순위로 '코로나(Corona)'가 .14, '환자(Patient)'가 .11, '병원(Hospital)'이 .1로 나타났다.

3. 토픽분석

본 연구에서는 단어 빈도분석에서 추출된 단어 8,469개 모두 포함하여 LDA 토픽모델링 분석을 실시하였다. Table 3은 총 10개 토픽으로 유사한 단어를 집합화한 결과를 보여주고 있다. '토픽 7 (의료환경/Medical condition)'이 33%로 가장 높은 백분율을 차지하였다. 그다음은 '토픽 5 (지역감염/Local infection)'가 두 번째 비중(25.2%)을 차지하였다. 토픽의 명명은 해당 토픽 내에 순위가 높은 단어가 아니라 다른 토픽에 포함되어 있지 않거나 혹은 해당 토픽에서 두드러지는 단어들의 조합으로 이름을 붙였다. 예를 들어, '코로나(Corona)', '간호사(Nurse)', '병원(Hospital)', '환자(Patient)'는 모든 토픽 내에서 순위가 높은 단어로 포함되어 있지만, '의료(Medical treatment), '확진(Confirmed)', '치료(Treatment)', '상황(Circumstance)' 등의 단어들의 조합이 '토픽 7'에서 두드러져 보였기에 토픽명을 '의료환경(Medical condition)'으로 하였다. '토픽 3'의 경우에는 다른 토픽과 달리 상위 순위에 '전달(Delivery)', '마음(Mind)', '응원(Cheering)'의 단어가 포함되었기에 토픽명을 '마음전달(Delivery of one's heart)'라고 결정하였다.

한편, LDA 시각화 결과에서는 토픽의 크기가 클수록 분석 기사 중 해당 토픽이 차지하는 비중을 의미하므로 가장 큰 크기의 토픽을 메인 토픽이라고 해석할 수 있다. 또한 본 연구에서는 공간적으로 근접해있는 다른 원들과 달리 크기는 작지만 토픽 간 거리가 많이 벌어지는 4개의 토픽 원을 확인할 수 있었다. 이들은 다른 토픽과 거리가 멀어서 판별타당도가 높고 주제가 뚜렷하다고 해석할 수 있다. 시각화 결과를 따로 제시하지는 않았지만 본 연구에서 메인토픽은 '토픽 7 (의료환경/Medical



Figure 2. N-gram network of the keywords.

Table 2. N-gram, Term Frequency-Inverse Document Frequency (TF-IDF), and Degree Centrality

Rank	N-gram			TF-IDF		Degree centrality	
	Keyword	Co-occurrence	n	Keyword	Value	Keyword	Value
1	Corona	Patient	223	Daegu	855.51	Nurse	.18
2	Emerging	Corona	213	President	757.64	Corona	.14
3	Dedicated	Hospital	194	Gwangju	549.88	Patient	.11
4	Corona	Virus	191	Manpower	540.45	Hospital	.10
5	Corona	Confirmed case	167	Isolation	508.10	Medical staff	.08
6	Nursing	Manpower	149	Nursing officer	489.82	Daegu	.07
7	Confirmed	Tested	140	Confirmed	480.03	Circumstance	.06
8	Corona	Confirmed	134	Protective clothing	479.91	Confirmed case	.05
9	Corona	Situation	133	Hospital	478.84	Working	.05
10	Patient	Treatment	132	Ward	473.96	Infection	.05
11	Doctor	Nurse	130	Patient	468.34	Medical treatment	.05
12	Infectious disease	Emerging	129	Infection	460.05	Doctor	.04
13	Screening	Testing station	128	Doctor	456.98	Support	.04
14	Nurse	Corona	125	Medical treatment	455.34	Thinking	.04
15	Dongsan	Hospital	124	Thinking	439.99	Nursing	.04
16	Isolation	Ward	121	Armed forces	437.64	Protective clothing	.04
17	Daegu	Gyeongbuk	119	Support	433.61	Manpower	.04
18	Medical treatment	Manpower	116	Working	432.59	Person	.04
19	Hospital	Nurse	112	Nursing	430.30	Mask	.04
20	Manpower	Shortage	110	Confirmed case	428.26	Isolation	.04

TF-IDF=term frequency-inverse document frequency.

Table 3. Latent Dirichlet Allocation (LDA) Topic Modeling

Key word rank	Topics									
	1	2	3	4	5	6	7	8	9	10
	Dispatch of personnel	Message for thanks	Delivery of one's heart	Prevention of infection	Local infection	Working circumstance	Medical condition	Status of infection	Medical support	Hospital situation
1	Corona	Nurse	Corona	Patient	Nurse	Corona	Nurse	Corona	Corona	Nurse
2	President	President	Delivery	Corona	Corona	Patient	Corona	Hospital	Medical staff	Isolation
3	Patient	Patient	Mind	Hospital	Daegu	Medical staff	Patient	Nurse	Daegu	Corona
4	Nurse	Corona	Nurse	Nurse	Patient	Nurse	Hospital	Patient	Hospital	Patient
5	Daegu	Daegu	Kit	Daegu	Hospital	Hospital	Medical staff	Confirmed	Doctor	Hospital
6	Doctor	Protective clothing	Medical staff	Medical staff	Medical staff	Daegu	Daegu	Confirmed case	Isolation	Quarantine
7	Medical staff	Working	Cheering	Confirmed case	Confirmed case	Medical treatment	Manpower	Medical staff	Patient	Ward
8	Medical treatment	Hospital	Daegu	Treatment	Circum-stance	Infection	Infection	Infection	Nurse	Support
9	Armed forces	Doctor	Duty free shop	Area	Working	Circum-stance	Medical treatment	Seoul	Confirmed case	Medical staff
10	Hospital	Message	Shopping	Mask	Doctor	Working	Confirmed case	Working	Support	Daegu
11	Nursing officer	Thank	Campaign	Confirmed	Ward	Nursing officer	Nursing	Tested	Medical treatment	Confirmed case
12	Wearing	Blue house	Mask	Infectious disease	Gwangju	Suport	Treatment	Area	Manpower	employee
13	People	Ward	Thanks-giving	Support	Medical treatment	Protective clothing	Working	Support	Screening	Admission
14	Message	Medical staff	Quarantine	Manpower	Thinking	Confirmed	Circum-stance	Site	Circum-stance	Nursing officer
15	Government	People	Mother	Isolation	Infection	Confirmed case	Support	Protective clothing	Clinic	Protective clothing
Top-30 most relevant terms for Topic λ=1	8.9% of tokens	3% of tokens	1.3% of tokens	7.4% of tokens	25.2% of tokens	5.3% of tokens	33% of tokens	10.2% of tokens	4.4% of tokens	1.3% of tokens

condition)'인 셈이고, 판별타당도 관점에서는 '토픽 1 (인력파견/Dispatch of personnel)', '토픽 2 (감사메세지/Message for thanks)', '토픽 3 (마음전달/Delivery of one's heart)', '토픽 10 (병원상황/Hospital situation)'이 나머지 6개 토픽과는 거리가 매우 멀고 동떨어져 있으므로 본 연구에서 특징적인 토픽인 것으로 확인할 수 있다.

논 의

본 연구는 최근 보건의료의 핵심문제인 코로나19와 그 대응을 위해 최일선에서 일하는 간호사에 대한 언론기사를 분석하여 언론을 통해 형성되고 있는 사회적 인식들을 파악하기 위해 시행되었다. 텍스트 마이닝과 토픽 모델링을 통해 도출된 결과를 종합하면, 간호사와 코로나를 다룬 언론기사들은 '감염 현황', '방역 노력', '의료자원'과 관련한 것으로 크게 구분할 수 있다. 이러한 내용이 독립적으로 다루어지기 보다는 국회의원 선거, 의사 파업과 같은 정치적 상황, 추석 등의 시기적 상황, 대구 신천지 사태와 같은 지역적 상황, 기업, 학교, 지자체 등의 행사 등 여러 상황과 복잡하게 맞물려 다루어졌다.

먼저 '감염 현황'은 코로나19에 감염된 환자, 지역, 사업장 등을 보도하는 내용으로, 이러한 보도와 함께 간호 또는 간호사가 언급되었다. 이는 단어 빈도분석과 TF-IDF에서 '환자(Patient)', '병원(Hospital)', '대구(Daegu)'나 '광주(Gwangju)'와 같은 지역명, '확진자(Confirmed case)' 등의 단어가 상위를 차지하고

있고 N-gram에서 '코로나- 환자(Corona-Patient)', '코로나-확진자(Corona-Confirmed case)', '확진-판정(Confirmed-Tested)', '코로나-사태(Corona-Situation)'가 함께 다빈도로 등장한 점, '지역감염(Local Infection)'과 '감염현황(Status of infection)'과 같은 토픽이 도출된 점 등을 통해 확인할 수 있다.

'방역 노력'은 코로나19 확산 예방을 위해 간호사를 포함한 의료진의 활동을 보도하는 내용으로, 단어 빈도분석과 TF-IDF에서 '근무(Working)', '치료(Treatment)', '격리(Isolation)', '지원(Support)', '방역(Quarantine)', '방호복(Protetive clothing)', '마스크(Mask)' 등의 단어가 다빈도 혹은 핵심적으로 등장하였고, N-gram '선별-진료소(Screening-Testing station)', '격리-병동(Isolation-Ward)', '환자-치료(Patient-Treatment)'와 같은 단어가 연결되었으며, '감염예방(Prevention of infection)', '의료지원(Medical support)', '병원상황(Hospital situation)', '감사메세지(Message for thanks)', '마음전달(Delivery of one's heart)' 등의 토픽이 도출된 것이 이를 뒷받침하는 자료이다. 우리나라에 코로나19가 출몰하고 대구 지역의 심각한 지역감염으로 이어졌던 지난해 1~3월, 코로나19 확산예방을 위해 신임간호장교와 전국 지역에서 온 자원봉사 간호사가 투입되었다. 전국의 간호사가 감염의 위기 속에서 국민의 생명을 지키기 위해 지역으로 자원봉사를 나섰고, 위기 속 간호사의 헌신에 대한 국민의 감사와 대통령의 치하가 언론의 주목을 받았다. 비록 분석결과로 도출되지는 않았으나, 희생, 헌신, 희망, 천사, 영웅, 투사 등의 용어들이 기사제목으로 등장하였다. 2020년 3월 중순부터 한 달간 보도된 영문뉴스를 국가별로 분석한 연구에서, 간호사의 중요성(중국), 희생(이스라엘), 위험한 근무조건을 항의하는 간호사의 목소리(미국, 캐나다), 간호사 시위와 간호사 부족(남미), 영웅과 천사(유럽), 간호사에 대한 보호부족(영국, 뉴질랜드, 호주) 등이 각 나라별로 주요한 이슈로 나타났다[16]. 요약하자면 세계 여러 나라의 언론은 코로나19 상황에서 간호사들의 방역 노력을 영웅적인 것으로 묘사하고 있지만 한편으로는 보호받지 못하는 희생자로 보도하고 있는 것을 알 수 있다. 그러나 우리나라 언론이나 소셜미디어에서는 코로나19로 인해 간호사들이 겪는 어려움에 대해 드물게 다루고 있다[11]. 우리나라는 국회의원 선거와 의사 파업 시기와 맞물려 간호사의 격무, 간호인력 확보에 대한 기사가 등장하였는데, '챌린지'와 같이 간호사의 노고를 알아주는 것은 고마운 일이나 열악한 근무환경과 인력부족의 문제 해결에 대해 정부가 더 적극적으로 나서달라는 간호계의 목소리를 전달하는 기사였다.

마지막으로 간호사를 포함한 의료진과, 환자를 치료할 병상 및 병원을 다루는 '의료자원'과 관련한 내용이 있었다. 이에 대한 근거는 단어빈도분석에서 '인력(Manpower)', '간호장교(Nursing officer)', '병상(Hospital bed)', '병원(Hospital)', '부족(Shortage)' 등이 다빈도로 등장하였고 N-gram에서 '전담-병원(Dedicated-Hospital)', '간호-인력(Nursing-Manpower)', '의사-간호사(Doctor-Nurse)', '의료-인력(Medical treatment-Manpower)', '인력-부족(Manpower-Shortage)'이 자주 연결되어 표현되었다는 점, '인력파견(Dispatch of personnel)'의 토픽이 도출된 점이다. 우리나라와 마찬가지로 세계 여러 나라는 코로나19와 관련된 의료인력 부족을 경험하고 있다. 세계 최고의 의료인력 수출국가인 필리핀의 경우, 의료인력 부족 문제를 해소하기 위해 의료자원봉사자를 모집하고 의료인력의 해외파견을 일시 금지했다[17]. 일부 국가에서는 간호 인력 확보를 위해 학기말 간호대학생을 조기에 간호사로 등록하거나 퇴직한 간호사를 다시 훈련시키기도 했다[18]. 우리나라 또한 대한간호협회를 중심으로 휴직하거나 퇴직한 간호사를 모집, 추가 교육하여 간호사가 부족한 현장에 배치하였다[19]. 우리나라의 간호인력 부족 문제는 코로나19 초기 대응시기뿐만 아니라 유행이 장기화되고 있는 현시점에서도 꾸준히 제기되고 있다. 최근 코로나19 업무를 담당하던 보건소 간호직 공무원의 과로 자살은, 죽음으로 내몰릴 수밖에 없는 환경에 처해진 공공보건기관 간호사의 현실을 보여준다. 보건소 간호사는 선별진료소의 운영, 검체 채취를 위한 가정방문, 확진자 후송, 역학조사, 자가격리자 관리, 백신접종과 이상반응 관리 등을 위해 주·야간 비상근무를 해야 하고 한 달 평균 100시간 이상의 시간외 근무를 수행하기도 한다[20]. 게다가 코로나19 대응업무는 감염병 전담부서가 아니더라도 보건직 간호사라면 누구나 업무 지원에 나서야 하는 상황이다[19]. 코로나19 환자 간호는 바이러스가 퍼지는 것을 예방, 억제하기 위해 추가적인 간호 처치와 예상치 못한 환자 변화상황에 대응해야 하기 때문에 이로 인한 업무부담이 높다[6,7]. 또한 치명적인 바이러스와 직접 접촉하는 상황에서 본인과 가족의 건강에 대한 걱정과 지속적인 환자 간호를 제공해야 하는 직업윤리 사이에서 심각한 스트레스를 경험한다[21,22]. Lucchini 등[23]은 간호활동점수(Nursing Activities Score)를 이용한 연구를 통해 중환자실의 코로나19 환자간호를 위해서는 간호사 대 환자의 비율이 1:1.5여야 한다고 제안한 바 있다. 환자를 돌보기 위해 필요한 시간과 노동강도를 고려하여 중증도 분류기준을 마련하고, 그에 따른 인력 기준이 제시되는 것이 바람직할 것이다. 또한, 코로나19와 같은 재난 상황에 대비한 간호 비상가동인력 운용체계도 구축될 필요가 있다[24]. 지역사회 감염위기를 조기에 경험한 대구 지자체의 경우 의료진의 소진과 이직으로 인한 의료체계 위기를 막

기 위해 환자 중증도에 따라 1:1 (최중증)~1:2.5 (중등증)의 간호 인력 기준을 제시하고 있다.

우리 사회는 코로나19 환자를 가장 가까이에서 돌보는 간호사에 대해 걸맞은 보상과 지원을 하고 있는가를 성찰해야 한다. 정부는 지난해 6월 코로나19 치료에 파견된 의료인력이 방역활동에 집중할 수 있도록 지원에 관한 지침을 마련하였고 올해 1월에는 중환자 전담 병상 간호사에 대한 수당과 야간 간호관리료의 인상대책을 발표하였다. 그러나 감염병 팬데믹이 장기화 되고 있는 상황에서 중환자실이 아닌 곳에서 감염병 예방에 투입되고 있는 간호사, 특히 공공보건기관 간호사에 대한 고려가 미흡하다는 지적이 있다. 지역보건법 상 보건소 간호인력의 배치기준은 지난 25여 년간 개정되지 않았고, 저출산, 고령화, 치매 관리 등 시대가 요구하는 필수간호서비스 제공을 위해 인력을 비정규직으로 채용하여 전국 공공보건기관의 간호사 중 비정규직 비율이 49.2%에 이른다[19]. 그 처우 또한 열악한데, 일예로 보건직 간호사의 특수업무수당은 월 5만 원으로 의사나 사회복지사와 비교하여 매우 낮은 수준이다[19]. 21세기 들어 SARS (Severe Acute Respiratory Syndrome), MERS (Middle East Respiratory Syndrome), 코로나19와 같은 신종감염병 대유행을 경험하면서 간호사에 대한 사회적 위상이 높아진 만큼, 간호사의 전문성과 역할에 걸맞은 근무조건과 처우가 마련되고 적정 간호인력이 배치되어야 할 것이다.

본 연구는 54개 언론사의 뉴스를 분석하는 빅카인즈를 사용하여 자료를 수집하였기 때문에, 국내에서 생산되는 모든 언론기사를 포함하였다고 확언하기 어렵다. 또한 뉴스는 사회현상을 다루는 객관적인 데이터로 유용하나 정서 관련 단어의 언급빈도가 적다는 제한점도 있다[25]. 한편, 본 연구가 활용한 텍스트 마이닝과 토픽 모델링은 양적 분석방법에 기반을 두기 때문에 기사의 개별 내용에 대한 질적, 심층 분석이 이루어지지 못했다. 마지막으로 데이터 정제에 많은 시간과 노력이 소요되어 분석된 언론기사가 최신성을 가지지 못하였다. 이러한 제한점에도 불구하고 본 연구는 코로나19가 국내에 유입된 지난해 1월부터 약 9개월간 언론에서 다루어진 코로나19와 간호사 관련 이슈가 무엇인지를 분석함으로써, 간호사 건강권, 간호인력 확보, 근무조건 개선 등 간호정책의 현안에 대한 경험적 근거를 제시하였다는 점에서 의의가 있다.

결론 및 제언

코로나19 팬데믹 상황은 간호계의 위기인 동시에 기회가 될 수 있다. 그동안 간호사의 사회적 가치를 떨어뜨려왔던 사회적 시선과 편견들을 넘어 간호에 대한 대중의 이해를 변화시켜 나가고 정책결정자와 의료리더로서 간호사의 목소리를 찾는 기회로 만들어 나가야 할 것이다.

본 연구에서는 코로나 유입단계부터 9개월 간 언론기사를 통해 드러난 간호사에 대한 사회적 인식을 파악하였고, 방역의 최일선에 선 간호사의 희생과 헌신에 대한 사회적 감사와 함께 열악한 근무조건과 처우 개선, 인력확보에 대한 목소리를 확인할 수 있었다. 즉, 코로나19 방역에서 간호사는 기여자인 동시에 희생자로 보도되었다. 21세기 들어 감염병이 대유행할 때마다 주된 역할을 해온 간호사가 더 이상 열악한 처우와 근무조건에서 외면받지 않도록 정부와 간호계가 함께 노력해 나가야 할 것이다. 추후 연구에서는 간호사에 대한 사회적 인식의 변화를 파악하기 위해 주요 시점별로 구간을 나누어 시계열적 분석을 시도할 것을 제안한다. 또한, 비중이 높은 토픽의 데이터를 추가적으로 분석하면 좀 더 심층적인 정보를 얻을 수 있을 것으로 생각된다.

REFERENCES

1. Hyun JH, Kim JH, Lee HY, Gwack J, Kim JE, Lee EY, et al. Contact tracing results of the first confirmed COVID-19 case in the Republic of Korea. Weekly Health and Disease. 2020;13(7): 352-358.
2. Misistry of Health and Welfare. Corona 19 outbreak status (overseas) [internet]. Daejeon: Ministry of Health and Welfare; c2021 [cited 2021, June 9]. Available from: http://ncov.mohw.go.kr/bdBoardList_Real.do?brdId=1&brdGubun=14&ncvContSeq=&contSeq=&board_id=&gubun=
3. Ministry of Health and Welfare. Corona 19 Domestic Occurrence. [internet]. Deajeon: Ministry of Health and Welfare; c2021 [cited 2021 June 9]. Available from: http://ncov.mohw.go.kr/bdBoardList_Real.do
4. Korea Disease Control and Prevention Agency. Corona 19 Vaccination Status. [Internet]. Chungju: Korean Disease Control and Prevention Agency; c2021 [cited 2021 June 0], Available from: https://kdca.go.kr/board/board.es?mid=a20501000000&bid=0015&list_no=713602&act=view
5. Im SH. This is South America. It's the fastest vaccination, but it's the world's No. 1 infection rate. What's going on in this country? Nownews. 2021 April 7.
6. Giuliani E, Lionte G, Ferri P, Barbieri A. The burden of not-weighted factors-Nursing workload in a medical Intensive Care Unit. Intensive and Critical Care Nursing. 2018;47:98-101. https://doi.org/10.1016/j.iccn.2018.02.009
7. Reper P, Bombart MA, Leonard I, Payen B, Darquennes O, Lab-

rique S. Nursing activities score is increased in COVID-19 patients. Intensive & Critical Care Nursing. 2020;60:102891. https://doi.org/10.1016/j.iccn.2020.102891
8. Yu SY. What is perceived the image of nurses? Comparison major and non-major students. Journal of Digital Convertgence. 2014;12(10):353-361. https://doi.org/10.14400/JDC.2014.12.10.353
9. Kalisch PA, Kalisch BJ. The changing Image of the Nurse. Menlo Park, CA: Addision-Wesley Publishing Co.; 1987.
10. Park SA, Park SJ, Lee C, Yun M, Hwang K. Image of nurses portrayed in internet newspapers. Culture and Convergence. 2017;39(6):677-700. https://doi.org/10.33645/cnc.2017.12.39.6.677
11. Yoon YM, Kim SK, Kim HK, Kim EJ, Jeong Y. Comparison of topics related to nurse on the internet portals and social media before and during the COVID-19 era using topic modeling. Journal of Muscle and Joint Health. 2020;27(3):255-267. https://doi.org/10.5953/JMJH.2020.27.3.255.
12. Seo D. Get It! Textmining with Python. Seoul: BJ Public; 2019. 204 p.
13. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. Journal of Machine Learning Research. 2003;3:993-1022.
14. Jelodar H, Wang Y, Yuan C, Feng X, Jiang X, Li Y, et al. Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. Multimedia Tools and Applications. 2019; 78:15169-15211. https://doi.org/10.1007/s11042-018-6894-4
15. DiMaggio P, Nag M, Blei D. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. Poetics. 2013;41(6):570-606. https://doi.org/10.1016/j.poetic.2013.08.004
16. Bennett CL, James AH, Kelly D. Beyond tropes: Towards a new image of nursing in the wake of COVID-19. Journal of Clinical Nursing. 2020;29(15-16):2753-2755. https://doi.org/10.1111/jocn.15346
17. Atique S, Bautista JR, Block LJ, Lee JJJ, Lozada-Perezmitre E, Nibber R, et al. A nursing informatics response to COVID-19: Perspectives from five regions of the world. Journal of Advanced Nursing. 2020;76:2462-2468. https://doi.org/10.1111/jan.14417
18. Jackson D, Bradbury-Jones C, Baptiste D, Gelling L, Morin K, Neville S, et al. Life in the pandemic: Some reflections on nursing in the context of COVID-19. Journal of Clinical Nursing. 2020;29(13-14):2041-2043. https://doi.org/10.1111/jocn.15257
19. Shin KR. COVID-19 counter measures and action plan for improvement of nursing treatment. HIRA Research. 2021;1(1): 103-107. https://doi.org/10.52937/hira.21.1.1.103
20. Blue House National Petition. Please don't let the nurses at the Corona-19 quarantine health center collapse from exhaustion [internet]. Seoul: Blue House; c2021 [cited 2021, June 29]. Available from: https://www1.president.go.kr/petitions/599451
21. Jiang Y. Psychological impact and coping strategies of frontline medical staff in Hunan between January and March 2020 during the outbreak of Coronavirus Disease 2019 (COVID-19) in Hubei, China. Medical Science Monitor. 2020;26:e924171-1-e924171-16. https://doi.org/10.12659/MSM.924171
22. Hwang N. Experience of COVID-19 in armed forces Daegu hospital. Journal of Military Nursing Research. 2020;38(2):68-73. https://doi.org/10.31148/kjmnr.2020.38.2.68
23. Lucchini A, Giani M, Elli S, Villa S, Rona R, Foti G. Nursing activities score is increased in COVID-19 patients. Intensive & Critical Care Nursing. 2020;59:102876. https://doi.org/10.1016/j.iccn.2020.102876
24. Oh EG. Perspectives on nursing profession for a post-COVID-19 new normal. Korean Journal of Adult Nursing. 2020;32(3): 221-222. https://doi.org/10.7475/kjan.2020.32.3.221
25. Park S, Do K, Kim H, Park G, Yun J, Kim K. An exploratory study of happiness and unhappiness among Korean based on text mining techniques. The Journal of the Korea Contents Association. 2018;18(7):10-27. https://doi.org/10.5392/JKCA.2018.18.07.010