텍스트 마이닝을 활용한
코로나 19 관련 뉴스 타이틀 분석

Analysis of news titles related to COVID-19 using text mining

- Github: https://github.com/youneedpython/youneedpython.github.io
- Email: youneedpython@gmail.com

2022年 06月

텍스트 마이닝을 활용한
코로나 19 관련 뉴스 타이틀 분석

[차 례]

[표 차례]

[그림 차례]

제1장 서론

1.1. 연구 배경

2019년 12월 중국에서 발생한 코로나바이러스감염증-19(Coronavirus Disease 2019, 이하 코로나 19)는 신종 호흡기 감염병이다. 우리나라는 최초 확진자가 2020년 1월 20일에 발생한 이후 2022년 5월까지 누적 확진자 약 1,800만 명, 누적 사망자 약 2만 명이 발생하였다[1-2].

전 세계적으로 코로나 19가 빠르게 확산되자 2020년 3월에 세계보건기구는 '감염병 세계적 유행'인 팬더믹(pandemic)을 선언했다. 코로나 19의 확산은 우리의 일상생활과 사회 전반에 많은 변화를 주었다. 이러한 시기에 사회적으로 관심 있는 의제가 무엇인지 정확하게 파악할 필요가 있다. 이를 위해 온라인 상 비정형 데이터를 분석하는 연구가 필요하다.

따라서 본 연구에서는 온라인 비정형 데이터를 텍스트 마이닝 기법을 활용하여 코로나 19 관련 사회적 이슈를 분석하고자 한다.

1.2. 연구 목적

연구 목적은 텍스트 마이닝 기법으로 뉴스 타이틀에서 코로나 19 관련 주요 이슈 키워드를 추출하여 파악하는 것이다. 코로나 19 관련 사회적 주요 이슈 키워드를 파악하고 이해도를 높여 효과적인 코로나 19 대응에 도움이 되고자 한다. 해당 목적을 실현하기 위해 첫째, 코로나 19 관련 뉴스 타이틀을 수집한다. 둘째, 텍스트 마이닝의 TF-IDF를 활용하여 주요 키워드를 추출한다. 셋째, 텍스트 네트워크 분석을 통해 이슈 키워드를 정확하게 파악한다.

연구 구성은 다음과 같다. 2장에서 코로나 19 관련 선행연구와 텍스트 마이닝 기법 중 TF-IDF와 텍스트 네트워크 분석을 소개한다. 3장에서는 연구 절차와 데이터 수집과 전처리를 설명한다. 4장에서는 TF-IDF 분석과 텍스트 네트워크 분석 결과를 설명하고, 마지막 5장에서는 결론을 제시한다.

제2장 이론적 배경

2.1. 코로나 19 관련 선행연구

코로나 19는 사회 전반적으로 영향력을 미치고 있기에 다양한 분야에서 연구가 진행되고 있다. 사회 분야로 진행된 연구를 살펴보면, 방역정책 실시에 따른 이해관계자들 간의 갈등과 코로나 19 확산 방지를 위한 봉쇄 정책으로 국가 간 갈등이 발생하였다. 또한 거리두기 강화로 사회적 비용이 증가하면서 전 세계적으로 정치적 편향 현상과 집단행동의 증가로 갈등이 더욱 심해 심화되었다[3-5].

코로나 19와 관련된 심리 연구를 살펴보면, 지속적인 사회적 단절 또는 고립은 불안, 우울, 스트레스 등과 같은 다양한 문제가 생길 수 있으며, 경기 침체로 인한 수입 감소는 극심한 분노와 스트레스와 밀접한 관련이 있음을 연구하였다. 또한 코로나 19가 장기화 되면서 우울과 불안 수준이 높게 나타났으며 남성보다 여성이 더 높은 수준으로 경험한 것으로 분석하였다[6-8].

언론과 관련된 연구로는 코로나 19 관련 뉴스 데이터를 일자 및 지역별 단어 빈도로 시각화하고, 특정 기간 내 발생한 주요 키워드와 코로나 19 감염자 및 확진자 추이를 분석하였다. 그리고 코로나 19 관련 뉴스에서 매체와 언론사의 정치 성향이 유의미한 영향이 있음을 확인하였다[9-10].

2.2. 텍스트 마이닝

텍스트 마이닝(text mining)은 비정형화 및 반 정형화 된 텍스트로부터 유용한 정보를 추출하는 기법으로 대량의 텍스트를 분석하고 정보를 탐색하는데 활용된다[11-12]. 본 연구에서는 데이터 분석을 위해 텍스트 마이닝 기법 중에서 TF-IDF와 텍스트 네트워크 분석 기법을 활용한다.

3.2.1. TF-IDF

특정 문서 내 주요 키워드를 추출하기 위한 기법인 TF-IDF(Term Frequency-Inverse Document Frequency)는 가중치를 부여하여 값을 계산한다. 즉, 특정 키워드가 특정 문서와 관련성이 높을 때 TF-IDF 값이 높아지며, 해당 문서의 주요 키워드가 된다. TF-IDF 값은 특정 단어의 상대적 등장빈도(Term Frequency)와 특정 단어가 전체 문서에서 등장하는 문서 비율의 역수(Inverse Document Frequency)를 곱하여 계산한다[13]. 본 연구에서 사용한 TF-IDF 공식은 아래와 같다.

$$F \quad IDF = TF(d,t) \times \log\left(\frac{D}{DF(t)}\right)$$

$F\ d,t)$: 서 d에 단어 t의 등장 횟수

$DF(t)$: 단어 t가 등장한 문서 수

D : 전체 문서 수

3.2.2. 텍스트 네트워크 분석

텍스트 네트워크 분석(text network analysis)은 텍스트 내 키워드를 노드(node), 노드 간 관계는 링크(link)로 표현한다. 키워드 간의 동시 등장빈도를 노드 간의 네트워크로 시각화하여 분석한다[14]. 텍스트 네트워크 분석의 핵심은 특정 키워드가 네트워크 내 하는 역할과 상대적인 영향력을 파악하는 것이다[15].

중심성(centrality)은 네트워크의 구조적 특성을 이해하기 위해 활용되는 대표적인 분석 지표이다. 중심성 분석을 통해 핵심 키워드가 무엇인지, 각 키워드가 네트워크 중심에 어느 정도 위치하는지를 알 수 있다. 중심성은 연결 중심성, 매개 중심성, 근접 중심성이 있다[16].

본 연구에서는 근접 중심성 지표로 네트워크 분석하였다.

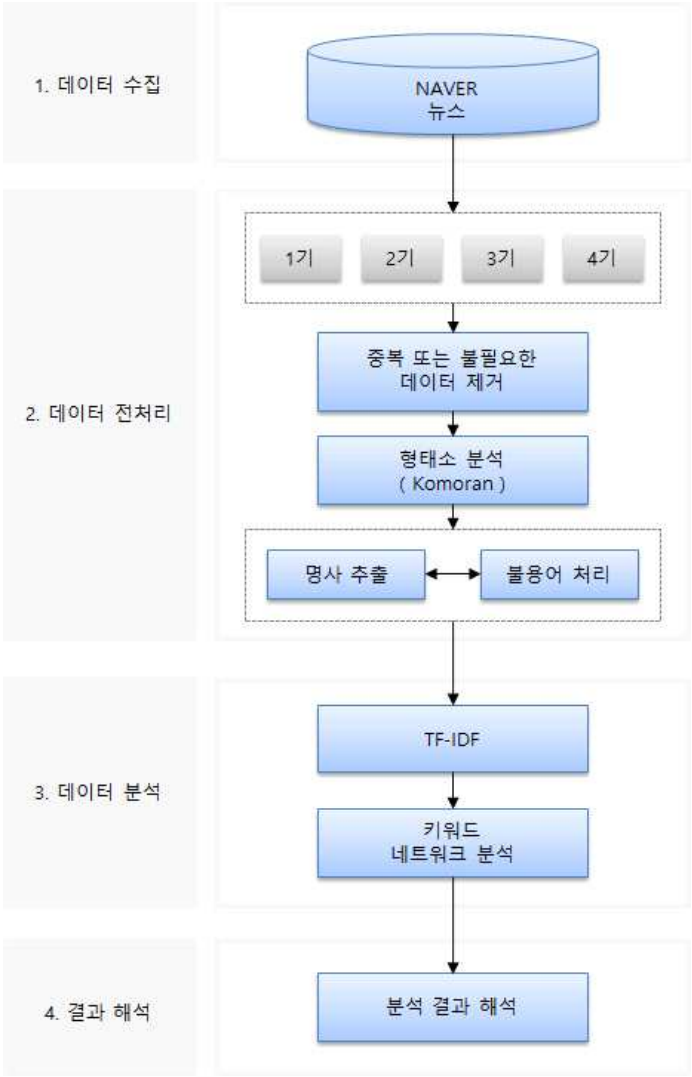
제3장 연구 방법론

3.1. 연구절차

본 연구는 포털사이트 네이버의 온라인 뉴스 타이틀을 대상으로 텍스트 마이닝 분석하여 코로나 19 관련한 주요 키워드를 파악한다. 이를 위한 연구절차는 [그림 1]과 같으며, 데이터 수집, 데이터 전처리, 데이터 분석, 결과 해석 단계로 구분된다.

구체적인 연구절차는 다음과 같다.

코로나 19 관련 네이버 온라인 뉴스를 수집하였다. 수집된 데이터는 4개의 시기로 분류한 후, 전처리 작업을 하였다. 전처리된 데이터를 TF-IDF 분석을 통해 상위 30개의 핵심 키워드를 추출하고, 키워드 간의 관계성 파악을 위한 네트워크 분석이 수행되었다. 마지막으로 분석된 결과를 해석하였다.



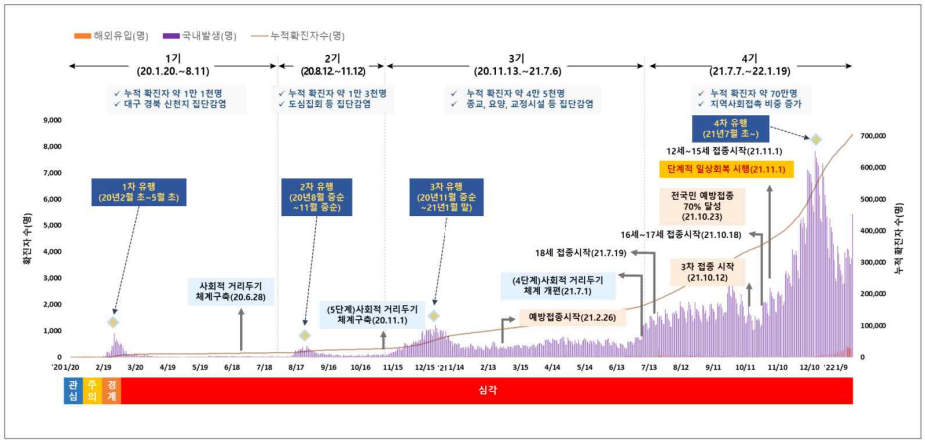
[그림 1] 연구절차

3.2. 데이터 수집

본 연구에서는 포털사이트 네이버의 온라인 뉴스 타이틀을 데이터 수집 대상으로 선정했다. 파이썬(Python)을 이용한 동적 웹 스크래핑(Dynamic Web Scraping) 방식을 활용하여 데이터를 수집하였다. 데이터 수집에 사용된 검색어는 '코로나 19', '코로나 바이러스'로 선정하였다. [그림 2]와 같이 온라인 뉴스의 언론사, 등록일, 제목을 데이터 세부 수집항목으로 수집하였다.

[그림 2] 데이터 수집 세부항목

데이터 수집을 위한 기간은 국내 첫 확진자 발생된 2020년 1월 20일부터 2022년 1월 19일이며, [17]의 코로나 19 유행시기 구분을 참고하여 1기~4기로 나누어 분석하였다([그림 3]).



[그림 3] 코로나 19 유행시기 구분

제1기는 국내 첫 코로나 19 확진자 발생한 2020년 1월 20일부터 부터 대구, 경북 지역 신천지와 전국적 집단발생이 시작된 2020년 8월 11일까지 총 205일이다. 제2기는 수도권 종교시설 및 대규모 도심 집회 등 집단발생이 시작된 2020년 8월 12일부터 2020년 11월 12일까지 총 93일이다. 제3기는 교정시설, 종교시설, 병원 및 요양시설 등 전국적 집단발생이 시작된 시점인 2020년 11월 13일부터 2021년 7월 6일까지 총 236일이다. 제4기는 오미크론형 변이 출현과 확산 시점인 2021년 7월 7일부터 2022년 1월 19일까지 총 197일이다([표 1]).

[표 1] 데이터 수집 기간

시기	기간	일수
제1기	2020.01.20 ~ 2020.08.11	205
제2기	2020.08.12 ~ 2020.11.12.	93
제3기	2020.11.13 ~ 2021.07.06	236
제4기	2021.07.07 ~ 2022.01.19	197
합계	**2020.01.20 ~ 2022.01.19**	**731**

수집된 등록일, 언론사, 기사제목은 [그림 4]와 같다.

	등록일	언론사	기사제목
0	2020.01.31.	부산일보언론사 선정	[전국 신종 코로나 현황] 확진환자 11명 정리(31일 오후 2시)
1	2020.01.31.	세계일보	수원 권선구 호매실동 시립금호어울림어린이집 폐쇄..원생 19명 '능동감시자...
2	2020.01.31.	경향신문언론사 선정	보스턴심포니 첫 내한공연 무산...신종 코로나바이러스 확산 우려
3	2020.01.31.	뉴스1	모모랜드, 신종 코로나 여파로 3월 日 팬미팅 잠정 연기
4	2020.01.31.	뉴시스언론사 선정	일본서 '신종 코로나' 확진자 14명으로 늘어
5	2020.01.31.	연합뉴스	태국서 신종코로나 첫 '2차 감염'..."중국 방문 안 한 택시기사"
6	2020.01.31.	연합뉴스	'주민 신종코로나 확진' 군산 모든 학교 졸업식 연기
7	2020.01.31.	이데일리	국내 '신종코로나' 확진자 4명 추가 발생...총 11명
8	2020.01.30.	연합뉴스언론사 선정	또 무증상 전파? 중국서 동창회 참석 6명 신종코로나 동시 확진
9	2020.01.31.	연합뉴스	중국 연구진 "신종코로나, 남자가 여자보다 더 잘 걸려"
10	2020.01.31.	뉴스1언론사 선정	신종 코로나 확산 우려에 대학가 졸업·입학식 줄취소(종합)
11	2020.01.31.	연합뉴스	이탈리아서도 신종코로나 첫 확진...20여개국 확산
12	2020.01.31.	서울신문	취소·중단...신종 코로나바이러스에 공연계도 비상
13	2020.01.31.	연합뉴스	민주평통, 신종코로나 비상에 내달 '1만명 전체회의' 또 취소
14	2020.01.31.	뉴시스	부산시, 신종 코로나 19명 능동감시...1명은 자가격리 조치
15	2020.01.31.	뉴시스	밀양시, 신종코로나 선별진료소 운영...대응체계 가동
16	2020.01.31.	연합뉴스	신종코로나 확산에 아이돌 일정 줄취소·연기 이어져
17	2020.01.31.	뉴스1	'신종 코로나' 전북대, 입학식 취소...졸업식은 대폭 축소
18	2020.01.31.	파이낸셜뉴스	北 "南서 '신종 코로나' 넘어올라"...금강산 시설 철거 연기 통보

[그림 4] 수집된 데이터

3.3. 데이터 전처리

분석하기 전 수집한 데이터는 전처리하였다. 전처리는 다음과 같이 5단계를 순서대로 진행했다.

(1) 수집된 데이터를 기간별로 분류
(2) 코로나 19와 관련 없거나 중복된 데이터 제거
(3) 형태소 분석
(4) 명사 추출
(5) 불용어 처리

첫째, 수집된 데이터를 4개의 기간으로 분류하였다. 기간별 분류된 데이터의 개수는 [표 2]와 같다.

[표 2] 기간별 데이터 규모

시기	기간	데이터 규모(개)
제1기	2020.01.20 ~ 2020.08.11	78,611
제2기	2020.08.12 ~ 2020.11.12.	36,000
제3기	2020.11.13 ~ 2021.07.06	108,000
제4기	2021.07.07 ~ 2022.01.19	52,000
합계	**2020.01.20 ~ 2022.01.19**	**274,611**

둘째, 코로나 19와 관련 없는 광고성 기사와 중복된 기사를 제거하였다.

셋째, KoNLPy 패키지의 Komoran 한글 형태소 분석기로 형태소 분석하였다. Komoran은 형태소와 품사 추출이 같이 되어 원하는 품사만 추출이 가능하다는 장점이 있다.

넷째, 품사 중 일반 명사와 고유 명사를 추출하였다.

마지막으로, 의미 분석에 기여하지 않는 숫자와 '명', '때' 등의 단어를 불용어 처리하였다. 데이터 전처리의 정확도를 위해 명사 추출과 불용어 처리는 반복하여 수행하였다.

제4장 데이터 분석 결과

4.1. TF-IDF 분석 결과

연구에서는 코로나 19 유행시기별 핵심적인 비중을 차지하는 키워드를 추출하여 분석하였다. 키워드 추출은 TF-IDF 값이 큰 상위 30개로 하였다. 각 시기별 핵심 키워드 분석 결과는 다음과 같다.

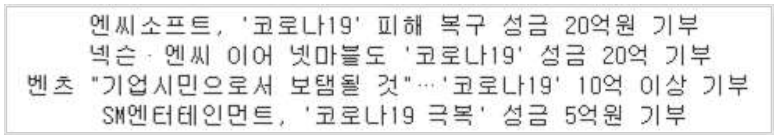
제1기의 핵심 키워드와 가중치 값을 [표 3]에 나타냈다. 제1기에는 전반적으로 확진·극복·지원·확산·감염과 같은 키워드가 다수 도출되었다. 이는 국내에 코로나 19 확진자가 처음 발생한 후, 코로나 19 상황에 대한 사람들의 관심을 보여준다. 제1기에만 도출된 키워드는 여파·피해·개발·대구·기부가 있다. 이는 코로나 19 증상과 집단 발생 상황 등 코로나 19를 이해하려는 현상이 나타나는 것을 알 수 있다.

[표 3] 제1기 TF-IDF 키워드 추출

순위	키워드	TF-IDF	순위	키워드	TF-IDF
1	확진	25362.33	16	지역	7412.43
2	극복	14956.01	17	추가	7400.74
3	지원	13093.58	18	**여파**	7395.08
4	확산	12027.59	19	검사	7331.83
5	감염	11361.19	20	백신	7224.39
6	대응	11167.69	21	**피해**	6935.04
7	신규	11081.19	22	해외	6930.54
8	발생	10061.93	23	위기	6850.75
9	치료	9971.00	24	정부	6848.49
10	사망	9218.44	25	세계	6703.93
11	방역	8683.73	26	**개발**	6694.06
12	종합	8625.03	27	**대구**	6545.51
13	환자	8457.39	28	한국	6471.40
14	국내	8368.57	29	**기부**	6456.97
15	치료제	7597.86	30	기업	6359.52

주) 파란색 키워드 : 각 시기별 처음 등장한 키워드

특히, 기부 키워드는 [그림 5]와 같이 코로나 19 성금 기부 기사가 많아 도출되었다.



[그림 5] 기부 관련 기사 타이틀

제2기에는 신규·발생·백신·방역·치료와 같은 키워드가 많이 도출되었다. 수도권에서 코로나 19 집단 감염이 처음으로 발생하면서 확산하는 코로나 19 상황에 대한 관심이 나타난 것으로 보인다. 특히, 제2기에만 나타난 키워드는 자릿수·긴급이 있다. 이는 집단 감염으로 코로나 19 확진자의 빠른 증가와 해외 코로나 19 백신 긴급 승인과 관련된 키워드로 분석된다.

[표 4] 제2기 TF-IDF 키워드 추출

순위	키워드	TF-IDF	순위	키워드	TF-IDF
1	확진	12889.07	16	대응	3695.16
2	신규	7442.37	17	국내	3649.54
3	확산	7169.80	18	추석	3622.48
4	발생	6247.05	19	위기	3494.87
5	백신	5956.43	20	속보	3449.44
6	감염	5924.90	21	임상	3422.43
7	방역	5159.72	22	환자	3375.26
8	극복	4822.33	23	사망	3244.43
9	치료	4820.94	24	병원	3126.95
10	지원	4772.63	25	진단	3102.48
11	검사	4736.79	26	서울	3097.11
12	지역	4128.91	27	자릿수	3072.33
13	치료제	3916.39	28	세계	3008.10
14	종합	3788.08	29	해외	2929.19
15	추가	3712.47	30	긴급	2913.15

주) 파란색 키워드 : 각 시기별 처음 등장한 키워드

제3기에는 누적·속보·사망과 같은 키워드가 다수 도출되었다. 코로나 19 확진자가 전국적으로 집단 발생되면서 누적 확진자 수와 사망자 수에 관심이 높아진 것으로 해석할 수 있다. 제3기에 유일하게 도출된 키워드로 키트·진단·병원·센터·자가가 있다. 이는 코로나 19 백신 접종과 진단 키트에 대한 관심이 높아지면서 처음 등장한 것으로 해석된다.

[표 5] 제3기 TF-IDF 키워드 추출

순위	키워드	TF-IDF	순위	키워드	TF-IDF
1	확진	38726.378	16	극복	11004.06
2	백신	34114.19	17	종합	10859.772
3	접종	29355.35	18	사망	10679.91
4	신규	25608.62	19	서울	10438.14
5	검사	18512.08	20	대응	10430.36
6	발생	17787.44	21	국내	10203.70
7	감염	15490.42	22	치료제	9523.52
8	확산	14403.70	23	지역	8810.02
9	지원	14234.13	24	**키트**	8652.75
10	추가	13771.16	25	**진단**	8621.30
11	예방	13239.09	26	**병원**	8510.61
12	방역	13171.59	27	정부	8339.98
13	치료	12160.26	28	의료	8183.58
14	누적	12085.39	29	**센터**	8160.92
15	속보	11289.94	30	**자가**	8022.77

주) 파란색 키워드 : 각 시기별 처음 등장한 키워드

제4기에는 검사·방역·추가·종합·대응과 같은 키워드가 도출되었다. 오미크론형 변이 출현과 확산되는 시기로 방역과 대응에 관심이 높아진 것으로 해석된다. 특히 제4기에 유일하게 나타난 키워드인 위드와 회복은 방역대책이 위드코로나로 변화되면서 일상회복에 대한 관심이 높아진 것으로 분석된다.

[표 6] 제4기 TF-IDF 키워드 추출

순위	키워드	TF-IDF	순위	키워드	TF-IDF
1	확진	19165.98	16	최다	5232.72
2	백신	13991.62	17	예방	4932.69
3	신규	13455.54	18	극복	4743.32
4	접종	13250.89	19	사망	4650.55
5	발생	8498.56	20	누적	4519.94
6	검사	8334.65	21	대응	4484.31
7	위드	8204.46	22	오후	4318.31
8	확산	8107.60	23	수도	4135.68
9	감염	7820.09	24	서울	4102.08
10	치료	7814.52	25	국내	4076.43
11	지원	7372.74	26	회복	4073.56
12	방역	7006.58	27	정부	4047.37
13	추가	6830.35	28	의료	4013.22
14	치료제	5881.60	29	추석	3893.02
15	속보	5481.76	30	임상	3827.83

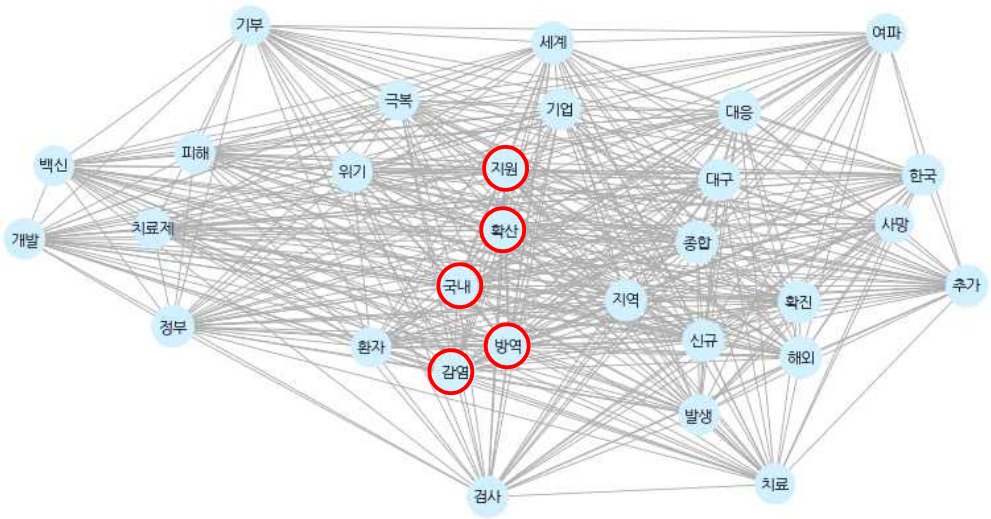
주) 파란색 키워드 : 각 시기별 처음 등장한 키워드

모든 시기에 공통적으로 도출된 키워드는 감염·검사·국내·극복·대응·발생·방역·백신·사망·신규·지원·추가·치료·치료제·확산·확진이 있다. 시기를 불문하고 코로나 19 치료와 방역에 높은 관심이 있음을 확인할 수 있다.

4.2. 텍스트 네트워크 분석 결과

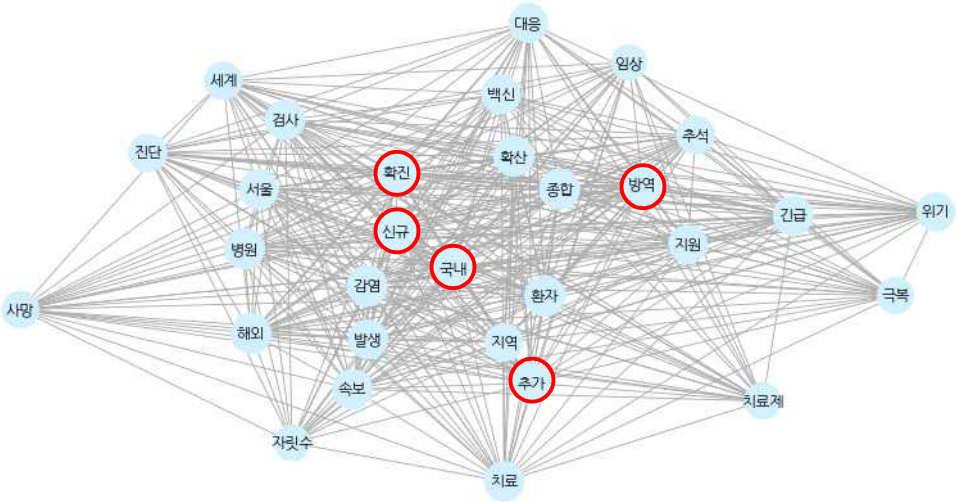
본 연구에서는 핵심 키워드들의 관계와 네트워크 구성을 알아보고자 시기별 키워드의 네트워크 분석을 하였다. TF-IDF 상위 30개의 키워드로 네트워크 노드를 구성한 후, 근접 중심성으로 네트워크 주요 키워드를 파악하였다.

근접 중심성으로 제1기 텍스트 네트워크를 분석한 결과, 상위 5개 키워드는 국내·감염·방역·지원·확산으로 나타났다. 이 키워드로 표현된 노드들이 제1기 텍스트 네트워크의 중심 역할을 하고, 이는 국내 코로나 19 확진자가 처음 발생한 이후 정부 대응에 높은 관심이 있음을 의미한다.



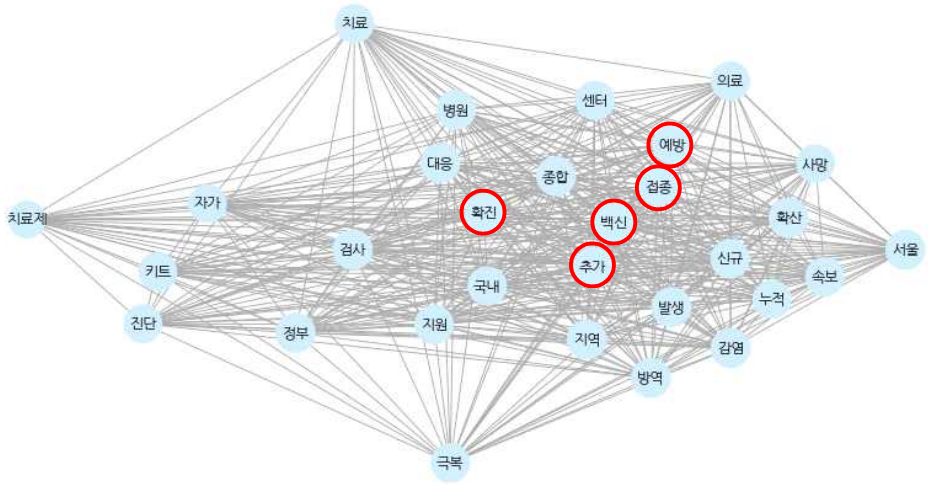
[그림 6] 제1기 텍스트 네트워크

근접 중심성으로 제2기 텍스트 네트워크를 분석한 결과, 상위 5개 키워드는 국내·신규·방역·추가·확진으로 나타났다. 이 키워드들이 제2기 텍스트 네트워크에서 가장 영향력이 높으며, 이는 수도권 종교시설 및 대규모 도심 집회로 코로나 19 확진자 수 증가가 사회적 이슈로 부각되었음을 나타낸다.



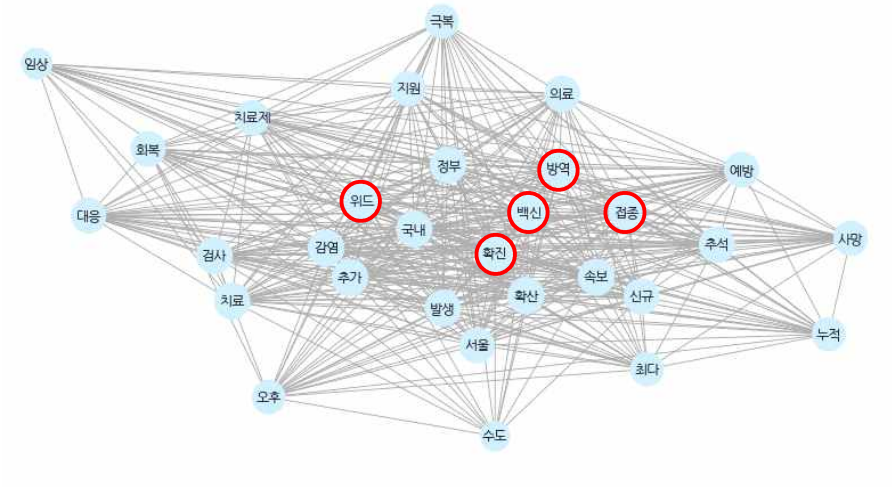
[그림 7] 제2기 텍스트 네트워크

근접 중심성으로 제3기 텍스트 네트워크를 분석한 결과, 상위 5개 키워드는 백신·예방·접종·확진·추가로 나타났다. 이 키워드들이 제3기 텍스트 네트워크에서 가장 큰 영향력을 미치며, 이는 코로나 19 백신 개발과 접종에 높은 관심이 있는 것으로 해석된다.



[그림 8] 제3기 텍스트 네트워크

근접 중심성으로 제4기 텍스트 네트워크를 분석한 결과, 상위 5개 키워드는 위드·방역·백신·접종·확진로 나타났다. 이 키워드들이 제4기 텍스트 네트워크에서 가장 중점적인 역할을 하며, 코로나 19 백신 접종률이 높아지면서 위드 코로나 정책이 시행됨에 따라 일상으로 복귀에 관심이 높은 것으로 분석된다.



[그림 9] 제4기 텍스트 네트워크

제5장 결론

5.1. 연구 결과 요약

본 연구는 온라인 뉴스 타이틀을 활용하여 코로나 19 상황에서 사회적인 주요 이슈를 분석하였다. 분석하기 위한 데이터 수집은 파이썬(Python)을 이용한 동적 웹 스크래핑 방식을 활용하였다. 수집된 데이터는 시기를 4개로 분류하고, 이후 전처리와 텍스트 마이닝 분석을 차례로 수행했다. 텍스트 마이닝 관련 여러 기법 중에서 TF-IDF와 텍스트 네트워크 분석을 활용하였다. TF-IDF로 각 시기별 30개의 주요 키워드를 추출하고, 이 키워드로 텍스트 네트워크 분석하였다.

[표 7]은 TF-IDF와 텍스트 네트워크 분석으로 추출된 코로나 19 시기별 핵심 키워드이다. 이를 종합적으로 분석한 결과는 다음과 같다.

[표 7] 시기별 핵심 키워드

시기	핵심 키워드
제1기	**국내**·감염·**방역**·지원·확산
제2기	**국내**·신규·**방역**·추가·확진
제3기	**백신**·예방·**접종**·**확진**·추가
제4기	위드·방역·**백신**·**접종**·**확진**

첫째, 제1기와 제2기에 공통적으로 출현된 핵심 키워드로는 국내와 방역이 있었다. 제1기와 제2기는 2020년 1월 20일부터 2020년 11월 12일까지로 2020년 한 해는 국내 방역에 높은 관심이 있었음을 알 수 있다. 이 시기에 시행되었던 국내 방역 지침은 다음과 같다. 2020년 2월 신천지 교회에서 시작된 1차 대유행으로 정부는 3월 21일에 교회, 클럽, 헬스장 등 다인이용시설의 운영을 통제하는 방역 지침을 발표했었다. 이후 5월 6일부터 방역 대응을 위한 생활 속 거리두기가 시행되면서 대중교통 이용 시 마스크 미착용일 경우 탑승이 거부되었다. 거리두기 명칭은 6월 28일 사회적 거리두기로 통일되었고, 수도권 감염 확산으로 8월 28일 강화된 사회적 거리두기가 시행되면서 방역 조치가 이전보다 강화되었다.

둘째, 제3기와 제4기에서 동시에 나타난 주요 키워드로는 백신·접종·확진이 있었다. 제3기와 제4기는 2020년 11월 13일부터 2022년 1월 19일까지로, 이 시기에는 확진자가 증가하면서 코로나 19 백신 사용 승인과 접종에 사회적인 관심이 높았음을 알 수 있다. 이와 관련된 뉴스 타이틀을 간략히 정리하면 다음과 같다. 2020년 12월에 영국이 세계 최초로 화이자 백신 사용을 승인하였고, 유럽연합이 코로나 19 백신 사용을 공식 승인했었다. 우리나라는 2021년 2월 26일부터 코로나 19 백신 접종이 요양병원과 요양시설 등의 만 65세 미만 입소자와 종사자를 대상으로 처음 시작되었다.

5.2. 시사점

본 연구는 코로나 19와 관련된 뉴스 타이틀에 텍스트 네트워크 분석을 활용하여 각 시기별 이슈가 되는 키워드를 추출했다. 이를 통해 각 시기별 사회적으로 관심도가 높은 키워드를 확인할 수 있었다.

또한 이후 텍스트 네트워크 분석을 수행하여 시간에 따른 코로나 19 관련 이슈가 어떻게 변화하는지를 분석한다면 사회적 현상을 파악하는데 도움을 줄 수 있을 것으로 기대된다.

5.3. 연구 한계

데이터 수집을 뉴스 타이틀 외에도 학술 논문, 페이스북 등 다양하게 수집하여 분석할 필요성이 있다. 수집된 데이터가 다양하다면, 이슈가 되는 사회 현상을 더 정확하게 파악할 수 있을 것으로 예상된다. 또한 텍스트 마이닝의 다른 분석 기법을 활용한다면 더 유의미한 분석 결과가 도출될 것으로 보인다.

[참고 문헌]

[1] WHO. COVID-19 Weekly Epidemiological Update(Edition 77, published 18 January 2022)[Internet], Available from: https://www.who.int/publications/m/item/weekly-epidemiological-updateon-covid- 19---18-January-2022.

[2] Coronavirus Disease-19. Republic of Korea. http://ncov.mohw.go.kr.

[3] Bae, K. B. (2022). COVID-19 and Social Conflict: Focusing on Topic Modeling. *Dispute Resolution Studies Review, 20*(1), 205-228.

[4] Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., & Yang, D. (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *Journal of Public Economics, 191*, 104254.

[5] Koetke, J., Schumann, K., & Porter, T. (2021). Trust in science increases conservative support for social distancing. *Group Processes & Intergroup Relations, 24*(4), 680-697.

[6] Wang, X., Gao, L., Zhang, H., Zhao, C., Shen, Y., & Shinfuku, N. (2000). Post earthquake quality of life and psychological well being: Longitudinal evaluation in a rural community sample in northern China. *Psychiatry and Clinical Neurosciences, 54*(4), 427-433.

[7] Carvalho Aguiar Melo, M., & de Sousa Soares, D. (2020). Impact of social distancing on mental health during the COVID-19 pandemic: An urgent discussion. *International Journal of Social Psychiatry, 66*(6), 625-626.

[8] Lee, Y. J. Kim, Hwang, H. H. Nam, S. K. Jung, D. S. (2021). A Longitudinal Comparative Study of Two Periods regarding the Influences of Psycho-Social Factors on Emotional Distress among Korean Adults during the Corona virus Pandemic(COVID-19). *THE KOREAN JOURNAL OF CULTURE AND SOCIAL ISSUES, 27*(4), 629-659.

[9] Hur, T. S. & Hwang. I. Y. (2022). Covid 19 News Data Analysis and Visualization. *Journal of the Korea Society of Computer and Information, 27*(4), 37-43.

[10] Park, J. H. (2020). A Comparative Study on the 'Corona19' News Frame Based on Ideological Orientation of Media. *Korean Society For Journalism And Communication Studies, 64*(4), 40-85.

[11] Feldman, R. & Dagan, I. (1995). Knowledge Discovery in Textual Database

s. *KDD, 95*, 112-117.

[12] Sebastiani, F. 2002. Machine learning in automated text categorization. *ACM Computing Surveys (CSUR), 34*(1), 1-47.

[13] Ramos, J. 2003. Using tf-idf to determine word relevance in document queries. *In Proceedings of the first instructional conference on machine learning, 242*, 133-142.

[14] Diesner, J. & K. M. Carley. (2005). Revealing Social Structure from Texts: Meta-Matrix Text Analysis as a Novel Method for Network Text Analysis. *Causal Mapping for Research in Information Technology*, 81-108.

[15] Chung, P. L., Ahn, H. C., Kwahk, K. Y. (2019). Identification of Core Features and Values of Smartphone Design using Text Mining and Social Network Analysis. *Korean Jouranl of Business Administration, 32*(1), 27-47.

[16] Freeman, L. C. (2005). Graphical techniques for exploring social network data. Models and methods in social network analysis.

[17] Korea Disease Control and Prevention Agency (KDCA). (2022). Two-year report of COVID-19 outbreak from January 20, 2020 to January 19, 2022 in the Republic of Korea.