

Covid 19 News Data Analysis and Visualization

Tai-Sung Hur*, In-Yong Hwang*

*Professor, Dept. of Computer Science, Inha Technical College, Incheon, Korea

*Student, Dept. of Computer Science, Inha Technical College, Incheon, Korea

[Abstract]

In this paper, we calculate the word frequency by date and region using news data related to COVID-19 distributed for about 8 months from December 2019 to July 2020, and visualized the correlation with the current state data of COVID-19 patients using the results. News data was collected from Big Kids, a news big data system operated by the Korea Press Promotion Foundation. The visualization system proposed in this paper shows the news frequency of the selected region compared to the overall region, the key keyword of the selected region, the region of the main keyword, and the date change of the selected region. Through this visualization, the main keywords and trends of COVID-19 confirmed and infected people can be identified for previous events.

▶ **Key words:** COVID-19, press analysis, Big Kinds, data analysis, visualization

[요 약]

본 논문에서는 2019년 12월부터 2020년 7월까지 약 8개월간 유통되었던 코로나19와 관련된 뉴스데이터를 이용하여 일자 및 지역별로 단어에 대한 빈도를 구하고, 결과를 활용하여 코로나19 환자에 대한 현황 데이터와의 상관관계를 시각화하였다. 뉴스데이터는 한국언론진흥재단에서 운영하고 있는 뉴스 빅데이터 시스템 '빅카인즈'에서 수집된 데이터를 활용하였다. 본 논문에서 제안하는 시각화 시스템은 지역과 기간을 선택하면 분석한 결과를 이용하여 전체 지역 대비 선택한 지역의 뉴스 빈도수, 선택한 지역의 주요 키워드, 주요 키워드의 지역별, 일자별 변화 등을 보여 주고 있다. 이러한 시각화를 통하여 이전에 발생하였던 사건에 대해 주요 키워드와 코로나19 확진자 및 감염자 추이를 확인할 수 있다.

▶ **주제어:** 코로나19, 언론보도 분석, 빅카인즈, 데이터 분석, 시각화

-
- First Author: Tai-Sung Hur, Corresponding Author: In-Yong Hwang
 - Tai-Sung Hur (tshur@inhac.ac.kr), Dept. of Computer Science, Inha Technical College
 - In-Yong Hwang (hiyong5978@gmail.com), Dept. of Computer Science, Inha Technical College
 - Received: 2022. 02. 24, Revised: 2022. 04. 05, Accepted: 2022. 04. 06.

I. Introduction

세계보건기구에서 2020년 3월 11일 코로나바이러스 감염증-19(COVID-19, 이하 코로나19)를 감염병의 세계적 대유행 단계인 글로벌 팬데믹으로 공식 선언하였다[1]. 2021년 5월 21일 기준 전 세계 220개국으로 확산되어 확진자 165,179,612명, 사망자 수는 3,574,776명을 기록하고 있고 사망률은 약 2.08%이다[2]. 2020년은 세계적으로 코로나19의 시대를 경험하고 있으며, 우리나라에서는 2020년 1월 20일 한 중국인 여성이 신종 코로나19 감염자로 확진이 되어 대한민국에 첫 번째 감염자가 되었다.

이런 상황 속에서 국내 언론들은 코로나19의 감염 확산 경과에 따라 관련 뉴스의 양도 다르게 나타나는 등 언론의 고유한 역할을 충실히 해내고 있다. 코로나19 관련 뉴스는 짧은 기간에 사회적 관심을 고조시킴과 동시에 뉴스의 양도 많으므로 뉴스의 가치가 매우 높다[3].

보통 특정 뉴스에서 사용한 단어의 빈도를 주로 측정한다[4], 세메트코와 발켄버그[5]나 아이엔저와 사이먼[6]의 연구 방법에 근거하여 언론 보도 프레임 분석 연구가 활발하게 이루어졌다. 하지만 프레임 분석은 전통적으로 시간이 많이 소요된다는 단점과 동시에 연구자의 주관적인 생각이 개입될 여지가 있어 분석결과의 신뢰도를 검증해야 한다는 한계가 있다[7].

이러한 문제점을 극복하기 위해 최근에는 시간과 비용이 절약되고 연구자의 개인이 최소화되어 객관성을 담보하기 때문에 컴퓨터를 이용한 자동화된 분석방법이 많이 사용되고 있다. 특히 방대한 양의 뉴스는 연구의 질을 높일 수 있으며, 자연어 처리 방식을 활용함에 따라 연구자간 신뢰도를 측정할 필요가 없는 빅카인즈(BIGKinds)를 이용한 뉴스 추출 방식은 최근 다양한 분야에서 많이 활용되고 있다[8].

본 논문에서는 한국언론진흥재단에서 운영하는 뉴스 빅데이터 시스템인 '빅카인즈'에서 수집되어 제공된 코로나19와 관련뉴스 데이터를 활용하여 지역 및 기간별로 단어를 분류하고, 분류된 데이터를 활용하여 이를 시각화할 것이다. 빅카인즈에서 제공된 데이터는 CSV형식으로 이전에 발생한 사건에 대해서만 시각화를 해볼 수 있지만, API 서비스를 연동하여 활용하면 최근 발생한 사건에 대해서도 코로나19 관련 뉴스에서 사용된 주요 키워드와 그에 따른 확진자 및 감염자 추이를 확인할 수 있다.

II. Related works

통계청에서 발간한 2020년 통계에 의하면 경상북도 지역은 65세 이상의 고령인구 비중이 20.7% 수준으로 2020년 4월 30일 0시 기준으로 경상북도 코로나19 사망률(3.8명)을 나타내고 있으며, 전국 사망률(2.3명)보다 높은 수준이다. 나이별 코로나19 사망자 비율은 80세 이상이 58.6%로 절반 이상을 차지하고 있다[9]. 이에 지역에 기반을 둔 코로나19 발생률과 사망률 관련 요인에 관한 연구가 진행되고 있다.

"경상북도 지역의 코로나19 발생률 및 사망률 관련 요인"에 관한 연구[10]는 사망에 영향을 미치는 요인을 확인하기 위해 종속변수로 확진 판정 후 사망 여부를 사용하였고 독립 변수로 나이, 성별, 기저질환 유무와 같은 개인 특성과 거주 형태와 같은 지역 특성을 이용하였다.

"COVID-19로 인한 사망자 수의 결정요인: 대유행 초기 단계에서 저소득 국가와 고소득 국가 간의 차이"에 관한 연구[11]는 코로나19 사망자수에 영향을 미치는 요인을 저소득 및 고소득 국가에 중점을 두었다. 그 결과 교통인프라 시스템, 기대수명 및 보건 관련 지출비율과 같은 요소가 고소득 및 저소득 국가의 사망자수 차이에 영향을 미치는 주요 요인으로 도출되었다.

디지털 정보 기술의 발전은 수많은 언론으로부터 매일 같이 발간되는 엄청난 양의 뉴스 정보를 빅데이터(big data)로 저장할 수 있게 되었으며, 이렇게 저장된 대규모의 뉴스 빅데이터는 단순히 뉴스 검색뿐만 아니라 원하는 뉴스를 빠르고 간편하게 추출하여 다양한 정보를 분석할 수 있도록 시스템을 구축하기도 한다.

한국언론진흥재단에서 구축한 뉴스 빅데이터인 빅카인즈는 기존에 제공되고 있는 단순 뉴스 검색 서비스인 KINDS에 빅데이터의 분석력을 도입하여 발전된 뉴스분석 서비스로 뉴스 정보 제공과 검색 서비스를 제공하고 있다. 빅카인즈에서는 비정형 텍스트로 구성되어 있는 뉴스를 정형화된 데이터로 전환하여 많은 사회 현상을 다양하게 분석할 수 있는 기초자료로 활용할 수 있도록 도움을 준다.

본 논문에서는 분석된 데이터를 Line 그래프와 Bar 그래프로 보여줄 뿐만 아니라 분석된 데이터를 이해하는데 도움을 줄 수 있도록 지도 및 네트워크 차트 활용하여 다양한 방법으로 데이터를 시각화하였다.

III. The Proposed Scheme

1. System Environment

본 시스템은 데이터 가공 및 분석을 위해 파이썬 (Python) 언어와 관련된 라이브러리를 활용하고 있으며, 분석된 데이터의 결과는 데이터베이스에 저장되도록 구현하였다. 저장되어 있는 데이터는 웹을 통해 확인할 수 있으며, 불러온 데이터는 이해하기 쉽도록 Chart.js와 D3.js 라이브러리를 이용하여 시각화하였다.

Table 1. Packages used in the system

Kind	Name	Content
Data Analysis	pandas	package for analyzing csv files
	pymysql	package for storing results in DB
Result Store	MariaDB	program for storing analyzed results
Data Visualization	php	package for obtaining analyzed results
	chart.js	package to show a basic chart
	d3.js	a package for representing maps and network charts

2. Collected Data

본 논문에서는 공공데이터 포털에 공개되어 있는 "한국 언론진흥재단_뉴스빅데이터_메타데이터_코로나"를 사용한 다[12]. 데이터 세트는 2019년 12월부터 2020년 7월까지의 데이터로 구성되어 있으며 일자, 키워드, 뉴스의 분류 등 약 200만개의 데이터를 사용한다.

코로나19와 관련된 뉴스 데이터와 연관 분석을 하기 위해 질병관리본부에서 보고한 대한민국의 코로나19 환자에 대한 현황 데이터를 이용한다. 코로나19 현황 데이터는 일자, 지역, 확진자 수, 사망자수, 격리해제 수의 데이터를 포함하고 있다[13].

코로나19 환자에 대한 현황 데이터는 2020년 2월 17일 이후로 취합된 데이터를 제공하고 있다. 하지만 뉴스데이터는 2019년 12월부터 수집된 데이터를 제공하고 있기에 사용자가 기간을 조건으로 검색할 때 2020년 2월 17일 이전의 기간을 선택하게 되면 코로나19 환자에 대한 현황 데이터는 시각화 하지 않도록 하였다.

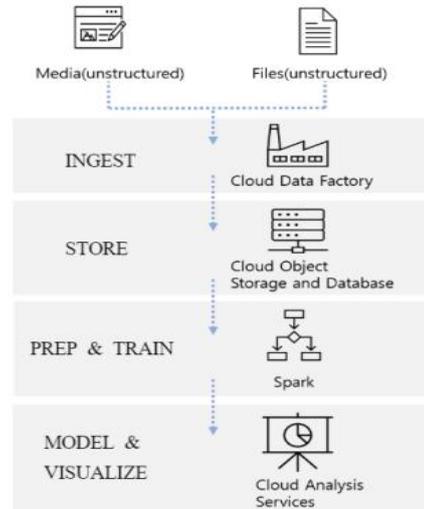


Fig. 1. Data Pipeline

3. Data Processing

본 논문에서 사용하는 뉴스데이터 중 일부는 데이터 형식이 다른 CSV 파일과 내용이 상이하여 분석 프로그램 실행 중 형식이 일치하지 않아 예외가 발생할 수 있다. 따라서 데이터 처리하기에 앞서 이러한 데이터가 있는지 확인 후 발견이 되면 정규화를 하였다.

```
File : covid19_20200501_20200531.csv
0      2020-05-01
1      2020-05-01
2      2020-05-01
3      2020-05-01
4      2020-05-01
5      2020-05-01
6      2020-05-01
7      2020-05-01
8      2020-05-01
```

Fig. 2. Data before normalization

Fig. 2는 2020년 5월의 데이터 중 일자를 출력한 것이다. 5월의 날짜 데이터는 형식이 "yyyy-mm-dd"의 형식으로 되어 있지만, CSV 파일 대부분은 날짜 데이터의 형식이 Fig. 3과 같은 "yyyymmdd"형식으로 구성되어 있어, Fig. 2처럼 다른 형식으로 발견되면 날짜의 포맷을 통일하기 위해 변환하였다.

```
File : covid19_20200501_20200531.csv
0      20200501
1      20200501
2      20200501
3      20200501
4      20200501
5      20200501
6      20200501
7      20200501
8      20200501
```

Fig. 3. Date data after format conversion

다음으로 키워드에 대한 데이터를 확인해 보면 키워드 데이터는 콤마(,)로 구분이 되어 줄 바꿈 없이 한 줄로 구성이 되어 있는 것이 정상이다. 하지만 일부 라인에서는 Fig. 4와 같이 콤마 대신 불필요한 줄 바꿈으로 키워드가 들어가 있는 경우를 발견하였다.

줄 바꿈이 되어 있는 데이터를 하나의 라인으로 변경하여 정규화하기에는 정상적으로 들어가 있는 라인에 영향이 생겨 분석된 결과에 영향을 줄 수 있어, 이러한 문제로 인해 발견된 데이터는 분석할 때 제외하고 처리하는 방법으로 해결을 하였다.

```

127531 회생자", "당선, 민주당, 전남, 10명, 목소리, 민선, 정치, 5 18민주, 묘지, 참해, 선대위, 선거, 국회, 선거대책위원회, 더불어민주당, 전남도, 홍선, 선거, 대책, 위원회, 5 18민주, 이남, 광주, 북구, 시, 이계호, 서삼석, 김현이, 주철현, 김희재, 소병철, 서동홍, 신정훈, 김승남, 윤재갑, 장선인, 해단남, 선철, 국환, 장선, 국남, 극복, 대통령, 경치, 국민, 경치, 경평, 발전, 국가, 군행, 심판, 국민, 동행, 투, 사퇴, 안정환, 구술, 문재인, 선공, 발전, 정외, 홍여춘, 지역, 주력, 역설, 민선, 개세, 국민, 주인, 단식, 직전, 세월, 참사, 회생자, 목숨, 해단식, 윤호, 박수, 유치, 원형, 방사광, 가속기, 호남", "전남 문재인, 홍여춘, 지역구, 국회의원, 대통령, 김희재, 주철현, 김현이, 소병철, 서삼석, 서동홍", "제21 당선인들이 16일 ..민심을 받들어 대통령의 정치를 맡겼다고 다짐했다. ..더불어민주당, 5 18민주묘지를 참배한 뒤 민주주의 문 앞에서 선대위 해단식을 하고 전남 지역민에게 감
1613472693428/11
127532 http://www.bigkinds.or.kr/news/newsDetailView.do?newsId=02100311.20200416092816001.2
127명 추가, 국제>일본, 국제>아시아, 국제>중남미, ..., "일본
127533 도쿄", "일본
127534 도쿄도
127535 오사카부", "549명, 일본, 코로나19, 신규, 확진자, 도쿄, 추가, NHK, 일본, 감염증, 신종, 코로나바이러스, 오사카부, 74명, 일본, 확진자, 코로나19, 누적, 9, 434명, 다이아몬드, 프렌세스호, 모함, 사망자, 1 도쿄도, 오사카부, 감염증, 도쿄, 추가, 신종, 누적, 신규, 포함, 다이아몬드, 보도, 확인", "16일 NHK 549명으로 확인됐다고 보도했다. 도쿄도의 신규 확진자가 127명, 오사카부가 74명이었다. 스운 포함)이다. 사망자는 총 191명이다. /감연하기자 geona@sedaily.com, http://www.sedaily.com/
127536 http://www.bigkinds.or.kr/news/newsDetailView.do?newsId=01100801.20200416071506001.2
제>중권, 중시, 경제>산업, 기업, 경제>금융, 재테크, ..., "계산
127537 회계
127538 미국
127539 일본
127540 경기", "스플라
127541 국민연
127542 인도
    
```

Fig. 4. Data before normalization

Table 2는 Fig. 4에서처럼 잘못 들어가 있었던 라인을 제거하고 나온 Record의 수를 파일별로 처리 전과 처리 후로 나타낸 것이다. 총 2,278,567건의 뉴스데이터가 공개되어 있었지만, 이 중 일자 혹은 키워드 잘못 들어가 있는 라인을 제외하고 약 788,372건을 사용하였다.

Table 2. Number of records in the file

File Name	Before	After
20191201_20200131.csv	17,606	17,605
20200201_20200229.csv	108,219	108,218
20200301_20200331.csv	184,089	184,088
20200401_20200430.csv	736,501	145,765
20200501_20200531.csv	123,542	122,482
20200601_20200630.csv	659,859	108,959
20200701_20200731.csv	448,751	101,255
Total	2,278,567	788,372

Table 3는 앞에서 언급한 문제에 대하여 처리한 뒤, 본 논문에서 제시하는 데이터 분석 프로그램을 활용하여 데이터베이스에 저장한 후 일자별, 지역별로 데이터를 그룹화 하여 단어의 합계가 가장 많은 상위 10개를 보여주고 있다.

Table 3. Number of words by date and region

Date	Region	Word Total
2020-05-28	Gyeonggi-do	193,051
2020-02-26	Daegu Metropolitan City	180,481
2020-08-19	Gyeonggi-do	177,097
2020-02-25	Daegu Metropolitan City	174,951
2020-08-18	Gyeonggi-do	174,897
2020-05-31	Gyeonggi-do	164,800
2020-02-27	Daegu Metropolitan City	164,392
2020-02-24	Daegu Metropolitan City	164,375
2020-03-02	Gyeonggi-do	163,498
2020-03-03	Gyeonggi-do	162,900

4. System Implementation

본 논문에서 제안하는 "코로나19 뉴스데이터 시각화" 시스템은 Linux, Apache, MariaDB, PHP를 이용한 웹 어플리케이션 기반의 시스템이다.

데이터 시각화는 Javascript 기반의 Chart.JS와 D3.js를 이용하였으며, 데이터베이스는 MariaDB로 구성하였다. 공공데이터 포털에 공개된 2019년 12월부터 2020년 8월까지 뉴스데이터와 코로나19 현황 데이터가 분석되어 테이블 형태로 저장되어 있다.



Fig. 5. Visualization web system

Fig. 5는 웹으로 구현된 시각화 시스템을 나타낸 것이다. 데이터를 처리하는 부분은 PHP와 HTML5, Javascript로 구현을 하였으며 4가지 모듈로 구성된다.

- 검색 모듈(Search Module): 데이터베이스에 저장된 값들을 그룹화 하여 사용자가 입력할 수 있는 형태로 만들어 놓았다. 사용자가 입력 값을 선택 후 검색하게 되면, 검색한 값들을 변수로 만들어 질의문 생성 모듈로 전달한다.

- 질의문 생성 모듈(Query Extract Module): 사용자로부터 입력받아 생성된 변수를 기반으로 질의문을 생성한다. 입력한 값에 따라 질의문의 형태는 변경된다.
- Json변환 모듈(Json Translate Module): Maria DB로 구현된 데이터베이스에서는 테이블의 형태 결과를 반환한다. 반환된 결과는 Json 형태의 데이터만 받아들이는 시각화 모듈에서는 그대로 사용할 수 없으므로, 시각화 모듈에서 사용할 수 있도록 Json형식으로 형태를 변환한다.
- 데이터 시각화 모듈 (Data Visualization Module): Javascript 기반의 시각화 모듈인 Chart.JS와 D3.JS API를 호출하며, API 형식에 맞도록 이전 Json 변환 모듈에서 제공해준 Json 데이터를 넣어주며, 변환된 시각화 결과를 웹페이지를 통해 보여준다.

사용자와 상호작용하는 부분은 HTML5를 이용해 구현하였다. 사용자가 검색을 하게 되면 Apache Web Server를 통해 데이터를 처리하고, 처리된 데이터는 사용자에게 시각화된 결과물의 형태로 보여주도록 되어 있다. Fig. 6은 앞에서 설명한 4가지 모듈에 대해 데이터가 처리되는 흐름을 보여주고 있다.

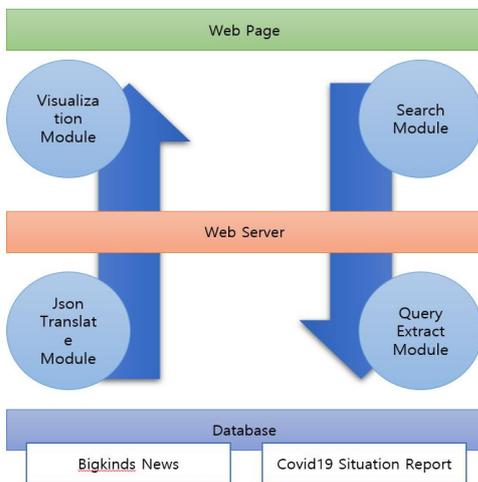


Fig. 6. Data processing flowchart

시각화 시스템은 사용자로부터 입력을 받아 원하는 데이터만을 이용해 결과를 보여줄 수 있다. 입력 도구는 지역 및 일자 등을 제공한다. 구현된 6가지 시각화에 대한 특징은 다음과 같다.

- 광역시도의 뉴스 수: 기간을 선택하게 되면 일별로 각 광역시도에 대해 뉴스의 변화량을 시간의 흐름에 따라 표시하게 된다. 키워드의 양이 많을수록 색이 진하게 표시가 된다.

- 지역의 뉴스 수: 선택한 지역 및 기간에 대해 일자별 키워드의 수를 보여준다.
- 지역과 전국의 뉴스 비율: 선택한 지역과 전국의 뉴스 비율을 Stack Bar 차트를 통해 확인할 수 있다. 차트의 하단 막대는 전국을 나타내며, 상단의 막대는 사용자가 선택한 지역을 나타낸다.
- 지역의 주요 키워드: 선택한 기간 내에 지역에서 가장 많이 사용한 키워드를 확인할 수 있다. 키워드는 선택한 기간에 가장 많이 사용한 상위 10위까지 표시가 된다.
- 키워드의 일별 변화량: "지역의 주요 키워드"에서 표시된 단어에 대해 일자별 변화량을 알 수 있다. 키워드마다 일별로 Line 차트를 이용해 표시되며, 특정 키워드가 갑자기 화제의 중심이 된 경우 중간부터 우상향 그래프로 표시가 될 수 있다.
- 일자별 코로나19 환자 변화: 선택한 지역 및 기간에 대해 코로나19 환자와 관련된 변화량을 알 수 있다. 검색 일자를 2020년 2월 17일 이후를 선택하면 그래프가 표시되며 확진자 수, 사망자수, 격리해제 인원 등에 대해 알 수 있다.

본 논문에서 구현된 시스템을 활용하여 2020년 3월 서울특별시로 검색할 시 데이터 시각화 시스템의 영역 중 "지역의 주요 키워드"에 Fig. 7에서 보는 바와 같이 "콜센터"가 관련된 단어에 포함되어 있음을 확인할 수 있다.

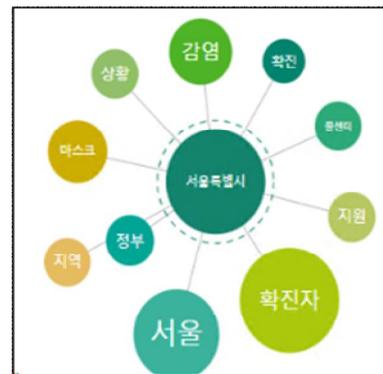


Fig. 7. Keywords in Seoul in March 2020

포털 사이트를 이용해 사용자가 선택한 기간에 대해 발행되었던 뉴스를 검색해보면 "콜센터 집단 감염사태"[14] 사건이 발생했음을 알 수 있으며, "콜센터"라는 단어가 포함된 뉴스의 비율이 급증 많음을 알 수 있다.

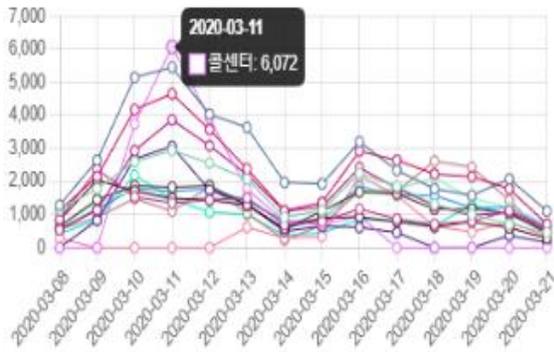


Fig. 8. Word Trends for March 2020

Fig 8은 2020년 3월 서울특별시로 검색을 했을 때 Fig 7에서 보여준 단어들의 일별 변화량을 나타낸 그래프이다. 3월 11일에 "콜센터"라는 단어가 6,072건 등장했음을 알 수 있다.

Fig. 9는 시각화 시스템에서 검색한 기간 동안의 코로나 19 확진자 수, 사망자수, 격리해제 인원수의 변화를 나타 내는 그래프이다. 그래프를 통해 "콜센터 집단 감염사태" 사건 이후 코로나19 확진자의 변화가 크게 변화된 것을 확 인할 수 있다.

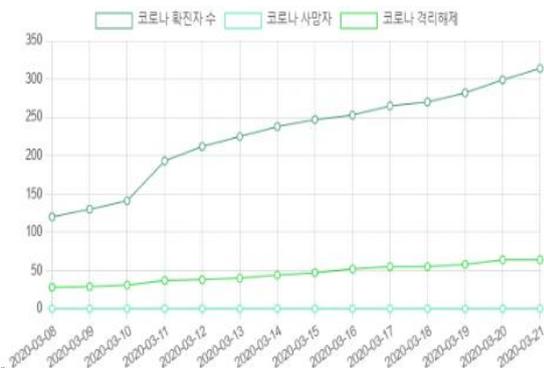


Fig. 9. Changes in the number of confirmed

Fig. 10은 검색한 기간에서 나온 지역별 키워드의 수를 지역별로 일별 변화량을 표시하였다. 색이 연할수록 해당 지역에서 사용했던 키워드의 수가 다른 지역에 대비되어 많이 없는 것이며, 이를 통해 뉴스의 양이 별로 없었음을 유추할 수 있다. 반대로 색이 진할수록 뉴스의 양이 많아 키워드 수가 많은 것이다.

이처럼 키워드의 수를 시간의 흐름에 따라 그래프로 확 인을 해보면 시간이 지날수록 색이 진해졌다가 연해짐을 알 수 있다. 또한 인근지역인 경기도에서도 뉴스가 급증했 음을 알 수 있다.

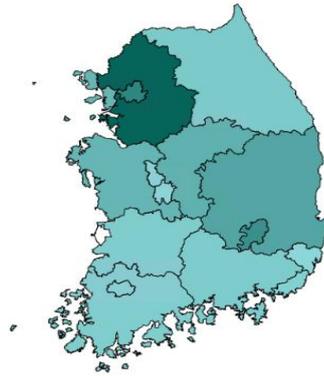


Fig. 10. Number of words nationwide in March 2020

IV. Conclusions

본 논문에서는 2019년 12월부터 2020년 7월까지 유통 되었던 코로나19와 관련된 뉴스 데이터를 이용하였으며, 정규화된 데이터를 바탕으로 주요 단어를 추출한 후 코로 나19를 분석하는 방법을 제안하였다.

제안 방법에서는 한국언론진흥재단에서 제공하는 공공 데이터를 이용하여 지역 및 기간별로 뉴스에서 사용하는 키워드와 빈도 수를 알아낸 후 이를 기반으로 시각화를 하 였다. 또한 비교분석을 위하여 코로나19 환자에 대한 현황 데이터를 같이 시각화하여, 이를 통해 특정 기간 동안 발 생한 사건에 대해 주요 키워드와 코로나19 확진자 및 감염 자 추이를 확인할 수 있다.

본 논문의 한계로는 2019년 12월부터 2020년 7월까지 의 데이터로 한정하여 분석 및 시각화를 하였기 때문에, 전 시기에 나타난 코로나19 관련 주요 키워드를 파악하 는 데에 한계가 있다. 따라서 코로나19 상황이 종료 되는 시 점까지의 뉴스를 추가하여 분석한다면 변화 추이를 알 수 있을 것이다. 또한 코로나19 관련 뉴스 뿐만 아니라 각계 각층에서 생산되는 정부 보도자료 및 SNS 등의 다양한 텍 스트를 통해서도 파악을 할 수 있다.

REFERENCES

[1] COVID-19 pandemic, "Declaration of public confidence in the COVID-19 pandemic", [https://en.wikipedia.org/wiki/ COVID-19_pandemic](https://en.wikipedia.org/wiki/COVID-19_pandemic)

[2] World Health Organization, <https://covid19.who.int/table> [accessed: May21, 2021]

[3] Mi Jeong Park, "The Press Reporting on the Infectious Diseases

Emergency Response and Protection of Personal Information. Journal of Media Law,” Ethics and Policy Research, 15(3), pp. 289-316, Dec 2016.

- [4] Gentzkow, M. & Shapiro, J. M, “What drives media slant? Evidence from US daily newspapers.” *Econometrica*, 78(1), pp. 5-71, Jan 2010.
- [5] Semetko, H. A., & Valkenburg, P. M, “Framing european politics: content analysis of press and television news,” *Journal of communication*, 50(2), pp. 93-109, 2000.
- [6] Iyengar, S., & Simon, A, “News coverage of the Gulf crisis and public opinion: Study of agenda-setting, priming and framing,” *ommunication Research*, 20(3), pp. 365-383, 1993
- [7] Babbie, R., *The practice of social research* (11th ed.). Boston, MA: Cengage Learning. 2007.
- [8] Eun Byul Lee & Jino John & Ji Sun Park, “A Study of Multicultural Space in Seoul : Analysing the Coverage of Foreign Communities with News Big Data Analytics ‘BigKinds’ for 27 Years,” *Journal of Media Economics & Culture*, vol.15, no.2, pp. 7-43, May 2017. DOI : 10.21328/JMEC.2017.5.15.2.7
- [9] Central Disease Control Headquarters, "Korea Centers for Disease Control and Prevention. Cases in Korea", http://ncov.mohw.go.kr/bdBoardList_Real.do?brdId=1&brdGubun=11&nevContSeq=&contSeq=&board_id=gubun=
- [10] D. H. Kim and S. J. Park, "Factors related to COVID-19 Incidence and Mortality rate in Gyeongsangbuk-do, Korea", *Korean Society for Agricultural Medicine and Community Health*, Vol.45, No. 4, pp. 235-244, Dec. 2020.
- [11] M. Valero and J. N. Valero-Gil, "Determinants of the number of deaths from COVID-19: differences between low-income and high-income countries in the initial stages of the pandemic", *SSRN Electronic Journal*, Jun. 2020
- [12] Korea Press Foundation, "Korea Press Foundation News Big Data Metadata COVID-19", <https://www.data.go.kr/data/15069309/fileData.do>
- [13] Joeeungen, "COVID-19 Situation Report in South Korea", https://github.com/joeeungen/coronaboard_kr
- [14] MBC News, "28 people infected with call center in Guro-gu, Seoul", https://imnews.imbc.com/replay/2020/nwtoday/article/5670350_32531.html

Authors



Tai-Sung Hur received the B.S degree in Dept. of Computer Science from Inha University in 1984, and M.S degree in Dept. of Computer engineering from Soongsil University in 1987, and Ph. D. degree in

Dept. of Computer engineering from Inha University in 1992. Dr. Hur has over 30 years of computer education. He is currently a Professor in the Dept. of Computer Science, Inha Technical College. He is interested in Data Science, Big data, Database and Internet of Things.



In-Yong Hwang received B.S. degree in 2022 from the Department of Computer Science Inha Technical College, Incheon Korea. He currently works for a company that produces web programs. His research interests include.

Data Science, Visualization and Web Program