

코로나19 데이터의 시각화

김선옥*, 김혜영^o

*한라대학교 정보통신소프트웨어학과,

^o한라대학교 정보통신소프트웨어학과

e-mail: sokim@halla.ac.kr, khy9843@gmail.com

Visualization of COVID-19 data

Sun-Ok Kim*, Hye-Young Kim^o

*Dept. of Information communication Software, Halla University,

^oDept. of Information communication Software, Halla University

● 요약 ●

본 논문은 실시간 코로나19 전국 확진자 수를 이용하여 데이터를 추출하고 분리하여 시각적으로 표현했다. 추출된 데이터를 각 지역에 따라 구분하고 전국 지도에 확진자 수를 표시하여 보기 쉽게 나타났다. 확진자의 경로는 코로나19 데이터에서 사후 관리를 위해 가장 중요한 데이터이다. 따라서 확진자의 경로는 정부에서 실시간으로 제공되고 있다. 하지만 이러한 데이터는 텍스트 형태로 되어있어 보기에 불편하다. 따라서 본 논문은 확진자 데이터를 분석하고 정리하여 지도에 표시하고자 한다. 이러한 시각화 방법은 확진자의 경로를 쉽게 파악 할 수 있고 감염 경로를 미리 예방 할 수 있게 한다.

키워드: 코로나19(COVID-19), 확진(definite diagnosis), 시각화(Visualization), 감염 경로(Infection route)

I. Introduction

1918년 스페인 독감이라는 인플루엔자 바이러스로 세계적으로 적게는 2천만명, 많게는 8천만명 정도가 독감으로 사망한 것으로 기록되어 있다. 또한 2012년에는 호흡기 감염증인 메르스가 사우디아라비아에서 발생되었고, 이후 중동지역에서 지속적으로 발생하고 있다. 메르스의 경우 중증 급성 호흡기 증후군과 유사하지만 치사율이 40%로 높은 정도의 질병이다. 2020년에는 세계적으로 covid-19 시대를 경험하고 있으며, 현재 우리나라에서는 2020년 1월 20일 한 중국인 여성이 신종 코로나바이러스 감염자로 확진되면서 대한민국에 첫 번째 감염자가 되었다. 그 이후 증상들이 빠른 속도로 나타나 전국적인 확진자수는 12월 23일 현재 52,550명 이다. 증상이 있는 사람들은 검사를 통해 양성 반응이 나오면 격리 조치가 실행되고 있다. 2월 이후 우리나라에서는 신천지 대구 교회에서 대규모 감염자가 나왔으며 이에 따라 추가 확진자와 여러 다른 지역의 감염에 대한 우려가 생겨났다. 따라서 학교들은 개학을 미루고 비대면 수업을 실시하게 되고, 학원들도 문을 닫는 등 현재에도 계속 지금 확산이 되고 있으며 1차 감염자 뿐 아니라 2,3차 감염자들도 늘어나고 있는 실태이다. 또한 코로나19 확진자의 중증 환자를 격리하거나 치료할 수 있는 병실 또한 부족한 상황이다. 하지만, 정부에서 제공하고 있는 확진자의 정보는 많은 사람들이 자신이 확진자와 접촉자로

분리 되는지에 대한 확신이 없어 검사를 하지 않는 경우가 많으며 이 사람들이 만일 무증상자일 경우 추가 확산을 막을 수 없다. 본 연구에서는 코로나19 데이터를 사용하여 확진자의 수와 경로를 파악하고 이를 시각화 할 것이다. 확진자의 수를 실시간으로 파악하여 확진자의 수를 지도에 표시하여 증상이 없는 사람들이 빠르게 검사를 하도록 유도하고, 감염 위험 지역을 쉽게 보여줌으로 인해 추가 감염을 예방할 것으로 기대된다.

II. Preliminaries

1. Related works

1.1 국내 동향

현재 정부에서 제공되는 코로나19 데이터는 코로나19에 대한 감염 정보를 제공하거나 확산 추세 예측에 사용 되고 있다. 사용자들의 휴대전화를 이용하여 위치 데이터를 취합하고 지도 서비스를 이용하여 이동 현황을 분석하고 있다. 하지만, 개인 보호를 위해 개인별 위치와 개인 식별 정보를 차단하고 있다. 하지만 이 경우 대한민국에서는

확진자의 경로를 정부에서 제공하고 있는 홈페이지에 표시되고 있다. 홈페이지의 경우 텍스트로 표현이 되고 있으며 체인점의 경우 장소만 제공되어 따로 지도에 검색해야 하는 번거로움이 있다. 또한 확진자가 마스크를 착용하지 않은채 1-2분 정도 마주쳤음에도 정부에서는 접촉자로 분류가 되지 않아 연락을 추가로 하지 않는 경우도 있다. 이러한 사람들은 접촉자로 분리 되지도 않고 감염이 되었을 경우에도 무증상자로 나타날 수도 있다. 코로나-19 예방을 위해 사업자들은 손쉬운 QR 코드와 NFC기능을 사용하여 전자 출입 명부를 활용하고 있다. 이때 QR코드는 일회용으로 발급받아 이를 인식하여 방문자 내역을 기록하여, 종이에 작성하고 있는 방문록으로 발생 할 수 있는 개인정보 침해를 해결 할 수 있는 방법이다. 하지만 애플리케이션이 설치되어있지 않은 사람이나 핸드폰 사용에 익숙하지 않은 사람들의 경우에는 종이에 작성하는 방문록으로 정보유출의 심각한 문제가 발생 할 수 있다.

III. The Proposed Scheme

1. Research environment

이 연구에서는 데이터를 활용하고, R프로그램을 사용하여 정리 분석하여 데이터를 가공하고 실시간으로 정보를 사용하고자 한다. 먼저 R프로그램을 사용하기 위해서 RStudio를 설치하였다. R프로그램에서 제공하는 패키지를 사용하여, 데이터를 가공하고 분석하고자 한다.

개발환경은 다음과 같다. RStudio은 3.6.3 버전을 사용하였으며, 대한민국 확진자의 데이터를 가져오기 위해 코로나 바이러스감염증-19의 정보를 알려주는 사이트인 코로나바이러스감염증-19 홈페이지를 이용하였다.

2. necessity

지난 1월 20일 중국인 여성이 신종코로나바이러스 감염자로 확인되면서, 대한민국에 첫 확진자가 생겼다. 그 후 2월 20일 신천지 대구 교회에서 대규모 확진자가 나오면서, 추가적인 지역 감염으로 인해 2차, 3차 확진자가 나오는 경우가 있었다. 이 시기에 대한민국은 빠른 대처가 부족했던 상황으로 정부에서 제공하고 있는 홈페이지를 통해 확진자의 경로를 확인을 할 수 밖에 없었다. 또한 유증상자를 대상으로 검사를 하게 되면서 대한민국에서는 처음 무증상자를 발견하게 되고 이에 대한 경로 또한 표시하게 되었다. 하지만 확진자의 수가 늘어나게 되면서 지역마다 위험 단계가 달라지고 있으며 그 지역에 확진자수를 표시하는 방법이 지역마다 다르다. 본 연구는 실시간 코로나 확진자의 데이터를 가져와 가공하고 이를 시각화하여 각 지역의 위험 단계를 시각적으로 보여주고 위험정도를 쉽게 예상 할 수 있으며 이를 통해 위험 지역의 방문을 자제할 수 있게 한다. 또한 감염 경로를 지도상에 표시하여 무증상자의 검사를 빠르게 하도록 유도하여 코로나-19 감염 확산을 방지하는 것을 목적으로 하고 있다.

3. research content

코로나-19의 데이터를 가공하고 분석하기 위한 라이브러리와 패키지는 다음과 같다.

설치된 패키지에 따라 설치된 라이브러리는 다음과 같다.

- library(rvest)
- library(R6)
- library(ggmap)
- library(ggplot2)
- library(raster)
- library(rgeos)
- library(maptools)

Table 1. 분석에 사용된 패키지

구분	종류	내용
웹 크롤링을 위한 패키지	rvest	R의 웹 스크래핑을 위한 패키지
	R6	기본 제공 참조 클래스와 유사한 참조 의미 체계를 사용하여 클래스를 만든다.
지도를 가져오기 위한 패키지	ggmap	지도 출력 및 경로 표현
	ggplot2	지도 출력 및 경로 표현
	raster	공간 데이터의 읽기, 쓰기, 조작, 분석 및 모델링
	devtools	R기반의 패키지 개발 쉽게 해주는 툴
	rgeos	API를 사용하는 오픈 소스
	maptools	지리 데이터를 조작하기 위한 도구 세트
rgdal	데이터 추상화 라이브러리에 대한 바인딩	

코로나-19 데이터는 정부사이트에 저장되어 있으며 본 논문은 이 데이터를 사용하여 분석할 것이며, 코로나-19 데이터를 시각화하기 위해 기초적인 데이터 추출에 함수를 사용한다. 이러한 절차에 따라 텍스트형태로 데이터는 추출된다. 아래의 내용은 코로나-19 데이터를 텍스트로 분류한 데이터들이다.

```
> count1
[1] "985" "955" "30" "53,533" "15,352" "37,425" "756" "103.25"
[9] "319" "314" "5" "16,050" "6,732" "9,171" "147" "164.89"
[17] "44" "43" "1" "1,615" "494" "1,092" "29" "47.34"
[25] "26" "26" "0" "7,607" "273" "7,134" "200" "312.21"
[33] "53" "53" "0" "2,483" "759" "1,703" "21" "84.00"
[41] "15" "14" "1" "956" "168" "784" "4" "65.63"
[49] "11" "11" "0" "754" "144" "604" "6" "51.15"
[57] "9" "9" "0" "595" "230" "343" "22" "51.87"
[65] "2" "2" "1" "133" "10" "123" "0" "38.85"
[73] "285" "277" "8" "12,744" "3,561" "8,982" "201" "96.18"
[81] "21" "21" "0" "1,033" "285" "73" "13" "67.05"
[89] "47" "46" "1" "939" "440" "484" "15" "58.71"
[97] "31" "29" "2" "1,347" "319" "1,007" "21" "63.46"
[105] "12" "12" "0" "725" "235" "483" "7" "39.89"
[113] "2" "2" "0" "527" "72" "450" "5" "28.26"
[121] "50" "50" "0" "2,169" "366" "1,742" "61" "81.46"
[129] "21" "21" "0" "1,145" "331" "811" "3" "34.06"
[137] "26" "26" "0" "330" "214" "116" "0" "49.20"
[145] "11" "0" "11" "2,381" "719" "1,661" "1" "-"
```

Fig. 1. 가공된 코로나-19 데이터

제공된 코로나-19 데이터는 html 형태로 되어있으며, 데이터를 가공하여 텍스트 형태로 수정하여 지역별로 구분하였다. 지역별로 구분된 데이터를 다시 지역별 확진자 수로 시각화 하였다. 아래는 대한민국을 시와 도로 구분하여 나타낸 함수이다.

시도 < α("합계","서울","부산","대구","인천","광주","대전","울산","세종","경기","강원","충북","충남","전북","전남","경북","경남","제주","검역")

이렇게 지역별로 나눈 데이터들을 데이터 프레임이라는 R데이터의 형태로 가공하여 시각화 하고자 한다.

다음은 코로나-19 데이터를 8X19로 분류하여 매트릭스 형태로 나타냈으며, 아래와 같다.

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19
1	985	319	44	26	53	15	11	9	2	285	21	47	31	12	2	50	21	26	11
2	955	314	43	26	53	14	11	9	1	277	21	46	29	12	2	50	21	26	0
3	30	5	1	0	0	1	0	0	1	8	0	1	2	0	0	0	0	0	11
4	53,533	16,050	1,615	7,607	2,483	956	754	595	133	12,744	1,033	939	1,347	725	527	2,169	1,145	330	2,381
5	15,352	6,732	494	273	759	168	144	230	10	3,561	285	440	319	235	72	366	331	214	719
6	37,425	9,171	1,092	7,134	1,703	784	604	343	123	8,982	735	484	1,007	483	450	1,742	811	116	1,661
7	756	147	29	200	21	4	6	22	0	201	13	15	21	7	5	61	3	0	1
8	103,25	164,89	47,34	312,21	84,00	65,63	51,15	51,87	38,85	96,18	67,05	58,71	63,46	39,89	28,26	81,46	34,06	49,20	-

Fig. 2. 코로나-19 데이터

가공된 코로나-19 데이터를 지역별로 구분하였고, 코로나-19 데이터를 발생 유형에 따라 6개(국내발생,"해외유입","확진환자","격리중","격리해제","사망자")로 구분하고 R프로그램을 사용하여 발생률, 발생합계, 시도를 추가하여 본 논문의 자료로 사용하였다.

또한 이들 자료를 가공하여 시도별로 구분된 데이터는 아래와 같다.

시도	합계	국내 발생	해외 유입	확진 환자	격리 중	격리 해제	사망 자	발생 률
1 합계	985	955	30	53,533	15,352	37,425	756	103.25
2 서울	319	314	5	16,050	6,732	9,171	147	164.89
3 부산	44	43	1	1,615	494	1,092	29	47.34
4 대구	26	26	0	7,607	273	7,134	200	312.21
5 인천	53	53	0	2,483	759	1,703	21	84.00
6 광주	15	14	1	956	168	784	4	65.63
7 대전	11	11	0	754	144	604	6	51.15
8 울산	9	9	0	595	230	343	22	51.87
9 세종	2	1	1	133	10	123	0	38.85
10 경기	285	277	8	12,744	3,561	8,982	201	96.18
11 강원	21	21	0	1,033	285	735	13	67.05
12 충북	47	46	1	939	440	484	15	58.71
13 충남	31	29	2	1,347	319	1,007	21	63.46
14 전북	12	12	0	725	235	483	7	39.89
15 전남	2	2	0	527	72	450	5	28.26
16 경북	31	30	0	2,169	366	1,742	61	81.46
17 경남	21	21	0	1,145	331	811	3	34.06
18 제주	26	26	0	330	214	116	0	49.20
19 검역	11	0	11	2,381	719	1,661	1	-

Fig. 3. 시도별 코로나-19 데이터

시도별 코로나-19 데이터를 막대그래프로 나타내면 다음과 같다.

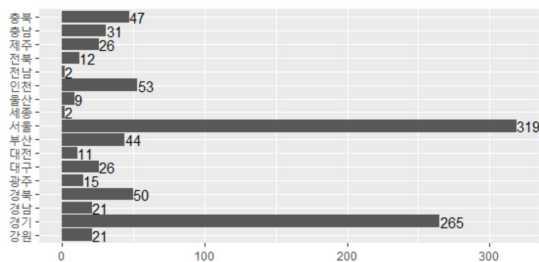


Fig. 4. 시도별 코로나-19

서울지역이 319명으로 현재(2020년12월24일) 제일 많은 확진자를 기록하고 있으며, 세종과 전남이 2명으로 가장 적은 확진자가 있음을 보여주고 있다. 가장 위험한 지역으로는 서울과 경기지역이라 할 수 있고, 인천과 경북지역도 많은 코로나-19의 확진자가 있음을 알 수 있다. 이처럼 데이터를 막대그래프로 표현하였으며, 지역별 위험도를 한눈에 확인하기 위해 시도별로 구분하여 지도에 나타내면 보다 시각적으로 확인할 수 있다.

다음은 코로나-19 데이터를 대한민국의 지도에 시각화하여 나타낸 것이다.

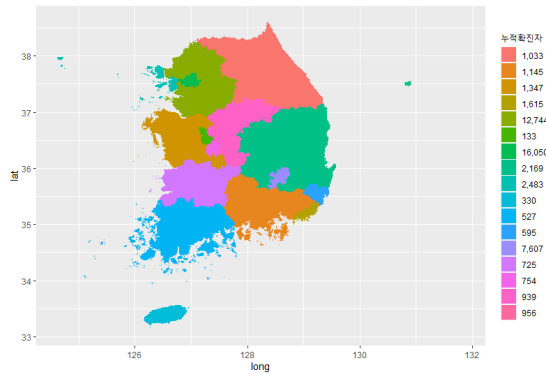


Fig. 5. 실시간(12월 24일) 확진자 현황 시각화

위의 그림처럼 데이터를 시각화 하면 색을 통해 보다 쉽게 코로나-19의 위험 지역을 파악 할 수 있다.

IV. Conclusions

본 논문은 지역별로 구분한 후 확진자의 수를 시각적으로 확인할 수 있게 하였다. 지역별로 서로 다른 색으로 구분하여 어느 지역이 위험 지역인지에 대한 확인을 쉽게 할 수 있으며, 이를 통해 위험 단계를 예측 할 수 있으며 그 위험지역 방문을 자제하게 하여 확산을 줄일 수 있다. 감염자의 경로를 사용하여 이를 지도에 표시하면 감염경로에 대한 데이터를 시각화하여 효율적인 관리와 증상이 없는 사람들의 잠재적 확진자의 경로를 빠르게 파악하여 예방할 수 있을 것이다.

REFERENCES

[1] 'August _ Monthly _ Corona 19_ Response _ Data _ Use _ Status _ and _ Cases', Korea Internet & Security Agency, 2020.8
 [2] 'Easy-to-learn R data analysis', written by Kim Young-woo, 2017.7
 [3] <http://ncov.mohw.go.kr/>
 [4] 'Social science data analysis using R' by Baek

Young-min,2015.4

- [5] Coronavirus Disease 2019 (COVID-19): Pandemic and the Challenge of Public Health,2020.10
- [6] The Effects of COVID-19 Risk Information Seeking and Processing on its Preventive Behaviors and Information Sharing by Minjung Park ,2020.10