

개인 맞춤형 감염병 정보 제공 시스템 개발

김민수*

*성균관대학교 경영대학원 MBA

*전남대학교 대학원 정보보호 박사과정 수료

miracle@mommaeat.com

Development of Personalized Infectious Diseases Information Provision System

Min-soo Kim*

*MBA of Business, Sungkyunkwan University

*Dr. of Information Protection, Chonnam National University

요 약

코로나19의 전 세계적인 확산은 사회·경제적으로 큰 변화를 가져왔다. 또한, 새로운 기회를 가져오기도 했다. ICT 기술 기반의 비대면 산업이 탄력을 기술개발이 극적으로 이루어지고 있으며, 기존의 기술들 또한 다양한 응용이 이루어지고 있다. 이에, 본 논문에서는 감염병에 대한, 개인맞춤형 정보를 제공해주기 위한 서비스를 개발하고 활용하는데 대해서 다루어 보고자 한다.

1. 서론

최근 코로나19로 인한 팬데믹은 확진자가 3천만 명이 넘었으며, 사망자 또한 100만명을 넘기는 지경에 이르렀으며, 관련 연구자들은 코로나로 인한 사망자가 올해안에 200만이 넘을 것으로 사라지지 않을 것이라고까지 예측하고 있다.[1] 감염병은 인류 역사상 지속적으로 존재해왔으며, 그 전파속도는 인류의 통신수단의 발전에 힘입어 통제가 어려운 수준까지 발전되어왔다. 또한, IT의 발전은 사용자들이 다양한 정보를 제공받는 것이 가능하게 만들었다. 이를 활용하여 감염병을 직접 진단하고 치료하는데 활용하는 방법[2, 3], 관련 정보를 분석하는 방법[4, 5] 등 다양한 감염병 관련 연구가 국내외에서 진행되고 있다. 그러나 이 정보들이 신뢰할 수 있고 개인에게 필요한 정보인지에는 고려해보아야 할 것이다. 본 논문에서는 개인에게 맞춤형 감염병 정보를 제공하기 위해 데이터를 수집하고 맞춤형 정보를 제공하는 개발하였다.

2. 데이터

본 연구에서 개인맞춤형 데이터 제공을 위해, 다양한 기관들로부터 다양한 종류의 데이터들을 크롤링하여 빅데이터를 매일 생성하고 이를 기준으로 데이터를 분석하는 프로세스를 개발하였다. 기본 데이터는 감염병 관련 공공데이터와 통계, 감염병과 관련된 SNS, 신문기사 등을 수집하여 활용하였다.



(그림 1) 감염병 관련 수집 데이터 소스

3. 시스템 구성

구현 시스템은 데이터를 수집하고 맞춤형 콘텐츠를 생산하는데 초점을 두었다. 감염병에 관련된 개인 맞춤형 콘텐츠는 각종 데이터를 수집하고 처리하여 생성되며 시스템 개념도는 그림 2와 같다.



(그림 2) 맞춤형 감염병 콘텐츠 생성

3.1 맞춤형 통계 기반 데이터

맞춤형 콘텐츠 데이터는 맞춤형 통계 데이터와 맞춤형 감염병 자가관리 정보로 구분된다. 시스템을 사용하는 사용자의 주소지 정보에 따라 지역별 확진자 정보데이터와 감염 대응 정보를 제공하고 지역별 감염병 위험도를 단계화하여 표현해 줌으로써 감염병 대응을 효율적으로 할 수 있도록 개발하였다. 감염병 위험도를 단계화하는데 표현한 파라미터는 다음과 같다.

<표 1> 맞춤형 통계 데이터 표현을 위한 주요 변수

| 주요 변수 | 활용 변수 |
|-------------|----------------|
| 확진자 수 | 감염률을 위한 기본 변수 |
| 감염률 | 감염 위험도 기본 변수 |
| 사망률 | |
| 지역별 인구통계 | 인구 대비 확진자 비율 |
| 지역별 확진자 수 | 타지역 대비 확진자 증감률 |
| 지역별 격리해제자 수 | - |
| 연령별 감염률 | 연령대별 감염 위험도 |
| 연령별 사망률 | 연령대별 사망 위험도 |

3.2 감염병 자가관리 정보

감염병 자가관리 정보는 수집된 확진자의 동선과 뉴스 기사 등에 따라 사용자의 거주지역 및 이동 지역을 기반으로 확진자 정보를 분석하고 개개인의 기저질환, 연령, 현재 상태, 감염병 통계 정보를 바탕으로 건강관리를 위한 정보를 제공할 수 있도록 구성하였다.

(1) 동선 수집 및 정보화는 각 지자체의 동선 제공 데이터를 크롤링하고 이를 GPS로 변환하여 만들어 냈다. 주소정보는 개체이름인식(NER)을 통해 추출하고 이를 휴리스틱한 방법을 통해, 행정안전부에서 제공하는 주소 데이터셋과 검색하여 GPS 정보를 추출하고 활용하였다.

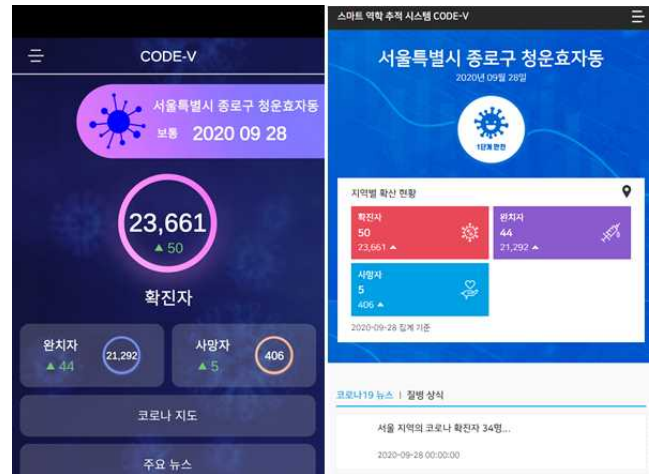
(2) SNS와 뉴스 기사는 크롤링 한 후 지역이나 위치에 관한 정보와 중요 키워드를 추출하였다. 지역과 위치에 관한 정보는 개체이름인식(NER) 추출 방식을 활용하고, 중요 키워드는 토픽모델링(Topic modelling) 알고리즘[5]을 이용하여 기사나 SNS에서 주요 토픽들을 발견하고 데이터를 분류하여 사용자에게 맞춤형 전달할 수 있도록 개발하였다.

(3) 개인 맞춤형 건강관리 정보는 개인의 기저질환, 연령, 지역, 현재 상태를 입력 받아 감염병 질환에 대해 개개인별로 얼마나 위험도가 높은지 정량화하여 표현하였다. 개개인별 감염병에 대한 위험도는

개인이 거주하는 지역 인구모델, 지역 감염률(지역별감염자 수/지역 인구), 지역발생 신규 확진자 수, 국내 발생 치명률(사망자 수/전체 확진자수), 개인의 연령(연령에 따른 치명률 통계), 성별(성별에 따른 위험도 예측치), 기저질환(사망자중 기저질환 보유율에 따른 가중치)에 따른 치명 예측치를 반영한 선형식을 구성하고 이에 정규화하여 개인별 맞춤형 위험도를 계산하고 위험도에 따른 맞춤형 콘텐츠를 제공할 수 있도록 서비스를 구성하였다.

4. 결론

본 논문에서는 개인 이용자들이 맞춤형 감염병 정보를 제공할 수 있도록 데이터를 생산하는 시스템을 개발하였다.



(그림 3) 개발된 앱과 모바일 웹

인류의 역사상 감염병은 꾸준히 발발하고 있으며 때론 치명적인 전염률과 사망률의 엄청난 수의 사람들이 사망하는 대제앙을 가져오기도 했다. 현재의 감염병은 과학기술과 교통의 발전으로 더 빠르게 전 세계에 전파되고 팬데믹(Pendamic)에 이르는 상황에 이르기까지 하고 있으나, 여전히 이에 대한 선제적 대응은 불가능하고 감염자를 격리하여 전파를 막고, 빠른 연구를 통해 극복할 방법을 찾아내는데 머무르고 있다. IT기술의 발전은 데이터의 빠른 공유와 전파가 가능하게 하고 있으며 이를 활용하면 감염병에 대해 개개인이 효율적으로 대응할 수 있는 정보를 제공하는 것이 가능할 것으로 여겨지며 이에 활용할 수 있는 시스템을 만들어 보고자 본 연구를 진행하였다.

5. 향후 과제

감염병 데이터를 개인이 입력한 정보와 공개된 정보를 바탕으로 활용하여 정보를 제공하는 시스템을 개발하였으나, 언어모델 등을 활용하여 검색 및 질의 응답이 가능한 시스템으로 확장할 예정이다.

4. 사사

“이 논문은 2020년도 정부의 재원으로 휴먼케어기술센터의 지원을 받아 수행된 연구사업임.

(과제명:스마트 역학추적시스템을 통한 감염성질환 정보 콘텐츠 개발)”

참고문헌

- [1] World Health Organization 코로나 사이트, https://www.who.int/emergencies/diseases/novel-coronavirus-2019?gclid=CjwKCAjw8MD7BRArEiwAGZsrBcGLjqQ6NH-S2lsGwLQ3o4JUrwfQMq9S_6LOEUSO9KHqvmKjsZU0XxoC9mMQAvD_BwE, 2020-09-26일 검색
- [2] 장미, 이복주, 강봉구, 서경민, “코로나 바이러스 확진자 데이터 기반 시뮬레이션 모델 학습 방법 제안,” 한국정보처리학회:학술대회논문집, pp. 334 - 337, 2020.
- [3] Singh, D., Kumar, V., Vaishali et al. Classification of COVID-19 patients from chest CT images using multi-objective differential evolution - based convolutional neural networks. Eur J Clin Microbiol Infect Dis 39, 1379 - 1389, 2020
- [4] Timo Möller, Anthony Reina, Raghavan Jayakumar, Malte Pietsch, COVID-QA: A Question Answering Dataset for COVID-19, 2020
- [5] Blei, David, Probabilistic Topic Models. Communications of the ACM. 55 (4): 77 - 84. 2012.