[1)]R-CDM 기반의 빅데이터 시각화 및 관리 시스템 구축

김승진*, 정창원**, 김태훈**, 이충섭**, 노시형**, 김지언**, 이고은**, 윤권하***
*원광대학교 의과학과
**원광대학교 의료융합연구센터
***원광대학교 의과대학 영상의학과
e-mail:{koch369369, mediblue, tae_hoonkim, cslee99, nosij123, kakashy112, 290eun, khy1646}@wku.ac.kr

Construction of Big Data Visualization and Management System Based on R-CDM

Seung-Jin Kim*, Chang-Won Jeong**, Tae-Hoon Kim**, Chung-Sub Lee**, Si-Hyeong No**, Ji-Eon Kim**, Go-eun Lee**, Kwon-Ha Yoon***
*Department of Medical Science, Wonkwang University
**Medical Convergence Research Center, Wonkwang University
***Department of Radiology, Wonkwang University School of Medicine and Hospital

요 약

본 논문은 R-CDM 의료영상정보를 기반으로 ELK Stack 기술을 적용하여 획득한 데이터의 분석 결과를 시각화하기 위한 시스템에 대해 기술한다. 제안한 시스템은 의료 빅데이터의 검색, 수집 그리고 분석 결과를 모니터링 할 수 있으며, 특히 대량의 데이터의 변화와 데이터간의 차이를 확인할 수 있다. 본 연구에서 제안한 시스템은 수집된 의료영상 빅데이터에 대해 적용하여 현황과 처리결과 그리고 실시간 분석결과에 대한 모니터링을 통해 관리의 효율성을 높여 실시간 검색 및 분석 서비스 분야에 기여할 것으로 기대된다.

1. 서론

최근 4차 산업혁명의 핵심기술 중 하나인 빅데이터 처리 기술에 관련된 산업은 거듭 성장하고 있다. 의료계 또한 의료영상을 기반으로 빅데이터 생성뿐 아니라 검색, 수집, 분석, 활용에 대한 중요성이 커지면서 데이터 시각화의 필요성이 요구된다[1,2].

본 논문은 영상의 공통데이터모델(Radiology-Common Data Model, R-CDM)을 기반으로 의료영상 데이터를 표준화하여 수집, 검색, 분석 그리고 시각화를 지원하는 관리 시스템을 구축하고자 한다. ELK(Elasticsearch Logstash Kibana, ELK) Stack[3] 기술을 적용함으로써 데이터를 검색 및 수집하고 분석결과는 다양한 표현방식으로 시각화하여 실시간 모니터링 결과를 보인다.

2. 영상 공통데이터 모델 (R-CDM)

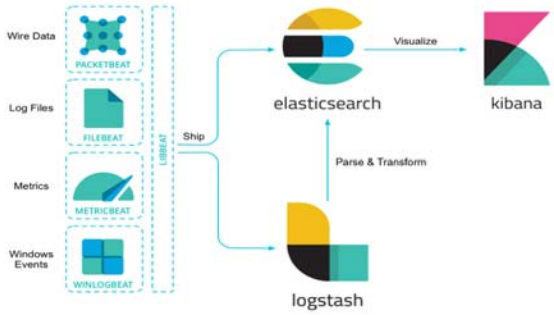
본 연구의 기반기술인 R-CDM은 영상에 관련된 정보를 머신러닝을 위한 데이터셋 구축 방안[4,5]을 기초로 의료용 디지털 영상 및 통신 표준(Digital Imaging and Communications in Medicine, DICOM)의 Header 정보와 기존의 OHDSI(Observational Health Data Sciences and Informatics)에서 제안한 OMOP 공통데이터모델의 확장모델을 제시하였다[6].

R-CDM은 Occurrence와 Image 테이블로 구성하였다. Occurrence 테이블은 영상 촬영기법, 시술방법, 촬영장비, 해부학적 위치, 조영제 사용유무, DICOM 영상정보 등의 내용이 포함되어 있으며, Image 테이블은 위상, 영상 개수, 해상도, DICOM 영상정보 등의 내용이 포함되어 있다.

R-CDM을 구성하는 정보를 효율적으로 관리할 수 있도록 분석한 결과를 시각화하는데 중점을 두었다.

3. 데이터 검색·수집·시각화

ELK Stack 기술은 데이터 분석 및 검색 기능을 담당하는 Elasticsearch, 수집 기능을 담당하는 Logstash, 다양한 표현방식으로 데이터를 시각화하며 실시간 모니터링 기능을 담당하는 Kibana로 구성된다.



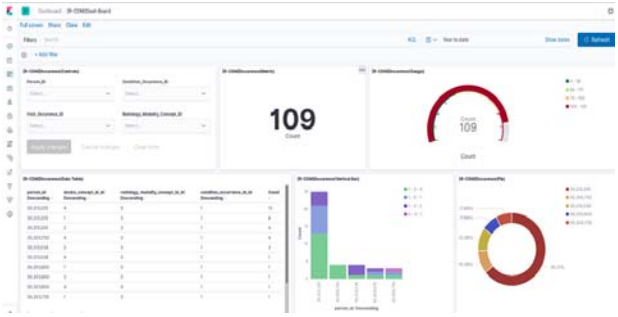
(그림 1) ELK Stack 구조

1) 이 논문은 2018년도 한국보건산업진흥원(HI18C1216)과 산업통산자원부(No.20001234)의 지원을 받아 수행된 연구임.

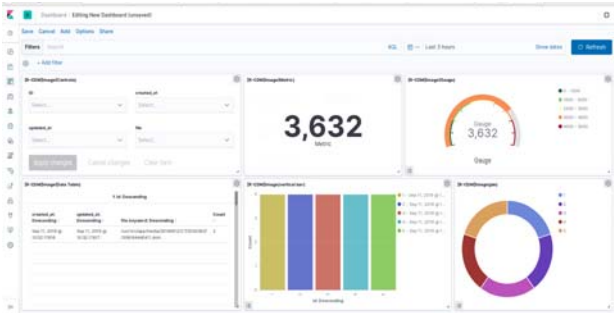
그림 1은 ELK Stack의 구조를 보인다. ELK Stack 기술을 활용하여 R-CDM 기반의 표준화 데이터를 수집하고 연구자의 요구사항에 따라 데이터를 검색·분석하는 기능을 갖는다. 분석결과는 Dashboard 공간에서 확인 가능하며 실시간 모니터링 기능으로 관리할 수 있다.

4. 데이터 시각화 및 실시간 모니터링 결과

그림 2는 R-CDM의 Occurrence 테이블, 그림 3은 R-CDM의 Image 테이블 데이터를 다양한 방식으로 시각화하여 Dashboard에 나타냈다. 텍스트 및 테이블 형식으로 표현한 빅데이터 정보는 검색·분석과정에 있어 한눈에 파악하기 어려울뿐더러 실시간으로 변화되는 데이터 확인이 불가능하다. 이러한 문제점을 본 논문에서는 ELK Stack 기술을 활용하여 연구자의 요구사항에 따른 부분적, 복합적, 전체적으로 검색·분석 결과에 대한 데이터를 시각화 데이터로 표현함으로써 해결하였다.

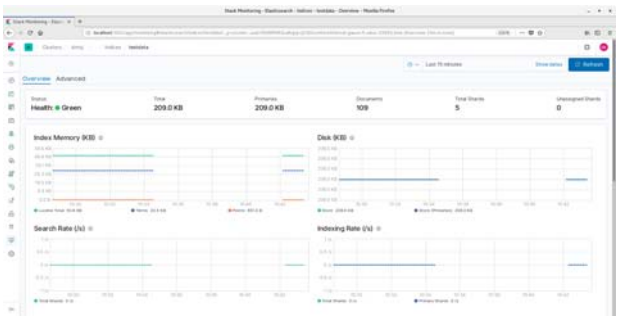


(그림 2) R-CDM Occurrence 테이블 데이터 시각화 표현



(그림 3) R-CDM Image 테이블 데이터 시각화 표현

그림 4는 R-CDM의 Occurrence 테이블, 그림 5는 Image 테이블에 대한 실시간 모니터링 예시를 보인다. 실시간 모니터링은 각 테이블에 대한 ELK Stack의 성능을 예의주시할 수 있으며 최적화된 상태로 유지하는데 필요한 정보를 제공한다.



(그림 4) R-CDM Occurrence 테이블 실시간 모니터링

(그림 5) R-CDM Image 테이블 실시간 모니터링

관리적 측면에 있어 Index Memory, Disk, Search Rate, Indexing Rate 등의 정보를 실시간 모니터링 할 수 있어 시스템 문제가 발생하게 되면 정확한 시점 및 문제점을 파악할 수 있어 빠른 대처가 가능하다.

5. 결론 및 향후연구

본 논문에서 제안한 시스템은 의료영상정보를 기반으로 ELK Stack 기술을 활용하여 데이터를 수집하고 수집한 데이터는 연구자의 검색 요구사항에 따라 다양한 방식으로 분석결과를 시각화하여 표현하였다. 또한 실시간 모니터링이 가능하여 관리적 측면에 있어서도 용이한 시스템으로 실시간 검색 및 분석 서비스 분야에 기여할 것으로 기대된다. 향후연구로는 연구자의 편의성에 중점을 둔 검색 서비스 개발과 머신러닝 기술을 적용하여 검색에 대한 데이터 분석결과를 예측할 수 있는 시스템을 구축하고자 한다.

참고문헌

[1] Sabyasachi Dash, Sushil Kumar Shakyawar, Big data in healthcare: management, analysis and future prospects. Journal of Big Data. Dec 2019.

[2] Belle, A., Thiagarajan, R., Soroushmehr, S.M., Navidi, F., Beard, D.A., Najarian, K.: Big data analytics in healthcare. BioMed. Res. Int. 10, 1 - 16 (2015)

[3] ELK Stack. https://www.elastic.co/kr/elk-stack

[4] Kohli, M. D., Summers, R. M., & Geis, J. R. (2017). Medical image data and datasets in the era of machine learning-whitepaper from the 2016 C-MIMI meeting dataset session. Journal of digital imaging, 30(4), 392-399.

[5] Chatap, Niranjan J., and Ashish Kr Shrivastava. “A survey on various classification techniques for medical image data.” International Journal of Computer Applications 97.15 (2014).

[6] Kim, S.J., et, al., “Construction of Monitoring System and Bigdata Visualization Based on the Radiology-CDM”. The Korean Society of Medical Informatics. Spring. (2019)