

# 의료 빅데이터를 위한 표준화 시스템 구축 및 성능평가<sup>1</sup>

김승진\*, 정창원\*\*, 노시형\*\*, 김지언\*\*, 김태훈\*\*, 전홍영\*\*, 이윤오\*\*\*, 윤권하\*\*\*\*

\*원광대학교 의과학과

\*\*원광대학교 의료융합연구센터

\*\*\*원광대학교병원 전산정보팀

\*\*\*\*원광대학교 의과대학 영상의학과

e-mail : {koch369369, nosij123, kakasky112, tae\_hoonkim, zip80}@wku.ac.kr, wonduke@empas.com, khy1646@wku.ac.kr

## Construction and Performance Evaluation of Standard System for Medical Big Data

Seung-Jin Kim\*, Chang-Won Jeong\*\*, Si-Hyeong No\*\*, Ji-Eon Kim\*\*, Tae-Hoon Kim\*\*,

Hong Yong Jun\*\*, Yun Oh Lee\*\*\*, Kwon-Ha Yoon\*\*\*\*

\*Dept. of Medical Science, Wonkwang University

\*\*Medical Convergence Research Center, Wonkwang University

\*\*\*Computing & Information Team, Wonkwang University hospital

\*\*\*\*Dept. of Radiology, Wonkwang University School of Medicine and Hospital

### 요 약

본 논문에서는 원광대학교병원 의료정보시스템의 임상데이터를 OHDSI 가 제안하는 공통데이터 모델로 변환하여 표준화 시스템 구축에 대해서 기술한다. 또한, 검색속도 향상을 위해 인덱싱 기법을 적용한 성능평가 결과를 보인다. 구축된 표준화 시스템은 다양한 임상연구에 활용될 것을 기대하고 있다.

### 1. 서론

최근 4 차 산업혁명을 바탕으로 한 의료산업의 변화에 따라 의료정보시스템의 표준화에 대한 연구가 활발하게 진행되고 있다. 특히, 빅데이터, 클라우드 서비스 기술 등을 활용한 의료 데이터 축적과 인공지능(AI) 기술을 접목한 데이터 분석 기술의 발전으로 의료산업에 대한 기대가 높아지고 있다. 기존의 의료 서비스는 진단과 치료에 관련된 의료서비스에서 개인 유전자 데이터, 임상정보, 실생활 환경 및 습관에 대한 정보 기반의 빅데이터를 활용 및 분석함으로써 환자 맞춤형 예방, 진단, 치료 그리고 예후에 이르기까지 토털의료서비스를 제공하기 위한 서비스로 변화되고 있다. 그러나 기관간의 데이터 이질성 극복하기 위한 표준화가 시급한 실정이다[1].

따라서 본 연구에서는 원광대학교 병원의료정보시스템(EMR, OCS, PACS 등)의 임상데이터를 OHDSI (Observational Health Data Sciences and Informatics)[2]에서 제안하는 공통데이터모델(CDM)로 변환하여 표준화 시스템을 구축한 내용을 기술한다[3][4]. 또한 대량의 데이터 검색속도를 향상시키기 위해 인덱싱 기법[5]을 적용하여 성능 평가한 결과를 보이고자 한다.

구축된 표준화 시스템은 국내외 다기관 국제연구에 기반이 될 것이며, 활발한 임상연구의 초석이 될 것으로 기대하고 있다.

### 2. 의료정보 표준화 시스템

현재 원광대학교병원 환자(1,780,000 여명)의 임상데이터를 확보하여 국제표준화(OHDSI)된 데이터로 변환(CDM) 하기 위해 그림 1 과 같이 데이터베이스 설계와 구축을 통해 저장하였다.

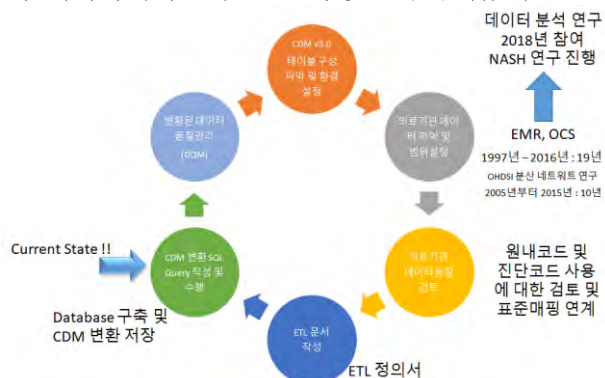


(그림 1) 의료 빅데이터 기반 표준화 시스템

표준화 시스템을 구축하기 위한 전체 과정은 그림 2 와 같은 절차를 통해 진행하였다. 먼저 OHDSI 의

<sup>1</sup> 이 논문은 2016 년도 정부(과학기술정보 통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. NRF- 2016M3A9A7918501)

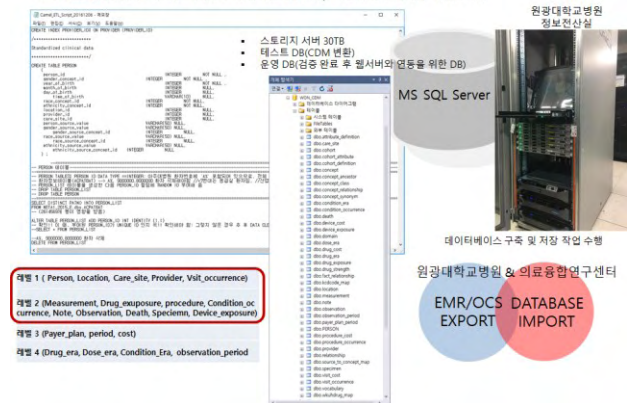
CDM 테이블 구성 파악 및 환경 설정을 통해 원광대학교병원 의료정보시스템의 데이터 파악 및 범위 설정을 진행하고 원내코드와 진단코드 사용에 대한 표준 매핑작업을 수행하였다. 그리고 ETL 정의서를 작성하고 CDM 으로 변환하기 위한 SQL Query 수행을 통해 데이터베이스에 변환·저장을 수행하였다.



(그림 2) 표준화 작업과정

현재 임상연구 진행에 필요한 레벨 2 까지 변환 작업을 수행하였으며 향후 추가적으로 레벨 3 에서 4 에 대한 데이터 변환 작업을 진행할 계획이다.

CDM 기반 데이터베이스 구축을 위해 ETL 정의서를 기반 Script 작성과 함께 각각의 테이블 생성



(그림 3) 구축된 의료 빅데이터 데이터베이스 결과

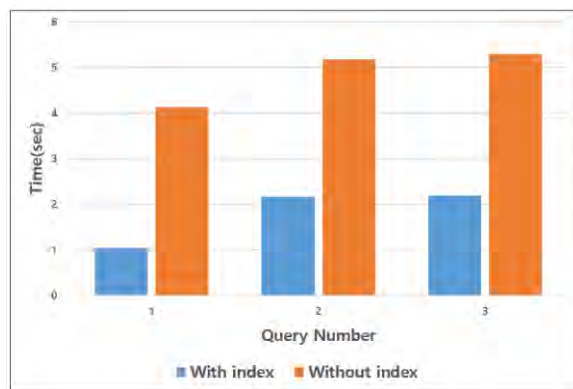
### 3. 검색 성능향상을 위한 인덱싱 기법 적용 성능평가

인덱싱 기법을 적용한 검색 성능평가를 실시한 결과는 표 1 과 그림 4 와 같다.

<표 1> 데이터 인덱싱(Data Index) 유무에 따른 데이터 추출 시간의 차이

Trial No.	With index(A) (sec)	Without index(B) (sec)	Ratios (B÷A)
1	1.04249	4.12153	4.0
2	2.15918	5.18182	2.4
3	2.19523	5.29545	2.4
Mean±SD	1.80±0.66	4.87±0.65	2.9

All data are represented as mean±standard deviation



(그림 4) 데이터 인덱싱 유무에 따른 환자데이터 추출 시간에 대한 성능평가

총 환자수 1,780,000 명중 지방 간(Fatty liver)에 해당하는 환자를 추출하는 쿼리(Query)문을 작성하여 50 번씩 3 세트로 수행한 결과의 평균값으로 인덱싱 기법을 포함한 데이터베이스의 성능을 평가하였다. 추출결과 107,786 명이 추출되었으며, 인덱싱이 포함된 데이터베이스는 포함되지 않은 데이터베이스보다 평균 2.92 배 빠른 속도를 확인하였다.

### 4. 결론 및 향후 연구

본 논문에서는 의료 빅데이터의 표준화 시스템 구축과 검색성능 향상을 위한 인덱싱 기법을 적용하여 성능을 평가한 결과를 보였다. 이에 따른 구축된 표준화 시스템은 근거기반의 다양한 임상연구에 활용될 것이다.

향후 연구내용으로는 구축된 의료 빅데이터 표준화 시스템의 임상데이터에 대한 데이터의 품질 관리를 위한 단계를 거쳐 다기관 임상연구를 진행하고자 한다. 또한 의료영상 및 생활습관 그리고 유전체에 관련된 데이터에 대해 기존 공통 데이터 모델에서 확장된 연구에 참여하여 의료 빅데이터 표준화 시스템구축과 임상응용 연구를 진행할 계획이다.

### 참고문헌

[1] A Study of Big Data Information Systems Building and Cases  
Choong-Kwon Lee, KISM Smart Media Journal Vol.4, No.3 pp. 56~62 ISSN : 2287-1322 [2015]

[2] OHDSI, <https://ohdsi.org/>

[3] Study on HIPAA PHI application method to protect personal medical information in OMOP CDM construction  
Hak-Ki Kim, Eun-Young Jung, Dong-Kyun Park, Korean Institute of Next Generation Computing Vol.13, No.6 [2017]

[4] Data Mapping and Design for PHR Establishment Using CDM in colorectal cancer patients.  
Sung-Jung Hong, Kyungpook University, Korea, [2016]

[5] An Indexing Approach in Large Video Databases  
Byung-Moon Moon, Hanyang University, Korea, [2015]