

계층적 LOD 개발 환경 구축 및 의료 정보 적용

문희경*

*원광대학교 컴퓨터공학과
e-mail:ybnjcw@wku.ac.kr

Construction of Hierarchical LOD Development Environment and Its Application of Medical Information

Hee-Kyung Moon*

*Dept of Computer Engineering, Wonkwang University

요 약

최근 ICT 기술과 의료 빅데이터를 활용한 다양한 연구가 활발하게 진행되고 있다. 이질적인 의료데이터의 공유와 확산을 위해 표준화 데이터 모델로 온톨로지 기반의 **Linked Open Data**가 대안으로 급부상하고 있다. 특히 의료 빅데이터의 분석을 위한 데이터 셋은 프로토콜화하기 어려운 문제점을 갖고 있다. 본 논문에서는 이러한 문제점을 해결하기 개발된 계층적 LOD 개발 환경 시스템을 기반으로 의료정보를 적용하기 위한 모델링에 중점을 두고자 한다. 본 연구는 의료 빅데이터의 검색과 분석연구 분야에 큰 영향을 줄 것으로 기대하고 있다.

1. 서론

최근 의료정보의 표준화 연구가 활발하게 진행되고 있다[1, 2]. 그러나 병원간 서로 다른 데이터구조로 인하여 표준화하기 어려운 문제점을 안고 개별적인 공통데이터 모델을 채택하여 의료 빅데이터를 구축하고 있는 실정이다. 의료정보의 활용 목적에 따라 데이터 구조화와 각 기관에서 사용하는 데이터 용어 및 코드체계가 표준화되어 있지 않아 다기관이 참여하여 연구하기에는 미흡한 부분이 있다. 본 논문은 이러한 문제점을 해결하기 위해 LOD(Linked Open Data)[3]를 기반으로 국제 표준인 OHDSI(Observational Health Data Sciences and Informatics)의 CDM(Common Data Model) 모델[4]과 매핑하여 고품질의 데이터셋 개발과 함께 의료정보를 적용하기 위한 새로운 방법에 대해서 제안하고자 한다. 이를 위해 계층분리 기반의 LOD 개발 시스템에 대해서 설명하고 의료정보의 데이터셋 설계에 중점을 두어 기술한다.

2. 관련 연구

병원관리 목적으로 의료전산화가 진행된지 25년 동안 수많은 의료정보가 수집되면서 빅데이터 분석 연구의 발판이 마련되었다. 헬스케어 ICT가 보급되면서 병원 기관의 의료정보는 디지털 플랫폼 상에서

급속하게 전자화되어 수집되었다. 최근 수집된 의료 빅데이터를 기반으로 임상결정지원 등 분석의 응용 영역이 확대되고 있는 추세이다[5]. 이와 함께 의료기관마다 개별 관리되고 있는 개인 의료정보를 의료기관간 정보 교류가 가능한 시스템으로 전환되고 있으며, 국내의 경우 2016년 10월 보건복지부에서 진료정보교류 표준을 고시하여 제정안을 마련하였다. 진료정보교류 표준 제정과 관련하여 의료정보 표준화와 의료 빅데이터 분석을 통한 새로운 신규 서비스 모델에 관한 연구가 활발하게 진행되고 있다.

따라서 본 논문에서는 각 기관별 의료정보의 상이한 문제와 이를 해결하기 위한 솔루션인 LOD기반 의료정보의 고품질을 위한 데이터셋 설계 방법을 제안한다.

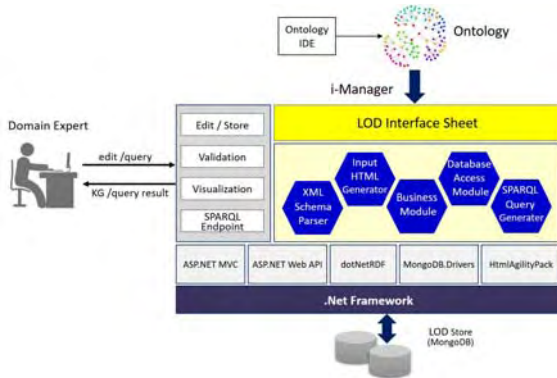
3. 계층적 LOD 개발환경 구축

본장에서는 LOD기반 의료정보를 효과적으로 개발하기 위해서 스키마-인스턴스 계층분리를 통합할 수 있는 인터페이스를 제공하는 시스템에 대해서 기술한다. 그리고 의료정보를 기반으로 한 계층분리 스키마에 대해 설명한다.

3.1 계층분리기반 LOD 개발 시스템 환경

의료정보의 LOD 데이터셋의 구조와 의미를 표현하

는 RDF에도 관점의 분리가 정의되어 있지만 기존의 LOD 개발은 이 관점의 분리를 활용하지 않았기 때문에 LOD 개발 과정의 시행착오와 데이터셋 검증에 문제를 가지고 있다. 계층분리기반의 시스템의 전체 소프트웨어 구조는 다음 그림 1과 같다.

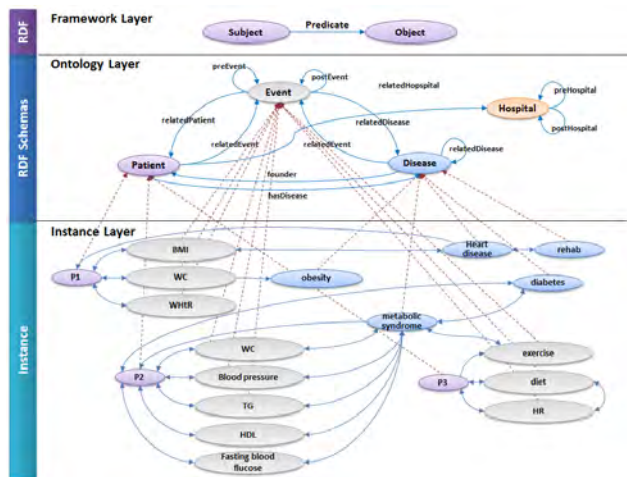


(그림 1) 계층분리 기반의 LOD 개발 시스템 환경

그림 1과 같이 온톨로지 계층과 인스턴스 계층을 분리하여 인스턴스 계층에 적합한 시스템을 구성하였다. 온톨로지 계층은 기존의 온톨로지 개발 도구를 활용하고 i-manager는 인스턴스 추출 생성에 적합하도록 전문화하였다.

3.2 LOD기반 계층 분리 Scheme

의료정보의 지식은 상호간 관계가 복잡하고 환자를 기반으로 관찰기간, 검체, 사망, 방문 발생정보, 수술/처치발생 정보, 약물 노출 정보, 장치/재료 노출 정보, 진단, 측정정보 그리고 비정형 데이터를 다루는 노트 그리고 관찰정보에 해당하는 임상데이터와 표준 의료기관 데이터 그리고 의료비용과 추출 요소와 표준 용어에 관련된 다양한 온톨로지 요소를 내포하고 있다.



(그림 2) 의료정보 LOD 개발의 계층 구조

그림 2는 의료정보의 LOD 표현의 일부를 보이고 있다. 사건(Event), 질환(Disease), 환자(Patient), 병원(Hospital)은 의료정보 도메인 온톨로지의 대표적인 클래스로 OHDSI의 CDM을 기반으로 한다.

환자인스턴스는 'P1', 'P2', 'P3'로 표현하였으며, 병원인스턴스는 이전과 이후로 표현하였고 질환(Disease)은 비만(obesity) 과 대사 증후군(metabolic syndrome) 그리고 심장 질환(heart disease), 당뇨(dianetes), 재활(rehab)로 나타났다. 그리고 질환 검출을 위한 주요인자인 체질량지수(BMI), 허리둘레(WC), 허리/키비(WHtR), 중성지방(TG), 혈압(Blood preasure), 운동량(exercise), 식단(diet), 심박수(HR), 고밀도리포단백질(High density lipoprotein) 등으로 표현하였다. 이를 기반으로 RDF와 RDF Schema와 같이 웹상에서 의료정보 도메인 온톨로지를 구현하는 기반으로 제공하고, LOD 인스턴스는 이 기반 위에서 운영되는 실제 임상데이터와 매핑된다.

4. 결론

본 논문에서는 계층적 LOD 개발환경을 기반으로 OHDSI의 CDM과 연계한 의료정보의 지식화를 위한 방법을 제시하였다. 이를 위해 계층적 LOD 개발 시스템 환경에 대해 설명하고 이를 기반으로 의료정보의 LOD 데이터셋 구조를 보였다.

향후 연구로는 설계사항을 토대로 LOD기반의 효과적인 의료정보 데이터셋을 개발하고 표준화된 시스템 개발하고자 한다. 이와 함께 개발된 LOD 데이터셋에 대한 성능평가 연구와 이를 기초로 의료 빅데이터 분석연구를 진행하고자 한다.

참고문헌

- [1] Y. Zhang, M. Qiu, CW. Tsai, M. M. Hassan, A. Alamri, "Health-CPS: Healthcare cyber-physical system assisted by cloud and big data." IEEE Systems Journal Vol.11, No.1, pp. 88-95, 2017.
- [2] RAC-ALBU, Elena Madalina, et al. Interoperability of Medical Data Through e-Health Service in Romania. In: International Conference on Exploring Services Science. Springer, Cham, 2016. pp. 683-692.
- [3] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann and S. Auer, "Quality Assessment for Linked Data: A Survey," Semantic Web, Vol.7, No.1, pp. 63-93, 2016.
- [4] OHDSI, <https://www.ohdsi.org/>
- [5] F. M. Suchanek and G. Weikum, "Knowledge Bases in the Age of Big Data Analytics," Proceedings of the Very Large Data Base Endowment, Vol.7, No.13, pp. 1713-1714, Aug. 2014.