

링크드 데이터를 활용한 척추정보 검색 서비스 Spine Information Search Service using Linked Data

서동민, 이승우, 이상호, 정한민, 성원경
한국과학기술정보연구원 소프트웨어연구센터

Dong-Min Seo, Seungwoo Lee, Sang-Ho Lee,
Hanmin Jung, Won-Kyung Sung
Dept. of S/W Research Center, Korea Institute
of Science and Technology Information

요약

차세대 웹 기술인 시맨틱 웹(Semantic Web)이 각광을 받음에 따라, 시맨틱 웹과 관련된 다양한 기술들이 활발히 연구되고 있다. 특히, 시맨틱 웹의 핵심 사례인 링크드 데이터(Linked Data)는 웹으로 접근 가능한 형태의 URI로 시맨틱 데이터를 표현하고 서로 연결함으로써, 오픈 데이터로서 상호운용성 극대화 및 데이터의 통합을 용이하게 한다. 이는 바이오인포매틱스 기술 발전과 결합하여 웹에서 바로 이용 가능한 많은 의학 링크드 데이터 생산과 이를 활용한 다양한 의학용어 검색 서비스 개발 결과를 가져왔다. 하지만, 기존 의학용어 검색 서비스들은 각 기관이 보유하고 있는 의학정보를 기반으로 서비스되기 때문에 다양하고 정확한 정보를 얻기 위해서는 많은 시간과 노력을 투자해야 한다. 그래서 본 논문에서는 척추 분야 의학 관련 링크드 데이터를 활용하여 정확도가 높고 확장이 용이한 척추정보 검색 서비스를 제안한다.

I. 서론

컴퓨터가 웹에 있는 데이터의 뜻을 이해하고 논리적 추론까지 할 수 있는 차세대 웹 기술인 시맨틱 웹이 각광을 받음에 따라, 시맨틱 웹과 관련된 다양한 기술들이 활발히 연구되고 있다. 특히, 링크드 데이터[1]는 차세대 웹의 가장 핵심이 되는 데이터 네트워크를 위한 기술로, 웹에서 자유롭게 데이터를 개방하여 연계할 수 있도록 하고, 이들 데이터가 다시 협업할 수 있게 하여 진정한 데이터 웹을 실현하고자 하는 것이다.

최근, 다양한 분야에서 시맨틱 웹의 실현을 위한 데이터 연계와 협업이 진행되고 있다. 일례로, 미국이나 영국 정부는 공공의 고부가가치 데이터 셋을 국민 모두가 활용할 수 있도록 개방한 data.gov와 data.gov.uk를 운영하고 있으며, 시스템적으로 보다 유용하게 활용할 수 있도록 모든 데이터 셋을 RDF로 표현했다. 또한, 미국의 인간게놈프로젝트를 계기로 최근 10년 동안, 컴퓨터를 활용해 생물학적 데이터를 수집, 관리, 저장, 평가, 분석하는 기술인 바이오인포매틱스가 크게 부상되고 생명과학계에서 데이터를 분석하고 신약 개발 프로세스를 가속화하는데 바이오인포매틱스가 활용되면서, 의학 관련 링크드 데이터가 지속적으로 증가하고 있다.

노령화 사회가 급속히 진행되면서 일반인들도 많은 의학정보를 접한다. 그래서 각 기관에서는 어려운 의학정보에 대한 이해를 돕기 위해서 다양한 의학용어 검색 서비스를 제공한다. 하지만, 기존 의학용어 검색 서비스들은 각 기관이 보유하고 있는 의학정보를 기반으로 서비스되기 때문에 다양하고 풍부한 정보를 얻기 위해서는 많은 시간과 노력을 투자해야 한다. 그래서 본 논문에서는 척추 분야 의학 관련 링크드 데이터를 활용해 사용자

친화적이면서 정확도가 높은 척추정보 검색 서비스를 제안한다. 본 논문에서 제안한 척추정보 검색 서비스는 별도의 의학정보 구축 없이 링크드 데이터를 연계하기 때문에 개발 비용이 적고, 향후 새로운 의학 관련 링크드 데이터 생성 시 확장이 용이하다.

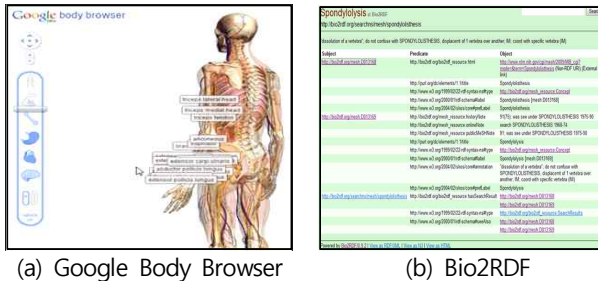
II. 관련 연구

사용자로부터 주어진 의학용어에 대한 다양하고 풍부한 정보를 제공하기 위해, 국내·외에서 KMLEI[2]와 같이 외부 의학정보 DB 및 웹을 활용해 많은 정보를 수집·제공하는 의학용어 검색 서비스가 개발되었다. 하지만, 기존 제안된 서비스들은 몇몇 의학정보 DB 및 웹에서 단순 키워드 검색을 통해 정보를 수집하기 때문에 정확도가 낮고 화면 구성이 일반 사용자들이 사용하기에는 불편해 활용도가 낮다.

그림 1(a)는 Google이 개발한 Body Browser[3] 서비스를 보여준다. 최근 이 서비스와 같이 장기, 뼈, 근육 등 인체 내부의 각종 부위를 세세하게 관찰할 수 있는 뷰어 제공 서비스들이 개발되고 있다. 하지만, 기존 개발된 서비스들은 뷰어 기능만 제공할 뿐 인체 각 부위에 대한 상세 정보를 제공하지 않고 소수의 정상 인체에 대한 정보를 제공하기 때문에 사용자에게 깊이 있는 이해를 제공하기에는 적합하지 않다.

Bio2RDF[4]는 2008년 의학 관련 링크드 데이터를 제공하는 Gene Ontology, OMIM, PubMed, Uniprot 등 총 40개의 의학정보 DB를 쉽게 연계할 수 있는 시스템을 제안했다. 이 시스템은 연계된 DB의 의학 관련 300억 링크드 데이터를 SPARQL을 통해 쉽게 활용할 수 있는 기능을 제공한다. 그림 1(b)는 Bio2RDF 검색 서비스를

통해 “Spondyloysis”에 대한 의학용어를 검색한 결과를 보여준다. 하지만, 검색 결과가 RDF 형태의 텍스트로만 제공되기 때문에 링크드 데이터에 대한 전문지식이 없는 일반인들이 활용하기는 어렵다.



(a) Google Body Browser

(b) Bio2RDF

▶▶ 그림 1. 기존 의학용어 검색 서비스

III. 척추정보 검색 서비스

1. 질환별 척추정보 데이터베이스 구축

척추정보 검색 서비스는 노령 척추 질환에 대한 적절한 진단과 치료, 재활에 도움을 주기위해 2009년부터 2012년까지 국가문제해결형 연구사업(NAP)을 통해 구축한 척추정보를 기반 데이터로 활용한다. 구축된 척추정보는 정상척추를 갖는 표본 27 set과 경추 및 요추 퇴행성 척추질환을 갖는 표본 311 set에 대한 표본정보, 단면 영상, 3차원 형상모델, 참조치수, 물성, 가상척추 모델 등으로 구성된다[5].

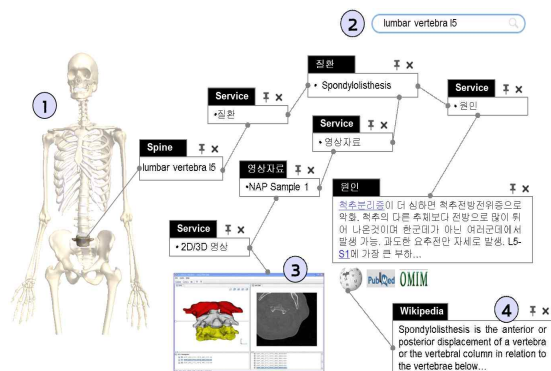
2. 척추 관련 링크드 데이터 연계

Bio2RDF 시스템에 연계된 의학 관련 DB를 통해 의학 용어에 대한 정의 및 설명 정보를 제공한다. 척추정보 검색 서비스는 빠른 시스템 응답 시간과 꼭 필요한 정보로 구성된 검색 결과를 제공하기 위해서, Bio2RDF 시스템에 연계된 40개의 모든 DB로부터 정보를 추출하지 않고 사용빈도가 높은 MeSH, OMIM, UniProt, EntrezGene DB만을 활용한다. 또한, 제시한 서비스는 SPARQL 인터페이스를 통해 링크드 데이터로부터 검색 결과를 구성하기 때문에 향후 필요한 정보를 갖고 있는 링크드 데이터 DB와의 추가적인 연계가 용이하다.

3. 척추정보 검색 뷰어

그림 2는 척추정보 검색 서비스를 보여준다. 척추정보 검색 서비스는 웹에서 운영될 수 있도록 WebGL을 기반으로 하며, Google Body Browser와 같이 ①의 3D 인체 또는 ②의 키워드 검색창을 통해 척추와 관련된 세부 또는 그룹화된 하나 이상의 뼈를 선택할 수 있다. 척추정보 검색 서비스는 Valuen-1→Servicen-1→Valuen 형태로 정보를 제공하는 방식으로, 사용자가 선택한 정보(Valuen-1)와 연관된 정보의 타입 정보(Servicen-1)를 제공하고 사용자가 선택한 Valuen-1의 Servicen-1에 대한 정보(Valuen)를 내비게이션 하는 방식으로 운영된다. 예를 들어, 그림 2에서와 같이 서비스는 (Spine) lumbar

vertebra l5에 대한 연관 정보로 동의어, 설명, 질환, 영상 자료를 제공하고 질환을 선택하면 관련 질환명 목록을 제공한다. 마찬가지로, 사용자가 선택한 (질환) Spondylolisthesis에 대한 연관 정보로 동의어, 원인, 증상/징후, 진단법, 치료법, 합병증, 영상자료가 제공되며 영상정보를 선택하면 ③과 같이 질환별 척추정보 데이터베이스에 구축된 척추정보를 가시화한다. 또한, 서비스는 ④와 같이 검색한 의학용어에 대한 풍부한 정보 제공을 위해 MeSH, OMIM, UniProt, EntrezGene DB 외에도 Wikipedia 검색 결과를 연계해서 제공한다.



▶▶ 그림 2. 척추정보 검색 뷰어

IV. 향후 연구

본 연구에서 제시한 서비스는 현재 프로토타입으로 현재 개발 중에 있다. 향후 연구로서 의학 논문, 특허, 기사 관련 링크드 데이터와의 연계를 통해 특정 의학용어에 대한 동향 및 분석 서비스를 제공하는 서비스로 확장하는 것을 고려하고 있다.

후 기

본 연구는 2013년 기초기술연구회 협동연구사업 국가 사회적 문제해결형 연구과제(NAP) 프로그램의 연구비 지원으로 수행되었음(NAP-09-2/P-13-JC-LU01-C01).

■ 참고 문헌 ■

- [1] C. Bizer, T. Heath, T. Berners-Lee, “Linked Data - The Story So Far”, IJSWIS. Vol.5 No.3, pp.1-22, 2009.
- [2] 가톨릭중앙의료원, “KMLE 의학 검색 엔진”, <http://www.kmle.co.kr>. 2007.
- [3] Google, “Google Body Browser”, <http://bodybrowser.googlelabs.com>, 2008.
- [4] F. Belleau, M.A. Nolin, N. Tourigny, P. Rigault, J. Morissette, “Bio2RDF: Toward a mashup to build bioinformatics knowledge system”, Biomed Inform. Vol.41 No.5, 2008. pp.706-716.
- [5] D. Seo, S. Lee, S. Lee, S. Lee, H. Jung, W. Sung, “The Development of Database Management System for Korean Spinal Disease Patients”, Proc. of SMA. Vol.1 No.1, pp.1-3, 2012.