

국내 Linked Open Data 발행 후 활용에 관한 연구

- 국가서지와 RISS Linked Data를 중심으로 -

A study on the analysis about practicality after publishing Linked Open Data: focused on National bibliography and RISS

이현정, 중앙대학교 대학원 문헌정보학과, caulis98@gmail.com

남영준, 중앙대학교 문헌정보학과, namyj@cau.ac.kr

Hyun Jung Yi, Dept. of Library and Information Science, Graduate School of Chung-Ang University

Young Joon Nam, Prof. of Library and Information Science in Chung-Ang University

Linked Data는 서로 분리돼 있는 데이터를 HTTP와 SPARQL을 이용하여 한 번에 질의하는 것이 가능하다. 많은 공공정보들이 여러가지 형식으로 공개되고 있지만 데이터의 재사용성을 높이고 공개된 다른 데이터들과의 활용도를 높이기 위해 Linked Open Data (LOD)로 발행되고 있다. 국내 LOD 중 국가서지 Linked Data와 RISS Linked Data를 중심으로 발행현황을 살펴보고 LOD 발행 후 외부 자원과의 연결을 시도한 활용서비스에 대하여 점검하고 활용방안에 대해 고찰하였다. LOD 발행에 있어 SKOS, FOAF, DC 등과 같은 국제 표준용어집을 사용하여 연계 용이성을 확보하고 LOD 연계에 대한 실질적 활용과 운용경험을 공유할 수 있는 플랫폼의 중요성을 제안하였다. 또한 국내의 LOD 플랫폼을 통해 홍보 및 네트워크를 구성하는 노력과 기술적인 운용에 대한 가이드의 필요성을 제안하였다.

1. 서론

도서관이 지향하는 서비스는 이용자에게 필요한 최적의 정보자원을 제공하는 것이다. 최근에 도서관이 인터넷 정보자원과 디지털 정보자원 등과 같이 새로운 형태의 정보자원을 적극적으로 수집하는 것도 이러한 선(善)을 달성하기 위함이다.

한편 공공정보 개방과 공유 확대를 위해 「공공데이터의 제공 및 이용 활성화에 관한 법률」이 2013년 7월에 공포되어 2013년 10월 시행되고 있다. 이에 따라 다양한 형태와 방법으로 여러 기관 및 분야의 공공정보가 법률 규정에 따라 공개되고 있다. 공공정보라고 명명되는 공공데이터는 그 자체가 갖고 있는 정보적 특성 때문에 연구자들을 비롯한 도서관 이용자들은 이를 중요한 정보자원으로써 인지하고 활용하고 있다. 특히 정부운영체에

서 발생하는 공공데이터의 특성 때문에 공공데이터가 갖고 있는 내용적 가치와 달리 공개된 많은 공공데이터들은 이를 재사용할 수 있는 형태적 가치는 상대적으로 낮은 편이다.

공공데이터 규모와 빈도의 대용량성으로 인해 공공데이터의 활용을 위해 이를 수집하고 분석하기 위해서는 최소한의 기계적 처리가 수반되어야 한다. 다른 영역에서 발행되는 공공데이터를 이용자가 하나의 시각으로 활용하기 위해서는 반드시 이종(異種)간 데이터의 체계적 수집과 활용적 연계가 필연적으로 수반되어야 한다. 링크드 데이터(Linked Data) 개념의 출현은 이와 같은 활용요구를 수용하기 위해 나타난 새로운 데이터 융합기술 가운데 하나이다. 즉 이용자가 공개된 공공데이터를 이용자 관점으로 자유롭게 정제하고 융합하며, 데이터 간 의미적 연결을 통해 활용할 수 있는 구조화된 데이터를 링크드 데이터의

발행이라 설명하고 있다. 우리나라의 경우도 국립중앙도서관의 국가서지, 한국교육학술정보원(KERIS)의 RISS 데이터베이스, 국사편찬위원회와 한국학중앙연구원의 한국사 데이터베이스, 국립수목원과 국립중앙과학관의 생물정보 등과 같은 데이터를 연계와 융합이 용이한 형태의 링크드 데이터 형태로 발행하는 사업을 수행하고 있다. 링크드 오픈 데이터(Linked Open Data: LOD)는 공공영역에서 국민들을 위한 개방정책을 링크드 데이터형태로 공개한 데이터를 의미한다. 우리나라를 비롯하여 세계 각국에서 정부운영계에서 생산된 공공데이터를 적극적으로 개방하는 전략은 단순한 정보 공개의 차원이 아니라 공개된 정보를 민간에서 적극적으로 활용하여 산업화, 자원화 전략이다. 즉 링크드 데이터는 국가차원의 정보활용 정책이라 할 수 있다. 우리나라는 2014년 4월에는 한국정보원진흥원(NIA)에서 ‘링크드 데이터 구축 공정 가이드’를 발간하는 등의 선도적 역할을 수행하며, 공공데이터를 기반으로 하는 새로운 정보 및 서비스 창출이 가능하도록 데이터 활용 기반을 조성하고 있다.

한편 링크드 데이터 형태로 발행한 데이터는 기본적으로 기계처리를 통해 연구자가 필요로 하는 형태로 해당 공공데이터가 연계될 수 있어야 한다. 그럼에도 불구하고 많은 기관에서 발행하는 링크드 데이터의 형태는 발행기관의 특성과 링크드 데이터발행에 대한 인식의 차이 때문에 링크드 데이터 발행수준과 발행형태, 활용을 위한 전제조건 등에서 많은 관점의 차이를 갖고 있다.

이러한 차이는 링크드 데이터 발행이 추구하는 이종간 데이터의 자유로운 융합을 위한 어려움으로 설명될 수 있다. 따라서 이 연구에서는 우리나라 국립중앙도서관(NL)에서 발행하는 국가서지를 비롯하여 KERIS의 RISS를 중심으로 우리나라 링크드 데이터 발행형태와 그것을 이용한 활용서비스를 분석하여 의도한 링크드 데이터 발행에 대한 활용방안을 제안하고자 한다.

2. 국내 Linked Data 발행사례

2장에서는 국내에서 Linked Data를 발행하여 서비스하고 있는 여러 기관의 사례를 개괄적으로 살펴보기로 한다. 소재목은 Linked Data의 고유명칭으로 한다.

2.1 국가서지 Linked Data

NL은 국가대표도서관으로서 자료를 수집하여 영구 보존하는 국가 서지와 관련된 저자명, 주제어를 RDF로 변환하고 Linked Data (LD) 형식으로 발행하여 누구나 자유롭게 활용할 수 있도록 제공하고자 하는 목적을 갖고 있다.

주제명(560,561건), 저자명(238,586건)의 통제어휘집 및 서지데이터(7,686,509책)를 LD로 발행하였다. 링크드 데이터 발행 원칙을 준수하여 표준용어집(FOAF, SKOS, Bibo, Dublin Core 등)을 작성하였으며, 이를 표준모형에 맞게 변환하였다. 국가서지 링크드 데이터 운영을 위한 플랫폼(<http://lod.nl.go.kr>)을 운영하고 있다.

2.2 RISS Linked Data

KERIS는 당 기관이 소장하고 있는 대학도서관 종합목록(RISS)의 일부를 포함한 학술정보를 Linked Data화하여 시범서비스를 구축하였다. Linked Data화 시범사업의 대상은 단행본과 해외학술지 MARC정보와 저자명, 주제명을 추출하였다. 소장정보에서는 대학(기관)의 코드정보와 도서관관련 코드 정보를 사용하였다. 저자정보에서는 저자 전거데이터를 사용하였다. 단행본 서명의 고유키를 URI로 설정하였고, data property로 서지정보를 주제명, 소장정보, 저자전거데이터를 object property로 관계를 생성한 Linked Data를 발행하였다.

2.3 Korean History Linked Open Data

한국사 LOD는 국사편찬위원회 인물 온톨로지와 시소러스, 한국학중앙연구원 민족문화대백과사전, 문화재청 유물/유적데이터와 각 유관 기관의 역사 데이터들을 통합한 데이터 혹은 링크 형태로 융합하여 LOD 기반의 한국사 연계 통합 DB로 구축되었다. 역사분야에 적합한 온톨로지 모델을 자체적으로 개발하였으며 활용성을 위하여 인물의 표현을 위한 FOAF 어휘 및 주제명 표현을 위한 SKOS 어휘 등을 사용했다.

2.4 생물정보 Linked Open Data

생물정보 LOD는 미래창조과학부와 한국정보화진흥원의 지원을 받아 2013년 국가DB 사업의 일환으로 구축되었다. 국립수목원과 국립중앙과학관으로부터 제공된 생물정보 데이터를 기초로 현재 균류, 식물류, 척추동물류 등 3가지 생물종 분류군에 대한 정보를 RDF형태로 제공하고 있다. 향후 분류군의 종류와 대상 생물종을 확대할 예정이며 시범서비스를 운영중이다.

2.5 서울 열린 데이터 광장 LOD

서울시는 열린 데이터 광장을 통해 다양한 공공 데이터를 공개하고 있으나, 데이터의 활용과 가공하는 과정은 복잡하고 개별적으로 정제하고 융합한 데이터는 공유하기가 어려운 문제가 있다. 서울시는 공공 데이터 활용과 공유 사례를 마련하고자 문화시설, 문화재, 지하철을 중심으로 Linked Open Data 시범사업을 추진하여 베타 서비스가 제공되고 있다. 원본 데이터는 데이터의 이름을 영문으로 변환하여 위키피디아 등의 외부 데이터와의 연결을 시도하고, 데이터 형식의 의미적 관계성(박물관, 아트홀, 기념관 등)을 정의하는 과정을 거치며, 관련 지번주소 또는 도로명 주소 등을 변환하고 위경도 좌표값을 통해 정제되고 상세화하여 양질의 데이터로 변환된다.

3. 국내 Linked Data 발행 후 활용사례

링크드 데이터는 기본적으로 이기종(異機種) 정보들의 연결을 최적화하기 위한 구조화된 데이터이다. 이 장에서는 국가서지와 RISS를 중심으로 LOD 발행 후 외부 자원과의 연결을 시도한 활용서비스에 대하여 점검하고 활용방안을 파악한다.

3.1 국가서지 Linked Data

국립중앙도서관은 인터링킹을 위한 외부 자원 연계 방법으로 다음의 세 가지 방법을 사용하였다. 각각의 방법은 국립중앙도서관 LOD 사이트의 인터링킹에서 예제를 확인할 수 있다.

- 서지의 연결

국립중앙도서관 국가 서지 Linked Data를 영국국립도서관, 캠브리지 대학도서관, 스웨덴 국립도서관, 노르웨이의 PODE 프로젝트와 데이터를 상호 연결하기 위하여 Endpoint와 SearchAPI 방법을 활용하여 인터링킹을 생성하였다. 국립중앙도서관에서 isbn, issn을 포함하는 서지데이터를 검색하여 textA로 추출하고 이후 각각의 인터링킹 대상 Endpoint를 통해 isbn, issn을 포함하고 있는 triple을 검색하여 textB를 추출한다. 생성된 textA, textB를 매칭하여 동일한지 확인 후 변환(owl:sameAs)하여 해외도서관 서지 인터링킹이 가능하도록 하였다.

- 주제명의 연결

일본국립도서관 홈페이지에서 제공하고 있는 RDF/XML 형식의 데이터를 다운로드 받아 NL 주제명과 매칭을 시도하였다. 일본국립도서관 주제명 중에서 영문으로 생성된 것은 LC 주제명과 sameAs 관계를 가지고 있으므로 LC 주제명을 비교하여 동일한 LC 주제명과 연결되는 NL 주제명과 일본국립도서관 주제명을 추출하여 주제명간에

skos:close Match 관계를 생성한다.

- 저자명의 연결

국제적인 전거 통제를 목적으로 OCLC에서 운영되는 VIAF(Virtual International Authority File)와 NL의 저자명의 연결을 시도하였다. 저자 관련 트리플에서 foaf:name, nlon:birth Year, nlon:deathYear를 비교하고 VIAF의 URI를 획득하여 owl:sameAs로 연결하였다.

3.2 RISS Linked Data

RISS LOD 운영플랫폼(<http://data.riss.kr>)은 구축된 데이터를 모두 제공하는 발행서비스와 그 중 일부 데이터를 통해 활용서비스를 구현한 것으로 나누어 운영하고 있다. 발행한 LOD를 DBpedia, Flickr, LCSH, OCLC WorldCat, BNB(영국국립도서관) 외부 데이터와 연결하여 내부 콘텐츠로 서비스하도록 구현한 사례를 잘 보여주고 있다. 현재는 일반사항을 설명하는 DBpedia 연결 부분은 실제 구현되지 않고 있으나 지리적 좌표(GeoNames) 연결을 통한 국내 도서관 소장정보와 해외 유사서지(WorldCat, LCSH, BNB)와의 연결, 이미지(Flickr)를 한 곳에서 볼 수 있다.

국가서지와 RISS 모두 해외 도서관 서지들과의 인터링킹을 시도하였으나 국립중앙도서관은 방대한 서지 레코드를 대상으로 동일 자원을 매칭하는 서비스를 구현하는 것이 가능하고, RISS는 레코드량이 많지 않은 Trial 서비스이므로 동일 자원이 아니더라도 주제어가 같은 유사서지로 연결하는 서비스를 보여주고 있어 차이가 있다.

3.3 현황 및 문제점

국립중앙도서관 국가서지 LOD는 외부연계에 관한 요구사항 정의 단계에서 여러 해외 LOD를 연계하려고 시도하였으나 정작 같은 주제의 국

내 LOD 외부자원인 RISS의 LOD 연계하지 않고 있다. 서비스 대상이 국내 이용자인 것을 감안할 때 해외 LOD 연계도 중요하지만 전국 대학종합목록인 RISS와 연계가 되어있지 않다는 것은 결국 상호참조와 소통의 부재를 가져오게 되고, 각각의 데이터베이스를 연계하여 서비스를 개발하고자 하는 이용자에게 부담을 가중시킨다.

제공중인 LOD를 이용하거나 및 서비스를 개발하는 사례로 볼 때 국가서지 LOD서비스는 발행 후 활용에 관해서는 어떤 레퍼런스를 제공하지 못하고 있으며 이용자의 창조적인 활용에 기대하고 있다. RISS는 Linked Data 활용 서비스를 제시하여 이용자 활용을 위한 가이드라인을 제시하고 있다.

4. Linked Open Data 활용방안

4.1 데이터간의 교합

Heath 등(2011)은 RDF 링크의 3가지 유형으로 Relationship Links, Identity Links, Vocabulary Links를 제시하였다. 관계를 기술하기 위한 Relationship Links는 데이터셋 내부에 존재하는 데이터를 다른 데이터셋에 기술된 개체와 연결하는 것이다. Identity Links는 다른 데이터 정보원에 존재하는 개체에 대한 추가적인 정보를 활용하기 위한 링크를 설정하는 것이다. owl:sameAs와 같은 속성을 사용하며 URI를 연결하는 것이 소셜 서비스상의 데이터 웹 환경에서 기본이 되고 있다. Vocabulary Links는 서로 다른 용어가 동일한 의미를 가지거나 서로 다른 용어 사이에 어떤 형태의 매핑이 필요한 경우에 활용된다.

Wiktionary에서 인터링킹(interlinking)의 사전적 의미는 ‘둘 또는 그 이상을 연결’이라고 매우 간단하지만 LOD에서 추구하는 인터링킹의 의미는 데이터베이스 내부에서의 다른 데

이터들끼리 연결하는 것뿐 아니라 외부 자원간 데이터들끼리 연결이 가능한 것이라 할 수 있다. 또한 인터링킹 자체가 LOD의 목적이자 방법이라고 할 수 있겠다. 인터링킹을 적극적으로 활용하여 서비스를 개발해야 할 것이다.

4.2 표준용어집의 사용

국가서지와 RISS의 데이터베이스는 국내 기관 및 대학교 데이터베이스와 연계되어 있어서 LOD 발행 시 표준 준수 및 연계가능성 검증이 필수다. 또한 현재 연계 기관 및 대학 이외 개인이나 LOD를 활용하고자 하는 기업 등에서 서비스 개발 시 다양한 매쉬업 서비스와 연계 가능한 공통기반 구성이 필요하다.

각자 nlon이나 keris라는 네임스페이스에 태그를 정의하여 사용하고 있으나 공통으로 사용할 수 있는 엘리먼트까지 재정의하여 사용하는 것은 공통모듈을 이용한 활용 시 문제가 될 소지가 있다. RISS의 경우에는 dc:creator 엘리먼트가 있음에도 keris: Author이라는 별도의 저자 엘리먼트를 사용하고 있다.

4.3 해외 데이터와의 연계 활용

영국의 BBC는 TV, 라디오 프로그램의 활용성을 높이기 위해 모든 콘텐츠 리소스 별로 URI를 부여하였으며, 기존의 서비스되고 있던 일반 웹페이지 별 리소스 단위로 재구성하였다. 기존의 콘텐츠를 활용하여 특별한 구축 노력 없이 쉽게 사이트를 구축할 수 있었다. 미국 의회 도서관에서는 도서관 정보를 LOD 형태로 서비스 하고 있다. 특히 표준어휘를 역참조 가능한 URI 통해 접근할 수 있도록 구성하여, 주제어, 세부분류, 장르/형식 등 관련된 데이터 제공하였다. (<http://id.loc.gov/authorities/{identifier}>) 사용자가 자신의 메타데이터를 링크할 수 있도록 하여 참여, 공유를 가능하도록 하였다.

4.4 플랫폼을 통한 활용 및 홍보

국가간 협력 네트워크는 좋은 경험과 사례를 배울 수 있는 방법일 것이다. Ding 등(2012)의 논문에서 35개국 이상이 참여하는 Open Government Partnership을 통한 협력 네트워크를 확인할 수 있다. 네트워크를 통한 영국, 미국, 브라질의 사례는 이미 많은 다른 국가들이 공공데이터를 개방하고 LOD를 발행하는 좋은 모델이 되고 있다.

각자 고유한 데이터베이스를 이용자에게 공개하고 활용할 수 있도록 소통하기 위한 도구로서 LOD는 중요한 의미를 지닌다. 하지만 국중과 RISS처럼 대부분이 중복되는 데이터를 가지고 있는 기관에서 중복된 데이터의 종합적 관리를 위해 국가도서관 차원에서 링크드 데이터 발행을 위한 최소한의 공통 기준요소를 구축하는 정책이 필요하다. 한편 우리나라 국가 데이터 포털에 LOD로 등록된 공공데이터의 개수는 “우리나라의 자유무역협정 발효국에 대한 수출입실적”을 포함하여 9개의 데이터셋이 등록되어 있다. 이 플랫폼은 국가차원의 공식 데이터 플랫폼으로써 현재는 양적인 증가정책을 추진하고 있으나 향후 링크드 데이터에 대한 질적인 개선전략도 함께 이루어질 것이다. 이와 같이 국내와 국제 LOD플랫폼에 적극적으로 도서관관련 LOD를 공개하고 등록하여 이에 대한 적용활용사례에 대한 분석을 모니터링하는 것도 필요하다.

4.5 기술적인 확산 방안

Open API와 LOD가 다른 점은 연계와 소통이다. 일반적으로 데이터를 받아서 제공해주는 Open API와 달리 LOD는 링크 메타데이터를 통해 지속적인 연계와 융합이 가능하고 서비스 단계가 풍부해 질 수 있다. 하지만 현재 발행된 LOD는 공공데이터에 한정되어 있고, 경쟁적으로 API에 대한 LOD화를 추진하고 있어,

실제 도서관분야를 비롯한 타 분야에서 발행한 데이터 연계를 통한 활용사례를 적극적으로 추진해야 한다.

API처럼 XML포맷의 데이터를 수신하여 파싱하는 dom 라이브러리와 같이 클라이언트 자체적으로 개발해서 사용한다면 활용성은 낮아진다. 공통적으로 LOD를 파싱하고 커스터마이징 태그에 대한 처리가 가능한, 언어별(java용, win32용, 기타언어용 등) LOD 클라이언트 라이브러리를 제공할 수 있다면 LOD에 대한 활용성은 비약적으로 높아질 것이다. API처럼 인증을 하거나 서비스 계약 또는 키를 받아서 접근하는 방식보다는 발행된 LOD를 접근하는 것이 쉬운 방법이기 때문이다. LOD 발행 및 확산을 유도하는 것 못지않게 LOD를 활용하여 실제 상황에서 실현될 수 있는 Linked Data 활용을 위한 활용플랫폼과 활용가이드라인을 제공하는 전략도 필요하다.

5. 결론

링크드 데이터의 기본적 철학은 이기종 데이터를 연계하여 이용자에게 필요한 새로운 시각의 데이터를 재사용하는 것이다. 이기종 데이터간의 연계 혹은 인터링킹을 위해 기존에 소장된 데이터의 연계지향 구조화 형태인 LOD 형태로 소장데이터를 발행하는 것이다. 각 기관별 발행의 기준과 요소는 기관별 특성에 따라 서로 다른 형태로 이루어져 있어 실제적으로 기관별 발행된 데이터의 연계기관이 의도한 타 LOD 이외에 의도하지 않는 데이터간의 연계에 대해서는 아직 본격적인 연계사례를 갖고 있지 않다. 특히 도서관분야의 LOD는 실제적으로 많이 발행되어 있지만 국가데이터포털에 등록도 되어 있지 않은 상태이다. 도서관 분야에서 구축된 국가서지 LOD와 RISS LOD에 대한 활용성과 이기종 데이터간의 연계 방법에 대

한 분석 등을 통해 활용성을 높여야 한다. 이를 위해서 이 연구에서는 LOD 연계활용을 위한 방안을 제안하였다. LOD 발행에 있어 SKOS, FOAF, DC 등과 같은 국제 표준용어집을 사용하여 연계 용이성을 확보하는 방법이다. 두 번째로 국내와 해외 LOD연계에 대한 실질적 활용사례를 만들어가는 것이다. 셋째는 국내외 국가 LOD 플랫폼의 활용과 그에 대한 결과를 홍보하고 해당 플랫폼에 등록하는 것이다. 마지막으로 기술적인 공통화된 활용플랫폼 구성과 운용에 대한 가이드라인을 만드는 것이다.

참고문헌

- Ding L, et al. Linked Open Government Data. IEEE Intelligent Systems 2012; 27(3): 11-15.
- Harth A, et al. Linked Data Management. Chapman and Hall/CRC. 2014.
- Heath T, et al. Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool. 2011.
- Krier, L, et al. Data Management for Libraries. ALA. 2013.
- 윤소영. 공공데이터 활용을 위한 링크드 데이터 국가 연계체계 구축에 관한 연구. 정보관리학회지 2013; 30(1): 259-284.
- 정효숙 외. 일반인을 위한 링크드 데이터 생성 시스템 개발 및 활용. 컴퓨터교육학회 논문지 2011; 14(2):47-59.
- 한국정보화진흥원. 링크드 데이터 구축공정 가이드(v1.0). 2014.
- 한국지역정보개발원. LOD(Linked Open Data) 기반의 공공데이터 개방. 2014년 지역정보화 동향분석 제5호(2014. 4월).
- 한용진 외. 테이블로부터 링크드 데이터 생성을 위한 패턴 충돌 해소. 한국 지능시스템 학회 논문지 2014; 24(3): 285-291.