

## [시맨틱웹] Linked Data 표준화 동향

### Linked Data 개요

2001년 Scientific American지에 Tim. Berners-Lee가 시맨틱웹의 비전을 소개한 지 벌써 9년이라는 시간이 흘렀다. 무성했던 시맨틱웹의 시시비비를 가리기 전에 시맨틱웹에서 강조하고자 했던 데이터의 연계와 재사용성의 강조를 위한 새로운 키워드이며 기술적 접근이 Linked Data이다. 때문에 많은 부분의 표준화 및 기술적 동향은 시맨틱웹의 그것과 동일하다. 하지만 Linked Data 측면에서는 SPARQL Endpoint 등의 데이터 연계가 더욱 강조되고 있다.

Linked Data는 차세대 웹의 가장 핵심이 되는 데이터 네트워크를 위한 운동이며, 실천강령이다. 웹에서 자유롭게 데이터를 개방하여 연계할 수도 있도록 하고, 이들 데이터가 다시 협업할 수 있게 하여 진정한 데이터 웹을 실현하고자 하는 것이다. 때문에 웹에서의 데이터 유통을 위해 HTTP를 사용하고, 연계 및 접근성을 보장하기 위해 RDF(Resource Description Framework)와 SPARQL(Simple Protocol and RDF Query Language)을 사용한다.

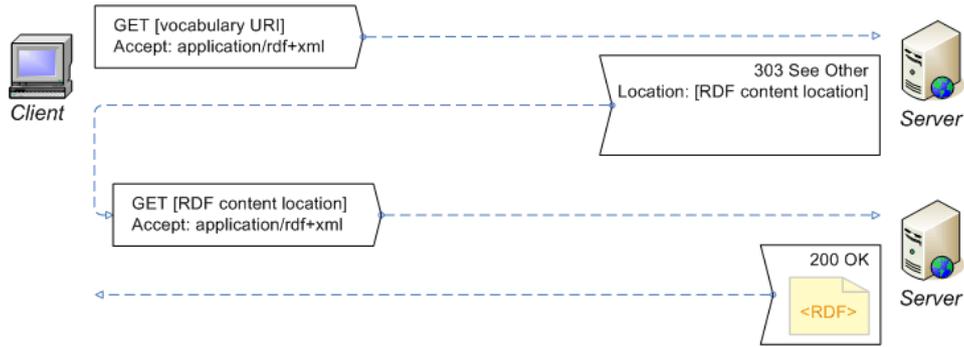
### Linked Data의 탄생

기술적 발전과 표준화의 노력에도 불구하고, 시맨틱웹의 비전과 성공은 많은 곳에서 회자되고 있다. 이를 일축하듯 시맨틱웹은 데이터의 웹을 위한 것이며, 데이터가 연계, 협업 될 때 진정한 의미로서의 시맨틱웹이 실현될 수 있음을 강조하던 Tim. Berners-Lee는 2009년 TED 컨퍼런스에서 Linked Data의 중요성에 대해 언급하며, 데이터 웹을 위해서는 데이터의 개방을 통한 연계, 협력이 이루어져야 함을 강조했다. 가공된 데이터가 아닌 RAW DATA가 더 많이 개방되고 연계되어야 시맨틱웹이 실현될 수 있고, 이 RAW DATA가 더 많이 Linked Data화 되어야 화려하고, 풍족한 정보의 꽃을 피울 수 있는 웹이 될 수 있음을 다시 한번 강조한 것이다.

### Linked Data의 4가지 원칙

1. 특정 개념을 URI로 명명한다.
2. HTTP를 통해 URI로 명명된 리소스에 접근할 수 있도록 한다.
3. URI로 접근했을 때 RDF 기반의 URI가 포함하고 있는 상세 정보를 제공한다.
4. RDF에 포함되어 있는 또 다른 관련 있는 개념으로 접근할 수 있도록 한다.

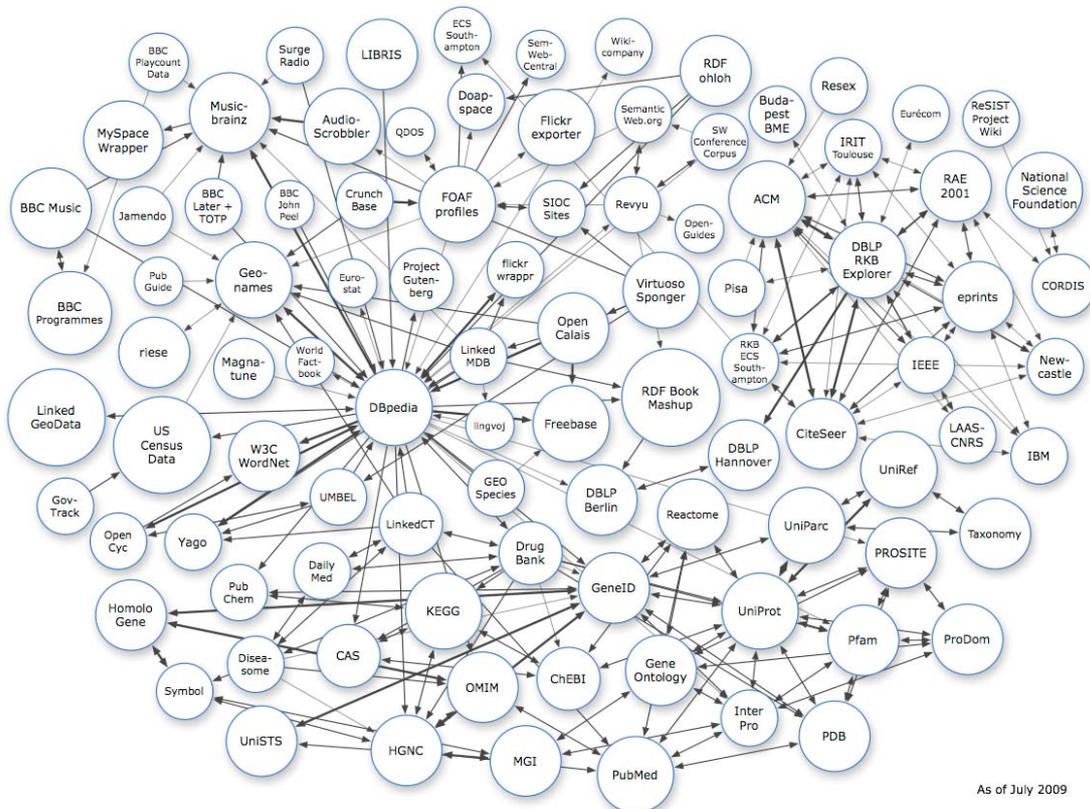
위의 4가지 단계가 Linked Data를 사용하는 기본 원칙이며, URI를 통해 특정 개념을 표현하고, HTTP 프로토콜을 이용해 해당 개념에 접근(lookup)할 수 있도록 하고 있다. 또한 RDF를 이용해 접근한 URI가 포함하고 있는 데이터들을 제공 받으며, 이 데이터 안에 포함되어 있는 또 다른 URI로의 접근을 통해 데이터를 연결시킬 수 있게 된다.



(출처: <http://www4.wiwiw.fu-berlin.de/bizer/pub/LinkedDataTutorial/>)

## Linked Data 동향

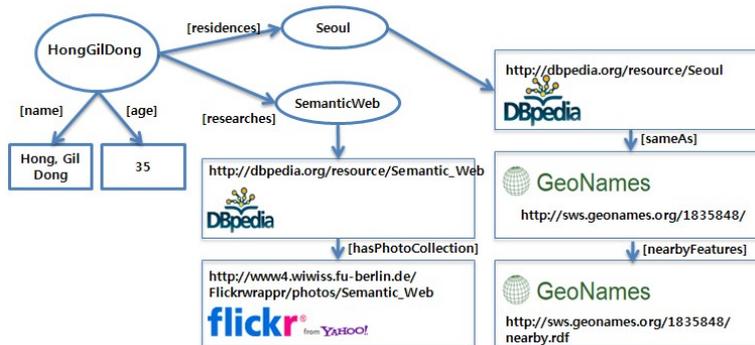
미국 정부는 고부가가치 공공의 개방형 데이터 셋을 국민 모두가 스스로 창조적으로 이용하도록 하기 위해 data.gov를 운영하고 있으며, 이를 RDF 형태의 기계가독형으로 발행하여 시스템적으로 보다 유용하게 활용할 수 있도록 하고 있다. 영국 정부(data.gov.uk)도 유사한 정책을 펴고 있다. 이를 통해 정부의 투명성을 높이고, 국민의 알 권리를 향상시키며 경제적, 사회적으로 데이터 개방 및 연계를 통한 가치 창출을 꾀하고 있다. 산업계에서는 BestBuy가 GoodRelations 온톨로지를 이용해 소상공인의 협력을 꾀하고 있으며, BBC와 NewYork Times 등도 자사의 콘텐츠 자산을 개방, 연계, 협업하기 위해 Linked Data에 앞장서고 있다. 대표적으로 활용되는 Linked Data는 DBpedia, MusicBranz, GeoNames 등이 있으며, 아래의 그림을 통해 확인할 수 있다.



(출처: <http://www.w3c.org>)

## Linked Data 예시

홍길동이라고 하는 사람에 대해 그가 거주하는(residences) 서울(Seoul)에 대한 정보는 DBPedia에서 제공하는 <http://dbpedia.org/resource/Seoul>로 연결할 수 있고, 이는 다시 GeoNames의 <http://sws.geonames.org/1835848/>로 연결할 수 있다. 마찬가지로, 연구하고 있는(researches) 시맨틱웹(Semantic Web)이라는 정보는 DBPedia의 [http://dbpedia.org/resource/Semantic\\_Web](http://dbpedia.org/resource/Semantic_Web)으로 연결할 수 있으며, 관련 사진 정보로의 지속적인 연결이 가능해진다.



(출처: <http://www.topquadrant.co.kr>)

## 결언

지구상에 존재하는 모든 데이터들이 디지털화되어 가고 있으며, 웹을 통해 확산되고 있다. 하지만 문서 중심의 별다른 규칙이나 표준 없이 생산되는 데이터들로 인해 몸살을 앓고 있으며, 실제로 원하는 데이터에 접근하기 위한 방법이 더욱 더 어려워지고 있다. Linked Data는 필요한 데이터를 연계하여 활용할 수 있게 함으로써 중복 데이터를 방지하고, 데이터의 신뢰도를 충족시키며, 데이터 재활용 가치를 높일 수 있게 해준다. 유비쿼터스 환경이 도래하며, 스마트폰 시대가 열리고, 소형화된 네트워크 기반의 랩탑 컴퓨터들의 시장점유율이 높아만 가는 네트워크 기반 디지털 시대에 살고 있는 우리에게 더욱 중요하게 다가오는 것이 디지털 데이터이며, 이 데이터들을 어떻게 생산, 유통 및 활용하느냐에 따라 우리나라 ICT의 미래가 달려 있다.

김성혁 (숙명여자대학교 문헌정보학과 교수, [ksh@sookmyung.ac.kr](mailto:ksh@sookmyung.ac.kr))