

링크드 데이터, 무엇을 어떻게 연결하는가? *

Linked data, what and how?

박진호(Park, Jin Ho)**

 0000-0003-2284-1217

목차

1. 서론
2. 데이터 중심 웹(web of data)
 - 2.1 웹과 시맨틱 웹
 - 2.2 시맨틱 웹과 링크드 데이터
3. 디지털 인문학 관점에서 링크드 데이터 활용
 - 3.1 디지털 인문학과 링크드 데이터
 - 3.2 링크드 데이터의 한계와 가능성
4. 결론

초록

본 논문은 웹의 진화에서 링크드 데이터와 시맨틱 웹의 중요성을 탐구하였다. 특히 링크드 데이터 기술이 무엇을 의미하는지와 구체적으로 어떻게 구축되고 보여지는지를 검토하였다. 링크드 데이터는 시맨틱 웹을 구현하기 위한 기술적 접근 중 하나로 볼 수 있으며, 웹이라는 플랫폼을 누구나 활용할 수 있는 거대한 데이터베이스로 만드는 방법이다. 링크드 데이터 기술은 지식의 네트워크를 구축하여 학제간 연구와 새로운 통찰을 가능하게 하지만, 데이터의 질과 일관성에 의존하며 데이터화 과정의 주관성을 내포한다. 앞으로 링크드 데이터는 기술적 진보와 함께 데이터의 표준화와 개방성을 중시해야 하며, 개인정보 보호와의 균형을 찾아야 한다. 또한, 이 기술이 가져올 윤리적, 사회적 책임에 대한 심도 있는 고찰이 요구된다. 링크드 데이터의 도입은 단순한 기술 변화를 넘어 학문적 연구와 산업 발전에 큰 기여를 할 것으로 보인다. 그러나 정보 사회가 전통적인 정보 접근 방식을 완전히 대체하지는 못했다는 점을 인식하고, 새로운 기술에 대한 비판적 이해와 안전한 활용이 중요하다.

주제어: 링크드 데이터, 시맨틱 웹, 링크드 오픈 데이터, 디지털 인문학

*본 연구는 2023 년 디지털인문학대회에서 발표한 ‘기타로 오토바이를 타자: 링크드 데이터, 무엇을 어떻게 연결하는가?’를 수정 및 보완한 것임.

** 한성대학교 지식정보문화트랙 조교수, jhp@hansung.ac.kr

Abstract

This paper explores the significance of Linked Data and the Semantic Web in the evolution of the web, specifically examining the implications of Linked Data technology and its implementation. Linked Data is viewed as a key technical approach for realizing the Semantic Web, transforming the web into an extensive database accessible to all. This technology fosters the creation of a knowledge network that facilitates interdisciplinary research and generates novel insights. However, its effectiveness relies heavily on the quality and consistency of the data, and it inherently involves the subjectivity of data processing. As we move forward, Linked Data must focus on the standardization and openness of data alongside technological advancements, while balancing these aspects with privacy concerns. Moreover, the adoption of this technology demands rigorous consideration of its ethical and social implications. The integration of Linked Data represents more than a mere technological shift; it promises substantial contributions to both academic research and industrial development. Nevertheless, it is crucial to acknowledge that the information society has not entirely supplanted traditional information access methods, emphasizing the need for critical engagement with and secure deployment of new technologies.

Keywords: Linked data, Semantic web, Linked Open Data, Digital Humanities

1. 서론

우리는 왜 대화가 가능할까? 무엇이 인간의 소통을 가능하게 할까?

우리가 같은 언어를 사용한다는 것도 여러가지 이유 중 하나로 볼 수 있다. 그렇다면 같은 언어를 사용하는 사람들은 소통이 가능할까? 그렇지 않다. 같은 말과 글을 사용한다고 하더라도 각자가 처한 경험, 상황에 따라 다른 생각을 한다. 특정 대화 주제나 대상에 대해서 서로 다른 개념을 갖고 있을 경우에도 우리는 제대로 소통할 수 없다. 다음 문장을 보자.

‘기타로 오토바이를 타자’

한국의 록 밴드 산울림을 알고 있는 사람이라면 이 문장이 어색하거나 문제가 있다고 생각하지 않을 것이다. 그렇지 않은 사람이라면 이 문장에 문제가 있다고 생각할 수 있다. 이유는 ‘기타’와 ‘오토바이’라는 단어가 통상의 개념에서 벗어났기 때문이다. 기타는 악기로 연주를 위한 도구이며, 오토바이는 이동수단이다. 위 문장은 각 단어가 갖는 본질, 개념에서 벗어났다. 기타는 탈 것이 아니기 때문에 ‘타다’라는 술어 사용은 적합하지 않다. 하지만 음악이라는 분야에서는 허용 가능한 표현이다. 즉 예술적 표현으로 받아들일 수 있다. 이렇듯 서로 같은 말과 글을 사용하지만 본질을 벗어나면, 개념이 다르면 우리는 소통이 불가능하다. 정보전달에 문제가 생긴다. 정보전달에 문제가 발생하면 사람들은 협업하거나 문제를 해결할 수 없다.

오늘날 가장 대표적인 소통, 정보전달 도구는 인터넷이다. 다양한 인터넷 도구 중에서도 누구나 손 쉽게 사용가능 하며, 널리 활용되는 대표적인 것이 웹(web)이다. 웹은 우리의 일상 생활, 작업 방식, 교육 방식, 세상과 소통 즉, 상호작용하는 방식을 근본적으로 변화시켰다. 정보는 웹에서 탄생하고, 풍부해지고, 재창조되고, 공유되고, 확산되며 소멸된다(박진호, 2013¹). 사실상 웹 없이 하루를 제대로 보내기 어려울 지경이다.

이런 웹에 중대한 변화가 일어나고 있다. 이 변화를 한 마디로 이야기하면 웹을 누구나 사용할 수 있는 글로벌 데이터베이스로 만드는 것으로 웹의 창시자인 팀 버너스 리(Tim Berners-Lee,

TBL)가 제안하였다(European Union, 2019)². 이제부터 웹은 단순한 문서 공유 수단에서 데이터 공유까지 가능한 수단으로의 변화를 의미한다.

디지털 인문학에서 데이터는 중요한 수단이다. 인문학은 전통적으로 텍스트, 이미지, 사운드와 같은 다양한 형태의 문화적 산물을 연구의 대상으로 삼아 왔다. 이러한 자료들이 디지털 형태, 특히 데이터 형태로 전환되면, 보다 광범위하고 정교한 분석이 가능하고, 그 과정에서 새로운 발견과 해석이 이루어질 수 있는 가능성을 열어준다.

웹이 문서 공유 중심의 플랫폼에서 데이터 공유 중심으로 진화한 것은 특히 디지털 인문학 분야에서 중대한 변화이다. 문서 전파를 용이하게 하는 공개 플랫폼으로 구상되었던 웹이 연결성과 상호작용성을 활용하여 보다 동적인 데이터 교환을 촉진하는 방향으로 발전하고 있고 이 변화를 대표하는 용어가 링크드 데이터(linked data), 시맨틱 웹(semantic web)이다. 이러한 변화는 디지털 인문학 학자들이 광범위한 기본 자료에 접근할 수 있게 할 뿐만 아니라, 다양한 데이터 세트를 연결하고 분석함으로써 새로운 연구를 진행할 수 있게 한다. 링크드 데이터와 시맨틱 웹 기술을 활용함으로써 연구자들은 역사적, 문화적, 언어적 경계를 넘나드는 패턴과 관계를 밝힐 수 있으며, 복잡한 인간 이야기에 대한 이해를 깊게 할 수 있다. 웹의 데이터 공간으로 전환은 인문학의 방법론적 도구를 풍부하게 하며, 기존 문서 중심 접근 방식의 한계에 의해 제약 받았던 새로운 통찰력과 학제간 연구를 촉진할 수 있다.

본 고에서는 데이터 중심 웹의 대표적인 방법이자 기술인 링크드 데이터가 무엇인지와 어떻게 동작하는지 살펴보고 그 유용성과 한계도 함께 검토하였다. 링크드 데이터의 이해에 있어서 기술적인 접근이 중요하지만 근본적으로 소통의 관점에서 그 구조와 작동방식을 검토하였다. 향후 디지털 인문학분야 연구자가 유용성과 한계를 분명히 이해하여 연구에 도움이 될 수 있기를 기대한다.

2. 데이터 중심 웹(web of data)

2.1 웹과 시맨틱 웹

웹의 창시자는 팀 버너스 리이다. 팀 버너스 리는 유럽입자물리연구소(Conseil Européen pour la Recherche Nucléaire, CERN)에 근무하던 1989년 3월 "Information Management: A Proposal"이라는 제목의 제안서를 작성하였다. 오늘날 웹의 원형으로 개발 목적은 연구 결과와 정보를 CERN 내의 다양한 그룹과 더 효율적으로 공유할 수 있는 시스템을 만드는 것이었다. 핵심은 하이퍼텍스트를 이용하여 정보를 연결하는 것이다(Berners-Lee & Timothy, 1989)

1993년에 CERN은 웹 기술을 전 세계에 무료로 제공하기로 결정했으며, 이는 웹의 빠른 확산에 크게 기여했다. 웹의 개발은 CERN에서의 연구와 정보 공유의 필요에 의해 추진되었지만, 이는 곧 전 세계적인 정보 접근의 혁신으로 이어졌다. 초기 웹은 주로 텍스트 기반의 정보를 다루었지만, 점차 멀티미디어 요소가 통합되면서 오늘날의 다양한 웹 경험을 가능하게 했다. 팀 버너스 리는 이후 웹의 표준화와 개방성을 촉진하기 위해 월드 와이드 웹 컨소시엄(W3C)을 설립하였고, 웹의 개발과 혁신을 지속적으로 이끌어왔다(CERN, 2014)³.

이런 웹의 모습은 HTML 문법을 따르는 문서들이 HTTP 프로토콜을 사용해서 서로 연결(hyperlink)된 거대한 문서집합으로 표현할 수 있다. 현재 모습으로도 웹은 훌륭한 정보자원으로

역할을 잘 수행하고 있다. 그런데 팀 버너스 리는 현재 웹의 좀 더 발전적인 활용 방안으로 시맨틱 웹이라는 새로운 개념을 제안하였다. 팀 버너스 리는 시맨틱 웹을 “현재 웹이 확장된 형태로, 잘 정의된 의미를 정보에 부여함으로써 사람과 컴퓨터의 협업을 보다 원활하게 할 수 있도록 하는 것”으로 정의하였다(Berners-Lee, Hedndler and Lassila, 2001⁴).

시맨틱 웹 정의에서는 이를 어디에 만들고, 어떻게 만들고, 왜 필요한지, 누구를 위한 것인지가 잘 드러난다. 정의와 용어에서 알 수 있듯이 시맨틱 웹은 현재의 웹이라는 플랫폼을 그대로 활용한다. 그리고 중요한 것은 누구를 위해 왜 하는가인데, 이는 사람을 위해 사람과 컴퓨터의 협업이 가능한 형태로 구축해야 함을 의미한다. 그 방법이 바로 잘 정의된 의미 정보를 부여하는 것이다. 핵심은 여기에 있다. 현재 우리가 사용하고 있는 웹은 인간의 가독성을 중심에 둔 형태로 기계가 처리하기에는 어렵다. 단적인 예는 다음의 코드를 비교함으로써 쉽게 이해할 수 있다.

```
<h1>링크드 데이터의 이해</h1>
```

```
<h2>박진호</h2>
```

<h1>과 <h2> 태그는 HTML에서 웹 문서 내의 제목을 정의하는 데 사용된다. <h1>은 주요 제목을 나타내며, 일반적으로 페이지에서 가장 중요하고 눈에 띈다. <h2>는 <h1>의 하위 제목으로, 덜 중요하고 <h1>보다는 덜 눈에 띄게 표시된다. 한글, 한국어를 사용하는 인간은 웹 문서에서 이 정보를 접하면 제목과 저자라고 쉽게 이해한다. 그런데 기계는 제목과 저자로 처리할 수 있을까? <h2>를 사람이름이라고 인식할 수 있을까? 다음 코드는 <h1>, <h2> 대신에 명확하게 태그가 표현하고자 하는 값이 무엇을 의미하는지 표현하고 있다.

```
<title>링크드 데이터의 이해</title>
```

```
<creator>박진호</creator>
```

위 예제는 소위 XML 형태로 HTML과 다르게 값의 의미를 짐작할 수 있다. XML 형태로 기계 처리도 가능하다. 하지만 XML로는 위의 시맨틱 웹 정의에서 말한 잘 정의된 의미를 정보가 부여되었다고 보기 어렵다. <title> 태그는 인간이 보기에는 그 의미를 이해할 수 있지만 기계가 그 의미를 알고 처리한다고 보기는 어렵다. XML은 구조화된 정보 전달로써 가치는 있지만 시맨틱 웹이 추구하는 의미처리에는 미흡하다.

의미기반 웹(semantic web)이 새로운 웹의 형태적 측면을 강조한 표현이라면 이전 웹은 구문기반 웹(syntax web)으로 볼 수 있다. HTML 태그 사용이 대표적인 구문 기반 웹의 핵심이다. 기존 웹을 플랫폼으로 그대로 사용하고 서로 연결된 새로운 데이터 웹을 만드는 것이 핵심이라면 시맨틱 웹이 이전 웹과 다른 가장 큰 특징은 의미전달이 가능한 새로운 표현언어에 있다. 이것이 RDF(Resource Description Framework)이며, 일반적으로 링크드 데이터로 데이터를 발행하는데 표준으로 활용된다. 시맨틱 웹은 현재의 웹을 확장하는 것으로 웹의 핵심기술을 그대로 활용한다. 단, 구문(syntax)이 아니라 의미(semantic)전달을 위해 HTML이 아니라 RDF라는 새로운 기술방식이 필요한 것이다.

2.2 시맨틱 웹과 링크드 데이터

시맨틱 웹과 링크드 데이터는 밀접한 관계를 가진다. 위의 정의에서 살펴본 것처럼 시맨틱 웹은 웹 데이터에 의미를 부여하여 기계가 처리할 수 있도록 설계된 웹의 확장된 형태이다. 링크드 데이터는 시맨틱 웹의 핵심 기술 중 하나로 이 두 개념은 서로를 보완하는 관계에 있다. 시맨틱 웹은 웹 데이터가 단순히 문서의 집합으로 존재하는 것이 아니라, 데이터 간의 명확한 관계를 통해 연결된 지식의 네트워크 형성을 추구한다.

링크드 데이터는 시맨틱 웹의 구현을 위한 구체적인 방법론이다. 이는 웹 상의 데이터를 표준화된 형식(RDF)으로 구조화하고, URI(Uniform Resource Identifier)를 사용하여 데이터 포인트 간에 명확한 링크를 생성함으로써 데이터 간의 관계를 명시적으로 표현한다. 이렇게 연결된 데이터는 서로를 참조하며, 이를 통해 사람과 기계가 데이터의 의미와 관계를 보다 명확하게 이해할 수 있게 된다. 결국, 시맨틱 웹은 링크드 데이터를 통해 실현될 수 있으며, 링크드 데이터는 시맨틱 웹의 이론을 실제로 구현하는 데 필수적인 기술이다. 이 두 개념은 함께 작동하여 웹을 단순한 정보의 저장소에서 의미와 관계가 풍부한 지식의 웹으로 변화시키는데 기여한다.

링크드 데이터는 데이터 제공자는 구축, 이용자는 활용해야 하는 최종 형태이다. 링크드 데이터로 데이터를 설계하고 발행하기 위해서는 기본적으로 지켜야 하는 원칙이 존재하며, 역시 팀 버너스 리가 제안하였고 아래와 같다(W3C, 2023⁵)

- 특정 개념(things)을 URI를 사용해서 명명함
- HTTP를 활용해서 URIs로 명명된 자원에 접근할 수 있도록 함
- URI로 정보를 찾았을 때 RDF(S), SPARQL과 같은 표준을 활용해서 유용한 정보를 제공해야 함
- 또 다른 URI를 포함하고 있어서 더 많은 것(개념, things)들을 탐색할 수 있어야 함

실제 발행된 링크드 데이터를 검토하면 위 원칙을 쉽게 이해할 수 있다. <그림 1>은 DBpedia의 정보 중 일부로 우리나라의 수도 ‘서울’에 대해 설명하고 있는 링크드 데이터 페이지이다. DBpedia는 2007년에 시작된 프로젝트로, 위키백과의 정보를 구조화하여 기계가 이해하고 처리할 수 있는 데이터로 변환하는 것을 목표로 한다 (DBpedia, n.d.⁶). DBpedia의 개발 배경은 웹의 데이터를 더 잘 이해하고 연결하는 데 있다. 이를 통해 사용자는 다양한 데이터 소스 간의 상호운용성을 높이고, 인공지능 및 데이터 분석 응용 프로그램에 활용할 수 있는 풍부한 데이터 집합을 제공받을 수 있다. DBpedia는 현재 다양한 도메인의 데이터를 포괄하고 있으며, 이를 통해 여러 다른 데이터 세트와의 상호 연계를 가능하게 한다. DBpedia는 현재도 계속 발전하고 있는 다양한 언어를 지원하는 글로벌 오픈 데이터로 중요한 역할을 하고 있다(Dojchinovski et al., 2018⁷).

<그림 1>의 URL은 “https://dbpedia.org/page/Seoul”이다. 링크드 데이터의 첫 번째 원칙은 특정 개념을 URI를 사용하여 명명하는 것으로 <그림 1>은 ‘서울’이라는 개념을 URI로 명명 즉 식별하고 있음을 보여준다. URI는 중복을 허용하지 않는 고유한 명명체제로 식별자 역할을 수행할 수 있다. 두 번째 원칙은 HTTP를 활용해서 URIs로 명명된 자원에 접근할 수 있도록 하는 것으로 <그림 1>에서 보는 것처럼 서울에 해당하는 URI를 통해서 ‘서울’ 정보에 접근할 수 있음을 확인할 수 있다. 세 번째 원칙은 URI로 정보를 찾았을 때 RDF(S), SPARQL과 같은 표준을 활용해서 유용한 정보를 제공해야 한다는 것이다.

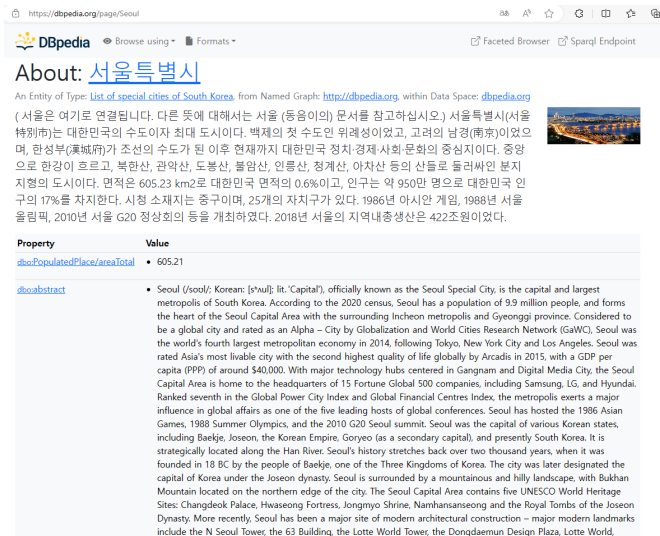


그림 1. DBpedia 에서 ‘서울’이라는 개념을 설명하고 링크드 데이터 페이지(출처:

DBpedia(<https://dbpedia.org/page/Seoul>)

현재 <그림 1>은 HTML 형태로 표현된 정보로 이 외 다양한 구조화된 형식으로 ‘서울’에 대한 정보를 제공한다(<그림 2 참조>).

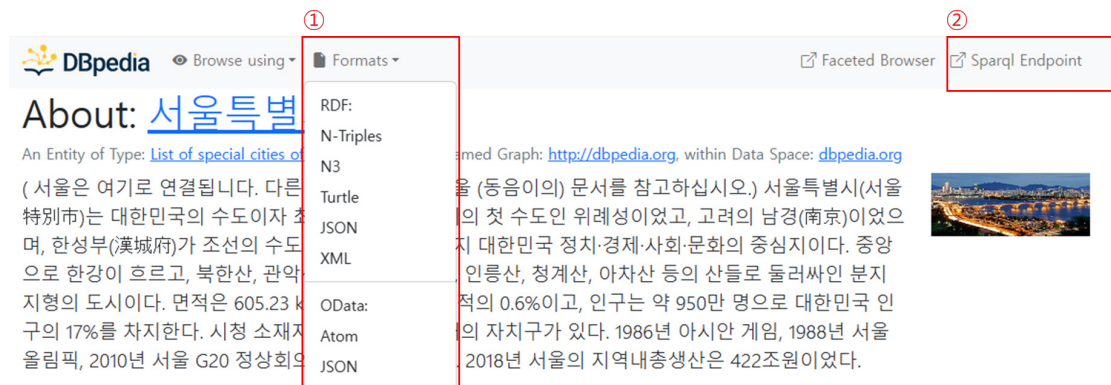


그림 2. DBpedia 에서 ‘서울’

(출처: DBpedia(<https://dbpedia.org/page/Seoul>))

<그림 2>에서 보면 두 가지를 확인할 수 있다. 첫 번째는 ‘Formats’ 서비스로 HTML 외 다양한 RDF 형식의 데이터 서비스를 확인할 수 있다. 두 번째는 Sparql Endpoint이다. <그림 3>은 RDF 형식 중 XML 형식의 데이터를 다운로드하여 확인한 결과 중 일부이다.


```

1 <?xml version="1.0" encoding="utf-8" ?>
2 <rdf:RDF
3   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
4   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
5   xmlns:owl="http://www.w3.org/2002/07/owl#"
6   xmlns:foaf="http://xmlns.com/foaf/0.1/"
7   xmlns:skos="http://www.w3.org/2004/02/skos/core#"
8   xmlns:dbp="http://dbpedia.org/property/"
9   xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
10  xmlns:dbo="http://dbpedia.org/ontology/"
11  xmlns:dcterms="http://purl.org/dc/terms/"
12  xmlns:georss="http://www.georss.org/georss/"
13  xmlns:gold="http://purl.org/linguistics/gold/"
14  xmlns:schema="http://schema.org/"
15  xmlns:prov="http://www.w3.org/ns/prov#"
16  xmlns:ns13="http://dbpedia.org/ontology/PopulatedPlace/" >
17  <rdf:Description rdf:about="http://dbpedia.org/resource/Caffe_Bene">
18    <dbo:wikiPageWikiLink rdf:resource="http://dbpedia.org/resource/Seoul" />
19    <dbp:locationCity rdf:resource="http://dbpedia.org/resource/Seoul" />
20    <dbo:locationCity rdf:resource="http://dbpedia.org/resource/Seoul" />
21  </rdf:Description>
22  <rdf:Description rdf:about="http://dbpedia.org/resource/Cairo">
23    <dbo:wikiPageWikiLink rdf:resource="http://dbpedia.org/resource/Seoul" />
24  </rdf:Description>
25  <rdf:Description rdf:about="http://dbpedia.org/resource/Calabash">
26    <dbo:wikiPageWikiLink rdf:resource="http://dbpedia.org/resource/Seoul" />
27  </rdf:Description>
28  <rdf:Description rdf:about="http://dbpedia.org/resource/Calgon_Carbon">
29    <dbo:wikiPageWikiLink rdf:resource="http://dbpedia.org/resource/Seoul" />
30  </rdf:Description>
31  <rdf:Description rdf:about="http://dbpedia.org/resource/Call_of_Duty:_Advanced_Warfare">
32    <dbo:wikiPageWikiLink rdf:resource="http://dbpedia.org/resource/Seoul" />
33  </rdf:Description>
34  <rdf:Description rdf:about="http://dbpedia.org/resource/Calvin_Smith">
35    <dbo:wikiPageWikiLink rdf:resource="http://dbpedia.org/resource/Seoul" />
36  </rdf:Description>
37  <rdf:Description rdf:about="http://dbpedia.org/resource/Calvin_Stamp">
38    <dbo:wikiPageWikiLink rdf:resource="http://dbpedia.org/resource/Seoul" />
39  </rdf:Description>
40  <rdf:Description rdf:about="http://dbpedia.org/resource/Cambodia_Angkor_Air">
41    <dbo:wikiPageWikiLink rdf:resource="http://dbpedia.org/resource/Seoul" />
42  </rdf:Description>
43  <rdf:Description rdf:about="http://dbpedia.org/resource/Camden_Pulkinen">
44    <dbo:wikiPageWikiLink rdf:resource="http://dbpedia.org/resource/Seoul" />

```

그림 3. DBpedia 에서 ‘서울’의 RDF/XML 중 일부

<그림 3>은 RDF 형식 중 XML 형식의 데이터를 다운로드하여 확인한 결과 중 일부이다. <그림 4>는 Sparql 서비스화면이다.

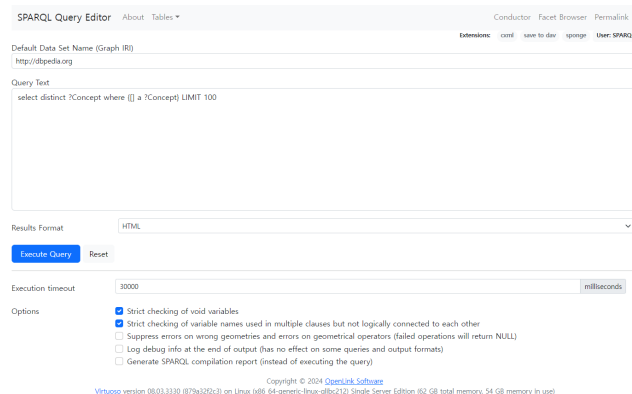


그림 4. DBpedia 의 Sparql 서비스 화면
(출처: DBpedia(<https://dbpedia.org/sparql/>))

Sparql은 RDF 데이터를 조회하기 위해 사용되는 질의 언어이다. SPARQL을 사용하면 RDF 데이터 집합에서 복잡한 질의를 수행하여 특정 정보를 추출하거나 데이터 간의 관계를 탐색할 수 있다.

마지막 원칙은 또 다른 URI를 포함하고 있어서 더 많은 것(개념, things)들을 탐색할 수 있어야 한다는 것이다. 이 마지막 원칙이 가장 핵심이다. 링크를 통해서 무엇을 연결하고 어떤 정보를 제공하는지를 확인하면 시맨틱 웹의 핵심인 ‘의미’를 부여하는 것이 어떤 것인지 확인할 수 있다. 우리에게 익숙한 현재의 웹, 즉 문서 중심의 웹에서 링크는 관련된 또다른 문서로의 이동을 의미

한다. 반면 시맨틱 웹, 링크드 데이터에서의 링크는 2가지 서로 다른 정보를 제공한다. 첫 번째는 문서 중심 웹과 같이 관련 정보, 또 다른 개념으로의 이동과 관련 데이터 제공이다. <그림 5>가 그 예이다.

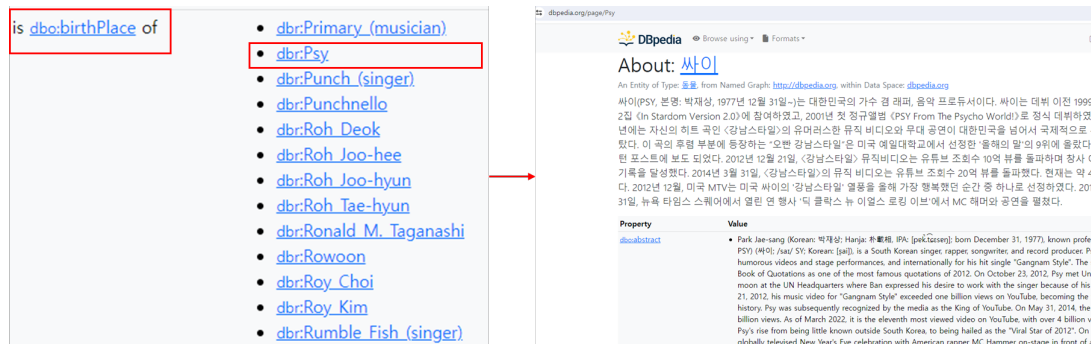


그림 5. 서울에서 태어난 사람을 연결하는 예시

(출처: DBpedia(<https://dbpedia.org/page/Psy>))

<그림 5>를 보면 ‘is dbo:birthplace of’ 즉 서울을 출생지로 갖는 사람에 대한 정보를 연결하여 보여주고 있다. 해당 정보 중 ‘dbr:Psy’를 연결하면 가수 ‘싸이’를 설명하는 또 다른 링크드 데이터 페이지로 이동함을 확인할 수 있다. 이는 기존과 같은 개념의 링크 정보이다. 그런데, ‘싸이’라는 개념, 객체와 ‘서울’이라는 개념, 객체를 연결해주는 속성인 ‘dbo:birthPlace’라는 속성에도 링크 정보가 있는 것을 확인할 수 있다. 연결된 정보를 확인해보면 <그림 6>과 같다.

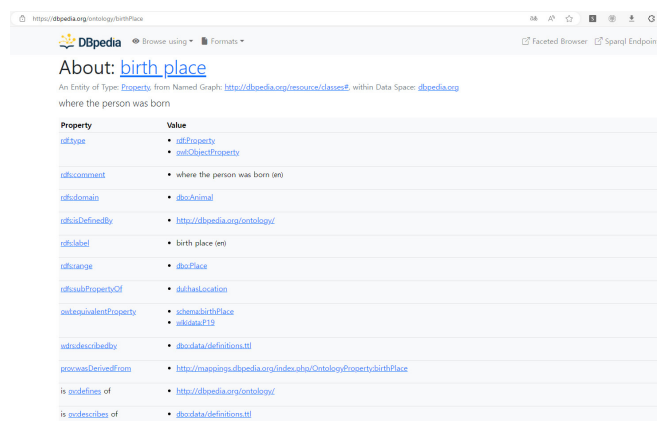


그림 6. ‘dbo:birthPlace’ 속성에 연결된 정보

(출처: DBpedia(<https://dbpedia.org/ontology/birthPlace>))

<그림 6>은 ‘dbo:birthPlace’ 속성에 대한 정보를 제공해준다. 이 요소의 정의가 무엇이고, 유형(type)이 무엇인지 등과 같이 해당 속성을 사용하는데 필요한 모든 규칙들이 정의되어 있다. 링크드 데이터는 기술대상인 개체나 개념 외에 이 개체와 개념을 표현하는 데 필요한 모든 규칙들도 URI를 기반으로 상세정보를 제공한다는 점이 기존 웹과 다른 점이다. 서두에 제시했던 것처럼 시맨틱 웹은 웹을 문서 정보가 아니라 데이터베이스로 만들기 위한 작업으로 볼 수 있다. 웹이라는 자유로운 공간에서 데이터베이스로 사용하기 위해서는 데이터베이스 구성에 필요한 규칙

들도 함께 공유되어야 한다. 이 점을 고려한다면 링크드 데이터의 연결이 갖는 특징을 쉽게 이해할 수 있다. 시맨틱 웹에 구현되는 링크드 데이터 형식의 데이터들은 설계도가 함께 표현되어 있고, 설계도 역시 링크드 데이터 발행 원칙을 그대로 준수하며, 웹 표준을 따른다. URI는 식별자이자, 연결을 가능하게 핵심 기술요소로 작동된다. 이런 특징은 RDF 기술 규칙에 따라 생성되는 데이터를 보면 쉽게 이해할 수 있다.

RDF는 데이터를 그래프(graph) 형태로 표현할 수 있는 모델이다. 그래프 데이터베이스와 유사하게 RDF는 데이터를 노드(node)와 엣지(edge, 노드 간의 관계를 나타내는 연결선)로 구성된 그래프 형태로 표현할 수 있다. 각 노드는 리소스나 속성값을 나타내며, 엣지는 노드들 사이의 관계를 표현한다. 지금까지 예제로 사용한 ‘서울’의 RDF 데이터를 예를 좀 더 검토하면 RDF로 표현 가능한 두 가지 형식을 발견할 수 있다. 먼저 HTML로 표현된 서울의 RDF 데이터 중 일부를 살펴보면 <그림 7>과 같다.

dbo:areaTotal	• 605210000.000000 (xsd:double)
dbo:country	• dbr:South Korea

그림 7. HTML로 표현된 서울 RDF 데이터 중 일부
(출처: DBpedia(<https://dbpedia.org/page/Seoul>))

<그림 7>의 데이터를 RDF 형식 중 하나인 N-Triples로 살펴보면 다음과 같다.

```
<http://dbpedia.org/resource/Seoul> <http://dbpedia.org/ontology/PopulatedPlace/areaTotal> "605.21"
<http://dbpedia.org/resource/Seoul> <http://dbpedia.org/ontology/country><http://dbpedia.org/resource/South_Korea> .
```

위 두 가지 형식을 간단하게 정리하면 다음과 같다.

URI – URI – Literal

URI – URI – URI

RDF는 아주 간단한 방식으로 데이터를 기술한다. 위와 같이 트리플(triple)형태로 데이터를 기술하며, 형태는 2가지에 불과하다. RDF 트리플은 주어(subject), 술어(predicate), 목적어(object)의 세 가지 구성요소로 구성된다. 주어는 기술대상 데이터 자원을, 목적어(object)는 주어와 관계 맺는 또 다른 자원 혹은 문자형태의 값(literal)을, 술어는 주어와 목적어의 관계를 설명하는 속성을 의미한다. 웹을 플랫폼으로 하기 때문에 이 3가지 요소는 URI를 기반으로 서술할 수 있다. URI는 유일성을 보장하기 때문에 이를 기반으로 표현된 데이터들은 정확하게 연결되고, 풍부한 데이터 생태계 구성이 가능한 것이다.

또 한가지 중요하게 고려해야 할 것은 상호운용성이다. 내가 RDF 형식으로 데이터를 발행하더라도 표준에 맞추어 데이터를 기술하지 않으면 다른 데이터와의 상호운용은 불가능하다. DBpedia ‘서울’ 페이지에서 사용하고 있는 다양한 속성들은 dbpedia ontology 모델을 준용한다. 그래서 각 속성들은 ‘dbo:’라는 네임스페이스를 사용한다. 예를 들면 ‘dbo:areaTotal’, ‘dbo:abstract’, ‘dbo:country’ 등 이다. 해당 온톨로지에서 정의하는 다양한 클래스와 속성 정보 역시 RDF 형식으로 내용을 확인할 수 있다.

DBpedia ⌕ Browse using 📄 Formats 🔗 Faceted Browser 🔗 Sparql Endpoint

About: <http://dbpedia.org/ontology/>

An Entity of Type: [Vocabulary](#) from Named Graph: <http://dbpedia.org/resource/classes#>, within Data Space: [dbpedia.org](#)

This ontology is generated from the manually created specifications in the DBpedia Mappings Wiki. Each release of this ontology corresponds to a new release of the DBpedia data set which contains instance data extracted from the different language versions of Wikipedia. For information regarding changes in this ontology, please refer to the DBpedia Mappings Wiki.

Property	Value
http://creativecommons.org/ns#license	http://creativecommons.org/licenses/by-sa/3.0/
owl:defines	xsd:anyURI xsd:boolean xsd:date xsd:dateTime xsd:double xsd:float xsd:gDay xsd:gMonth xsd:gMonthDay xsd:gYear xsd:gYearMonth xsd:integer xsd:negativeInteger xsd:nonNegativeInteger xsd:nonPositiveInteger

그림 8. DBpedia ontology 에서 사용하는 용어집 중 일부
(출처: DBpedia(<https://dbpedia.org/ontology/>))

<그림 8>은 DBpedia ontology에서 사용하는 다양한 용어들을 설명하고 있는 용어집이다. 정보 자원을 기술할 때 내가 기술하는 대상 정보자원의 유형에 적합한 표준 온톨로지를 사용함으로써 데이터의 상호운용성이 높아지는 것이다.

현재 다양한 정보자원 관리 기관과 조직들이 다양한 형태의 정보자원을 링크드 데이터 형태로 발행하고 있으면 이는 거대한 웹 데이터베이스를 형성하고 있다. <그림 9>는 2007년 5월에 시각화된 링크드 오픈 데이터 다이어그램이고 <그림 10>은 2023년 9월에 작성된 것이다. 해당 그래프는 전세계 링크드 데이터를 대상으로 시각화한 것이다. 해당 그래프는 Dublin City University, NUI Galway, University College Cork, University College Dublin의 연구자들이 포함된 Insight Center for Data Analytics에서 유지 관리하고 있다.

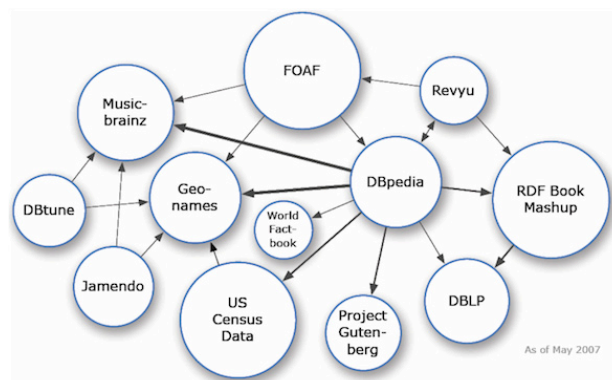


그림 9. 2007 년 5 월 LOD 클라우드 다이어그램
(출처: The Linked Open Data Cloud (<https://lod-cloud.net/>))

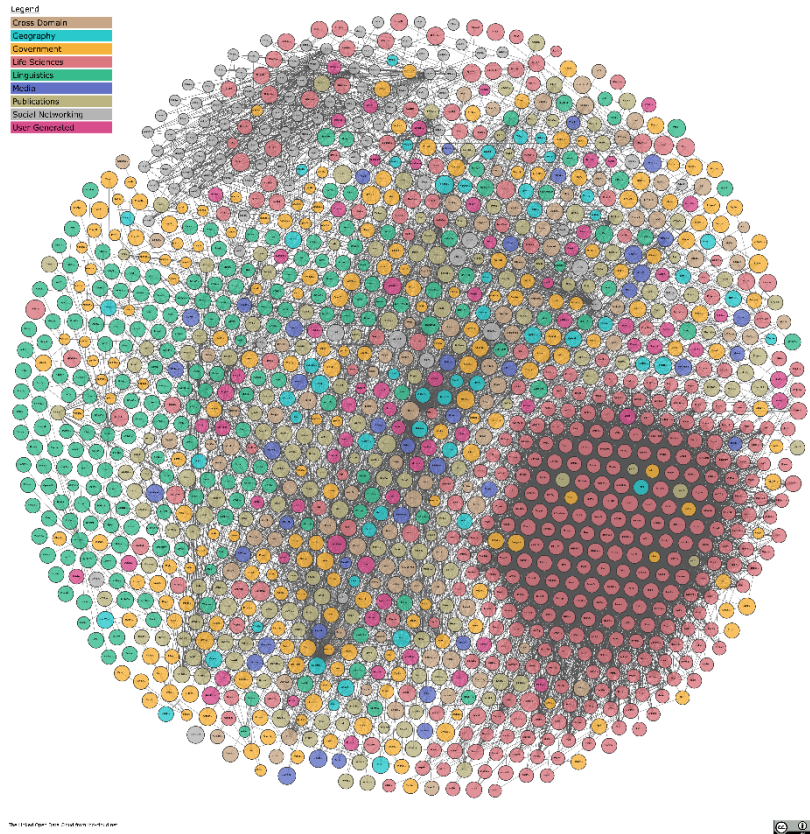


그림 10. 2023 년 9 월 LOD 클라우드 다이어그램
(출처: The Linked Open Data Cloud (<https://lod-cloud.net/>))

링크드 데이터는 개방형 데이터(open data) 형태로 지속적으로 증가하고 있으며, 데이터 기반의 새로운 웹 생태계를 만들어가고 있다. 오픈 데이터는 저작권에 상관없이 누구나 자유롭게 데이터를 활용, 재활용, 재배포할 수 있는 데이터를 말한다. 일반적으로 링크드 데이터는 링크드 오픈 데이터(Linked Open Data)로 불리는데 이는 링크드 데이터와 오픈 데이터의 합성어이다. 웹이라는 개방된 생태계를 활용한다는 점에서 그리고 링크드 데이터가 다른 데이터와 연결이 중요하다는 점을 감안할 때 이 둘은 떼어낼 수 없는 관계에 있다고 볼 수 있다. 링크드 데이터는 개방형 생태계와 만날 때 그 가치가 증가한다고 할 수 있다.

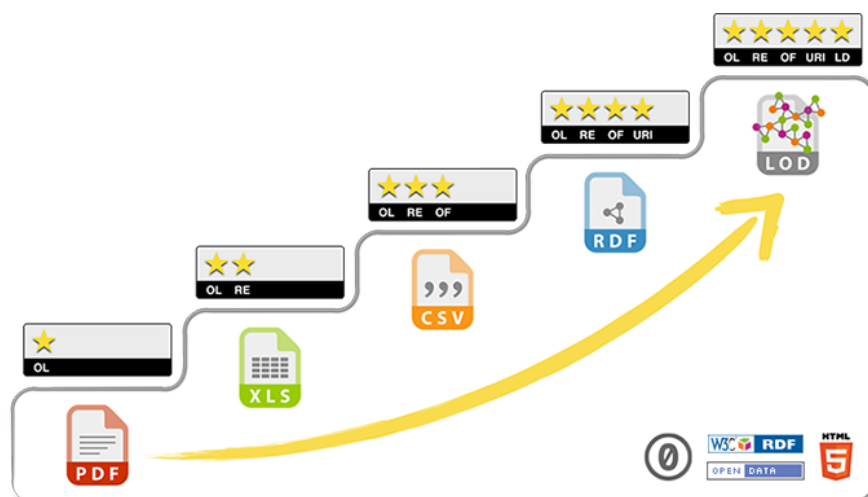


그림 11. 5Star open data⁸

(출처: 5 ★ OPEN DATA(<https://5stardata.info/en/>))

<그림 11>은 팀 버너스 리가 5개 별점으로 오픈 데이터의 단계를 제안한 것이다. 오픈 데이터 중 가장 높은 수준이자 지향점은 LOD이다. 별점 1점부터 5점까지는 단순히 데이터를 개방하는 것이 아니라 누구나 자유롭게 활용할 수 있는 정도와 데이터가 의미기반의 구성되어 있는 정도에 따라 구분하였다.

지금까지 시맨틱 웹, 그 중에서도 링크드 데이터에 대해서 간략하게 검토해 보았다. 실제 링크드 데이터 변환과 발행을 위해서는 다양한 기술적 장치 도입을 포함한 노력이 필요하다. 여기서는 링크드 데이터의 이해를 위해 개념을 중심으로 검토하였다.

3. 디지털 인문학 관점에서 링크드 데이터 활용

3.1 디지털 인문학과 링크드 데이터

디지털 인문학은 진화하고 있는 학문분야로 볼 수 있다. 기술발전과 새로운 문화형성에 따라 그 개념은 변화하고 있다고 볼 수 있다. 다양한 정의들이 존재하지만 일반적으로 디지털 인문학은 디지털 기술과 인문학의 교차점에 있는 분야로 볼 수 있다. 필연적으로 디지털 도구를 활용한 학문 연구와 접근이 필요하다. 이와 함께 디지털 형태의 정보자원 혹은 디지털 도구를 활용한 오래된 정보자원의 변화 등을 통해 기존 연구에 대한 재해석, 보다 정확하고 근거에 기반한 연구결과 도출이 가능하다. 즉 디지털 인문학 작업은 질적 연구에 정량적 방법을 적용한 새로운 길을 열었다고 볼 수 있다.

링크드 데이터는 웹을 기반으로 누구나 자유롭게 사용할 수 있는 개방형 데이터베이스이다. 덤프 형태(dump)의 데이터세트 제공이나, 오픈 API(open API)와 같이 널리 활용되고 있는 개방방식에 비하면 늦은 출발이라고 볼 수 있지만, 확산속도는 대단하다. 여러가지 이유가 있겠지만 기본적으로 링크드 데이터의 원천 데이터는 이미 현재 웹을 구성하고 있는 다양한 데이터베이스들에 기반하고 있다는 점이 가장 큰 영향으로 볼 수 있다. 대부분의 웹 정보자원들은 데이터베이스 형태로 구조화되어 관리되고 있는 상태이므로 이를 링크드 데이터 원칙에 맞추어 표준 용어집을 사용하여 발행하고 공유하는 일은 완전히 새로운 일을 시작하는 것과는 다르다.

링크드 데이터의 확산은 디지털 인문학 입장에서 자유롭게 활용할 수 있는 고품질 데이터의 확보라는 관점에서 보면 긍정적이다. 특히 웹을 플랫폼으로 하는 대규모 데이터세트라는 점에서 매력적이다. 링크드 데이터의 본질적인 구조는 잘 정의되고 표준화된 개념과 속성에 기반을 두고 있으며, 이는 다양한 플랫폼과 응용 프로그램에서 데이터에 대한 일관된 해석을 보장한다. 이러한 특성은 데이터세트가 신뢰할 수 있고 기계 처리에 적합하며, 학술 및 산업에서 응용에 대한 유용성을 향상시킬 수 있다는 점에서 더 중요하다.

링크드 데이터는 일관된 의미 구조 덕분에 데이터세트가 본질적으로 상호 운용 가능하다. 이는 다양한 도메인에서 공통의 이해와 공유라는 점에서 중요하며, 서로 다른 학문의 통찰을 활용하는 연구를 촉진할 수 있다. 또한, 링크드 데이터 원칙을 채택함으로써, 맥락과 유용성이 풍부한 데이터세트 생성을 촉진할 수 있다.

그런데 링크드 데이터의 이 의미 구조는 기술하고자 하는 대상에 대한 정의와 기술규칙에 근거

한다. 기술적인 관점에서 보면 링크드 데이터는 우리가 알아야 하는 새로운 것임에는 분명하다. 하지만 설명대상을 명확하게 정의하고, 해당 대상이 속하는 범주에 적합한 속성에 기반해서 기술하는 방식은 새롭지 않다. 즉, 활용할 데이터 형태나 공유 방식은 변화가 있다고 할 수 있지만 데이터가 담는 내용과 구성방식은 이전과 다르지 않고, 인간이 이해하기 쉬운 방식으로 인간의 시선으로 만들어진 정보가 인간의 언어로 규칙하에 기술된다는 점에서 한계가 존재한다. 이런 규칙을 우리는 표준이라고 부른다.

이런 점에서 보면 링크드 데이터는 문자로 된 데이터세트의 하나로 볼 수 있고, 그 구축 방식에 있어서 표준을 강조하고 있어 완전히 새로운 데이터 구축방식이라고 보기는 어렵다. 즉, 우리가 이해하기 어려운 완전히 새로운 방식의 데이터 구축은 아니라고 볼 수 있다. 따라서 링크드 데이터를 읽고 이해하는 것은 어렵지 않으며, 양질의 데이터세트를 확보하여 적절히 활용하는 것이 더 중요하며, 링크드 데이터가 갖는 한계가 무엇인지를 명확하게 인지하는 것이 더 중요하다고 할 수 있다.

3.2 링크드 데이터의 한계와 가능성

링크드 데이터는 다양한 학문 분야에 걸쳐 데이터세트 간 접근성과 연결성을 향상시키는 큰 가능성을 제공한다. 데이터를 범용 표준으로 구조화하고 쉽게 탐색 가능한 관계를 통해 연결함으로써, 연구자들이 격리된 데이터세트 내에서는 명백하지 않을 수 있는 새로운 통찰과 상관관계를 발견할 수 있게 한다. 그러나 이러한 통합은 검토할 가치가 있는 본질적인 도전과 한계를 드러낸다.

먼저 생각해 볼 수 있는 것은 이 기술이 연결하는 데이터의 질과 일관성에 대한 의존성이다. 링크드 데이터는 다양한 데이터 소스를 효과적으로 연결할 수 있지만, 이러한 연결의 유용성과 신뢰성은 기반 데이터의 질에 좌우된다. 데이터 형식, 정확성 또는 완성도에서의 불일치는 오해의 소지가 있는 결론이나 불완전한 분석으로 이어질 수 있다. 이는 특히 디지털 인문학에서 문제가 되는데, 문화적 또는 역사적 해석과 같은 해석적 데이터는 종종 주관적 요소를 포함하고 있어 표준화된 데이터 형식으로 원활하게 번역되지 않을 수 있다.

또한, 개방성은 데이터 개인정보 보호 및 보안에 대한 우려를 야기한다. 개방성과 민감한 정보 보호 필요성 사이의 균형은 중요한 문제이다. 디지털 인문학자들은 종종 문화적으로 민감하거나 개인을 식별할 수 있는 데이터를 다루게 된다. 엄격한 통제와 명확한 윤리 지침 없이, 이러한 데이터의 확산은 윤리적 위반 또는 오용으로 이어질 수 있다.

기술적 장벽도 도전 과제이다. 링크드 데이터를 발행하거나 활용하기 위한 도구는 점점 더 사용자 친화적이 되고 있지만, 기술적 전문 지식이 없는 이들에게는 여전히 상당한 지식 격차가 존재한다. 효과적으로 링크드 데이터를 조작하고 질의하기 위해 일정 수준의 프로그래밍 또는 기술 지식이 필요한 것은 이러한 풍부한 연결된 데이터세트에 접근하고자 하는 사람들에게 장벽이 될 수 있다. 링크드 데이터의 확장도 문제다. 더 많은 데이터가 연결됨에 따라, 광범위한 데이터세트에서 질의의 성능과 정확성을 유지하기 위해서는 점점 더 정교한 기술과 인프라가 필요하며, 기술 자원에 대한 상당한 투자 없이는 복잡하거나 매우 큰 데이터세트 활용에 어려움을 겪을 수밖에 없다.

물론, 위에서 제시한 문제들이 모두 링크드 데이터이기 때문에 발생하는 것은 아니다. 링크드 데이터가 아니더라도 데이터베이스, 형태에 상관없이 기계처리 가능한 데이터세트를 구축해서 발

행하는 모든 기관이나 사람은 고민해야 하는 사항들이다. 또한 이렇게 공유된 데이터들이 공통으로 갖는 문제로 볼 수 있다. 사실상 오늘날 모든 데이터, 데이터세트가 갖는 문제로 볼 수 있다.

오히려 인간이 공유를 목적으로 다수의 정보를 정형화하고 표준화시키면서 필연적으로 발생할 수밖에 없는 문제로 보는 것이 정당할 수 있다.

그렇다면 링크드 데이터가 갖는 문제는 해당 기술의 한계라는 관점보다는 표준을 기반으로 데이터를 기술하는 것이 갖는 문제점을 검토하는 것이 바람직할 것이다. 지금까지 살펴본 링크드 데이터의 개념과 특징에 근거하여 본 논문에서는 다섯가지 관점에서 한계와 문제점을 제안하고자 한다.

첫 번째는 복잡성의 축소문제이다. 표준 기반 접근 방식은 인간 경험과 문화적 특징을 사전에 정의된 범주와 용어로 축소한다. 이러한 접근은 원 데이터의 풍부함을 과도하게 단순화하고 왜곡할 수 있으며, 표준화된 프레임워크에 맞지 않는 맥락과 의미를 제거한다. 두 번째는 일반화, 보편화된 프레임워크를 강제하면서 발생하는 문제를 들 수 있다. 데이터 표준의 맥락에서 이는 지역적 또는 맥락적 변화에 관계없이 특정 언어적 및 개념적 프레임워크를 보편적으로 강요할 수 있는 점이다. 이는 지배 문화나 조직이 유효하다고 간주하는 용어와 범주만을 사용해야 하는 문제를 야기시킬 수 있다. 비슷한 관점에서 검토할 수 있는 세 번째 문제는 유연성 부족이다. 표준은 그 본성상 안정적이고 일관성이 있도록 설계된다. 그러나 지식의 동적이고 진화하는 분야를 다룰 때 이것은 한계가 될 수 있다.

네 번째는 권력 역학과 통제 문제이다. 어떤 것이 표준화되는지에 대한 의사 결정 과정은 특정 그룹이 분류 시스템을 정의하고 통제할 권한을 가진 권력 역학을 포함한다. 이는 사회 내 기존의 권력 구조를 강화할 수 있으며, 데이터 표준을 통제하는 사람들이 지식의 조직과 접근 방식을 통제할 수 있음을 의미한다. 마지막은 데이터의 동질화 문제로 표준 사용은 다양하고 이질적인 데이터 세트를 통일된 구조에 맞추어 동질화하는 결과를 낳을 수 있다. 이는 다양한 문화의 데이터를 제한적으로 나타내는 데 제한을 받을 뿐만 아니라 혁신적인 통찰을 생성할 수 있는 잠재력에도 영향을 미칠 수 있다.

물론 표준화가 갖는 장점을 무시할 수는 없다. 링크드 데이터의 경우 다양한 표준용어들을 사용을 허용하고 상호간의 연결도 가능하다. 표준에 존재하지 않는 용어들을 사용할 수도 있다. 하지만 이 경우 상호운용성과 활용성이 낮아지는 것은 감수해야 한다. 오히려 링크드 데이터는 하나의 개체에 대해서 다양한 관점들을 갖는 데이터들이 모아지고 이를 활용할 수 있다는 점에서 지금까지의 문제를 해결할 수 있는 가능성을 갖고 있기도 하다. 또한 그 동안 언어적 차이, 문화적 이질성, 기술적 한계 등으로 접근하기 어려웠던 데이터를 활용할 수 있게 됨으로써 갖는 장점이 더 크다고 할 수 있다.

링크드 데이터가 줄 수 있는 다양한 혜택에 대해서 W3C Library Linked Data Incubator Group은 도서관 링크드 데이터 최종보고서를 발행한 바 있다(W3C, 2011)⁹. 최종 보고서에서는 “도서관(Library)’이라는 용어를 도서관, 박물관, 기록관 등 모든 문화유산 기관을 포함하는 용어로 더 넓게 보고 지식정보자원을 관리하고 서비스하는 기관들이 모두 참조할 수 있도록 작성되었음을 밝힌 바 있다. 보고서에서는 링크드 데이터의 혜택, 전통적 도서관 데이터의 문제, 도서관에서 링크드 데이터 실현을 위한 권고사항을 정리하여 제안하였다. 이중 링크드 데이터가 주는 혜택에 대한 내용을 정리하면 <표 1>과 같다.

표 1. 링크드 데이터의 혜택(출처: W3C, 2011)

대상	혜택
정보 접근자 (연구자, 학생, 이용자)	· 도서관과 외부 기관과의 정보자원 연결로 보다 풍부한 탐색 활동이 가능하게 함 · 이용자와 애플리케이션이 지속적인 탐색활동이 가능하여 도서관의 컬렉션을 웹 상의 보다 큰 정보 세계로 연결해 줄 수 있음 · 웹에 구조화된 데이터가 추가되는 것으로 검색엔진 최적화(SEO)를 통해 향상된 결과 제공이 가능함 · 다양한 분야의 지식베이스에 연결이 가능하여 학제 간 연구에 도움을 줄 수 있음
정보관리 조직 (도서관 등)	· 상향식 데이터 발행으로 자관의 정보자산을 기술하는 일의 가치를 높일 수 있음 · 예산 문제로 상세한 수준의 데이터 기술이 부족한 문제를 극복할 수 있음 · 도서관이 관리시스템 등 주류 솔루션을 벗어나 데이터 연결이 가능한 다수의 시스템 선택이 가능하고, 표준 연결데이터 활용으로 보다 다양한 개발자 그룹과 소통이 가능함 · 클라우드 기반의 데이터 활용 접근을 위한 출발점이 될 수 있으며, 독립형 시스템에 비해 비용대비 효과가 큼 · 작은 기관, 개인들이 적은 비용으로 가시적인 데이터 연결 효과를 볼 수 있음, 즉 공개된 데이터로 누군가 가치 있는 성과를 낼 수 있는 기회를 제공해 줄 수 있음
정보활동 주체 (사서, 기록담당자 등)	· 글로벌 공유 데이터를 통해 편목에서의 중복 노력을 줄이고, 자원 기술의 효율성, 재활용성을 증가시킴 · 타 지식분야의 잘 정리된 기술내용을 활용할 수 있어 사서 등 본인의 전문분야에 대한 기술노력에 집중할 수 있음 · 구문, 형식이 아니라 데이터의 의미를 기술하는 것으로 향후 지속가능한 체계로 데이터관리가 가능함
관련 개발자(사)	· 도서관에 특화된 데이터에 집중할 필요가 없이, HTTP와 같은 웹 프로토콜을 활용함으로써 일관된 방식의 데이터 검색과 결합이 가능함 · LD 방식으로 도서관계를 벗어나 새로운 시장에 활용할 수 있으며, RDF, HTTP를 활용함으로써 특정 소프트웨어 활용의 부담에서 벗어날 수 있음

도서관을 중심으로 기술된 내용이지만 위에 언급한 것처럼 정보를 생성하거나 관리하고 서비스하는 기관으로 확대하여 해석해도 무리가 없다. 연구자를 포함한 이용자들이 얻을 수 있는 가장 큰 혜택은 다양한 분야 간의 정확한 연결이 가능하다는 점이다. 정보자원을 관리하는 기관 입장에서 보면 링크드 데이터를 통해서 예산 한계에 상관없이 풍부한 정보기술이 가능하다는 점과 해당 기관의 데이터를 통해서 가시적인 성과창출 확인이 가능하다는 점이다. 정보활동 주체자인 데이터 관리자의 경우는 데이터에 대한 중복 노력을 방지하면서도 데이터 품질을 높일 수 있는 장점이 있다. 또 연구자 개발자들은 특정 도메인에 특화된 기술적 배경지식 없이도 웹 기술을 활용한 데이터 탐색과 활용, 시장 창출이 가능하다.

이렇듯 링크드 데이터는 표준 준수와 공개된 기술과 플랫폼을 활용한다는 점에서 혜택과 가능성은 물론 고려해야 할 문제를 함께 갖고 있다. 기술이라는 측면에서 갖는 장점을 최대한 활용하되, 표준이 갖는 문제점에 대한 인문학적 고찰도 필요하다 할 수 있다.

4. 결론

본 논문은 웹이 문서 공유 중심에서 데이터 공유 중심으로 전환되는 과정에서 링크드 데이터와 시맨틱 웹의 중요성을 조명하였다. 특히 디지털 인문학에서의 링크드 데이터 활용은 연구 방법을

혁신적으로 변화시킬 잠재력을 가지고 있음을 검토했다. 링크드 데이터는 데이터 간의 명확한 연결을 통해 지식의 웹을 구축하고, 이는 학문 간의 경계를 넘어 새로운 통찰과 해석의 기회를 제공한다. 하지만 이 기술은 데이터의 질과 일관성에 크게 의존하며, 문화적, 역사적 해석을 데이터화하는 과정에서의 주관성이라는 문제를 안고 있다.

앞으로 링크드 데이터의 발전은 기술적인 진보와 함께 데이터의 표준화 및 개방성이 중요한 과제로 남아 있다. 데이터의 개방성과 개인정보 보호의 균형을 찾는 것, 그리고 다양한 데이터 소스의 통합을 통한 새로운 가치 창출이 중요하다. 또한, 이러한 기술들이 제공하는 편리함과 함께 디지털 인문학자들이 직면할 윤리적, 사회적 책임에 대해 깊이 고민하고 대비해야 할 필요가 있다.

이러한 맥락에서 본 논문은 디지털 인문학을 포함한 다양한 분야의 발전을 위해 링크드 데이터가 어떻게 기여할 수 있는지를 간략히 탐구하였다. 링크드 데이터, 시맨틱 웹은 더욱 풍부하고 다양한 학문적 연구, 학제간 연구, 이중 데이터의 결합을 통한 새로운 산업 육성 등이 가능해질 것임을 전망한다. 링크드 데이터의 도입은 단순히 기술적인 변화를 넘어서 학문 연구의 패러다임을 확장하는 새로운 가능성을 열어줄 것이다.

단, 정보화 사회가 인간의 정보문제 해결 방식을 완전히 변화시켰다고 보기는 어렵다. 인간은 여전히 도서관과 같은 장소, 전문가라는 사람, 전문적 지식의 결과물을 활용한다. 단, 인터넷을 통해서 다양한 검색서비스를 통해서 많은 정보를 빠른 시간안에 탐색하고 편리하게 활용할 수 있는 시대에 살고 있다. 물리적 공간이 아닌 가상공간의 정보 보고가 필요하고, 현실의 전문가, 현장에서의 만남 대신 소셜미디어를 통한 연결을 통해 문제를 해결한다. 링크드 데이터는 웹의 표준 기술을 활용해서 인지하고자 하는 대상을 식별하고, 정의하고 대상 기술에 적합하다고 권위자들이 정한 속성에 맞추어 만들어진 데이터이다. 이전에 정보에 대해서 요구하던 권위와 신뢰성을 같은 방식으로 요구하고 있는 것이다. 학문에서, 산업에서, 일상에서 새로운 기술에 대한 이해와 활용은 필요하지만 그 한계와 문제점을 비판적으로 바라볼 수 있는 시선이 안전한 기술의 활용과 발전을 보장할 수 있다.

¹ 박진호(2013). “도서관은 웹에 존재하는가? : 링크드데이터, 글로벌 데이터베이스”. 국가전자도서관 세미나. 경기도: 국가기록원

² European Union. “The origin of Linked Data”. <https://data.europa.eu/en/news-events/news/origin-linked-data>

³ CERN. “World Wide Web born at CERN 25 years ago”. <https://home.cern/news/news/computing/world-wide-web-born-cern-25-years-ago>

⁴ Berners-Lee, T., Hendler, J., & Lassila, O. (2001). “The semantic web”. *Scientific american* 284-5 34-43. <https://doi.org/10.1145/3591366.3591376>

⁵ W3C. “LinkedData”. <https://www.w3.org/wiki/LinkedData>

⁶ DBpedia. “About DBpedia”. <https://www.dbpedia.org/about/>

⁷ Dojchinovski, M., Hernandez, J., Ackermann, M., Kirschenbaum, A., & Hellmann, S. (2018). “Dbpedia nif: Open, large-scale and multilingual knowledge extraction corpus”. arXiv preprint arXiv:1812.10315.

⁸ Hausenblas, Michael. “5 ★ OPEN DATA”. <https://5stardata.info/en/>

⁹ W3C. “Library linked data incubator group final report: W3C incubator group report 25 October 2011.”. <http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>

참고문헌

- 박진호(2013). “도서관은 웹에 존재하는가? : 링크드데이터, 글로벌 데이터베이스”. 국가전자도서관 세미나. 경기도: 국가기록원
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). “The semantic web”. Scientific american 284-5 34-43. <https://doi.org/10.1145/3591366.3591376>
- CERN. “World Wide Web born at CERN 25 years ago”. <https://home.cern/news/news/computing/world-wide-web-born-cern-25-years-ago>
- DBpedia. “About DBpedia”. <https://www.dbpedia.org/about/>
- Dojchinovski, M., Hernandez, J., Ackermann, M., Kirschenbaum, A., & Hellmann, S. (2018). “Dbpedia nif: Open, large-scale and multilingual knowledge extraction corpus”. arXiv preprint arXiv:1812.10315.
- European Union. “The origin of Linked Data”. <https://data.europa.eu/en/news-events/news/origin-linked-data>
- Hausenblas, Michael. “5 ★ OPEN DATA”. <https://5stardata.info/en/>
- W3C. "Library linked data incubator group final report: W3C incubator group report 25 October 2011." . <http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>
- W3C. “LinkedData”. <https://www.w3.org/wiki/LinkedData>