

오픈소스 도구를 이용한 기록정보 링크드 오픈 데이터 구축 절차 연구

A Study on the Procedure for Constructing Linked Open Data of Records Information by Using Open Source Tool

하승록 (Seung Rok Ha)*
임진희 (Jin Hee Yim)**
이해영 (Hae-young Rieh)***

초 록

웹을 통해 제공되는 서비스는 현재 문서중심에서 데이터 중심으로 변화를 겪고 있으며, 그 변화의 중심에는 Linked Open Data(LOD)가 존재한다. 본 연구는 이러한 흐름에 부응하여 기록정보의 LOD 구축을 위한 구체적인 절차와 방법을 살펴보고자 하였다. 또한 소규모 아카이브의 서비스 지속가능성을 염두에 두고, 오픈소스 소프트웨어를 활용하여 LOD 구축 절차를 진행하는 범례를 제시하고자 했다. 이에 본 연구에서는 LOD 구축을 위한 5단계의 프레임워크를 제안했다. 그리고 일상아카이브인 '인간과 기억 아카이브'의 일기 기록물 컬렉션을 수집하여, 제안된 5단계 프레임워크에 따라 오픈소스 소프트웨어인 Protege와 Apache Jena Fuseki를 활용하여 POC(Proof of concept)를 진행하였다. 오픈소스를 활용하여 기록정보의 LOD를 구축한 뒤, 상호연결(Interlinking)과 SPARQL 검색을 통해서 외부 LOD와 연결되는 모습을 확인할 수 있었다. 또한 기록정보의 LOD 구축 절차 진행과정의 경험을 바탕으로 내용정보 기술의 품질 향상, 아키비스트의 역량 고도화, 기록정보의 접근성 향상을 위한 상호연결 고도화, LOD 서비스의 수준 결정, LOD 구축을 위한 도구 선정 등, 기록관 LOD 구축을 위한 필요요건을 제시했다.

ABSTRACT

Recently, the web service environment has changed from document-centered to data-oriented focus, and the Linked Open Data (LOD) exists at the core of the new environment. Specific procedures and methods were examined to build the LOD of records information in accordance with this trend. With the service sustainability of small-scale archive in consideration, an exemplification on LOD building process by utilizing open source software was developed in this paper. To this end, a 5-step service framework for LOD construction was proposed and applied to a collection of diary records from 'Human and Memory Archive'. Proof of Concept (POC) utilizing open source softwares, Protege and Apache Jena Fuseki, was conducted according to the proposed 5 step framework. After establishing the LOD of record information by utilizing the open source software, the connection with external LOD through interlinking and SPARQL search has been successfully performed. In addition, archives' considerations for LOD construction, including improvement on the quality of content information, the role of the archivist, were suggested based on the understanding obtained through the LOD construction process of records information.

키워드: 기록정보, 링크드 오픈 데이터, 온톨로지, 오픈소스 도구, Protege, Apache Jena records information, Linked Open Data, LOD, ontology, open source tool, Protege, Apache Jena

* 명지대학교 기록정보과학전문대학원 기록관리전공(gktmdfhr@gmail.com) (제1저자)

** 정보인권연구소 연구위원(yimjhkr@empas.com) (교신저자)

*** 명지대학교 기록정보과학전문대학원 교수(hyrieh@mju.ac.kr) (공동저자)

■ 논문접수일자: 2017년 2월 19일 ■ 최초심사일자: 2017년 3월 8일 ■ 게재확정일자: 2017년 3월 27일
■ 정보관리학회지, 34(1), 341-371, 2017. [http://dx.doi.org/10.3743/KOSIM.2017.34.1.341]

1. 머리말

1.1 연구배경 및 목적

2009년 웹의 창시자 Tim Berners-Lee는 TED 강연에서 차세대 웹을 언급하며 웹의 발전가능성을 이야기하였다. 그는 웹에 있는 현재의 '문서'들이 '데이터'로 바뀌고, 그 '데이터'들이 모두 개방되고 공유된다면, 강력하고 유의미한 변화가 일어날 것이라 말했다(Berners-Lee, 2009). 이전 웹에서의 검색은 대부분 용어 매치를 통해 이루어졌고, 검색하는 사람의 의도를 반영하지 못했다. Berners-Lee는 시맨틱 웹의 개념을 처음으로 제시하면서(Berners-Lee, Hendler, & Lassila, 2001), 이용자의 검색의도를 복합적으로 반영하는 검색의 필요성을 주장하였다. 이후 시맨틱 웹의 구현을 위한 연구들이 지속적으로 진행되어 왔으며, 시맨틱 웹을 구현하기 위한 방법으로 온톨로지 언어인 RDF(Resource Description Framework)와 OWL(Web Ontology Language)이 개발되었다. 그리고 2009년 Berners-Lee는 시맨틱 웹을 실현하기 위한 방법으로 링크드 오픈 데이터(Linked Open Data: LOD)의 개념을 제시하였고, 이것은 차세대 웹이 지향하는 흐름이 되었다.

최근 국내 중앙 정부와 지자체, 민간분야에서도 이러한 흐름에 맞물려 LOD 도입을 위한 시도를 진행하고 있다. 그 사례로 국사편찬위원회는 2013년부터 LOD기반의 한국사 통합DB를 구축하여 한국사LOD서비스를 제공하고 있다. 이 서비스는 역사 교과서 및 한국사와 관련된 주요 주제어를 선정하여 원문, 이미지, 해설 등의 연관정보를 즉각적으로 연계하고 시각화된 화면으로

제공한다. 기록학계에서도 기록물의 공유와 개방 그리고 더 나은 검색을 위한 온톨로지 구축과 시맨틱 웹에 대한 연구들이 진행되고 있다.

그러나 현실적으로 아카이브의 입장에서는 LOD 구축 사업의 예산을 마련하기 어려운 경우가 많다. 또한 LOD 구축을 통해 기록정보의 확장과 공유가 이뤄졌을 때 아카이브의 입장에서 얻을 수 있는 이득이 그리 크지 않고, 기존의 기록물 정리 및 서비스만으로도 충분하다고 생각할 수 있다. 하지만 기록정보는 다른 정보와는 다르게 한 곳의 기록관에만 고유하게 존재하는 경우가 많다. 따라서 개별 기록관을 찾아가야 그 기록관에서 소장하고 있는 기록을 알 수 있는 상황이 대부분이다. 그러므로 많은 기록관들이 기록의 통합서비스가 가능하도록 발전해야 한다는 측면에서, LOD 구축을 통한 기록정보의 확장과 공유는 큰 방향성이 될 것이다. 현재 웹의 발전도 서로 다른 사이트 및 기관의 데이터들이 벽(Silo 현상)을 깨고 나와 공유되는 방향으로 진행되고 있다. 또한 구글이나 네이버 등 대부분의 검색엔진과 포털 사이트도 한 플랫폼에서 다양한 사이트의 정보와 기능을 활용하는 방향을 지향하고 있다. 앞으로 아카이브의 기록정보도 더 이상 아카이브의 데이터베이스 안에만 고립되어 있지 말고, LOD라는 도구를 통해 다양한 장소와 사이트에서 자유롭게 공유되고 연결되어야 한다.

기존의 연구들이 LOD 구축을 통해 얻을 수 있는 효용성에 초점을 두었다면, 본 연구의 목적은 그 이면에 있는 구축의 구체적인 절차와 방법을 살펴보는 것에 있다. 그리고 그 과정에서 도출되는 사항들을 확인하고, 그를 통해 LOD 구축을 위해 필요한 요건을 제시하고자 한다.

이를 위해 먼저 LOD의 기반인 온톨로지를 구체적으로 어떤 방법론에 따라 구현해야 할 것인지 제시하고, 나아가 LOD 구축을 위한 프레임워크를 제시하고자 한다. 그리고 실제 기록정보 데이터셋을 수집하고, 오픈소스 소프트웨어를 활용하여, 본 연구에서 제안한 LOD 구축 프레임워크에 따라 기록정보 LOD 발행을 진행하고자 한다. 그리고 이러한 POC(Proof Of Concept) 진행과정의 경험을 바탕으로 기록관 LOD 구축의 필요 요건들을 제시하고자 한다. 본 연구에서 제시된 LOD 구축 절차 연구와 기록관 LOD 구축의 필요 요건에 대한 논의는 아키비스트의 기록정보 LOD 구축에 대한 구체적인 이해를 도울 것으로 기대된다.

1.2 선행연구

초기 온톨로지에 대한 연구는 온톨로지 언어에 대한 소개 및 비교 연구들이 주를 이루다가, 온톨로지의 발전에 따라 구체적으로 온톨로지를 활용하기 위한 시스템의 구축방법론으로 연구흐름이 이어져왔다(이병길, 김희섭, 2013). 온톨로지에 관한 기록학계의 선행연구들도 온톨로지 모델링 및 구축을 통해 검색시스템의 재현율과 정확률의 향상을 위한 연구방향이 주를 이루었다. 그러나 기록학계의 온톨로지 관련 연구는 그 수가 매우 적고, 연구방향도 기록정보 검색시스템의 질적 향상과 기록정보서비스의 변화에 초점이 맞춰져 있었다. 최근에 와서야 기록정보의 개방과 공유를 위한 링크드 데이터(Linked Data)에 대한 논의까지 연구영역이 확장되고 있다.

Noy와 McGuinness(2001)는 온톨로지가 무

엇인지, 왜 구축되어야 하는지, 어떻게 구축해야 하는지에 대한 핵심개념과 구축방법론을 제시했다. 이 연구에서 제시된 7단계 방법론은 처음 온톨로지를 접하고, 구축하려는 개발자에게 도움이 되는 지침들을 설명하고 있다.

박옥남(2012)은 국가기록원의 전거데이터 셋을 링크드 데이터화하는 연구를 진행하였다. 국가기록원은 다양하고 많은 기록물을 보유하고 있으므로 기록물의 활용도를 높이기 위해서 국가기록원이 가진 데이터의 개방 및 연계가 필요하다고 주장했다. 이를 위해 국가기록원 전거데이터의 구조 및 검색시스템을 분석하고, 전거데이터 세트의 온톨로지를 설계하여 RDF/OWL로 구축하였다. 이 연구는 기록물의 개방과 공유를 위해 온톨로지를 설계하고 링크드 데이터로 구축하는 방법을 연구한 것에 의의가 있다.

이병길, 김희섭(2013)은 온톨로지 기반의 검색 시스템을 개발하고 그 성능을 평가하였다. 현행 기록관리 프로세스에 따라 체계적으로 관리되고 있는 기록물인 '새마을운동' 컬렉션을 수집하고, 데이터를 분석 및 분류한 뒤, Protege를 사용하여 온톨로지 검색시스템을 구현하였다. 이 연구는 수집한 기록물의 메타데이터 요소를 기반으로 온톨로지에 반영할 요소를 선정하는 과정에서, 기존 기록물의 메타데이터 항목 중에서 기록물의 내용 항목이 빈약하다는 것에 주목하고, 온톨로지에 내용 항목을 추가로 반영한 점에 의의가 있다.

이현정(2015)은 국내의 도서관 분야의 링크드 데이터 발행 현황과 필요성에 연구했다. 그리고 공개된 링크드 데이터의 SPARQL Endpoint 주소를 통해, 범용 쿼리를 적용한 멀티소스 검색 시스템을 구현했다. 이 연구는 SPARQL 검색

을 모르는 일반인도 링크드 데이터에 접근하여 검색할 수 있도록 하는 기초를 제공하는 것에 의의가 있다.

이정현, 이윤용, 방기영, 김용(2015)은 노무현 대통령 재임기간에 생성된 전자문서와 웹 기록을 대상으로 온톨로지 설계 및 RDF 변환 작업을 진행하고, 기록정보 서비스의 확장 프로세스와 인터페이스를 제시했다.

이유빈, 이해영(2017)은 시맨틱 웹의 사례를 찾아 검색 방법, 카테고리 및 페싱, 검색 결과 탐색 및 재검색, 상세정보 제시 등의 측면에서 특징을 분석하고 사용자 평가를 진행했다. 그리고 그 결과를 바탕으로 기록물 검색 시스템을 위한 인터페이스를 제안했다.

1.3 연구범위 및 방법

일반적으로 데이터의 LOD 구축은 기업을 통해서 진행되는데 작은 아카이브의 경우 세 가지 문제점이 존재한다. 첫째, 큰 규모의 아카이브나 기관과는 다르게 작은 아카이브는 LOD 구축 사업을 위한 예산 확보가 어려운 경우가 대다수이다. 둘째, 소장한 기록물에 대하여 완전히 디지털화 하지 못했거나 디지털 자료에 대한 기술이 온전히 진행되지 못한 부분들이 있어, 필요한 경우 기록물의 기술에 대한 보완을 해가며 LOD 구축을 진행해야 할 수도 있다. 셋째, 기록의 아이템 혹은 시리즈가 가지는 의미적인 내용에 대한 평가를 가장 정확하게 내릴 수 있는 존재는 결국 아키비스트이다. LOD 구축을 담당하는 직원이 기록물의 특성에 대한 설명을 듣는다하여도 아키비스트가 직접 구축하는 것이 품질을 높이는 방안이 될 것이다.

따라서 이러한 문제점들을 극복하기 위해 아키비스트가 LOD 구축 사업을 이끌어간다는 전제하에 본 연구를 다음과 같이 진행하였다. 먼저 인간과 기억 아카이브에서 수집한 “5월12일 일기수집” 기록 컬렉션을 연구에 활용할 수 있도록 동의를 구하고 접근권한을 얻었다. 인간과 기억 아카이브는 민간 분야의 기록관리 활성화를 위해 2013년에 한국국가기록연구원이 주축이 되어 설립한 아카이브이다. 5월12일 일기수집 컬렉션은 영국의 University of Sussex의 Mass Observation 프로젝트에서 1937년 5월 12일부터 일반인들의 일기들을 받아 아카이빙한 것을 우리나라도 같이 협업하기로 하여, 5년째 진행되고 있는 수집컬렉션으로 기록의 공유가 의미 있다. 또한 일기 기록은 그 내용이 한 시대의 의식주와 다양한 영역의 생활사 연구를 위한 사료로 기능하기 때문에(최효진, 임진희, 2015) 행정문서 위주의 공공기록보다 비교적 기록의 내용이 중요한 의미를 가진다. 그리고 현재 인간과 기억 아카이브의 기록컬렉션은 공공기록관리시스템이 아니라 ICA에서 제공하는 오픈소스 기록관리시스템인 AtoM으로 관리되고 있고, 오픈소스 발행 플랫폼인 Omeka를 통해 전시되고 있다. 따라서 공유를 위한 기초가 되어 있는 데이터 세트라 판단하여, 연구의 대상으로 선정하였다.

둘째, 데이터 모델링을 위한 온톨로지 구축방법론과 LOD 구축방법론을 구성하였다. 온톨로지 개발을 위한 방법론으로는 스탠포드 대학교에서 만들어진 지침서인 “ontology development 101”을 본 연구의 목적에 맞게 수정하여 온톨로지 구축 방법론으로 활용하였다. 이 절차는 다음 장에서 상세히 설명한다. 그리고 이러한 7단

계 온톨로지 구축 방법론을 중심으로, Bauer와 Kaltenböck(2011)가 주장한 LOD 구축을 위한 8단계와 한국정보화진흥원(2014b)이 제시한 LOD 구축 방법론을 참고하여, LOD를 구축하고 발행하는 5단계의 구축 절차를 구성하였다. 이 절차는 다음 장에서 상세히 설명한다.

셋째, 아카이브에서 LOD 구축 사업에 예산을 편성해야 하는 부담 없이, 아키비스트가 직접 LOD를 구축하여 공개할 수 있도록, LOD 구축과정에서 활용할 오픈소스 소프트웨어를 선정하였다. 본 연구에서는 온톨로지 저작도구로 Protege를 소개하고, RDF파일을 저장하고 웹에 서비스할 수 있는 기반을 조성하는 RDF 저장소로 Apache Jena Fuseki를 활용하였다.

넷째, 본 연구에서 활용하기로 한 5단계의 LOD 구축 절차에 따라, 각 단계를 오픈소스 도구를 활용하여, 수집한 일기 기록물을 LOD로 구축하였다. 단, 본 연구는 내부 웹 서버인 Localhost 환경에서 연구를 진행하여, 실제로 외부 웹 서버를 통해 LOD를 서비스하는 단계까지는 실행하지 못한 한계점이 존재한다.

다섯째, 오픈소스 소프트웨어를 활용하여 기록정보의 온톨로지 구축과 LOD 구축에 대한 POC (Proof Of Concept)를 진행하면서 얻은 경험을 통해 아카이브가 기록정보의 LOD를 구축할 때 필요로 하는 요건을 확인하였다.

2. 온톨로지와 LOD의 구축 방법

2.1 온톨로지

온톨로지(Ontology)는 본래 철학의 한 분야

로, 존재의 본질에 관한 학문이다. 1990년대 초부터 온톨로지는 컴퓨터 분야에서 사용되었고, 공유된 개념화에 대한 형식적이고 명시적인 명세라고 정의되었다(Gruber, 1993).

온톨로지 언어 중 가장 대표적인 것은 1999년 W3C 권고안으로 채택된 RDF와 2004년 W3C 권고안으로 채택된 RDF Schema, OWL(Web Ontology Language) 등이 있다. 그 중 OWL은 다른 온톨로지 마크업 언어들보다 표현력이나 추론능력이 뛰어나다는 평가를 받고, 가장 널리 사용된다. OWL은 온톨로지 헤더(Header), 클래스(Class), 속성(Property), 객체(Individual) 등 4개의 구성요소를 가진다(伊藤健太郎, 佐藤勇紀, 濱崎俊, 2007). 온톨로지 헤더는 OWL 온톨로지 문서 전체의 정보를 기술하는 요소이고(Antoniou & Harmelen, 2008), 클래스는 비슷한 속성을 지니고 있기 때문에 하나의 군으로 모아지는 객체들을 의미한다(박현근, 2005). 속성은 객체간의 관계 혹은 객체와 데이터 값의 관계를 표현하는 어휘다. OWL속성은 3가지로 나뉘는데, 객체와 객체를 연결하는 Object Property, 객체와 데이터 값을 연결하는 Data Property, 그리고 참조를 위한 Annotation Property가 있다.

OWL에서는 컴퓨터가 이해할 수 있도록 특별한 의미를 지닌 속성을 나타내는 어휘들이 존재한다. 속성의 역 관계를 정의하는 역 속성(Inverse of), 속성이 이행적이라는 것을 나타내는 이행속성(Transitive Property), 속성의 정의역(Domain)과 공역(Range)이 바뀌어도 의미가 다르지 않다는 대칭속성(Symmetric Property), 정의역이 단 하나의 공역과 연결되는 함수속성(Functional Property), 함수속성의 반대개념

으로 하나의 정의역이 여러 개의 공역을 가질 수 있으나 공역은 단 하나의 정의역만 가질 수 있는 역함수속성(Inverse Functional Property)이 있다(노상규, 박진수, 2007). OWL은 이렇게 특별한 속성을 통해서, 컴퓨터가 직접 명시되지 않은 내용까지 추론할 수 있도록 설정한다.

온톨로지 구축방법은 이론적, 경험적으로 다양하게 제안되어 왔다. 대표적으로 Skeletal Methodology, Methontology, ONIONS, CommonKADS, Ontology Development 101, On-To-Knowledge Methodology와 같은 방법론들이 존재한다(김은경, 2004).

Ontology Development 101은 온톨로지 구축의 필요성과 개념설명을 하고, 와인과 음식 분야를 예시로 들어 온톨로지 개발 방법론을 소개하기 때문에 처음 온톨로지를 개발하는 사람도 이해하기가 쉬운 구축방법론이다. 따라서 본 연구는 온톨로지 개발방법론으로 Ontology Development 101를 사용하였다.

Ontology Development 101에서는 온톨로지 개발지침을 7단계로 제시하고 있다(Noy & McGuinness, 2001). 하지만 현재 Protege의 구동환경과 맞지 않는 부분이 있어, 그 부분을 수정하여 연구를 진행하였다. 본 연구에서 수정하여 활용한 온톨로지 구축의 일곱 단계는 다음과 같다.

첫째, 온톨로지의 분야와 범위를 결정한다. 그러기 위해서는 온톨로지가 포함하려는 범위, 온톨로지의 사용목적, 온톨로지가 제공하는 정보의 종류, 온톨로지의 이용자 등에 대한 질문이 필요하다.

둘째, 존재하는 온톨로지의 재사용을 고려한

다. 특정 주제 영역의 온톨로지 구축 시, 관련 영역에 대한 온톨로지가 구축되어 있는지 조사한다. 온톨로지 재사용은 온톨로지 구축시간의 절감, 공용된 개념 및 어휘 사용으로 인한 상호 운용성 확보 및 개념의 명확화가 가능하다는 장점이 있다.

셋째, 온톨로지에서 중요한 단어들을 열거한다. 이 단계에서는 우선 개념의 중복, 단어사이의 관계, 개념의 분류(클래스인지 속성인지) 등을 고려하지 않고 온톨로지에서 표현하고 싶은 모든 단어들을 열거하는 것이 중요하다.

넷째, 클래스들 간의 계층구조를 정의한다. 의미적으로 같은 범위에 속하는 단어를 하나의 개념으로 정의하여 클래스를 생성하고, 클래스들 간의 관계를 정리하여 계층구조를 정의한다. 만약 클래스A가 클래스B의 상위 클래스라면, B의 모든 객체들도 A의 객체가 된다.

다섯째, 클래스의 속성들을 정의한다. 클래스 혼자서는 개념을 설명할 수 없다. 따라서 클래스를 정의하고 난 뒤에는 개념의 내부구조를 기술할 수 있는 설명을 해야 한다. 클래스의 속성은 하위클래스로 상속되므로, 가장 일반적인 클래스에 속성을 연결해야 한다.

여섯째, 속성의 패킷(제약조건)을 정의한다. 속성의 관계 수(cardinality)를 정의하고, 속성의 유형(문자, 숫자, Boolean 등)을 정의하며, 속성의 정의역과 공역을 설정한다.

일곱째, 객체를 생성한다. 이 단계에서는 클래스를 선택하여 객체를 생성하고, 속성을 채워 넣는다. 그렇지만 보통 일반적인 LOD 구축에서는 만들어진 온톨로지 스키마에 따라 원본데이터에서 자동으로 객체를 입력할 수 있도록 프로그램 사용한다.

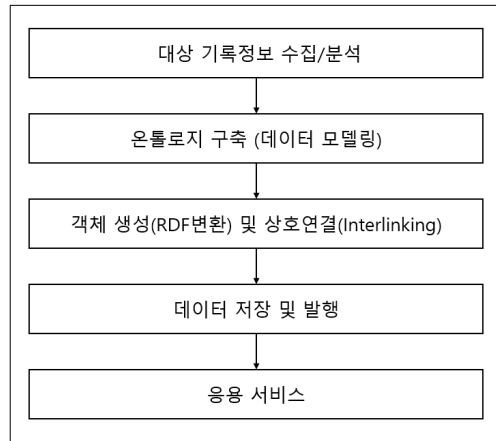
2.2 Linked Open Data(LOD)

LOD는 Linked Data(연결된 데이터)와 Open Data(개방된 데이터)의 핵심개념을 포괄하는 개념으로 링크드 데이터 발행원칙에 맞춰 데이터를 개방하는 것을 의미한다(한국정보화진흥원, 2014a). Berners-Lee(2006)는 링크드 데이터의 4가지 원칙을 제시함으로써 링크드 데이터의 개념을 설명하였다. 그가 말하는 링크드 데이터란 웹상의 데이터에 URI(Uniform Resource Identifier)를 부여해 이름을 정하고, RDF로 데이터를 기술할 뿐만 아니라, 데이터 간의 관계를 부여하고, HTTP를 통해 웹에 발행함으로써 기계가 데이터를 이해하고 처리할 수 있도록 하는 방법이다(김성혁, 2011).

Bauer와 Kaltenböck(2011)는 Berners-Lee가 주장하는 LOD를 발행하기 위해서 다음과 같은 8단계를 거쳐야 한다고 제안했다. 그 8단계는 데이터를 분석하고, 데이터를 정제하며, 데이터 모델을 설계하고, 어휘선정하고, 라이선스를 결정하며, 데이터를 RDF로 변환(RDF화)하고, 외부와 내부(본인)의 데이터를 연계하며, 데이터를 발행하는 것이다. 한국정보화진흥원(2014b)은 정보자원의 효율적 활용을 위한 데이터 웹 구축을 위해 웹에 공개 가능한 데이터를 발굴하고, 링크드 데이터로 변환하는 활동을 지원하기 위해 표준화된 구축절차를 준비단계, 구축 및 발행 단계로 나누고 각 세부사항에 대한 자세한 절차들을 가이드로 제시한다.

본 연구에서는 Bauer와 Kaltenböck(2011)가 주장한 LOD 구축을 위한 8단계와 한국정보화진흥원에서 발행한 가이드에서 제시하는 준비단계와 구축 및 발행절차를 참고하고, 앞에서 설

명한 7단계의 온톨로지 구축 방법론을 중심으로 하여, 총 5단계의 기록정보 LOD 구축 프레임워크를 구성하여 활용하였다(〈그림 1〉 참조).



〈그림 1〉 제안된 5단계 LOD 구축 프레임워크

첫째, 대상 기록정보 수집 및 분석단계다. 이 단계에서는 LOD로 구축하려는 대상 기록에 대한 상세한 조사를 진행한다. 구축 대상인 기록의 형태 및 포맷을 조사하고, 사용가능한 메타데이터의 특징들을 조사하며, 기록에 대한 저작권 등을 확인한다. 또한 구축 및 연계 대상의 건수 등을 확인하고, 기록 반출을 위한 계획을 수립하여 원천 데이터를 입수한다(한국정보화진흥원, 2014b). 이 단계에서 중요한 점은 기존 기록물 기술과는 다르게 아키비스트가 기록물의 내용까지 파악하여, 주제나 장소, 인물과 같은 전거 데이터를 기술해야 한다는 점이다. LOD환경에서 기록물이 다른 LOD와 많은 연계가 가능해지기 위해서는 기록물의 구조와 맥락을 담은 메타데이터뿐만 아니라 내용에 대한 메타데이터도 풍부해져야 한다.

둘째, 온톨로지 구축 단계이다. 이 단계는 수집

한 기록물을 가지고 데이터 모델링을 하는 단계로서, 앞서 설명한 스탠포드 대학교의 Ontology Development 101과 같은 7단계의 온톨로지 구축방법론을 활용하여 기록물의 명세화와 용어 설계 등을 진행하고, Protege와 같은 온톨로지 저작도구를 사용하여 온톨로지를 구축한다.

셋째, 객체 생성(RDF 변환) 및 상호연결(Interlinking) 단계이다. 이 단계에서는 구축한 온톨로지(스키마)와 매핑 규칙 등을 통해서 기록물의 데이터를 RDF화하는 작업을 진행한다. 데이터를 RDF화하는 방법은 보통 두 가지가 있는데, 정적인 구조적 데이터를 편집도구를 이용하여 변환하는 방법과 관계형데이터베이스의 데이터를 RDF로 변환(RDB-to-RDF)하는 방법이 있다. CSV나 엑셀, XML 파일과 같은 정적인 구조적 데이터 형태의 기록물들은 Protege의 플러그인 기능인 Cellfie와 ConvertToRdf, RDFizer, RDF Refine, Ontotrans, Virtuoso와 같은 편집 도구를 이용할 수 있다. 그리고 RDB(관계형 데이터베이스)형태로 반입한 기록물도 마찬가지로 매핑 도구를 이용하고 편집도구를 사용하여 RDF로 변환하는데, RDB-to-RDF 방법은 기존에 만들어진 데이터베이스의 스키마에 따라 자체적으로 변환되기 때문에, 데이터베이스 스키마가 정확하게 설계되어야 한다(한상은, 2014). 이렇게 데이터를 RDF 객체(Individual)로 변환하여 온톨로지에 올린 다음에는, 각 객체의 관계 확인 및 수정작업을 진행해야 한다. 이 과정에서 의미의 연결을 위해 다른 LOD의 데이터와 상호 참조할 수 있도록 상호연결(Interlinking)하는 작업이 수행되어야 한다.

넷째, 데이터 저장 및 발행 단계이다. 이 단계

는 RDF로 변환된 데이터를 저장하고 발행하는 단계이다. 보통 RDF 저장소를 사용하여 RDF 데이터를 저장한다. RDF 저장소란 도메인 지식을 정의한 온톨로지를 저장하고 관리하며, RDF 데이터를 쿼리할 수 있는 인터페이스와 추론기능 등을 제공하는 저장소(repository)이다. RDF 저장소는 시맨틱 웹 분야에서 가장 많이 성장한 산업 부분이고, JAVA와 Oracle 등이 시장에 진입하면서 산업이 더욱 성장하고 있다(안세열, 2011).

마지막으로 응용서비스 단계이다. 이 단계는 구축된 LOD를 단순히 다운로드하여 사용하는 것이 아니라, 시각화와 검색 등 응용서비스를 지원하는 부분이다. 이 단계에서는 소프트웨어를 사용하여 확장 응용서비스가 가능한데, 대표적으로 LOD 검색을 가능하게 하는 SPARQL Endpoint, Pubby 등이 있고, 데이터를 시각화시켜주는 LODLive, RelFinder 등이 있다. 하지만 이 단계는 프로그래밍 지식이 필요하므로, 소규모 아카이브가 자력으로 구축하기는 어려운 한계점이 있다. 따라서 자관의 정보지원팀과 협의가 필요하다.

LOD를 구축한다는 것은 데이터를 웹에 공개하여 의미를 공유한다는 뜻이고, 그것은 URI를 통한 데이터 접근 및 이용이 가능한 환경에서 데이터의 의미를 공유하고 재사용할 수 있다는 것을 의미한다(김영천, 2011). 이러한 데이터의 재사용이 효과적으로 이뤄지기 위해선 표준용어집으로 사용가능한 데이터세트가 존재해야 한다. 대표적으로 위키피디아와 DBpedia, 더블린코어(DC), DCTERMS, SKOS, FOAF 등이 표준 데이터 세트 및 표준어휘로 사용된다. DC와 DCTERMS는 링크드 데이터 환경에

서 가장 대표적으로 사용하는 어휘로써, 본 연구에서도 많은 속성들에 DC와 DCTERMS의 어휘를 사용하여 온톨로지의 상호운용성을 확보하였다. SKOS는 W3C에서 시맨틱 웹 기반의 지식 구조화 프레임워크 표준으로 개발한 것으로, SKOS를 이용하면 지식 표현 체계를 시맨틱 웹으로 표현하는 것이 가능하다(박옥남, 2012). 본 온톨로지는 SKOS를 사용하여 주제에 대한 개념을 표현하고자 했다. FOAF는 사람과 사람 사이의 관계를 명세하는 어휘로, 본 연구에서는 인물에 대한 객체를 표현하기 위해 사용했다.

2.3 구축도구

온톨로지 저작도구는 TopQuadrant사의 Topbraid와 스탠포드 대학교의 Protege가 대표적이다. 본 연구에서는 오픈소스 소프트웨어를 사용하여 링크드 데이터를 구축하고자, 온톨로지 저작도구로 Protege를 선택했다.

Protege는 스탠포드 대학교에서 온톨로지 구축을 위해 개발한 도구로, 자바(JAVA)기반의 오픈소스 기반의 편집기이다. Protege는 RDF/XML, OWL/XML, Turtle 등 다양한 언어로 온톨로지 파일 작성이 가능하고, 클래스와 속성의 정의 및 관계 설정, 속성의 제약 등 OWL의 기본 구성요소에 대한 편집기능을 갖추고 있다. 그 외에도 시각화나 추론, 데이터 삽입과 관련하여 다양한 종류의 기능을 플러그인 형태로 사용할 수 있도록 제공하고 있다(박중호, 2011). Protege는 1999년 Protege-2000이 개발된 이후, 2017년 현재 5.1버전까지 개발되었고, 최근에는 Webprotege라고 하여 웹사이트를 통해 온톨로지를 쉽게 생성하고 업로드하며, 수정하

고 공유할 수 있으며, 친숙한 인터페이스를 가진 개발환경이 구축되었다. Webprotege의 가장 큰 장점은 하나의 프로젝트에 여러 명의 사용자가 협업하여 온톨로지를 구축해나갈 수 있는 점이다. 프로젝트 내에서 메모 및 토론, 코멘트, 수정 이력 등 다양한 기능을 통해서 효율적인 공동 작업이 가능해졌다. 하지만 이전의 Protege에서 제공하던 강력한 플러그인 기능들은 사용할 수 없다는 한계점이 존재한다.

본 연구에서는 Protege 5.0 버전을 사용하여 온톨로지를 구축하였다. Protege 5 버전부터는 엑셀 스프레드시트를 매핑규칙을 통하여 해당 온톨로지의 객체(Individual)로 생성해주는 플러그인 소프트웨어인 Cellfie를 기본적으로 지원한다. 이전 버전에서 Cellfie의 이름은 Mapping Master이었는데, 이 플러그인은 엑셀 스프레드시트 안의 데이터를 맨체스터 구문(Manchester Syntax)을 기반으로 한 매핑규칙(변환규칙)을 통해서 Protege로 작성중인 온톨로지 안에 클래스와 객체로 집어넣는 기능을 한다. Cellfie 이외에도 구조화된 데이터를 매핑규칙을 통해 RDF 파일로 변환시켜주는 소프트웨어로 OpenRefine, D2R Server, ConvertToRdf, RDFizer, Ontotrans 등이 있다. 하지만 본 연구에서 사용하는 대상 기록의 기술정보는 엑셀 형식으로 AtoM 기록 관리시스템에서 추출했고, Protege라는 하나의 프로그램 안에서 온톨로지 구축과 객체생성 및 상호연결까지 마칠 수 있다는 점, 그리고 프로그램 간의 종속성 등을 고려하여 Cellfie를 연구에서 사용하였다.

RDF 데이터를 저장하는 RDF 저장소로는 Apache Jena, Virtuoso, AllegroGraph, OWLIM, Ontobase 등이 있다. 본 연구에서는 대표적인

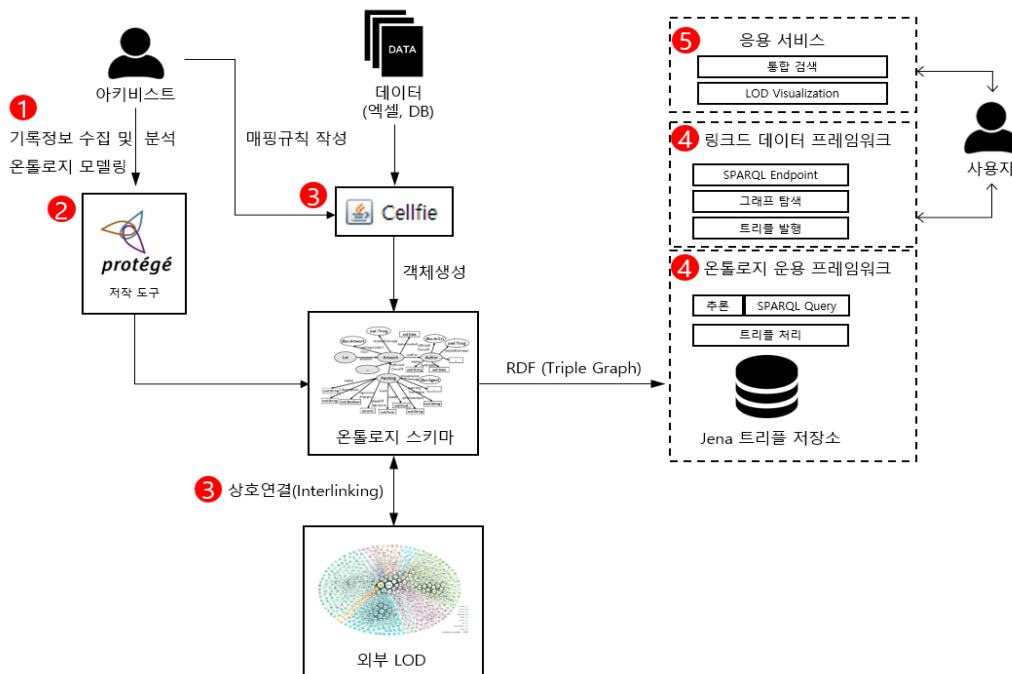
RDF 저장소인 Apache Jena Fuseki를 사용하여 발행한 링크드 데이터를 활용하였다. Apache Jena는 JAVA기반의 오픈소스 소프트웨어로, RDFS와 SPARQL 처리를 위해 개발된 프레임워크이다(정갑주, 2015). 이 프레임워크는 하나의 어플리케이션을 일컫는 것이 아니라, 저장 API와 추론 API, RDF API와 온톨로지 API, SPARQL API를 통틀어 일컫는 개념이다. 본 연구에서는 저장 API인 TDB나 SDB를 설치하지 않고, Jena Fuseki 서버만 가동하여 구축한 Linked Data에 대한 SPARQL 질의를 실행해보았다. Fuseki 서버는 HTTP 프로토콜을 통하여 웹 서비스와 SPARQL Endpoint를 제공하는 도구이다. Fuseki 서버는 Jena의 다른 SPARQL 도구인 ARQ와 다르게 웹 브라우저를 통해 이용자친화적인 환경에서 RDF 데이터

의 업로드와 수정, 데이터세트 관리, SPARQL 질의 등을 제공한다.

3. 오픈소스 도구를 이용한 LOD 구축 사례

3.1 대상 기록의 특징

본 연구는 기록정보의 온톨로지를 구축하고 LOD로 발행하는 연구를 진행하기 위해 접근을 허락받은 인간과 기억 아카이브의 “5월12일 일기컬렉션”을 수집하고, 앞서 제안한 5단계 LOD 구축 절차에 따라 오픈소스 소프트웨어를 활용하여 LOD를 구축해보는 POC(Proof of concept)를 진행하였다(<그림 2> 참조).



<그림 2> LOD 구축 프레임워크(한국정보화진흥원 지식자원활용부, 2014a, p. 28의 그림 재구성)

5월12일 일기 컬렉션은 다양한 연령, 지역, 직업의 시민들이 매년 5월12일 하루 동안 경험한 일이나 스쳐간 생각과 느낌을 글, 사진, 동영상, 음성으로 기록하여 그 결과물을 ‘인간과 기억 아카이브’에서 보관하고 있는 기록 컬렉션이다(최효진, 임진희, 2015). 이 컬렉션은 ICA에서 오픈소스 기록관리시스템으로 제공하고 있는 AtoM을 통해 관리되고 있고, 오픈소스 웹 발행 플랫폼인 Omeka를 통하여 전시하고 있다.

AtoM은 기록을 효율적으로 기술하고 열람, 검색하기 위해 ISAD(G), ISAAR(CPF), ISDIAH, ISDF와 같은 ICA 기술표준과 DACS와 RAD 등의 국가표준 그리고 더블링크어, MODS 등 다양한 메타데이터 요소를 제공한다(안대진, 김익한, 2015). 5월12일 일기컬렉션이 가지고 있는 내용정보, 구조정보, 맥락정보는 기술요소로 정리되어 있으며, 5월12일 일기컬렉션에서 주로 사용되는 기술 요소는 <표 1>과 같다.

<표 1> 5월12일 일기기록 기술 요소

식별영역	참조코드
	제목
	일자
	기술계층
	규모 및 매체
맥락영역	생산자 이름
	기록보존소
내용과구조영역	범위와 내용
	정리시스템
접근점	주제 접근점
	이름 접근점
기술통제영역	아키비스트 주기
주기사항	주기
디지털 객체	유형

구축한 온톨로지(스키마)와 매핑 규칙 등을

통해서 기록물의 데이터를 RDF화하는 작업을 진행하기 위해서는 기록물 기술 데이터가 엑셀과 같은 구조적 데이터 형식이나 관계형데이터베이스 형식으로 존재해야 한다. 인간과 기억 아카이브에서 사용하는 소프트웨어인 AtoM과 Omeka는 XML과 EAD(Encoded Archival Description) 형식으로 기록정보를 내보내기(export)할 수 있는 기능을 갖추고 있다. 하지만 Omeka의 경우에는 기록물 컬렉션 단위나 시리즈 단위로 기록정보를 내보내기할 수 없고, 건 별로 내보내기 해야 하는 단점이 존재한다. 그리고 보통의 RDF 변환 소프트웨어는 엑셀과 데이터베이스 형태의 데이터를 변환시켜주는 프로그램이기 때문에, 인간과 기억 아카이브에서 내보내기한 XML과 EAD 데이터를 다시 엑셀로 변환시켜줘야 하는 문제점이 발생한다. 따라서 본 연구는 소프트웨어를 사용하지 않고 대상 기록을 직접 살펴보면서, LOD로 발행되어 개방되었을 때 보다 활발한 공유가 가능한 접근점을 가진 기록 100건을 선정하였다. 그리고 그 기술정보를 엑셀 스프레드시트로 옮기고, 앞서 제안한 5단계 LOD 구축 절차에 따라 연구를 진행하였다(<그림 2> 참조). 여기서 접근점은 이용자가 정보검색 시 원문검색이나 주제어 검색을 선호하는 점을 고려하고, 외부 LOD와 상호연결할 수 있는 접근점을 고려한 것이다. Gracy(2015)는 MARC와 EAD로 표현된 정보를 다른 데이터 세트의 정보와 연결하는 방법을 연구하며 이름, 장르, 주제, 직업과 같은 접근점을 분석했다. 특히 링크드 데이터로 변환할 때 도움이 되는 이름이나 주제어가 포함된 영역(전기, 초록, 기술 등)에 대해 연구했다. 본 연구에서 구축한 LOD와 한국관광공

ReferenceCode	Level	Upper Level	Title	Creator	Date	ExtentAndMedium	Repository	Note	Media Type	RelatedOrganization	Location	Subject	RelatedPerson
HUMA KR C2	collection		8월12일 일기수집 collection		20130512		영양가/역사자료						안락민
HUMA KR C2 1-5-12	series	HUMA KR C2	제1호 기록물 일기수집 일기		20130512	제1호 기록물 일기수집 일기	영양가/역사자료						
HUMA KR C2 sub-series-3	subseries	HUMA KR C2 1-5-12	장소정보		20130512	영양가/역사자료	영양가/역사자료			생활유지물			
HUMA KR C2-20160426KAR14	item	HUMA KR C2 sub-series-3	가죽수첩 제복포의 디자인자료	윤민철	20130512	그림물기 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			제복포	이우
HUMA KR C2-20160426KAR22	item	HUMA KR C2 sub-series-3	제복포의 디자인자료		20130512	그림물기 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			제복포	이우
HUMA KR C2-20160426KAR17	item	HUMA KR C2 sub-series-3	제복포의 디자인자료	정유환	20130512	그림물기 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			제복포	이우
HUMA KR C2-20160426KAR13	item	HUMA KR C2 sub-series-3	염색, 인화용 디자인자료		20130512	그림물기 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			이형이다공물	이우
HUMA KR C2-20160426KAR30	item	HUMA KR C2 sub-series-3	염색, 인화용 디자인자료	정유환	20130512	그림물기 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			63필름	이우
HUMA KR C2-20160426KAR38	item	HUMA KR C2 sub-series-3	염색, 인화용 디자인자료		20130512	그림물기 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			염색	이우
HUMA KR C2-20160426KAR11	item	HUMA KR C2 sub-series-3	가죽수첩 디자인 자료	정유환	20130512	그림물기 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			염색	이우
HUMA KR C2-20160426KAR35	item	HUMA KR C2 sub-series-3	가죽수첩 디자인 자료		20130512	그림물기 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			염색	이우
HUMA KR C2-20160426KAR23	item	HUMA KR C2 sub-series-3	가죽수첩 디자인 자료	정유환	20130512	그림물기 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			염색	이우
HUMA KR C2 2-5-12	series	HUMA KR C2	기술계층		20130512		영양가/역사자료						
HUMA KR C2-20160426KAR291	item	HUMA KR C2 2-5-12	기술계층	정유환	20130512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-20160426KAR292	item	HUMA KR C2 2-5-12	기술계층	정유환	20130512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-20160426KAR293	item	HUMA KR C2 2-5-12	기술계층	정유환	20130512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-20160426KAR311	item	HUMA KR C2 2-5-12	기술계층	정유환	20130512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-20160426KAR328	item	HUMA KR C2 2-5-12	기술계층	정유환	20130512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-20160426KAR324	item	HUMA KR C2 2-5-12	기술계층	정유환	20130512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2 2-5-12	series	HUMA KR C2	기술계층		20130512		영양가/역사자료						
HUMA KR C2 sub-series-17	subseries	HUMA KR C2 2-5-12	기술계층		20130512		영양가/역사자료						
HUMA KR C2-201409-MK204	item	HUMA KR C2 sub-series-17	기술계층	정유환	20140512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-201409-MK201	item	HUMA KR C2 sub-series-17	기술계층	정유환	20140512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-201409-MK197	item	HUMA KR C2 sub-series-17	기술계층	정유환	20140512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-201409-MK196	item	HUMA KR C2 sub-series-17	기술계층	정유환	20140512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-201409-MK207	item	HUMA KR C2 sub-series-17	기술계층	정유환	20140512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-201409-MK206	item	HUMA KR C2 sub-series-17	기술계층	정유환	20140512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-201409-MK205	item	HUMA KR C2 sub-series-17	기술계층	정유환	20140512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2 sub-series-18	subseries	HUMA KR C2 2-5-12	기술계층		20140512		영양가/역사자료						
HUMA KR C2-201409-MK408	item	HUMA KR C2 sub-series-18	기술계층	정유환	20140512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-201409-MK409	item	HUMA KR C2 sub-series-18	기술계층	정유환	20140512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-201409-MK407	item	HUMA KR C2 sub-series-18	기술계층	정유환	20140512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-201409-MK406	item	HUMA KR C2 sub-series-18	기술계층	정유환	20140512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-201409-MK405	item	HUMA KR C2 sub-series-18	기술계층	정유환	20140512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-201409-MK404	item	HUMA KR C2 sub-series-18	기술계층	정유환	20140512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-201409-MK403	item	HUMA KR C2 sub-series-18	기술계층	정유환	20140512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-201409-MK402	item	HUMA KR C2 sub-series-18	기술계층	정유환	20140512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-201409-MK401	item	HUMA KR C2 sub-series-18	기술계층	정유환	20140512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-201409-MK398	item	HUMA KR C2 sub-series-18	기술계층	정유환	20140512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2-201409-MK400	item	HUMA KR C2 sub-series-18	기술계층	정유환	20140512	기술계층 1권 (A4 1페이지)	영양가/역사자료	유물수보장 2016.08.14	text/pdf			기술계층	이우
HUMA KR C2 sub-series-14	subseries	HUMA KR C2 2-5-12	기술계층		20140512		영양가/역사자료						

〈그림 3〉 대상 기록정보 수집 및 분석

사 LOD의 연결점을 분석했을 때도 장소와 같은 주제가 유용한 것으로 나타났다. 따라서 본 연구는 이름, 주제, 장소와 같은 내용정보를 중심으로 대상 기록을 선정했다.

3.2 대상 기록정보 수집 및 분석

앞서 언급하였듯이 대상 기록컬렉션 중에서 선정한 100건의 기록물들은 1개의 컬렉션 레벨, 3개의 시리즈 레벨, 15개의 서브시리즈 레벨, 81개의 아이템 레벨로 구성되어 기록물의 계층을 LOD에서 표현할 수 있도록 설정하였다. 그리고 〈표 1〉에서 제시된 기술 요소를 활용하여 〈그림 3〉과 같이 엑셀 스프레드시트에 복사하였다.

‘참조코드’, ‘기술계층’, ‘제목’, ‘생산자’, ‘일자’, ‘규모및매체’, ‘저장소’, ‘기록유형’ 등의 요소는 인간과 기억 아카이브의 AtoM에 기술된 값을 그대로 사용하였다. ‘상위계층’ 요소는 기

존 기록물 기술에는 없는 요소지만, 온톨로지에 엑셀 데이터의 값을 입력할 때 기록물의 계층관계(isPartOf)를 표현하기 위해서 설정하였다. ‘관련조직’, ‘장소’, ‘주제’, ‘관련인물’ 등 기록의 내용정보에 대한 요소는 연구자가 일기의 전문을 보고 판단하여 기술했다.

아키비스트는 대상 기록정보를 수집하고 분석하는 과정에서 기록의 맥락과 구조, 내용을 파악하여 LOD 환경에서 자관의 데이터가 많은 상호연결을 가질 수 있도록 접근점을 생성해야 한다. 앞서 기술 요소를 설정한 것처럼 일기의 내용 속에 있는 주제나 장소, 기록물 생산과 관련된 인물 등의 항목은 중요한 접근점이 될 수 있다. 또한 기록 기술의 내용을 검수하여 접근점으로 활용할 값(value)에 대한 판단도 내려야 한다. 본 연구에서 활용한 대상 기록은 유치원생의 그림일기에 대한 기록이 많다. 그 중 하나를 살펴보면 일기의 내용에는 송리산이라 표기되어 있지만, 해당 산은 존재하지 않는

경우가 있다. 이런 경우 아키비스트가 유치원의 위치와 가까운 산이 속리산이라는 점과 기록생산자의 연령 등을 감안하여 잘못 표기한 것이라 판단하고, 장소에 대한 접근점을 속리산으로 부여할 수 있다.

3.3 온톨로지 구축

기록의 수집과 분석이 끝나면, 그것을 토대로 온톨로지를 구축한다. 온톨로지 구축은 앞에서 구성한 7단계에 따라 작업을 진행했다.

첫째는 본 온톨로지의 분야와 범위를 결정하였다. 본 연구의 온톨로지 범위는 추후 일기기록 분야에서 재사용이 가능한 온톨로지를 제작하는 것이 목적이 아니라, LOD 구축의 절차를 보여주기 위함이므로 인간과 기억 아카이브의 일기컬렉션이 가지고 있는 기록물의 기술과 내용에만 한정하여 온톨로지를 구축하고자 했다. 그리고 본 연구의 온톨로지가 제공하는 정보는 본래 기록물에 기술되어 있던 구조·맥락정보와 연구자가 직접 선정한 내용 정보라고 정의하였다.

이렇게 온톨로지의 구축범위를 정한 후, Protege를 사용하여 온톨로지 헤더(Ontology Header)를 작성했다. 온톨로지 헤더에는 dc:title, dcterms:issued, rdfs:comment 등과 같이 Annotation Property를 사용하여 '해당 온톨로지가 무슨 이름인지, 언제 발행되었는지, 어떤 내용에 대한 것인지'와 같이 상세한 내용들을 기술했다.

둘째는 존재하는 온톨로지의 재사용을 고려하는 절차이다. 본 연구는 더블린코어(DC)와 DCTERMS, SKOS, FOAF 등의 어휘를 재사

용하고, Schema.org의 범용 온톨로지를 재사용하기로 결정했다. Schema.org란 인터넷상의 구조화된 데이터를 위한 스키마로써, 구글·MS·야후 등의 기업과 다양한 사람들이 참여하여 만든 온톨로지이다. 이 온톨로지는 Event, Place, Product, Action 등 다양한 개념의 클래스들을 표현하고 있다. 본 연구에서는 장소에 대한 개념을 표현하기 위해 사용했다.

최근 ICA(세계기록관리협회)에서는 ISAD(G), ISAAR(CPF), ISDF, ISDIAH 등의 기술표준을 준용하면서, 온톨로지에서도 사용 가능한 차세대 기술표준으로 Record in Context(RIC)를 발표했다. 하지만 현재까지는 개념모델(Conceptual Model)만 발표되었고, OWL형식의 RIC-Ontology는 공개되지 않아, 본 연구에서 온톨로지 모델로 활용되지 못했다.

셋째는 온톨로지에서 중요한 단어들을 열거하는 단계이다. 이 단계에서는 클래스와 속성으로 사용가능한 표현들을 정리하였다. 사실 AtoM의 기술 요소가 기록의 구조, 맥락, 내용적인 측면에 대해서 범주화하여 분류해놓은 항목들이기 때문에, 앞서 <표 1>에서 선정한 기술 요소들과 그 값들을 중요한 단어로 선정하여 열거할 수 있다. 또한 기록물의 내용에서 찾을 수 있는 관련조직, 장소, 인물, 주제들도 중요한 단어로 열거했다.

넷째는 클래스와 클래스의 관계를 정의하는 단계이다. 이 단계에서 총 5개의 클래스를 정의하였다. 클래스는 의미적으로 비슷한 범위 안에 속하는, 각기 고유한 URI를 갖는 객체(Individual)의 집합체이다. 예를 들어 그 하위로 Italy, USA, England와 같은 객체들이 모여 Country라는 클래스를 형성하는 것이다. 따라서 선정한 100

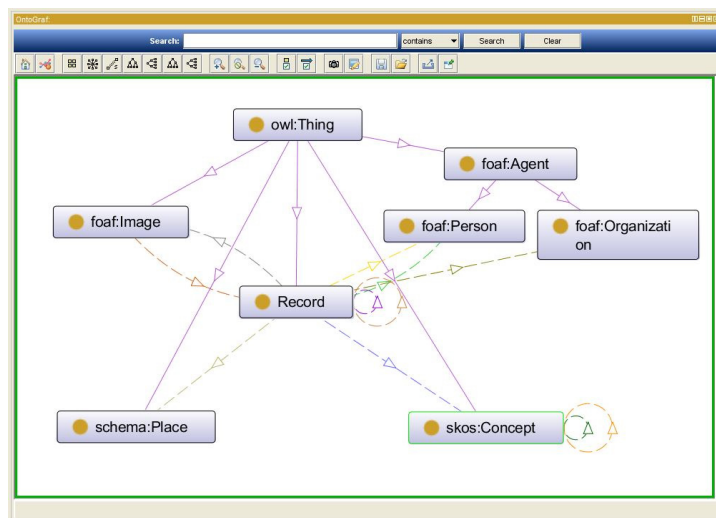
개의 기록들이 각각 고유한 URI를 가지는 하나의 객체(자원)라면, 그러한 객체들은 일기 기록물이라는 하나의 클래스로 묶을 수 있다. 또한 인물·조직·장소·주제 등의 기술 요소들도 고유한 URI를 가지면서 전거레코드와 같은 역할을 하는 접근점이 될 수 있으므로, 객체로 생성하고 클래스로 묶었다. 나머지 요소들은 단지 기록물을 설명하는 역할로서, 따로 URI를 갖지 않는 것으로 정의했다. 예를 들어 제주도, 제주도, 속리산, 전주, 서울 등의 데이터는 하나의 URI를 가지는 객체로 독립하여 다른 정보와의 연결점으로 사용하였다. 또한 인물과

조직은 하나의 상위클래스를 가질 수 있으므로 foaf:Agent라는 상위클래스를 생성하여 그 밑에 포함시켰다. 클래스명은 앞서 언급했던 것처럼 재사용 온톨로지 어휘에서 의미적으로 동일한 클래스의 이름을 재사용하였다. 최종적으로 생성한 클래스의 정의는 <표 2>와 같고, 클래스간의 관계는 <그림 4>와 같이 정의하였다. 그러나 제목, 일자, 주기사항 등은 객체로 생성하지 않고, 단순히 기록물을 설명하는 속성으로 사용했다.

다섯째는 클래스의 속성(Property)을 정의하는 단계이다. 먼저 클래스에 속한 객체와 다른

<표 2> 대상 기록의 온톨로지 클래스(Class) 정의

클래스명	사용어휘	의미
foaf:Agent	foaf:Organization	조직
	foaf:Person	인물
Record	본 온톨로지 고유정의	기록물
schema:Place	schema.org	장소
skos:Concept	skos	주제



<그림 4> 대상 기록의 클래스 관계

클래스의 객체를 연결하는 Object Property를 정의하기 위해서 클래스 간의 관계를 살펴보았다. 본 연구에서 설계한 온톨로지는 모든 객체들이 일기기록(record)을 기술하는 요소로 사용되는 모델이므로, 최상위 클래스인 owl:Thing의 밑에 있는 모든 클래스들이 diary:Record 클래스를 중심으로 연결되어 있는 형태이다. 즉, 기록물-인물, 기록물-조직, 기록물-장소, 기록물-주제, 기록물-기록물, 주제-주제 등 6가지 관계가 존재하는 다중개체모형이다(〈그림 4〉 참조).

본 연구에서 정의한 모든 Object Property는 더블린 코어나 DCTERMS, Schema.org, SKOS 등 다른 온톨로지 어휘를 재사용하여 정의했다.

객체를 기술하는 데이터 값을 표현하는 Data Property에는 앞서 3.1에서 열거한 항목 중 Object Property로 선정되지 않은 일자와 주기 사항, 참조코드, 제목, 형태, 규모 및 매체, 기술 계층, 기록보존소 등의 항목들을 사용했다. 해당 요소들은 단지 기록물 객체를 설명하는 역할이기 때문에 Data Property로 사용한다. 하지만 온톨로지 모델링에 따라 일자 등과 같은 요소와 그 값도 URI를 부여한 객체로서 다룰 수 있다. Data Property도 Object Property와 마찬가지로 더블린코어나 DCTERMS와 같은 다른 온톨로지의 어휘를 재사용하였다.

참조를 위해 사용하는 Annotation Property에서는 OWL이 기본적으로 가지고 있는 주석 속성을 제외하고, 온톨로지의 제목을 제시하는 dc:title과 온톨로지의 발행일을 제시하는 dcterms:issued를 생성하여 온톨로지 헤더 정보에 대한 주석을 달았다. 속성을 정의하면서 한 속성에서 사용한 어휘는 다른 속성에서 중복하여 않도록 유의해야 한다.

여섯째는 속성의 제약조건과 정의역 및 공역을 정의하는 단계다. 본 연구에서는 2가지 제약조건을 설정했다. 먼저, 1개의 기록물 컬렉션은 여러 개의 기록시리즈를 가질 수 있으므로 기록물의 포함 관계를 나타내는 속성인 dcterms:hasPart에 대해서 역함수(Inverse Functional) 제약조건을 설정하였다. 그리고 기록은 집합적 기술의 원칙에 따라 상위 계층의 기술을 계승받으므로 이행(Transitive) 제약조건도 설정했다. 그리고 dcterms:hasPart의 역속성인 dcterms:isPartOf는 제약조건으로 함수(Functional)속성을 사용했다.

또한 Object Property와 Data Property의 정의역(Domain)과 공역(Range)에 대해서 설정하였다. 〈표 3〉은 본 연구의 온톨로지서 정의한 속성들과 정의역 및 공역의 설정을 정리한 것이다.

마지막으로 일곱 번째는 객체(Individual)를 생성하는 것이다. Protege에서 직접 객체를 생성하여 속성을 연결할 수 있지만, 데이터의 양이 많아지면 수동으로 입력하는 것은 한계가 있다. 때문에 보통 이 단계에서는 데이터 변환 프로그램을 사용하여 객체를 자동 생성하는 경우가 많다.

3.4 객체 생성 및 상호연결

지금까지가 온톨로지 모델을 구축하기 위한 단계였다면, 이번 단계는 구축한 온톨로지 모델에 데이터를 입력하여 링크드 오픈 데이터를 구축하는 단계이다. 이 단계는 객체를 생성하는 단계와 생성한 객체를 외부 LOD와 상호연결하는 단계로 나뉜다.

〈표 3〉 대상 기록의 온톨로지 속성(Property) 정의

유형(Type)	이름	정의역(Domain)	공역(range)	제약조건
Object Property	dc:subject	diary:Record	skos:Concept	
	dcterms:contributor	diary:Record	foaf:Person	
	dcterms:creator	diary:Record	foaf:Person	
	dcterms:hasPart	diary:Record	diary:Record	• InverseFunctional • Transitive
	dcterms:isPartOf	diary:Record	diary:Record	Functional
	dcterms:relation	diary:Record	foaf:Organization	
	event:place	diary:Record	schema:Place	
	skos:broader	skos:Concept	skos:Concept	
	skos:narrower	skos:Concept	skos:Concept	
owl:sameAs				
Data property	dc:date			
	dc:description			
	dc:identifier	diary:Record		
	dcterms:title			
	dcterms:hasFormat	diary:Record		
	diary:ExtentAndMedium	diary:Record		
	diary:level	diary:Record		
	diary:respository	diary:Record		
Annotation Property	dc:title			
	dcterms:issued			

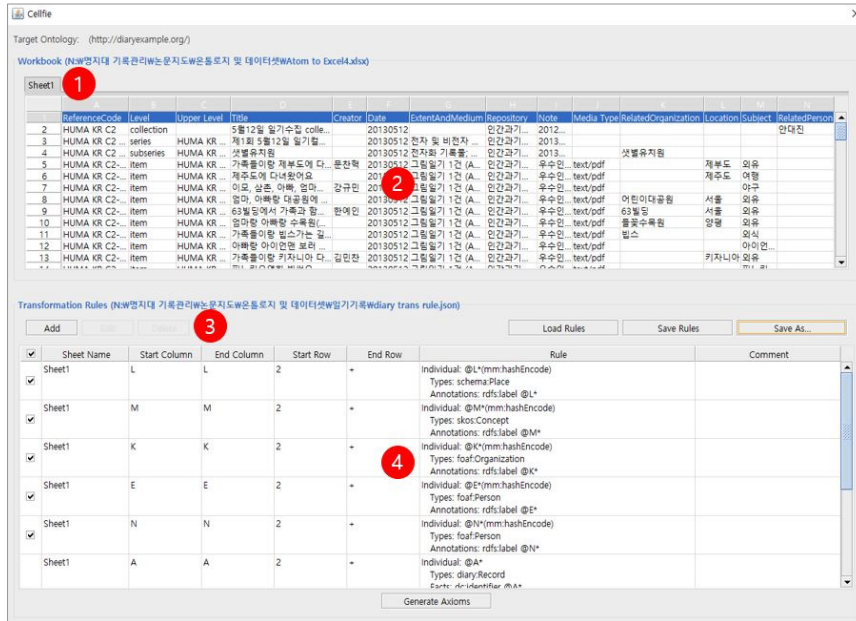
본 연구에서는 외부데이터를 입력하여 객체를 생성해주는 프로그램으로 Cellfie를 사용하였다. Cellfie는 엑셀 스프레드시트를 매핑규칙을 통하여 온톨로지의 객체로 생성해주는 Protege의 플러그인 소프트웨어이다.

Cellfie를 실행하면 〈그림 5〉와 같은 화면이 나오는데 그림의 ①번은 시트 탭으로서, 엑셀 내의 시트 탭과 동등한 탐색이 가능하다. ②번은 엑셀 내의 데이터를 볼 수 있도록 행과 열을 표시해주는 뷰어이다. ③번은 매핑(변환)규칙을 편집하는 메뉴인데, 추가와 편집, 삭제가 가능하다. ④번은 매핑 규칙 목록을 표시하고 있는 브라우저인데, 체크박스의 체크를 통해서 원하는 규칙만 실행할 수 있다. 이 때 규칙 사이

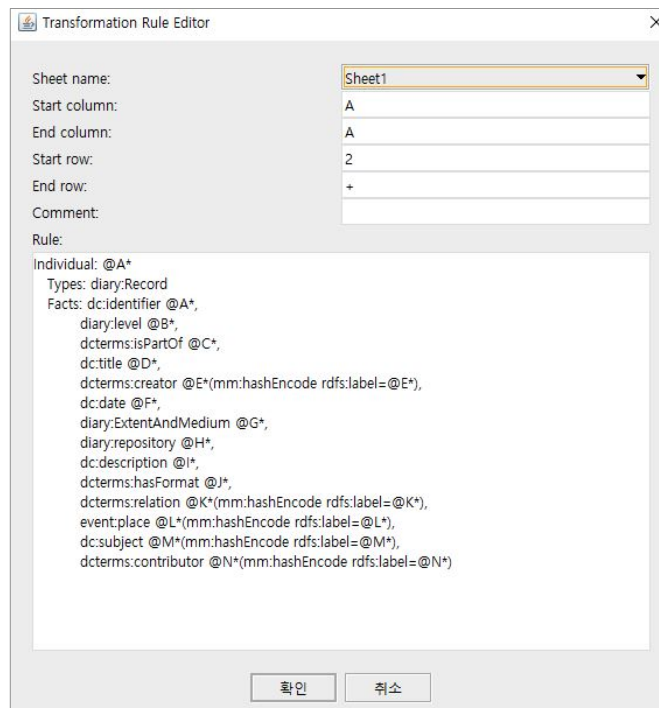
의 순서는 상관이 없다.

본 연구에서는 엑셀 스프레드시트의 E열(생산자), K열(관련조직), L열(장소), M열(주제), N열(관련인물)에 대한 매핑규칙을 작성하고, A열(기록물)에 대한 매핑규칙을 작성하였다.

새로운 매핑규칙을 만들 때는 〈그림 6〉에서 제시된 바와 같이 기준이 되는 데이터의 시트를 Sheet name에서 선택하여 규칙을 추가한다. 만약 본 연구와 같이 시트가 단 하나만 존재하는 것이 아니라 여러 개의 시트가 존재한다면 매핑규칙을 정의할 때 이 부분에 유의해야 한다. 보통 EAD(Encoded Archival Description) 파일을 엑셀 파일로 변환한다면, 전거데이터에 대한 정보들이 각기 다른 시트에 담기므로 여러 개의



<그림 5> Cellfie 실행화면



<그림 6> Cellfie Rule 편집화면

시트가 존재하게 된다. 다음으로, <그림 6>의 Start column과 End column, Start row와 End row 메뉴에서는 각각 시작과 끝의 열과 행을 정의하는데, 특히 1행은 기술요소 중 제목이 들어가게 되므로 2행부터 시작하게 된다. 끝 부분에 '+' 기호를 쓰면, 데이터가 있는 행까지 매핑규칙이 적용되어 데이터가 변환된다. Rule 부분에는 매핑(변환)규칙을 작성하게 되는데, 매핑 규칙의 작성은 OWL문장을 사람이 읽기 쉽도록 변형해주는 맨체스터 구문(Manchester Syntax)을 기반으로 한다(O'Connor, Halaschek-Wiener, & Musen, 2010). 또한 맨체스터 구문은 다양한 제약조건들을 구현할 수 있는데, 주로 괄호를 통해서 제약조건을 정의한다(Josef, 2016).

다음, 앞의 3.2에서 제시된 대상 기록정보 수집 및 분석 단계에서 정리한 엑셀 파일을 온톨로지에 삽입(변환)하기 위해서 다음과 같이 매핑규칙을 작성했다. 먼저, 생산자(E열)와 관련조직(K열), 장소(L열), 주제(M열), 관련인물(N열)에 대한 객체를 생성하는 매핑 규칙을 작성하였다. 이는 본 LOD의 기반이 되는 온톨로지 모델에서 주제, 장소, 관련조직, 관련인물 클래스들이 전거레코드와 같은 역할을 하면서 Record 클래스를 기술할 때 상호참조 되는 형태이기 때문에 필요한 과정이다. 따라서 객체를 생성하는 매핑 규칙에서도 Record 클래스를 제외한 다른 클래스들의 객체를 먼저 생성한 후, Record 클래스의 객체를 생성할 때 이들과 연결해주었다.

둘째, 기록물에 대한 객체를 생성하는 매핑 규칙을 작성하였다(<그림 6> 참조). 이 단계에서는 맨체스터 구문의 제약조건을 활용하여 첫 번째 단계에서 생성된 객체의 값을 전거레코드처럼 상호참조 할 수 있도록 설정하였다. 예를

들어 기록물(A열)의 객체를 생성할 때 주제(M열)의 값이 존재한다면 그것을 기록물의 주제로 상호참조되도록 하는 것이다.

매핑규칙을 통해 객체를 생성하였으면 LOD의 상호연결(Interlinking)을 위한 작업을 진행해야 한다. 이 작업이 풍부하게 이뤄질수록 링크드 데이터는 더 많은 LOD와 연결되어 공유되고 개방된다. 그리고 이 작업에서는 앞서 매핑규칙에서 실행하지 못했던 주제와 주제 간의 상하관계를 SKOS 어휘를 통해 마치 시소러스처럼 설정할 수 있다. 보통 내부 객체간의 관계에 대한 정의가 끝나면, 데이터의 풍부한 공유와 개방을 위해서 외부 LOD와의 상호연결 작업을 진행한다.

외부 LOD와 상호연결하기 위해서는 먼저 대상 객체를 선정하고, 해당 객체와 의미적으로 동일한 외부 LOD의 객체를 검색해야 한다. 이 작업은 구글 등 인터넷 검색을 통해서 이뤄질 수도 있고, DBpedia 혹은 Wikipedia와 같이 유명한 LOD사이트에서 찾아낼 수도 있다. 이렇게 대상 객체와 동일한 의미를 가진 객체의 URI를 찾았다면 이를 Protege를 통해 연결시켜준다.

예를 들어, 외부 LOD와 상호연결을 정의한 예시 중에 '제주도'라는 장소 객체를 외부 LOD와 상호연결하기 위하여, 구글 검색을 통해 제주도에 대한 외부 LOD 데이터를 검색하였다. 검색을 통해서 DBpedia, KData, 주소데이터, 한국관광공사 등의 LOD사이트에서 본 '제주도'라는 객체를 찾아냈다. 이렇게 찾아낸 객체의 URI들을 새로운 객체로 온톨로지 내에서 생성하고, 이를 본 연구의 온톨로지에서도 '제주도'라는 객체와 연결시킴으로써 상호 연결하였다.

인터넷에서 상호연결을 위한 객체 검색 시

유의할 점은, 실제 객체의 고유한 URI와 웹에서 서비스하는 페이지의 URI가 다르다는 점이다. 한국관광공사LOD의 제주도에 대한 객체의 URI는 “http://data.visitkorea.or.kr/resource/CATEGORY:390000”이지만, 해당 객체를 웹에서 서비스하고 있는 페이지의 URI는 “http://data.visitkorea.or.kr/page/CATEGORY:390000”이다. 이렇게 대다수의 LOD사이트에서 객체 고유의 URI와 객체의 정보를 보여주는 웹서비스 페이지의 URI가 서로 다른 모습을 볼 수 있다. 상호연결 시에는 이 점을 주의하여 웹 서비스 페이지의 URI가 아니라, 객체의 고유한 URI를 연결해야 한다.

3.5 데이터 저장 및 발행

앞에서는 Protege와 Cellfie를 활용하여 데이터 모델링작업과 객체생성(RDF변환) 및 상호연결의 작업을 진행하였다. 이번 단계에서는 앞의 작업을 통해 얻은 RDF 파일을 RDF저장소에 저장하고, SPARQL Endpoint 검색을 통해서 LOD구축이 완료된 기록정보가 어떻게 확장되고 연결되는지 확인해보았다.

연구에서 활용하는 RDF 저장소는 Apache Jena Fuseki Server이다. Apache Jena Fuseki Server는 여러 가지 방법으로 실행시킬 수 있는데, 일반적으로는 명령프롬프트(cmd)를 이용하여 작동시키는 Standalone방식이 보편적이다(정갑주, 2015). 하지만 이 경우에는 외부 웹에서 도메인을 통해 접속할 수 없다는 문제점이 존재한다. 본 연구는 웹 서버에 Apache Jena 소프트웨어를 설치하는 것이 아니라 실험 PC에 소프트웨어를 설치하는 Standalone방식을 선

택했다. 이렇게 내부 웹 서버인 Localhost 환경에서 데이터의 저장과 SPARQL 검색 등의 연구를 진행함에 따라 실제 LOD 환경에서 기록 정보가 연결되는 것을 보여주지 못하였다는 한계가 있다.

Jena Fuseki를 Localhost 웹 환경에서 구현하기 위해서는 명령 프롬프트(cmd)로 실행하는데, 한 가지 유의할 점은 Jena Fuseki Server를 구동하기 전에 Oracle사의 JAVA가 설치되어 있어야 한다. Jena는 Java기반의 시스템이기 때문에, JAVA가 설치되어 있지 않으면 오류가 나면서 실행되지 않는다.

가동된 Jena Fuseki Server에서는 데이터 세트를 관리하고, 새로운 데이터를 추가하며, SPARQL Query 실행하는 등 여러 가지 작업을 실행할 수 있다. 업로드된 RDF 파일은 SPARQL Query를 통해 검색과 수정이 가능하다. 앞서 여러 차례 언급한 SPARQL은 Simple Protocol And Rdf Query Language의 약칭으로, 2008년 W3C에서 최종표준안으로 제정된 질의 언어이다. SPARQL 질의 언어는 다중 매칭, 다양한 조건적 매칭, 검색 조건의 제약 등을 지원하고, 트리플(SPO) 형태의 추론에 용이한 구조를 지원함으로써 효과적인 시맨틱 검색을 유도한다(하상범, 한은영, 최호준, 2005). 트리플(SPO) 형태란 시맨틱 웹의 모든 RDF 자원이 주어부(Subject), 술어부(Predicate), 목적부(Object) 형태의 트리플(SPO) 구조로 이루어져 있는 것을 의미한다.

SPARQL은 SELECT, ASK, CONSTRUCT, DESCRIBE 등 4개의 Query 형식이 있다. 대표적으로 SELECT는 SQL과 같이 찾고자 하는 데이터를 검색하기 위해 사용한다. SPARQL의 구

문은 PREFIX, {SELECT, ASK, DESCRIBE, CONSTRUCT}, WHERE}절로 구성된다. 변수의 앞에는 ?를 사용하며, WHERE절은 {S P O }의 형태로 주어, 술어, 목적어.(마침표)를 넣어 검색을 실행한다(이현정, 2015).

본 연구의 실제 실행단계에서는 SPARQL 검색의 사례로 한국관광공사LOD를 상호연결하여 살펴보았다. 한국관광공사LOD는 관광지, 숙박, 음식점, 축제 등의 정보를 포함하고 있어서, 상호연결하여 다양하게 활용할 수 있는 LOD이다. 이러한 LOD를 본 연구에서 구축한 LOD와 '장소' 객체를 통해 상호연결 한다면, SPARQL 검색을 통해서 "제주도가 가진 다른 언어로 된 명칭들에 대한 데이터"를 가져올 수 있다. 이러한 질의(query)는 <표 4>와 같이 작성했다.

먼저 SPARQL에서 사용할 어휘들에 대해서 Prefix를 선언하고, 전체를 검색(SELECT *) 할 것이라 표현하였다. WHERE 이하의 조건 절에서는 본 연구의 LOD와 상호연결된 외부 LOD를 찾기 위한 질의를 작성했다. WHERE 절에서 중간에 세미콜론(:)을 사용하였는데, 이는 하나의 문장을 주어부(혹은 술어부나 목적어부)로 만들어주는 기호이다. 그리고 WHERE절 안에 SERVICE 명령어를 사용하였는데, 이는 외부 LOD가 SPARQL Endpoint서비스를 제공

하고 있으면 그 서비스를 통해 SERVICE < >안의 조건에 해당하는 정보를 검색해서 값을 가져오는 명령어이다.

이러한 질의(query)의 결과는 <그림 7>과 같다. 결과화면에는 <표 4>의 질의문에서 지정한 변수 네 가지가 항목으로 출력되며, '일기기록 Data'와 '연결된 외부 LOD' 항목의 값, 즉 URI를 클릭하면 해당 LOD데이터로 이동이 가능하다. <그림 7>의 결과화면에서 '외부 LOD의 이름'과 '다른 명칭들'의 목록에 나와 있는 데이터 값은 현재 본 연구의 LOD에서 보유하고 있는 데이터가 아니라 '제주도' 객체와 상호연결된 외부 LOD의 데이터를 불러온 것이다. 이것은 LOD의 상호연결을 통해 데이터가 확장되고 연결된다는 점을 증명한다.

<표 4>에서 사용한 skos:prefLabel은 색인어 및 다국어 레이블을 표현하는 SKOS의 어휘를 사용한 것이다. 이처럼 다른 온톨로지의 어휘를 재사용하면 그 의미가 공유되므로, 의미의 상호운용성이 증가하고 SPARQL 검색시 편리하다는 이점이 있다. 만약 <표 3>의 diary:level처럼 각자의 온톨로지에서 고유하게 정의한 어휘를 속성으로 사용하게 되는 경우, 해당 속성의 URI를 따로 찾아보는 작업을 거쳐야 하는 번거로움이 있다.

<표 4> 상호연결된 외부 LOD 데이터를 가져오기 위한 SPARQL Query

```

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
SELECT *
WHERE { ?일기기록Data owl:sameAs ?연결된외부LOD .
SERVICE <http://data.visitkorea.or.kr/sparql>
{ ?연결된외부LOD rdfs:label ?외부LOD의이름 ; skos:prefLabel ?다른명칭들 . } }
    
```

일기기록Data	연결된외부LOD	외부LOD의이름	다른명칭들
1 <http://diaryexample.org/4cc804471acbda432f35f1639a722bb8>	<http://data.visitkorea.or.kr/resource/CATEGORY:390000>	"제주도"@ko	"제주도"@ko
2 <http://diaryexample.org/4cc804471acbda432f35f1639a722bb8>	<http://data.visitkorea.or.kr/resource/CATEGORY:390000>	"제주도"@ko	"濟州道@zh-hant"
3 <http://diaryexample.org/4cc804471acbda432f35f1639a722bb8>	<http://data.visitkorea.or.kr/resource/CATEGORY:390000>	"제주도"@ko	"Jeju-do"@es
4 <http://diaryexample.org/4cc804471acbda432f35f1639a722bb8>	<http://data.visitkorea.or.kr/resource/CATEGORY:390000>	"제주도"@ko	"Jeju-do"@fr
5 <http://diaryexample.org/4cc804471acbda432f35f1639a722bb8>	<http://data.visitkorea.or.kr/resource/CATEGORY:390000>	"제주도"@ko	"Jeju-do"@en
6 <http://diaryexample.org/4cc804471acbda432f35f1639a722bb8>	<http://data.visitkorea.or.kr/resource/CATEGORY:390000>	"제주도"@ko	"Jeju-do"@de
7 <http://diaryexample.org/4cc804471acbda432f35f1639a722bb8>	<http://data.visitkorea.or.kr/resource/CATEGORY:390000>	"제주도"@ko	"濟州特別自治道"@ja
8 <http://diaryexample.org/4cc804471acbda432f35f1639a722bb8>	<http://data.visitkorea.or.kr/resource/CATEGORY:390000>	"제주도"@ko	"济州道@zh-hans"
9 <http://diaryexample.org/4cc804471acbda432f35f1639a722bb8>	<http://data.visitkorea.or.kr/resource/CATEGORY:390000>	"제주도"@ko	"Чечкудо"@ru

〈그림 7〉 상호연결된 외부 LOD 데이터를 가져오기 위한 SPARQL Query 결과

이상으로 오픈소스 도구를 이용하여 기록정보의 수집 및 분석, 온톨로지 구축, 객체 생성 및 상호연결, 데이터 저장 및 발행 등 기록정보의 링크드 오픈 데이터 구축 절차에 대해 POC (Proof Of Concept)를 진행해보았다.

4. 기록관 LOD 구축의 필요 요건

4.1 내용정보 기술 품질의 고도화

기록이 활동을 증거하기 위해서는 내용, 맥락, 구조의 세 가지 요소를 갖추어야 하나, 기록의 기술은 내용보다는 주로 맥락과 구조에 대한 항목들로 이루어진다. 그러나 앞서 3.1에서 언급한 것처럼 일반 이용자들은 기록의 주제나 내용을 통해 검색하는 것을 선호한다. 또한 내용정보가 다른 정보들에 비해 링크드 데이터 환경에서 외

부 LOD와 연결하기 유용하다. 따라서 기록의 내용정보에 대한 기술이 잘 관리된다면, 링크드 데이터를 구축할 때, 보다 용이하게 기록정보가 확장되고 공유될 것이다. 예를 들어, 수집한 일기기록 중 가족과 영화를 보러갔다는 내용의 일기가 있었는데, 일기의 전문을 읽어보니 영화의 제목이 '아이언맨3'라는 것을 파악할 수 있었다. 만약 그것을 주제로 기술했다면, LOD 구축 시에 해당 주제를 DBPedia의 '아이언맨3'와 상호연결하여 기록정보를 확장시킬 수 있다. 이러한 상호연결 과정을 거치게 되면 아카이브는 따로 데이터에 대한 기술을 하는 과정 없이 '아이언맨3'에 대한 상세한 내용을 이용자에게 제공할 수 있고, 필요한 때에 SPARQL검색을 통해서 해당 객체에 대한 정보를 가져올 수 있다.

공공기관의 아카이브가 아니라 인간과 기억 아카이브와 같은 일상 아카이브의 경우에는 기록의 맥락이 아닌 내용정보의 중요성이 더욱 강조된다. 그러나 인간과 기억 아카이브에서

수집한 일기 기록물 컬렉션을 살펴보면 구조와 맥락정보가 중점적으로 기술되어 있고, 일부 건에 대해서만 주제어를 선정하여 기술하고 있었다. 이는 공공기관의 아카이브와 일상 아카이브 모두 기록물의 기술이 기능분류체계(혹은 출처분류) 아래에서 이뤄지고 있고, 구조정보와 맥락정보를 중시하고 있기 때문이라 생각한다. 하지만 인간과 기억 아카이브와 같은 일상 아카이브는 보통사람들의 일상을 미세한 영역으로 범주화하며, 개인의 행위와 경험을 사회적 맥락으로 해석하고 기록하는 곳이다(곽건홍, 2011). 따라서 아키비스트가 기록물 자체에 대한 내용을 파악하고, 그 내용이 가지는 사회적 맥락까지 고려하여 내용정보를 기술해야, 후대에 일상사를 연구할 때에 기록을 활용하기가 용이하다.

그렇지만 현실적인 여건 때문에 모든 기록의 내용을 파악하고 그 사회적 맥락을 해석하기란 어려운 일이다. 기록의 주제어를 자동으로 추출하는 방법을 내용정보를 기술하는 방법을 생각해볼 수 있지만, 내용의 주제와는 동떨어진 결과 나올 수 있는 한계점이 존재한다. 결국 기록의 내용정보에 대한 품질 향상을 위해서는 기록의 내용을 가장 잘 파악한 전문가가 판단하고 기술해야 할 것이다.

하나의 예를 들어 보면, 수집한 대상 기록물 중에서 어느 교사의 일기가 있었다. 만약 이 대상 기록의 내용을 제대로 파악하지 않고 단순히 제목과 생산자(직업)만 파악하여 기술했다면, 이 일기는 교사, 일기 등의 키워드로만 외부 LOD 데이터와 연결될 수 있었을 것이다. 하지만 아키비스트가 일기의 내용과 사회적 맥락을 파악하여 내용정보를 기술했다면, 세월호 침몰

사건에서 아무것도 실천하지 못한 교사의 분노, 혼란스러움, 교사라는 직업에 대한 정체성 혼란 등 더욱 많은 연결점을 찾을 수 있을 것이다. 이렇게 아키비스트가 기록의 내용을 파악하고 관련된 사회적 맥락을 고려하여 내용정보를 기술한다면, 앞서 언급한 바와 같이 더 풍성한 LOD 연결이 가능해질 것이다.

하지만 기록의 내용정보에 대한 기술이 없는 기록들에 대하여, 아키비스트가 모든 내용을 파악해서 연결점을 생성하기란 매우 어려운 일이다. 따라서 아키비스트는 기록물의 내용에 대하여 직접적으로 관련이 있는 업무담당자가 기록 생산 단계에서 주제어 등 기록의 내용정보에 대한 기술을 남길 수 있도록 독려해야 한다. 사실 온나라시스템에서도 기록의 생산단계에서 주제어를 기술하기 위한 항목이 존재했지만 실제로 많이 사용되지 않았다. 아키비스트는 이런 상황과 업무담당자의 인식을 개선하기 위하여 업무담당자에게 요구하기도 하고, 인식의 개선을 위한 교육을 마련하는 등 노력이 필요하다.

4.2 아키비스트의 역량 고도화

기록학계의 온톨로지와 LOD에 대한 대부분의 연구는 그것들의 구축을 통해서 얻을 수 있는 기록정보서비스의 질적 향상에 초점이 맞춰져 있었다. 따라서 기록정보를 LOD로 구축하는 구체적인 절차와 방법에 대한 학술적인 연구는 부족한 편이었다. 하지만 이 영역에 대한 학술적인 연구가 이뤄져야, 기록정보의 LOD 구축에서 아키비스트는 어떤 역할을 해야 할지, 어떤 지식을 함양해야 할지, 앞으로 어떻게 기

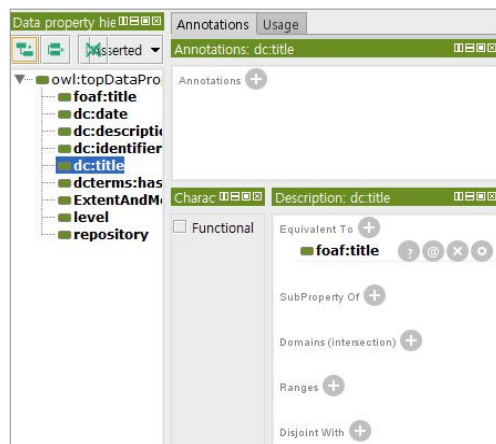
록물 기술을 함으로써 LOD 환경 속에 적응할 것인지 고민해볼 수 있을 것이다.

아키비스트가 기록정보의 LOD 구축에서 기여해야 할 부분은 기록정보를 분석하여 LOD 환경 속에서도 기록의 내용과 구조, 맥락을 모두 표현할 수 있는 온톨로지 모델을 구현하는 것이다. 소규모 아카이브의 경우에는 예산 확보가 어렵기 때문에 본 연구에서 제시한 것처럼 오픈소스 소프트웨어를 사용하여 직접 기록정보의 LOD를 구축할 수도 있겠지만, 일반적으로 데이터를 LOD로 구축하는 업무는 기업이 맡아서 진행한다. 이러한 경우 아키비스트가 LOD에 대한 이론적 배경과 절차에 대한 지식을 가지고 있다면, 소장하고 있는 기록정보에 더 적합한 LOD를 구축할 수 있도록 주도적으로 기업과 협의할 수 있을 것이다. 아키비스트는 출처의 원칙, 원질서 존중의 원칙 및 기술계층 등 기록정보가 다른 정보들과 다르게 갖는 고유한 특징들이 LOD에 온전히 반영될 수 있도록, LOD 구축에서 주도적인 역할을 할 수 있는 지식 역량을 강화해야 할 것이다.

4.3 기록정보의 접근성 향상을 위한 상호 연결 고도화

상호연결(Interlinking)은 3가지 종류가 있다. 데이터와 관련된 정보를 연결하는 관련 링크(Relationship Links), 객체와 동일하거나 추상적인 개념을 식별하기 위해서 다른 데이터 셋과 연결하는 동질 링크(Identity Links), 용어의 의미에 대한 정의를 나타내기 위해 연결하는 어휘링크(Vocabulary Links)가 있다(Heath & Bizer, 2011).

관련 링크로서 본 연구에서 사용한 예시는 제주도에 대한 다른 언어의 명칭들을 skos:prefLabel을 통해서 제주도와 연결시켜준 것이다. 이렇게 데이터에 대한 주석과 같은 설명을 연결시켜주는 것이 관련 링크이다.



〈그림 8〉 어휘링크 예시

어휘 링크는 dc나 skos, foaf와 같이 다른 재사용 어휘들 중에서, 온톨로지에서 사용한 속성(혹은 클래스)과 동일한 의미를 가진 어휘를 연결시켜주는 링크이다. 예를 들어서 dc의 title이나 foaf의 title은 모두 같은 의미를 지니는 어휘이다(〈그림 8〉 참조). 본 연구에서는 dc의 title을 선택하여 사용하였는데, 어휘 링크를 통해서 dc:title과 foaf:title의 의미가 같다고 설정해두면, 추후 SPARQL 검색에서도 foaf:title로 검색이 가능해진다. 추후 Record in Context - Ontology (RiC-O)가 발표된다면 RiC-O에서 사용하는 어휘들을 어휘링크를 통해 연결해줌으로써, 기록의 특성에 대한 보다 상세한 표현이 가능해지고, 기록정보에 관한 온톨로지의 특징을 확고히 가져갈 수 있을 것이다. 보통 어휘 링크는

owl:equivalentProperty 혹은 rdfs:subClassOf, skos:broadmatch 등의 어휘로 표현된다.

마지막으로 동질링크는 본 연구에서 가장 중점적으로 사용한 owl:sameAs와 같은 동일한 의미를 가진 객체를 연결시켜주는 것이다. 이와 같은 세 가지의 상호연결이 LOD 구축에서 풍성하게 정의되면, 표현하려는 데이터의 의미를 명확하게 정의할 수 있고, 데이터의 중복 구축을 방지하며, 연결된 다른 데이터세트의 잠재된 지식을 탐험하며 확장해나갈 수 있다.

본 연구는 구축한 LOD와 연관이 있을 것 같은 객체를 구글 검색이나 다른 LOD 사이트에서 검색하여 상호 연결하였다. 이렇게 상호연결을 정의하면서 여러 가지 데이터세트를 한 사이트에서 갖추고 있는 데이터 허브 사이트의 필요성을 느낄 수 있었다. 현재 정부에서는 정부 3.0을 표방하며 공공데이터포털(www.data.go.kr)을 통해 국가적 차원으로 공공데이터를 개방하고 있다. 공공데이터포털은 일종의 데이터 허브 사이트의 역할을 하는데, 2017년 2월 현재, 표준데이터 31세트와 오픈API 2,189건, 파일데이터 19,517건을 공개하고 있다. 파일데이터의 경우 RDF파일로 발행되지 않은 것이 많았지만, 표준데이터는 모두 RDF, JSON, XML, XLS, CSV 형식, 그리고 그리드, 지도, 차트, OPEN API형식으로 제공하여 활용할 수 있도록 서비스하고 있다. 표준데이터와 OPEN API로 제공하는 데이터들은 모두 기록정보 LOD와 연결하여 사용하기 좋은 데이터 세트라 판단된다. 가장 활용신청 건수가 높은 OPEN API 데이터세트는 국내 관광정보 서비스이다. 일기 기록물의 내용을 통해서 추출해낸 장소 정보를 이러한 데이터세트와 상호연결한다면, 보다 많

은 지식들을 브라우징할 수 있는 기회를 이용자에게 제공할 수 있을 것이다.

상호연결은 본 연구에서처럼 LOD 구축자가 직접 정의하는 것이 가장 정확하지만, 데이터의 양이 방대해지면 그것은 사실적으로 불가능하다. 따라서 프로그램을 사용하여 자동으로 데이터세트에서 의미 있는 객체를 생성하고, 이와 유사한 객체를 다른 데이터세트에서 찾아 추천해주는 자동화 시스템을 사용해야 할 것이다.

앞으로 이런 LOD 환경이 점차 확대되어, 상호연결이 고도화되어간다면, 소장하고 있는 기록정보가 없으면서도 공개된 LOD 데이터들을 가져와 주제별로 자신만의 아카이브를 구축하는 형태도 등장할 수 있을 것이라 예상된다. 이러한 형태의 아카이브는 실질적으로 데이터를 소유하지는 않지만, 여러 LOD 데이터의 연결을 통해서 사람들에게 흥미롭고 새로운 접근점을 제공할 수 있을 것이다. 이렇게 LOD의 상호연결이 고도화된다면, 데이터 큐레이션 형태의 아카이브를 통해 새로운 의미를 가진 데이터가 생성되고 제공될 수 있을 것이다.

4.4 LOD 서비스 수준의 결정

기록정보의 LOD를 구축한 다음에는, LOD를 서비스하는 수준에 대해서도 고려해봐야 한다. 링크드 데이터를 발행하여 서비스하는 첫 번째 방법은 RDF파일을 웹에 업로드하여 공개하는 것이다. 공공데이터포털의 데이터세트 중 상당수가 이러한 기본적인 형태로 발행되고 있었다. 하지만 이것은 이용자가 웹을 통해서 데이터를 연결할 수 없으므로 진정한 의미에서 LOD는 아니라고 생각된다.

두 번째는 본 연구에서처럼 RDF 저장소에 RDF 데이터를 저장하여 검색과 활용이 가능하도록 SPARQL Endpoint를 제공하는 것이다. 본 연구에서는 앞에서 살펴보았듯이 RDF/XML 형태의 OWL문서를 Jena Fuseki Server를 통해 RDF 저장소에 업로드하였고, SPARQL 질의가 가능한 상태로 발행하였다.

세 번째는 이용자가 LOD 데이터에 대하여 친화적으로 접근할 수 있도록 웹페이지를 제작하여 발행한 LOD를 서비스하는 것이다. 본 연구에서 살펴본 한국사 LOD와 한국관광공사 LOD가 이런 형태인데, SPARQL 검색뿐만 아니라 기존의 키워드 검색처럼 검색이 가능하고, 해당 객체에 대한 상세내용을 보여주는 HTML 웹페이지를 제공한다. 하지만 보통 이런 식으로 서비스하는 LOD의 웹페이지 URL은 객체가 가진 고유한 URI와 다른 경우가 많다. 따라서 상호연결 시에는 이 점을 주의하여 객체의 URI를 찾아 연결해야 한다.

마지막으로 시각화 소프트웨어를 사용하거나 OPEN API 형태로 LOD 데이터에 대한 이용과 접근을 제공할 수 있다. 이러한 서비스 방법들은 LOD 구축과는 별개로 프로그래밍 작업이 필요하기 때문에, 아카이브의 입장에서는 개발을 위한 회의가 필요한 사항이다.

본 연구에서는 웹 서버를 구축하여 LOD를 서비스하는 단계까지는 진행하지 않았다. 하지만 이미 웹페이지를 통해 인터넷 서비스를 제공하고 있는 아카이브라면 서버에 RDF 저장소를 설치하여 LOD서비스를 제공할 수 있다. 또한 아마존 웹 서비스와 같이 클라우드 컴퓨팅 서비스를 활용하여 서버를 구축하고 서비스를 제공할 수 있을 것이다. LOD를 어떤 수준으로 서비

스하느냐는 아카이브의 예산과 기술력, LOD의 범위와 규모 등을 고려하여 결정해야 한다.

4.5 LOD 구축을 위한 도구 선정

본 연구에서는 오픈소스 소프트웨어를 사용하여 구체적인 LOD 구축 절차를 진행했다. LOD를 구축하는 방법은 여러 가지가 있는데, 그 중에서 오픈소스 소프트웨어를 활용하여 구축하는 방법을 연구하게 된 이유는, LOD 구축에 큰 예산을 집행하지 못하는 소규모 아카이브에서 지속가능성을 가지고 LOD 서비스를 진행하는데에 이 방법이 최선이라고 판단했기 때문이다. 하지만 오픈소스 소프트웨어의 문제점은 기능의 제약이 많다는 점이다. 예를 들어 Protege의 Individual 메뉴에서는 생성된 객체들에 대해서 검색기능을 지원하고 있지 않아서, 생성된 객체의 수가 많아진다면 탐색하기가 어려울 수도 있다. 또한 Prefix로 선언한 어휘들을 사용할 때 용어의 정확한 명칭을 사용해야 하는데, 웹에서 해당 어휘들을 정의하고 있는 페이지를 참조하여 입력하는 과정에서 어휘의 철자가 틀리는 오류가 발생할 수도 있다. 이러한 한계점들을 상용 소프트웨어들은 어느 정도 보완하고 있기 때문에, 예산의 여유가 있는 아카이브의 경우에는 본 연구와 같이 오픈소스 소프트웨어를 선택하지 않고, 기능과 편의성을 고려하여 LOD 구축 도구를 선정할 수 있을 것이다. LOD 구축과 관련하여 가장 많이 사용되는 상용 소프트웨어는 TopQuadrant's사의 TopBraid Composer이다. TopBraid Composer는 온톨로지 모델링에서부터, 추론 기능, 데이터 업로드, SPARQL Query, Localhost 서버를 통한 SPARQL Endpoint 제공,

웹 페이지 발행까지 본 연구에서 진행한 모든 과정을 하나의 프로그램 안에서 해결할 수 있는 강력한 소프트웨어이다. 따라서 아카이브 자체 내에서 LOD를 구축할 수 있는 기술력을 가진 인력이 있고 LOD의 구축 범위가 넓지 않으며 예산 집행이 가능하다면, TopBraid Composer를 사용하여 자관의 LOD를 구축하는 것도 고려해볼만하다. 또한 공공기관의 기록관 같은 경우, LOD 구축 사업으로 진행할 수 있는 예산을 확보할 수 있다면 LOD 구축 사업을 공고하여 전문기업을 통해 구축하는 것도 하나의 방법이다.

외부 업체를 통해서 LOD를 구축을 진행한다면, 아카이브는 먼저 LOD 구축을 위한 준비를 갖춰놓고, LOD 구축의 노하우를 가진 업체와 협의를 통해서 완성도 높은 LOD 구축을 지향해야 할 것이다. 이를 위해서 본 연구에서 진행한 POC의 도구와 절차, 방법 등을 참고하여 프로토타이핑 식으로 LOD를 구축해보면서, 기록정보의 기술에 대한 품질 개선 방향이나 LOD 구축 시 요구사항 등을 정리하여 업체와 원활하게 협의해야 할 것이다.

5. 결론

링크드 오픈 데이터(Linked Open Data)는 데이터의 의미 연결을 통해서, 인지하지 못한 영역까지 무한한 지식 브라우징의 기회를 제공한다. 이러한 기회는 서로 다른 영역의 지식들이 복합되어 창조적인 콘텐츠 서비스를 창출해내고, 새로운 형태의 아카이브를 출현시킬 수 있을 것이다. 본 연구는 새로운 웹 환경에 부응하는, 기록정보의 LOD 구축을 위한 절차와 방법에 대

해서 연구하였다. 또한 소규모 아카이브의 서비스 지속가능성을 염두에 두고 오픈소스 소프트웨어를 활용하여 LOD 구축 절차를 진행하는 방법을 제시하고자 하였다.

이에 본 연구는 LOD 구축절차로 대상 기록 정보 수집 및 분석, 온톨로지 구축, 객체 생성 및 상호연결, 데이터 저장 및 발행, 응용 서비스 등 총 5단계의 LOD 구축절차를 제안하였다. 그리고 제안한 5단계 절차의 구체적인 내용과 방법을 설명하기 위하여 인간과 기억 아카이브의 일기 기록물을 수집하였고, 오픈소스 소프트웨어인 Protege와 Apache Jena Fuseki를 활용하여 POC(Proof of concept)를 진행하였다. 그리고 상호연결(Interlinking)과 SPARQL 검색을 통해서 본 연구에서 구축한 LOD와 외부 LOD가 연결되고 확장되는 모습을 확인하였다.

또한 본 연구는 기록정보의 LOD 구축 절차를 진행하면서 얻은 경험을 바탕으로 기록관 LOD 구축의 필요요건들을 제시하였다. 먼저 LOD 환경에서 풍성한 상호연결이 가능해지기 위해서는 기존의 맥락과 구조정보를 중요시하던 현황에서 벗어나, 기록의 내용정보에 대한 기술의 품질 향상이 필요함을 확인하였다. 그리고 LOD 구축에 있어서 아카이브가 LOD에 대한 지식을 함양하여 LOD 구축에 있어서 주도적인 역할을 하며, 기록의 생산단계에서 내용정보에 대한 기술이 이뤄질 수 있도록 노력해야 함을 제시하였다. 또 공공데이터포털 같은 LOD 데이터 허브 사이트를 통해서 관련 링크와 동질 링크, 어휘 링크와 같은 상호연결을 풍성하게 정의하고, 이러한 상호연결의 고도화가 데이터 큐레이션 형태의 새로운 아카이브 출현에 기여할 수 있을 것이라고 거론하였다. 또한 아카이

브의 성격, 기술력, 예산, LOD 구축 대상의 규모와 범위 등을 고려하여 LOD 서비스 수준을 결정하고, LOD 구축 도구를 선정해야 함을 언급하였다.

하지만 본 연구는 앞서도 언급하였듯이 내부 웹 서버환경에서 연구를 진행했기 때문에 기록정보의 LOD가 확장되고 연결되는 실제적인 서비스의 사례를 보여주지 못했다. 따라서 본 연구의 LOD 구축결과만으로 기록관 LOD 구축의 필요요건을 확인하고 논의하기엔 제한점이 있을 수 있다. 후속연구에서는 이러한 한계점을 보완하여 기록정보의 연결과 확장을 확인할 수 있는 풍부한 사례들이 연구되고 제시되기를 희망한다.

최근 Omeka에서 새로운 버전인 Omeka S (Beta)를 선보였다. 이 새로운 버전은 관리하는 자원(Resource)을 누구나 쉽게 LOD로 공유할 수 있도록 다양한 기능을 제공한다. 다른 분야의 데이터들과 달리 기록정보는 아직까지 LOD로 구축된 시범사례가 많지 않다. 그런데 많은 아카이브, 도서관, 박물관에서는 이미 오픈소스 웹 전시 플랫폼인 Omeka를 사용 중이다. 만약 Omeka S가 성공적으로 정착된다면 아카이브에서 Omeka를 통해 관리하고 있던 기록정보들이 비교적 쉽고 저렴하게 LOD로 구축될 수 있을 것이다.

이렇게 기록정보의 더 많은 공유와 개방이

이뤄질 수 있는 환경들이 조성되고 있다. 그렇지만 이미 저장된 대량의 기록에서, 그것도 기능분류체계 아래에서 맥락을 중심으로 기술되어 있는 수많은 기록에서 내용정보를 새롭게 기술하는 것은 지난한 문제이다. 하지만 본 연구에서 계속 강조했던 것처럼 기록의 내용정보는 LOD로 구축을 위해서 반드시 필요한 부분이다. 따라서 앞으로는 대량의 기록정보에서 내용정보를 추출해낼 수 있는, 빅데이터 분석에 대한 학술적인 연구가 필요할 것이다. 비전자문서라도 OCR(Optical Character Recognition)을 통해서 전문을 디지털화한 후, Text Mining과 같은 빅데이터 기법을 통해서 내용분석(Content Analysis)을 하게 되면, 대량의 기록정보에 대한 내용정보를 비교적 손쉽게 추출하여, 기록정보의 LOD 구축에 큰 영향을 끼칠 것이라 예상된다.

앞으로 여러 기록관의 많은 기록물들이 LOD를 통해 외부의 다양한 데이터와 연결되어, 많은 이용자들이 지금까지 확인하기 어려웠던 기록물들을 쉽게 확인할 수 있기를 기대한다. 또한 다양한 정보들과의 연계를 통해 기록 속에 숨어있는 보물과 같은 정보가 사회 구성원들에게 발견되어 의미있는 역할들을 하기 바라며, 기록관이 앞으로 명실공히 기록의 보존 및 정보 제공 기관으로 사회에서 인식되고 그 본분을 이행해 내기도 기대한다.

참 고 문 헌

- 곽건홍 (2011). 일상 아카이브(Archives of everyday life)로의 패러다임 전환을 위한 소론. 기록학연구, 29, 3-33.

- 김성혁 (2011). 도서관 자료 공개와 활용을 위한 링크드 데이터 기술. 국회도서관보, 48(3), 24-33.
- 김영천 (2011). 온톨로지를 이용한 시맨틱 웹 기술의 사례연구. 석사학위논문, 건국대학교 정보통신대학원, 정보통신학과.
- 김은경 (2004). 시맨틱웹을 위한 온톨로지 구축방법에 관한 비교연구. 석사학위논문, 중앙대학교 대학원, 문헌정보학과.
- 노상규, 박진수 (2007). 온톨로지. 서울: 가즈토이.
- 박옥남 (2012). 기록물 전자통제 기반 Linked data 구축에 대한 연구. 한국비블리아학회지, 23(2), 5-25. <https://doi.org/10.14699/kbiblia.2012.23.2.005>
- 박중호 (2011). 온톨로지(Ontology) 기반 한국어 동사 의미망 구축 방법 연구. 박사학위논문, 청주대학교 대학원, 국어국문학과.
- 박현근 (2005). Owl 시맨틱웹 기반 온톨로지 상에서의 규칙 - 사실 생성에 관한 추론. 박사학위논문, 중앙대학교 대학원, 컴퓨터공학과.
- 안대진, 김익한 (2015). AtoM 시스템의 구축과 커스터마이징 방법에 관한 연구. 기록학연구, 45, 5-50.
- 안세열 (2011). 온톨로지 기반 지식 검색 시스템 개발. 석사학위논문, 고려대학교 컴퓨터정보통신대학원, 디지털정보·미디어공학전공.
- 이병길, 김희섭 (2013). 새마을운동 기록물의 개체기반 온톨로지 검색시스템 설계 및 평가. 한국기록관리학회지, 13(3), 67-97. <https://doi.org/10.14404/jksarm.2013.13.3.067>
- 이유빈, 이해영 (2017). 온톨로지 기반의 기록물 검색 시스템을 위한 인터페이스 제안. 한국기록관리학회지, 17(1), 217-244.
- 이정현, 이운용, 방기영, 김용 (2015). Linked open data기반 대통령기록관 기록정보 서비스 확장에 관한 연구. 한국기록관리학회지, 15(2), 55-82. <https://doi.org/10.14404/jksarm.2015.15.2.055>
- 이현정 (2015). 링크드 데이터 활용을 위한 멀티소스 검색시스템 구축에 관한 연구. 박사학위논문, 중앙대학교 대학원, 문헌정보학과.
- 정갑주 (2015). JENA ARQ and FUSEKI Basics. Retrieved from <https://www.youtube.com/watch?v=8F3TWWJvgmBU>
- 최효진, 임진희 (2015). 개인 일기의 연구 자료로서의 가치와 전망. 기록학연구, 46, 95-152.
- 하상범, 한은영, 최호준 (2005). Owl 기반의 sparql을 이용한 시맨틱 검색. 한국정보과학회 학술발표논문집, 32(2), 706-708.
- 한국정보화진흥원 지식자원활용부 (2014a). 2014 링크드오픈데이터 국내 구축 사례집. 서울: 한국정보화진흥원.
- 한국정보화진흥원 지식자원활용부 (2014b). 링크드 데이터 구축 공정 가이드 v1.0. 서울: 한국정보화진흥원.
- 한상은 (2014). 오픈데이터의 링크드데이터(Linked data) 발행방법 간 유용성 비교평가에 관한 연구.

석사학위논문, 성균관대학교 대학원, 문헌정보학과.

Antoniou, G., & Van Harmelen, F. (2008). *A semantic web primer*. Cambridge, Mass.: MIT Press.

Bauer, F., & Kaltenböck, M. (2011). *Linked open data: The essentials*. Retrieved from <https://www.semantic-web.at/LOD-TheEssentials.pdf>

Berners-Lee, T. (2006). *Linked data*. Retrieved from <https://www.w3.org/DesignIssues/LinkedData.html>

Berners-Lee, T. (2009). *The next web*. Retrieved from https://www.ted.com/talks/tim_berners_lee_on_the_next_web

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web, *Scientific American*, 284(5), 28-37.

Gracy, K. F. (2015). Archival description and linked data: A preliminary study of opportunities and implementation challenges. *Archival Science*, 15(3), 239-294.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199-220.

Josef, H. (2016). *Cellfie-plugin user's guide*. Retrieved from <https://github.com/protegeproject/cellfie-plugin/wiki/User%27s-Guide>

Heath, T., & Bizer, C. (2011). *Linked data: Evolving the Web into a global data space*. San Rafael, Calif.: Morgan & Claypool.

Noy, N. F., & McGuinness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology*. Retrieved from <http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf>

O'Connor, M. J., Halaschek-Wiener, C., & Musen, M. A. (2010). *Mapping master: A flexible approach for mapping spreadsheets to OWL*. In Patel-Schneider P.F. et al. (Eds.). *The semantic Web - ISWC 2010*. ISWC 2010. Lecture Notes in Computer Science, vol 6497. Springer, Berlin, Heidelberg

공공데이터포털 홈페이지. Retrieved from <https://www.data.go.kr>

인간과기억아카이브 홈페이지. Retrieved from <http://hmarchives.org/>

주소데이터LOD 제주도 웹페이지. Retrieved from <http://jusodata.kr/resource/Jeju-do>

한국관광공사LOD 제주도 웹페이지. Retrieved from <http://data.visitkorea.or.kr/resource/CATEGORY:390000>

한국사LOD 홈페이지. Retrieved from <http://lod.koreanhistory.or.kr/lodIntro.do>

DBPedia 제주도 웹페이지. Retrieved from <http://ko.dbpedia.org/page/제주도>

KData 제주도 웹페이지. Retrieved from http://data.kdata.kr/administrative_section/Jeju

Mass Observation 프로젝트 홈페이지. Retrieved from
<http://www.massobs.org.uk/write-for-us/12th-may>
Omeka Homepage. Retrieved from <http://omeka.org>

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

- An, Dae-Jin, & Kim, Ik-Han (2015). A study on constructing and customizing an AtoM system. *The Korean Journal of Archival Studies*, 45, 5-50.
- An, Seyeol (2011). Development of an ontology-based knowledge search system: The case of KT call center. Master's thesis, Graduate School of Computer & Information Technology at Korea University.
- Bak, Jong-Ho (2011). The study of establishing a method for semantic network of Korean verb based on ontology. Ph.D. dissertation, Graduate School of Cheongju University, Department of Korean Literature.
- Choi, Hyo Jin, & Yim, Jin Hee (2015). Value and prosect of individual diary as research materials: Based on "The 12th May Diaries Collection." *The Korean Journal of Archival Studies*, 46, 95-152.
- Ha, Sangbum, Han, Eunyoung, & Choi, Hojun (2005). OWL-based semantic search using SPARQL. *Korean Institute of Information Scientists and Engineers Workshop Presentation File*, 32(2), 706-708.
- Han, Sang Eun (2014). Evaluating usability of linked data publishing methods on open data: Focusing on the RDF triple store and RDB-to-RDF LD publishing methods. Master's thesis, Graduate School of Sungkyunkwan University, Department of Library and Information Science.
- Jeong, Gab-Ju (2015). JENA ARQ and FUSEKI Basics. Retrieved from <https://www.youtube.com/watch?v=8F3TWJvgmBU>
- Kim, Eunkyong (2004). A comparative study on methodologies of building ontologies towards semantic web. Master's thesis, Graduate School of Chung-Ang University, Department of Library and Information Science.
- Kim, Sung-Hyuk (2011). Linked data technologies for disclosing and utilizing library materials. *National Assembly Library Review*, 48(3), 24-33.
- Kim, Young Cheon (2011). A case study on the semantic web technology using ontology.

- Master's thesis, Graduate School of Konkuk University, Department of Information and Communications.
- Kwak, Kun-Hong (2011). A study on a paradigm shift to archives of everyday life. *The Korean Journal of Archival Studies*, 29, 3-33.
- Lee, Byung Gil, & Kim, Heesop (2013). Design and evaluation of an individual instance-based ontology retrieval system for archival records of the "saemaul movement." *Journal of Korean Society of Archives and Records Management*, 13(3), 67-97.
<https://doi.org/10.14404/jksarm.2013.13.3.067>
- Lee, Jeong Hyeon, Lee, Youn Yong, Bang, Ki Young, & Kim, Yong (2015). A study on the extension of archival information service based on linked open data in the presidential archives. *Journal of Korean Society of Archives and Records Management*, 15(2), 55-82.
<https://doi.org/10.14404/jksarm.2015.15.2.055>
- Lee, Yu-Been, & Rieh, Hae-Young (2017). A suggestion of interface for ontology-based record retrieval system. *Journal of Korean Society of Archives and Records Management*, 17(1), 217-244.
- National Information Society Agency (2014a). 2014 linked open data internal build case. Seoul: National Information Society Agency.
- National Information Society Agency (2014b). Linked data build guide v1.0. Seoul: National Information Society Agency.
- Park, Hyeongeun (2005). Reasoning on production of rules and facts using OWL-based semantic web ontology. Ph.D. dissertation, Graduate School of Chung-Ang University, Department of Computer Science and Engineering.
- Park, Ok Nam (2012). The design and development of linked data from authority data in national archives of Korea. *Journal of the Korean Biblia Society for Library and Information Science*, 23(2), 5-25. <https://doi.org/10.14699/kbiblia.2012.23.2.005>
- Rho, Sangkyu, & Park, Jinsoo (2007). *Ontology*. Seoul: God's Toy Business.
- Yi, Hyun Jung (2015). A study on the implementation of multi-source search system for linked data utilization. Ph.D. dissertation, Graduate School of Chung-Ang University, Department of Library and Information Science.

