

1997 외환위기 지식그래프: 디지털 아카이브의 관계 중심적 접근

A Knowledge Graph of the Korean Financial Crisis of 1997: A Relationship-Oriented Approach to Digital Archives

이유경(Yu-kyeong Lee)¹, 김학래(Haklae Kim)²

E-mail: yuyu506@cau.ac.kr, haklaekim@cau.ac.kr



¹ 제 1저자 중앙대학교 일반대학원 문헌정보학과 기록관리학전공 석사과정
² 교신저자 중앙대학교 사회과학대학 문헌정보학과 교수

논문접수 2020-10-21
최초심사 2020-10-21
게재확정 2020-11-09

ORCID

Yu-kyeong Lee
<https://orcid.org/0000-0002-3983-8869>
Haklae Kim
<https://orcid.org/0000-0002-2616-421X>

© 한국기록관리학회

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

• 이 논문은 2019년도 중앙대학교 CAU GRS 지원에 의하여 작성되었음.

초 록

정보기술의 발전에 따라 아카이브의 디지털화가 가속화되고 있다. 그런데 전통적인 방식의 디지털 아카이브는 기록을 효과적으로 검색하고 연계하고 이해하는 데 한계가 있다. 본 논문은 디지털 아카이브의 활용성을 극대화하기 위한 방안으로 관계 중심의 지식그래프 방식을 제안한다. 디지털 아카이브의 사례인 ‘1997 외환위기 아카이브’의 특징을 검토하고, 아카이브에 포함된 모든 개체와 개체 사이의 관계는 RiC-O(Records in Contexts-Ontology) 기반의 지식그래프로 구축한다. 본 연구의 결과인 외환위기 지식그래프는 1997 외환위기 아카이브의 모든 개체를 기계가 처리할 수 있는 형식으로 구축한다. 디지털 아카이브와 비교해 지식그래프 접근은 개체의 정보, 개체 사이의 관계를 정확히 탐색할 수 있고, 이를 통해 의미검색, 지능형 서비스에 활용될 수 있다.

ABSTRACT

Along with the development of information technology, the digitalization of archives has also been accelerating. However, digital archives have limitations in effectively searching, interlinking, and understanding records. In response to these issues, this study proposes a knowledge graph that represents comprehensive relationships among heterogeneous entities in digital archives. In this case, the knowledge graph organizes resources in the archives on the Korean financial crisis of 1997 by transforming them into named entities that can be discovered by machines. In particular, the study investigates and creates an overview of the characteristics of the archives on the Korean financial crisis as a digital archive. All resources on the archives are described as entities that have relationships with other entities using semantic vocabularies, such as Records in Contexts-Ontology (RiC-O). Moreover, the knowledge graph of the Korean Financial Crisis of 1997 is represented by resource description framework (RDF) vocabularies, a machine-readable format. Compared to conventional digital archives, the knowledge graph enables users to retrieve a specific entity with its semantic information and discover its relationships with other entities. As a result, the knowledge graph can be used for semantic search and various intelligent services.

Keywords: 디지털 아카이브, 1997 외환위기, 지식그래프, RiC-O

Digital archives, Korean financial crisis of 1997, Knowledge graph, RiC-O

1. 서론

아카이브(archives)는 장기 보존의 가치를 가진 기록을 의미하며, 때론 기록을 보관하는 기관이나 장소를 뜻하기도 한다(Pearce-Moses, 2005, p. 30). 물리적 공간에 기반한 전통적 아카이브는 기록의 보존이 핵심이며, 이용자의 자유로운 접근에 제한적이다(김유승, 2010, p. 35). 반면, 최근 기록이 전산화되어 보존됨에 따라 디지털 방식의 아카이브에 대한 중요성이 증가하고 있다. 디지털 아카이브에서 이용자는 시공간의 제약 없이 아카이브에 접근하여 자유롭게 기록을 검색하고 열람할 수 있다(이병길, 김희섭, 2013, pp. 68-69).

그러나 현재 서비스되고 있는 디지털 아카이브는 기능적 측면에서 여전히 제약이 있다. 첫째, 검색이 제한적이다. 대부분의 디지털 아카이브는 키워드 기반 검색을 지원한다. 단편적인 검색어의 활용은 이용자의 검색 의도를 정확하게 반영하는 데 어려움이 있다(이정희, 김희섭, 2007, p. 344). 기록의 맥락을 둘러싼 정보를 입체적으로 파악하려면 다양한 검색의 조합이 필요할 수 있지만, 일반 이용자가 활용 방법을 효과적으로 이해하기 쉽지 않을 수 있다(안주현, 2012, p. 422). 둘째, 기록의 연계가 제한적이다. 관련 있는 기록이 여러 디지털 아카이브에 분산되어 존재하면 통합적으로 연계해 제공되어야 한다(서해숙, 2013, p. 238). 하지만 현재 서비스되고 있는 디지털 아카이브는 기관마다 메타데이터 스키마가 다르고, 기록의 연계를 고려하지 않은 경우가 많다(현문수, 2013, p. 135). 마지막으로, 기록의 이해가 제한적이다. 디지털 아카이브는 기본적으로 기록을 원자료 형태로 제공한다. 일반 이용자가 가공되지 않은 전문적인 기록의 맥락을 이해하거나 해석하는 것은 어려움이 있다(박진경, 김유승, 2017, p. 239).

지식그래프(knowledge graph)는 개념, 사물, 사건과 같은 의미 있는 개체의 상호 연결된 정보를 표현하기 위한 기술이다. 지식그래프는 의미를 갖고 있는 개체와 개체 사이의 관계를 맥락으로 표현하고, 기계가 처리할 수 있는 인공 언어로 표현할 수 있다. 예를 들어, ‘일어나다’의 의미는 특정한 사건의 발생, 또는 어떤 개체의 움직임이나 상태로 중의적으로 해석된다. 이때, 특정 개체가 명시적으로 선언되면, ‘일어나다’의 의미는 명확해질 수 있다. 즉, ‘외환위기’와 ‘일어나다’가 함께 사용되는 경우 ‘특정한 사건이 발생한다’라는 맥락의 이해가 가능하다. 지식 그래프는 지식 모델을 통해 개체 사이의 관계를 표현하고, 웹 온톨로지 언어와 같은 인공 언어를 사용하여 지식 모델을 기계가 처리할 수 있도록 한다. 디지털 아카이브에 지식그래프 적용은 의미 검색, 관계 기반의 분석을 위한 방안으로 고려할 수 있다.

본 연구는 ‘1997 외환위기 아카이브(이하 외환위기 아카이브)’¹⁾의 특성을 분석하고 개선 방안을 검토한다. 1997년에 촉발된 외환위기(IMF 위기, 이하 ‘외환위기’)는 금융 및 기업의 대규모 구조조정, 대량 실업과 노동 유연화 등 한국 사회에 큰 변화를 만든 사건이다(지주형, 2011, p. 444). 외환위기 아카이브는 외환위기와 관련된 다양한 기록을 수집하고 공개하는 디지털 아카이브로 온라인에서 관련 기록을 열람할 수 있고, 전체 데이터는 개방된다. 그러나 기록학적 가치에 관계없이, 외환위기 아카이브는 이용자가 기록을 효과적으로 탐색하고 활용하는 데 제약이 있다.

본 연구는 외환위기 아카이브를 개선하기 위해 지식그래프 기술을 적용한다. 외환위기 아카이브에 기술된 자료는 RiC-O(Records in Contexts-Ontology) 모델을 통해 기계가 처리할 수 있는 형식으로 구축한다. 외환위기와 관련된 기록, 인물, 조직, 사건 등을 식별할 수 있는 개체로 정의하고, 개별 개체는 URI(Uniform Resource Identifier)를 부여한다. 논문의 구성은 다음과 같다. 2장은 지식그래프와 Records in Contexts 관련 연구를 소개한다. 3장은 1997 외환위기 아카이브를 소개하고, 기능적 측면에서 한계점을 분석한다. 4장은 데이터 수집과 지식그래프 구축 방법을 설명한다. 5장은 지식그래프의 활용 사례를 소개하고, 6장에서 연구 결과를 정리한다.

1) 1997 외환위기 아카이브. 검색일자: 2020. 11. 10. <https://97imf.kr/>

2. 관련 연구

2.1 지식그래프

‘지식그래프’는 구글의 검색서비스 브랜드명에서 비롯되어, 최근에는 링크드 데이터와 그래프 형식의 지식베이스를 통칭하는 용어로 사용되고 있다(김학래, 2017, p. 78). 링크드 데이터(Linked Data)란 웹에 분산된 데이터를 효과적으로 공유하고 통합하기 위하여 데이터를 의미 수준에서 연결하는 기술이다(김학래, 2017, pp. 49-52). 이를 위해 모든 개체는 URI로 식별하고 HTTP 프로토콜로 접근할 수 있도록 한다. 개체에 대한 정보는 주어(Subject)-술어(Predicate)-목적어(Object)의 진술문(statement) 형식으로 구성된 RDF (Resource Description Framework) 그래프 구조로 기술한다(이정현, 이윤용, 방기영, 김용, 2015, p. 59). 개체에 대한 정보를 실질적으로 기술하기 위해서는 어휘(vocabulary)가 필요하며, 분야별로 공유되는 개념을 사람과 기계 모두에게 명확하도록 정의한 온톨로지(ontology)를 활용한다(김학래, 2017, p. 37). 이렇게 발행된 데이터는 의미 수준에서 기계의 처리가 가능해지고, 의미적으로 연관된 또 다른 데이터와 그래프로 연결되고 통합되어 지식베이스를 구성한다. 다시 말해, 지식그래프는 의미 검색(semantic search)의 출발점이 될 수 있고, 서로 다른 기관에서 발행된 데이터 간 연계와 통합을 지원할 수 있다.

링크드 데이터는 기록자료의 표현과 관리를 위한 중요한 기술로 인식되고 있다(Gracy, 2015). 기록자료의 표현은 다양한 방식으로 제안되고 있고, 디지털 문화유산 및 기록관리 분야에 적용되고 있다(Ziku, 2020). Matienzo, Roke, & Carlson(2017)은 디지털 아카이브의 기술(description)을 위해 Schema.org²⁾의 적용과 확장 방법을 설명하고, Schema 어휘와 ISAD(G), ISAAR-CPF, DACS 사이의 기초적인 매핑 관계를 제안하고 있다. Mazzini와 Ricci(2011)는 아카이브 데이터의 표현을 위한 EAC-CPF 온톨로지 모델을 제안하고 있다. 데이터 모델은 기록자료를 표현하기 위한 90개 클래스와 30개 속성을 포함하고 있지만, 기존 어휘의 재사용과 연계 방안을 소개하고 있지는 않다.

국내 기록학계에서 지식그래프 구축과 관련된 연구는 의미 검색의 성능 평가, 온톨로지 모델의 설계, 링크드 데이터의 상호연결(interlinking)을 주요 키워드로 전개되어왔다. 먼저 이정희, 김희섭(2007)과 이병길, 김희섭(2013)은 각각 대학 전자기록물과 새마을운동 기록물을 대상으로 시맨틱 검색시스템을 구축하고 성능을 평가한다. 두 연구는 키워드 기반 검색시스템에 비해 재현율 및 정확률이 향상되었음을 보고하며 의미 검색의 우수성을 확인한다. 위 선행연구는 의미 검색을 통해 디지털 아카이브의 검색기능을 향상할 수 있음을 보여준다.

온톨로지 모델 설계에 중점을 둔 연구는 다음과 같다. 이정현 외(2015)는 대통령기록물과 대통령 재임 중 사건을 연계할 수 있는 온톨로지를 설계하고, 이를 LOD(Linked Open Data; 링크드 오픈 데이터)로 발행하기 위한 서비스 흐름도 및 웹 인터페이스 화면을 제안한다. 이은옥, 박희진(2017)은 패킷 분석을 바탕으로 ‘이산가족 찾기 기록’의 온톨로지를 제안하고 모델의 품질을 평가한다. 최진경(2020)은 과학기술성과를 대상으로 온톨로지를 설계하고 웹 구현 화면의 프로토타입을 소개하고 있다. 위 연구들은 특정 분야 기록에 특화된 온톨로지 어휘를 제시한 의의가 크지만, 상호운용성의 측면에서 약점을 지닐 우려가 있다. 본 연구는 지식그래프 구축에 기존 온톨로지 어휘를 재사용하며, 특히 기록학계의 국제 표준으로 개발 중인 Records in Contexts-Ontology를 활용하여 디지털 아카이브 간의 범용성 확보를 도모한다.

링크드 데이터의 상호연결에 중점을 둔 연구는 다음과 같다. 박옥남(2012)은 국가기록원의 계층별, 주제별, 생산기관별 검색이 연계되지 않는 경직성을 지적하고, 주제 및 생산기관 전거데이터를 링크드 데이터로 구축하여 기록물 데이터와 상호연결한다. 하승록, 임진희, 이혜영(2017)은 기록정보 LOD 구축 프레임워크를 5단계로 제안하고 개념 증명(Proof of Concept)을 진행하는 과정에서, ‘5월 12일 일기 컬렉션’을 LOD로 구축하고 한국관광공사LOD

2) Schema.org. Retrieved November 10, 2020, from <http://schema.org>

서비스와 상호연결한다. 하승록, 안대진, 임진희(2017)는 대규모 상호연결을 위해 자동화의 불가피성을 강조하며, 국가기록원 중앙영구기록관리시스템(CAMS) 데이터 일부를 LOD로 구축하고 디비피디아(DBpedia)와의 상호연결 자동화 실험을 진행한다. 위 연구들은 디지털 아카이브가 고립되지 않고 외부 데이터와 연계될 때 그 가치가 더욱 높아짐을 보여주고 있다. 선행연구의 연장선상에서 본 연구는 지식그래프를 위키데이터(Wikidata)³⁾와 자동화된 방식으로 상호연결하며, 나아가 연결된 외부 데이터를 재사용하여 데이터 강화(enrichment)를 시도한다.

2.2 Records in Contexts

Records in Contexts(이하 'RiC')는 기록을 기술(describe)하기 위한 국제 표준이다. RiC은 크게 세 부분으로 구성된다. 첫째, 기록을 기술하는 데 필요한 개체(entity), 속성(attribute), 관계(relation)를 정의하는 개념 모형 RiC-Conceptual Model(이하 'RiC-CM')이다. 둘째, RiC-CM을 공식적, 기술적(technical)으로 표현하기 위한 온톨로지 RiC-Ontology(이하 'RiC-O')이다. RiC-CM이 정의한 원칙에 따라 기록을 기술하여 RDF 그래프 형태로 구현할 수 있도록 어휘와 규칙을 제공한다. 셋째, 적용을 위한 가이드라인 RiC-Application Guidelines (RiC-AG)이다(Clavaud, 2020, pp. 16-19). 이 중 RiC-CM v0.2의 프리뷰와 RiC-O v0.1이 2019년 12월 발표되었다. 차후 개발 일정은 ICA EGAD RiC-O 공식 웹사이트에 업데이트되고 있으며 2020년에 RiC-O v0.2가, 2021년에 RiC-CM과 RiC-O의 v1.0이 공개될 예정이다(ICA EGAD, 2020a).

RiC-CM의 개체는 4계층으로 구분되는데, '기록자원(Record Resource)', '사본(Instantiation)', '행위자(Agent)', '활동(Activity)'을 비롯하여 총 22개 개체를 포함한다. 개체는 속성을 지니는데, 상위 개체의 속성은 하위 개체에 그대로 상속된다. 예를 들어 최상위 개체 '대상(Thing)'의 속성인 '식별자(Identifier)', '이름(Name)', '기술 주기(Descriptive Note)'는 모든 개체가 공통적으로 지닌다(ICA EGAD, 2019b, p. 21). 개체와 개체 사이의 관계는 5계층으로 구분되며, 최상위 관계 '~와 관련이 있다(is related to)'에서 시작하여 하위 계층으로 갈수록 좁은 의미를 지니게 된다(ICA EGAD, 2019b, p. 53).

이와 같은 구조 속에서, '기록(Record)' 및 '기록집합(Record Set)' 개체는 다양한 개체들과 관계로 연결되어 그래프 또는 네트워크를 형성하며 다차원적(multidimensional)으로 기술된다. 기록이 생산, 이용되는 실제 세계는 시공간에 위치한 다양한 사람들과 객체들이 역동적으로 상호작용하는 네트워크의 일종으로 이해할 수 있다. 따라서 다차원적 기술 방식은 실제 세계의 복잡한 맥락을 반영하는 데에 이점을 지닌다(ICA EGAD, 2016, pp. 9-10).

RiC-O는 RiC-CM이 정의한 개체, 속성, 관계를 기계가 처리 가능한 형식으로 표현한다. RiC-O는 OWL(Web Ontology Language; 웹 온톨로지 언어) 기반의 온톨로지로서, '클래스(owl:Class)', '데이터타입 속성(owl:DatatypeProperty)', '오브젝트 속성(owl:ObjectProperty)'을 지니고, 접두어(prefix)로는 'rico'를 사용한다. RiC-O의 클래스 및 속성은 대체로 RiC-CM에 상응하나, 모델의 정확성과 유연성을 위해 일부 클래스 및 속성이 추가되기도 하였다(ICA EGAD, 2019a). RiC-O를 통해 RDF 데이터를 발행하면, 기록학 관점에서 명확하게 표현된 어휘를 활용하여 지식 그래프를 구축할 수 있다. 이를 통해 RiC-CM이 표방한 다차원 기술을 RDF 그래프로 구현하여, 기록의 맥락을 효과적으로 표현하고 기록에 대한 이해의 깊이를 더할 수 있다.

RiC을 활용한 지식그래프 구축 프로젝트로, ANF(Archives nationales de France; 프랑스국립아카이브)는 전거레코드를 포함하여 소장 중인 기록 메타데이터(archival metadata)를 RiC-O를 준수하는 RDF 형태로 변환하기 위한 작업을 진행하고 있다. ANF는 Sparna사와 함께 EAD 및 EAC-CPF 형식의 기록 메타데이터를 RiC-O 기반으로 자동 변환하는 소프트웨어 'RiC-O Converter'를 개발하여 깃허브에 공개하기도 하였다. 또한 국제적 협동조합 SNAC(Social Networks and Archival Context Cooperative) 역시 소장 중인 전거레코드를 포함한 기록 메타데이터를 RiC-O의 일부 요소들을 활용하여 지식그래프로 구축할 예정이다(ICA EGAD, 2020b). 스위스의 시청각문화유

3) Wikidata. Retrieved November 10, 2020, <http://wikidata.org>

산 디지털 아카이브 Memobase 또한 RiC-O를 이용해 자관의 기록 메타데이터를 지식그래프로 구축하고 위키데이터, 지오네임스(GeoNames) 등 외부와 연계하여 데이터를 강화할 계획을 밝히고 있다(Memobase, 2020).

국내에서 RiC을 활용하여 진행 중인 프로젝트는 아직 찾아볼 수 없지만, 학술연구가 꾸준히 이루어지고 있다. 먼저, RiC의 개념적 소개와 분석을 진행한 연구를 살펴볼 수 있다. 박지영(2016; 2017a; 2017b)은 RiC의 개념적 특성 및 개발 방향 분석, ISAD(G)와의 비교, 기록공동체의 리뷰 검토 등을 수행한다. 현문수, 설문원(2018)은 국내의 공공 전자기록 조직모형의 재설계 방향을 도출하는 과정에서 RiC-CM의 특징을 분석하며 국제 동향을 검토한다. 박선희(2019)는 RiC의 특징과 한계점을 분석하고 국내 적용을 위한 시맨틱 기반 기록관 통합모델을 제안하고 있다.

한편, RiC-CM을 적용하여 기록 기술을 시도한 연구도 살펴볼 수 있다. 신미라, 김익한(2019)은 개념적 차원과 논리적 차원에서 RiC-CM에 대한 관계형 데이터 모델과 속성 그래프 기반 모델링을 제안한다. 김수현, 이성숙(2020)은 국가기록원 영구기록물 기술 방식의 한계를 복수출처 반영, 기록물 철-건 및 생산기관 간 연관 관계 표현 측면에서 분석하고, RiC-CM 기반의 해결 방안을 제시한다. 전예지, 이해원(2020)은 RiC-CM v0.2의 개정 내용을 분석하고 프로테제(Protege)를 이용해 온톨로지로 구현한 사례를 제시한다.

RiC은 물론 개발 중인 표준으로서 더욱 많은 연구가 필요하지만, 외부 데이터와의 폭넓은 연계 및 기록 맥락의 복잡성을 반영하는 다차원적 기술 측면에서 높은 잠재력을 지니고 있다. 그러나 국내에서 RiC-O 모델을 적용하여 지식그래프를 구축한 사례는 드물다. 본 연구는 외환위기 아카이브를 대상으로 RiC-O의 적용방안을 검토하고, 외부 데이터의 의미적 연계 방안을 제안한다.

3. 1997 외환위기 아카이브

3.1 개요

외환위기 아카이브는 1997년 발생한 외환위기에 대한 대중적 이해와 정보 공유를 위해 투명사회를 위한 정보공개센터를 중심으로 구축되었다. 외환위기 아카이브는 1997년 외환위기가 사회에 엄청난 영향을 주었지만, 국가적 위기를 전반적으로 이해할 수 있는 기록을 찾을 수 없다는 문제의식에서 출발했다. 투명사회를 위한 정보공개센터는 국제통화기금(IMF: International Monetary fund)에 외환위기와 관련된 기록의 비밀해제를 요청했고(2017년), 흩어져있던 기록을 정리하고 전산화하여 디지털 아카이브를 개발했다(김조은, 2019, p. 209). 외환위기 아카이브는 오픈소스 소프트웨어인 옴에카(Omeka)를 이용해 디지털 자원의 공유와 전시 등 다양한 온라인 서비스를 제공한다. 디지털 자원은 5가지 유형을 포함한다.

- 기록: 국제기구, 정부기관, 기업, 노동계, 연구기관, 언론사 등에서 수집한 다양한 유형의 정보자료
- 인물 및 조직: 외환위기와 관련된 국내외 주요 인물 및 기관 정보
- 연표: 외환위기 전후의 일자별 주요 사건 정보
- 용어: 외환위기를 이해하는데 필요한 경제 용어 정보
- 콘텐츠: 기록을 바탕으로 한 주제별 해설

기록은 PDF, XLS, JPG 등의 형식으로 원문이 공개되거나, 원문에 접근할 수 있는 외부 링크를 제공한다. 다만 저작권에 의해 원문 접근이 제한된 기록이 일부 존재한다. 용어, 인물 및 조직, 연표에 대한 정보는 한국은행

경제용어사전, 한국민족문화대백과사전, 위키피디아 등 외부 정보원이 제공하는 해설을 참조하고 있다. 외환위기 아카이브에서 기록, 인물 및 조직, 연표, 용어는 키워드 기반의 통합검색, 고급검색, 페이스(facet) 검색이 가능하며, 각각의 목록을 다운로드 받을 수도 있다.

3.2 한계

외환위기 아카이브는 여러 기관에 파편적으로 존재하는 기록을 통합하여 접근성을 높이고, 원문과 목록을 적극적으로 공유한다. 그러나, 외환위기 아카이브는 기능적 측면에서 다음과 같은 한계점을 갖고 있다.

첫째, 키워드 기반 검색은 이용자의 검색 의도를 정확하게 반영하지 못한다. 예를 들어, 1997년 경제부총리를 지낸 “임창열”이 생산한 기록은 성명을 키워드로 검색할 수 있다. 그러나 생산자명이 “부총리”로 입력된 기록인 “IMF 자금지원 협의 관련 부총리 발표문”은 검색되지 않는다. “부총리”와 “임창열”은 동일한 사람을 나타내지만, 기계는 그 의미와 관계를 정확하게 표현하지 않으면 이해할 수 없다.

둘째, 확장성 있는 표준을 사용한 메타데이터 기술이 필요하다. 외환위기 아카이브는 다수의 기관에서 수집한 기록을 바탕으로 구축되었고, 이를 더블린 코어(Dublin Core) 중심으로 기술한다. 그러나 단순성을 특징으로 하는 더블린 코어 기반 통합 방식은 메타데이터의 손실을 야기할 우려가 있다(임태원, 2014, pp. 41-42). 예를 들어 IMF 기록이 본래 계층적으로 기술되어 있을지라도, 더블린 코어 변환 과정에서 기술의 평면적 축소가 불가피하다. 이러한 문제를 해결하기 위해서는 의미 수준에서 상호운용성을 확보할 수 있는 데이터 모델로 확장이 필요하다. 나아가, 의미 수준에서의 상호운용성을 확보하면, 외환위기 관련 인물, 조직, 연표, 용어 정보를 외환위기 아카이브가 직접 구축하지 않더라도 외부 데이터와의 연계 및 데이터 강화를 통해 제공할 수 있게 된다.

마지막으로 외환위기에 대한 배경지식이 없으면 기록을 이해하기 어렵다. 예를 들어, “IMF 자금지원 협의 관련 부총리 발표문”은 부총리, 자금지원, 발표문에 대한 맥락정보가 있어야 이해할 수 있다. 현재 외환위기 아카이브는 주제별 해설을 제공하는 콘텐츠로 이를 보완하고 있으나, 콘텐츠가 모든 기록을 포괄하기는 현실적으로 어렵다. 따라서 기록에 대한 이해와 활용을 돕기 위해 기록과 기록을 둘러싼 개체의 관계 정보가 다차원적으로 제공될 필요가 있다.

4. 지식그래프 구축

4.1 데이터 수집

외환위기 아카이브는 디지털 자원의 데이터셋을 공유하고 있고, 누구나 자유롭게 데이터를 활용할 수 있다. 데이터셋은 소장기록, 인물·조직, 연표, 용어 총 세 종류로 구성되고, XLSX 파일로 제공된다. 데이터셋의 기술은 더블린 코어 메타데이터 요소를 중심으로 이루어져 있고, 옴에카 소프트웨어에서 제공하는 ‘Item Type Metadata’가 확장되어 있다.

본 연구는 외환위기 아카이브의 기록과 그 맥락을 지식그래프로 표현하기 위해 소장기록, 인물·조직, 연표를 구축 범위로 정의한다. 수집된 소장기록은 5,381건, 인물·조직은 각각 52건, 23건, 연표는 474건을 포함하고 있다.⁴⁾ 수집된 데이터셋은 메타데이터 항목을 분석하여 지식그래프 구축에 사용할 범위를 결정한다.

4) 외환위기 아카이브의 데이터셋은 지속적으로 업데이트되고 있으나, 본 연구는 2020년 8월 5일 홈페이지에 공개된 데이터셋을 수집, 활용하였다.

메타데이터 항목 중 “Item Type Metadata:자원유형”은 자원유형을 구분하기 위한 항목으로, “기록”, “인물”, “조직”, “연표” 중 하나를 값으로 지닌다. 이 항목은 인물·조직 데이터세트에서 해당 자원이 인물인지 조직인지 구분해줄 수 있지만, 소장기록 및 연표 데이터세트에서는 큰 의미를 지니지 않는다. 따라서 소장기록 및 연표 데이터세트는 “Item Type Metadata:자원유형” 항목을 구축 범위에서 제외한다. 또한, 메타데이터 항목은 존재하나 데이터 값이 입력되지 않은 항목도 구축 범위에서 제외한다. 소장기록 데이터세트는 7개 항목, 인물·조직 데이터 세트는 4개 항목, 연표 데이터세트는 1개 항목이 값이 존재하지 않아 제외된다.

따라서 소장기록 데이터세트는 전체 22개 중 14개 항목, 인물·조직 데이터세트는 전체 13개 중 9개 항목, 연표는 전체 12개 중 10개 항목을 사용한다. 다만 외환위기 진행 과정에서 개편된 정부기관의 계승 관계가 외환위기의 맥락을 이해하는 데 중요한 정보가 될 수 있다고 보고, 계승 관계 표현을 위한 항목을 인물·조직 데이터세트에 추가한다. 데이터 값은 국가기록원의 ‘기록물 생산기관 변천정보’⁵⁾를 참고하여, ‘재정경제원’, ‘재정경제부’, ‘금융감독원’ 3건에 한해 시험적으로 입력하여 사용한다.

4.2 RiC-O 어휘의 적용

외환위기 아카이브의 데이터세트는 적절한 온톨로지 어휘를 적용해 지식그래프로 변환된다. 온톨로지 어휘는 새롭게 정의하지 않고 표준 어휘를 재사용하여 지식 모델의 상호운용성을 확보한다. 이런 목적에서 기록학 관점에서 명확한 의미 표현은 RiC-O를 우선적으로 적용하고, BIBFRAME⁶⁾과 범용적 표현을 위한 Schema.org를 적용한다. 기존 메타데이터 항목과 온톨로지 어휘의 주요 매핑 현황은 <표 1>과 같다.

<표 1> 주요 어휘 매핑 현황

데이터세트	메타데이터 항목 ⁷⁾	속성 매핑	정의역(rdfs:domain)	공역(rdfs:range)
소장기록; 인물/조직; 연표	dc:identifier	rico:identifiedBy	rico:Record; rico:Instantiation; rico:Person; rico:CorporateBody; rico:Event	rico:Identifier
소장기록	dc:source	rico:hasProvenance	rico:Record; rico:RecordSet	rico:Agent
소장기록	dc:source	rico:includedIn	rico:Record; rico:RecordSet	rico:RecordSet
소장기록	dc:title	rico:title	rico:Record; rico:Instantiation	rdfs:Literal
소장기록	dc:creator	rico:createdBy	rico:Record	rico:Agent
소장기록	dc:date	rico:hasBeginningDate	rico:Record	rico:Date
소장기록	itm:기록유형	rico:hasContentOfTypes	rico:Record	rico:ContentType
소장기록	itm:기록형태	rico:hasDocumentaryFormType	rico:Record	rico:DocumentaryFormType
소장기록	dc:language	rico:hasLanguage	rico:Record	rico:Language
소장기록	itm:소주제	rico:hasMainSubject	rico:Record	rico:Thing
소장기록	itm:대주제	rico:hasSubject	rico:Record	rico:Thing
소장기록	dc:publisher	rico:publishedBy	rico:Record	rico:Agent
소장기록	dc:description	rico:scopeAndContent	rico:Record	rdfs:Literal
소장기록	dc:format	rico:instantiationExtent	rico:Instantiation	rdfs:Literal
소장기록	dc:relation	schema:url	rico:Instantiation	rdfs:Resource

5) 기록물 생산기관 변천정보. 검색일자: 2020. 11. 10.

<http://theme.archives.go.kr/next/organ/mainView.do>

6) BIBFRAME Ontology. Retrieved November 10, 2020,
from <https://id.loc.gov/ontologies/bibframe.html>

7) ‘dc’는 더블린 코어를, ‘itm’은 Item Type metadata를 의미한다.

데이터세트	메타데이터 항목 ⁷⁾	속성 매핑	정의역(rdfs:domain)	공역(rdfs:range)
인물/조직: 연표	dc:description: itm:참고자원: itm:참고자원URL	rico:descriptiveNote	rico:Person; rico:CorporateBody; rico:Event	rdfs:Literal
인물/조직	dc:title	rico:hasAgentName	rico:Person; rico:CorporateBody	rico:AgentName
인물/조직	itm:바이오그래피	rico:history	rico:Person; rico:CorporateBody	rdfs:Literal
인물/조직	File	schema:image	rico:Person; rico:CorporateBody	rdfs:Resource
인물/조직	itm:소속	rico:isMemberOf	rico:Person	rico:CorporateBody
인물/조직	itm:직급	rico:occupies	rico:Person	rico:Position
인물/조직	itm:홈페이지	schema:sameAs	rico:Person	rdfs:Resource
인물/조직	isSuccessorOf	rico:isSuccessorOf	rico:CorporateBody	rico:CorporateBody
연표	dc:title	rico:hasName	rico:Event	rico:Name
연표	itm:관련인물조직	rico:involves	rico:Event	rico:Agent
연표	dc:date: itm:연표구분	rico:isAssociatedWithDate	rico:Event	rico:SingleDate; rico:DateRange
연표	itm:관련기록: itm:관련기사	rico:resultsIn	rico:Event	rico:Record; rdfs:Resource

지식그래프의 주요 클래스와 속성은 다음과 같다. 첫째, ‘기록(rico:Record)’ 클래스는 외환위기의 상황을 이해할 수 있는 증거로서 외환위기 아카이브가 수집한 정보를 의미한다. 기록은 그것이 담고 있는 정보가 물리적 매체에 실체화된 ‘사본(rico:Instantiation)’을 하나 이상 지니며(rico:hasInstantiation), ‘기록집합(rico:RecordSet)’에 속할 수 있다(rico:includedIn). 예를 들어, “IMF-한국 합의서 원본” 기록은 PDF 형식의 사본을 지니고, 그 출처에 따라 “재정경제부 기록집합”에 속한다.

둘째, ‘행위자(rico:Agent)’ 클래스는 외환위기와 관련하여 업무나 활동을 수행한 ‘인물(rico:Person)’, ‘조직(rico:CorporateBody)’ 또는 ‘직위(rico:Position)’를 의미한다. 기록 및 기록집합은 행위자를 출처로 지니고(rico:hasProvenance), 행위자에 의해 생산된다(rico:createdBy).

셋째, ‘사건(rico:Event)’ 클래스는 외환위기를 배경으로 발생한 사건을 의미한다. 사건은 행위자와 연관되고(rico:involves), 기록이 생산되도록 이끈다(rico:resultsIn). 예를 들어, “[1997.11.21] 재정원, 국제통화기금(IMF) 유동성 조절자금 지원요청” 사건은 “IMF”, “재정경제원”과 같은 행위자와 연관되며, “한국의 대국제통화기금 유동성 조절자금 지원 요청 주요 내용”을 비롯한 일련의 기록을 발생시켰다.

외환위기 아카이브의 데이터세트에서 기록, 행위자, 사건은 파편적인 정보이지만, 지식 모델의 적용으로 각 개체는 관계 중심의 그래프로 표현된다.

4.3 데이터 강화(Data Enrichment)

지식그래프에서 데이터 강화는 외부 데이터에 존재하는 동일한 개체의 정보를 재사용하는 것을 말한다. 데이터 강화를 통해 원본 데이터세트에 존재하지 않는 정보를 새롭게 구축하지 않고 활용할 수 있다. 본 연구는 인물과 조직 데이터를 위키데이터로부터 강화한다.

작업에는 ‘오픈리파인(OpenRefine) 3.3’에 ‘RDF extension 1.2.0’을 설치하여 사용한다. 작업도구로 오픈리파인을 선택한 까닭은 무료 오픈소스 소프트웨어이고, 대규모 데이터의 정제에 용이하며, 확장기능을 설치하여 데이터 강화 및 RDF 변환 작업을 자동화된 방식으로 진행할 수 있기 때문이다.

데이터 강화는 세 단계 과정으로 진행된다. 첫째, 데이터 값에 포함된 오류나 공백을 정제한다. 예를 들어 원본 데이터세트에서 인물로 잘못 분류되어 있던 “AMF(아시아통화기금)”를 조직으로 수정한다. 데이터 정제는 소장기록 및 연표 데이터세트도 공통적으로 수행한다.

둘째, 데이터 조화(reconciliation) 기능으로 위키데이터에서 일치하는 개체를 탐색한다. 자동으로 탐색된 일치율은 인물과 조직에서 각각 78%, 43%이다. 그러나 여기에는 동명이인과 잘못 연결되거나, 위키데이터에 동일 개체가 존재함에도 자동으로 연결되지 않는 사례가 포함되어 연구자가 이차적으로 검증을 진행하였다. 인물은 부적합한 일치를 제거하고(4건), 복수로 제시된 후보 중 적합한 항목을 선택하여(4건) 최종 결과는 다시 78%로 조정되었다. 마찬가지로 조직을 검토한 결과 일치 항목이 추가되어(11건) 최종 결과는 91%로 조정되었다. 최종적으로도 일치되지 않는 항목은 위키데이터에 해당 개체가 존재하는 않는 경우에 해당한다.

셋째, 위키데이터와 일치된 개체의 데이터를 강화한다. 데이터 조화 기능으로 일치된 개체는 위키데이터에 정의된 속성정보를 원본 데이터에 추가할 수 있다. <그림 1>은 인물 데이터세트의 데이터 강화 작업을 보여준다. ①의 “Benjamin B. King”은 위키데이터에 개체가 존재하지 않아 연결되지 않은 상태이다. 반면, ②의 “김대중(Kim Dae-Jung)”은 데이터 조화 서비스에서 참조할 수 있는 개체를 찾아 연결하고 있다. 연결이 이루어지면 ③과 같이 해당 개체의 국적(country of citizenship), 소속정당(member of political party), 직업(occupation) 등 속성정보를 강화할 수 있다. 이러한 정보는 원본 데이터에는 포함되지 않았지만, 위키데이터 개체와의 의미적 연결을 통해 확장이 가능해진 것이다. 데이터 강화 결과, 인물과 조직 데이터세트에서 각각 6개와 7개의 속성이 새롭게 추가되었다.

id	DublinCore:Iden	DublinCore:Title	country of citize	country of citize	member of politi	member of politi	occupation	occupation 2	occupation 3
1.	KC-O-00001	Benjamin B. King Ben E. King (79) B. B. King (76) Create new Item							
2.	KC-O-00002	Kim Dae-jung Choose new match		Kim Dae-jung (Q45785) South Korean politician (1924-2009)			politician Choose new match	journalist Choose new match	
3.	KC-O-00003	Kim Young-saeng Choose new match					politician Choose new match		
4.	KC-O-00004	Chun Doo-hwan Choose new match	South Korea Choose new match		Democratic Justice Party Choose new match		politician Choose new match		
5.	KC-O-00005	Roh Tae-woo Choose new match	South Korea Choose new match		Democratic Justice Party Choose new match		politician Choose new match		
6.	KC-O-00006	Roh Moo-hyun Choose new match	South Korea Choose new match		Democratic Party Choose new match	Uri Party Choose new match	judge Choose new match	politician Choose new match	lawyer Choose new match
7.	KC-O-00007	Chae-ik Kim Choose new match					economist Choose new match		
8.	KC-O-00008	엄낙용 Choose new match							
9.	KC-O-00009	김인호 Choose new match							
10.	KC-O-00010	Yoon Jeung-hyun Choose new match	South Korea Choose new match		Bareun Party Choose new match		politician Choose new match		
11.	KC-O-00011	정덕구 Choose new match	South Korea Choose new match				politician Choose new match		

<그림 1> 위키데이터 연계를 통한 데이터 강화

강화된 메타데이터 항목은 <표 2>와 같이 온톨로지 어휘 매핑을 진행한다. 원본 데이터세트와 마찬가지로 RiC-O 어휘를 우선 적용하고, 더욱 명료한 의미를 표현하는 데 필요한 경우 다른 어휘를 예외적으로 사용한다. 예를 들어, ‘sex or gender’ 항목은 schema:gender 속성을 사용한다. ‘ISNI’ 항목은 rico:identifiedBy 속성을 사용하고, 공역값(range)은 식별자(rico:Identifier)와 ISNI (bf:Isni)로 확대한다.

<표 2> 강화된 데이터의 어휘 매핑 현황

유형	메타데이터 항목	속성 매핑	정의역(rdfs:domain)	공역(rdfs:range)
인물	member of political party	rico:isMemberOf	rico:Person = schema:Person	rico:CorporateBody = schema:Organization
인물	position held	rico:occupies	rico:Person = schema:Person	rico:Position
인물	place of birth	schema:birthPlace	rico:Person = schema:Person	rico:Place = schema:Place
인물	sex or gender	schema:gender	rico:Person = schema:Person	rico:DemographicGroup = schema:GenderType
인물	occupation	schema:hasOccupation	rico:Person = schema:Person	schema:Occupation
인물	Country of citizenship	schema:nationality	rico:Person = schema:Person	rico:Place = schema:Country
인물	position held	rico:existIn	rico:Position	rico:CorporateBody
조직	inception	rico:beginningDate	rico:CorporateBody = schema:Organization	rdfs:Literal
조직	headquarters location	rico:hasLocation	rico:CorporateBody = schema:Organization	rico:Place = schema:Place
조직	ISNI	rico:identifiedBy	rico:CorporateBody = schema:Organization	rico:Identifier = bf:Isni
조직	parent organization	rico:isHierarchicallyInferiorTo	rico:CorporateBody = schema:Organization	rico:CorporateBody
조직	logo image	schema:logo	rico:CorporateBody = schema:Organization	rdfs:Resource
조직	official website	schema:sameAs	rico:CorporateBody = schema:Organization	rdfs:Resource
조직	country	rico:containedBy	rico:Place = schema:Place	rico:Place = schema:Country
조직	headquarters location; country	rico:hasPlaceType	rico:Place = schema:Place; rico:Place = schema:Country	rico:PlaceType

4.4 지식그래프 변환

정제된 데이터세트는 지식 모델을 적용하고 그래프의 유효성과 데이터 값의 정확성을 점검한다. 모든 개체는 고유한 URI를 부여하고, 데이터세트는 RDF 형식의 그래프로 변환 및 추출하여 그래프 데이터베이스에 저장한다. 그래프 데이터베이스는 RDF 데이터를 위한 표준 질의언어 SPARQL 및 시각화 도구를 제공하는 Ontotext사의 ‘GraphDB SE’ 시험판을 사용한다.

구축된 외환위기 지식그래프의 규모는 <표 3>과 같다. 명시적 진술문(explicit statement)은 지식 모델을 적용해 직접 구축된 데이터이고, 추론된 진술문(inferred statement)은 그래프 데이터베이스에서 추론을 통해 찾아낸 데이터이다.

<표 3> 외환위기 지식그래프의 규모

구분	명시적 진술문	추론된 진술문
소장기록	118,826	126
인물·조직	1,757	116
연표	6,829	90
합계 (공통 진술문 포함)	127,412	332
합계 (공통 진술문 제외)	127,161	172

소장기록 데이터세트는 명시적 진술문 118,826개, 추론된 진술문 126개로 총 118,952개이다. 인물·조직 데이터세트는 명시적 진술문 1,757개, 추론된 진술문 116개로 총 1,873개이다. 연표 데이터세트는 명시적 진술문 6,829개, 추론된 진술문 90개로 총 6,919개이다. 합산 결과는 명시적 진술문 127,412개, 추론된 진술문 332개로 총 127,744개이다. 이 중 데이터세트에 공통으로 포함된 명시적 진술문 251개와 추론된 진술문 160개, 총 411개를 제외한 규모는 명시적 진술문 127,161개, 추론된 진술문 172개로 총 127,333개에 해당한다.

또한 <표 4>에서와 같이, 데이터 강화를 통해 인물 데이터세트에서 59개, 조직 데이터세트에서 56개의 개체가 위키데이터로부터 추가되었다. 확장된 개체의 총 규모는 중복되는 개체(3개)를 제외하고 112개에 해당한다.

<표 4> 위키데이터로부터 확장한 개체의 규모

클래스	규모	개체			합계 (A + B - C)
		인물 데이터세트 (A)	조직 데이터세트 (B)	중복 (C)	
rico:CorporateBody		7	5		12
rico:DemographicGroup = schema:GenderType		1	-	-	1
rico:Identifier =bf:Isni		-	16	-	16
rico:Place = schema:Country		4	4	2	6
rico:Place = schema:Place		17	10	1	26
rico:Position		16	-	-	16
schema:Occupation		14	-	-	14
rdfs:Resource		-	21	-	21
계(중복 제외)		59	56	3	112

5. 지식그래프의 활용

외환위기 지식그래프는 이용자의 의도에 맞는 의미적 탐색을 지원한다. 키워드 검색 기반의 기존 외환위기 아카이브는 상이한 문자열로 표현된 동일 개체, 혹은 동일한 문자열로 표현된 상이한 개체를 인식하는 데 한계를 지녔다. 예를 들어, 기계는 동일 인물일지라도 “임창열”과 “임 부총리”, “부총리 겸 재정경제원 장관 임창렬”, “임창렬”을 모두 별개의 개체로 처리하게 된다. 그러나 외환위기 지식그래프는 모든 개체를 URI로 식별함으로써 어떠한 문자열로 표현되든 “임창열”이라는 한 명의 인물을 인식할 수 있다. <그림 2>는 외환위기 지식그래프에서 SPARQL을 활용한 의미적 탐색의 예시를 보여준다. “임창열”의 URI를 사용하여 “임창열”이 생산한 기록(?기록)에 대한 정보를 질의한다. <그림 3>은 해당 질의의 결과를 보여준다. 생산자명이 “부총리”로 입력되어 기존에는 검색되지 않았던 “IMF 자금지원 협의 관련 부총리 발표문”이 검색결과에 포함된 것을 확인할 수 있다.

한편, 외환위기 지식그래프는 관련 기록, 인물, 조직, 사건을 관계 기반으로 재구성한다. 개체와 개체 사이의 관계를 네트워크 형태로 기술하여 외환위기 당시의 정황을 입체적으로 재구성할 수 있다. 이를 통해 기록의 의미와 맥락을 효과적으로 이해할 수 있다.

```

1 PREFIX rico: <https://www.ica.org/standards/RiC/ontology#>
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3 select ?식별자명 ?기록명 ?생산일자 ?소주제명 ?기록형태명 where {
4   ?기록 rico:createdBy <http://knowledge.cau.ac.kr/97imf/agent/9021d16d09ca7a43483cae8ac98a1be5> .
5   ?기록 rico:identifiedBy ?식별자 .
6   ?식별자 rdfs:label ?식별자명 .
7   ?기록 rdfs:label ?기록명 .
8   ?기록 rico:hasBeginningDate ?생산일 .
9   ?생산일 rdfs:label ?생산일자 .
10  ?기록 rico:hasMainSubject ?소주제 .
11  ?소주제 rdfs:label ?소주제명 .
12  ?기록 rico:hasDocumentaryFormType ?기록형태 .
13  ?기록형태 rdfs:label ?기록형태명 .
14  ?기록 rico:hasInstantiation ?사본 .
15 }

```

〈그림 2〉 외환위기 지식그래프 질의문 예시

	식별자명	기록명	생산일자	소주제명	기록형태명
1	"KC-R-00045"	"Summary of Side Letter ; 이면각서 원문"	"1997-12-03"	"1차협상"	"일반문서"
2	"KC-R-00061"	"IMF 자금지원 협의 관련 부총리 발표문"	"1997-12-03"	"1차협상"	"연설문"
3	"KC-R-00047"	"대통령, 국회의장, 후보자 각서"	"1997-12-03"	"1차협상"	"일반문서"
4	"KC-R-00058"	"임창열 부총리가 IMF 총재에게 보낸 서신 "	"1997-12-04"	"1차협상"	"일반문서"
5	"KC-R-00092"	"IMF지원과 경제정책방향"	"1997-12-17"	"1차협상"	"일반문서"
6	"KC-R-00102"	"IMF 자금지원 협의 관련 부총리 발표문"	"1997-12-03"	"1차협상"	"연설문"
7	"KC-R-02513"	"Deputy Prime Minister Special Announcement (97.12.24)"	"1997-12-24"	"재벌"	"연설문"
8	"KC-R-04018"	"IMF 합의의향서"	"1998-00-00"	"1차협상"	"기사"

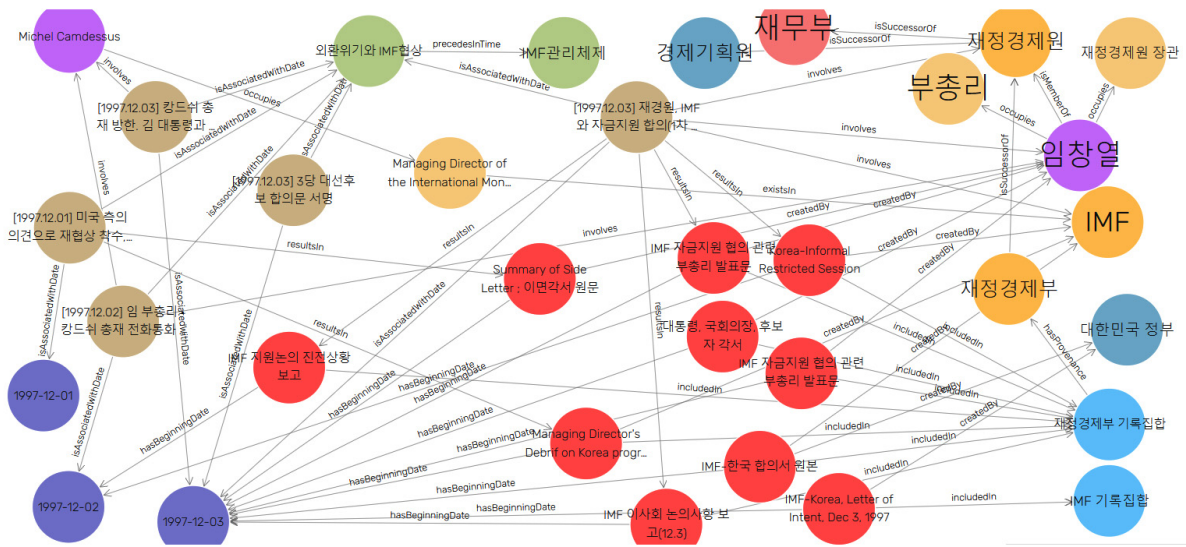
〈그림 3〉 외환위기 지식그래프 질의결과 예시

지식그래프의 복잡한 관계는 시각화 도구를 이용하여 더욱 효과적으로 파악할 수 있다. <그림 4>는 외환위기 지식그래프의 시각화 사례로, 1997년 12월 1일부터 한국 정부와 IMF의 ‘제1차 협상’이 타결된 3일 동안 발생한 주요 사건, 행위자, 그리고 기록 간의 관계를 보여준다. 12월 1일에는 “미국 측의 의견으로 재협상 착수, 이면합의 (side letter) 요구” 사건이 발생하였다. 이는 12월 3일에 “Summary of Side Letter; 이면각서 원문”과 같은 기록이 만들어진 배경이 된다(rico:resultsIn). 12월 2일 발생한 “임 부총리·강드쉬 총재 전화통화 합의 계속 난항 (이하 생략)” 사건은 “임창열”과 “Michel Camdessus”라는 인물과 연관된다(rico:involves). 12월 3일에 이루어진 “재경원, IMF와 자금지원 합의(1차 의향서 승인)” 사건은 “재정경제원”, “임창열”, “IMF”와 연관되며, IMF 이사회의 비공식 회의 관련 일련의 기록 및 “IMF 자금지원 협의 관련 부총리 발표문[5페이지]” 기록의 배경이 된다.

지식그래프에 표현된 사실 정보는 기록 사이의 관계를 탐색하는 데 다시 활용된다. 예를 들어 12월 3일 “임창열”이 생산한 기록 중 “재정경제부 기록집합”에 속하는(rico:includedIn) 관계를 통해 “IMF 자금지원 협의 관련 부총리 발표문[6페이지]” 기록, “대한민국 정부”가 생산한 기록 중 “재정경제부 기록집합”에 속하는 “IMF-한국 합의서 원본” 기록, “IMF 기록집합”에 속하는 “IMF-Korea, Letter of Intent, Dec 3” 기록을 탐색할 수 있다. 출처뿐만 아니라 유형, 주제, 언어 등을 기준으로 한 집합화와 탐색도 가능하다.

각각의 노드(node)를 확장하면 추가적인 정보를 탐색하고 다른 노드와 연결되는 공통점을 발견할 수도 있다. 즉, “임창열”은 “부총리” 및 “재정경제원 장관”, “Michel Camdessus”는 “Managing Director of the International Monetary Fund”라는 직위 정보를 파악할 수 있다. 마찬가지로 “재정경제부”는 “재정경제원”의 계승자(rico:isSuccessorOf)임을 확인할 수 있다.

Visual graph ⓘ



<그림 4> 외환위기 지식그래프의 시각화 사례 - '제1차 협상' 관련 기록, 행위자, 사건

6. 결론

디지털 아카이브는 물리적 공간 중심의 전통적 아카이브보다 이용자의 접근이 수월하고 활용성이 높다. 그러나 기록의 검색, 연계, 이해 측면에서 여전히 한계를 갖고 있다. 본 연구는 디지털 아카이브의 활용성 제고 방안으로 외환위기 아카이브를 지식그래프로 구축하고 실증하였다. 본 연구가 구축한 지식그래프는 기능적 측면에서 외환위기 아카이브를 개선하고 있다.

첫째, 의미 검색을 위한 데이터를 제공한다. 외환위기 아카이브에 포함된 기록, 인물, 조직, 사건 등 모든 개체는 지식 모델을 통해 의미 관계가 정의되어 이용자의 의도에 맞는 정확한 검색이 가능하다. 둘째, 확장성 있는 표준을 사용하여 데이터를 연계한다. 본 연구는 시맨틱 웹 표준 OWL, RDF, RDFS를 따르고 RiC-O를 중심으로 BIBFRAME, Schema.org와 같은 기존 온톨로지 어휘를 적용하였다. 이를 통해 기록을 풍부하고 명확하게 기술할 수 있으며, 서로 다른 메타데이터 스키마를 사용하는 기관의 데이터와의 통합이 이루어질 수 있도록 의미적 수준에서의 상호 운용성을 확보하고자 하였다. 데이터 측면에서 보면, 인물 및 조직 개체를 외부의 위키데이터와 연결하여 데이터를 강화함으로써 데이터의 중복 구축 없이 기록의 맥락에 대한 정보를 더욱 풍부하게 할 수 있다. 셋째, 외환위기 에 대한 전문적 배경지식을 갖추지 않았더라도 기록을 쉽게 이해할 수 있다. 개체 간 관계가 의미적 수준에서 정의되고 다차원적으로 기술되어 일반 이용자가 외환위기와 관련된 사건, 그에 관여한 행위자 및 그 행위자와 관련되는 또 다른 행위자 등 포괄적인 관계망 속에서 기록의 의미를 이해할 수 있다.

본 연구는 외환위기 아카이브를 지식그래프로 구축하기 위해 지식 모델을 정의하고, 주요 데이터셋을 그래프로 변환하고 있다. 그러나 외환위기 아카이브의 활용성을 더욱 극대화하기 위해서는, 소장기록의 실제 내용을 지식그래프로 구축하는 것이 필요하다. 현재 구축된 지식그래프는 메타데이터에 바탕을 두고 있고, 상세한 정보는 데이터타입 속성 '범위 및 내용(rico:scopeAndContent)'을 통해 문자열 형태로 제공된다. 개체와 개체 사이의 관계성을 상세화시키기 위해 텍스트에서 새로운 개체를 인식하고 추출하는 것이 필요하고, 국어사전 및 역사 정보를

포함하는 데이터와의 추가적인 연계가 필요하다. 향후 연구는 외환위기 아카이브의 사건과 개체를 구체화하기 위한 방안, 지식그래프로 구축한 데이터를 공동으로 활용할 수 있는 기술을 검토한다. 더불어, RiC과 Schema.org의 상호 연계를 위한 연구가 필요하다.

참고문헌

- 김수현, 이성숙 (2020). RiC-CM을 적용한 영구기록물 기술방안 연구. 한국기록관리학회지, 20(1), 115-137.
<https://doi.org/10.14404/JKSARM.2020.20.1.115>
- 김유승 (2010). 아카이브 2.0 구축을 위한 이론적 고찰. 한국기록관리학회지, 10(2), 31-52.
<https://doi.org/10.14404/JKSARM.2010.10.2.031>
- 김조은 (2020). 시민사회단체에서 아카이브를 만든다는 것: '1997 외환위기 아카이브' 구축 사례. 한국기록관리학회지, 20(1), 207-212. <https://doi.org/10.14404/JKSARM.2020.20.1.207>
- 김학래 (2017). 지식그래프. 서울: 커뮤니케이션북스. <https://doi.org/10.979.11288/05141>
- 박선희 (2019). 기록물 맥락정보 향상 및 통합시스템 개발에 관한 연구 -RiC-CM 및 RiC-O를 중심으로. 기록과 정보·문화연구, 9, 55-96.
- 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>
- 박지영 (2016). 차세대 기록물 기술표준에 관한 연구. 한국기록관리학회지, 16(1), 223-245.
<https://doi.org/10.14404/JKSARM.2016.16.1.223>
- 박지영 (2017a). ISAD(G)에서 RiC-CM으로의 전환에 관한 연구. 한국기록관리학회지, 17(1), 93-115.
<https://doi.org/10.14404/JKSARM.2017.17.1.093>
- 박지영 (2017b). RiC에 대한 기록공동체의 리뷰를 통해 본 기록물 기술표준 개선을 위한 제안. 기록학연구, 54, 81-109.
<https://doi.org/10.20923/KJAS.2017.54.081>
- 박진경, 김유승 (2017). 참여형 디지털 아카이브 활성화 방안 연구. 한국비블리아학회지, 28(2), 219-243.
<https://doi.org/10.14699/KBIBLIA.2017.28.2.219>
- 서해숙 (2013). 진도아리랑 디지털 아카이브 구축의 필요성과 설계. 인문콘텐츠, 29, 219-240.
- 신미라, 김익한 (2019). RiC을 적용한 아카이브 시스템 데이터 모델링 연구. 한국기록관리학회지, 19(1), 23-67.
<https://doi.org/10.14404/JKSARM.2019.19.1.023>
- 안주현 (2012). 필사본 음식 조리서를 활용한 디지털 아카이브 구축. 嶺南學, 21, 395-430.
<https://doi.org/10.36034/YNCDOI.2012..21.395>
- 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>
- 이은옥, 박희진 (2018). 이산가족 찾기 기록 패킷 기반 온톨로지 모델 설계에 관한 연구. 한국기록관리학회지, 18(4), 231-257.
<https://doi.org/10.14404/JKSARM.2018.18.4.231>
- 이정현, 이윤용, 방기영, 김용 (2015). Linked Open Data 기반 대통령기록관 기록정보 서비스 확장에 관한 연구. 한국기록관리학회지, 15(2), 55-82. <https://doi.org/10.14404/JKSARM.2015.15.2.055>
- 이정희, 김희섭 (2007). 대학 전자기록물을 위한 온톨로지 기반 검색시스템 설계 및 구현. 정보관리학회지, 24(3), 343-362.
<https://doi.org/10.3743/KOSIM.2007.24.3.343>
- 임태원 (2014). 디지털 문화유산자원 통합서비스를 위한 데이터 모델 연구. 석사학위논문. 명지대학교 기록정보과학전문대학원.
- 전예지, 이해원 (2020). RiC-CM v0.2 분석을 통한 온톨로지 모델링에 관한 연구. 한국기록관리학회지, 20(1), 139-158.
<https://doi.org/10.14404/JKSARM.2020.20.1.139>

- 지주형 (2011). 한국 신자유주의의 기원과 형성. 서울: 책세상. <https://doi.org/10.978.897013/8039>
- 최진경 (2020). 국내 과학기술성과 아카이브의 관리와 활용을 위한 시맨틱 기술 연구. 석사학위논문. 서울대학교 대학원, 협동과정 기록학전공.
- 하승록, 안대진, 임진희 (2017). 기록정보 LOD 구축을 위한 의미 상호연결 자동화 실험 연구. 한국기록관리학회지, 17(4), 177-200. <https://doi.org/10.14404/JKSARM.2017.17.4.177>
- 하승록, 임진희, 이해영 (2017). 오픈소스 도구를 이용한 기록정보 링크드 오픈 데이터 구축 절차 연구. 정보관리학회지, 34(1), 341-371. <https://doi.org/10.3743/KOSIM.2017.34.1.341>
- 현문수 (2013). 민간·공동체 아카이브 네트워크 구축을 위한 기술정보 공유 방향 연구. 한국기록관리학회지, 13(3), 127-150. <https://doi.org/10.14404/JKSARM.2013.13.3.127>
- 현문수, 설문원 (2018). 차세대 공공 전자기록의 조직 모형 개발을 위한 방향 탐구. 기록학연구, 56, 183-212. <https://doi.org/10.20923/KJAS.2018.56.183>
- Clavaud, F. (2020. 1. 28). ICA-Records in Contexts Conceptual Model and Ontology. Presented at the Conference That Took Place at the ANF, France. from https://f.hypotheses.org/wp-content/blogs.dir/2167/files/2020/02/20200128_2_RecordsInContexts_englishVersionAdded1003.pdf
- Gracy, K.F. (2015). Archival Description and Linked Data: A Preliminary Study of Opportunities and Implementation Challenges. Archival Science, 15(3), 239-294.
- ICA EGAD (2016). Records in Contexts: Conceptual Model For Archival Description - Consultation Draft v0.1. from <https://www.ica.org/sites/default/files/RiC-CM-0.1.pdf>
- ICA EGAD (2019b). Records in Contexts: Conceptual Model For Archival Description - Consultation Draft v0.2 (preview). from https://www.ica.org/sites/default/files/ric-cm-0.2_preview.pdf
- Matienzo, M. A., Roke, E. R., & Carlson, S. (2017). Creating a Linked Data-friendly Metadata Application Profile for Archival Description. In Proceedings of the 2017 International Conference on Dublin Core and Metadata Applications (DCMI'17). Dublin Core Metadata Initiative, 112-116.
- Mazzini, S. & Ricci, F. (2011). EAC-CPF Ontology and Linked Archival Data. In SDA, 72-81.
- Pearce-Moses, R. (2005). A Glossary of Archival and Records Terminology. Chicago, IL: Society of American Archivists.
- Ziku, M. (2020). Digital Cultural Heritage and Linked Data: Semantically-informed Conceptualisations and Practices with a Focus on Intangible Cultural Heritage. Liber Quarterly, 30(1), 1-16. <http://doi.org/10.18352/lq.10315>

[웹사이트]

- ICA EGAD (2019a). International Council on Archives Records in Contexts Ontology (ICA RiC-O) version 0.1. Retrieved November 10, 2020, from https://www.ica.org/standards/RiC/RiC-O_v0-1.html
- ICA EGAD (2020a). Next Steps. Retrieved November 10, 2020, from <https://ica-egad.github.io/RiC-O/next-steps.html>
- ICA EGAD (2020b). RiC-O Projects and Tools. Retrieved November 10, 2020, from <https://ica-egad.github.io/RiC-O/projects-and-tools.html>
- Memobase (2020). AUCH BEIM DATENMODELL SETZEN WIR AUF INNOVATION. Retrieved November 10, 2020, from <http://memobase.ch/de/memobase2020>

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- Ahn, Ju Hyeon (2012). Digital Archive using Korean Culinary Manuscripts. Youngnamhak, 21, 395-430. <http://dx.doi.org/10.36034/yncoi.2012..21.395>

- Choi, Jin gyeong (2020). A Study on Semantic Technology for Management and Utilization of Science and Technology Achievements Archives in Korea. Unpublished master's thesis. Seoul National University, Graduate School. Interdisciplinary Program in Archival Studies, Seoul.
- Ha, Seung-rok, An, Dae-Jin, & Yim, Jin-hee (2017). An Experimental Study on the Automatic Interlinking of Meaning for the LOD Construction of Record Information. *Journal of Korean Society of Archives and Records Management*, 17(4), 177-200. <https://doi.org/10.14404/JKSARM.2017.17.4.177>
- Ha, Seung-rok, Yim, Jin-hee, & Rieh, Hae-young (2017). A Study on the Procedure for Constructing Linked Open Data of Records Information by Using Open Source Tool. *Journal of the Korean Society for Information Management*, 34(1), 341-371. <https://doi.org/10.3743/KOSIM.2017.34.1.341>
- Hyun, Moonsoo & Seol, Moon-won (2018). An Exploration of the Direction of Development of the Next Generation Conceptual Model for Organizing Public Digital Records. *The Korean Journal of Archival Studies*, 56, 183-212. <https://doi.org/10.20923/KJAS.2018.56.183>
- Hyun, Moonsoo (2013). A Study on Sharing Descriptive Information for Establishing Private and Community Archive Network. *Journal of Korean Society of Archives and Records Management*, 13(3), 127-150. <https://doi.org/10.14404/JKSARM.2013.13.3.127>
- Jeon, Ye Ji & Lee, Hyewon (2020). A Study on the Ontology Modeling by Analyzing RiC-CM v0.2. *Journal of Korean Society of Archives and Records Management*, 20(1), 139-158. <https://doi.org/10.14404/JKSARM.2020.20.1.139>
- Ji, Joo Hyung (2011). Origin and Formation of Korean Neoliberalism. Ch'aek Sesang. <https://doi.org/10.978.897013/8039>
- Kim, Haklae (2017). Knowledge Graph. Seoul: CommunicationBooks, Inc. <https://doi.org/10.979.11288/05141>
- Kim, Joeun (2020). Building the Archives in a Civil Society: 'The Archives of 1997 Korean Financial Crisis.' *Journal of Korean Society of Archives and Records Management*, 20(1), 207-212. <https://doi.org/10.14404/JKSARM.2020.20.1.207>
- Kim, Soohyun & Lee, Sungsook (2020). A Study on Archive Description Using RiC-CM. *Journal of Korean Society of Archives and Records Management*, 20(1), 115-137. <https://doi.org/10.14404/JKSARM.2020.20.1.115>
- Kim, You-Seung (2010). A Theoretical Study on Establishing Archive 2.0. *Journal of Korean Society of Archives and Records Management*, 10(2), 31-52. <https://doi.org/10.14404/JKSARM.2010.10.2.031>
- 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, Eun Uk & Park, Heejin (2018). A Study on a Facet-Based Ontology Design for Archival Records of Finding Dispersed Families. *Journal of Korean Society of Archives and Records Management*, 18(4), 231-257. <https://doi.org/10.14404/JKSARM.2018.18.4.231>
- 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, Junghee & Kim, Heesop (2007). A Design and Implementation of Ontology-based Retrieval System for the Electronic Records of Universities. *Journal of the Korean Society for Information Management*, 24(3), 343-362. <https://doi.org/10.3743/KOSIM.2007.24.3.343>
- Park, Jinkyung & Kim, You-seung (2017). A Study on Participatory Digital Archives. *Journal of the Korean Biblia Society for Library and Information Science*, 28(2), 219-243. <https://doi.org/10.14699/KBIBLIA.2017.28.2.219>
- 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>
- Park, Sun-hee (2019). A Study on Improving Record Contextual Information and Developing Integrated System -Focusing on RiC-CM and RiC-O. *The Korean Journal of Archival, Information and Cultural Studies*, 9, 55-96.
- Park, Zi-young (2016). Analyzing the Next-generation Archival Description Standard: “Record in Context” of ICA EGAD. *Journal of Korean Society of Archives and Records Management*, 16(1), 223-245.
<https://doi.org/10.14404/JKSARM.2016.16.1.223>
- Park, Zi-young (2017a). Transition of Archival Description from ISAD(G) to Record in Context Conceptual Model. *Journal of Korean Society of Archives and Records Management*, 17(1), 93-115.
<https://doi.org/10.14404/JKSARM.2017.17.1.093>
- Park, Zi-young (2017b). Improving Archival Descriptive Standard Based on the Analysis of the Reviews by Archival Communities on Ric-CM Draft. *The Korean Journal of Archival Studies*, 54, 81-109.
<https://doi.org/10.20923/KJAS.2017.54.081>
- Seo, Hae-Sug (2013). Necessity and Design of Construction of Digital Archive of Jindo-Arirang. *Humanities Contents*, 29, 219-240.
- Shin, Mira & Kim, Ikhan (2019). A Study in the Data Modeling for Archive System Applying RiC. *Journal of Korean Society of Archives and Records Management*, 19(1), 23-67. <https://doi.org/10.14404/JKSARM.2019.19.1.023>
- Yim, Tae Won (2014). A Study on Data Modeling for Integrated Services of Digital Cultural Heritage Resources: Focusing on the Europeana Data Model(EDM). Unpublished master's thesis, Myongji University, Graduate School of Records, Archives & Information Science. Seoul.