

## 북한 인문학 시맨틱 데이터 아카이브 구축과 활용\*

홍종욱\*\* · 김도민\*\*\* ·

강수연\*\*\*\* · 홍수현\*\*\*\*

### [초 록]

북한 인문학 시맨틱 데이터 아카이브는 양적 접근과 질적 접근, 그리고 정확성과 확장성의 조화를 원칙으로 삼아 설계되었다. 첫째, 인물 - 문헌 - 기관을 세 축으로 삼아 관계형 데이터베이스를 구축했다. 둘째, 각 '항목' 사이의 의미적(semantic) '관계'를 표현할 수 있는 온톨로지를 설계했다. 셋째, 축적된 정보를 시각화하는 방식은 관계망 그래프를 채택했다. 넷째, 하이퍼링크(hyperlink)를 통해 서로 연결되는 '북한 인문학 위키' 페이지를 설정했다.

\* 이 논문은 2020년도 서울대학교 통일평화연구원의 재원으로 통일기반구축사업의 지원을 받아 수행된 결과물임.

\*\* 서울대학교 인문학연구원 부교수(제1저자)

\*\*\* 서울대학교 인문학연구원 선임연구원(교신저자)

\*\*\*\* 서울대학교 국사학과 석사과정

주제어: 디지털 인문학, 네트워크 그래프, 위키 문서, 북한 역사학, 비선형적 글쓰기  
Digital Humanities, Network Graph, Wiki Article, North Korean Historiography,  
Nonlinear Writing

북한 인문학 시맨틱 데이터 아카이브 구축과 활용은 ‘공유’와 ‘표현’이라는 디지털 인문학의 특징을 체감하는 과정이었다. 첫째, 디지털·웹 환경이 보장하는 소통과 협업 가능성을 충분히 활용함으로써 질 높은 공동연구를 수행할 수 있었다. 개인연구의 집합으로서의 공동연구를 넘어 데이터 수집에서 최종적인 시각화까지 생각의 과정을 함께 하는 공동연구를 지향했다. 디지털·웹 환경이 일련의 과정을 시각적으로 공유하면서 작업을 진행할 수 있도록 도왔다.

둘째, 선형적 글쓰기를 넘어서 인문학의 새로운 표현 방식을 시도할 수 있었다. 북한 인문학 아카이브는 연구자가 놓치기 쉬운 의미관계까지 포착하여 네트워크 그래프로 드러냄으로써 인간의 직관을 넘어서는 이미지를 표현할 수 있었다. 위키 문서 역시 하이퍼링크, 타임라인, 전자 지도 등 디지털·웹 환경 고유의 생산력을 충분히 활용함으로써 선형적 글쓰기를 넘어 새로운 인문학적 표현을 시도할 수 있었다.

## 1. 북한 인문학 시맨틱 데이터 아카이브 구상

이 글은 ‘북한 인문학 시맨틱 데이터 아카이브’ 설계와 구축 경험을 공유하고 활용 사례를 소개한 뒤 과제와 전망을 정리하는 것을 목표로 한다. 서울대 인문학연구원은 서울대 통일평화연구원이 주관하는 통일기반구축사업의 일환으로 2020년도부터 3개년 계획으로 북한 인문학 시맨틱 데이터 아카이브(이하 북한 인문학 아카이브) 구축 사업을 진행하고 있다.

북한 인문학 아카이브는 ‘역사학’, ‘고고학’, ‘한국문학’, ‘외국문학’, ‘서양고전’, ‘동양고전’ 등 6개 분과학문을 대상으로 삼아 북한 인문학에 관한 최초의 학제적 연구를 시도한다. 북한 인문학 아카이브는 문헌 인문학 연구자와 디지털 인문학 연구자의 협업을 꾀한다. 문헌 인

문학자들이 중심이 되어 북한 인문학 정보를 종합적으로 수집한 뒤, 디지털 인문학 연구자와 함께 연구 및 교육에 본격적으로 활용할 수 있는 데이터베이스를 설계하고 구축한다.

여기서 기존의 북한 인문학 관련 디지털 아카이브 현황을 살펴보자. 첫째, 북한 자료만을 제공하는 방식으로는 통일부에서 구축한 디지털 아카이브인 통일부 북한자료센터와 통일부 북한정보포털이 있다.<sup>1)</sup> 통일부 북한자료센터는 『로동신문』, 『근로자』 등 북한 주요 문헌의 목차에 대한 문자열 방식의 검색을 제공한다. 북한의 주요 인문학자의 직위나 그들이 작성한 글을 확인할 수 있는 유용성이 있다. 다만 북한자료센터는 사전 허가를 받은 방문객에 한하여 목차 검색 및 원문 복사를 허용하는 폐가제 형태를 취하고 있다. 다음으로 통일부 북한정보포털은 북한의 군사, 경제, 사회, 교육문화, 용어, 법령정보, 인물, 남북한 비교통계 등을 제공한다. 북한과 관련한 다양한 정보를 웹상에서 제공하지만, 인문학과 관련된 자료는 거의 포함되어 있지 않다.

둘째, 한국 근현대사 자료 중 일부로서 북한 자료를 포함하여 웹상에서 제공하는 기관으로는 국립중앙도서관, 국사편찬위원회, 미국의 우드로윌슨센터 등이 있다. 국립중앙도서관은 해외 한국 관련 기록물 중 미국이 북한에서 노획한 RG 242 (노획 북한 문서, National Archives Collection of Foreign Records Seized, 1675~1983)<sup>2)</sup>를 제공한다. 북한 인문학과 관련해서는 해방 이후 6·25 전쟁 이전에 생산된 『로동신문』 등 90여 종의 신문과 『조선여성』 등 320여 종의 잡지가 포함되어 있다. 그런데 이 사이트는 잡지 권호별 목차까지 제공함에도 불구하고, 정작

1) 통일부 북한자료센터 웹사이트(<https://unibook.unikorea.go.kr>), 통일부 북한정보포털 웹사이트(<https://nkinfo.unikorea.go.kr>) 참조(이상 검색일 2021.2.24).

2) 이흥환(2020), 「문서군 가이드: RG 242 (노획 북한 문서, National Archives Collection of Foreign Records Seized, 1675-1983)」, 국립중앙도서관 웹사이트 (<https://www.nl.go.kr/NL/contents/N20403010000.do>, 검색일 2021.2.24).

검색에서는 목차를 포함하지 않아 기초적인 활용에 어려움이 있다.<sup>3)</sup> 다음으로 국사편찬위원회에서 구축한 자료 가운데 일부 북한 인문학자가 식민지기 및 해방 직후에 남긴 글을 확인할 수 있다.<sup>4)</sup>

해외 기관 중 미국의 우드로윌슨센터(Woodrow Wilson Center)가 북한 관련 자료를 수집·제공한다.<sup>5)</sup> 우드로윌슨센터는 1991년 이후 국제 냉전사 프로젝트(The Cold War International History Project, CWIHP)를 센터 내에 설치하고 여러 정부가 생산한 냉전 관련 역사자료를 수집해왔다. 국제 냉전사 프로젝트는 정치적·언어적 한계 때문에 접근하기 힘들었던 과거 공산권 문서를 영어로 번역하여 원문과 함께 제공한다. 또한, 2006년부터 우드로윌슨센터는 한국 북한대학원대학교와 공동으로 북한 국제 문서 프로젝트(North Korea International Documentation Project, NKIDP)를 수행하면서, 북한 주재 동유럽, 중국, 소련 대사관 그리고 공산권 주재 북한 대사관 문서를 제공하고 있다. 우드로윌슨센터는 다양한 국가에서 생산한 다양한 언어의 자료를 영문으로 바꿔 키워드 검색까지 가능한 디지털 아카이브를 구축해놓았다.<sup>6)</sup>

이상 국내외 북한 인문학 관련 디지털 아카이브를 살펴봤다. 통일부 북한자료센터는 기초 자료라고 할 수 있는 신문 및 잡지를 다수 제공하지만, 디지털 원문은 제공하지 않고 검색 기능도 불완전하다. 더욱이 직접 방문자만 이용할 수 있는 탓에, 코로나바이러스감염증-19 상

---

3) 구체적인 사례를 제시하면 다음과 같다. SA 2005 『조선여성』 창간호(1946)에 박정애가 쓴 글 「조선민주건설을 위한 여성의 횃불이 되라」가 19쪽에 실려 있으며 이를 목차로 제공하지만, 검색 창에 ‘박정애’를 입력하면 이 잡지가 검색되지 않는다.

4) 국사편찬위원회 한국사 데이터베이스 웹사이트(<http://db.history.go.kr>, 검색일 2021.2.24) 참조.

5) 미국 우드로윌슨센터 웹사이트([www.wilsoncenter.org](http://www.wilsoncenter.org), 검색일 2021.2.24) 참조.

6) 미국 우드로윌슨센터 디지털 아카이브(<https://digitalarchive.wilsoncenter.org>, 검색일 2021.2.24) 참조.

황에서 이용에 심각한 제한이 따르고 있다. 그밖에 북한 관련 정보를 제공하는 국내 웹 사이트와 미국의 아카이브 등은 다양한 자료를 이용할 수 있지만, 단순 문자열(n-gram) 검색을 통해 추출된 문헌 자료에 대해 디지털 화상 파일을 제공하는 데 그쳐 디지털 아카이브 나름의 장점을 충분히 살리고 있지 못하다.

북한 인문학 아카이브는 기존 디지털 아카이브의 성과를 충분히 활용한 위에 새롭게 시맨틱 데이터 아카이브를 지향한다.<sup>7)</sup> 다량의 복잡한 자료를 RDF (Resource Description Framework) 형식에 따라 데이터 사이의 의미적(semantic) 관계에 유의하여 정리한다. 정리된 항목과 관계를 네트워크 그래프로 시각화한다. 아울러 공유와 협업의 플랫폼으로 주목받는 위키를 도입하여 논문과 같은 선형적 글쓰기와 다른 새로운 글쓰기 방식을 시도한다. 이제까지 대부분의 온라인 아카이브가 오프라인 아카이브를 보조하는 역할에 그친 데 반해, 비대면 연구 및 교육 환경에 걸맞은 북한 인문학 플랫폼이 구축될 것으로 기대한다.

## 2. 북한 인문학 시맨틱 데이터 아카이브 설계와 구축

### 2.1. 아카이브 설계와 쟁점

북한 인문학 아카이브를 설계하는 단계에서 선택과 고려가 필요한 쟁점이 있었다.

먼저 양적 접근과 질적 접근을 어떻게 조화시킬 것인가라는 문제다. 디지털 인문학이라고 하면 많은 경우 빅 데이터를 수집하고 분석함으로써 아날로그 환경에서 포착하기 어려웠던 논점을 발견하는 방법론

7) 시맨틱 데이터에 대해서는 김현 · 임영상 · 김바로(2016), 『디지털 인문학 입문』 (한국외국어대학교 지식출판원, pp. 147-162) 참조.

을 떠올린다. 한편 종이에 글을 쓰듯 북한 인문학에 관한 지식과 정보를 디지털 환경에 축적하여 공유하는 방법론을 생각해 볼 수 있다. 전자가 양적 접근이라면 후자는 질적 접근이다. 양적 접근이 예컨대 구글을 검색하여 데이터를 축적한다면, 질적 접근은 구글에 검색될 인문학 데이터의 생산을 지향한다.

북한 인문학 아카이브는 데이터의 의미적(semantic) 관계 구축을 중시한다는 점에서 질적 접근에 친화적이다. 다만 이미 구축된 다양한 디지털 아카이브의 자료를 충분히 활용하기 위해서라도 북한 인문학 아카이브는 양적 접근과 질적 접근의 조화를 꾀했다. 관계형 데이터베이스(RDB)를 통해 데이터를 축적하는 동시에 ‘북한 인문학 위키’ 페이지도 작성했다. 또한, 최종적인 시각화에서도 문자열 검색 결과를 표시한 뒤 항목 사이의 의미적 관계망을 그래프로 표시했다.

다음으로 정확성과 확장성을 어떻게 조화시킬 것인가라는 문제다. 북한 인문학 아카이브는 순도 높은 자료의 축적을 기본으로 하지만, 궁극적으로는 대중 참여형 플랫폼을 지향한다. 여기서 확장성의 모순이 발생한다. 전문 연구자가 아닌 이가 아카이브 구축에 참여하면 정보의 순도를 보장하기 힘들다. 그렇다고 질 높은 자료의 축적만 고집하면 대중과의 쌍방향 소통은 어렵다.

정확성과 확장성의 조화를 위해 온톨로지에 기반한 관계형 데이터베이스(RDB)와 개방형 위키 페이지를 동시에 구축하는 방식을 취했다. 사업에 참여한 전문 연구자가 구축한 데이터 아카이브를 참고하여 외부 참여자가 북한 인문학 위키를 편찬하고, 위키에 쌓인 정보를 관리자가 검증하여 다시금 데이터베이스에 반영하는 형식이다. 북한 인문학 위키를 작성하기 위해서는 접속 아이디가 필요한데, 아이디 부여 범위는 점차 넓혀갈 계획이다.

양적 접근과 질적 접근, 그리고 정확성과 확장성의 조화를 원칙으로 삼아 북한 인문학 아카이브를 설계했다. 첫째, 인물 - 문헌 - 기관을

세 축으로 삼아 관계형 데이터베이스를 구축했다. 기존의 문헌 연구와 디지털 아카이브에서 확인할 수 있는 다양한 인물 정보, 그리고 학술지 기사목록 등 문헌 정보를 충분히 활용했다. 인물과 문헌 정보를 중심으로 데이터베이스를 구축하면서 또 하나의 축인 기관과의 연관성을 놓치지 않고자 했다. 둘째, 관계형 데이터베이스는 온톨로지에 입각해 구축했다. 평면적인 표에 데이터를 입력하는 것이 아니라, 인물, 문헌, 기관 등 각 ‘항목’ 사이의 의미적(semantic) ‘관계’를 표현할 수 있는 온톨로지를 설계했다.

셋째, 축적된 정보를 시각화하는 방식은 관계망 그래프를 채택했다. 전문 연구자가 아니더라도 쉽게 접근할 수 있도록 검색 창에서는 단순 문자열(n-gram) 검색을 제공한다. 여기에서 추출된 항목을 클릭하면 관련 항목 사이의 의미적 관계가 네트워크 그래프로 출력되도록 설계했다. 넷째, 북한 인문학 위키 페이지를 설정했다. 서술형 위키 문서 작성에는 관계형 데이터베이스에 축적된 정보를 활용할 수 있다. 위키 문서는 하이퍼링크(hyperlink)를 통해 서로 연결된다. 위키 소프트웨어에는 작성 이력이 자동으로 기록되므로 진척 상황을 쉽게 확인할 수 있다.

## 2.2. 온톨로지 확립

북한 인문학 아카이브 온톨로지는 인물과 문헌 등의 메타 데이터를 기본적으로 주어, 동사, 목적어 형식으로 의미적 관계로 표현한다. 북한 인문학의 특징을 잘 보여주는 ‘항목’(class)과 ‘관계’(relation)가 무엇인지에 대한 연구팀의 토론 및 회의에 근거하여 데이터 모델을 설계했다.

‘항목’과 ‘관계’는 자명해 보였지만, 실제 데이터 구축 과정에서는 무엇을 항목으로 세우고 그것 사이의 관계는 어떻게 설정할 것인가를

거듭 고민해야만 했다. 예컨대 학교, 직장 등의 ‘기관’(institution)과 학회 등의 ‘단체’(group)를 별도 항목으로 설정했다. 당초 민간 영역이 거의 존재하지 않는 북한 사회 특성상 ‘단체’ 항목을 설정하지 않았으나, 많은 학자의 식민지기 경력에 나오는 각종 학우회나 진단학회 등의 학회는 역시 기관과 구별할 필요가 있다고 판단해 따로 항목을 두었다.

또한, ‘논쟁’(controversy)과 ‘학설’(argument) 항목을 설정했다. 논쟁은 어느 지역 어느 시대에도 존재한다. 다만 많은 경우 논쟁의 시작과 끝이 불분명하고 쟁점과 구도가 명료하지 않아 분절하여 파악하기가 어렵다. 그러나 북한 학계의 경우 특히 건국 초기에서 1960년대까지는 과학원 등이 주도하여 체계적으로 논쟁을 진행하고 ‘총화’를 통해 결론을 짓는 관행이 충실하게 지켜졌다. 북한 인문학 아카이브에서는 사회주의 학술의 특징을 고려하여 ‘논쟁’과 ‘학설’ 항목을 세움으로써 북한 인문학 고유의 양상을 포착하고자 했다.

이러한 과정을 거쳐 항목과 그 속성(표 1), 관계와 그 속성(표 2)을 잠정적으로 확립했다.

〈표 1〉 항목과 속성(2020년 11월 기준)

항목(Class)	속성(Attribute)		
인물 (Person)	식별자(ID)	대표명(Name)	한글명(Korname)
	한자명(Chiname)	가나명(Japname)	영문명(Engname)
	이칭(Alias)	성별(Sex)	생년(BirthYear)
	몰년(DeathYear)	연구영역(Field)	
단체 (Group)	식별자(ID)	대표명(Name)	한글명(Korname)
	한자명(Chiname)	가나명(Japname)	영문명(Engname)
	이칭(Alias)		
기관 (Institution)	식별자(ID)	대표명(Name)	한글명(Korname)
	한자명(Chiname)	가나명(Japname)	영문명(Engname)
	이칭(Alias)	유형(Type)	관련국(Country)
	설립년(EstablishedYear)	폐지년(AbolishedYear)	

항목(Class)	속성(Attribute)		
공간 (Location)	식별자(ID)	대표명(Name)	한글명(Korname)
	한자명(Chiname)	가나명(Japname)	영문명(Engname)
	이칭(Alias)	유형(Type)	위도(Latitude)
	경도(Longitude)	형태(Form)	
사건 (Event)	식별자(ID)	대표명(Name)	한글명(Korname)
	한자명(Chiname)	가나명(Japname)	영문명(Engname)
	이칭(Alias)	유형(Type)	발생시간(Time)
원고 (Article)	식별자(ID)	대표명(Name)	원제(Title)
	한자명(Chiname)	번역출처(Tresource)	수록볼륨(Cvolume)
	집필일(Wdate)	범주(Category)	비고(Explanatory)
	핵심어	시작쪽(Beginpage)	종료쪽(Endpage)
	목차		
서적 (Book)	식별자(ID)	대표명(Name)	원제(Title)
	면수	발행일자	부수
권호 (Volume)	식별자(ID)	대표명(Name)	한글명(Korname)
	한자명(Chiname)	가나명(Japname)	영문명(Engname)
	권호(Edition)	부제(Subtitle)	출판사(Phouse)
	발행연도(Pyear)	발행월일(Pdate)	부수
학술지 (Journal)	식별자(ID)	대표명(Name)	한글명(Korname)
	한자명(Chiname)	가나명(Japname)	영문명(Engname)
신문 (Newspaper)	식별자(ID)	대표명(Name)	한글명(Korname)
	한자명(Chiname)	가나명(Japname)	영문명(Engname)
논쟁 (Controversy)	식별자(ID)	대표명(Name)	한글명(Korname)
	한자명(Chiname)	가나명(Japname)	영문명(Engname)
	이칭(Alias)		
학설 (arguMent)	식별자(ID)	대표명(Name)	한글명(Korname)
	한자명(Chiname)	가나명(Japname)	영문명(Engname)
	이칭(Alias)		
용어 (Term)	식별자(ID)	대표명(Name)	한글명(Korname)
	한자명(Chiname)	가나명(Japname)	영문명(Engname)
	대주제(majortheme)	중주제(mediumtheme)	소주제(minortheme)

<표 2> 관계 및 관계속성(2020년 11월 기준)

항목(Class)	관계(Relation)	관계 속성
인물(Person) → 인물(Person)	hasSon : 아들일다	
	hasDaughter : 딸일다	
	isSpouseOf : 배우자로삼다(양방향)	
	isCousinOf : 사촌이다(양방향)	
	hasMaster : 스승으로삼다	
	hasMet : 만나다	
인물(Person) → 단체(Group)	isMemberOf : 구성원이다	직위, 기간
인물(Person) → 기관(Institution)	gradutes : 졸업하다	
	takesCourseAt : 수학하다	기간
	worksAt : 근무하다	직위, 기간
인물(Person) → 공간(Location)	livesIn : 거주하다	
	isBornIn : 출신이다	
	visits : 방문하다	
기관(Institution) → 기관(Institution)	isAdjunctTo : 부속이다	
	isPredecessorOf : 전신이다	
기관(Institution) → 공간(Location)	isLocatedIn : 위치하다	
사건(Event) → 인물(Person)	isRelatedTo : 관련 있다	
사건(Event) → 공간(Location)	isSetIn : 벌어지다	
논쟁(Controversy) → 용어(Term)	isAbout : 에_대한_것이다	
학설(arguMent) → 논쟁(Controversy)	isInvolvedIn : 포함되다	
원고(Article) → 학설(arguMent)	insists : 주장하다	
원고(Article) → 인물(Person)	writer : 저술자이다	
원고(Article) → 원고(Article)	cites : 인용하다	
	isTranslationOf : 번역본이다	
원고(Article) → 서적(Book)	isChapterOf : 수록되다	
	cites : 인용하다	
원고(Article) → 권호(Volume)	isPartOf : 게재되다	
원고(Article) → 용어(Term)	mentions : 언급하다	

항목(Class)	관계(Relation)	관계 속성
서적(Book) → 학설(arguMent)	insists : 주장하다	
서적(Book) → 인물(Person)	creator : 저작자이다	
	editor : 편집자이다	
서적(Book) → 기관(Institution)	publisher : 간행처이다	
서적(Book) → 서적(Book)	isPhotoprintOf : 영인본이다	
	isReprintOf : 번각본이다	
권호(Volume) → 기관(Institution)	publisher : 간행처이다	
학술지(Journal) → 권호(Volume)	includes : 포함하다	

북한 인문학 시맨틱 데이터 아카이브의 뼈대를 이루는 온톨로지 설계는 인문학적 통찰과 상호 토론이 필요한 공동 작업이자 연구 과정이었다. 제1차 연도는 역사학과 고고학, 그 가운데서도 주로 역사학을 대상으로 했다. 따라서 위 온톨로지는 주로 역사학계의 학문 관행을 반영한 결과라고 할 수 있다. 앞으로 문학, 고전학 등으로 대상 분야를 넓혀 감에 따라 더 범용적인 온톨로지를 얻기 위해 항목과 관계에 대한 수정이 필요할 것이다. 이렇게 얻은 ‘표준적 데이터 모델’에 대한 경험은 북한 인문학은 물론 한국 근대 지식을 분절화하여 디지털·웹 환경에 기록하는 데 중요한 밑거름이 될 것이다.

### 2.3. 아카이브 구축

대상 자료의 입력은 크게 인물 정보와 문헌 정보의 두 방향에서 진행했다. 인물 정보는 ① 북한 인문학 관련 문헌의 저자 및 역자 목록화 → ② 인물에 관한 각종 정보 수집 및 정리 → ③ 수집한 정보를 설계된 온톨로지에 따라 구글 스프레드시트에 입력 → ④ 토의를 통해 ‘기관’, ‘단체’ 등 항목 수립 기준 마련 → ⑤ 합의된 기준에 따라 1차 입력 데이터를 검토 및 수정하는 과정을 거쳤다.

인물 정보를 수집하는 단계에서는 생몰연월, 가족관계 등 인물의 기본정보와 더불어 인물의 학력, 활동, 교유관계 등의 정보를 다양한 1·2차 자료를 토대로 파악했다. 다만 북한 자료에 대한 접근이 제약된 탓에 상당수의 정보가 불확실하거나 파편적이었다. 특히 식민지기에 뚜렷한 족적이 없거나 북한 정계에서 활동하지 않은 북한 인문학자의 경우 정보가 거의 없거나 때로는 수집한 정보들이 서로 충돌하는 경우가 많았다. 이를 해결하기 위해 먼저 출처가 명확한 1차 사료를 기준으로 삼았다. 1차 사료에서 정보가 파악되지 않는 경우 신뢰성이 높다고 여겨지는 2차 자료를 토대로 사실관계를 정리했다.

사실관계가 충돌하는 경우는 입력자의 판단에 따랐다. 그리고 충돌하는 사실의 근거와 입력자의 판단 과정은 모두 기록을 남겨두었다. 온톨로지에 따른 데이터베이스에는 입력자가 판단한 결과물만 반영되어 있지만, 서술형 위키 문서에서는 사실의 엇갈림 등을 풀어서 설명할 수 있을 것이다. 나아가 북한 인문학 아카이브에 접속하는 누구든지 추가적으로 의견을 개진하고 사료를 제시할 수 있고, 이 과정에서 검증된 정보를 데이터베이스에 반영함으로써 더 신뢰성 있는 아카이브를 구축할 수 있을 것이다.

문헌 정보도 설계된 온톨로지 내용을 토대로 관련 정보를 입력했다. 인물 정보와 마찬가지로 ① 북한 인문학 관련 주요 문헌 수집 및 엑셀을 활용한 기초적인 정보 정리 → ② 수집한 기초 정보를 설계한 온톨로지에 맞춰 여러 연구자가 분담하여 구글 스프레드시트 상에 입력 → ③ 실제 입력 과정에서 문제가 된 ‘항목’과 ‘관계’를 어떻게 수정할지, 입력 방식을 어떻게 통일할지 등을 토의 → ④ 입력이 완료된 데이터를 다시 개별 연구자가 분담해 검토하는 과정을 거쳤다.

우선 북한 인문학 관련 문헌의 저자명, 제목, 발행기관, 발행일자 등 기본적인 정보를 수집했다. 실제 입력 과정에서는 이러한 정보를 어떠한 문자와 맞춤법을 사용해 입력할 것인지를 비롯하여 세세한 모든

부분이 토의 대상이 되었다. 이미 설계된 온톨로지를 일부 수정할 필요도 있었는데, 예를 들어 ‘isTranslationOf : 번역본이다’는 초기 온톨로지 설계에는 포함되지 않았으나 원문의 정보를 정리 및 입력하는 과정에서 추가된 ‘관계’이다. 건국 초기 북한 역사학을 대표하는 학술지 『역사제문제』에 실린 논문 중 상당수가 소련 잡지에 실린 글의 번역본이었기 때문이다. 이후 개별 연구자가 1차 입력이 완료된 데이터를 다시 검토하여 오류를 최소화했다.

현재까지 입력을 완료한 문헌은 조선역사편찬위원회에서 발행한 『역사제문제』(1948~50년)와 과학원에서 발행한 『역사과학』(1955~94년)에 실린 논문이다.<sup>8)</sup> 이밖에 『역사과학』 1955년 8월호에 실린 「해방후 10년 간에 발표된 역사 논문 및 단행본 목록」을 토대로 1946년부터 1955년까지 『정로』, 『로동신문』, 『근로자』, 『문화유물』 등에 수록된 역사학 관련 논문 및 단행본 정보를 입력했다.

북한 인문학 시맨틱 데이터 아카이브 구축 제1차 연도는 계획대로 ‘역사학’과 ‘고고학’ 분야를 대상으로 데이터를 구축했다. 위와 같은 입력 과정을 거쳐 6,400여 건(항목 Node 2,900여 건, 관계 Edge 3,500여 건)의 시맨틱 데이터를 구축하였다. 자세한 내용은 <표 3>과 같다.

온톨로지에 입각해 입력한 데이터 시트의 내용을 그 형식에 맞추어 관계형 데이터베이스(RDB, SQL)에 옮겼다. 관계형 데이터베이스로 이전된 데이터를 편집, 가공하여 기초 데이터베이스를 구축했다. 관계형 데이터베이스에 구축된 데이터를 그래프 데이터베이스(GraphDB, No-SQL)에 옮겼다. 데이터 시각화는 그래프 데이터베이스를 바탕으로 관계망 그래프를 작성하는 방식을 취했다.<sup>9)</sup>

8) 『역사과학』의 경우 1962년 과학원출판사, 1964년 사회과학원출판사, 1977년 사회과학출판사, 1978년 과학백과사전출판사, 1988년 과학백과사전종합출판사와 사회과학출판사, 1993년 과학백과사전종합출판사 등으로 발행처가 바뀌었다.

9) 전반적인 기술 환경은 한국학중앙연구원 디지털인문학연구소가 제공하는 솔루션

<표 3> 항목과 관계 데이터 규모(2020년 11월 기준)

항목	노드(Node)	항목	엣지(Edge)
인물(Person)	1042	인물	16
		단체	74
		기관	253
		공간	-
단체(Group)	51	-	-
기관(Institution)	156	기관	16
		공간	5
공간(Location)	4	-	-
사건(Event)	2	인물	4
		단체	-
		기관	-
		공간	-
논쟁(Controversy)	4	용어	-
학설(Argument)	13	논쟁	15
권호(Volume)	210	기관	176
서적(Book)	107	인물	90
		기관	52
		학설	1
원고(Article)	1319	인물	1153
		학설	19
		기관	58
		서적	9
		권호	1289
		용어	-
학술지(Journal)	11	권호	235
		기관	2
신문(Newspaper)	5	권호	3
		기관	2
용어(Term)	-	용어	-
노드 데이터 총 2,900 여건		엣지 데이터 총 3,500 여건	
총 6,400 여건			

션을 통해 마련되었다.

마지막으로 ‘북한 인문학 위키’ 페이지를 작성했다. 위키는 다수의 사용자가 웹 브라우저 상에서 복잡한 HTML을 사용하지 않고 간단한 마크업 언어(markup language)를 다루는 것만으로도 웹문서를 작성하고 관리할 수 있게 고안되었다. 위키는 디지털 환경에서의 글쓰기 특징을 잘 보여주는 동시에, 아날로그 환경에서의 글쓰기 형식이 지닌 미덕 또한 동시에 갖추고 있는 플랫폼이다.<sup>10)</sup> 제1차 연도에는 샘플 페이지로서 식민지 조선과 북한에 걸쳐 활약한 역사학자 김광진의 위키 문서를 작성했다. 인물 위키 문서에는 생애, 연구, 논저 등 인문학적 정보를 대상으로 한 텍스트 기술과 더불어, 타임라인, 전자 지도, 네트워크 그래프 등 멀티미디어 형식의 정보 제공 플랫폼을 포함했다.

### 3. 북한 인문학 시맨틱 데이터 아카이브 활용

2020년 11월 북한 인문학 시맨틱 데이터 아카이브 웹 사이트를 개설했다<sup>11)</sup>. 웹 사이트 홈페이지 메뉴에서 ‘Ontology’를 클릭하면 온톨로지 정보를 확인할 수 있다. ‘Project Wiki’를 클릭한 뒤 다시 ‘데이터’를 클릭하면 ‘북한 인문학 위키’로 연결된다. 현재는 샘플로서 김광진 위키 문서만 있다. ‘Search Data’를 클릭하면 네트워크 그래프 출력을 위한 검색창이 뜬다. 아래에서는 네트워크 그래프와 위키 문서 활용 사례를 설명하겠다.

---

10) 위키 플랫폼에 대해서는, 김지선 · 장문석 · 류인태(2021a), 「공유와 협업의 글쓰기 플랫폼, 위키」, 『한국학연구』 60 참조.

11) 북한 인문학 시맨틱 데이터 아카이브 웹사이트(<http://dh.aks.ac.kr/~nkh>, 검색일 2021.2.24) 참조.

### 3.1. 관계망 그래프 ① : 인물과 기관

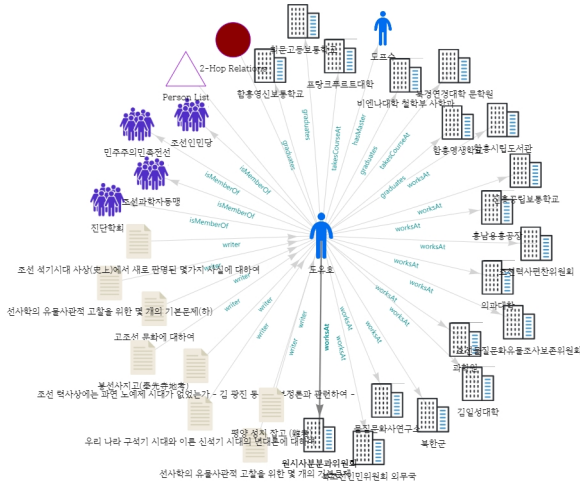
여러 정보 사이의 의미적 관계를 체계화한 온톨로지에 기반하여 산출되는 네트워크 그래프는 기존의 문자열(n-gram) 검색 결과와 어떠한 ‘차별점’을 보여줄 수 있을까? 특히 “인간의 두뇌로 온전히 통제하기 어려운 다량의 복잡한 정보를 입력하여”<sup>12)</sup> 구축한 시맨틱 데이터 아카이브가 산출하는 네트워크 그래프라는 ‘이미지’가 기존의 ‘역사상(像)’에 던지는 의미는 무엇일까? 이를 다음의 구체적인 사례를 중심으로 설명하고자 한다.

검색창에 북한의 주요 역사학자인 도유호를 검색하면 관련 글, 단체, 기관, 사건 등 여러 ‘항목’이 도유호와의 ‘관계’에 따라 그 주변에 무작위로 흩어진 형태로 산출된다. [그림 1]은 이를 항목별로 정리한 것이다.

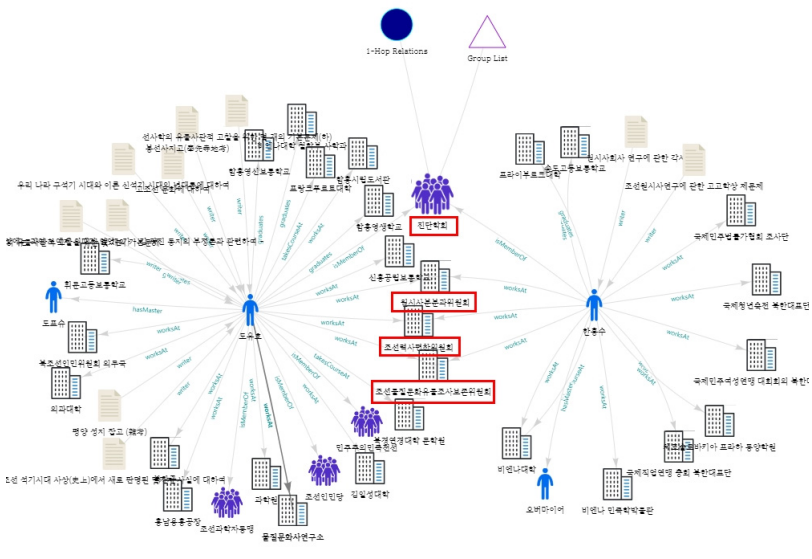
이 그래픽은 일견 기존의 문자열 검색에서 도유호를 검색했을 때 이를 포함한 항목들을 표시되는 것과 유사해 보인다. 그러나 문자열 검색 방식에서 검색된 항목들이 도유호와 어떠한 관계인지를 설명하지 못하는 데 반해, 이 그래픽에서는 ‘항목’ 사이의 ‘관계’가 드러난다. 예컨대 도유호가 함흥영신보통학교, 함흥영생학교, 휘문고등보통학교, 비엔나대학을 ‘graduates’ 즉 졸업했고, 프랑크푸르트대학과 북경연경대학에서 ‘takesCourseAt’ 즉 수학했음을 한 눈에 확인할 수 있다. 나아가 도유호가 참여한 진단학회를 클릭하면 이번에는 진단학회를 중심으로 한 그래픽이 산출되고 다시 ‘2-Hop Relations’를 클릭하여 2단계까지 관계망을 출력하면 [그림 2]가 출력된다.

---

12) 장문석·류인태(2021b), 「디지털 인문학과 한국문학 연구(1) — 작가 연구를 위한 시맨틱 데이터베이스 설계」, 『민족문학사연구』 75(계재예정) 참조.

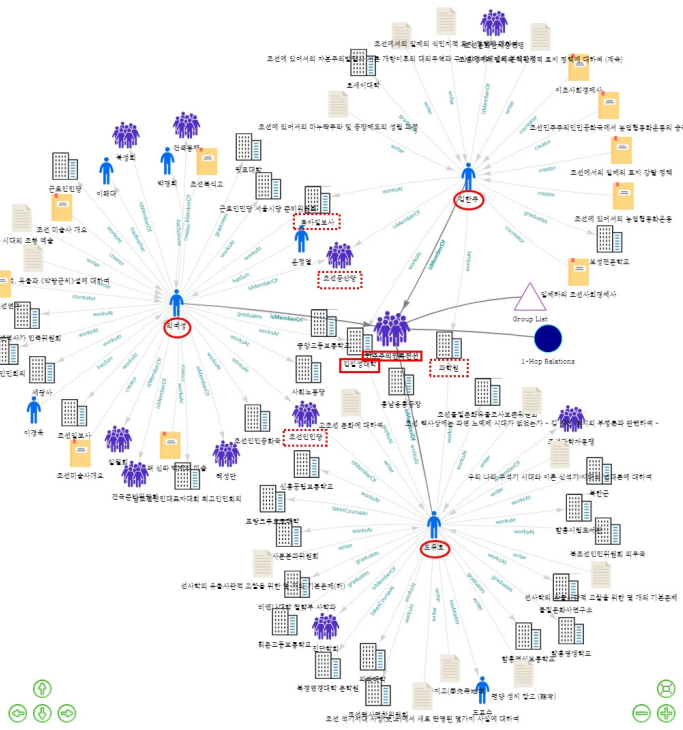


[그림 1] 도유호 관계망



[그림 2] 진단학회 등을 매개로 한 도유호와 한흥수 관계

이 그래픽에서는 도유호가 참여하거나 근무한 여러 단체와 기관 즉 진단학회뿐 아니라 원시사분분과위원회, 조선력사편찬위원회, 조선물질문화유물조사보존위원회 등에서 한홍수라는 학자도 함께 활동했음을 보여준다. 물론 예컨대 진단학회에서 두 사람이 같은 시기에 활동했는지까지는 알 수 없다. 더 상세한 정보는 위키 문서 등을 통해 제공할 수 있을 것이다. 또한, 진단학회에 참여했던 북한 학자들은 훨씬 더 많지만, 현재까지 입력된 데이터가 한정적이라 한홍수만 확인된다는 한계가 존재한다.



[그림 3] 김일성종합대학 등을 매개로 한 도유호, 김한주, 리여성의 관계

같은 방식으로 민주주의민족전선이라는 단체를 클릭한 뒤 2단계 관계망을 출력하면 [그림 3]이 표시된다. 이 그래픽은 여러 단체와 기관을 매개로 리여성, 김한주라는 학자가 도유호와 깊은 관계를 맺고 있었음을 보여준다. 먼저 일차적으로 도유호, 리여성, 김한주는 북한 정권 수립 이전에 민주주의민족전선에서 함께 활동했고, 나중에는 김일성종합대학에서 함께 근무했음을 알 수 있다. 나아가 리여성과 김한주는 조선공산당과 동아일보사에서, 김한주와 도유호는 과학원에서, 리여성과 도유호는 조선인민당에서 각각 함께 활동 혹은 근무했음을 보여준다.

이러한 검색과 클릭을 거쳐 산출된 관계망은 기존의 문자열(n-gram) 방식에서 도유호를 검색했을 때 제공하는 결과물을 모두 포함한다. 그리고 문자열 검색 결과에는 없는 도유호가 각 항목과 맺은 관계를 제시한다. 나아가 도유호가 관계 맺은 기관이나 단체를 클릭했을 때 그 항목을 중심으로 관계된 새로운 인물이 표시되면서, 도유호와 새로운 인물이 어떤 항목을 매개로 관계를 맺고 있는지를 보여준다. 의미적(semantic) 관계에 기반한 데이터 구축은 기존의 문자열(n-gram) 검색 방식을 넘어 새로운 다차원적 관계를 그래픽으로 제시해 준다.

구체적으로 시맨틱 데이터 아카이브에서 검색어 입력과 클릭을 통해 산출되는 그래픽 이미지는 기존의 역사상 혹은 역사서술과 어떠한 차별점을 보여줄 수 있을까? 기존의 역사 인물에 대한 서술은 대체로 연구 논문이라는 아날로그 형태의 기록물로서 작성된다. 북한 인문학 아카이브 구축 과정에서 이용한 도유호 관련 정보도 전문 연구자가 출처를 명확히 하여 밝혀낸 논문에서 가져 왔다.<sup>13)</sup> 다만 기존 논문에서는 연구자의 의도에 따라 다양한 정보가 배치된다. 따라서 독자 역시 논문 작성자가 제시하는 선형적(linear) 역사상을 받아들이게 된다.

13) 한창균(2013), 「도유호와 한홍수: 그들의 행적과 학술 논쟁(1948~1950)」, 『한국고고학보』 87, 한국고고학회, pp.114-115.

시맨틱 데이터 아카이브에서는 예컨대 도유호에 접근한 이용자는 스스로의 판단에 따라 관심 있는 기관이나 인물을 클릭하게 된다. 혹은 우연히 어떤 항목을 클릭함으로써 도유호와 관계된 새로운 항목과 마주하게 된다. 기존의 아날로그 형태의 연구 논문이 선형적 정보라면, 온톨로지에 기반한 데이터를 통해 제시되는 정보는 ‘비선형적’(非線型的, non-linear)이다. 시맨틱 데이터 아카이브에서는 “연구자의 통찰에 근거해 분절된 인문학 데이터의 편찬과 그것을 효과적으로 다룰 수 있는 컴퓨터 기술이 만남으로써, 아날로그 환경에서는 기대하기 어려운” 새로운 역사상의 가능성을 제시한다.<sup>14)</sup>

### 3.2. 관계망 그래프 ② : 학설과 논쟁

1950~60년대 북한 역사학계에서는 삼국시대 사회성격 논쟁, 근현대 시기구분 논쟁, 민족부르주아지 논쟁 등 다양한 학설 논쟁이 벌어졌다. 북한의 학설 논쟁은 국가 기관에 의해 제기되고 일정한 토론을 거친 뒤 정해진 형식에 따라 결론을 맺는 경우가 많다. 따라서 논쟁의 참가자, 매체, 쟁점 등을 분절화하여 관계망으로 표시하는 것이 가능하다. 이 점에 착안하여 북한 인문학 아카이브에서는 ‘학설’과 ‘논쟁’을 항목으로 세워 그 사이의 관계를 표시했다.

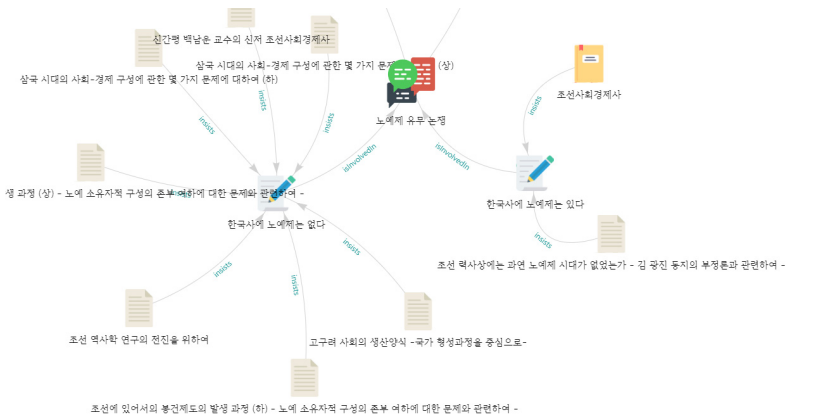
북한의 고대사 연구에서 중요한 쟁점 가운데 하나는 노예제였다. 한국사에 노예제는 있었는가, 있었다면 언제인가라는 문제였다. 검색 창에서 ‘노예’로 검색하면 ‘노예제 유무 논쟁’이라는 항목과 더불어 ‘한국사에 노예제는 있다’, ‘삼국시대는 노예제다’, ‘순장은 노예제의 특징이다’ 등을 비롯해 9개의 학설 항목이 표시된다. 그밖에 제목에 ‘노예’를 포함하는 논문 15편도 제시된다. 여기까지는 단순 문자열 검색 결과와 동일하다. 예컨대 논문 가운데는 「조선에 있어서의 봉건제도

14) 비선형적·조합적 읽기라는 문제의식은, 장문석·류인태(2021b) 참조.

의 발생 과정 (상) — 노예 소유자적 구성의 존부 여하에 대한 문제와 관련하여 — 처럼 노예제 논쟁과 관련 있는 글도 등장하지만, 「《문화통치》시기 일제의 식민지노예교육정책 《개편》 책동」처럼 그다지 관계 없는 논문도 포함된다. 조건을 제한하는 검색 기능을 아직 갖추지 못한 아쉬움이 있다.

하지만 데이터 자체는 의미적으로 구조화되어 축적되어 있기 때문에, 표시된 항목을 클릭해 나가면 시맨틱 데이터 아카이브의 장점을 확인할 수 있다. ‘노예’로 검색해 표시된 네트워크 그래프에서 ‘노예제 유무 논쟁’ 항목을 클릭하면 해당 논쟁이 ‘한국사에 노예제는 있다’와 ‘한국사에 노예제는 없다’라는 두 가지 대립되는 학설로 구성되어 있음을 확인할 수 있다. 여기서 관계망을 2단계까지 클릭하면 [그림 4]와 같이 두 학설을 주장한 책과 논문이 출력된다.

‘노예제 유무 논쟁’을 구성하는 학설 ‘한국사에 노예제는 있다’를 클릭하면 해당 학설은 백남운의 책 『조선사회경제사』와 도유호의 논문 「조선 역사상에는 과연 노예제 시대가 없었는가」에서 주장되었음



[그림 4] ‘노예제 유무 논쟁’의 구조

을 확인할 수 있다. 마찬가지로 이와 대립되는 학설 ‘한국사에 노예제는 없다’는 「조선에 있어서의 봉건제도의 발생 과정 (상) — 노예 소유자적 구성의 존부 여하에 대한 문제와 관련하여 —」를 비롯하여 김광진이 쓴 여러 글에서 주장되었음을 확인할 수 있다. 인물을 중심으로 생각한다면 한편에 김광진, 다른 한편에 백남운과 도유호가 서서 노예제 유무 논쟁을 벌였음을 알 수 있다.

김광진의 여러 글은 ‘한국사에 노예제는 없다’ 이외에 ‘하호는 농노다’, ‘삼국시대는 봉건제다’, ‘순장은 노예제의 특징이 아니다’ 등의 학설도 주장했다. 이들 학설은 모두 ‘삼국시대 사회성격 논쟁’을 구성하는 학설이다. 또한 ‘순장은 노예제의 특징이 아니다’는 ‘고조선 사회성격 논쟁’을 구성하기도 한다.

여기서 ‘삼국시대 사회성격 논쟁’을 클릭하면 해당 논쟁은 [그림 5]와 같이 ‘하호는 농노다’ 대 ‘하호는 노예다’, ‘삼국시대는 봉건제다’ 대 ‘삼국시대는 노예제다’, ‘순장은 노예제의 특징이 아니다’ 대 ‘순장은 노예제의 특징이다’라는 세 쌍의 대립하는 학설로 구성되어 있음을 확인할 수 있다. 각 쌍에서 전자가 김광진의 논문 「삼국 시대의 사회 — 경제 구성에 관한 몇 가지 문제에 대하여」의 주장이고, 후자는



[그림 5] ‘삼국시대 사회성격 논쟁’의 구조

도유호의 논문 「조선 력사상에는 과연 노예제 시대가 없었는가」의 주장이다. ‘삼국시대 사회성격 논쟁’에서 김광진이 하호를 농노라고 보아 삼국시대를 봉건제라고 주장한 데 반해, 도유호는 하호를 노예라고 보아 삼국시대를 노예제라고 주장한 것이다. ‘노예제 유무 논쟁’뿐만 아니라 ‘삼국시대 사회성격 논쟁’에서도 김광진과 도유호는 대립했다.

‘삼국시대 사회성격 논쟁’은 노예제냐 봉건제냐의 논쟁으로 이해되는 경우가 많지만, 그 이면에는 한국사에 노예제가 존재하느냐를 다루는 노예제 유무 논쟁이 자리 잡고 있었다. 이러한 양상을 잘 보여주는 것이 바로 삼국시대가 노예제 사회라고 주장하는 도유호와 삼국시대는 봉건제이고 한국사에 노예제는 존재하지 않는다고 주장하는 김광진의 대립이었음을 [그림 5]에서 확인할 수 있다. 또한 ‘삼국시대 사회성격 논쟁’이 ‘고조선 사회성격 논쟁’과 연결되어 있다는 사실 역시 논자와 학설 등의 항목과 관계를 통해 드러난다.

논쟁을 둘러싼 매체, 기관의 관계가 드러나고, 이는 다시 새로운 논자, 논쟁과의 관계로 이어진다. 김광진과 도유호가 논쟁을 벌인 장은 『력사과학』이라는 학술지였다. 그리고 두 사람은 함께 김일성종합대학의 교원이었다. 김광진은 도쿄상과대학을 졸업했는데 논쟁 상대였던 백남운 역시 같은 학교 출신인 것이 확인된다. 김광진과 백남운은 북한에서뿐만 아니라 식민지기부터 논쟁을 주고받은 사실이 그래프에 드러난다. ‘노예제 유무 논쟁’의 한축을 이룬 ‘한국사에 노예제가 있다’는 주장이 실린 백남운의 책 『조선사회경제사』는 1933년에 출간되었다.

또한 백남운은 『동아일보』(1934.5.4)에 실은 「보전학회논집에 대한 독후감」에서 조선시대는 아시아적 생산양식이 아니라고 주장했다. 이는 김광진이 「이조 말기 조선의 화폐 문제」라는 글에서 주장한 조선시대는 아시아적 생산양식이라는 학설과 대립하며 ‘아시아적 생산양식 논쟁’을 구성한다. 김광진의 글은 바로 보전학회논집 제1호에 실렸

다. 북한에서 김광진과 백남운이 벌인 ‘노예제 유무 논쟁’은 식민지기까지 거슬러 올라가 ‘아시아적 생산양식 논쟁’이라는 또 하나의 논쟁과도 얽혀 있었음을 확인할 수 있다. 김광진이 도쿄상과대학 선배인 백남운과 식민지기에 벌인 논쟁이 서로 다른 논자, 학설과 결합하며 북한 역사학으로 이어진 사실을 관계망 그래프를 통해 확인할 수 있다.

### 3.3. 북한 인문학 위키

네트워크 그래프 상에 표시된 인물 ‘김광진’ 항목을 클릭하면 ‘북한 인문학 위키’ 가운데 김광진 위키 문서로 연결된다.

김광진 위키 문서는 개요, 생애, 연구 등의 서술 부분을 본체로 한다.<sup>15)</sup> 네트워크 그래프로 표시된 정보 등이 문장으로 풀어져 설명되어 있는 셈이다. 온톨로지와 그에 기초한 네트워크 그래프로는 충분히 표시되지 않았던 시간적 관계나 충돌되는 사실 등도 말로 풀어 설명할 수 있다. 텍스트 서술 가운데는 고유명사를 중심으로 다양한 하이퍼링크(hyperlink)가 삽입되어 있다. 독자는 하이퍼링크를 따라 위키 안과 밖의 문서로 이동하면서 자신의 관심에 따라 지식의 맥락을 구성하는 비선형적 읽기를 경험할 수 있다.<sup>16)</sup>

또한, 위키가 지원하는 다채로운 확장 기능을 활용하여 타임라인, 전자 지도, 네트워크 그래프를 장착했다. 타임라인에는 ‘김광진과 그의 시대’라는 표제로 생애의 주요 사건을 표시해 주는 ‘시간 정보’ 그래프를 담았다. 전자 지도에는 평양에서의 출생과 성장, 도쿄 유학, 일본에서의 여행, 귀국 후 경성 즉 서울에서의 교원 생활, 중일전쟁기 평양으로 낙향 등의 ‘공간 정보’를 지도 위에 구현했다. 아울러 연구자라는

15) 김광진 관련 정보는, 홍종욱(2016), 「보성전문학교에서 김일성종합대학으로: 식민지 지식인 김광진의 생애와 경제사 연구」(『역사학보』 232)를 참조했다.

16) 위키 플랫폼에 대해서는, 김지선·장문석·류인태(2021a) 참조.



[그림 6] 김광진 위키 문서

김광진의 정체성에 주목하여 논저 목록을 표로 제시했다. 끝으로 김광진을 중심으로 한 여러 항목의 관계를 보여주는 네트워크 그래프도 포함시켰다.

#### 4. 북한 인문학 시맨틱 데이터 아카이브와 디지털 인문학

북한 인문학 시맨틱 데이터 아카이브 구축과 활용은 ‘공유’와 ‘표현’<sup>17)</sup>이라는 디지털 인문학의 특징을 체감하는 과정이었다.

첫째, 디지털 · 웹 환경이 보장하는 소통과 협업 가능성을 충분히 활용함으로써 질 높은 공동연구를 수행할 수 있었다. 북한 인문학 아카이브의 방향과 구조는 토론과 논쟁을 통해 정해졌다. 온톨로지 설계는 각자 자신이 접한 자료의 성격에 기반해 의견을 개진하고 절충하는

17) 류인태, 「디지털 인문학은 인문학이다」, 『인문논총』 77-3, 2020.8., p. 393.

과정이었다. 구체적인 입력은 여러 연구자가 동시에 접속하여 작업이 가능한 데이터 시트상에서 이루어졌다. 개인연구의 집합으로서의 공동연구를 넘어, 데이터 수집에서 최종적인 시각화까지 생각의 과정을 함께 하는 공동연구를 지향했다. 디지털·웹 환경이 일련의 과정을 시각적으로 공유하면서 작업을 진행할 수 있도록 도왔다.

둘째, 선형적 글쓰기를 넘어서 인문학의 새로운 표현 방식을 시도할 수 있었다. 북한 인문학 아카이브는 연구자가 놓치기 쉬운 의미 관계까지 포착하여 네트워크 그래프로 드러냄으로써 인간의 직관을 넘어서는 이미지를 표현할 수 있었다. 위키 문서 역시 하이퍼링크, 타임라인, 전자 지도 등 디지털·웹 환경 고유의 생산력을 충분히 활용함으로써 선형적 글쓰기를 넘어 새로운 인문학적 표현을 시도할 수 있었다. 연구자가 정해 놓은 길을 따라 서둘러 결론에 이르기보다 무한히 연결되고 확장하는 지식의 관계망 속에서 사람과 세상에 대한 성찰을 얻는 것이야말로 인문학의 이념에 걸맞은 실천이라고 생각된다.

공유와 표현이라는 디지털 인문학의 특징을 본격화하기 위해 북한 인문학 아카이브는 많은 보완이 필요하다. 첫째, 네트워크 그래프를 정교화할 필요가 있다. 기초 질의어(query)를 구축하고 그에 바탕하여 조건 검색이 가능하도록 수정한다면 더 효율적인 결과를 출력할 수 있을 것이다. 둘째, 위키 문서를 확충할 필요가 있다. 현재 인물 위키에 더해 문헌 위키를 하나의 축으로 더한다면 북한 인문학에 대한 더 포괄적인 정보를 제공할 수 있을 것이다. 셋째, 아카이브에 축적된 데이터를 표 형태로 출력할 필요가 있다. 데이터 다운로드가 가능해지면, 외부 데이터 시트가 아니라 위키 페이지 상에서 직접 데이터를 축적하는 것이 가능해질 것이다.

북한 인문학 시맨틱 데이터 아카이브가 북한 인문학 연구와 교육을 위한 디지털·웹 플랫폼으로 자리 잡고, 나아가 한국 인문학의 쇄신을 위한 의미 있는 도전이 되기를 기대한다.

## 참고문헌

### 【는 저】

- 김현 · 임영상 · 김바로(2016), 『디지털 인문학 입문』, 한국외국어대학교 지식출판원.
- 김지선 · 장문석 · 류인태(2021a), 「공유와 협업의 글쓰기 플랫폼, 위키」, 『한국학연구』 60.
- 류인태(2020), 「디지털 인문학은 인문학이다」, 『인문논총』 77-3.
- 장문석 · 류인태(2021b), 「디지털 인문학과 한국문학 연구(1) -작가 연구를 위한 시맨틱 데이터베이스 설계-」, 『민족문화사연구』 75(게재예정).
- 이흥환(2020), 「문서군 가이드: RG 242 (노획 북한 문서, National Archives Collection of Foreign Records Seized, 1675-1983)」, 국립중앙도서관 웹사이트(<https://www.nl.go.kr/NL/contents/N20403010000.do>, 검색일 2021.2.24.).
- 한창균(2013), 「도유호와 한홍수: 그들의 행적과 학술 논쟁(1948~1950)」, 『한국고고학보』 87.
- 홍종욱(2016), 「보성전문학교에서 김일성종합대학으로: 식민지 지식인 김광진의 생애와 경제사 연구」, 『역사학보』 232.
- 국사편찬위원회 한국사 데이터베이스 웹사이트(<http://db.history.go.kr>, 검색일 2021.2.24).
- 미국 우드로윌슨센터 디지털 아카이브(<https://digitalarchive.wilsoncenter.org>, 검색일 2021.2.24).
- 미국 우드로윌슨센터([www.wilsoncenter.org](http://www.wilsoncenter.org), 검색일 2021.2.24).
- 북한 인문학 시맨틱 데이터 아카이브(<http://dh.aks.ac.kr/~nkh>, 검색일 2021.2.24).
- 통일부 북한자료센터 웹사이트(<https://unibook.unikorea.go.kr>, 검색일 2021.2.24).
- 통일부 북한정보포털(<https://nkinfo.unikorea.go.kr>, 2021.2.24).

원고 접수일: 2021년 2월 10일

심사 완료일: 2021년 2월 20일

게재 확정일: 2021년 2월 21일

ABSTRACT

---

Construction and Utilization of the Semantic Data  
Archive for North Korean Humanities

Hong, Jong-wook\* · Kim, Domin\*\* ·  
Kang, Sooyeon\*\*\* · Hong, Suhyeon\*\*\*

The North Korean humanities semantic data archive was designed based on the principle of a qualitative and quantitative approach, and a balance between accuracy and scalability. First, a relational database was built comprising the following three pillars: people, literature, and institutions. Second, an ontology that can express the semantic ‘relation’ between each ‘class’ was designed. Third, a relationship network graph was adopted as a way of visualizing the accumulated information. Fourth, a “North Korean Humanities Wiki” page that is linked to each other through hyperlinks was constructed. The establishment and utilization of the North Korean humanities semantic data archive was accompanied by a process of experiencing the characteristics of digital humanities such as ‘sharing’ and ‘expression’. First, the quality of joint research could be renewed by fully utilizing the communication and collaboration possibilities guaranteed

---

\* First Author, Associate Professor, Institute of Humanities, Seoul National University

\*\* Corresponding Author, Senior Researcher, Institute of Humanities, Seoul National University

\*\*\* Graduate Student, Department of Korean History, Seoul National University

by the digital and web environment. While emphasizing the importance of joint research, the experience of not just fitting into the results of thought but participating in the actual process of thought was a novel experience. The digital and web environment helped to visually share the thought process and proceed with the work. Second, it was possible to attempt a new way of expressing humanities beyond linear writing. The North Korean Humanities Archive was able to express an image that goes beyond human intuition by capturing the semantic relations that are likely to be obscured by the researcher's intuition and by revealing them as a network graph. Wiki documents facilitated experimentation with new humanistic expressions that went beyond linear writing by fully utilizing the unique productivity of digital and web environments such as hyperlinks, timelines, and electronic maps.

