

영어권 한국학 연구 지형도 구축을 위한 온톨로지 기반 데이터 모델링^{*}

이동학 · 이창희 · 김영인^{**}

-
1. 서론
 2. 기존 유관 학술 데이터베이스 검토
 3. 온톨로지 설계
 - 3.1. 온톨로지 설계 원칙
 - 3.2. 온톨로지의 기본적 구조
 - 3.3. 온톨로지 구성 요소의 체계화 및 관계 정의
 4. 결론
-

■ 국문요약

본 연구는 영어권 학술장에서 축적된 한국학 연구성과를 온톨로지 기반 데이터 모델로 구조화하는 과정과 그 설계 논리를 제시한다. 이를 위해 20세기 이후 국내외에서 영어로 작성된 전근대 시기 한국학 연구 학술지 논문 및 단행본을 전수 조사하여 목록화하고 Triple 형태의 시맨틱 데이터 모델을 설계하였다.

기존 유관 학술 데이터베이스인 Signs@40, Connected Papers, WEAS, CAJ를 검토하여 토픽 모델링, 인용 네트워크, 주제 분류 체계 등 이들 데이터 베이스에서 활용되는 방법론을 분석하고, 이를 바탕으로 영어권 한국학 연구성

* 이 논문은 2024년 대한민국 교육부와 한국연구재단의 글로벌 인문·사회융합연구 지원사업의 지원을 받아 수행된 연구임(NRF-2024S1A5C3A02042827).

** 이동학(제1저자): 고려대학교 국어국문학과 박사수료 / lee98dhk@naver.com
이창희(공동저자): 고려대학교 국어국문학과 박사수료 / po77777@korea.ac.kr
김영인(공동저자): 고려대학교 국어국문학과 석사수료 / enqn0124@korea.ac.kr

과의 특성을 반영하고 기초 메타데이터와 세부 정보를 정교하게 수집할 수 있는 온톨로지를 설계하였다. 온톨로지는 Author, Item, Journal, Publisher의 네 개 핵심 클래스를 중심으로 하며, reference, review, isAuthorOf, list, publish, issue의 여섯 가지 관계를 통해 연구성과의 생산·유통·인용 구조를 표현하였다.

이는 학술논문, 단행본, 번역서, 서평 등 다양한 유형의 연구성과를 포괄하면 서도 국내외 학술장의 상호 연결망과 인용 관계, 연구 동향을 다층적으로 분석 할 수 있는 토대를 제시한 것으로, 본 연구에서 제안한 데이터 모델은 향후 한국학의 글로벌화와 국내외 학술장 간 소통 증진을 위한 통합 플랫폼의 기초적 구조로 기능할 수 있을 것이다.

주제어: 영어권 한국학, 온톨로지, 데이터 모델링, 디지털인문학, 시맨틱 데이터

1. 서론

본고는 영어권 학술장에서 축적된 한국학 연구성과를 온톨로지 기반 데이터 모델로 구조화하는 과정과 그 설계 원리를 제시하는 데 목적이 있다. 보다 구체적으로 설명하자면, 국내 및 국외에서 영어로 작성된 동아시아·한국학 연구 학술지 및 한국학을 주제로 다룬 단행본 등의 연구 성과 가운데 '전근대 시기 한국학'을 연구주제로 다룬 20세기 이후의 출판물¹⁾을 대상으로 전수 조사 및 목록화를 수행하고 데이터 간 상호 연결성 및 맥락을 부여하는 Triple 형태의 데이터 모델을 설계하는 것이다. 본고는 고려대학교 한자한문연구소 국제인문학센터에서 수행 중인 2024년 한국연구재단 글로벌인문사회융합연구지원사업 연구과제 「영어권 학술장 한국학 연구 성과의 조사·분석 및 디지털 환경에서의 한국

1) 본고에서는 이를 '영어권 한국학 연구성과'로 축약하여 언급하고자 한다.

학 연구 지형도 구축」(연구기간: 2024.06.01.~2027.05.31.)의 3개년 수행기간 가운데 2차년도 과제 목표인 ‘온톨로지 디자인과 데이터 모델링 및 데이터셋 편찬’의 연구 경과를 중점적으로 다룬다. 본 연구과제는 한국·미주 및 유럽의 한국학 전문가들과 DB분야 전문가의 공동 작업을 통해 20세기 이후 미주 및 유럽 등 영어권 학술장과 국내에서 영어로 발간된 한국학 연구성과를 전수 조사 및 정리하고²⁾ 디지털 도구를 통한 분석·시각화를 거쳐 열람 가능한 형태로 웹상에 구현하는 것을 목표로 한다. 이를 통해 한국 및 국외 영어권 학술장에서 상호 열람 가능한 공개된 ‘영어권 전근대 한국학 데이터베이스’를 구축하여 한국학의 글로벌화와 글로벌 한국학 연구 협업을 위한 통합 플랫폼을 마련하고자 하는 것이 궁극적 목적이다.³⁾

미주 및 유럽 등 영어권 학계에서 제출된 한국학 연구성과의 현황을 검토하는 선행연구는 다음과 같이 제출되어 왔다. 정두희(1996)⁴⁾, 홍승기(1998)⁵⁾는 구미의 주요 출판부에서 출판된 9권의 연구서적과 5권의 연구서적을 중심으로 비판적으로 검토 및 정리하였고 조지형(2002)⁶⁾은 『하버드 한국학 자료 목록』을 정성적·정량적으로 접근하여 미국 내 한국학의 현황과 문제점을 분석하였다.⁷⁾ 이은자(2006)⁸⁾는 영국 대학

-
- 2) 본 연구과제는 전통 한국학에 집중하여 문학·사학·철학·예술·과학·교육·문화 등 전 분야에 걸친 조사를 수행하였다.
 - 3) 연구과제 시작 시점인 2024.06.01. 기준으로 전임교수 연구책임자 1명, 박사급 동연구원 5명, 대학원생 연구원 6명으로 구성된 연구팀을 구성하였고 이후 진행 과정에서 추가 연구원을 충원하여 2025년 12월 기준 총 9인의 연구원을 중심으로 영어권 학술장 연구성과의 전수 조사와 온톨로지 설계 및 데이터셋 구축을 병행하고 있다.
 - 4) 정두희, 「미국에서의 한국사 연구의 경향」, 『구미한국학연구의 경향과 평가』, 서강대 인문학연구소, 1996.
 - 5) 홍승기, 「한국사 연구에 보이는 구미학풍의 영향」, 『진단학보』 86, 진단학회, 1998.
 - 6) 조지형, 「미국에서의 한국학의 흐름과 전망: 안과 밖의 생산적 대화를 위하여」, 『미국사연구』 15, 미국사연구회, 2002.

을 중심으로 한국학 교육과 연구 현황을 검토하였고 김영(2009)⁹⁾은 영국 런던대학 SOAS의 교육과 연구현황을 분석하였다. 김종명(2013)¹⁰⁾은 2012년도 한국학중앙연구원 정책과제(AKSR 2012- P06) “해외한국학 연구업적 리뷰 기반 연구, 2007~2012”의 최종 결과물로 2007년~2012년에 미국을 비롯한 북미주의 한국학계에서 출판 또는 발표된 학술서, 학술논문 및 박사학위논문의 현황을 정리하여 제출하였다. 심재훈(2013)¹¹⁾은 상기 정책과제 수행의 일환으로 영어권 주요 학술지 4종에 실린 한국학 관련 논문 95편을 분석하였다. 이송희(2023)¹²⁾는 고전문학·한문학 분야를 대상으로 1960년부터 2022년까지 출간된 번역서, 개설서, 연구논저 등 단행본과 2000년부터 2022년까지 한국학 관련 전문학술지(해외 4종, 국내 3종)에서 출간된 학술논문의 현황을 조사하였고, 로스킹(2025)은¹³⁾ 북미 영어권 학계에서 2009년 이후 제출된 한국 고전학(古典學)의 주요 연구자와 연구 성과를 검토하고 주목할 만한 연구 성과를 일곱 가지 주제로 분류하였다.

-
- 7) 『하버드 한국학 자료 목록』 중 학위논문의 경우 총 7,011편 가운데 5,968편, 서적은 26,388편 가운데 23,206편, 학술지 발표 논문은 총 49,744 편 가운데 45,409 편이 영어로 작성되어, 세계 한국학의 절대다수가 영어로 이루어지고 있음을 보여준다. 조지형(2002), 178~180면 참고.
 - 8) 이은자, 「영국 대학에서의 한국학 교육 및 연구 현황」, 『한중인문학연구』 17, 한중인문학회, 2006.
 - 9) 김영, 「영국에서의 동아시아 한국학의 연구동향」, 『한국학연구』 21, 인하대학교 한국학연구소, 2009.
 - 10) 김종명, 「서문: 미국의 한국학 연구 동향, 2007~2012」, 『역사와 담론』 67, 호서사학회, 2013.
 - 11) 심재훈, 「추세, 안착, 공명: 영어권 주요 학술지에 나타난 한국학 연구, 2007~2012」, 『역사와 담론』 67, 호서사학회, 2013.
 - 12) 이송희, 「국내외 한국고전문학 분야 영어 간행물 현황 분석: 한국한문학의 해외 확산과 소통을 위한 탐색」, 『Journal of Korean Culture』 60, 고려대학교 한국언어문화학술학관연구소, 2023.
 - 13) 로스킹, 「북미 한국 고전학 연구 동향과 전망」, 『한국한문학연구』 94, 한국한문학회, 2025.

이상의 선행연구를 검토하였을 때 일정 지역 및 시기를 중심으로 연구 현황의 조사가 이루어져 한국 학계에서 해외 한국학계의 연구 동향을 파악할 수 있게끔 한 의의가 있다. 또한 한국 고전학 연구 성과에 대한 일곱 가지 주제 분류가 가능하다는 것은 다수의 한국학 연구 성과가 꾸준히 축적되어 왔다는 방증이다.

그러나 미주, 유럽 등 영어권과 국내에서 영어로 발간된 학술지 전체, 인문·과학·예술 등 전 분야를 대상으로 한 '영어권 한국학'의 전체적인 흐름을 파악하는 데는 한계가 있었다. 학술적 필요성에 의해 관련 연구 성과는 꾸준히 제출되어 왔으나 개인 연구자가 단독으로 수행하기는 실질적으로 불가능하고 프로젝트로 완결지지를 수 있는 작업도 아니기 때문이다. 이에 이상의 선행연구에서 이룩한 성과에 더하여 현재까지 제출된 '영어권 한국학' 연구성과를 총망라하고자 하는 것이 본 연구의 주된 목표 중 하나이다.

한편 국내 및 해외 한국학계의 상호 연결과 소통은 증가하고 있지만 두 학술장 간의 불충분한 학문적 이해는 여전한 과제로 남아있다. 이는 '한국학'이라는 범주 내의 연구임에도 불구하고 참조하는 연구 성과와 연구 방향이 다르다는 점에서 기인한다. 예를 들어 JaHyun Kim Haboush는 여성학 관련 저명한 학자로 국내 학계와는 달리 영어권 학계에서는 반드시 참고해야 할 저자로 인식되어 있다.¹⁴⁾

이러한 현상은 국내외 학술장의 연구 시각과 방법론, 담론이 근본적으로 다른 방향으로 형성되어 있음을 시사한다. 이러한 맥락에서 디지털

14) Google Scholar에 "JaHyun Kim Haboush"를 검색하여 피인용수 1위(Women and Confucian cultures in premodern China, Korea, and Japan, 214회), 2위 (The rise of neo-Confucianism in Korea, 198회)가 국내 발간 연구성과(학술지, 학위논문)에서 인용된 횟수를 살펴보면 각각 214회 중 12회, 198회 중 38회이다. (검색일자 2025.12.25. 기준)

환경에서의 통합적 구현은 더욱 중요한 의미를 지닌다. 국내외 학자가 저술한 영어 한국학 연구성과의 참고문헌, 연구 동향, 상호 연결망을 통합적으로 보여준다면 양 학술장의 연구사적 흐름과 상호 연결 관계를 가시적으로 파악할 수 있으며, 그 과정에서 단절된 지점 또한 확인할 수 있다. 이는 궁극적으로 상호 학술장에 관한 이해를 심화하고 연구 시각을 확장하는 선순환으로 이어질 것으로 기대된다.

이를 위해 본고에서는 디지털 환경에서의 구현을 전제로, 그 기초 자료가 되는 데이터베이스를 어떠한 방식으로 모델링할 것인지를 핵심 문제로 설정하였다. 연구성과의 전반적 동향을 다층적으로 제시하기 위해서는 데이터 수집과 가공을 위한 효과적인 틀이 선행적으로 마련되어야 하며, 이를 위해 데이터셋 편찬에 앞선 단계로서 온톨로지 설계가 필요하다.

온톨로지란 본래 철학에서 존재론을 지칭하는 용어로 존재의 유형과 속성과 본질, 각 존재들 간의 관계, 관계를 통해 맺어진 존재들이 구성하는 총체 등을 연구하는 분야이다. 이것이 정보기술 분야의 개념으로 등장하면서 '정보화의 대상이 되는 세계를 전자적으로 표현할 수 있도록 구성한 데이터 기술 체계'¹⁵⁾의 의미로 사용되었다. 이를 본고의 연구대상인 '국내 및 국외에서 영어로 작성된, 전근대 시기 한국학을 연구주제로 다룬 20세기 이후에 출판된 연구성과'에 적용시켜 본다면 개별 연구성과의 저자, 주제어, 참고문헌 등을 마크업하고 연구성과의 메타데이터와 다양한 세부 정보를 포괄할 수 있는 속성값을 부여함으로써 각 논문의 기본 정보와 상호 인용 관계망을 확인할 수 있는 시맨틱 데이터 구

15) 김현·안승준·류인태, 「데이터 기반 인문학 연구 방법의 모색: 문중 고문서 아카이브와 디지털 인문학의 만남」, 『횡단인문학』 창간호, 숙명여자대학교 숙명인문학연구소, 2018, p.23.

조로 구현할 수 있다. 즉 학계에 제출된 다양한 유형의 연구성과를 최대한 포용하면서도 연도별·지역별 저자 및 연구성과의 인용 추이 혹은 주제·분야별 연구 흐름의 통시적 변화 등을 도출해낼 수 있는 기초 데이터베이스를 구축하는 것이 본고에서 설계한 온톨로지의 방향성이다.

본고는 이상에서 설명한 온톨로지의 구축을 위한 과정과 고려한 지점을 설명하고 온톨로지의 구조와 맥락에 대해 설명하고자 다음과 같은 목차를 구성하였다.

2장에서는 온톨로지를 설계하기 위해 참고한 유관 학술 데이터베이스를 대상으로 전반적인 구조와 주요 내용을 간략히 소개하고, 분석의 주안점을 제시하고자 한다.

3장에서는 온톨로지의 설계를 위해 설정한 원칙과 기준을 제시하고, 온톨로지의 구조와 세부 속성, 관계를 중심으로 각 요소별 설계의 이유와 맥락을 논의할 것이다. 이를 통해 영어권 학술장에서 제출된 방대한 연구성과에 대한 지형도 구축의 기초적 모델을 제시하고자 한다.

2. 기존 유관 학술 데이터베이스 검토

본 연구과제에서 구축하고자 하는 것은 디지털 환경에서 영어권 학술장 한국학 연구성과의 연구 지형도이다. 즉 학술정보의 연구 경향과 상호 영향관계 및 세부적 내용을 디지털 환경에 효과적으로 구현하는 것이 목적이라 할 수 있다. 이는 '학술사'를 웹상에 구현하는 통시적 작업이기도 하지만 동시에 영어권 한국학 연구의 축적과 변화를 하나의 지식정보 체계로 재구성하는 시도이기도 하다.

이러한 목적 하에서 온톨로지의 설계를 위한 사전작업으로 학술정보

를 다룬 기존의 데이터베이스와 디지털인문학 성과를 참고하였다. Signs라는 페미니즘 저널에 대해 토픽 모델링을 토대로 분석하여 정보를 게시한 Signs@40¹⁶⁾, 영어·불어·독어권을 중심으로 출간된 동아시아학 주요 연구성과에 대한 해제목록 DB인 WEAS¹⁷⁾, 학술 연구성과의 상호 인용 연결망을 보여주는 Connected Papers¹⁸⁾, 중국의 대규모 학술정보 데이터베이스인 CNKI¹⁹⁾를 대상으로 데이터 스키마의 구조 및 인터페이스의 측면에서 검토하였다.

이에 본 절에서는 언급한 네 가지 데이터베이스를 간략히 리뷰하고 본 연구과제의 온톨로지 설계에 참고한 지점들에 대해 논의하고자 한다. 각 사례는 방대한 학술 데이터를 구조화 및 시각화하는 접근법을 제시하고 있으므로 온톨로지 설계 시 고려한 지점에 대해 논의하고자 한다.

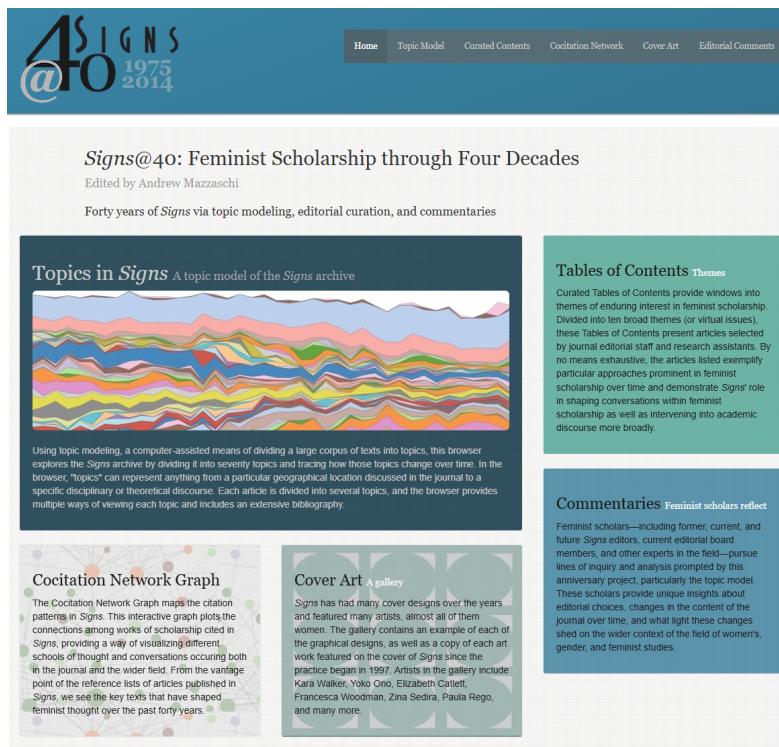
16) URL: <https://signsat40.signsjournal.org/>

17) URL: <https://www.connectedpapers.com/>

18) URL: <https://weas.info/database/>

19) URL: <https://www.cnki.net/>

2.1. Signs@40: 토픽 모델링 기반 학술지 아카이브 분석 플랫폼



〈그림 1〉 Signs@40 홈페이지

Signs@40는 1975~2014년 발간된 미국 펜실베니아대학의 여성학 학술지 Signs를 대상으로 토픽 모델링을 통해 분석한 프로젝트이다. JSTOR에 제공된 학술지 파일을 기반으로 본문에 나타난 단어 빈도 패턴을 바탕으로 70개의 주제로 학술지의 내용을 분류하고, 특정 단어의 빈도수와 시기별 주제의 연구 동향을 파악할 수 있도록 시각화 자료를 제공하였다.

원형 아이콘으로 표시된 주제군에 마우스를 올리면 해당 주제에 속한 단어들을 확인할 수 있고 각 단어들의 빈도수를 확인할 수 있으며, 주제 간 유사도에 따른 군집 표현을 볼 수 있다.

또한 Signs에 수록된 논문의 인용 네트워크 그래프(Cocitation Network Graph)를 제공하고 있다. Signs에서 4회 이상 인용되고, 다른 자료와 함께 4회 이상 공동 인용된 자료를 대상으로 하여 223개의 노드와 426개의 엣지 데이터를 구축하였다. 이는 학문을 다양한 참여자 혹은 자료 간의 대화로 보고, 함께 인용되는 자료들은 특정 분야 내의 한 흐름이나 학파를 대표한다고 하였다.²⁰⁾

이는 대규모의 영어권 한국학 연구성과를 분석하고 시각화하고자 하는 본 연구과제의 목적에 참고할 만한 데이터베이스이다. 여성학 학문 동향을 단어와 주제 중심으로 구현하면서도 인용 관계망을 통해 학맥의 흐름을 추적하도록 한 것은 ‘학술사’를 디지털 환경에 효과적으로 구현한 모범 사례라 할 만하다.

2.2. Connected Papers: 학술 연구성과의 유사성 기반 네트워크망

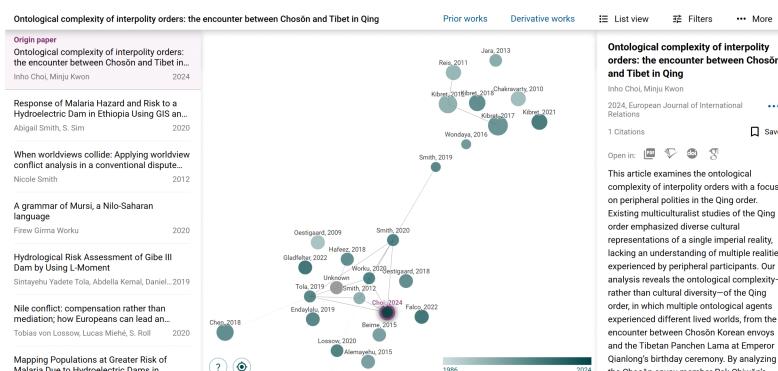
Connected Papers는 특정 논문을 기점으로 관련 연구의 인용 양상을 네트워크로 탐색할 수 있게 해주는 온라인 툴이다. 사용자가 특정 논문을 검색하면 약 5만 건의 학술문헌 데이터를 분석하여 해당 논문과

20) “The assumption operative in this type of graph is that scholarship can be viewed as conversations among various participants represented by particular sources; sources that tend to be cited together are being put into conversation with one another and are thus representative of a particular strain or school of thought within a field.” Signs@40의 Cocitation Network Graph-Exploring the Graph의 설명 참고.

인용 관계가 강한 논문들을 선별한 뒤 노드-링크 그래프 형태로 구현한다.

그래프의 노드는 개별 논문이고, 노드의 크기는 해당 논문의 인용 횟수, 색상은 출판 연도를 반영한다. 논문들은 상호 유사성에 따라 배열되고 유사성은 공동 인용(Co-citation)과 서지 결합(Bibliographic Coupling)에 따라 측정된다. 공동 인용은 두 논문이 동일한 논문들에 인용되었는가를 의미하고 서지 결합은 두 논문이 공통으로 참고하는 문헌이 많은지를 보는 것이다. 이는 사용자가 입력한 논문을 중심으로 개념적·내용적으로 유사한 논문들을 그래프로 연결하여 보여주는 개념 기반의 유사성 네트워크라 할 수 있다.

이는 방대한 자료 속에서 핵심 논문과 내용적으로 관련된 연구의 군집을 직관적인 그래프 형태로 빠르게 파악할 수 있다는 장점이 있다. 다만 한국학 분야에서 오류가 발견되기도 하였는데, 예를 들면 다음과 같다.



〈그림 2〉 Connected Papers의 유사성 기반 네트워크망 조회 사례

위 그래프는 Choi, 2024²¹⁾를 조회한 것이다. 해당 연구성과에 연결

21) Choi, Inho and Minju Kwon. "Ontological complexity of interpoly orders: the encounter between Choson and Tibet in Qing." *European Journal of International Relations*.

된 논문들로는 Tola, 2019²²⁾; Smith, 2020²³⁾; Worku, 2020²⁴⁾ 등이 있다. 이들은 조회한 논문과 학문 분야에서 연관성이 적다.

Choi는 조선 사신과 티벳의 판첸라마가 청 건륭제의 생일 연회에서 마주친 사례를 들어, 청 제국 질서를 중심으로 한 주변 정치체들은 문화적 다양성을 넘어선 존재론적 복잡성을 지니고 있었다는 내용을 다룬다. 반면 Tola는 에티오피아 오모 강에 위치한 Gibe III 댐의 수문학적 위험성을 분석한 공학 논문이고, Smith는 Gibe III 댐 주변의 인구 집단이 말라리아 위험에 더 크게 노출되어 있다는 연구이다. Worku는 에티오피아 오모 계곡 하류에서 사용되는 니로-사하라어계 언어인 무르시 (Mursi)에 관한 박사학위 논문이다. 이는 에티오피아 오모 유역과 관련된 연구와 한국학 관련 연구성과가 함께 조회된 경우로 Choi가 군집의 맥락에 맞지 않게 포함되었음을 확인할 수 있었다.

Connected Paper의 데이터베이스는 Semantic Scholar Paper Corpus와 연결되어 있고²⁵⁾, Semantic Scholar 팀은 과학 분야의 논문, 저자, 과학적 개념, 인용 관계 등 다양한 정보로 이루어진 약 2억 8천만 개의 노드로 구성된 문헌 그래프를 구축하였다.²⁶⁾ Connected

International Relations 31 (2024): 28-52.

- 22) Tola, Sintayehu Yadete, Abdella Kemal and Daniel Reddythota. "Hydrological Risk Assessment of Gibe III Dam by Using L-Moment." *Water science and engineering* 5 (2019): 22.
- 23) Smith, Abigail and Sunhui Sim. "Response of Malaria Hazard and Risk to a Hydroelectric Dam in Ethiopia Using GIS and Remote Sensing." (2020).
- 24) Worku, Firew Girma. "A grammar of Mursi, a Nilo-Saharan language." (2020).
- 25) "Our database is connected to the Semantic Scholar Paper Corpus (licensed under ODC-BY). Their team has done an amazing job of compiling hundreds of millions of published papers across many scientific fields." Connected Papers의 About > How does it work? 5항 참고.
- 26) Ammar, Bridger Waleed, Dirk Groeneveld, Chandra Bhagavatula, Iz Beltagy, Miles Crawford, Doug Downey, Jason Dunkelberger, Ahmed Elgohary, Sergey Feldman, Vu A. Ha, Rodney Michael Kinney, Sebastian

Paper는 애초에 과학계 연구자들을 이용자로 상정하고 있고²⁷⁾, 기반 데이터베이스 역시 과학 연구성과 중심이어서 인문학 연구성과를 조회하는 과정에서 오류가 발생한 것으로 추정된다.

해당 툴은 연구성과 간의 직접적 인용 관계망을 보여주지 않고 유사성을 토대로 군집을 형성하는 것이지만 유사성 측정의 근간이 되는 것은 참고문헌에 관련된 정보였다. 이는 다량의 문헌을 시각화하는 데이터베이스를 구축할 때 참고문헌이 중요한 자료임을 보여주는 사례였다. 참고문헌은 본 연구과제에서도 온톨로지 스키마의 주요한 하나의 축으로 작용한다. 이를 과학이 아닌 인문학, 그 중에서도 전통 한국학의 실정에 맞게 수집 및 가공하는 작업이 온톨로지 및 데이터 스키마 설계 과정에서 요청되었다.

Kohlmeier, Kyle Lo, Tyler C. Murray, Hsu-Han Ooi, Matthew E. Peters, Joanna L. Power, Sam Skjonsberg, Lucy Lu Wang, Christopher Wilhelm, Zheng Yuan, Madeleine van Zuylen and Oren Etzioni. "Construction of the Literature Graph in Semantic Scholar." North American Chapter of the Association for Computational Linguistics (2018).

27) "Connected Papers is a unique, visual tool to help researchers and applied scientists find and explore papers relevant to their field of work." Connected Papers의 About 참고.

2.3. WEAS Database: 구미 동아시아학 문헌 해제형 메타데이 터 DB



〈그림 3〉 WEAS 홈페이지

WEAS 데이터베이스는 2013년~2016년 교육부 토대연구 지원사업으로 추진된 『영어·불어·독어권 동아시아학 주요 연구성과 DB 구축』의 결과물이다. 이 데이터베이스는 북미와 유럽에서 출간된 동아시아학 분야의 학술단행본과 논문을 망라하여 총 9,411건(단행본 3,032권, 논문 6,379편)의 데이터를 수록하였다.

각 데이터는 저자, 제목, 발행연도, 출판사 등의 서지정보와 목차 및 해제로 구성되어 있다. 제목과 목차, 해제는 한국어로 번역되어 있다.

데이터베이스는 검색 기능과 카테고리별 탐색, 학술지별 탐색을 제공한다. 카테고리별 탐색은 분야, 지역, 시대로 구성되어 있고 분야는 주제에 따라 철학 및 종교, 언어학 및 문학, 역사, 예술, 기타로 구분하였다. 지역은 동아시아학의 주요 연구성과를 다루는 만큼 중국·한국·일본·동아시아·기타로 구분하였고 시대는 중국·한국·일본으로 나누어 각각 선사부터 왕조별, 근·현대 등으로 구분하였다.

학술지별 탐색은 학술지별로 수록된 논문 목록과 메타데이터를 제공

한다.

영어권 한국학 연구성과를 총망라하고자 하는 본 연구진의 시각에서 WEAS 데이터베이스는 논문의 기본 메타데이터 제공 방식에서 참고할 만한 것이었다. 우선 주제 분류의 방식에서 연구성과를 하나의 분야에 분속시키지 않고 여러 분야가 중첩되어 있을 경우 여러 개의 분류를 병기하였다. 한국학 분야에서 제출된 연구성과를 검토해 보면 동·서 교류사, 불교 수용사, 의학 문헌사, 외교사, 동서 철학 비교 등 복합적 연구 주제를 다루고 있는 논문들이 다수여서 하나의 주제 하에 분속시키기 어려웠다. 이에 관련된 분류를 병기하는 해당 DB의 방법은 참고할 만하고 현재 온톨로지 설계와 함께 진행중인 구체적 분류체계의 마련을 위해 고려할 부분이기도 하다.

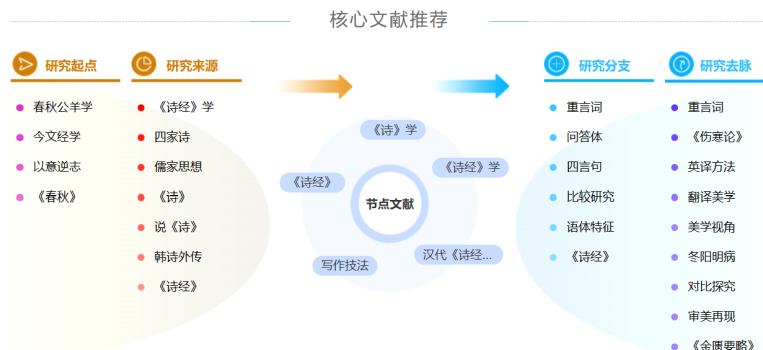
2.4. CAJ: 주제 분류 및 지식 연결형 대규모 학술논문 DB

China Academic Journals(CAJ)는 CNKI(China National Knowledge Infrastructure, 중국학술정보국가인프라)의 하위 서비스로, 중국 본토 발간 학술지의 99.9% 이상을 수록하며, 총 1만 종 이상의 저널이 10개 계열(Series)과 168개 주제(Subject), 3,000여 개 세 부분야로 분류되어 있다.

주요 특징으로는 주제, 인용색인, 발간 주기, 핵심 저널, 키워드, 초록 등을 통한 다양한 검색 방식과 XML 기술을 활용한 신속·정확한 검색, 정렬 및 그룹화, 지식 네트워크 등이 언급되었고 지식 네트워크는 인용, 유사 주제, 저자 소속 등 연구 효율 향상을 위한 네트워크를 제공한다고 하였다.²⁸⁾

28) eastview> China Academic Journals 설명 참고.

온톨로지의 설계를 위해 참고한 부분은 지식 네트워크이고, 그 중 '核心文献推荐'과 '引文网络'의 구조에 주목하였다. 아래는 金前文(2010)에 대한 CAJ의 일부분이다.



〈그림 4〉 CAJ Subject Tree

해당 논문의 核心文献推荐(Subject Tree)은 연구성과에 나타난 문헌 및 인명, 개념어의 층위를 구분하여 제시하고 각 요소를 클릭하면 관련된 논문의 링크로 이동할 수 있다. 명칭에서도 알 수 있듯 논문에 제시된 핵심 문헌들을 제시하였고 研究起點(ROOT)에서 점차 상위 단계로 올라가면서 보다 넓은 범주의 키워드로 통합되고 다시 研究分岐(BRANCH)과 研究去脉(LEAVES)에서 파생되는 모습을 보인다.

위 사진에서는 研究起點(ROOT)에 春秋公羊學, 今文經學, 以意逆志, 『春秋』가 언급되었다. 연구의 시작점(起點)이 되는 학문적 개념과 자료를 나타내는 것으로 보인다. 研究來源(TRUNK)에서는 『詩經』學, 四家

<https://www.eastview.com/resources/journals/caj/#:~:text=China%20Academic%20Journals%20,subjects>

29) 金前文, 「漢賦寫作技法与漢代《詩經》學」, 『荊楚理工學院學報』 25, 2010.

詩, 儒家思想, 『詩經』 등이 제시되었는데 '來源'이라는 명칭에서 보이듯 연구가 근원을 두고 있는 곳이면서도 起點보다는 상위 층위에 있는 것으로 보인다.

節点文献(Node Literature)에서는 『詩』學, 『詩經』學, 漢代 『詩經』, 『詩經』, 寫作技法이라는 5개의 節点(Node)를 제시하였다. '노드'는 네트워크에서 각각의 연결점을 뜻하는데 여기에서는 본 논문에 나타난 학술사상의 주요 지점으로, 다음 단계로 나아가기 위해 거쳐야 하는 주요한 축으로 보인다. 이후 研究分岐(BRANCH)와 研究去脉(LEAVES)에서는 연구의 관점과 주제, 문헌, 개념, 장르 등이 다양하게 뻗어나가는 모습을 확인할 수 있다.

이는 하나의 학술논문 내에 언급된 학문적 개념과 문헌자료가 중요도와 역할에 따라 각자의 층위를 갖고 분포한다는 점을 시각화하면서 후속 연구에서 어떤 개념으로 연결되어 파생되는지 나타내는 일종의 개념사·이론사 구현 시도로 보인다.

한편 CAJ는 논문을 중심으로 한 인용·파인용 관계망을 제공한다.



〈그림 5〉 CAJ 인용관계망

위 사진은 引文網絡(Citation Network)의 분석 화면으로 도식의 중심에는 節点文獻(Node Literature)가 위치한다. 좌측에는 二級參考文獻(Sub-reference)과 參考文獻(Reference), 우측에는 引証文獻(Cited Literature), 二級引証文獻(Sub-cited Literature)이 있다. 상하에 각각 共引文獻(Co-citing Literature)과 同被引文獻(Co-cited Literature)이 있다.

參考文獻은 節点文獻에서 직접 인용한 학술논문 혹은 단행본이고 二級參考文獻은 參考文獻에서 인용한 자료들 중 연구의 배경과 토대가 되는 문헌을 정리한 것이다. 引証文獻은 節点文獻을 인용한 논문이고 二級引証文獻은 引証文獻을 다시 인용한 후속 논문으로 節点文獻이 학술사적 발전 양상을 보여준다. 共引文獻은 節点文獻과 같은 참고문헌을 가진 문헌들로 동일한 학문적 배경을 가진 연구성과이다. 同被引文獻은 節点文獻과 후속 연구인 引証文獻에서 함께 인용된 참고문헌들이다.

이처럼 CAJ의 지식 네트워크는 ‘核心文獻推薦’과 ‘引文網絡’이라는 두 개의 축으로 전개되는데 전자는 논문에 등장하는 학술적 개념과 문헌자료를 중심으로, 후자는 참고문헌을 중심으로 학문적 연원과 후속 연구의 전개 및 파생 양상을 구현하였음을 알 수 있다.

이상으로 온톨로지 설계를 위해 참고한 4개의 학술 데이터베이스를 검토하였다. Connected Papers와 CAJ는 대량의 연구성과 데이터베이스 구축을 위해 개별 논문과 참고문헌의 인용 관계를 중심적으로 활용하였고 Signs@40에서는 대표적 연구성과에 대해 Cocitation Network Graph를 제공함을 알 수 있었다. WEAS는 동아시아학의 영역에서 연구성과를 어떠한 주제 하에 분속시킬 것인가의 정리된 분류 체계를 보여주었다.

3. 온톨로지 설계

영어권 한국학 연구성과를 온톨로지 기반 데이터 모델로 구조화하는 과정은 단순한 데이터 구축을 넘어 학술사적 전개 양상을 디지털 환경에서 재현하는 작업이다. 즉 연구성과 전체를 포용하면서도 각 연구성과가 국내외에서 어떠한 양상으로 제출되어 왔는지, 국내와 국외 연구성과의 연속성 및 상호 영향관계는 어떠한지, 주로 참고한 문헌은 어떠한지 등을 통합적 네트워크로 재구성하고자 하는 것이다.

본 장에서는 온톨로지 설계의 원칙을 서술하고, 이에 기반하여 온톨로지의 기본적 구조와 개별 요소 및 관계의 설계 맥락에 대해 논의하고자 한다.

3.1. 온톨로지 설계 원칙

3.1.1. 연구 성과의 유형과 형태적 다양성 수용

본 연구의 변별점은 데이터셋 구축에 앞서 국내외에서 제출된 영어권 한국학 연구성과에 대한 전수 조사에 있다.

자료 수집의 범주에 대해 설명하자면 국내에서 발간된 영문학술지, 해외에서 발간된 영문학술지, 국내외의 단행본과 잡지를 대상으로 하였다. 학문 범주로는 문학·사학·철학·예술·과학·교육·문헌 등 한국학 전 분야에 걸친 조사를 수행하였고 현재의 한류·미디어 등에 관해 제출된 연구성과보다는 '전통 한국학'에 초점을 맞추어 진행하였다. 시기적으로는 우선 1894년 이전의 전근대 시기를 대상으로 한 연구로 한정하였으나, 그 이후를 대상으로 하였더라도 문헌과 문체가 전근대 시기

의 영향력 하에 있다면 조사 및 검토 과정에서 연구진의 판단에 따라 포함시켰다.

조사 방법은 우선 국내·국외 학술지의 홈페이지에 접속하여 개별 학술지에 수록된 논문 중 전근대 이전의 한국학을 연구한 논문을 수집하였다. 이후 Choson, Korea, Vernacular 등 한국학 관련 검색어를 통해 보충 작업하였다. 저자(편자/역자), 제목, 저널명, 권호, 발행기관, 페이지, 출간연도 등의 기초적인 메타데이터를 수집하고 해당 연구가 '전통 한국학'에 속하는지 모호할 경우 별도로 표기해 두고 2차적인 검토와 검증을 진행하였다. 단행본의 경우 아시아학회의 News Letter, 학술논문 중 Book review에 언급된 저서를 참고하였고, 하버드대학 도서관의 HOLLIS에 접속하여 한국학 관련 검색어를 입력하여 조사하였다.

이상의 과정을 거쳐 연구 개시일인 2024.06.01.부터 2025.11.8.까지 약 3,549건³⁰⁾의 학술연구성과를 다음과 같이 수집하였다.³¹⁾

저자명	논문명	저널명	호수	발행기관	페이지	출간연도
Sang Yil Kim David Chung Richard D. McBride	Han and Han: Phenomenological Interpretation of Hanism Confucianism and Christianity: Morphological Anxiety Hidden Agendas in the Life Writings of Kim Yusin	ACTA KOREANA	vol.1	계명대학교 한국학연구소	19~25	1998
Adrian Buzo	Worthy Ancestors and Succession to the Throne: On the Office Ranks of the King's Ancestors in Early Silla Society by Peter Banaschak (review)	ACTA KOREANA	vol.1	계명대학교 한국학연구소	41~55	1998
John I. Goulder	Chong Hyagong: Korea's Challenge to Orthodox Neo-Confucianism by Mark Setton (review)	ACTA KOREANA	vol.1	계명대학교 한국학연구소	101~142	1998
James H. Grayson	Korean Adoption and Inheritance: Case Studies in the Creation of a Classical Confucian Society by Mark A. Peterson (review)	ACTA KOREANA	vol.1	계명대학교 한국학연구소	156~157	1998
John Goulder	Tracing the Historical and Cultural Roots of Korean Ethno-Nationalism	ACTA KOREANA	vol.2	계명대학교 한국학연구소	160~163	1998
SaeHyang P. Chung	New Findings on Some Possible Artistic Sources of Kim Hongdo's "Lunch"	ACTA KOREANA	vol.2	계명대학교 한국학연구소	171~173	1998
Kang-nam Oh	Korean Shamanist Ritual: Symbols and Dramas of Transformation by Daniel A. Kister (review)	ACTA KOREANA	vol.2	계명대학교 한국학연구소	23~46	1999
Joe Jungno Ree	The Korean Alphabet: Its History and Structure ed. by Young-Key Kim-Renaud (review)	ACTA KOREANA	vol.2	계명대학교 한국학연구소	67~89	1999
Hyangsoon Yi	The Memoirs of Lady Hyegyong: The Autobiographical Writings of a Crown Princess of Eighteenth-Century Korea by Hyegyong-gung Hong Ssi (review)	ACTA KOREANA	vol.2	계명대학교 한국학연구소	156~159	1999
		ACTA KOREANA	vol.2	계명대학교 한국학연구소	184~186	1999

〈그림 6〉 국내영문학술지 목록 예시(계명대학교 ACTA KOREANA)

30) 추가 조사와 검토를 진행함에 따라 총 논저의 개수는 변동될 수 있음을 밝힌다.

31) 수집된 연구성과에 관한 보다 구체적인 사항은 추후 별도의 논문에서 다룰 예정이다.

저자명	논문명	저널명	호수	발행기관	페이지	출간연도
Rutt, Richard	Songs of Flying Dragons: A Critical Reading (review)	Korean Studies	1	University of Hawai'i Press	275-278	1977
Lee, Byung Won	Structural Formulas of Melodies in the Two Sacred Buddhist Chant S	Korean Studies	1	University of Hawai'i Press	111-116	1977
Kawashima, Fujiya	Historiographic development in South Korea: State and society from	Korean Studies	2	University of Hawai'i Press	29-56	1978
Lee, Kwang-rin	Trends in study of modern Korean history in South Korea	Korean Studies	2	University of Hawai'i Press	57-65	1978
Shultz, Edward J.	Confucian Gentlemen and Barbarian Envoys: The Opening of Korea,	Korean Studies	2	University of Hawai'i Press	200-202	1978
Hatada, Takashi(역자Morris, V. Di An)	Interpretation of the King Kwanggaeto's Inscription	Korean Studies	3	University of Hawai'i Press	1-17	1979
Shultz, Edward J.	Military Revolt in Koryo: The 1170 Coup d'Etat	Korean Studies	3	University of Hawai'i Press	19-48	1979
Nathan, Leonard; Ueda, Makoto	Celebration of Continuity: Themes in Classic East Asian Poetry (review)	Korean Studies	3	University of Hawai'i Press	135-139	1979
Best, Jonathon W.	Notes and Questions Concerning the Samguk Sagi's Chronology of	Korean Studies	3	University of Hawai'i Press	125-134	1979
Ch'oe, Yong-ho	An Outline History of Korean Historiography	Korean Studies	4	University of Hawai'i Press	1-27	1980
Sohn, Ho-min	The State of the Art in the Historical-Comparative Studies of Japan	Korean Studies	4	University of Hawai'i Press	29-50	1980
Janelli, Dawnhee Yim	A Type Index of Korean Folktales (review)	Korean Studies	4	University of Hawai'i Press	137-139	1980
Nelson, Sarah M.	The Prehistory of Korea (review)	Korean Studies	4	University of Hawai'i Press	139-141	1980
Vos, Frits	Tales of the Extraordinary: An Inquiry into the Contents, Nature, and	Korean Studies	5	University of Hawai'i Press	1-2	1981
Yun, Chang Sik	The Structure of the Kuun mong [A dream of nine clouds]	Korean Studies	5	University of Hawai'i Press	27-41	1981
Pihl, Marshall R.	Pansori: The Korean Oral Narrative	Korean Studies	5	University of Hawai'i Press	43-62	1981
Kirkland, J. Russell	The "Horse-riders" in Korea: A Critical Evaluation of a Historical Theor	Korean Studies	5	University of Hawai'i Press	109-128	1981
McLeod, Russell; Bouchez, Daniel	Anthology of Korean Literature: From Early Times to the Nineteenth	Korean Studies	5	University of Hawai'i Press	129-139	1981
Haney, Youngsook Kim	Korean Women: In a Struggle for Humanization (review)	Korean Studies	5	University of Hawai'i Press	144-153	1981
Lee, Peter H.	The Korean Buddhist Canon: A Descriptive Catalogue (review)	Korean Studies	5	University of Hawai'i Press	158-160	1981
Jr., Robert E. Buswell	The Identity of the Pöpchin pyöhaen nok [Dharma collection and s	Korean Studies	6	University of Hawai'i Press	1-16	1982
Clark, Donald N.	Chosön's Founding Fathers: A Study of Merit Subjects in the Early Yi	Korean Studies	6	University of Hawai'i Press	17-40	1982
Kwon, Yonung	The Royal Lecture and Confucian Politics in Early Yi Korea	Korean Studies	6	University of Hawai'i Press	41-62	1982
Ledyard, Gari	Hong Taeyong and His Peking Memoir	Korean Studies	6	University of Hawai'i Press	63-103	1982

〈그림 7〉 해외영문학술지 목록 예시(하와이대학 Korean Studies)

원서명(한국어)	번역서명(한자어)	원작기(한국어)	번역가	출판사(한자어)	출간년도	
간장록	A Korean War Captive in Japan, 1597-1600: The Writings of Kang-Hang	강창기	Haboush, Jellyn Kim, Robinson, Kenneth R.	Columbia University Press	2013	
동경대전	Salvation through Dissent: Tonghak Heterodoxy and Early Modern Korea	최재우	George K. Krammer	University of Hawai'i Press	2013	
일제강점기 역사문화논문집	Imperatives of Culture: Selected Essays on Korean History, Literature, and Society from the Japanese Colonial Era		Hanscom, Christopher et al.	University of Hawai'i Press	2013	
사는 도모의 회상	Reflections of a Zen Buddhist Nun: Essays by Zen Master Kim Hyo		Jin Y. Park	University of Hawai'i Press	2014	
성가사집	A Handbook of Korean Zen Practice: A Mirror on the Son School of Buddhism (Songga Isegi)	봉강경	John Jorgenson	University of Hawai'i Press	2015	
법정기유수집解題	Korea's Great Buddhist Confucian Debate		기희	Muller, A. Charles	University of Hawai'i Press	2015
자성록	A Korean Confucian Way of Life and Thought: The Chasongrok (Record of Self-Reflection) by Yi Hwang (Segey)	이왕경	Edward J. Chung	University of Hawai'i Press	2015	
임동집	Seeking Order in a Turbulent Age: The Writings of Dong Taeh, a Korean Neo-Confucian	정도전	David M. Robinson	University of Hawai'i Press	2016	
고려도경	A Chinese Traveler in Medieval Korea: Xu Jing's Illustrated Account of the Xuante Embassy to Koryo	서금	Sam Varneschi	University of Hawai'i Press	2016	
지눌 대집	Numerous Awareness is Never Dark: The Korean Buddhist Master China's Excerpts on Zen Practice	지눌	Robert J. Ewell, Jr.	University of Hawai'i Press	2016	
언어교과서 1	The Analytics of Dasein, Volume 1: A Korean Syntactic Reading	정의용	Hongkyu Kim	Oxford University Press	2016	

〈그림 8〉 번역서 목록 예시

저자/편자	서명	출판사	출간년도
Hwising Cho	The Power of the Brush: Epistolary Practices in Chosön Korea	University of Washington	2020
Shinyoung Kwon	Moral Authoritarianism: Neighborhood Associations in the Three Koreas, 1931-1945	University of Hawaii Press	2023
Maya K. H. Stiller	Carving Status at Kungmangsan: Elite Graffiti in Premodern Korea	University of Washington	2021
Barbara Wall	The Dynamic Essence of Transmedia Storytelling: A Graphical Approach to the Jo Brill		2024
Sixiang Wang	Boundless Winds of Empire: Rhetoric and Ritual in Early Choson Diplomacy	Columbia University Press	2023
David Fedman, Eleana J. Kim, Forces of Nature: New Perspectives on Korean Environments		Cornell University Press	2023
CedarBough Saej, Andrew DaInvented traditions in North and South Korea		University of Hawai'i Press	2021
Paul S. Cha	Balancing Communities: Nation, State, and Protestant Christianity in Korea, 1884	University of Hawaii Press	2022
Sojin Lim, Niki J.P. Alsford	Routledge Handbook of Contemporary South Korea	Routledge	2022
Wayne Patterson	William Franklin Sands in Late Choson Korea: At the Deathbed of Empire, 1896-1905	Rowman and Littlefield	2021

〈그림 9〉 단독저서 목록 예시

연구 성과의 유형에 대해 정리하자면 가장 큰 분류는 논문과 단행본이다. 학술지 형태로 발간된 학술논문 내에는 연구논문 외에 잡지 형식의 단문, 문헌 목록이 있고 다른 논문에 대한 Article Review와 단행본에 대한 Book Review, 저명한 학자를 초청하여 진행한 Interview, 학술지의 개별 호에 수록된 여러 논문을 소개하는 편집자 서평인 Editor's Note가 있다.

단행본은 한국학 관련 주제에 대한 연구서, 개별 한문 자료에 대한 영문 번역서, 여러 자료에 대한 영문 번역을 수록한 Anthology, 연구논문을 모아 하나의 단행본으로 구성한 논문 선집이 있다. 그 외에는 미주 및 유럽의 대학에서 제출된 영문 학위 논문이 있다.

개별 학술논문은 출간 시기와 학술지별로 다양한 형태서지적 특성을 보인다. 저자와 제목, 초록, 키워드, 서론·본론·결론, 참고문헌의 정형화된 형태가 있는 반면 초록이 없는 경우, 키워드 혹은 참고문헌이 없는 경우, 각주가 없고 미주의 형태로 제공되는 경우 등이 있다. 참고문헌은 원전 자료를 Primary Source로 구분한 경우와 구분하지 않고 알파벳 순서로 다른 논문 및 단행본과 함께 나열한 경우가 있다. 2000년대 중반 이후의 연구 성과는 다소 정형화되어 가는 경향을 보이나 반드시 그렇지는 않음을 확인하였다.

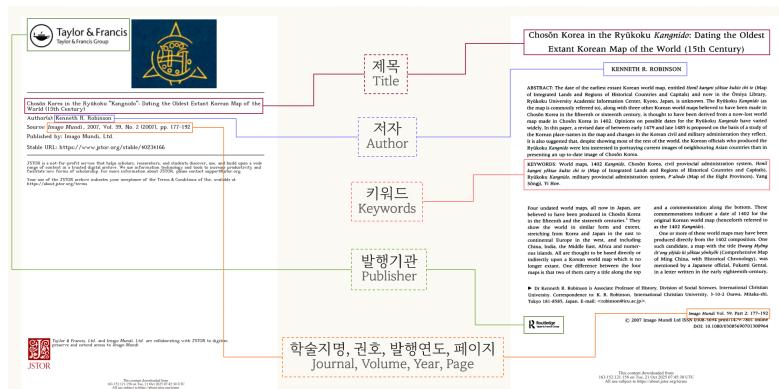
본고의 온톨로지 설계에서는 이러한 다양한 유형과 형식적 요소를 포용하고자 하였다. 이는 연구성과의 유형별·요소별 추이를 보기 위함도 있지만, 새로운 연구성과가 제출될 경우 이를 포함시킬 수 있는 통용 가능한 플랫폼 구축을 목적으로 하기 때문이다. 연구 성과의 유형과 서지적 형태의 불규칙성을 막론하고 데이터셋에 광범위하게 수용하면서도 유형별 데이터를 체계적으로 구분 및 분석 가능한 구조적 틀을 고안하고자 하였다.

3. 1. 2. 메타데이터의 체계적 정리

본고의 데이터 설계 목적은 ① 선행 연구와 후속 연구 속에서의 학술 사적 위치 ② 피인용 횟수와 인용 맥락의 검토 ③ 영어권 학계에서 원전 자료(실록, 일기, 문집 등)가 연구 대상으로 다루어진 추이 ④ 주제별 연구 동향 ⑤ 학술지·발행기관·지역별 연구 동향 등에 대한 다층적·다면적 지형도를 구축하고자 함이다.

이를 위해서는 개별 연구성과의 메타데이터에 대한 상세한 수집이 선행되어야 한다. 추후 디지털 도구를 활용한 분석 및 시각화를 위한 기초 데이터를 마련하기 위함이다. 이에 온톨로지는 메타데이터를 충실히 정리하고 다시 이를 세부적으로 분류하는 방향으로 설계되었다.

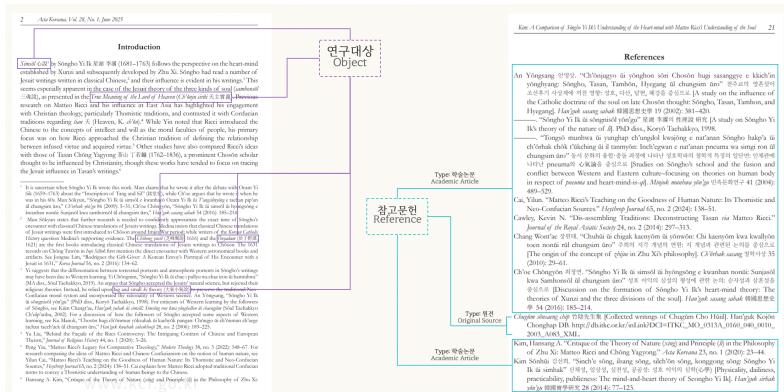
아래는 본고에서 수집한 메타데이터이다.



〈그림 10〉 연구 성과의 메타데이터

개별 연구 성과의 메타데이터로는 제목, 저자, 키워드, 발행기관, 학술지명, 권호, 발행연도, 수록 페이지를 설정하였다. 다만 키워드의 경우 논문에 따라 명시되어 있지 않은 경우가 있어, 키워드가 명시된 경우

에 한하여 수집 대상 데이터로 삼았다.



〈그림 11〉 연구 성과의 내용적 요소

연구 성과의 내용을 구성하는 요소로는 연구대상과 참고문헌을 설정하였고, 실제 데이터의 수집은 연구성과 말미에 제시된 참고문헌(Reference list)을 활용하였다. 참고문헌이 명시되어 있지 않은 경우에는 각주에서 수집하였다. 참고문헌 및 각주에는 학술 연구성과와 한시·산문·소설 등의 작품, 작품이 수록된 책·문집·총서 등의 원전 자료가 함께 언급되어 있다. 이에 모든 연구성과의 참고문헌 데이터를 확보하고 그 속에서 원전 자료를 구분하는 방향으로 데이터 수집의 틀을 고안하였다.

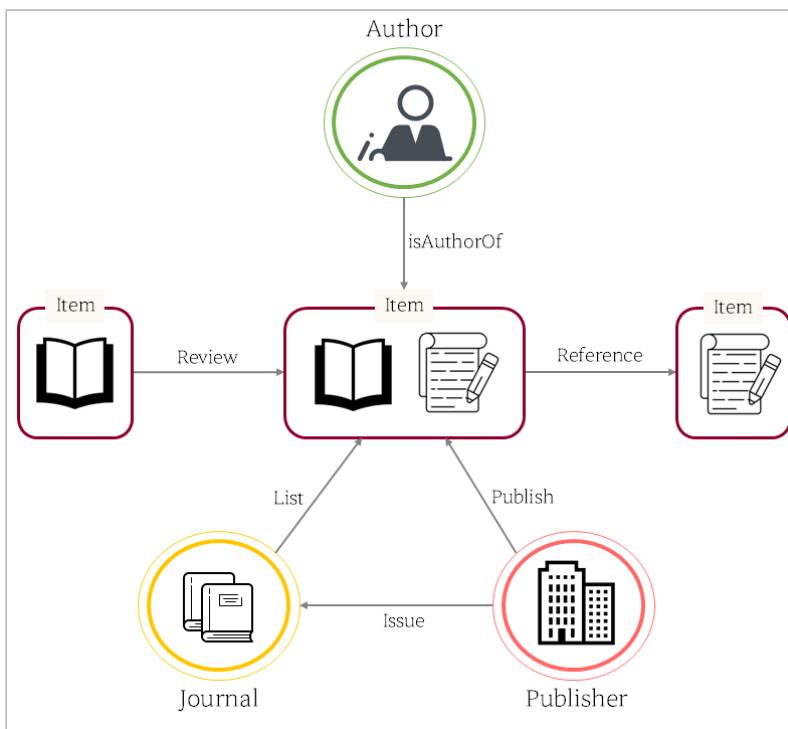
본 연구의 수행 과정에서 메타데이터의 수집을 넘어 연구성과의 내용을 데이터화하기 위해 토픽 모델링으로 본문을 분석하고자 시도하였다. 그러나 1960~1990년대에 발표된 다수의 논문은 기계 판독이 가능한 OCR 처리 PDF가 제공되지 않아, 원문 이미지를 대상으로 별도의 OCR 처리를 시도하였다. 이 과정에서 원본 스캔의 해상도 저하, 인쇄

품질의 불균일성 등으로 인해 텍스트 인식 정확도가 현저히 낮아, 본문을 분석 가능한 수준의 텍스트 데이터로 안정적으로 추출하는 데 한계가 있었다.

또한 OCR 처리가 완료된 PDF 파일을 대상으로 몇 차례 토픽 모델링을 시도하였으나 빈도가 높은 단어로 Chosön, Korea, Vernacular 등이 반복적으로 추출되었다. 그러나 이는 개별 연구논저의 내용과 학술적 맥락이 반영되지 않은 자주 등장하는 ‘단어’에 불과하였다. 또한 모든 논문의 OCR 상태가 균일하지 못하여 추출된 단어 역시 논문에 등장한 빈도를 있는 그대로 반영한다고 볼 수 없었다.

따라서 본 연구는 방대한 분량의 영어권 한국학 연구성과 전체를 목록화하고 데이터셋으로 구축하는 첫 시도로서, 연구성과의 학술사적 분석을 가능하게 하는 기초 자료로서 메타데이터와 참고문헌 정보를 중심으로 이를 체계적으로 정리하는 데 주안점을 두었다.

3.2. 온톨로지의 기본적 구조



〈그림 12〉 온톨로지의 기본적 구조

본고에서는 영어권 학술장을 구성하는 중심 요소를 ‘저자(Author)’, ‘연구성과(Item)’, ‘학술지(Journal)’, ‘발행기관(Publisher)’라는 4개의 노드로 설정하고 이들 사이의 6종의 관계로 온톨로지를 구성하였다.

스키마에서 가장 중점이 되는 부분은 ‘연구성과(Item)’이다. 연구성과는 학술장에서 제출된 단행본과 논문 및 다양한 유형의 관련 서적을 포괄하는 개념으로, 한 연구성과는 다른 연구성과의 참고문헌이면서 개

별 연구성과이기도 하다. 학술사는 선행 연구를 참고하고 인용하며 축적되어 전개되어온바, 디지털 환경에서 이를 구현하기 위해 연구성과 및 연구성과 간 관계가 중심이 되는 구조를 고안하였다.

연구성과를 둘러싼 주요 요소로는 저자와 학술지 및 발행기관이 있다. 저자의 경우, 개별 저자의 저술 횟수를 파악하고 각 저술이 다시 학계에 얼마나 영향을 주었는지를 확인할 수 있는 지표인 피인용 횟수를 확인한다면, 이를 기반으로 학계에서 어떤 인물을 중심으로 연구성과가 제출되어 왔는가, 국내에서 영향력 있는 학자가 국외의 한국학계에서도 영향력이 있는가, 반대로 국외에서 영향력 있는 학자가 국내 한국학에서도 영향력이 있는가 등 학술사의 일면을 조명할 수 있다.

학술지와 발행기관의 경우, 한국학 연구에서 비중 있게 인용된 혹은 영향력 있는 학술지는 무엇인지, 발행기관은 어느 국가와 지역에 위치하는지, 한국학 연구의 중심지와 상대적 주변은 어디인지 등을 다면적으로 분석하기 위해 개별 노드로 설정하였다.

이상의 네 개 노드로 구성된 구조는 외형상 단순해 보일 수 있으나, 영어권 학술장에서 생산된 한국학 연구성과를 전반적으로 포괄하면서도 각 연구성과의 메타데이터를 충실히 정리하여 심층적인 데이터 분석을 위한 기초를 마련하고자 하는 방향에서 설계되었다.

3.3. 온톨로지 구성 요소의 체계화 및 관계 정의

온톨로지를 구성하는 개별 클래스와 세부 속성을 설계할 때 고려한 지점은 연구성과 전체를 포괄적으로 수용할 수 있는 틀을 만들고, 각 연구성과가 지니는 학술사적 맥락과 전후 관계를 분석 가능한 형태로 구조화하는 것이었다. 이에 아래와 같이 온톨로지의 구성 요소를 체계화하였다.

◎ Item 클래스

속성명	설명	예시
id	식별자	IT001 (추후 일괄 부여)
name	대표명	연구성과001 (추후 일괄 부여)
title	연구성과의 제목	Changing Symbolic Landscapes of Buddhist Temples in Sabi, the Royal Capital of the Paekche Kingdom
author	연구성과의 저자 혹은 번역자	Lee Byongho
type	논문/잡지/단행본/원전자료/디지털자료	논문
subtype	학술논문/학위논문/서평(review)/연구서/번역서	학술논문
Journal	논문을 수록한 학술지의 명칭	ACTA KOREANA
Publisher	학술지 혹은 단행본의 발행기관	계명대학교 한국학연구원
date	학술지 혹은 단행본의 발행연월	2023.6.
VolNo	학술지 권호	26(1)
page	학술지 페이지	1-22
table	단행본의 목차	-
ISBN	단행본의 ISBN	-
keyword	학술논문의 키워드	Royal authority, symbolic landscape, defensive installation, communication route, spatial distribution.
role	연구성과/참고문헌/both	연구성과
area	연구성과의 주제가 속한 분야	분야 확정 후 기입 예정
target	본 자료를 인용한 연구성과	-
url	인터넷자료 주소	-

연구성과 및 참고문헌을 포괄하는 클래스이다. 제목과 저자, 학술지 명, 발행연월 및 기관, 권호, 페이지, 키워드, 단행본의 목차와 ISBN 등 메타데이터를 속성값으로 부여하였다. 그 외에 각 연구성과의 유형을 학술논문 · 잡지 · 단행본 · 원전자료 · 디지털자료로 1차 분류하고, 논문은 학술지논문 · 학위논문 · 서평(review)으로, 단행본은 연구서 · 번역서³²⁾로 2차 분류하여 제출된 연구성과의 모든 유형을 포괄 및 분류할

수 있도록 하였다.

개별 연구성과가 다른 논문 혹은 단행본의 참고문헌으로 인용된 경우 role 속성에 both 값을 부여하는 것으로 하였다. 참고문헌으로만 등장한 연구성과는 ‘참고문헌’, 다른 논문이나 단행본에 인용되지 않은 연구성과는 ‘연구성과’로 표기하도록 하였다.

본고에서 다루는 연구성과는 ‘전근대시기 한국학’을 대상으로 한다. 즉 한시 · 산문 등의 작품과 문집 · 일기 · 실록 · 언해서 · 불경 · 불화 등 고전 자료를 다룬다는 특징이 있다. 참고문헌에 언급된 고전 자료를 하나의 데이터로 다루어 출간 시기별, 학술지별, 지역별, 저자별로 다룬 고전 자료의 양상을 구체적으로 파악할 수 있게끔 하는 것이 기준 학술 데이터베이스와의 변별점이다. 이에 참고문헌에 고전 자료가 언급된 경우 type에 ‘원전자료’를 표기하도록 하였다.

keyword는 각 논문의 주제 혹은 내용을 명시적으로 보여줄 수 있는 정보로, 학술논문에 명시된 경우에 한하여 수집하도록 하였다. 주제어가 있으면 적고, 없으면 공란으로 두도록 하였다. LLM 혹은 토픽모델링을 통해 모든 논문의 키워드를 직접 추출하는 방안도 논의되었으나 이는 임의로 추출한 정보가 저자의 의도와 일치할지 알 수 없는, 연구진이 개입하여 얻은 정보이다. 따라서 저자가 직접 선정한 키워드만을 키워드가 명시된 경우에만 수집하여 데이터의 객관성을 담보하고자 하였다.

target은 Item 클래스가 연구성과와 참고문헌을 동시에 관리하기 때문에 참고문헌이 어느 연구성과에 인용되어 있는가를 식별하기 위해 만든 속성이다. 예를 들어 연구성과 A와 B가 있고 B가 A를 인용했다면 Item 클래스 A의 속성값인 target에 B의 제목을 표기하는 것이다.

32) 연구서는 전근대 한국학을 주제로 다룬 연구서이고 번역서는 여러 한문 작품을 영어로 번역하여 수록한 번역선집(Anthology) 혹은 개별 작품에 대한 번역서이다.

한편 area는 개별 연구성과의 주제와 연구 내용을 아우르는 분야·분류이다. 본 연구는 전근대 이전의 한국학을 주제로 다룬 영어권 학술장의 연구성과를 전수 조사하였으므로 문학, 사학, 철학, 예술, 과학, 교육, 문헌 등 전 분야에 걸친 수집을 진행하였다. area를 속성값으로 둔 목적은 분야별로 제출된 연구성과의 개수와 추이를 추후 통시적으로 살펴보기 위한 기초 데이터를 수집하기 위함이다.

다만 검토 결과 하나의 분야에 분속시키기 어려운 경우가 많았다. 예를 들어 'Activities of the Painters Accompanying the Late Choson Envoys to Japan'³³⁾은 조선후기사에 속하면서도 한·일 외교사, 서화미술사 등의 내용을 담고 있다. 이를 조선후기사로 분류한다면 외교 및 미술사와 관련된 내용을 포괄할 수 없는 것이다. 이러한 경우 논문이 속한 분야를 모두 입력해 두면 해결할 수 있다.

문제는 본고에서 다루는 연구성과를 포괄하기 위해서는 기존의 학술 분류체계가 아닌 새로운 분야·분류를 설정해야 한다는 점이다. 존 둘이 도서관 십진분류법은 한국 학술분류의 실정에 맞지 않았다. 한국에 맞게 변용한 한국십진분류(KDC) 역시 적용하기에 적절치 않았는데, 한국학 범주 바깥의 분류는 사용하기 어렵고 세부 분류 역시 지나치게 상세하였다.

KCI의 학술표준분류 역시 한계가 존재하였다. 예를 들어 한국 불교에 관한 연구는 반드시 철학이 아니더라도 불교미술·사찰·불교수용사·종파 등으로 다양하게 전개되어 왔다. KCI 학술표준분류상 한국불교는 인문학>철학>한국철학>한국불교철학에 위치해 있다. KCI 분류상 인문학>불교학>지역불교및불교사연구가 존재하기는 하나, 본고에서 다루는 한국학 분야의 불교는 '한국불교'가 중점인 데 반하여, KCI 분류상의 한국불교는 전 세계의 불교 중 한국의 불교이므로 애초에 충위가 맞

33) Hong Sun-pyo(2000), 『Korea Journal』 40-4, 한국학중앙연구원 출판부.

지 않는다.

앞서 서술한 문제의식을 기반으로 연구진은 추후 논의를 통해 영어권 학술장에 제출된 한국학 연구성과를 포괄하고 분류할 수 있는 분야를 확정할 예정이며, 분야가 확정되면 area를 기입하는 작업에 착수할 것이다.

◎ Journal 클래스

속성명	예시
id	JR001 (추후 일괄 부여)
name	학술지001 (추후 일괄 부여)
JournalName	ACTA KOREANA
Publisher	계명대학교 한국학연구원

Item을 수록한 학술지를 포괄하는 클래스이다. 학술지명과 학술지를 발간한 발행기관명을 표기하도록 하였다. 연구성과를 가장 활발하게 발간한 학술지가 무엇인지, 시기별 추이는 어떠한지를 알기 위해 연구성과를 수록한 학술지의 총 목록이 정리되어야 하겠다. 해당 클래스는 이를 위한 기초 데이터를 저장하는 역할을 한다.

◎ Author 클래스

속성명	예시
id	AU001 (추후 일괄 부여)
name	저자001 (추후 일괄 부여)
AuthorName	Lee Byongho
alternativeName	-

연구성과의 저자 혹은 번역자를 담는 클래스이다. 한국학 연구성과의 지형도를 구성하는 주요 요소 중 하나는 저자로, 저자가 저술한 연구성과의 개수가 많고 각 연구성과가 이후의 연구에 많이 인용되었다면 이는

저자가 영어권 학계에 영향을 미친 주요 인물이라는 점을 나타낼 수 있다. 한편 본 연구에서는 연구성과의 번역자 역시 학술적 생산 과정에 참여한 주체로 간주하여, 저자 클래스에 포함하여 설계하였다.

◎ Publisher 클래스

속성명	설명
id	PB001 (추후 일괄 부여)
name	발행기관001 (추후 일괄 부여)
type	연구소
PubName	계명대학교 한국학연구원
latitude	35.8559055953885
longitude	128.484579003028

학술지 혹은 단행본의 발행기관을 담는 클래스이다. 발행기관의 type은 연구소, 학과, 학회, 국공립기관, 사설기관, 대학으로 구분하였다.³⁴⁾ 각 발행기관의 위도와 경도를 속성값으로 부여하였는데, 이는 추후 연구성과가 발간된 지역과 상호 영향관계를 지도 상에 매핑하기 위해 수집하고자 하는 것이다.

◎ 클래스 간 관계 설계

Source (A)	Target (B)	Relation
Item [role:연구성과]	Item [role: 참고문헌]	reference (참조하다)
Item	Item	review (리뷰하다)
Author	Item	isAuthorOf (저자이다)
Journal	Item [type: 논문]	list (수록하다)
Publisher	Item [type: 단행본]	publish (단행본을 발행하다)
Publisher	Journal	issue (학술지를 발행하다)

34) 본 연구에서는 발행기관의 type을 상호 배타적인 범주로 설정하기보다는, 발행 단위의 성격을 설명하기 위한 분류로 사용하였다. 이에 따라 연구소, 국공립기관, 대학 등 일부 유형은 기관의 실제 기능과 발행 방식에 따라 중첩될 수 있으며, 분류 시에는 최종 발행 명의를 기준으로 하였다.

각 클래스는 Item을 중심으로 관계를 맺도록 설계하였다. Item에 수록된 연구성과가 참고문헌을 통해 다른 연구성과를 인용한 경우에는 'reference' 관계를 부여하였으며, 이는 연구성과 간 인용 관계를 명시적으로 표현하고 인용 관계망을 구현하기 위한 핵심 관계가 된다. 또한 Item에 수록된 논문이 단행본이나 타 논문에 대한 리뷰 논문인 경우 'review' 관계를 부여하여, 영문 한국학 연구성과에서 리뷰 논문의 비중이 높다는 자료적 특성을 반영하도록 하였다.

Item과 저자인 Author는 'isAuthorOf' 관계로 연결하였으며, Publisher는 단행본을 발행하는 경우 'Publish' 관계로, 학술지를 발행하는 경우에는 'Issue' 관계로 연결하였다. 한편 학술지에 수록된 논문에 대해서는 Journal이 Item을 'list'하는 관계를 설정하여, 학술지 단위의 수록 구조를 표현하였다.

4. 결론

본고는 영어권 학술장에서 제출된 전근대 한국학 연구성과를 온톨로지 기반 데이터 모델로 구조화하는 과정과 설계 맥락을 제시하였다.

본 연구는 미주 및 유럽 등 영어권과 국내에서 영어로 발간된 한국학 연구성과를 총망라할 수 있는 데이터 수집의 틀을 마련하였다. 특정 지역이나 시기에 한정될 수밖에 없는 기존 선행연구의 한계를 극복하고 이를 디지털 도구로 분석하기 위한 기초 데이터 모델을 견고하게 구축하고자 하였다.

이는 연구성과의 유형과 형태적 다양성을 수용하는 방향으로 이루어져 학술논문, 학위논문, 단행본, 번역서, 서평 등 다양한 유형의 연구성

과를 포함하면서도 각각의 특성을 반영한 세부 분류 체계를 마련하였다. 이는 새로운 연구성과가 지속적으로 추가될 수 있는 확장 가능한 플랫폼 구축을 가능하게 한다.

또한 저자, 제목, 키워드, 학술지명 등 기본 서지정보뿐만 아니라 참고문헌과 참고문헌에 언급된 원전 자료, 발행 기관의 지리적 정보 등을 상세히 수집하여 각 연구성과의 학술사적 위치, 피인용 추이, 문헌별·지역별·시기별 연구 동향 등을 다각도로 분석할 수 있는 데이터셋 구조를 구축하였다.

향후 연구 과제는 영어권 한국학 연구 성과를 포괄할 수 있는 분야 및 분류 체계를 확정하고 이를 개별 연구 성과에 적용하면서 각 연구 성과의 내부 내용적 요소에 접근할 수 있는 방안을 보다 충실하게 보완하는 것이라 하겠다.

영어권 한국학 연구성과를 전수 조사하고 이에 대한 온톨로지 기반 데이터 모델링을 진행한 것은 디지털 환경에서 한국학 연구 지형도를 구축하여 국내외 학술장 간의 불충분한 학술적 이해를 해소하고 단절된 부분을 발견하여 상호 소통을 증진시키기 위한 선구적 시도로, 한국학의 글로벌화와 글로벌 한국학 연구 협업을 위한 중요한 토대가 되기를 기대한다.

참고문헌

* 학술논문

- 정두희, 「미국에서의 한국사 연구의 경향」, 『구미한국학연구의 경향과 평가』, 서강대 인문학연구소, 1996.
홍승기, 「한국사 연구에 보이는 구미학풍의 영향」, 『진단학보』 86, 진단학회,

1998.

- 조지형, 「미국에서의 한국학의 흐름과 전망: 안과 밖의 생산적 대화를 위하여」, 『미국사연구』 15, 미국사연구회, 2002.
- 이은자, 「영국 대학에서의 한국학 교육 및 연구 현황」, 『한중인문학연구』 17, 한중인문학회, 2006.
- 김영, 「영국에서의 동아시아 한국학의 연구동향」, 『한국학연구』 21, 인하대학교 한국학연구소, 2009.
- 김종명, 「서문: 미국의 한국학 연구 동향, 2007~2012」, 『역사와 담론』 67, 호서사학회, 2013.
- 심재훈, 「추세, 안착, 공명: 영어권 주요 학술지에 나타난 한국학 연구, 2007~2012」, 『역사와 담론』 67, 호서사학회, 2013.
- 이송희, 「국내외 한국고전문학 분야 영어 간행물 현황 분석: 한국한문학의 해외 확산과 소통을 위한 탐색」, 『Journal of Korean Culture』 60, 고려대학교 한국언어문화학술확산연구소, 2023.
- 로스킹, 「북미 한국 고전학 연구 동향과 전망」, 『한국한문학연구』 94, 한국한문학회, 2025.
- 김현 · 안승준 · 류인태, 「데이터 기반 인문학 연구 방법의 모색: 문중 고문서 아카이브와 디지털 인문학의 만남」, 『횡단인문학』 창간호, 숙명여자대학교 숙명인문학연구소, 2018.

* 인터넷자료

Signs@40 <https://signsat40.signsjournal.org/>

Connected Papers <https://www.connectedpapers.com/>

WEAS DB <https://weas.info/database/>

CAJ <https://cajviewer.cnki.net/>

고려대학교 한자한문연구소 국제인문학센터>사업소개

https://kuhjhm.korea.ac.kr/hjhm/research/global_intro.do

eastview <https://www.eastview.com/>

Abstract

Ontology-Based Data Modeling for Constructing a Research Landscape of Korean Studies in the English-Speaking Academy

Lee, Dong-hak & Lee, Chang-hee & Kim, Young-in*

This paper articulates the design principles and methodological process involved in constructing an ontology-based data model for structuring scholarly research on pre-modern Korea produced within the Anglophone academic sphere. It undertakes a comprehensive survey and cataloguing of English-language journal articles and monographs on pre-modern Korean studies published since the twentieth century, both within and beyond Korea, and proposes a semantic data model predicated on RDF-style triples.

As a methodological point of departure, this paper examines existing scholarly databases—Signs@40, Connected Papers, WEAS, and CAJ—focusing on their respective approaches to topic modeling, citation network analysis, and subject classification. Drawing on this comparative analysis, the proposed ontology is designed to capture the distinctive characteristics of Anglophone Korean Studies while enabling the systematic and fine-grained collection of both foundational bibliographic information and detailed scholarly relations.

The ontology comprises four core classes—Author, Item, Journal, and Publisher—and six relational properties (reference, review, isAuthorOf, list, publish, and issue). This structure accommodates a wide range of scholarly

* (First Author) Lee, Dong-hak: Ph.d candidate, Korea Language and Literature, Korea University / E-mail: lee98dhk@naver.com

(Co-Author) Lee, Chang-hee: Ph.d candidate, Korea Language and Literature, Korea University / E-mail: po77777@korea.ac.kr

(Co-Author) Kim, Young-in: Master's student, Korea Language and Literature, Korea University / E-mail: enqn0124@korea.ac.kr

outputs, including journal articles, monographs, translations, and reviews, while providing a robust framework for multi-layered analysis of transnational academic networks, citation practices, and evolving research trends.

By establishing a scalable and interoperable semantic infrastructure, this paper offers a foundational model for an integrated digital platform that can facilitate cross-regional scholarly communication and contribute to the further globalization of Korean Studies.

Key Words

Anglophone Korean Studies, Ontology, Data Modeling, Digital Humanities, Semantic Data

논문접수일: 2025.12.10. 심사완료일: 2025.12.19. 게재확정일: 2025.12.30.

