

특집 논문

북한 인문학 데이터 아카이브와 디지털 인문학 교육

: 빅데이터 혁신융합대학 교과목 개발을 중심으로*

류인태 (고려대학교)**

허수 (서울대학교)***

홍종욱 (서울대학교)****

국문요약

이 글은 북한 인문학 데이터 아카이브를 설계·구축한 논리와 이를 매개로 진행된 디지털 인문학 교육 경험을 정리한 것이다.

첫째, 북한 인문학 데이터 아카이브는 학술 정보가 내포한 입체적 성격과 그것들 사이의 복잡한 지식 관계망을 효과적으로 표현하기 위해, 시맨틱 데이터 프로세싱(Semantic

DOI URL: <https://doi.org/10.35369/jpus.15.1.202306.11>

* 이 논문은 2022년도 서울대학교 통일평화연구원의 재원으로 통일평화기반구축 사업의 지원을 받아 수행된 결과물임. 이 논문의 게재료는 서울대학교 인문학 연구원에서 지원 받았음.

** 제1저자

*** 제2저자

**** 교신저자(hongtsky@snu.ac.kr)

Data Processing) 방법론에 근거하여 설계·구축하였다.

둘째, 서울대학교 인문대학은 교육부의 혁신융합대학 사업의 일환으로 디지털 인문학 교육 과정을 개설·운영하고 있는데, 그 가운데 ‘빅데이터 역사정보’ 과정의 하나로 ‘시맨틱 데이터 편찬’ 교과를 도입·개설하였다.

셋째, ‘시맨틱 데이터 편찬’ 교과 강의는 북한 인문학 데이터 아카이브와 연계하여 운영함으로써, 북한 인문학 및 한국 근대사에 관한 도메인 지식과 디지털 인문학 방법론인 시맨틱 데이터 처리를 동시에 익히는 융합교육이 가능하도록 했다.

넷째, ‘시맨틱 데이터 편찬’ 교과의 수강생은 웹 환경에서 북한 인문학과 한국 근대사에 관한 지식을 다룸으로써, 아날로그 환경에서 지식을 소비하는 수동적 입장에서 벗어나 디지털 환경에서 지식을 생산·유통하는 능동적 경험을 할 수 있었다. 아울러 북한 학계가 통시적·공시적으로 지식의 국제적 네트워크 속에 존재함을 확인하고 냉전과 분단 상황에서 전개된 북한 인문학의 특수성을 이해할 수 있었다.

주제어: 북한인문학, 북한인문학 데이터아카이브, 빅데이터 혁신융합대학, 디지털 인문학

I. 들어가며: 북한 인문학 데이터 아카이브의 교육적 활용

1. 통일·평화기반구축사업과 북한 인문학 데이터 아카이브

서울대학교 인문학연구원은 2020년도부터 2022년도에 걸쳐 서울대학교 통일평화연구원이 주관하는 통일·평화기반구축사업¹⁾의 일환으로 ‘북한 인문학 데이터 아카이브’ 구축 사업(이하 본 연구)을 진행하였다. 해당 연구에 참여한 연구자들은 각기 ‘역사학과 고고학’, ‘한국문학과 외

1) 통일·평화기반구축사업에 대해서는 서울대학교 통일평화연구원 웹사이트 해당 페이지 참조(<https://ipus.snu.ac.kr/sun>).

국문학', '동양고전과 서양고전' 등 개별 인문학 분야에 관한 전문적 이해를 바탕으로 북한의 인문학술 자료를 수집한 뒤, 그에 관한 내용을 디지털 환경에서 데이터로 편찬하고 큐레이션하는 학제적 연구를 시도하였다. 현재까지 진행된 해당 연구의 결과물은 데이터베이스 기반의 웹사이트로 구현되어 있다.²⁾

북한 인문학 데이터 아카이브는 북한에서 그동안 이루어져 온 인문학 연구의 동향을 파악할 수 있는 기초 자료로서 여러 학술지와 그에 실린 논문들 그리고 여타 단행본 등의 학술정보를 정리해서 일반 연구자에게 공개함과 동시에, 필요한 경우 유관 정보를 검색할 수 있는 환경까지 갖추어 놓았다. 특히 현재 서비스되고 있는 대부분의 학술 리포지터리가 단순 문자열(n-gram) 검색을 통해 문헌 자료의 메타데이터 정도만 출력해서 보여주는 수준에 머무르고 있는 것과 달리, 본 연구는 학술지와 논문 간 수록 관계 및 연구자 간 학연과 직연으로 복잡하게 얽힌 북한 학술 정보의 지식 관계망을 효과적으로 들여다볼 수 있게끔 하기 위해, 시맨틱 데이터(semantic data) 기반의 아카이브를 구현하고자 하였다.

이러한 형식의 학술 결과물을 웹상에 구현하는 연구 과업은, 북한에서 그동안 축적되어 온 역사, 문학, 고전에 관한 지식을 다루는 전통적 성격의 인문학 연구 활동이자, 동시에 디지털 환경에서 유관 지식을 데이터로 변환·활용함으로써 기존 인문학 연구의 외연을 더욱 확장하고자 하는 실험적 시도라 할 수 있다. 향후 통일을 대비한 지식을 디지털 환경에서 꾸준히 확충하고 그것을 매개로 북한 인문학을 다루는 전 세계 연구자들 사이의 네트워크를 마련하는 허브로서, 본 연구가 기능할 수 있을

2) <북한 인문학 데이터 아카이브> (<http://dh.aks.ac.kr/~nkh/index.html>). 아카이브 구축 과정에 대해서는 이미 한 차례 학계에 보고한 바가 있다. 홍종욱·강수연·홍수현·김도민, 「북한 인문학 시맨틱 데이터 아카이브 구축과 활용」 『인문논총』 제78권 1호 (2021).

것이라는 기대가 있다.

2. 인문학 교육에서 북한 인문학 데이터 아카이브 활용 가능성

통일·평화기반구축사업의 일환으로 진행된 연구 프로젝트로서 북한 인문학 데이터 아카이브 구축 연구는 2023년 1월 공식적으로 종료되었다. 한편으로 해당 연구에 참여한 연구자들을 중심으로 추후 유관 데이터를 꾸준히 수집·축적해 나감으로써, 북한 인문학 데이터 아카이브를 장기적으로 확장·활용하려는 계획이 있다.

예컨대 본 연구의 장기적 방향 가운데 하나는 아카이브에 기 구축된 데이터와 향후 구축할 데이터를 CSV, JSON, RDF 등의 웹 표준 형식에 맞추어 공공데이터로 가공해서 연구자 및 대중에게 공유하는 것이다.³⁾ 해당 목표를 현실화하기 위해서는, 단순히 북한 인문학에 관한 자료를 많이 수집하여 데이터로 변환한다고 해서 능사가 아니라, 북한 인문학과 관련된 지식을 체계적으로 정리하기 위한 데이터 모델(data model)을 먼저 정교하게 디자인하고, 그에 근거한 데이터를 일정 이상 규모로 가공·편찬한 다음 개별 데이터를 효과적으로 보여줄 수 있는 분석 및 시각화 결과물 형식을 함께 고민해야 한다. 소위 ‘데이터 리터러시(data literacy)’의 맥락에서 인문 데이터 처리(humanities data processing)에 관한 폭넓은 접근이, 북한 인문학 데이터 아카이브를 업데이트해 나가는 과정에서 지속적으로 수행되어야 함을 의미한다.

그런데 그 과정은 연구의 차원에서뿐만 아니라 인문학 교육의 차원에서도 고민할 수 있다. 북한 인문학에 관한 데이터를 다루는 과정에 학생

3) 공공데이터로서 인문학 데이터의 구축 및 가공의 필요성과 그 의미에 관해서는 김바로, 「〈공공데이터법〉과 인문데이터: 공공기관 보유 인문데이터 공개 신청 사례를 중심으로」 『한국고전연구』 제57호 (2022) 참조.

들이 참여한다는 것은, ①북한 인문학 자료를 수집하는 경험을 통해 북한의 학술문화·풍토에 관한 기초 이해를 도모할 수 있고, ②북한 인문학 데이터 모델 디자인에 참여함으로써 인문학 데이터를 가공하기 위한 기본적 문제의식을 익힐 수 있으며, ③북한 인문학 데이터를 편찬하는 본격적 과업을 수행함으로써 북한에 관한 지식을 능동적으로 접하는 가운데, ④북한 인문학 데이터를 분석·시각화하는 과제를 통해서 인문적 성격의 데이터 시각화에 관한 이해와 경험을 확충할 수 있음을 의미한다. 이는 곧 ‘북한 인문학’이라는 도메인 지식을 공부하면서 동시에 ‘데이터 리터러시’라는 디지털 방법론을 익힐 수 있는 소위 ‘융합적’ 성격의 교육이 가능함을 암시한다.

그리고 현재 디지털 인문학 교육(digital humanities education)의 일환으로 데이터 리터러시에 관한 정규 교육 및 세미나와 워크숍 그리고 콜로키엄 형식의 학술행사가 전 세계의 대학(원)에서 꾸준히 개최되고 있음을 고려할 때,⁴⁾ 북한 인문학 데이터 아카이브를 구축한 방법론 및 절차를 교육적 차원에서 활용하고자 하는 것은 자연스러운 일이다. 연구와 교육의 선순환적 프로세스가 갖는 순기능을 고려할 때, 북한 인문학 데이터 아카이브의 교육적 활용은 오히려 더욱 적극적으로 고려해야 할 지점이라 할 수 있다. 더욱이 디지털 환경과 데이터 처리 기술을 접목한 인문학 교육의 방향 및 내용에 관한 논의가 그 어느 때보다 활발한 최근의 시류를 생각할 때도, 북한 인문학 아카이브의 교육적 활용 방안을 고민하는 것은 필연적 흐름이며, 현 시점에서 반드시 고민해야 할 과제라 하겠다.

4) 미국 학계의 디지털 인문학 교육에 관해서는, 이지수·이혜은, 「미국 대학의 디지털 인문학 교육 프로그램 연구」 『정보관리학회지』 제36권 4호 (2019) 참조.

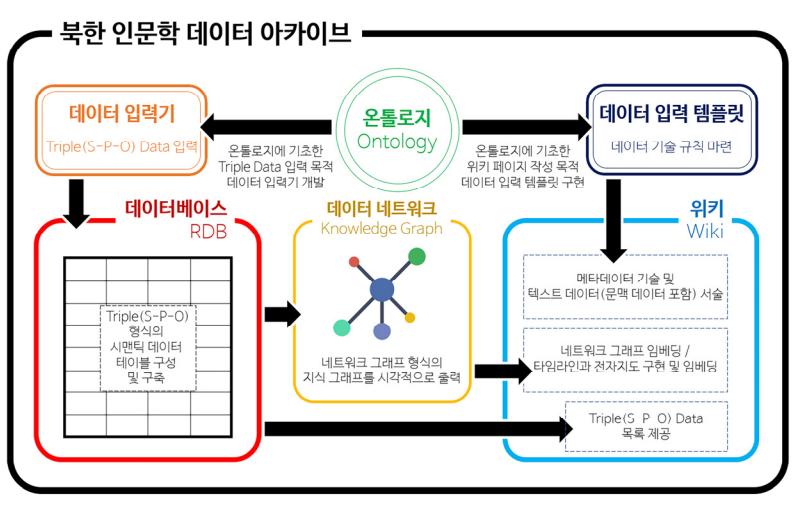
II. 북한 인문학 데이터 아카이브의 설계와 구축

1. 시맨틱 데이터 프로세싱: 인문 데이터 처리 방법론

복잡한 지식을 입체적으로 탐구해야 하는 상황에서 관계형데이터베이스(RDB)와 문자열(n-gram) 검색에 기초한 데이터 처리 기술은 그 한계가 뚜렷하다. 특히 인문학 연구는 각종 문헌을 통해 발견되는 다양한 지식이 입체적으로 결합되는 양상이 보편적이고, 그러한 여러 지식 사이의 의미적 연결 고리를 어떻게 탐색할 것인가의 문제의식이 중심에 자리하기 때문에, 인문학 연구에서 데이터 처리 기술을 활용할 때는, 그러한 맥락에서 유효할 수 있는 구체적 방법론을 고민할 필요가 있다.

그리고 인문학 연구에서 데이터 처리 기술을 활용하고자 하는 근본적 의도는, 학계에서 통용되는 여러 질적(qualitative) 논의의 줄기를 압축하고 체계화할 수 있는 표준적 성격의 데이터 스키마(schema)를 개발하여, 해당 스키마를 기준으로 준용 가능한 형식의 학술 데이터를 디지털 환경에서 편찬·공유하고자 하는 목적과 관련이 있다. 예컨대 인문학 연구자들은 전통적으로 개인에 의한 ‘논문(academic paper) 작성’과 ‘학회지(academic journal) 게재’라는 학술적 인준 형식에 의존해 연구 지식을 공유하는 방식을 취해 왔는데, 디지털 환경에서는 ‘개인적 차원의 글쓰기(writing)’ 결과물을 발표하는 것이 아니라 다수의 연구자가 참여해 편찬한 ‘협업 기반의 데이터셋(dataset)’을 폭넓게 개방·공유함으로써, 유관 지식을 자유롭게 유통하고 그로부터 다채로운 맥락의 학술적 담론을 이끌어 내는 것이 가능하다. 특정 지식이나 분야를 대상으로 한 정교한 데이터셋이 디지털 환경에서 체계적으로 구축될 수 있다면, 소위 ‘공공데이터’로서 해당 학술 데이터가 유관 분야에서 활용될 여지는 무궁무진한 것이다.

〈그림 1〉 북한 인문학 데이터 아카이브 구축 개념도



북한 인문학 데이터 아카이브 편찬 연구는, 디지털 환경에서의 인문학 연구라는 특수성을 고려해 다양한 맥락의 정보가 복잡하게 연결되는 북한 인문학의 세계를 정밀한 체계의 데이터로 재현하고자 하였으며, 웹에서 통용되는 표준적 성격의 포맷을 준용해 학술 데이터를 편찬하고자 하였다. ‘표현’과 ‘공유’는 디지털 인문학의 중심 철학이라 할 수 있는데, 북한 인문학 데이터 아카이브는 그러한 디지털 인문학 연구의 큰 줄기를 기준으로 삼은 것이다.⁵⁾ 그에 따라 시맨틱 데이터 프로세싱(Semantic Data Processing)⁶⁾ 방법론에 입각한 온톨로지(ontology) 디자인과 지식그

5) ‘표현’과 ‘공유’로 대표되는 디지털 인문학의 줄기와 그 구체적 사례에 관해서는 ‘김지선·유인태, 『지식의 공유와 표현 그리고 디지털 인문학: 해외 디지털 인문학 연구 사례 검토』 『인문논총』 제79권 2호 (2022), 391~427쪽.’ 내용을 참고할 수 있다.

6) 시맨틱 데이터 프로세싱은 개별 데이터 사이의 의미적 관계를 전제로 하나의 커다란 지식그래프로서 모든 데이터를 연결하는 데이터 처리 방법론을 말함

래프(knowledge graph) 구현 그리고 위키(wiki)를 활용한 데이터 편찬 환경 마련 등을 데이터 아카이브 구현 과정에서의 필수적인 요건으로 삼았다. <그림 1>은 그러한 여러 요소의 유기적 관계를 표현한 북한 인문학 데이터 아카이브의 구축 개념도이다.

인문학 연구에서 시맨틱 데이터 프로세싱을 활용해 종합적인 성격의 데이터 아카이브를 구현하고자 하는 의도는, 여러 유형의 문헌에 담긴 다채로운 층위의 정보를 균질한 형식으로 정리하되, 그 사이의 복잡한 의미적 관계를 디지털 환경에서 최대한 재현하기 위한 목적에서 비롯하는 것이다. 양적 연구가 아니라 질적 연구로 표방되는 인문학 연구 본연의 맥락을 데이터 처리 기술과 연계한다고 할 때, 가장 적합한 방법론 가운데 하나라 할 수 있다.

북한 인문학 자료의 외연이, 문학, 역사, 철학 등 다채로운 영역의 지식 세계를 포괄하는 만큼 아날로그 환경에서의 선형적 읽기로는, 그와 같은 다양한 맥락으로 얽힌 정보들 사이의 유의미한 지점을 발견하기가 쉽지 않다. 시맨틱 데이터 프로세싱을 통해 편찬한 데이터를 매개로 디지털 환경에서의 비선형적 읽기를 전개함으로써, 북한 인문학과 관련된 수많은 정보 사이의 의미적 관계를 디지털 환경에서 효과적으로 재현할 수 있다.

다. 그래프를 구성하는 각각의 노드(node)로서 식별자(id)를 포함해 대상자원의 고유한 속성(DataProperties)을 정의하고, 개별 노드와 노드 데이터 사이의 의미적 관계(ObjectProperties) 또한 엣지(edge) 데이터로 정의함으로써, 여러 정보와 그것들 사이의 의미적 관계를 Triple(S-P-O, Subject-Predicate-Object) 형식의 데이터로 구축하는 것이 핵심이다. Triple(S-P-O) 형식의 데이터를 다루는 기술들이 다양한 맥락에서 시맨틱 데이터 프로세싱에 활용될 수 있는데, 데이터를 RDF나 OWL 형식으로 가공할 경우 SPARQL 기반의 의미 추론 검색이 가능하고, Graph DB로 구축할 경우 그래프 질의어(Cypher Query)를 활용해 데이터 사이의 의미적 관계를 네트워크 그래프 형식으로 출력해 살펴볼 수 있다.

2. 온톨로지 디자인과 지식그래프 구현

시맨틱 데이터 프로세싱을 위해서는 기본적으로 Triple(S-P-O) 형식의 데이터를 편찬해야 하는데, 구체적으로는 RDF와 OWL 포맷의 데이터 구축 또는 GraphDB 구현을 전제해야 하며, 그것을 위해서는 온톨로지(Ontology)⁷⁾ 디자인이 반드시 이루어져야 한다. 일반적인 데이터베이스 구축의 경우 데이터 모델(Data Model) 디자인을 전제로 한다면, 시맨틱 데이터 구축은 시맨틱 웹(Semantic Web) 기술을 전제로 하기 때문에 더욱 확장된 개념의 데이터 모델로서 도메인 온톨로지 설계가 필수라고 할 수 있다.

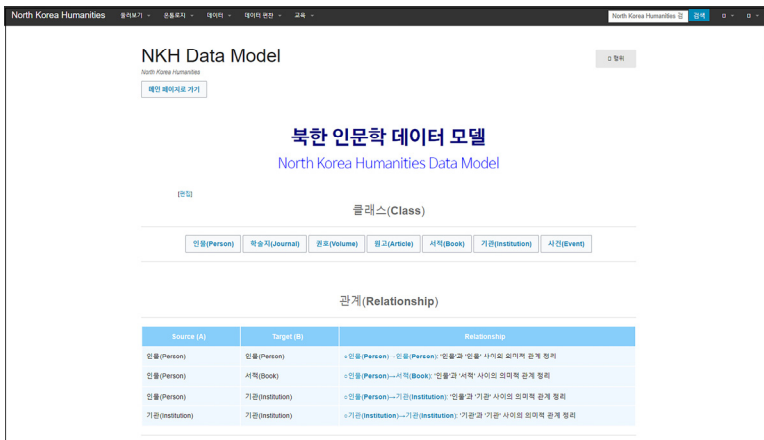
북한 인문학 자료는 기본적으로 학술지(Journal)와 해당 학술지의 구체적 권호(Volume) 그리고 특정 권호에 게재된 논문(Article)을 중심으로 구성된다. 논문은 아니지만 주요한 연구 서적(Book)과 논문 및 서적을 저술한 인물(Person) 그리고 유관 단체(Group) 및 기관(Institution) 또한 중요한 정보다. 더 나아가 연구자로 활동한 인물들 사이에 벌어진 사건(Event)과 인물의 활동지나 기관의 소재지로서의 공간(Place) 또한 학술적 차원에서 검토해야 하는 중요한 정보로서, 그러한 것들이 모두 온톨

7) 온톨로지는 특정한 분야에서 다루어지는 복잡한 정보 요소들의 세계를 명시적으로 기술하고 보여줄 수 있게끔 하는 일종의 체재를 의미한다. 관련해서 그루버(Thomas A. Gruber)의 'an explicit formal specification of a shared conceptualization'이라는 정의가 가장 보편적으로 언급되는데, 컴퓨터가 이해할 수 있는 기계가독(Machine-readable) 형식으로 데이터를 조직하는 것이 핵심이다. Triple(S-P-O) 형식의 데이터 구조는 기계가독형 데이터를 만들기 위한 일종의 뼈대라 할 수 있으며, 이에 따라 온톨로지는 대체로 클래스(Class), 인스턴스(Instance), 속성(Property, DataProperties), 관계(Relation, ObjectProperties) 등의 요소로 구성된다. 클래스는 대상 세계의 범주(classification)를 의미하며, 인스턴스는 범주를 구성하는 개별 요소로서의 실체(entity)를 가리키며, 속성은 클래스와 인스턴스의 특성(feature)을 나타내며, 관계는 클래스 및 인스턴스 사이의 의미적 관계를 정의한다.

로지 체계 내에서 각각의 클래스(Class)로 구성된다.

클래스에 소속된 개체들은 서로 간에 의미적 관계(Relationship, Object-Properties)를 구성하는데, 해당 관계 정의는 Triple(Subject-Predicate-Object) 형식 즉 개별 클래스에 속하는 주어(S)와 목적어(O)로서의 개체들 사이에서 서술어 역할을 한다. 예컨대 어떤 인물(Person)이 특정 논문(Article)의 저자(writer)라거나, 특정 논문(Article)이 어떠한 학술지 권호(Volume)에 일부분으로 수록되어 있다거나(isPartOf) 하는 정보들이 여기에 해당한다. 클래스(Class)와 관계(Relationship)에 대한 설정을 바탕으로 개별 클래스에 속하는 개체들에 공통으로 부여되는 속성(Attribute, DataProperties)을 정의하면, 기본적인 온톨로지 디자인의 일개가 갖추어진다. <그림 2>는 그와 같은 클래스와 관계 그리고 속성에 관한 구체적인 정보를 정리한 북한 인문학 데이터 모델을 소개하는 페이지다.

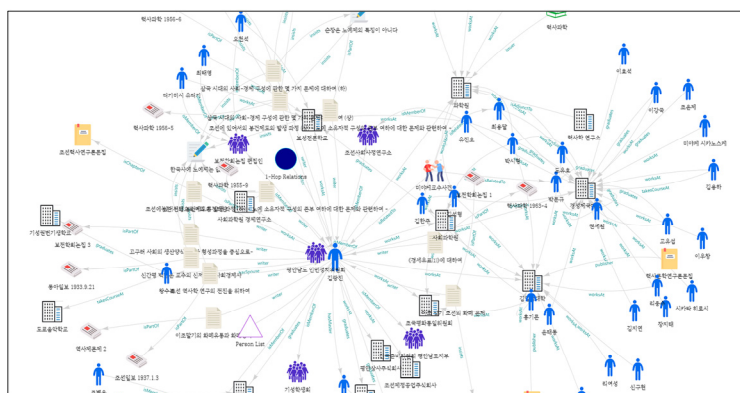
<그림 2> 북한 인문학 데이터 모델(NKH Data Model) 페이지⁸⁾



8) <북한 인문학 데이터 모델> 페이지, <<http://dh.aks.ac.kr/~nkh/wiki/index.php>>

데이터 모델(온톨로지)을 바탕으로 Triple(S-P-O) 형식으로 입력된 데이터셋을 활용해 데이터 네트워크를 구현할 수 있다. <그림 3>은 북한 인문학 데이터 네트워크 시각화 페이지(<http://dh.aks.ac.kr/~nkh/search/nkhSearch.html>)에서 특정 데이터를 검색했을 때 지식그래프 형태로 출력되는 결과물의 예시이다. 한국학중앙연구원 디지털인문학연구소에서 개발한 RDB(SQL) 연동 vis.js 기반 데이터 시각화 프로그램을 활용해 구현한 것으로, 복잡하게 얽혀 있는 인문 데이터 사이의 의미적 관계를 직관적으로 확인할 수 있다는 것이 특징이라 하겠다.

<그림 3> 북한 인문학 데이터 네트워크 검색을 통한 출력 예시



3. 미디어위키와 데이터 편찬

시맨틱 데이터 프로세싱에 기초한 데이터 아카이브의 특징은 데이터

NKH_Data_Model}

- 9) 본 연구의 데이터 네트워크를 구현하는 데 활용한 소프트웨어는 한국학중앙연구원 디지털인문학연구소에서 만든 <Semantic Database Visualization> (http://dh.aks.ac.kr/Edu/wiki/index.php/Semantic_Database_Visualization)이다.

사이의 의미적 관계를 다층적으로 정의할 수 있다는 것인데, 그러한 작업을 진행하기 위해서는 여러 사람이 참여해 유관 분야의 데이터를 다각도로 기술(description)하고 지속적으로 편찬해 나가기 위한 웹 기반의 데이터 구축 플랫폼이 필요하다. 위키(Wiki)는 그러한 과업을 수행하는 데 있어서 최적화된 소프트웨어 가운데 하나라 할 수 있다.

위키는 간단한 마크업 언어로 정보를 편찬할 수 있게끔 한 소프트웨어 엔진이자, 편찬된 정보가 체계적으로 정리·누적될 있게끔 한 데이터베이스이자, 그 모든 것들이 웹 환경에서 다수의 인원에 의해 구동될 수 있게끔 한 웹사이트이다.¹⁰⁾ 대중적으로 널리 알려진 위키피디아(Wikipedia) 사례에서도 확인할 수 있듯이, 데이터를 편찬해 나가는 협업 기반의 연구를 디지털 환경에서 전개할 경우, 가장 효과적으로 활용할 수 있는 지식플랫폼 가운데 하나가 위키이다. 참여자 각자가 정리한 정보와 지식을 온라인상의 공동 공간에 수집·관리·체계화하는 데 있어서 유용한 기능 및 도구를 제공하기 때문이다.

〈그림 4〉 북한 인문학 위키 메인 페이지¹¹⁾



10) 김지선·장문석·유인태, 「공유와 협업의 글쓰기 플랫폼 위키」 『한국학연구』 제60집 (2021), 376쪽.

11) 〈북한 인문학 위키〉 〈http://dh.aks.ac.kr/~nkh/wiki/index.php〉

위키는 다양한 프로그래밍 언어로 개발된 여러 위키 소프트웨어가 상용화·운용되고 있는데, 본 연구는 그 가운데서도 전 세계에서 가장 보편적으로 사용되는 미디어위키(MediaWiki)를 활용했다. 미디어위키는 위키 마크업(Wiki Markup)이라 불리는 언어를 사용하며 위키에서 생성된 정보를 관계형 데이터베이스(RDB)에 저장해 관리한다.¹²⁾ XML(eXtensible Markup Language) 형식으로 개별 문서 데이터를 출력(내보내기 기능)해주는 기능이 있어서, 통일된 형식의 스키마를 지정하고, 그것을 준용한 형식의 데이터 편찬과 관리에 있어서 매우 유용하다.

Ⅲ. 빅데이터 혁신융합대학 역사정보 마이크로 디그리

1. 혁신융합대학 개관

‘디지털 신기술 인재양성 혁신융합대학 사업’은 교육부 주관 하에 수도권 대학과 지방대 및 전문대가 컨소시엄을 구성해서 신산업 분야의 인재를 양성하는 사업이다. 혁신융합대학 사업은 2021년 상반기에 시작하여 2026년까지 6년간, 8개 분야 10만 명의 인재를 양성할 것을 목표로 삼으며, 이를 위해 참여 대학이 교육 및 학위과정을 공동으로 개발·운영하는 데 중점을 두었다.¹³⁾

서울대학교는 빅데이터와 차세대 반도체 분야에서 컨소시엄 주관대학으로 선정되었다. ‘빅데이터 혁신융합대학 사업’(이하 ‘사업’)에는 서울대

12) 김지선·장문석·유인태, 「공유와 협업의 글쓰기 플랫폼 위키」, 376쪽.

13) 혁신융합대학에 관한 개괄적 내용은, “빅데이터 혁신융합대학” 웹사이트의 〈사업단 소개〉 페이지(https://bigdatahub.ac.kr/?page_id=217)를 참고하였다.

를 비롯하여 경기과학기술대학교, 경상국립대학교, 서울시립대학교, 숙명여자대학교, 전북대학교, 한동대학교의 7개교가 참여하고 있다. 빅데이터 혁신융합대학 사업단(이하 ‘사업단’, 단장 김흥기) 본부는 사업 신청 단계에서부터, 빅데이터 인재양성에서 인문학의 성찰성과 인문 콘텐츠의 중요성에 주목하여 인문대학에 사업 참여를 제안하였다.

서울대학교 인문대학은 그동안 대학인문역량강화 사업, 즉 ‘코어’ 사업에서 인문데이터과학연계전공을 운영한 경험이 있다. 또한 인문대학 산하에 ‘디지털 위원회’를 창설해서 디지털 기술을 매개한 연구 및 교육에 관한 제반 사안을 심의해 왔다. 이런 바탕 위에서 사업단의 요청이 있자, 인문대학 학장단은 인문학과 디지털 신기술의 접목이 인문학 연구 및 교육의 바람직한 방향이라고 판단하여 참가를 결정하였다. 이에 따라 인문대학은 사업 신청 단계에 이미, 다수 교원의 참여 속에 31개의 관련 교과목 개발안을 수합하여 신청서에 반영한 바 있다.¹⁴⁾

본격적인 사업 출범 후 사업단은 ‘빅데이터 분야의 다양성을 함양한 핵심 인재를 양성하는 혁신적 분산·융합대학 모델 구축’을 비전으로 삼았다. 또한 구체적 목표로 ‘빅데이터의 본질적 다양성’을 이해하고, 그 ‘다양한 분석방법론을 통합적으로 이해’하며, ‘전공·지역·젠더·전문성 면에서 다양한’ 인재를 양성하는 것 등을 표방하였다.

2. ‘빅데이터역사정보’ 교과

사업단은 새로운 교육과정 및 교과목 개발을 독려하는 차원에서, 2021년 9월, ‘연계전공 및 교과과정 개발을 위한 기획연구’(이하 ‘기획연구’) 과제와 ‘교과목 모듈화 기획 및 개발’(이하 ‘교과목 개발’) 과제를 발주하

¹⁴⁾ 허수, 「교수논단 디지털 신기술 혁신공유대학 사업」 『서울대 인문대학 소식지』 제53권 (2021년 가을), 7쪽.

였다. 인문대학에서도 사업에 관심을 가진 교수들이 이 과제에 참여하였다. 후자의 교과목 개발은 개인 차원에서 참여하였으며, 전자의 기획연구에는 기획팀을 구성하여 팀 차원에서 참여하였다. 특히 인문대학은 기획연구와 관련해 ‘언어인지’·‘역사정보’·‘디지털 스토리텔링’·‘빅데이터 문화연구’(이상 가칭)의 4개 교과과정을 개발하는 목표를 세웠다. 각 기획팀은 개별 과정에 필요한 내용을 논의하여 2022년 1월, 개별 과정에 관한 설계안을 담은 보고서를 사업단에 제출하였다.¹⁵⁾

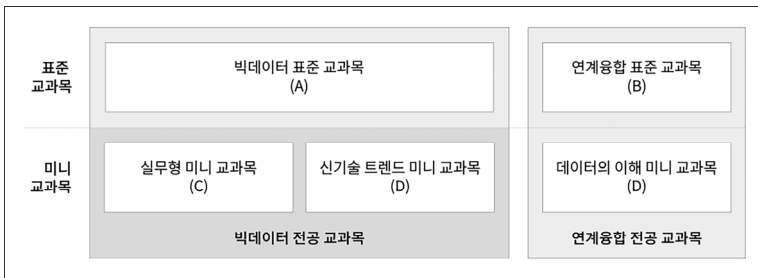
인문대학이 기획한 4개의 교과과정 가운데 승인이 가장 먼저 이루어진 것은 ‘역사정보’ 교과였다. ‘역사정보’는 2023년 1월, ‘빅데이터역사정보’(Big Data in Historical Informatics)라는 이름으로 ‘서울대 교과인증과정’이 되었다. 당시 승인된 교과인증과정은 학부 25개, 대학원 4개였다. 서울대 교무처가 작성한 “「서울대학교 교과인증과정」 운영 가이드라인”에 따르면, 교과인증과정은 “급변하는 기술 및 교육환경에 대응하는 미래 인재 양성을 위해 전문화된 지식을 유연하게 학습하고 인증 받을 수 있는 최소 단위의 교과과정”이다.¹⁶⁾ 참여 학생이 9학점 이상 15학점 이내의 교과과정을 이수하면, 이수내역을 졸업증명서(학위수여증명서) 및 성적증명서 등을 통해 인증 받게 된다.¹⁷⁾

‘빅데이터역사정보’(이하 ‘역사정보’)는 국사학과가 주관기관이며, 한국사학 전공, 동양사학 전공, 서양사학 전공 등에서 5명의 전임교원이 운영에 참여하고 있다. 교육 목표로는 “디지털 환경에 걸맞은 새로운 역사교육 모델 창안”과 “역사정보 빅데이터를 활용한 ‘역사적 사고력 함양’ 모

15) ‘역사정보’는 ‘언어인지’와 함께 기획팀을 구성하여 『언어인지 및 역사정보 연계 융합 마이크로 디그리 과정 개발 결과보고서 (2022년 1월 14일)』를 제출하였다.
 16) 서울대학교 교무처 교무과, “「서울대학교 교과인증과정」 운영 가이드라인”(2023.1.30.)
 17) 서울대학교 학칙 제77조의2(교과인증과정 이수) <<https://www.law.go.kr/LSW//schlPubRulInfoP.do?schlPubRulSeq=2200000101435>> 등을 참조.

색” 등을 앞세웠다. 소속 학생들은 12학점 이상을 이수해야 하는데, 이를 통해 역사정보 빅데이터에 대한 접근방안과 분석력을 갖추고, 이를 졸업 후 자신의 연구와 직무를 수행하는 데 활용할 수 있도록 하였다.

〈그림 5〉 빅데이터 혁신융합대학 교과목 구성도



‘역사정보’ 과정의 교과목은 총 23학점이며, 3학점 과목 4개, 1학점 과목 11개로 구성된다. 이 중에서 2023년 1학기에 개설된 과목은 7개, 15학점이다. ‘빅데이터 개론 I’, ‘프로그래밍 기초’, ‘기초통계학’(이상 3학점)은 〈그림 5〉의 ‘(A) 빅데이터 표준 교과목’에 해당하며,¹⁸⁾ ‘언어데이터과학’(3학점)은 인문대학 언어학과 전공 과목으로 ‘(B) 연계융합 표준 교과목’에 해당한다. 주로 1학점으로 된 미니교과목 ‘빅데이터로 읽는 한국근대사 1’, ‘미디어위키를 활용한 역사 데이터 편찬’, ‘디지털 역사학과 동양사의 이해’ 등은 ‘(D) 연계융합 전공 교과목’에 해당한다. 이외에 ‘역사GIS입문’, ‘한국사 디지털 원문자료의 이해’, ‘서양사 디지털사료의 이해’ 등 8개의 미니교과목들도 개발되고 있거나 조만간 개발될 예정이다.

한편, 본 과정에서는 빅데이터의 형태와 취급 방식에 따라 ‘역사정보

18) 교과목 구성도는 “빅데이터 혁신융합대학” 웹사이트의 〈표준교과목〉 페이지 (https://bigdatahub.ac.kr/?page_id=483) 를 참고하였다.

텍스트마이닝’, ‘역사지리정보’, ‘시맨틱데이터 편찬’ 세 가지 갈래로 세부 분야를 구분하고 있다. ‘역사정보 텍스트마이닝’ 분야는 텍스트 중심 데이터의 이해와 취급을, ‘역사지리정보’ 분야는 지리정보 중심 데이터의 이해와 취급을 지향하며, ‘시맨틱데이터 편찬’ 분야는 데이터의 연계와 자기표현 및 공유 등 데이터의 조직화와 이를 통한 소통을 지향한다.¹⁹⁾

교과목 개발이 상당부분 진행될 경우, 당초의 구상대로 세 개 분야에 대응한 세 개의 이수체계를 완비해서 수강생들에게 제시하고자 한다. 이렇게 되면 수강생들은 자신에게 맞는 이수 과정을 선택할 수 있다. 예컨대 자신의 관심과 적성에 따라 특정 분야에 집중된 이수과정을 선택할 수 있다. 혹은 세 분야에 두루 걸쳐 초급 단계의 과목을 수강함으로써 빅데이터역사정보에 관한 개론적 지식과 역량을 갖추 수도 있다.

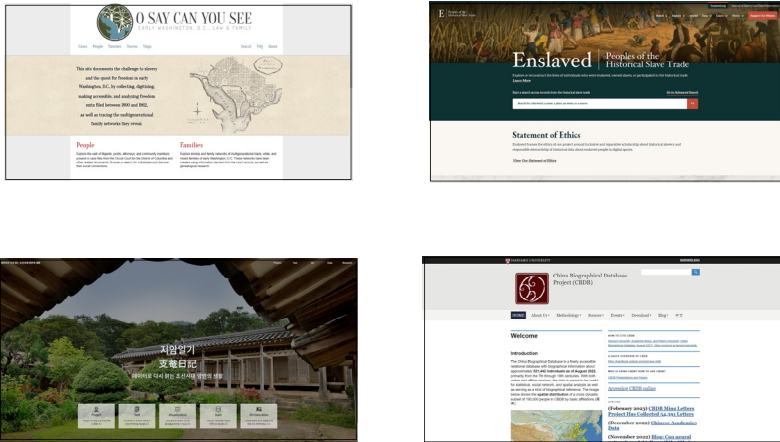
3. 시맨틱 데이터 프로세싱과 역사 데이터 편찬 교육

‘빅데이터역사정보’ 과정의 세 분야인 ‘역사정보 텍스트마이닝’, ‘역사지리정보’, ‘시맨틱데이터 편찬’ 가운데서도 ‘시맨틱데이터 편찬’ 유관 교과는 디지털 인문학의 다양한 지류 가운데서도 시맨틱 데이터 프로세싱을 접목한 인문학 연구 방법론과 관련이 있다. 앞서 II 장에서 시맨틱 데이터 프로세싱에 관한 내용을 소개하였지만, 인문학 연구가 전통적으로 유지해 온 질적(qualitative) 연구의 시각을 데이터 처리의 차원에서 연장하려고 하는 목적이 시맨틱 데이터 프로세싱 방법론에 내포되어 있다.

19) 세 분야에 관한 내용은 『언어인지 및 역사정보 연계융합 마이크로 디그리 과정 개발 결과보고서 (2022년 1월 14일)』의 ‘III. 역사정보 마이크로 디그리 과정’ 부분을 참조하였다.

20) 좌측 상단부터 시계 방향으로, <O Say Can You See> (<<https://earlywashingtondc.org/>>), <Enslaved> (<<https://enslaved.org/>>), <CBDB> (<<https://projects.iq.harvard.edu/cbdb>>), <데이터로 읽는 17세기 재지사족의 생활> (<<http://jiamdiary.info/>>).

〈그림 6〉 시맨틱 데이터 프로세싱을 접목한
국내의 디지털 역사학 연구 사례²⁰⁾



이와 관련해 역사학 연구에서 보편적으로 통용되는 질적 관점의 연장선상에서 역사 데이터를 편찬하는 경험은, 데이터 기반의 역사학 연구를 전개하는 데 있어서 가장 기초적인 소양에 해당한다. 데이터를 잘 다룬다는 것은 다른 누군가가 구축해놓은 데이터를 가져와서 잘 사용하는 것을 가리키기도 하지만, 기본적으로는 나 자신이 직접 데이터를 편찬하고 잘 가공하는 것을 의미하기 때문이다. 시맨틱 데이터 프로세싱은, 아날로그 환경에서 이루어져 온 인문학 연구의 질적 차원을 디지털 환경의 데이터 처리 과정으로 연장한다는 측면에서 여타 디지털 인문학 방법론과는 다른 특수성을 지니며, 그렇기 때문에 학부나 대학원 과정에서 디지털 인문학 유관 교과를 개설·운영할 경우 빠뜨려서는 안 되는 기초 교육 프로그램에 해당한다.

〈그림 6〉에 제시한 여러 디지털 역사학 연구 프로젝트는 그러한 문제의식에 입각해 시맨틱 데이터 프로세싱을 시도한 대표적 사례라 할 수

있으며, 현 시점에서도 유관 방법론을 접목한 연구가 활발히 이루어지고 있다.

IV. <미디어위키를 활용한 역사 데이터 편찬> 교과 소개

1. 교과 개요

역사 데이터 편찬 교육은, 디지털 인문학 교육의 일환으로서 디지털 환경에서 역사 지식을 체계적으로 구성하고 자유롭게 유통하는 실재로서 데이터를 능동적으로 다룰 수 있는 역량을 기르는 데 초점을 두어야 한다. 디지털 환경에서 데이터를 수집·정리하는 과정과 축적된 데이터를 효과적으로 다루는 기술을 익히게 되면, 계량적 방법론에 입각해 대규모 역사 데이터를 분석하거나 창의적 관점에 입각한 역사 데이터 시각화 및 큐레이션 연구를 수행할 수 있다. 아날로그 환경에서는 좀처럼 시도하기 어려운 방식의 학술적 접근을 디지털 환경에서 새롭게 시도할 수 있다는 장점이 있다.

역사 데이터 편찬 과정의 구체적 도구로서 위키(wiki)를 활용하고자 하는 것은, 위키가 텍스트를 포함해 여러 미디어 형식의 자료를 대상으로 하이퍼링크 기능을 지원함으로써 멀티미디어 환경에서 정보와 지식을 효과적으로 조직·표현할 수 있는 손쉬운 환경을 제공하기 때문이다. 위키의 마크업(markup) 언어는 하이퍼링크의 무궁한 연결성과 다채로운 확장 기능(extension) 지원을 바탕으로, 정보를 정리하고 지식을 서술하는 데 있어서 여러 형식의 미디어를 복합적으로 활용하고 그것들이 입체

적으로 연결될 수 있는 가능성을 제시한다. 그리고 다수의 참여자가 각기 작성한 정보가 실제 링크로 연결되는 결과는, 각자의 몫이지만 동시에 함께 하는 과업으로서 역사 데이터 편찬 과정을 경험하게끔 한다. 역사를 비판적으로 이해하고 유관 정보를 수집·정리하고, 유관 내용을 위키 환경에서 기록하는 전반적 과정은 ‘데이터’를 매개로 유기적으로 연결된다.

마크업 기반의 위키 문법을 익히고, 위키를 활용한 역사학 연구 사례를 살펴보는 과정을 통해 참여자는, 위키 환경에서 역사 데이터를 다각도로 다루기 위한 구체적 방안을 고민할 수 있다. 그리고 그러한 고민이 이루어짐과 함께 실제 역사문화 자원을 대상으로 위키 기반의 데이터 편찬 실습을 진행할 경우, 디지털 환경에서 역사 데이터를 편찬하기 위한 기본적 이해를 습득할 수 있다. 단순히 미디어위키 문법과 확장기능 사용법을 터득하는 것이 목표가 아니라, 위키를 통해 역사학 데이터를 직접 편찬하는 경험을 하고, 그로부터 데이터 기반 역사학 연구의 기초를 이해하는 것이야말로 역사 데이터 편찬 교육의 핵심적인 목적이라 할 것이다.

〈미디어위키를 활용한 역사 데이터 편찬(Creating History Data Based on MediaWiki)〉 강의는 빅데이터 혁신융합대학의 마이크로디그리(micro degree) 교과과정의 일환으로서 미니교과목(1학점, 5주강의)으로 개설되었다. 강의의 절반은 사전 녹화가 이루어졌으며, 플립 러닝(Flipped Learning) 형식으로 사전 녹화한 영상을 학생들이 시청하고 현장강의에 참여하면, 현장에서는 토론과 함께 데이터 처리를 실습하는 방식으로 강의를 진행하였다.

2. 5주간 교육 과정과 내용

15주 이상 이루어지는 3학점 단위의 일반적인 학부 강의와 달리 본 교과의 경우 5주간 진행되는 1학점 단위 마이크로디그리 강의이기 때문에, 교육 과정을 설계함에 있어서 강의 내용을 기능적으로 압축해서 커리큘럼을 디자인하는 것이 필요하다. 학생들에게 전달해야 할 15주 강의 내용 가운데 1/3 분량을 잘라낸다고보다도, 완결된 형식의 15주 강의 내용을 미리 디자인하고 그것을 1/3로 압축하는 형태에 더 가깝다.

그리고 본 교과의 경우 ‘미디어위키(MediaWiki)’라는 소프트웨어 활용 실습이 중심이 되지만, 그에 앞서 인문학 자원을 디지털 환경에서 데이터 화하고 처리하는 맥락이 전제되기 때문에 ‘디지털 인문학(digital humanities)’에 관한 기초적 이해가 우선적으로 전달되어야 한다. 미디어위키를 다루기 위한 기초 지식과 경험을 습득하고, 그로부터 실제 미디어위키를 활용해 역사 데이터를 편찬하는 실습을 진행하는 데만 해도 적지 않은 시간이 투입되는데, 디지털 인문학에 관한 개론적 내용을 함께 전달해야 하기 때문에 강의 구성에 있어서 더욱 꼼꼼한 접근이 필요하다.

그러한 고민을 바탕으로 1주차는 디지털 인문학에 관한 개론적 내용에 관한 전달을, 2~3주차는 미디어위키에 대한 기초적 이해와 문법에 관한 전달을, 4~5주차는 미디어위키를 활용해 역사 데이터를 편찬하는 실습을 각기 진행하는 것으로 총 5주 강의를 디자인하였다. 미디어위키라는 소프트웨어를 다루되 역사 데이터 편찬의 맥락에서 고민하고, 그 과정에서 이루어지는 여러 문제의식을 디지털 인문학의 관점에서 바라보도록 하는 것이 초점이라 하겠다. 하단의 표는 그에 관한 내용을 정리한 것이다.

〈표 1〉 〈미디어위키를 활용한 역사 데이터 편찬〉 주차별 교과 내용

주차	주제	소주제	내용	강의형식
1주차	디지털 인문학과 역사 데이터	디지털 인문학의 기초 (1H)	○디지털 인문학 개괄 -디지털 인문학의 형성과 전개 -디지털 인문학의 특징과 의미	동영상 (30M) 이론
			○디지털 환경의 진화와 인문학 연구 -디지털 환경과 인문학 연구의 변화 -디지털 환경과 인문학의 연결 양상	현장강의 (30M) 이론
		해외의 디지털 역사학 연구 사례 (1H)	○해외의 디지털 역사학 연구 사례 -아카이브와 데이터베이스 구축 사례 -데이터 시각화/큐레이션 구현 사례	동영상 (30M) 이론
			○해외 디지털 역사학 사례 분석 -아카이브와 데이터베이스 사례 -데이터 시각화와 큐레이션 사례	현장강의 (30M) 이론
		데이터 기반 역사학 연구 (1H)	○국내의 디지털 역사학 연구 사례 -삼일운동 데이터베이스 구축 사례 -지암일기 데이터아카이브 구현 사례	동영상 (30M) 이론
			○국내 역사학 데이터 살펴보기 -삼일운동 데이터의 구조와 특징 -지암일기 데이터의 구조와 특징	현장강의 (30M) 이론
2주차	위키와 역사데이터 편찬 연구	위키와 글쓰기 그리고 하이퍼 텍스트 (1H)	○위키(Wiki) 개괄 -위키의 탄생과 역사 -위키의 기능과 활용 사례	동영상 (30M) 이론
			○위키, 전자글쓰기, 데이터 -하이퍼텍스트와 위키 문서 -위키의 '내보내기'와 XML 데이터	현장강의 (30M) 이론
		위키와 데이터 그리고 집단지성 (1H)	○위키데이터와 집단지성 -위키데이터의 현황과 특징 -위키데이터의 구성과 활용	동영상 (30M) 이론
			○위키데이터 다루기 -위키데이터 쿼리(SPARQL) 기초 -위키데이터 쿼리(SPARQL) 응용	현장강의 (30M) 이론
		위키 기반 역사학 연구사례 (1H)	○위키 기반 디지털 역사학 연구 사례 -역사문헌 기초 아카이브 구축 사례 -역사자원 데이터베이스 구축 사례	동영상 (30M) 이론
			○위키 기반 역사 데이터 구축과 가공 -조선왕조실록 전문사전 위키와 XML데이터 가공	현장강의 (30M) 이론
3주차	미디어	미디어	○미디어위키 문법의 기초1	동영상

주차	주제	소주제	내용	강의형식
	위키의 문법과 활용	위키 문법의 기초 이해1 (1H)	-새 문서 생성 및 문단과 문장 구성 -글자 입력 및 하이퍼링크 활용	(30M) 이론
			○나의 위키 페이지 작성 실습1 -새 문서 생성 및 문단과 문장 구성 -글자 입력 및 하이퍼링크 활용	현장강의 (30M) 이론
		미디어 위키 문법의 기초 이해2 (1H)	○미디어위키 문법의 기초2 -표 만들기, 주석 달기와 출처 표기 -문서 분류와 위키 문법 적용 취소	동영상 (30M) 이론
			○ 나의 위키 페이지 작성 실습2 -표 만들기, 주석 달기와 출처 표기 -문서 분류와 위키 문법 적용 취소	현장강의 (30M) 이론
미디어 위키와 멀티미디어 리소스 (1H)	○멀티미디어 자료의 삽입과 활용 -이미지, 동영상, 문서 삽입 -외부 웹 리소스 임베딩	동영상 (30M) 이론		
	○나의 위키 페이지 작성 실습3 -이미지, 동영상, 문서 삽입 -외부 웹 리소스 임베딩	현장강의 (30M) 이론		
4주차	역사위키 문서 편찬실습 (응용)	북한인문학 위키와 역사자료 수집 (1H)	○북한인문학 위키와 역사 정보 -북한인문학 데이터 편찬 프로젝트 -북한인문학 데이터 위키와 역사정보	동영상 (30M) 이론
			○근대 역사자료 수집과 정리 -사료의 개념 및 이해 -역사자료 찾기와 인용	현장강의 (30M) 이론
		역사위키 문서편찬 실습 (1H)	○역사정보 위키 문서 작성 맥락 -인물 정보 페이지 작성 규칙/맥락 -기타 정보 페이지 작성 규칙/맥락	동영상 (30M) 이론
			○역사정보 위키 문서 작성 실습 -인물 정보 페이지 작성 실습 -기타 정보 페이지 작성 실습	현장강의 (30M) 이론
역사위키 문서 편찬과 Triple Data (1H)	○위키 문서 작성과 데이터 기술 -위키 문서 작성과 데이터 모데 -Markup과 Triple 데이터 기술	동영상 (30M) 이론		
	○위키에서 Triple 데이터 구축 실습 -위키 문서 서술정보 마크업 실습 -위키 문서 Triple 데이터 구축 실습	현장강의 (30M) 이론		
5주차	역사위키 문서 편찬실습	역사정보 활용 타임라인	○Timeline 구현 및 임베딩 -Timeline.js 활용 spread sheet 작성 -web 게시와 Timeline 임베딩	동영상 (30M) 이론

주차	주제	소주제	내용	강의형식
	(심화)	구현실습 (1H)	○Timeline 구현 및 임베딩(실습) -Timeline.js 활용 spread sheet 작성 -web 게시와 Timeline 임베딩	현장강의 (30M) 이론
5주차	역사위키 문서 편찬실습 (심화)	역사정보 활용 네트워크 그래프 구현실습 (1H)	○Network Graph 시각화 -Network Graph의 개념과 이해 -Network Graph 시각화 구현	동영상 (30M) 이론
		위키와 데이터 그리고 역사학 (1H)	○실습 결과물에 대한 평가와 토론 ○데이터 편찬 작업과 역사학 연구의 관계 에 관한 논의 진행	현장강의 (1H) 이론

3. 평가 기준과 그 의미

디지털 인문학 교과는 여타 인문학 교과와 비교할 시 실습의 비중이 크기 때문에, 이론과 실습의 균형을 고려하되 실습을 통해 이론에 관한 숙지를 확인하는 방식으로 평가 방안을 설정하는 경우가 일반적이다. 강의론을 통해 전달받은 특정한 지식을 얼마나 잘 이해하고 있는지를 시험하는 데 목적을 두기보다는, 강의에서 배운 기초적인 내용을 바탕으로 자신만의 지식을 얼마나 창의적으로 기술하고 있는지를 확인하는 것이 중요하다.

그리고 본 강의의 경우 5주라는 짧은 시간 내에 진행되기 때문에 그러한 특성을 고려해 별도로 시험(test)을 안배하지 않고, 3가지 맥락 총 4개의 과제를 부여하는 것으로 평가 내용을 마련하였다. 아래는 개별 과제 항목과 각각의 과제에 대한 구체적 내용이다.

○ 과제1: Wiki Document 작성

-4주차 2번째 소주제 ‘역사위키 문서 편찬 실습’ 강의를 통해 개인별로 부여한 위키 문서를 작성하도록 한다. 역사 인물, 사건, 공간, 문헌 등에 관한 정보를 수집·정리해서 약속된 형식에 따라 위키 문서를 작성한다.

○ 과제2: Triple(S-P-O) Data 기술

-4주차 3번째 소주제 ‘역사 위키문서 편찬과 Triple Data’ 강의를 통해 개인별로 부여한 위키 표제어의 유관 Triple Data를 기술하는 과제를 수행한다. 표제어와 관련된 역사 인물, 사건, 공간, 문헌 등에 관한 정보를 수집·정리해서 약속된 형식에 따라 데이터를 구축한다.

○ 과제3: Timeline 시각화 구현

-5주차 1번째 소주제 ‘역사 정보 활용 타임라인 구현 실습’ 강의를 통해 개인별로 작성한 위키 문서와 Triple Data를 바탕으로 Timeline 시각화 구현 과제를 수행한다. 시간 정보를 기준으로 표제어와 관련된 다양한 맥락의 역사적 사실과 유관 정보를 정리해서 Timeline 형식으로 시각화한다.

○ 과제4: Network Graph 시각화 구현

-5주차 2번째 소주제 ‘역사 정보 활용 네트워크 그래프 구현 실습’ 강의를 통해 개인별로 작성한 위키 문서와 Triple Data를 바탕으로 Network Graph 시각화 구현 과제를 수행한다. 표제어와 관련된 다양한 맥락의 유관 정보를 정리해서 Network Graph 형식으로 시각화한다.

과제1(위키 문서 작성)은 역사학 연구를 하는 데 있어서 raw data를 수집·정리하는 과업과 관련이 있다. 사료 또는 2차 자료를 읽고, 그로부터

찾고자 하는 정보와 관련된 기초적 내용을 발췌해서 기술하는 것이 여기에 해당한다. 가장 기초적인 훈련이라 하겠다. 과제2(트리플 데이터 기술)는 질적 관점에 입각해 raw data를 가공하는 과업이다. 기초적인 데이터베이스를 구축하는 과정이 여기에 해당하는데, 선형적(linear) 텍스트로부터 비선형적(non-linear) 데이터를 추출하는 훈련이라 하겠다. 과제3, 4(시각화 구현)는 편찬한 데이터를 적극적으로 활용해 결과물을 표현하는 작업으로서 그 의미를 갖는다. 통상적으로 시각화(visualization)는 창의적 표현을 내포하는 경우가 많은데, 여기서는 데이터 스키마에 기초한 데이터 표현 역량을 길러주는 데 초점이 있다.

4. 수업 진행 내용과 결과

〈미디어위키를 활용한 역사 데이터 편찬〉 강의는 2023년 1학기 3월 한 달 간 진행되었다. 수강생들은 앞서 소개한 교육 과정에 참여해 4가지 과제를 수행하는 데 집중하였다.

수강생들은 대체로 북한 학자가 쓴 글을 이전에 접한 경험이 없었으며, 유튜브나 위키피디아에서 제공하는 일부 정보를 제외하고는 북한에 관한 자료를 접해 본 경험 또한 없었다. 오프라인 환경에서 북한에 관한 자료를 찾는 것이 쉽지가 않기에, 과제를 진행하면서 1차적으로는 웹에서 자료를 수집하였고, 웹에서 입수한 정보를 바탕으로 서울대학교 내 특수자료실에서 확인할 수 있는 자료를 통해 사실관계를 검토하는 양상을 보였다.

북한 자료를 전문적으로 연구하는 연구자가 아니기에, 학부생으로서 수강생이 자료 수집 단계에서부터 겪었을 어려움을 짐작할 수 있다. 한편으로 수강생이 북한에 관한 전문적 이해를 얼마나 갖추고 있는지의 여부와 별개로 온오프라인에서 찾은 정보를 수집·검토·정리하게끔 하는

것만으로도 유효한 교육적 효과를 거둘 수 있다. 아래는 위키 페이지 작성을 위한 자료 수집 과정에 관해 한 수강생이 남긴 소감이다.

“여러 북한의 웹사이트가 한국에서 막혀있어서 자료 수집에 어려움이 있었습니다. 제가 조사를 맡은 ‘도상록’은 그나마 한국에 관련 논문이 있고, 위키피디아와 나무위키에 관련 문서가 있었지만, 중국의 바이두와 같은 포털 사이트에서는 ‘도상록’에 관한 문서를 거의 찾아볼 수 없었으며, 북한 사이트는 접근하기 어렵고, 일본 웹사이트의 경우 제가 일본어를 몰라서 찾아볼 수 없었기에 자료조사를 깊게 하진 못하였습니다. 뿐만 아니라, 3월 20일자 <조선의 오늘>에서 ‘도상록’과 관련된 자료를 게재하기 전에, 나무위키나 여러 기사에 사용된 ‘도상록’ 사진의 첫 출처가 어디인지 확인하기가 어려워 사진 첨부를 하지 못하고 있었습니다. <조선대백과사전>의 ‘도상록’ 사진을 올려야 하나 고민까지 하였습니다. 이처럼 오래된 자료를 찾아보기 힘들다는 것이 큰 어려움 중의 하나였습니다. 그 외에, 학교 특수자료실은 제가 직접 들어가서 여러 자료를 확인할 수 있는 것이 아니라, 자료를 따로 요청해서 요청한 자료만 확인할 수 있게 되어 있었습니다. 그러다 보니 웹의 참고자료 부분에서 확인한 자료 이외에 또 다른 ‘도상록’에 관한 자료가 있는지 확인할 방법이 없었습니다. ‘도상록’이 북한에서 한 연구가 여러 가지가 있는데, 논문을 직접 읽어보고 내용을 정리하면 좋았을 텐데, 제 물리학 지식이 이러한 논문을 쉽게 읽어낼 정도는 아니었고, 논문 원문도 찾기가 어려워서 연구 관련 부분은 자세히 작성하지 못한 것이 마음에 걸립니다.”

수강 학생의 입장에서 여러 아쉬움을 위주로 소감을 풀어냈지만, 그러한 과정을 통해 수집한 자료를 어떻게 검토할 것인지 고민하고, 또 고민을 거쳐 나온 정보와 지식을 위키에 어떻게 정리할 것인지를 생각하는 과정 자체가 역사 정보를 다루는 훈련이라 할 수 있다. <그림 7>은 그러한 사유의 결과물로서 수강 학생이 작성한 위키 페이지 예시이다.

21) <북한 인문학 위키>의 ‘도상록’ 페이지 <<http://dh.aks.ac.kr/~nkh/wiki/index.php/도상록>>

〈그림 7〉 수강 학생이 작성한 위키 페이지 예시²¹⁾



과제물로서 위키 문서를 작성하는 경험은 여타 인문학 강의에서 이루어지는 보고서(레포트) 작성 및 제출과는 사뭇 다른 교육적 효과를 불러 일으킨다. 보고서는 일반적으로 작성해서 제출하고 나면 그것으로 끝이지만, 여러 정보를 기술한 위키 페이지는 평가가 끝나고 강의가 종료된 이후에도 웹상에 그 내용이 계속 남는다. 나의 지식 활동이 ‘완결된’ 것으로서 평가를 끝으로 오프라인 상에서 사라지는 것이 아니라, ‘미완결’된 것으로서 온라인에 계속 유지·보존된다고 생각하면, 과제로 주어진 정보와 지식을 기술하는 동인과 맥락이 다를 수밖에 없다. 아래는 그에 관한 수강 학생의 소감이다.

“보고서(레포트)는 교수님과 수강생만 읽고 묻혀버리기 마련입니다. 그래서 보고서를 작성할 때와 위키 문서를 작성할 때의 마음가짐이 다른 것 같습니다. 보고서는 과제로써 대충하고 끝내자는 느낌이 큰데, 위키 문서는 앞으로 여러 연구자분과 수강생분들, 또는 대중들이 볼 것이라는 생각을 하니 책임감을 더 느끼고 작성을 하게 되었습니다. 또한, 위키 페이지 작성은 단순히 글을 쓰는 것이 아닌, 표, 그림, 타임라인 등 여러 가지 내용을 같이 작성

해볼 수 있으므로, 작성할 때 훨씬 재미있었습니다.”

위키 문서 작성은 수강 학생들로 하여금 온라인상에서의 지식 활동이 ‘종료’없이 늘 ‘시작’과 ‘과정’으로서의 의미를 갖는다는 것을 자연스럽게 익히도록 한다. 그리고 그것은 온라인상에서 내가 수집·검토·정리·서술하는 정보와 지식이 ‘개인’인 나의 지식 활동으로서만 중요한 것이 아니라, 늘 누군가의 관심과 이해 가운데서 비판적으로 검증·확인될 수 있음을 깨닫도록 한다. 아날로그 환경과는 다른 개방적이고 확장적인 지식 유통 체제를 갖춘 디지털 환경의 특수성을 생각해볼 수 있다.

위키 문서 작성을 바탕으로 수강생들은 <그림 8>, <그림 9>과 같이 Triple Data를 기술(description)하고, 소규모지만 Timeline 형식의 데이터 시각화 결과물을 구현하는 작업도 과제로 수행하였다. 짧은 기간이라 수강생들이 그에 관한 전문 지식을 얻기에는 한계가 있지만, 한편으로 선형적 텍스트로부터 비선형적 데이터를 추출하고, 유관 맥락에서 데이터를 시각적으로 표현해보는 경험을 해보는 것 자체만으로도 역사 정보를 새롭게 이해할 수 있는 채널이 된다.

<그림 8> 수강 학생이 과제로 기술한 Triple Data 예시22)

TripleData [편집]		
Source (A)	Target (B)	Relationship
Person: 도상록	Institution: 영성고등보통학교	A graduates B
Person: 도상록	Institution: 계교중학교	A graduates B
Person: 도상록	Institution: 도쿄제국대학	A graduates B
Person: 도상록	Institution: 송도고등보통학교	A worksAt B
Person: 도상록	Institution: 신경공업대학	A worksAt B
Person: 도상록	Institution: 경성대학	A worksAt B
Person: 도상록	Institution: 김일성대학	A worksAt B
Person: 도상록	group: 일본수학물리학회	A isMemberOf B

〈그림 9〉 수강 학생이 과제로 구현한 Timeline 시각화 예시²³⁾



아래는 한국의 근대 역사(북한 유관 자료) 자료를 대상으로 위키 문서 작성 및 여러 데이터 처리 과제를 수행한 결과 학생들이 남긴 소감이다.

“저는 중국의 로컬학교에서 초·중·고를 나왔기 때문에, 한국 역사에 대해서 제대로 배워본 적이 없습니다. 한국 근대사와 관련해서는 박물관에서 본 내용과 유튜브에서 본 내용들 정도로만 알고 있었습니다. 하지만 이번 수업을 통해서, 수업 초반에는 한국 근대사에 관련해서 찾아볼 수 있는 디지털 인문학 자료들을 접할 수 있게 되었고, 수업 중반에는 북한에 대한 자료조사가 왜 한국 근대사를 이해하는 데 도움이 되는지 알게 되었습니다. 나아가서 수업 후반에 ‘도상록’에 대해 자료조사를 하며, 한국 근대사에 대해 더 자세히 알게 되었습니다. 그가 참여했던 삼일운동, 식민지 시절의 한국, 경성대학의 역사, 다양한 이유로 인한 학자들의 대규모 탈북 등 여러 가지 사건을 접

22) 〈북한 인문학 위키〉의 ‘도상록’ 페이지 내 TripleData 섹션 〈<http://dh.aks.ac.kr/~nkh/wiki/index.php/도상록#TripleData>〉

23) 〈북한 인문학 위키〉의 ‘도상록’ 페이지 내 Timeline 섹션 〈<http://dh.aks.ac.kr/~nkh/wiki/index.php/도상록#Timeline>〉

할 수 있는 계기가 되었습니다.”

“현재 한국 근대사 파악에 대한 현실과 문제점들에 대해 알 수 있어 좋았습니다. 남북분단이라는 한국의 특수성 때문에 근대사는 반쪽짜리 역사가 될 수밖에 없는 이유에 대해 이해하게 되었고, 직접 위키 페이지를 작성하면서 우리나라의 역사지만 해외의 자료들을 파악해야만 한다는 점이 씁쓸한 현실을 직시하게 했습니다. 그동안 사회주의 계열 독립운동가들에 대해 잊고 살았는데, 다시 한 번 그들에 대해 관심을 가질 수 있는 계기가 되었습니다.”

V. 디지털 환경에서의 북한 인문학 교육의 의미

1. 북한 인문학 데이터 아카이브: 교육 플랫폼으로서 활용

북한 인문학 데이터 아카이브를 활용한 강의는 ‘북한의 인문학 자료 수집과 정리’라는 아날로그 인문학과 ‘시맨틱 데이터 프로세싱’이라는 디지털 인문학의 결합을 지향하였다. 사전 동영상 강의 제작 및 시청을 접목한 플립 러닝을 적용하고 디지털 인문학에 관한 이론과 데이터 편찬 실습의 조화를 꾀하는 등 여러 실험적인 방법을 도모했지만, 여타 일반적인 성격의 인문학 강의와 비교할 때 가장 큰 차이는 소위 논문 성격의 보고서 제출이 아니라 위키 페이지 작성을 과제로 부여한 점에 있었다. 이것은 인문학 교육에서 어떠한 의미를 지닐까.

1990년대를 전후로 인문학 연구자 사이에서 컴퓨터 이용이 보편화되기 시작하자, 펜으로 종이에 쓰는 기존의 글쓰기와 워드 프로세서를 이용하는 글쓰기 사이의 차이에 관한 논의가 생겨났다. 아날로그 환경에서의 글쓰기가 선형적 글쓰기라면 디지털 환경에서의 글쓰기는 조합형 글

쓰기라는 측면이 주목되었다.²⁴⁾ 그로부터 다시 30년이 지난 오늘날에는 컴퓨터로 글을 쓰는 것이 극히 당연한 일이 되어, 그 이전의 아날로그 글쓰기와는 비교조차 하기 어려운 상황이 되었다. 한편으로 그러한 변화 가운데 학술적 글쓰기 환경은 어떻게 달라졌는가의 질문을 던진다면, 그 결과는 다소 의외다. 적어도 인문학 분야에서는 단행본이나 학술논문이 여전히 ‘글쓰기’ 형식의 주류라서 기존 아날로그 환경의 글쓰기와 큰 차이가 없고, 오히려 그것을 더 강화한 느낌마저 있다.

인문학 분야에서 오랜 기간 지속되어 온 글쓰기의 전통은 존중해야 하겠지만, 디지털 기술을 기반으로 웹 환경이 일상생활까지 침투한 오늘의 상황에서는, 그러한 전통적 방식이 사회 일반의 지식과 정보 유통 방식에서 벗어나는 측면을 내포하는 것 또한 사실이다. 이러한 모순적 상황을 극복하고자 북한 인문학 데이터 아카이브는 참여자 혹은 독자와 만나는 주요한 통로로 북한 인문학 위키를 설정했다. 위키는 아날로그와 디지털 글쓰기의 장점을 결합한 매체라 해도 과언이 아니다. 위키는 장절을 세우고 각주를 달고 문장으로 풀어쓴다는 점에서 아날로그 환경의 선형적 글쓰기를 계승한다. 동시에 풍부한 하이퍼링크를 매개로, 새롭게 기록되는 내용이 기존에 기록·축적된 내용과 연결되게끔 함으로써 비선형적 독해의 가능성을 연다. 더불어 전자지도나 타임라인 등 다양한 멀티미디어 자원을 임베딩함으로써 참여자와 독자의 흥미를 불러일으킨다.

또한 독자적으로 개발하여 도입한 데이터 입력기는 지식 그래프와 위키 문서의 결합이라는 북한 인문학 데이터 아카이브의 고유한 구조를 상징한다. 수강자가 직접 입력한 Triple Data는 관계형 데이터베이스에 반영되어 기존에 입력된 Triple Data와 함께 일종의 지식 그래프인 데이터 네트워크로 웹상에서 시각화된다. 위키의 하이퍼링크와 함께 지식 그래

24) 김지선·장문석·유인태, 「공유와 협업의 글쓰기 플랫폼 위키」, 371~419쪽.

프는, 새롭게 구축한 데이터가 기존 축적된 정보와 연결되어 지식의 외연을 확장한다는 실감을 데이터 편찬 참여자에게 안겨 준다.

데이터 아카이브 편찬 과업에 참여한 학생들은 그러한 교육 과정을 대학에서 학문을 배우고 실천하는 기회로 삼을 수 있으며, 지식의 소비자가 아니라 지식의 생산과 유통에 직접 참여하는 능동적 경험으로 수용할 수 있다. 자신이 주도적으로 생산한 인문 지식이 웹상에서 지속적으로 유통·공유될 것이라 데 대한 생각은 곧 미래 환경에서 지식 생산자의 역할과 책임이 무엇인지를 느끼고 고민하는 과정이기도 하다.

2. 데이터 편찬: 북한 이해와 통일 인식의 매개

북한 인문학 데이터 아카이브를 활용한 대학 교육은 디지털 인문학의 방법론을 도입하여 젊은 세대에게 친숙하고 자연스러운 형태로 북한 관련 자료와 연구를 접할 기회를 제공한다. 다만 디지털 인문학 방법론이 학문에 접근하기 위한 벽을 낮추는 효과가 있다는 것은 북한 인문학뿐만 아니라 어떤 주제라도 해당하는 이야기이다. 여기서는 한발 더 나아가 북한이라는 탐구 대상이 지닌 특성이 디지털 인문학이라는 방법론과 친화성을 지님으로써 연구 대상과 방법론 사이에 상승효과를 가져오는 측면을 논하고자 한다.

첫째, 북한 인문학 위키를 활용한 자료와 연구의 수집 및 정리는 북한의 학문이 고립된 채 형성, 전개되었다는 통념을 깨뜨린다. 강의에서는 보고서 제출 대신 북한 인문학 관련 위키 페이지 작성을 과제로 부과한다. 위키 페이지 작성을 위해 수강생이 수집·정리한 자료의 출처를 보면 한국은 물론 일본, 중국, 미국, 러시아, 유럽 등 다양한 나라에 걸쳐 있다. 시기적으로도 식민지 시기 자료에서 21세기 이후 생산된 연구 성과까지 포괄한다.

한 수강생은 북한 핵물리학의 아버지라고 불리는 도상록의 위키 페이지를 작성했다.²⁵⁾ 근거로서 제시된 자료 가운데 한국 국가기록원이 소장하고 있는 독립운동 판결문이 눈에 띈다. 18세의 도상록이 함흥에서 3.1 운동에 참여한 기록의 소재를 알려주는 URL(링크)이 각주에 부여되었다. 또한 도쿄제국대학 물리학과에서 수학한 도상록이 일본의 수핵물리학회에서 발표한 논문도 확인할 수 있다. 동 학회지 1927년 호에 따르면 논문 제목은 「저온에서 수소 가스의 양자역학적 이론」이고 당시 도상록은 개성에 있는 송도 고등보통학교 교원이었음을 알 수 있다. 동 학회지가 오픈 액세스 상태이므로 논문 내용 전체를 열람할 수 있는데, 웹 환경이기에 가능한 결과라 하겠다.²⁶⁾

다른 수강생이 작성한 김두봉 위키 페이지에서는 북한의 어문 정책을 확립한 한글학자 김두봉이 중국에 망명하여 임시정부는 물론 좌익 계열의 조선의용대 등에서 활동한 사실도 관련 연구가 실린 웹 사이트의 링크를 통해 쉽게 확인할 수 있다.²⁷⁾ 김두봉과 도상록을 포함한 다수의 북한 학자가 세계 여러 나라 학계와 교류했다. 예컨대 건국 초기 북한을 대표하는 역사학자인 이청원 위키 페이지에는 6·25 전쟁 직전 소련에 유학할 당시 작성된 이력서, 성적표 등의 사진이 디지털 파일로 아카이빙되어 있다.²⁸⁾

북한 관련 자료와 연구를 수집하는 과정에서 수강생은 북한 인문학이

25) <북한 인문학 위키>의 ‘도상록’ 페이지 <<http://dh.aks.ac.kr/~nkh/wiki/index.php/도상록>>

26) 일본 수핵물리학회지 1927년호 <https://www.jstage.jst.go.jp/article/subutsukai-shi1927/15/1/15_1_15/_pdf/-char/ja>

27) <북한 인문학 위키>의 ‘김두봉’ 페이지 <<http://dh.aks.ac.kr/~nkh/wiki/index.php/김두봉>>

28) <북한 인문학 위키>의 ‘이청원’ 페이지 <<http://dh.aks.ac.kr/~nkh/wiki/index.php/이청원>>

독립운동을 포함한 식민지 제국의 경험과 유산 위에 존재하는 동시에 사회주의 국가를 중심으로 하는 국제 교류 속에서 전개되었다는 사실을 자연스럽게 이해할 수 있다. 정보의 공개와 유통이 자유로운 디지털 환경의 특징을 접목한 연구와 교육이 북한 사회와 북한 인문학과 관련해 그동안 간과해 온 측면을 환기했다 평가해도 무리가 되지 않을 것이다.

둘째, 북한 인문학 위키를 활용한 연구 및 교육을 통해 북한 사회 그리고 북한 관련 자료가 지니는 특수성을 확인할 수 있다. 북한 사회가 지닌 폐쇄성 탓에 북한이 생산한 문헌 자료를 북한 바깥에서 안정적으로 접하기는 어렵다. 가까스로 구한 자료도 그것을 활용하는 차원에서는 신중할 필요가 있다. 북한 학계에서는 학자의 숙청이나 문헌의 날조가 드물지 않게 일어나기 때문이다. 북한 인문학 데이터 아카이브에서는 자료를 아카이빙할 때 관련 연구를 충분히 참고해 정확한 자료를 구하고, 나아가 숙청과 날조의 사실을 밝히려 노력하였다. 예컨대 이청원 위키 페이지에서는 이청원 숙청 후 다른 연구서에서 그의 이름이 삭제된 부분을 찾아내어 자료로 남겼다.²⁹⁾

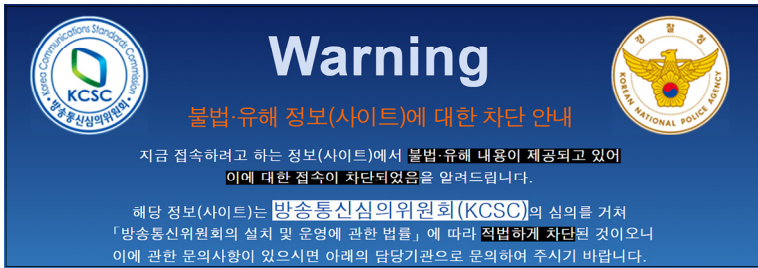
이러한 측면 때문에 북한 학계를 조망하는 연구도 북한 바깥에서 이루어지는 경우가 많다. 앞서 살핀 도상록 위키 페이지는 임정혁이 쓴 『현대 조선의 과학자들』이라는 연구를 많이 참조했다. 저자 임정혁은 재일 조선인으로 일본에서 조선총련이 깊게 관여하는 조선대학교 교원이다. 해당 문헌 역시 일본어판을 한국어로 번역한 것이다.³⁰⁾ 식민과 냉전 와중에 벌어진 한국인 디아스포라를 통해야 비로소 북한 사회를 들여다볼 수 있는 셈이다.

29) <북한 인문학 위키>의 ‘이청원’ 페이지 <<http://dh.aks.ac.kr/~nkh/wiki/index.php/이청원>>

30) 일본어판은 ‘任正燾, 『現代朝鮮の科学者たち』, 彩流社, 1997.’이며, 한국어판은 ‘임정혁(김향미 역), 『현대 조선의 과학자들』, 교육과학사, 2003.’이다.

여러 언어로 생산된 인문학 자원을 디지털화하고 아카이빙하는 과업은 해당 지역의 정부와 학계가 주도한다. 북한 인문학 자료의 아카이빙은 동일한 한국어를 사용하는 남한 사회에서 일정한 공감대가 형성되어 있어서, 통일부 북한 자료 센터 등을 중심으로 조금씩 진전되고 있는 상황이다. 다만 한국에서도 냉전과 분단의 현실은 엄연하다. 북한 자료는 여전히 ‘특수 자료’로 분류되어 이용에 제한이 많다.

〈그림 9〉 한국 정부의 북한 관련 웹사이트 차단 화면



이런 상황에서 한국에서 웹을 통해 북한 자료를 수집하고 이용하는 것은 수월성과 어려움을 동시에 지닌다. 웹을 이용할 경우 물리적 한계를 넘어 세계 각국이 구축·제공하는 디지털 아카이브를 넘나들 수 있지만, 한편으로 웹은 정부의 판단과 검열에 쉽게 노출된다는 단점 또한 지닌다. 한국에서 웹을 이용해 북한 관련 자료를 탐색하다보면, 정부의 인터넷 검열에 따라 〈그림 9〉와 같은 접속 차단 화면과 맞닥뜨리는 경우가 허다하다.

이와 관련해 제한적이기는 하나 북한이 자체적으로 인문학 자원을 디지털화해 웹을 통해 공개한 사례도 있다. 예컨대 일본에서는 조선총련과 관계가 있는 것으로 판단되는 조선언론정보기지(KPM)라는 회사가 북한 학술 자료 등의 기사 목록과 내용을 웹 서비스하고 있다.³¹⁾ 하지만 이

웹 사이트는 한국 내에서는 접속이 어렵다. 앞서 언급한 임정혁이 교원으로 있는 조선대학교에 관한 정보를 수집하고자 해당 학교의 웹 사이트를 탐색하면 마찬가지로 <그림 9>와 같은 접속 차단 화면으로 연결된다.

한국 사회에서는 중국과 북한의 인터넷 검열을 입을 모아 비판하지만, 우리 정부의 인터넷 검열에 대한 심도 있는 논의는 부족하다. 북한 인문학 데이터 아카이브의 실천은 한국 사회의 냉전적 현실을 포함하여 북한 자료가 지니는 특수성을 실감하는 계기로 작동한다.

VI. 나오며: 향후 전망과 과제

북한 인문학 데이터 아카이브는 구상 단계부터 이미 교육적 활용을 염두에 두었다. 사업계획서에서는 추진 배경으로서 먼저 북한 인문학 분석을 통해 북한 사회에 대한 이해력을 높이겠다고 밝혔다. 단기적, 정책적 북한론으로는 우리 사회 일각의 북한 혐오를 불식시키기 어렵다는 판단 아래 북한 사회 나름의 상식을 구성하는 인문학적 축적을 분석함으로써 북한에 대한 생산적 비판을 시도하였다. 다음으로 디지털 기술을 활용함으로써 젊은 세대 사이에 통일·평화 연구의 저변을 확대하겠다고 밝혔다. 20대 혹은 대학생의 통일과 평화에 대해 무관심이 지적되는 상황에서 디지털 웹 환경을 활용해 통일·교류협력 대상으로 북한을 수용할 계기를 제공하고자 한 것이다.³²⁾

31) <조선언론정보기지> 웹사이트 <<http://www.dprkmedia.com>>. <조선언론정보기지>에 대해서는 2006년 통일뉴스 기사 참조. 「북 언론통합사이트KPM 개통」, 『통일뉴스』(온라인) 2006년 2월 22일, <<http://www.tongilnews.com/news/articleView.html?idxno=63209>>

이 글은 북한 인문학 자료를 대상으로 디지털 환경에서 데이터 아카이브를 설계·구축한 논리와, 해당 아카이브를 매개로 진행한 디지털 인문학 교육에 관한 내용을 함께 정리한 것으로서 의미를 갖는다. 유관 내용은 아래 네 가지 정도로 압축할 수 있다.

첫째, 북한 인문학 데이터 아카이브는 시맨틱 데이터 프로세싱(Semantic Data Processing) 방법론에 근거하여 설계·구축하였다. 일반적인 디지털 아카이브가 단순 문자열(n-gram) 검색을 통해 문헌 자료의 메타데이터 정도만 제공하는 것과 달리, 북한 인문학 데이터 아카이브는 북한 학술 정보의 지식 관계망을 효과적으로 표현할 수 있는 의미론적 기반의 데이터 아카이브를 지향하였다. 이를 위해 온톨로지를 설계하여 지식그래프를 구현하도록 하였는데, 위키를 활용한 데이터 편찬 과업과 해당 결과물을 결합하였다.

둘째, 서울대학교 인문대학은 교육부의 혁신융합대학 사업의 일환으로 디지털 인문학 교육 과정을 개설·운영하고 있다. 2021년에 시작된 교육부의 ‘디지털 신기술 인재양성 혁신융합대학 사업’의 일환으로 서울대학교를 주관대학으로 하는 7개 대학 컨소시엄은 ‘빅데이터 혁신융합대학 사업’을 추진하고 있다. 이에 발맞춰 서울대학교 인문대학은 4개의 디지털 인문학 교과과정을 개설하였는데 그 중 하나가 ‘빅데이터 역사정보’이다. 이 글에서 다루는 ‘시맨틱 데이터 편찬’은 ‘역사정보 텍스트마이닝’, ‘역사지리정보’와 함께 ‘빅데이터 역사정보’ 교과과정을 구성하는 세 분야 중 하나다.

셋째, ‘시맨틱 데이터 편찬’ 교과 강의는 기본적으로 북한 인문학 데이터 아카이브를 연계하여 운용함으로써, 북한 인문학 및 한국 근대사에

32) 서울대학교 통일평화연구원 웹사이트 ‘통일기반구축사업’ 페이지 <https://ipus.snu.ac.kr/wp-content/uploads/2023/03/17.-2022-통일평화기반구축사업-결과보고서_인문학연구원홍중욱.pdf>

관한 도메인 지식과 디지털 인문학 방법론으로서 시맨틱 데이터 처리에 관한 내용을 동시에 익히는 융합교육이 가능하도록 하였다. 관련해서 빅 데이터 혁신공유대학의 마이크로 디그리 과정의 일환으로 〈미디어위키를 활용한 역사 데이터 편찬〉 강의를 진행하고 있다. 해당 강의에서는 위키 문법 습득, Triple Data 생성, Timeline 작성, Network Graph 구현 등의 디지털 인문학 방법론과 더불어 북한 인문학 자료를 수집·정리하는 역량을 기른다.

넷째, ‘시맨틱 데이터 편찬’ 교과목의 수강생은 웹 환경에서 북한 인문학과 한국 근대사에 관한 지식을 다룸으로써, 아날로그 환경에서 유관 지식을 소비하는 수동적 입장에서 벗어나 디지털 환경에서 인문 지식을 생산·유통하는 능동적 경험을 익힐 수 있다. 아울러 북한 학계가 통시적·공시적으로 지식의 국제적 네트워크 속에 존재함을 확인함과 동시에 냉전과 분단 상황에서 전개된 북한 인문학의 특수성을 이해할 수 있다. 수강생은 위키 페이지 작성 과제를 수행하는데, 위키는 아날로그 환경의 선형적 글쓰기와 디지털 환경의 조합형 글쓰기 양쪽의 장점을 모두 갖춘 미디어다. 무엇보다 위키에 작성·표현된 지식이 웹을 통해 개방·공유된다는 점에서, 수강생은 우리 사회의 지식 생산에 참여한다는 자부심과 책임감을 느낄 수 있다. 또한 웹 환경에서 국경과 시대를 넘나들며 자료를 수집, 정리함으로써 북한의 학문이 고립된 채 형성·전개되었다는 통념을 깨뜨릴 수 있다. 동시에 북한 인문학 자료를 다루는 데 있어서 남북한 사회 모두 냉전의 제약을 받고 있다는 점 또한 확인할 수 있다.

【참고문헌】

김현·임형상·김바로, 『디지털 인문학 입문』, 서울: HueBooks, 2016.

임정혁(김향미 역), 『현대 조선의 과학자들』, 파주: 교육과학사, 2003.

김바로, 「〈공공데이터법〉과 인문데이터 -공공기관 보유 인문데이터 공개 신청 사례를 중심으로」 『한국고전연구』 제57호, 2022, 167~192쪽.

김지선·장문석·유인태, 「공유와 협업의 글쓰기 플랫폼 위키」 『한국학연구』 제60집, 2021, 371~419쪽.

김지선·유인태, 「지식의 공유와 표현 그리고 디지털 인문학 : 해외 디지털 인문학 연구 사례 검토」 『인문논총』 제79권 2호, 2022, 391~427쪽.

이지수·이혜은, 「미국 대학의 디지털 인문학 교육 프로그램 연구」, 『정보관리학회지』 36권 4호, 한국정보관리학회, 2019, 107~128쪽.

장문석·김윤진·이지수·송가배·고자연·김지선, 「디지털 인문학과 지식의 공동생산 - 위키 플랫폼과 〈한국 근대 지식인 아카이브〉 편찬」 『인문논총』 제78권 1호, 2021, 75~124쪽.

허수, 「교수논단 디지털 신기술 혁신공유대학 사업」 『서울대 인문대학 소식지』 53(2021년 가을), 7쪽.

홍종욱·강수연·홍수현·김도민, 「북한 인문학 시맨틱 데이터 아카이브 구축과 활용」 『인문논총』 제78권 1호, 2021, 13~41쪽.

서울대학교 교무처 교무과, “「서울대학교 교과인증과정」 운영 가이드라인”

「북 언론통합사이트KPM 개통」, 『통일뉴스』(온라인) 2006년 2월 22일

〈<http://www.tongilnews.com/news/articleView.html?idxno=63209>〉

일본 수하물리학회지 1927년호

〈https://www.jstage.jst.go.jp/article/subutsukaishi1927/15/1/15_1_15/_pdf/-char/ja〉

빅데이터 혁신융합대학 홈페이지 〈<https://bigdatahub.ac.kr/>〉

조선언론정보기지 홈페이지 〈<http://www.dprkmedia.com>〉

통일평화연구원 홈페이지 〈<https://ipus.snu.ac.kr/>〉

북한 인문학 시맨틱 데이터 아카이브 〈<http://dh.aks.ac.kr/~nkh/index.html>〉

지암일기 데이터 아카이브 〈<http://jamdiary.info/>〉

한국학중앙연구원 디지털인문학연구소 제작 Semantic Database Visualization

〈http://dh.aks.ac.kr/Edu/wiki/index.php/Semantic_Database_Visualization〉

CBDB 홈페이지 〈<https://projects.iq.harvard.edu/cbdb>〉

Enslaved 홈페이지 〈<https://enslaved.org/>〉

O Say Can You See 홈페이지 〈<https://earlywashingtondc.org/>〉

North Korean Humanities Data Archive and Digital Humanities Education

Ryu, Intae (Korea University)

Hur, Soo (Seoul National University)

Hong, Jong-wook (Seoul National University)

Abstract

This article summarizes the experience of creating the North Korean Humanities Data Archive(NKHDA) and using it as a platform for digital humanities education. The relevant information is organized in four parts.

First, the North Korean Humanities Data Archive(NKHDA) was designed and created based on the Semantic Data Processing methodology to effectively represent the three-dimensional nature of academic information and the complex network of knowledge network between them.

Second, as part of the Ministry of Education's Innovative Convergence University project, the College of Humanities at Seoul National University(SNU) has established and operated digital humanities education courses. Among them, 'Semantic Data Compilation' was introduced and launched as a part of the 'Big Data Historical Information' curriculum.

Third, the course 'Semantic Data Compilation' is operated in conjunction with the North Korean Humanities Data Archive(NKHDA). The course is designed to provide students with domain knowledge on North Korean humanities and modern Korean history, as well as semantic data processing as a digital humanities methodology.

Fourth, by accessing knowledge about North Korean humanities and Korean modern history on the web, students taking the 'Semantic Data Compilation' course are able to move from a passive position of consuming relevant knowledge in the analog environment to an active experience of producing and distributing humanities knowledge in the digital environment. In addition, students are able to see that North Korean academia exists in an international network of knowledge, both synchronously and publicly, and understand the specificity of North Korean humanities as it developed in the context of the Cold War and the division of Korea.

Keywords: North Korea humanities, North Korean Humanities Data Archive, big data innovation convergence university, digital humanities

류인태(Ryu, Intae)

고려대학교 한자한문연구소에서 연구교수로 재직 중이다. 조선후기에 작성된 다양한 유형의 한문 기록을 대상으로 한 번역 및 연구를 꾸준히 수행하고 있다. 그와 동시에 디지털 인문학 방법론을 적용한 실험적 연구와 교육 또한 지속적으로 기획·참여하고 있다. <북한 인문학 데이터 아카이브> 구현 프로젝트 연구팀에서는 데이터 처리와 관련된 여러 실무를 맡았다. 주요 논문과 저서로는 「데이터로 읽는 17세기 재지사족의 일상: 『지암일기(1692-1699)』 데이터베이스 편찬 연구」(2019), 「디지털 인문학은 인문학이다」(2020), 『디지털로 읽고 데이터로 쓰다』(공저, 2023) 등이 있다.

허수(Hur, Soo)

서울대학교 대학원 국사학과 부교수로 재직 중이다. 한국 근대 사상사를 전공하고 개념사 연구 및 디지털 역사학 연구를 진행 중이다. 주요 저서와 논문으로는 『이돈화 연구: 종교와 사회의 경계』(2010), 「근대 전환기 '개벽'의 불온성과 개념화: 동학, 천도교를 중심으로」(2021) 등이 있다.

홍종욱(Hong, Jong-wook)

서울대학교 인문학연구원 부교수로 재직 중이다. 식민지 조선과 북한을 포함한 20세기 한국 지식인의 사회사상, 역사 인식을 연구하고 있다. 디지털 인문학에도 관심을 가져 <북한 인문학 데이터 아카이브> 구축에 참여했다. 주요 논저로는 『戰時期朝鮮の轉向者たち: 帝國/植民地の統合と亀裂』(2011), 「1930년대 마르크스주의 역사학의 아시아 인식과 조선 연구」(2021), 「주체 사관에서 인민과 민족의 자리」(2022) 등이 있다.