

디지털 시대, 인문학 활동의 디지털 전환

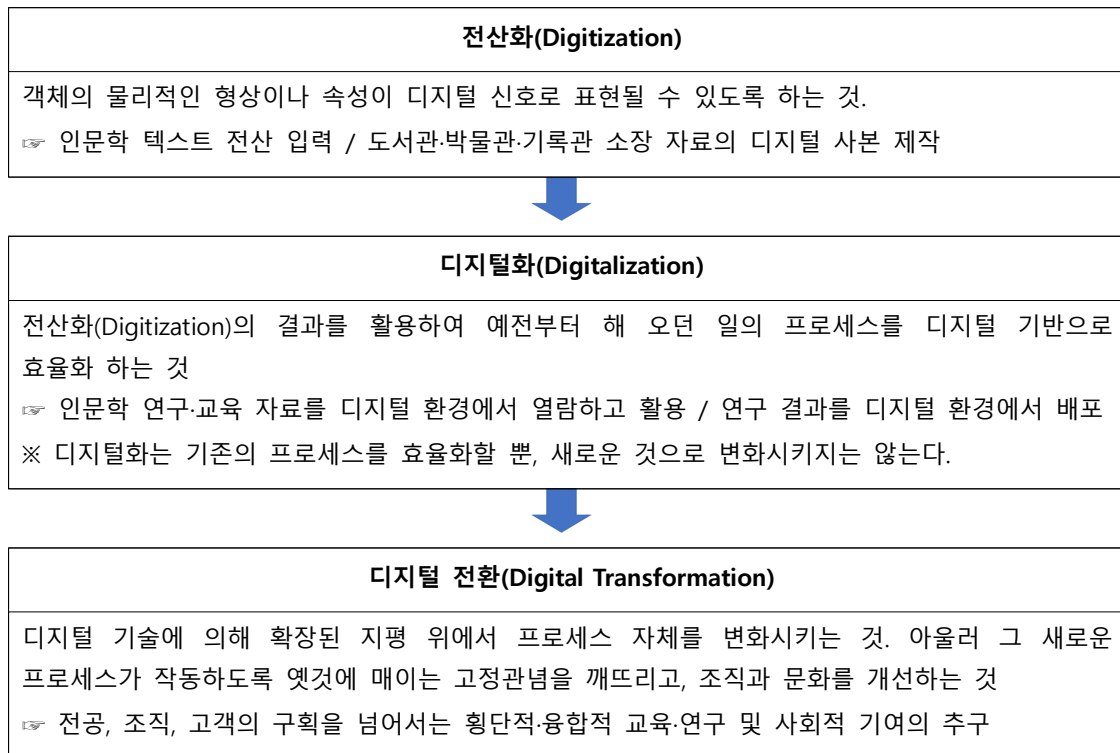
한국의 사례: 한양도성 타임머신 프로젝트

김 현 (金 炫)¹

1. 디지털 인문학: 디지털과 인문학의 조우에 관한 현재의 화두

디지털 세계에서 변화해 가는 인문학 활동의 과거와 현재를 ‘디지털화’(Digitalization)와 ‘디지털 전환’(Digital Transformation)이라는 개념으로 구분해 보고자 한다.

※ 디지털 세계로 이행하는 3가지 단계



한국의 인문학계가 ‘전산화(Digitization)’ 또는 ‘디지털화(Digitalization)’ 라는 이름으로 디지털 기술을 대면하게 된 것은 1990년대의 일이다. 한 세대를 지나는 동안 인문학의 교육·연구

¹ 韓國學中央研究院 人文情報學 教授. 디지털人文學研究所長

환경은 급변하였다. 학자들은 연구 자료의 상당 부분을 온라인상에서 찾아볼 수 있게 되었고, 초등학생들도 학교에서나 가정에서 컴퓨터를 학습의 동반자로 삼게 되었다. 그러나 그 변화가 아무리 대단한 듯이 보여도, 바뀐 것은 지식을 담는 매체와 그 매체를 사용하는 방법일 뿐 인문학 활동의 내용과 성격이 크게 달라진 것은 아니다. 디지털 기술로 인해 교육과 연구의 방법이 효율화 되었을 뿐이다.

지금까지 인문학 분야에서 ‘디지털화(Digitalization)’로 불리던 일은 주로 인문학 연구나 교육 자료를 디지털 환경에서 보다 신속하고 편리하게 이용할 수 있게 그 콘텐츠를 데이터베이스 담고 온라인상에서 접근할 수 있게 하는 일이었다. 이 일은 사실상 학자나 교사 학생들과 같은 인문학 활동 당사자들이 아니라 이들의 요구에 따라 정보기술자들이 용역으로 한 일이 대부분이었다. 수요자와 실행자가 다를 경우 이들이 할 수 있는 일은 ‘기존의 일을 개선하는’ 수준에서 크게 벗어날 수 없는 것이 당연하다.

한국의 인문학계는 이제 인문학 활동의 ‘디지털 전환(Digital Transformation)’이라는 새로운 화두를 접하고 있다. ‘디지털 인문학(Digital Humanities)’이라는 이름으로 수행하는 인문학 활동의 ‘디지털 전환’이 예전의 ‘디지털화’와 다른 점은 그 일의 수행 주체가 바로 학자, 교사, 학생들과 같은 인문학 활동 당사자들이라는 점이다.

인문학 활동의 당사자들의 직접 디지털 환경에서 운용할 수 있는 인문지식 데이터를 생산하고, 그 데이터를 기반으로 새로운 융합적 지식의 생산과 횡단적 응용을 시도하며, 거기에서 도출된 요구에 따라 새로운 데이터를 탐색하고 가공하는 일이 이른 바 ‘디지털 인문학’이라고 하는 실천적인 학술 활동의 골자이다. 이 활동을 통해 궁극적으로 얻고자 하는 것은 기존의 것을 그대로 디지털 세계에 옮겨 놓는 것이 아니다. 디지털이 아니던 시절에는 불가능했고 그래서 생각해 본 적도 없지만, 이제는 디지털이기 때문에 가능해진 새로운 일들이 실제로 이루어지도록 하는 것이다.

전통적인 인문학 연구와 교육이 주로 학자 개인의 활동 위주였던 데 반해, 학제적 소통과 횡단적 응용을 지향하는 디지털 인문학은 산·학·연의 여러 분야의 연구자들이 협력할 수 있는 융합연구 프로젝트를 적극적으로 지향한다. 현재 한국에서 태동하고 있는 디지털 인문학²

활동의 예시로서 한국학중앙연구원 디지털인문학연구소에서 수행한 연구 사례를 소개하기로 한다.

2. 디지털 환경의 인문학 활동 사례: 한양도성 타임머신

1) 한양도성 타임머신 프로젝트

‘한양도성 타임머신’ 프로젝트는 대한민국 정부가 추진하는 콘텐츠산업 3대 혁신전략³ 과제의 일환으로 2020년부터 2022년 말까지 대한민국의 문화재청이 주도한 문화유산 디지털 콘텐츠 개발 사업이다. 이 사업은 “한양도성 600년 문화유산을 3D 데이터로 만들고, 그 데이터를 민간에서 활용할 수 있도록 개방형 플랫폼으로 구축하여, 이동통신사•포털•게임제작사 등이 다양한 가상현실 콘텐츠로 활용할 수 있도록 하는 것”을 목표로 하였다.

한양도성 타임머신은 이 이전까지 한국에서 시도된 적이 없었던 거대 규모의 “문화유산 3D 디지털 콘텐츠 개발 프로젝트”이다. 대한민국의 수도 서울의 옛 모습인 조선시대의 한양을 3차원 가상 세계에서 볼 수 있게 것 만도 충분히 의미 있는 일이지만, 이 프로젝트는 또 하나의 도전적인 과업을 안고 출발했다. 물리적인 형상을 가진 건조물만 3차원 가상현실로 재현하는데 머물지 않고, 그 안에서 살아갔던 사람들의 삶과 역사를 디지털 데이터로 확인할 수 있게 하는 과업을 병행하였다. 이를 위해 36명의 인문학 전문연구자들이 유의미한 기록물들을 수집하고, 분석하고, 데이터화 하는 데이터 큐레이터의 역할을 수행하였다.⁴

2 한국 대학의 디지털인문학연구소:

- 한국학중앙연구원 디지털인문학연구소 <http://dh.aks.ac.kr/>
- 한림대학교 디지털인문학연구소 <https://sites.google.com/site/dhhallym/>
- 서울대학교 AI연구원 인공지능 디지털인문학센터 https://aiis.snu.ac.kr/sub1_4_30.php
- 카이스트 디지털인문사회과학센터 <https://dhcss.kaist.ac.kr/>
- 고려대학교 인문융합연구원 디지털인문센터 <https://kucdh.korea.ac.kr/kucdh/index.do>
- 고려대학교 민족문화연구원 디지털인문학센터 <https://riks.korea.ac.kr/organization/digitalhumanities>
- 아주대학교 디지털휴머니티 연구센터, 디지털역사 연구센터 <https://www.ajou.ac.kr/human/research/research01.do>
- 한국외국어대학교 디지털인문학한국학연구소 <http://dhks.hufs.ac.kr/>

3 ‘콘텐츠산업 3대 혁신전략’, 대한민국 정책 브리핑, 2020. 3. 17.

<https://www.korea.kr/special/policyCurationView.do?newsId=148866166>

한양도성 프로젝트에 참여한 인문학 연구자들은 역사학, 민속학, 문화인류학, 전통의상, 전통음식학, 한문학, 그리고 디지털 인문학 전공자들이다. 이들은 다양한 연구 자료에서 찾아낸 지식정보가 직접 컴퓨터 시스템 상에서 활용될 수 있도록 기계가독형 데이터를 생산하였다. 한양도성의 문화유산과 관련이 있는 인문지식을 데이터로 전환하기 위해서 연구팀이 택한 방법은 옛 문헌에 기록된 내용 속에서 사실을 이야기하는 지식의 노드와 그 노드 사이의 관계를 추출하여, RDF(Resource Description Framework)⁵ 형식의 시맨틱 네트워크 데이터를 만드는 것이었다.

2) 한양도성 타임머신 프로젝트의 디지털인문학 연구

A. 한양도성 시맨틱 데이터 아카이브 구축 프로세스

한양도성 타임머신 프로젝트 인문학연구자들은 시맨틱 데이터 아카이브 구축 업무를 다음과 같은 프로세스로 진행하였다.

① 기초 연구 자료 선정

- 한양도성의 문화유산의 역사적·문화적 배경과 관련 지식을 얻을 수 있는 다양한 문헌자료를 조사하여 그 내용을 분석, 정리
- ※ 조사 대상 문헌의 종류: 고문헌, 학술서적, 논문, 교양서적, 전시도록, 사진첩, 신문기사, 박물관 유물, 아카이브 소장 기록물, 웹 자원



② 디지털 스토리텔링 주제 선정

- 한양도성 당해 연도 대상 사업지역의 공간성과 관련이 있는 시간적 사건, 인물을 중심으로 하는 스토리 발굴
- 이 스토리가 그 안에 적절한 수준의 문맥 요소들(인물, 사건, 장소, 사물, 문헌,

⁴ 한국의 문화재청이 기획한 ‘한양도성 타임머신 프로젝트’의 실무는 (주)PCN(컴퓨팅 시스템 전문 회사)를 중심으로 하는 컨소시엄이 수행하였다. 한국학중앙연구원 디지털인문학연구소는 이 컨소시엄의 일원으로 디지털인문학 연구에 기반한 시맨틱 데이터 아카이브 구축 업무를 담당하였다.

⁵ RDF는 W3C(World Wide Web Consortium)가 제안하는, 웹상의 데이터 교환을 위한 표준 모델이다. RDF 구문의 형식은 웹의 하이퍼링크 구조를 확장한 것으로, 연결의 출발점과 도착점뿐 아니라 그것들 사이의 관계도 URI로 명명한다. (Resource Description Framework (RDF), <https://www.w3.org/RDF/>)

개념용어 등)을 포함하고 있는지 검토하여 데이터 큐레이션 출발점으로 선정.
※ 추상적인 이야기, 사실보다는 해석이나 판단에 치중하는 이야기는 제외



③ 데이터 큐레이션
- 스토리를 구성하는 문맥요소를 찾아서 이것을 노드화 하고 그 요소들 사이의 관계를 정하는 방식으로 시맨틱 데이터 네트워크 개발
※ 온라인 협업 환경 상에서 연구자 개개인의 큐레이션이 다른 연구자의 큐레이션 결과와 연결될 수 있도록 하고, 이러한 확장에 오류와 중복이 없는지 상시 점검

B. 온톨로지 개발 (EKC Data Model 2021)

데이터 큐레이션을 위한 RDF 문에 적용할 온톨로지⁶ 스키마는 한국학중앙연구원 디지털인문학연구소에서 2016년에 처음 제정하고, 매년 확장해 오고 있는 EKC(Encyves of Korean Culture) 데이터 모델을 기반으로 하였다.⁷ 연구팀은 새 프로젝트의 기초 연구 자료와 스토리 주제가 선정된 시점에 그 프로젝트의 데이터 큐레이션에 적용할 온톨로지 초안을 제정하고, 개정된 온톨로지에 입각하여 데이터 큐레이션을 수행한다. 아울러 온톨로지 어휘의 관리를 담당하는 태스크 포스를 구성하여 온톨로지 어휘의 사용 현황을 모니터링하고, 새로운 어휘 제정의 요구가 있을 때 이를 판단, 제정, 공시하는 일을 수행한다.

§ EKC Data Model 2021 Class 정의

 http://dh.aks.ac.kr/hanyang2/wiki/index.php/Ontology:EKC_2022:Class

§ EKC Data Model 2022 Object Property 정의

 http://dh.aks.ac.kr/hanyang2/wiki/index.php/Ontology:EKC_2022:Relation

⁶ 온톨로지란 정보화의 대상이 되는 세계를 전자적으로 표현할 수 있도록 구성한 데이터 기술 체계이다. 넓은 의미에서는 모든 정보화의 틀이 다 온톨로지일 수 있겠지만, 대상 자원을 '클래스'(class)로 범주화하고, 각각의 클래스에 속하는 개체(individuals)들이 공통의 '속성'(attribute)을 갖도록 하고, 그 개체들이 다른 개체들과 맺는 '관계'(relation)를 명시적으로 기술하는 것이 가장 일반적인 온톨로지 설계 방법이라고 할 수 있다.

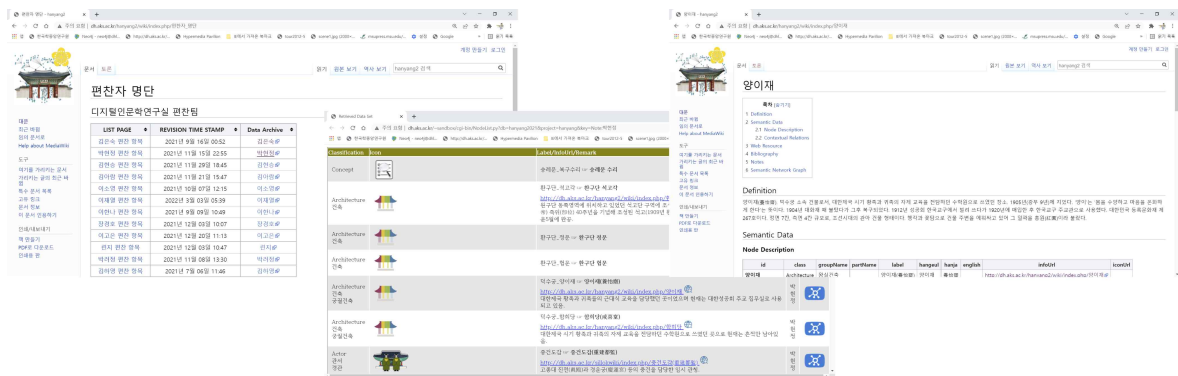
⁷ EKC 온톨로지의 도메인: 한국의 전통문화에 관한 역사적 사실 관계 및 그 사실의 문헌적 근거에 관한 지식)

C. 온라인 협업 플랫폼의 운영

한양도성 데이터 아카이브의 시맨틱 데이터는 모두 왕실의례, 전통복식, 전통음식, 전통미술을 연구하는 전문연구자들이 직접 문헌자료를 읽고, 그 내용을 분석해서 추출한 데이터이다. 여러 연구자들이 각각의 전문분야에서 추출한 데이터를 종합하고 정리하는 과정에서 중복되거나, 불명확한 데이터를 검토 정리하고, 필요한 경우 새로운 온톨로지 어휘를 추가하기도 한다. 그리고 다시 보완된 온톨로지에 입각해서 새로운 데이터를 발굴하는 작업을 수행한다. 연구팀이 이러한 형태의 협력연구를 위해 활용한 디지털 연구 지원 시스템은 다음의 두 가지이다.

§ 한양도성 위키

미디어 위키(MediaWiki)TM 기반의 ‘한양도성 위키’는 한양도성 데이터 아카이브 연구자들의 데이터 큐레이션을 위한 온라인 협업 시스템이다. 기초연구자료 선정, 스토리 선정, 온톨로지 확장 등 한양도성 프로젝트 안에서 수행된 모든 작업의 결과가 이 위키 시스템에 공시된다. 또한 개별 큐레이터들은 자신이 데이터 큐레이션의 과정에서 생산한 데이터를 이 위키 페이지에 기록하여 다른 연구자들이 그 데이터를 공유할 수 있도록 하고 있다.

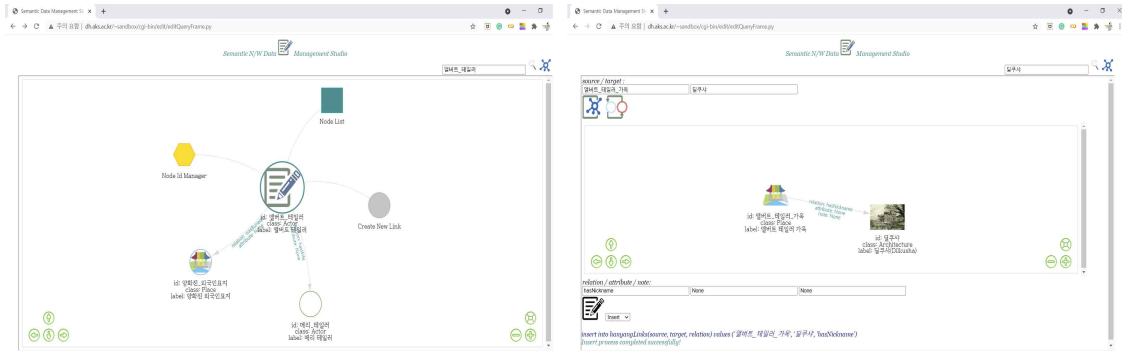


한양도성 위키: 위키 플랫폼 기반의 데이터 편찬 협업 시스템

§ 시맨틱 네트워크 데이터 관리기

한국학중앙연구원 디지털인문학연구소에서 개발한 ‘시맨틱 네트워크 데이터

관리기'(Semantic Network Data Management Studio, SN-DMS)는 노드와 링크로 구성된 시맨틱 데이터를 실시간으로 관리할 수 있는 소프트웨어이다.



시맨틱 네트워크 데이터 관리기

3) 인문지식 시맨틱 데이터 편찬 사례: 무신년진찬(戊申年進饌)

한양도성 타임머신 프로젝트에 참여한 인문학 연구자들은 2020년부터 2022년말까지 약 3년에 걸쳐, 조선시대(1392~1010) 한양의 역사와 조선왕실 문화를 중심으로 하는 시맨틱 데이터를 생산하였다: 70,000여 건의 노드와 100,000여 건의 링크 데이터.

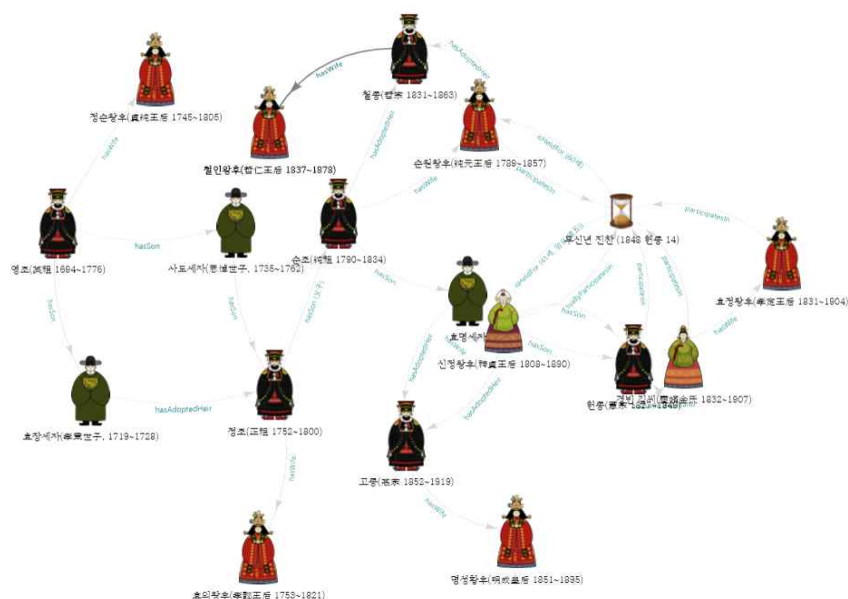
이 데이터의 내용과 성격에 대한 이해를 돕기 위해 한 가지 예시를 보이도록 하겠다.

A. 무신년진찬의 장소, 기록물, 인물

1848년 한양도성에 있는 조선왕실의 궁궐 창덕궁(昌德宮)과 창경궁(昌慶宮)에서 조선의 왕 헌종(憲宗, r. 1834-1849)은 그의 할머니 순원왕후(純元王后, 1789-1857)와 어머니 (神貞王后, 1809-1890)의 생신을 축하하고 장수를 기원하는 잔치를 열었다. 이 잔치의 준비와 진행, 마무리에 관한 구체적인 사실들은 『무신년진찬의궤(戊申進饌儀軌)』에 기록되어 있으며, 잔치의 주요 장면들은 「무신진찬도병(戊申進饌圖屏)」에 그림으로 묘사되어 전한다.



무신년진찬의 장소와 기록물



무신년진찬의 참석자와 왕실계보

B. 무신년진찬의 복식(服飾)

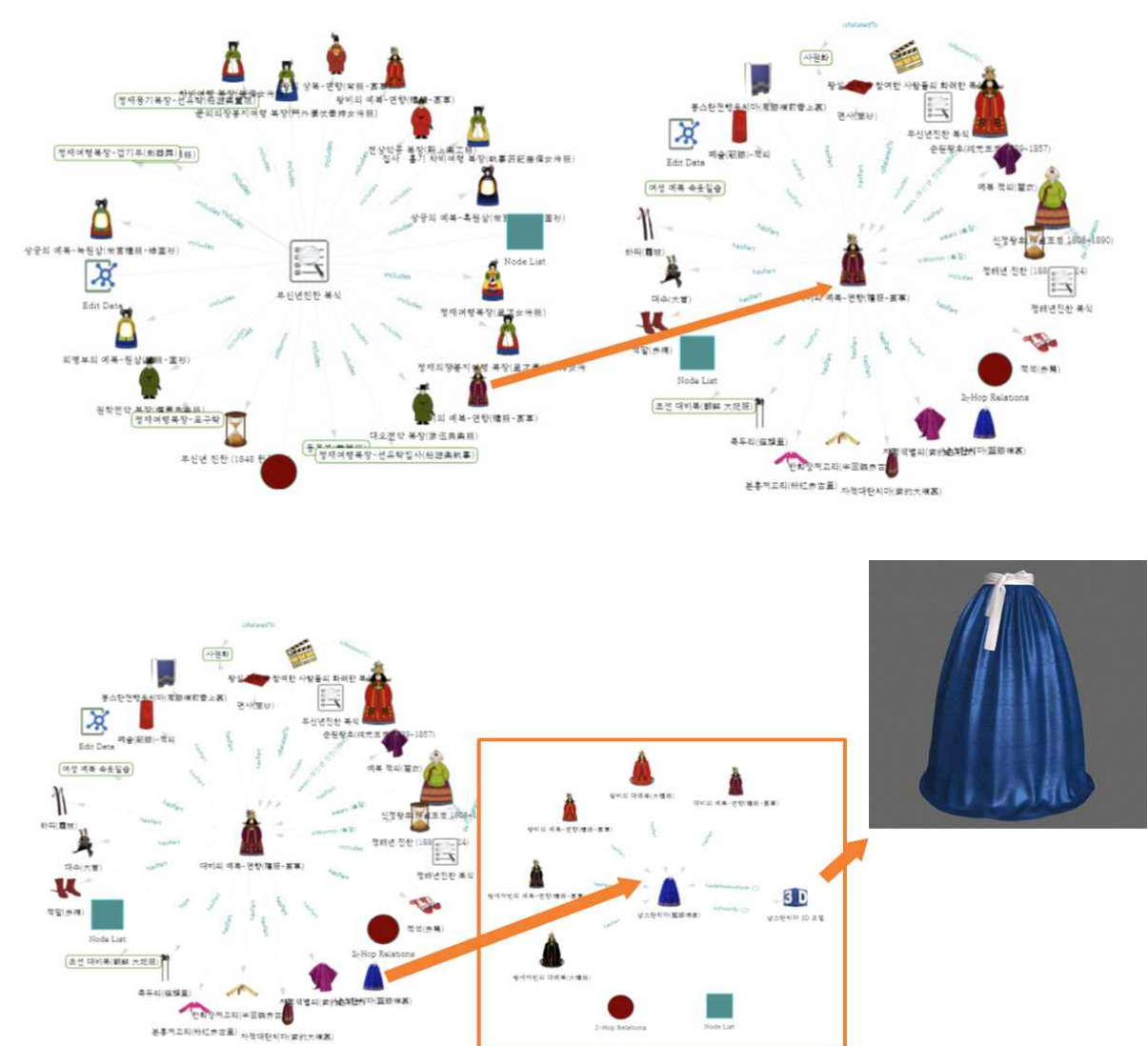
§ 고증자료

『戊申進饌儀軌』 : 「儀衛」, 「工伶」, 「樂器風物」, 『經國大典』, 『國朝五禮儀』, 『尙方定例』, 왕실인물 초상화, 박물관 소장품

§ 복식(服飾) 데이터의 관계성 정의 (Object Property)

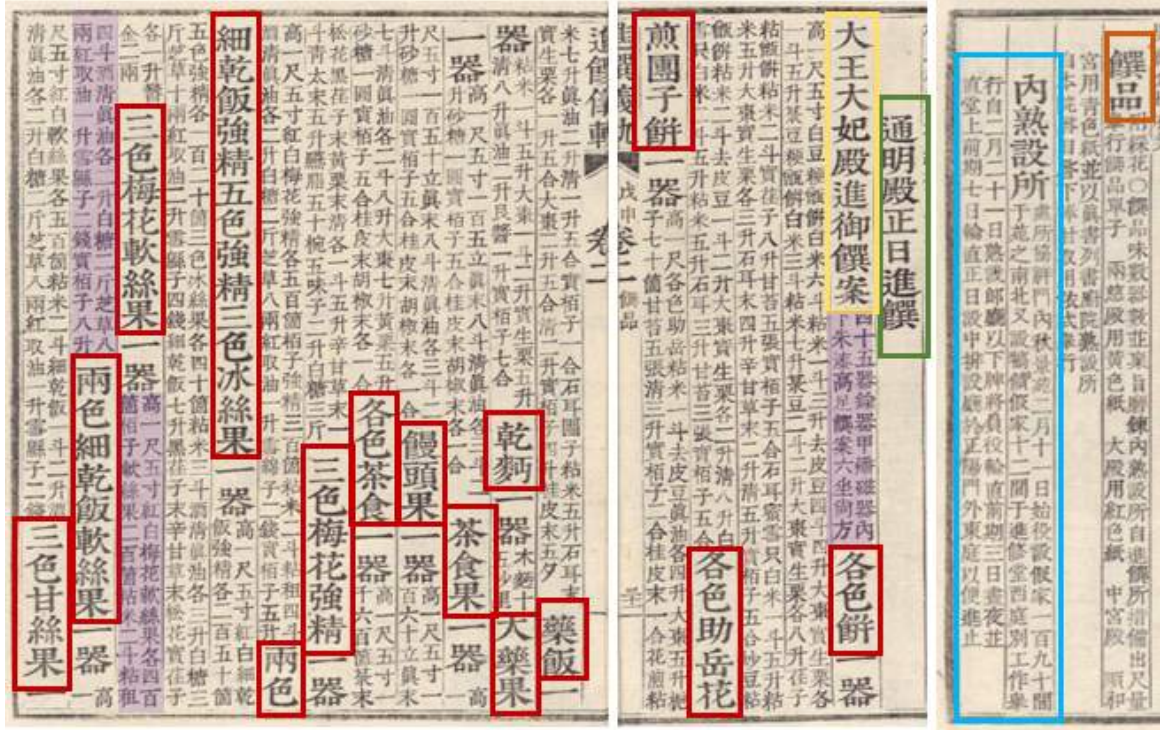
Relation	Domain	Range	Inverse Relation	Attribute	Note
wears	Actor	Clothing>Costume			[Actor] A wears [Costume] B
isWornIn	Clothing	Event/Concept			[Costume] A is worn in [Event] B
isDepictedIn	Any	Record/Object	depicts		A is depicted in [Record/Object] B
isMentionedIn	Any	Record	mentions		A is mentioned in [Record] B
documents	Record	Event			[Record] A is documentation of [Event] B
hasPart	Any	Any	isPartOf		[Costume] A includes [Clothing Item] B as a part

§ 무신년진찬 복식(服飾) 데이터 관계망



C. 무신년진찬의 음식(飮食)

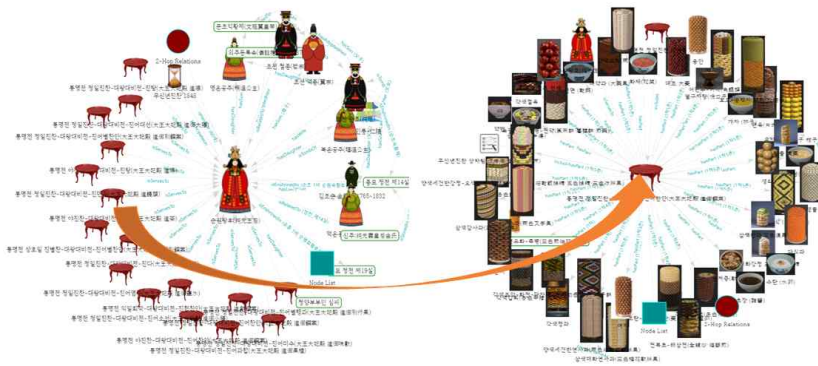
§ 고증자료: 『戊申進饌儀軌』 「饌品」



§ 음식(飮食) 데이터의 관계성 정의 (Object Property)

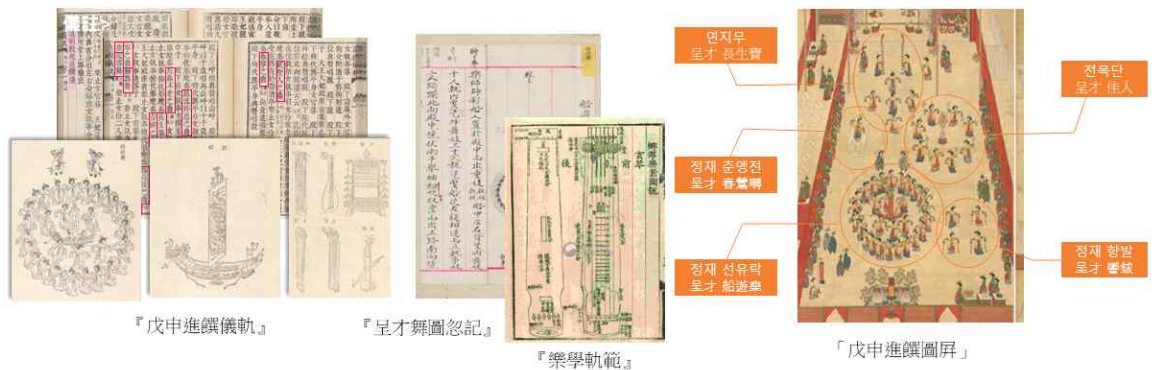
Relation	Domain	Range	Inverse Relation	Attribute	Note
hasPart	Food>Menu	Food>Individual Food		Hight of Individual Food B	[Menu] A include [Individual Food] B
hasIngredient	Food>Individual Food	Food>Ingredient		Amount of Ingredient	[Individual Food] A uses [Ingredient] B
isGarnishedWith	Food>Menu	Object			[Menu] A is garnished with [Decoration Object] B
isProvidedIn	Food>Menu	Event			[Menu] A is provided in [Event] B
isServedIn	Food>Individual Food	Object			[Individual Food] A is served in [Object:bowl/dish] B
isServedOn	Food>Individual Food/Menu	Object			[Individual Food/Menu] A is served on [Object:food table] B
isServedTo	Food>Menu	Actor			[Menu] A is served to [Actor] B

§ 무신년진찬 음식(飮食) 데이터 관계망



D. 무신년진찬의 정재(呈才)⁸

§ 고증자료: 『戊申進饌儀軌』 : 「儀註」, 「樂器風物」, 「工伶」,
 『呈才舞圖忽記』, 『樂學軌範』, 「戊申進饌圖屏」



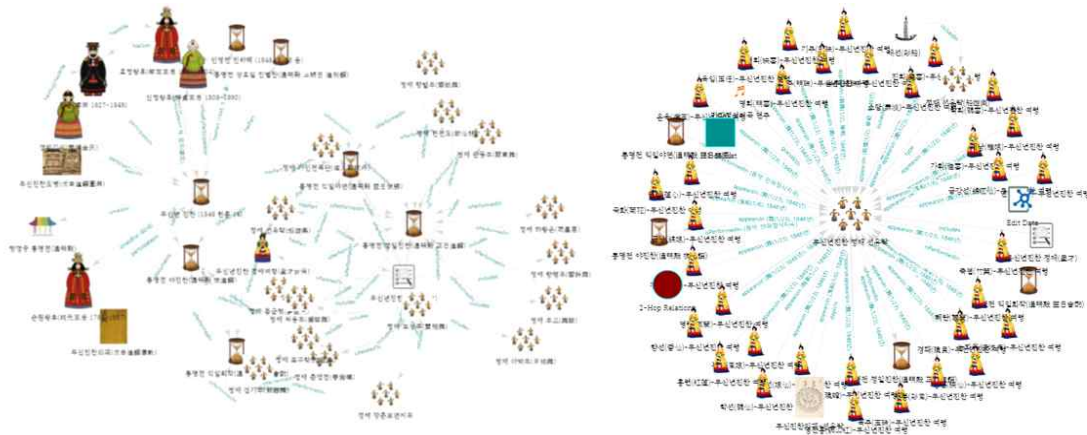
§ 정재(呈才) 데이터의 관계성 정의 (Object Property)

Relation	Domain	Range	Inverse Relation	Attribute	Note
hasPerformance	Event	Work	isPerformedIn		[Event] A has [Performance] B
isPerformedAt	Event/Concept	Place			[Performance] A is Performed at [Place] B
isPerformedBy	Event/ Work>Performance	Actor			[Event/Work] is Performed by [Actor]
appearsIn	Actor	Work>Performance			[Actor] A appears in [Performance] B
plays	Actor	Object			[Actor] A plays in [Object:musical

⁸ 궁중이나 지방 관아에서 공연했던 악(樂)·가(歌)·무(舞)의 종합예술

					Instrument] B
isPreviousInSequenceTo	Event/Work/Concept	Event/Work/Concept	isNextInSequenceTo		A precedes B (in temporal or logical sequence)
isUsedIn	Object	Event/Concept	uses	Timespan of Event	[Object] A is used In [Event] B
goes With	Object	Object			[Object] A is used with [Object] B
isDepictedIn	Any	Record/Object	depicts		A is depicted in [Record/Object] B
isMentionedIn	Any	Record	mentions		A is mentioned in [Record] B

§ 무신년진찬 정재(畵才) 데이터 관계망



3. 인문지식 데이터 아카이브의 수혜자와 미래 고객

한양도성 타임머신 프로젝트의 수행자들은 이 프로젝트이 결과물의 옛 도시의 물리적인 형상만을 보여주는 것이 아니라, 그 도시에 어떤 사람들이 살았고 그들이 어떤 관계를 유지했으며 어떤 문화를 즐겼는지를 알게 하는 것을 목표를 세웠으며, 이 목표를 이루기 위해 디지털 기술자와 인문학 연구자들의 협업 조직을 만들었고, 문화유산과 관련이 있는 인문지식을 디지털 데이터로 전환하는 노력을 기울였다.

인문지식을 기계가독적인 데이터로 전환하는 노력의 1차적인 수혜자는 이 프로젝트에 참여한 연구자 자신들이다. 수십 년간 한국사와 고전을 연구해온 선임연구원들은 데이터 큐레이션 활동을 통해 학제적인 안목을 넓힐 수 있었고, 과거에 놓쳤던 중요한 사실들을 발견할 수 있었다고 자평한다. 신입 연구원들(대학원 학생)은 데이터 큐레이션 활동을 통해 옛

문헌과 유물의 세계에 몰입하고 많은 것을 배울 수 있었다. 무엇보다 중요한 것은 디지털 네이티브 세대의 젊은 학생들이 자신들에게 익숙한 디지털 언어로 한국의 전통문화를 탐구할 수 있는 새로운 학습방법의 가능성을 확인할 수 있었다는 점이다. 한양도성 데이터 아카이브는 이들 연구자들에게 인문학 연구와 교육의 지속적인 동반자로 기능할 것이다.

한양도성 데이터 아카이브는 작지 않은 공공예산이 투입된 연구사업의 결과물인 만큼 공익적인 기대효과가 있어야 할 것이다. 전통적으로 인문지식의 소통 수단이었던 미디어 ‘글’과는 다른 모습으로 편찬된 이 새로운 콘텐츠의 고객은 누구일까?

데이터를 시각적으로 보여줄 수 있다는 점에 주목하는 이들은 이것이 ‘대중’들의 접근을 용이하게 할 것으로 기대할 수 있다. 그러나 아카이브 속의 데이터가 큰 규모로 축적되면, 그래프의 복잡도는 폭발적으로 증가하게 되고, 사람의 눈으로 그 관계성을 추적하는 일이 어려워지게 된다.⁹ 그러면 시맨틱 아카이브의 고객은 해당 분야의 교수나 연구원 같은 전문가 그룹에 한정되는 것일까? 세분화된 전공영역에서 깊이 파고드는 연구를 하는 이들은 자신의 주의력이 다방면으로 확산되는 것을 원하지 않을 수도 있다.

저작물의 고객을 ‘대중’과 ‘전문가’로 나누는 종래의 이분법적 사고로는 시맨틱 데이터 아카이브의 고객을 이야기하기 힘들다. 오히려, 지식의 소비자가 동시에 새로운 지식의 생산자 되는 디지털 세계 안에는 ‘전문가’의 학술과 ‘대중’의 즐길거리 사이에서 중재적 역할을 하는 매개적 커뮤니케이터들이 다양한 형태로 존재한다는 점에 주목할 필요가 있다. 유튜버, 블로거, 위키백과의 기고자들, 디지털 콘텐츠 제작자들..... 시맨틱 데이터 아카이브를 가장 먼저 반기고 이용할 사람들은 디지털 지식 콘텐츠 시장의 프로슈머들일 것이다.

고등연구기관의 학술활동이 전공이라는 이름으로 극단적으로 세분화되어 있기 때문에 전문 연구자들의 저작물만을 가지고 분야를 넘나드는 횡단적 지식을 탐구하는 것은 쉽지 않다. 만일 그 정도 수준의 심도 있는 지식 데이터가 전공의 벽을 넘어서 꼬리를 물고 이어져서 그 지식 세계의 다채로운 면모에 접근할 수 있는 길을 열어 준다면, 대중이 관심을 갖는 2차적

⁹ 궁극적으로 네트워크 데이터에서 유용한 정보를 찾는 것은 사람의 눈으로 할 일이 아니라 소프트웨어의 도움을 받아서 할 일이다. 네트워크 데이터베이스 상에서 SPARQL (Simple Protocol and RDF (Resource Description Framework) Query Language) 검색 명령을 이용하면, 아카이브 속에 담긴 데이터들의 복잡한 연관관계를 추적해서 의미 있는 결과를 얻을 수 있다.

콘텐츠를 만드는 사람들에 큰 도움이 될 것을 기대할 수 있다. 3D 가상현실을 무대로 하는 콘텐츠 유통 플랫폼-메타버스의 세계에서는 오락물로서의 게임 콘텐츠뿐 아니라, 다양한 수준의 교양, 교육 콘텐츠가 개발•운용될 것이다. 시맨틱 데이터 아카이브는 그러한 오락물과 교양물, 교육 프로그램의 개발이 필요로 하는 스토리라인과 에피소드의 디테일을 제공할 수 있다.

여기서 한 걸음 더 나아가 생각할 것은, 지금은 위에서 언급한 특정인들이 프로슈머의 역할을 수행하지만 디지털 환경에서 태어난 미래 세대들은 누구나 디지털 프로슈머로 자라날 것이라는 점이다. 이들 디지털 원주민들은 복잡하게 얽힌 디지털 클라우드의 세계에서 자신의 호기심이 이끄는대로 다양한 지식의 연결고리들을 추적하고, 거기서 얻은 자료들을 가지고 자신의 이야기를 담은 또 다른 디지털 콘텐츠로 만들어낼 수 있는 능력자들이다.

그리고, 우리가 궁극적으로 기대하는 시맨틱 데이터 아카이브의 타겟 오디언스는 인공지능이다.

2022년말 세상을 놀라게 하고, 반년도 안 돼 더욱 고도화된 모습을 보인 ‘Chat GPT’¹⁰는 가까운 미래에 인문학의 영역에서도 인공지능이 교육과 연구의 동반자가 될 가능성을 보여 주었다. 인공지능은 이미 방대한 양의 인문학 텍스트를 학습하였고, 그 내용과 관련한 인간의 질문에 거침없이 대답한다. 우리는 인공지능의 박식함과 말주변에 놀라면서, 한 편으로 그 대답 속의 모호성과 부정확한 오류에 대해 걱정하고 있다. 인공지능이 신뢰할 수 있는 인문학 공부의 동반자가 될 수 있게 하기 위해서는 인간의 지성에 의해서 검증된 ‘명시적이고 신뢰할 수 있는 데이터’가 기계적 심화학습(Deep Learning)의 내용과 방향성을 제어할 수 있어야 한다. 그것은 특정 분야의 정리된 지식을 집중적으로 교육하는 ‘지도학습’(supervised learning), 또는 ‘지도 학습’보다는 느슨하게 ‘주어진 지식’과 인공지능 스스로 습득한 지적능력을 균형적으로 융합할 수 있게 하는 것을 ‘강화학습’(reinforcement learning)의 방법으로 가능하다.

한양도성 타임머신 시맨틱 데이터 아카이브를 구성하는 RDF(Resource Data Framework) 형식의 데이터는 인간의 언어 속에 모호하게 담겨있던 인문학 지식을 컴퓨터가 인식할 수

¹⁰ <https://openai.com/blog/chatgpt>

있는 명시적인 데이터로 변환한 것이다. 이처럼 특정 분야의 전문가들에 의해서 정밀하게 만들어 데이터를 집중적으로 교육하는 방식으로 인공지능의 ‘전문성’과 ‘정확성’을 제고시킬 수 있다.

오해하지 말아야 할 점은 믿을 수 있는 기초 데이터가 제공되지 않았는데 컴퓨터가 스스로 인간을 가르칠 무엇을 찾아낼 일은 결코 없다는 것이다. 인공지능 기술이 남용되는 곳에서는 모호한 데이터로부터 도출된 잘못된 판단을 마치 새로운 발견인 것처럼 오해하거나 과장하는 일도 발생한다. 이러한 오류를 통제하는 역할도 시맨틱 데이터 아카이브 속의 명시적인 데이터에 기대할 수 있는 일이다.