

데이터 기반 인문학 연구 방법의 모색*

문중 고문서 아카이브와 디지털 인문학의 만남

김 현** · 안승준*** · 류인태****

|| 초록 ||

단순히 인문학 자료를 디지털화 하거나, 연구 결과물을 디지털 형태로 간행하기보다는, 정보 기술의 환경에서 보다 창조적인 인문학 활동을 전개하는 것을 ‘디지털 인문학’이라고 한다. 무엇이 ‘디지털 인문학적 연구’인가 라는 물음에 답을 줄 만한 연구 모델은 아직 명확하게 정립되어 있지 않다. 하지만, 최근 수년 사이 ‘디지털 인문학 연구’로 보고된 여러 형태의 인문 분야 연구에서 가장 두드러진 공통점을 찾는다면, 그것은 종래에 인간의 언어로만 기술되었던 인문 지식의 요소들을 기계가독적(machine readable) 데이터로 전환하여, 컴퓨터가 그 정보의 해석을 도울 수 있게 하는 것이라고 할 수 있다.

필자는 이 글에서, ‘무엇이 디지털 인문학인가’하는 원론적 논의에서 한 걸음 더 나아가 인문 지식을 데이터로 전환하는 것은 어떠한 작업이며, 그것을 통해 우리는 학술적 성과를 기대할 수 있는지를 설명하고자 한다. 이를 위해 한국학중앙연구원 장서각과 인문정보학 교실에서 수행하고 있는 데이터 기반 문중고문서 연구를 사례로 제시하였다.

디지털 인문학의 프레임워크 위에서 만들어진 새로운 형식의 고문서 디지털 아카이브는 연구자들이 직접 그 디지털화된 문서의 내용을 데이터로 취급할 수 있게 한다. 그리고 데이터의 분석, 정리, 해석을 통해 찾아낸 ‘사실’과 ‘이야기’를 확장된 데이터로 축적해 나

* 이 글은 숙명인문학연구소 창립 기념 학술대회(2017. 6. 9.) 발표문 “소통과 융합의 세계, 그곳에 서의 인문학”(김현) 및 제59회 역사학회(2016. 10. 28) 발표문 “고문서 연구를 위한 데이터 기술 모델”(김현, 안승준, 류인태)의 연구를 통합·보완하여 작성한 논문임.

** 한국학중앙연구원 한국학대학원 인문정보학 교수

*** 한국학중앙연구원 장서각 고문서연구실장

**** 한국학중앙연구원 한국학대학원 박사과정 (인문정보학 전공)

2 | 횡단인문학 창간호

아갈 수 있게 함으로써, '자료'의 아카이브에서 '지식'의 아카이브로의 진화를 지향한다.

고문서 내용을 데이터로 전환하여 기계적 가독성을 부여하는 것은 일차적으로 과거의 생활사에 대한 인문학적 연구에 기여하겠지만, 향후 이러한 데이터의 광범위한 축적으로 만들어질 인문학 빅데이터는 현대에 발생하는 사회과학적, 산업적 빅데이터와 연계져서 현대사회에 대한 이해와 과거 역사의 탐구 사이의 연결고리를 만들어내는 것도 가능하게 할 것이다. 인문지식이 데이터로 소통될 수 있도록 하는 노력을 통해 우리의 인문지식은 전통적인 인문학의 경계를 넘어서는 세계와도 교섭할 것이며, 인간과 사회에 대한 이해로서 인문학의 외연을 더욱 의미 있게 확장해 갈 수 있을 것이다.

주제어: 디지털 인문학, 문중고문서, 온톨로지, 디지털 스토리텔링, 디지털 아카이브, 디지털 큐레이션, 인문학 빅데이터

1. 4차 산업혁명 시대, 인문학의 과제

“3차 산업혁명 시대에 컴퓨터는 ‘프로세스’의 효율화를 위한 도구로 쓰였다. 인력과 시간의 낭비를 줄이고 고객의 편의성을 증대시키기 위해 더 많은 일에 컴퓨터를 도입해 온 것이다. 그런데, 바로 그곳에 우리가 예전에 전혀 기대하지 않았던 부산물이 남는 것을 알게 되었다. 전산 시스템에 남은 프로세스의 흔적, 바로 ‘데이터’(Data)이다.

3차 산업혁명 시대의 컴퓨터가 프로세스 효율화의 도구였다고 한다면, 4차 산업혁명 시대의 컴퓨터는 데이터의 해석과 활용의 도구이다. 데이터의 해석을 집적하여 자동적으로 유효한 수준의 의사결정이 이루어질 수 있게 한 것을 ‘인공지능’이라 하고, 그 지능을 가지고 인간을 보조할 수 있게 한 기계 장치를 ‘로봇’이라 한다.

인간들의 삶이 컴퓨터에 의존함으로써 얻게 된 방대한 데이터, 이른바 빅 데이터(Big Data)라고 하는 것은 바로 ‘인간들의 다양한 삶의 자취’이다. 빅 데이터의 구성 요소는 ‘개체’와 그 개체들 사이의 ‘관계’이다. 예전에는 인간들이 삶 속에서 만들어내는 수많은 관계(사람과 사람 사이의 관계, 사람과 사물 사이의 관계, 또 인간 주변에서 만 들어지는 사물과 사물 사이의 관계 등)를 일일이 포착하기가 어려웠지만, 그 삶의 많은 부분이 컴퓨터에 의존함으로써 그 실상을 들여다볼 수 있는 데이터가 만들어지게 되었고, 그것이 의미 있는 ‘빅 데이터’로 간주되게 된 것이다.”¹⁾

디지털 환경에서 데이터의 소통이 증대되고, 이로 인해 과거에는 다른 영역인 줄 알았던 분야의 일들이 하나로 묶여서 놀라운 시너지 효과를 낸다고 한다. 이른바 '4차 산업 혁명 시대의 사회.' 그곳에서 인문학이 새롭게 관심을 갖고 추구해야 할 일과 그 방법은 무엇인지에 대해 생각할 필요가 있다. 데이터의 오남용을 경계하고 반성하는 사고는 '데이터의 시대'에 꼭 필요한 것이지만, 우리의 보다 긴급한 과제는 가치 있는 인문학적 지식이 디지털로 소통하는 이 사회에서 의미 있게 쓰일 수 있는 길을 찾는 일이다.

전통적으로 인문학적 연구는 '글'로 쓰인 자료를 조사하여 읽고, 그 내용을 해석하고 분석한 결과를 새로운 '글'로 표현하는 형태로 이루어져 왔다. 그러한 배경에서 인문학 연구자들이 일반적으로 생각하는 '인문학 연구'와 '디지털'의 관계는 연구에 도움이 되는 기초 자료를 디지털 매체에 저장하여 접근성을 향상시키고, 검색을 용이하게 하는 것, 그리고 연구 결과로 생산된 새로운 글을 디지털 문서로 간행하여 그 보급을 용이하게 하는 것이었다. 이러한 일들은 현재에도 '인문학 자료의 전산화'라는 이름으로 활발하게 진행되고 있다고 할 수 있다. 하지만, 이와 같은 '전산화'의 결과물은 비록 그것이 디지털 매체에 디지털 정보로 기록되었다고 해도, 그것의 내용을 탐구하는 '해석'과 '분석'은 어디까지나 독자의 사고를 통해서만 이루어질 수 있는 일이고, 컴퓨터 시스템의 도움을 받아 새로운 사실을 발견하거나 지식을 증진시키는 효과를 낳는 것은 아니었다.

단순히 인문학 자료를 디지털화 하거나, 연구 결과물을 디지털 형태로 간행하기보다는, 정보 기술의 환경에서 보다 창조적인 인문학 활동을 전개하는 것을 '디지털 인문학'이라고 한다.²⁾ 디지털 인문학은 아직 그 역사가 짧고, 그에 분야에 종사하는 연구자도 많지 않을 뿐 아니라, 무엇이 '디지털 인문학적 연구'인가 라는

1) 김현, 「4차 산업혁명 시대의 인문학」, 『전통문화』 42호, 전통문화연구회, 2017. 5.
<http://xuanflute.com/nayaboard/board/view.php?boardname=diary&no=112&page=1>

2) 디지털 인문학은 디지털 환경(digital environment)에서 수행하는 인문학 연구와 교육, 그리고 그 연구와 교육의 성과를 우리 사회에 확산시키는 활용 노력을 포함하는 말이다. 2008년 미국의 인문학재단(National Endowment for the Humanities: NEH)이 디지털 인문학 지원단(Office of Digital Humanities)을 설치하고 각 대학의 디지털 인문학 연구에 대한 지원을 강화하면서, 디지털 인문학은 미국뿐 아니라 세계 대학의 새로운 관심사로 부상하고 있다.

물음에 답을 줄 만한 연구 모델도 명확하게 정립되어 있지 않다고 할 수 있다. 하지만, 최근 수년 사이 '디지털 인문학 연구'로 보고된 여러 형태의 인문 분야 연구에서 가장 기초적인 공통점을 찾는다면, 그것은 종래에 인간의 언어로만 기술되었던 인문 지식의 요소들을 기계가독적(machine readable) 데이터로 전환하여, 컴퓨터가 그 정보의 해석을 도울 수 있게 하는 것이라고 할 수 있다.

필자는 이 글에서, '무엇이 디지털 인문학인가'하는 원론적 논의에서 한 걸음 더 나아가 인문 지식을 데이터로 전환하는 것은 어떠한 작업이며, 그것을 통해 우리는 학술적 성과를 기대할 수 있는지를 설명하고자 한다. 이를 위한 예시로서 보이고자 하는 연구 사례는 한국학중앙연구원 장서각에서 소장하고 있는 문중고문서에 관한 디지털 인문학 연구이다.

2. 고문서 디지털 아카이브

한국학중앙연구원은 1981년부터 부속 기구인 장서각을 중심으로 우리나라 여러 문중에서 소장해 온 고문서를 수집하고 정리, 보존하는 사업을 수행해 오고 있다. 그 과정에서, 2011년에는 30년간의 고문서 수집 사업의 성과로 집적된 문중 고문서 5만여 건의 고해상도 이미지를 수록한 데이터베이스를 제작하였다. '한국고문서자료관'이라고 명명한 이 디지털 아카이브는 서비스 자료의 활용도를 높이기 위해 고문헌 용어 용례 사전, 고지명 용례 사전, 이두 용례 사전 등 고문서를 읽을 때 참조할 수 있는 용어 해설 데이터와 각 문중 주요 인물들의 가계를 담은 종합보인 만가보(萬家譜)의 인물 데이터를 함께 제공하고 있다.

한국 고문서 자료관 디지털 아카이브의 구축으로 인해 고문서 연구자들은 그들의 연구 자원에 쉽게 다가갈 수 있게 되었다. 그렇지만 여기까지의 일은 디지털 기술로 고문서 연구의 환경을 개선한 것이지, 디지털 환경에서 고문서상의 지식

을 찾고 그것을 연구한 것은 아니다. 한국 고문서 자료관 소장 자료를 온라인 상에서 서비스할 수 있도록 하기까지의 일이 이른 바 '인문학 자료의 디지털화'라면, 그 이후에 디지털 기술을 통해 탐구해야 할 과제는 고문서의 내용을 데이터로 삼는 '디지털 인문학' 연구이다.³⁾

'고문서 자료의 디지털화' 수준을 넘어서서 디지털적인 방법으로 고문서의 내용을 분석할 수 있는 연구 환경을 제공하고, 그 연구를 통해서 얻어낸 결과를 또 다시 데이터로 서비스할 수 있는 새로운 형식의 고문서 디지털 아카이브는 디지털 기술의 발전만으로 기약할 수 있는 것이 아니다. 고문서를 자료로 삼는 연구를 수행할 때, 연구자들 스스로 디지털적인 데이터 분석 및 서술 방법을 적극적으로 활용하고, 그렇게 함으로써 그 연구의 성과들이 곧바로 고문서 아카이브의 새로운 데이터로 축적될 수 있게 하는 노력이 필요하다.

이러한 이유에서 한국학중앙연구원의 고문서연구실과 인문정보학교실의 연구자들은 고문서 연구를 위한 디지털 환경, 즉 고문서 연구자들이 직접 데이터의 생성과 분석, 시각화를 수행할 수 있게 하는 고문서 아카이브 구현 방향을 모색하고 있다.⁴⁾ 이는 '고문서'라는 실물보다 그 안에 기술된 고문서의 내용을 다룰 수 있는 방법을 강구하고자 하는 것이며, 디지털 환경에서 디지털 기술로, 고문서에 적합한 글의 내용을 분석하고, 새로운 의미를 찾아 그것을 재구성하는 방법을 찾고자

3) 김현, 「디지털 인문학과 고문헌 자료 연구」, 『열상고전연구』 제50집 (2016. 4. 열상고전연구회) p. 30.

4) 한국학 분야의 고문헌 자료에 대한 시맨틱 웹 기반 데이터 모델 설계 연구는 한국학중앙연구원 한국학대학원 인문정보학교실과 장서각 고문서연구실 연구자들에 의해 2013년부터 지속적으로 탐구되어 왔으며, 다음과 같은 몇 차례의 데이터베이스 편찬 및 데이터베이스 설계 연구에 적용함으로써 그 내용을 검증하고 보완하여 왔다.

* 한글 박물관 디지털 아카이브 데이터베이스 모델 설계: '국립한글박물관 디지털 아카이브 구축 기본 구상', 2013. 12. 한글박물관 (연구책임자 김현)

* 선원속보 시맨틱 데이터베이스 모델 설계: '규장각 장서각 연계 기록유산 DB 구축 사업', 2014. 12. 장서각 (연구책임자 안승준)

* 광산김씨 문중 고문서 시맨틱 데이터베이스 모델 설계: '門中古文書 디지털 아카이브 구현 연구', 2015. 8. (김하영, 한국학대학원 석사학위 청구논문)

* 부안김씨 문중 고문서 시맨틱 데이터베이스 모델 설계: '관계의 발견을 위한 디지털 스토리텔링 데이터 모델', 2015. 11. 장서각 (연구책임자 김현)

* 부안김씨 문중 고문서 시맨틱 데이터베이스 모델 설계: '조선시대 고문서 시맨틱 웹 DB 설계 기초 연구', 2016. 10. (류인태, 2016 세계한국학대회 발표논문)

하는 것이다.

3. 문중 고문서 온톨로지 설계

1) 분석 대상 자료: 부안 김씨 문중 고문서

고문서 아카이브에서 취급하는 정보가 고문서 실물의 서지적 데이터에 머물지 않고, 그 문서에 쓰인 내용, 나아가 그것으로부터 유추할 수 있는 사실과 이야기를 포함할 수 있기 위해서는 데이터 기술 형식이 그 목적에 부합하도록 설계되어야 한다. 이러한 취지의 데이터 기술 모델을 찾기 위해 분석 대상으로 삼은 다양한 고문서 자료 중의 일부가 부안 김씨 문중 고문서이다.

부안 김씨 고문서는 ‘분재기(分財記)’, ‘명문(明文)’, ‘소지(所志)’, ‘입안(立案)’, ‘고신(告身)’ 등 다양한 유형의 문서를 포함하고 있는데, 각각의 문서들이 해당 형식의 고문서가 지니는 특징을 잘 보여주는 대표적 사례로 평가받고 있다.⁵⁾ 2016년 현재 한국학중앙연구원 장서각(藏書閣)의 고문서 자료관 아카이브에 등재되어 있는 부안 김씨 문중 고문서는 740여 건이다. 그 가운데 7종 17편의 고문서를 본 연구의 1차적인 분석 대상 자료로 삼았다. 이 17편의 고문서는 부안 김씨 여러 인물들 중 조선시대 중·후기에 살았던 ‘김홍원-김명열-김번·김문-김수종’ 4대가 남긴 문서로, 여러 대에 걸친 생활사의 연속과 변화를 살필 수 있는 자료들이다. 아울러 그 생활사의 주역으로 살아간 문중 인물들의 혈연관계를 보다 구체적으로 알려주

5) 부안 김씨 문중 고문서가 지닌 자료로서의 가치에 대해서는 다음의 서술을 참고할 수 있다. “우반동 김씨란 우반동에 세거(世居)하는 부안 김씨라는 뜻으로 이 책에서 소개하는 우반동의 주인공들이 바로 이들이다. 이 우반동 김씨 종가(宗家)에는 현재 수많은 고문서(古文書)가 소장되어 있는데 1970년대 초에 정구복 교수에 의해 처음으로 발굴된 후 그 학술적 가치가 널리 인정되어 한 국정신문화연구원에서 두 차례에 걸쳐 자료집으로 간행한 바 있다. 그리고 이 문서의 문화재적 가치 또한 크게 인정받아 1987년에 보물 900호로 지정되기에 이르렀다.” -전경목, 『고문서를 통해 서 본 우반동과 우반동 김씨의 역사』, 신아출판사, 2001. p.3

는 부안 김씨 족보를 분석 대상 자료로 추가하였다. 본 연구에서 사용한 자료는 ‘김홍원 후대의 정보를 가장 세밀하게 기록하고 있는 ‘지평공파보(持平公派譜)’이다.

[표 1] 분석 대상 자료: 부안 김씨 문중 고문서 7종 17편과 부안 김씨 족보

구분	자료 유형	대상자료
01	고신	1660년 김명열의 고신
02		1664년 김명열의 고신
03		1662년 김번의 고신
04		1666년 김문의 고신
05	호적	1675년 김번 가족의 준호구
06		1855년 김용관 가족의 호구단자
07	소지	1837년 고생원택 노 분옥의 소지
08	입안	1675년 예조의 김수종 계후입안
09	매매명문	1677년 김번 토지매매명문
10		1837년 고생원택 노 부록 토지매매명문
11		1838년 고우진 토지매매명문
12		1840년 최재영 토지매매명문
13		1854년 고진사택 노 재록 토지매매명문
14	배지	1837년 고생원택 노 부록 배지
15	분재기	1672년 김명열 별급문서
16		1677년 김번 5남매 화회문기
17		1688년 김번 분재문기
18	족보	우반동 부안 김씨 지평공파보

2) 문중 고문서 온톨로지

온톨로지(Ontology)란 정보화의 대상이 되는 세계를 전자적으로 표현할 수 있도록 구성한 데이터 기술 체계이다. 원래 온톨로지라는 말은 철학에서 ‘존재론’이라고 번역되는 용어로서 ‘존재에 대한 이해를 추구하는 학문’의 의미를 갖는 말이었다. 그러한 용어가 정보과학 분야에서 중요한 개념으로 등장하게 된 것은 인간이 세계를 이해하는 틀과 컴퓨터가 정보화 대상(콘텐츠)을 이해하는 틀 사이에 유사성이 있다고 보았기 때문이다. 넓은 의미에서는 모든 정보화의 틀이 다 온톨로지일 수 있겠지만, 대상 자원을 ‘클래스’(class)로 범주화하고, 각각의 클래스에 속

8 | 횡단인문학 창간호

하는 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.⁶⁾

7종 17편의 부안 김씨 고문서와 지평공파보(持平公派譜)에 담긴 내용을 디지털 데이터로 기술하기 위해서는 그 데이터 요소와 그 요소들 사이의 관계를 담아 낼 수 있는 그릇을 먼저 설계해야 한다. 이 같은 의미의 데이터 기술 설계를 웹 온톨로지 언어(OWL, Web Ontology Language)⁷⁾의 어휘와 문법을 준용하여 구현하였다.

[표 2] 문중 고문서 온톨로지 설계 프레임워크⁸⁾

온톨로지 구성 요소 (관장 용어)	용도 ⁹⁾	Web Ontology Language (OWL)
Class 클래스	a group of individuals that belong together because they share some properties. 공통의 속성을 가진 개체들을 묶는 범주	owl: Class
Individual 개체	Instances of classes 클래스에 속하는 개체	owl: NamedIndividual
Relation 관계	relations between pairs of individuals (같거나 다른 클래스에 속하는) 개체 사이 관계	owl: ObjectProperty
Attribute 속성	relations between individuals and data values 개체가 속성으로 갖는 데이터 값	owl: DatatypeProperty
Relation Attribute 관계 속성	attributes related to relations 관계 정보에 부수되는 속성	N/A in OWL
Domain 영역	A domain of a property which limits the individuals to which the property can be applied 특정 관계 또는 속성이 적용될 수 있는 클래스를 한정	rdfs: domain
Range 범위	The range of a property limits the individuals that the property may have as its value 특정 관계 또는 속성이 Data 값으로 삼을 수 있는 클래스를 한정	rdfs: range

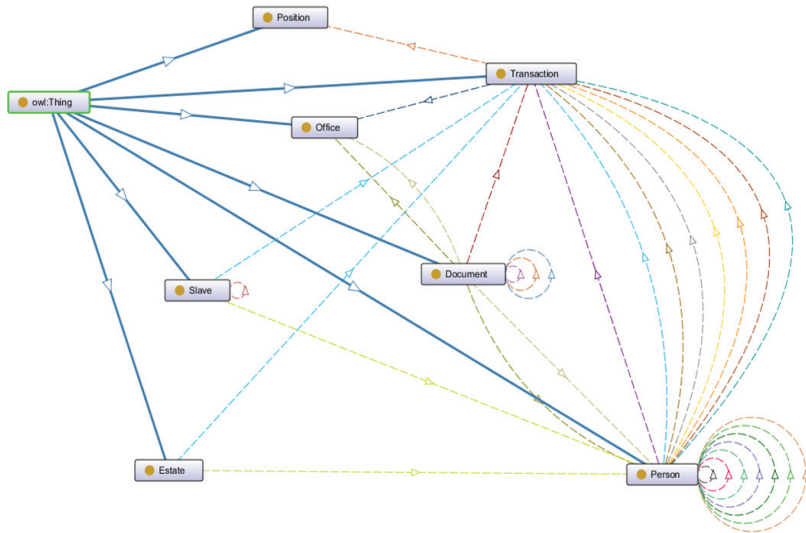
6) 김현, 『디지털 인문학 입문』 (2016, HUEBOOKS) p. 164.

7) W3C(World Wide Web Consortium)가 제안한 온톨로지 기술 언어. 현재 통용되는 권고안은 2004년에 발표되었다.

3) 클래스(Class) 설계

[표 3] 문중 고문서 데이터 클래스

Class Name	국문 클래스 명칭	설명
Document	문서	고문서
Transaction	내용	고문서에 기록된 행위(상속, 매매, 차용 등)의 대상이 되는 목적물 또는 고문서를 통해 처리된 사안의 구체적 내용
Person	인물	고문서의 발급·수급 및 고문서에 기록된 행위의 주체·객체로서의 인물 또는 내용 개체를 이루는 개별 요소로서의 인물
Office	관부	고문서의 발급자·수급자로서의 관부 또는 내용 개체를 이루는 개별 요소로서의 관부
Estate	토지	내용 개체를 이루는 개별 요소로서의 토지
Slave	노비	내용 개체를 이루는 개별 요소로서의 노비
Position	관직	내용 개체를 이루는 개별 요소로서의 관직



[그림 1] 문중 고문서 데이터 클래스 다이어그램

9) OWL Web Ontology Language Overview, W3C Recommendation

(<http://www.w3.org/TR/owl-features/>)

10) http://digerati.aks.ac.kr:82/index.php?title=인문정보학_온톨로지_설계_가이드라인

(1) 문서(Document) 클래스

‘문서(Document) 클래스는 아카이빙(기록물 관리)의 대상이 되는 문서 그 자체에 대한 정보를 담기 위한 틀이다. 본 연구에서는 부안 김씨 문중 고문서 중, ‘토지·노비 상속’과 관련된 분재기, ‘토지 매매’와 관련된 토지매매명문, ‘관직 임명’과 관련된 고신, ‘호적 조사와 관련된 호구단자 및 준호구, ‘행정 청원’과 관련된 소지, ‘행정 증명’과 관련된 입안 등 7개 분야의 문서를 대상으로 삼아 고문서 온톨로지 설계를 수행하였다. 그리고 문서 유형에 따라 아래와 같은 5종의 하위 클래스를 설계하였다.

- R0: 묶음 문서 → 점련문서 또는 가상의 묶음 문서
- R1: 실행 문서 → 재산 상속(분재기), 토지 매수 및 매도(토지매매명문), 매매명문 작성의 권한 위임(배지), 현금 및 현물 차용(차용증빙수기)과 같이 구체적 실행 내용을 담고 있는 문서
- R2: 공증 문서 → 재산 상속 또는 토지매매와 같은 구체적 실행에 대한 공증을 위한 문서(소지, 입안, 조목, 초사류 등)
- R3: 호구 문서 → 호구 조사 내용 및 호적에 기록된 내용을 공적으로 증명하는 문서(호구단자, 준호구)
- R4: 인사 문서 → 임명 사실을 증명하는 문서(고신, 교지 등)

(2) 내용(Transaction) 클래스

본 연구에서 설계한 온톨로지의 특별한 측면은 기록물 관리 대상으로서의 ‘문서(Document)와 해당 문서에 담긴 구체적 분석 대상으로서의 ‘내용(Transaction)을 별도의 클래스로 나누어 범주화하였다는 점이다. ‘문서’(Document) 클래스가 문서 그 자체가 지닌 형식적 정보를 정리하기 위한 틀이라면, ‘내용’(Transaction) 클래스는 문서가 담고 있는 구체적 내용을 정리하기 위한 틀이라고 할 수 있다.

재산의 상속 사실을 기록한 분재기의 경우 ‘상속 행위와 상속된 재산’, 토지의 매매 사실을 기록한 토지매매명문의 경우에는 ‘매매 행위와 매매된 토지’, 호구 조사 및 호구 내용을 증명하는 호구단자와 준호구의 경우에는 ‘호구 조사와 호구 내용’, 관직 임명 사실을 증명하는 고신의 경우에는 ‘임명 사건과 임명된 관직’ 등이 본 ‘내용’(Transaction) 클래스의 개체가 된다. 이처럼 ‘문서’(Document) 클래스에 따라 각기 다른 성격의 ‘내용’(Transaction)이 기술될 수 있으므로 아래와 같은 7개의 하위 클래스를 설계하였다.

- B1: 노비 분재
- B2: 토지 분재
- B3: 노비 매매
- B4: 토지 매매
- B5: 호구 가족
- B6: 호구 노비
- B7: 관직 임명

(3) 인물(Person) 클래스

‘인물’(Person) 클래스는 고문서와 직·간접적으로 관련된 사람들을 개체로 갖는다. 이 때 노비는 재산으로서의 성격을 강조하여 독립적인 ‘노비’(Slave) 클래스로 설정되기 때문에, ‘인물’(Person) 클래스에 포함되지 않는다. ‘인물’(Person) 클래스는 노비를 제외하고, 고문서의 수취자와 발급자 및 고문서에 기록된 다양한 행위 주체 및 객체가 되는 인물들을 정보화하고, 해당 인물들 사이의 관계를 담아 내기 위한 틀이라 할 수 있다. 대상 자료가 문중 고문서임을 고려할 때 ‘인물’(Person) 클래스의 중심이 되는 인물들은 대부분 한 가계(家系)의 구성원으로서, 해당 가문의 족보로부터 구체적 관계 정보를 추출할 수 있다. 이 점을 고려하여 다음과 같이 11개 하위 클래스(Sub Class)를 설계하였다.¹¹⁾

- N1: 남자 자손 (동성)
- N2: 남자 자손 부인 (이성)
- N3: 남자 자손 부인의 아버지와 그의 선대 및 방계 혈족 (이성)
- N4: 여자 자손 (동성)
- N5: 여자 자손 남편 (이성)
- N6: 여자 자손 남편의 아버지와 그의 선대 및 방계 남자 (이성)
- N7: 여자 자손과 남편 사이에서 태어난 아들과 그 직계 후손 (이성)
- N8: 여자 자손과 남편 사이에서 태어난 아들의 부인 (이성)
- N9: 여자 자손과 남편 사이에서 태어난 딸 (이성)
- NR: 친척. 가계 상의 위치를 알 수 없으나 문서에 친척으로 기록된 인물
- NA: 지인. 문서상 친척 관계를 확인할 수 없는 인물

(4) 관부(Office) 클래스

‘관부’(Office) 클래스는 문서를 발급하거나 수급하는 주체로서의 행정기관을 정의하기 위한 틀이다. 현재는 ‘중앙 관부’와 ‘지방 관부’ 등 2 개의 하위 클래스를 두었으나, 향후 고문서 간 관계 연구를 통해 관청 사이의 유의미한 변별점이 새롭게 발견될 경우 새로운 하위 클래스의 설정이 가능할 것이다.

- G1: 지방 관부
- G2: 중앙 관부

11) 가계도 구현을 위한 문중 인물의 데이터 모델은 조선왕실 족보(『선원보』 및 『선원속보』) 인물 망 데이터베이스를 위해 설계한 ‘왕실 족보 온톨로지’로 처음 설계되었다. 이후 평산 김씨 고문서를 대상으로 한 연구(김하영, 『門中古文書 디지털 아카이브 구현 연구』, 한국학중앙연구원 한국학대학원 석사학위 청구논문, 2015.)를 통해 외손의 부인과 외손녀 등 하위 클래스를 추가하고, 그에 근거하여 hasFosterDaughter(수양녀, 시양녀) 관계어를 추가하는 등 문중 가계를 담아낼 수 있는 틀로 점차 확장하였다. 본 연구 또한 그와 같은 기존 연구를 통해 확장되어온 문중 인물 데이터 모델을 기본적으로 적용하였다.

(5) 토지(Estate) 클래스

‘토지(Estate) 클래스는 하나의 독립적 실체로 식별 가능한 토지나 가옥 즉, 부동산 성격의 대상을 개체로 갖는다. 토지와 직접적으로 관련된 문서로는 ‘분재기’, ‘토지매매명문’ 등이 있다.

개체로서의 토지는 ‘분재기’의 내용 항목에 해당하는 ‘토지 분재’나 ‘토지매매명문’의 내용 항목에 해당하는 ‘토지 매매’와 직접적 관련이 있기 때문에, ‘내용’(Transaction) 클래스 개체의 부분(isUnitOf)으로 관계어를 기술할 수 있다.

고문서에서 주로 언급되는 토지는 크게 답(畓)과 전(田)으로 나눌 수 있으며, 이외에 추가적으로 거론되는 부동산 성격의 대상으로 가옥이 있기 때문에, ‘답(畓)’, ‘전(田)’, ‘실(室)’ 3개의 하위 클래스를 설계하였다.

- L1: 답(畓)
- L2: 전(田)
- L3: 실(室)

(6) 노비(Slave) 클래스

‘노비(Slave) 클래스는 하나하나의 노비를 개체로 갖는다. 노비와 직접적으로 관련된 문서로는 ‘분재기’, ‘노비매매명문’, ‘토지매매명문’, ‘배지’ 등을 들 수 있다.

개체로서의 노비는 ‘분재기’의 내용 항목에 해당하는 ‘노비 분재’나 ‘노비매매명문’의 내용 항목에 해당하는 ‘노비 매매’와 직접적 관련이 있으며, ‘내용’(Transaction) 클래스 개체의 부분(isUnitOf)으로 관계어를 기술할 수 있다. 또한 노비 가운데 일부는 ‘토지매매명문’이나 ‘배지’와 관련하여 소유주를 대신하여 문서의 수발급자 역할을 한 경우가 많이 발견되기 때문에, 그와 관련하여 문서의 수·발급자를 정의하는 관계어(issuedBy, issuedTo)와 함께 기술할 수도 있다. 노비의 하위 클래스는 ‘성별 확인이 안된 노비’, ‘노(奴)’, ‘비(婢)’ 등 3 가지로 설계하였다.

- S0: 성별 확인이 불가능한 노비
- S1: 노(奴)
- S2: 비(婢)

(7) 관직(Position) 클래스

관직이 중요한 데이터로 취급되는 문서로는 ‘고신’, ‘유지’, ‘교지’, ‘입안’ 등을 들 수 있다. 개체로서의 관직은 ‘고신’ 또는 ‘유지’의 내용 항목에 해당하는 ‘관직 임명’과 직접적 관련이 있으며, ‘내용’(Transaction) 클래스의 객체에 대해 ‘직위’(appointsAs) 및 ‘임지’(placesTo)로 관계를 기술할 수 있다.

- P1: 관직

4) 관계(Relation, ObjectProperty) 설계

앞서 정리한 7개 클래스의 개체들은 동일한 클래스, 또는 다른 클래스의 개체들과 여러 맥락에서 유의미한 연관 관계를 맺을 수 있다. 예컨대 ‘문서’ 클래스에 속하는 특정 고문서 개체는 ‘인물’ 클래스에 속하는 특정 인물이 청원하여 ‘관부’ 클래스에 속하는 특정 관부에서 발급된 것이며, 동시에 ‘노비’ 클래스에 속하는 특정 노비와 ‘토지’ 클래스에 속하는 특정 토지를 처분하는 것과 관련된 내용을 담고 있는 것일 수 있다. 이러한 형태로 7개 클래스에 속하는 많은 개체들 사이의 유의미한 관계를 명시적으로 기술할 경우, 그로부터 학술적으로 유의미한 정보나 지식을 도출해낼 수 있다. 아래의 모든 관계성은 A에서 B로 향하는 방향성을 갖는다.

[표 4] 문중 고문서 데이터 관계자(Relation, Object Property)

Relation	Description	Domain	Range
A hasPart B (B isPartOf A)	문서 A는 문서 B를 부분으로 삼는다. ※ 접편문서 또는 가상적인 묶음 문서(R0)와 그것에 속하는 개별 문서(R1, R2, R3, R4) 사이의 관계를 표현한다.	문서 (Document)	문서 (Document)
A hasEvidence B (B isEvidenceOf A)	문서 A는 문서 B를 증빙 서류(Evidential Document)로 하는 관계가 있다. ※ 실행 문서(R1)와 그것에 부수하여 발생한 공증 문서(R2) 사이의 관계를 표현한다.	문서 (Document)	문서 (Document)
A isRelatedTo B	A hasRelation B : 문서 A는 문서 B와 관계가 있다. ※ 관련 사건의 인과 관계 등 문서 사이에 의미상 유관한 관계가 있는 경우 이를 표시하기 위한 관계어이다.	문서 (Document)	문서 (Document)
A issuedBy B (B isIssureOf A)	문서 A는 인물 또는 관부 또는 노비 B에 의해 발급되었다.	문서 (Document)	인물(Person), 노비(Slave), 관부(Office)
A issuedTo B (B isRecipientOf A)	문서 A는 인물 또는 관부 또는 노비 B에게 발급되었다.	문서 (Document)	인물(Person), 노비(Slave), 관부(Office)
A isProofOf B (B isProvedBy A)	문서 A는 내용 B를 증명하는 기록이다.	문서 (Document)	내용 (Transaction)
A isUnitOf B (B hasUnit A)	토지 또는 노비 A는 내용 B에 속하는 개별 요소이다.	토지(Estate), 노비(Slave)	내용 (Transaction)
A ownedBy B (B owns A)	토지 또는 노비 A는 인물 B의 소유이다.	토지(Estate), 노비(Slave)	내용 (Transaction)
A isOffspringOf B (B hasOffspring A)	노비 A는 노비 B의 자식이다.	노비 (Slave)	노비 (Slave)
A hasSon B (B isSonOf A)	인물 A는 인물 B를 아들로 두었다.	인물 (Person)	인물 (Person)
A hasWife B (B isWifeOf A)	인물 A는 인물 B를 아내로 두었다.	인물 (Person)	인물 (Person)
A hasAdoptedHeir B (B isAdoptedHeirOf A)	인물 A는 인물 B를 계후자(양자)로 두었다.	인물 (Person)	인물 (Person)
A hasDaughter B (B isDaughterOf A)	인물 A는 인물 B를 딸로 두었다.	인물 (Person)	인물 (Person)

A hasFosterDaughter B (B isFosterDaughterOf A)	인물 A는 인물 B를 수양녀(시양녀)로 두었다.	인물 (Person)	인물 (Person)
A hasRelative B	인물 A는 인물 B에 대해 친척이다.	인물 (Person)	인물 (Person)
A knows B	인물 A는 인물 B에 대해 지인이다.	인물 (Person)	인물 (Person)
A isGiverOf B (B hasGiver A)	인물 A는 노비 분재 또는 토지 분재 B의 증여자이다. ※ 개별 인물과 분재기에서 드러나는 내용 요소 사이의 관계를 표현한다.	인물 (Person)	내용 (Transaction)
A isInheritorOf B (B hasInheritor A)	인물 A는 노비 분재 또는 토지 분재 B의 상속자이다.	인물 (Person)	내용 (Transaction)
A isSellerOf B (B hasSeller A)	인물 A는 노비 매매 또는 토지 매매 B의 판매자이다. ※ 개별 인물과 노비 또는 토지 매매명문에서 드러나는 내용 요소 사이의 관계를 표현한다.	인물 (Person)	내용 (Transaction)
A isBuyerOf B (B hasBuyer A)	인물 A는 노비 매매 또는 토지 매매 B의 구매자이다.	인물 (Person)	내용 (Transaction)
A isUnitOf B (B hasUnitA)	인물 A는 호구 B의 구성원이다. ※ 개별 인물과 준호구 또는 호구단자에서 드러나는 내용 요소 사이의 관계를 표현한다.	인물 (Person)	내용 (Transaction)
A placesIn B	임명 내용 A는 피임자를 관부 B에 임용한다(는 내용을 담고 있다). ※ 내용 클래스 가운데서도 관직 임명 요소가 어떠한 특정 관부와 관련이 있는지 그 관계를 표현한다.	내용 (Transaction)	관부 (Office)
A appointsAs B	임명 내용 A는 피임자를 관직 B에 임용한다(는 내용을 담고 있다). ※ 내용 클래스 가운데서도 관직 임명 요소가 어떠한 특정 관직과 관련이 있는지 그 관계를 표현한다.	내용 (Transaction)	관직 (Position)

5) 개체 속성(Attribute, DatatypeProperty) 설계

‘개체 속성’은 특정 클래스(Class)에 속하는 단일 ‘개체(Individual)’가 지닌 특성을 데이터 값으로 표현하는 요소이다. 7개 클래스에 속하는 각각의 개체들이 지니는 속성을 아래의 형식과 같이 정의하였다.

(1) ‘문서’(Document) 클래스 개체 속성

- id	식별자
- name	명칭
- description	설명
- doc_type	문서 유형 ※ 분급문기, 별급문기, 화회문기, 소지, 입안, 명문 등
- issued_year	발급년
- issued_by	발급자
- issued_to	수급자
- bound	점련 여부
- doc_owner	문서 소장처
- URL	Web 자원 주소

(2) ‘내용’(Transaction) 클래스 개체 속성

[노비 분재 또는 토지 분재]

- id	식별자
- name	명칭
- owner	증여자
- distributee	상속자
- trans_year	상속 연도

[노비 매매 또는 토지 매매]

- id	식별자
- name	명칭
- seller	매도자
- buyer	매수자

- proxy_seller	매도대리인
- proxy_buyer	매수대리인
- purpose	매도사유
- price	대가
- deal_year	매매 연도

[호구 가족 또는 호구 노비]

- id	식별자
- name	명칭
- reporter	보고자(증명자)
- receiver	피보고자(수급자)
- report_year	보고 또는 증명 연도

[관직 임명]

- id	식별자
- name	명칭
- appointee	피임명자
- appointed_date	피임명일

(3) '인물'(Person) 클래스 개체 속성

- id	식별자
- name	한글 이름
- chi	한자 이름
- midname	자(字)
- nickname	호(號)
- alias	이칭

※ 족보상의 이름과 고문서 상의 이름이 다를 경우

- birth	생년
- birth_sexagerary	생년(간지)
- death	졸년
- death_sexagerary	졸년(간지)
- clan	본관
- grave	묘지
- shrine	제향지
- posts	관력
- conferment	시호

(4) ‘관부’(Office) 클래스 개체 속성

- id	식별자
- name	한글 명칭
- chi	한자 명칭
- district_type	관할 유형(중앙/지방)

(5) ‘토지’(Estate) 클래스 개체 속성

- id	식별자
- name	명칭
- location	위치(현대주소)
- latitude	위도
- longitude	경도
- code	자호 ※ 양전 시 붙이는 전답의 구분 번호(千字文 순서)
- address	지번
- scale	크기

- section 배미
- extent 두락수(마지기)
- norm 결부수(복/속)

(6) ‘노비’(Slave) 클래스 개체 속성

- id 식별자
- name 한글 이름
- chi 한자 이름
- birth 생년
- sexagenary 생년(간지)
- parent 부모

(7) ‘관직’(Position) 클래스 개체 속성

- id 식별자
- name 한글 명칭
- chi 한자 명칭
- pos_type 관직유형(문반/무반)
- rank 품계

6) 관계 속성(Relation Attribute) 설계

‘관계 속성’은 ‘개체(Individual)’가 아니라, 개체와 개체 사이의 ‘관계(Relation)’에 부여된 ‘속성(Attribute)’이다. 개체의 고유한 속성은 아니면서, 한 개체가 다른 개체와 관계를 맺을 때 발생하는 특수한 정보가 있을 수 있는데, 지식의 관계 구조에서는 이러한 정보가 중요한 의미를 가질 수 있기 때문에 관계 속성을 다룰 수

있는 장치를 마련하였다.¹²⁾

(1) ‘문서’(Document) 클래스 개체 사이의 관계 속성

- A hasPart B : bound 접련 여부

(2) ‘노비’(Slave) 클래스 개체 사이의 관계 속성

- A isOffspringOf B : order 차서

(3) ‘토지’(Estate), ‘노비’(Slave) 클래스 개체와 ‘내용’(Transaction) 클래스 개체 사이의 관계 속성

[토지 매매 또는 토지 분재]

- A isUnitOf B : ex_owner, new_owner

 전 소유주, 새 소유주

[노비 매매]

- A isUnitOf B : ex_owner, new_owner, age

 전 소유주, 새 소유주, 매매 또는 분재 당시 연령

[노비 분재]

- A isUnitOf B : ex_owner, new_owner, age, note, workplace

 전 소유주, 새 소유주, 상속 당시 연령, 비고(혼인질/도망질), 외거지

12) ‘관계 속성’은 Web Ontology Language (OWL)에서 직접 지원하지 않는다. 하지만 ‘관계 속성’을 사용할 때 대상 세계를 보다 용이하게 기술할 수 있기 때문에 인문 정보 데이터 모델을 설계할 때 ‘관계 속성’을 사용하도록 권장하고 있다. (※ 참고: http://digerati.aks.ac.kr:82/index.php?title=인문정보학_온톨로지_설계_가이드라인)

[호구 노비]

- A isUnitOf B : owner, age
소유주, 호구조사 당시 연령

(4) ‘토지’(Estate), ‘노비’(Slave) 클래스 개체와 ‘인물’(Person) 클래스
개체 사이의 관계 속성

- A ownedBy B : sequence 소유 순서

(5) ‘인물’(Person) 클래스 개체 사이의 관계 속성

- A hasSon B : legitimacy, order	적서구분, 차서
- A hasDaughter B : legitimacy, order	적서구분, 차서
- A hasWife B : legitimacy, order	처첩구분, 순서(후처의 경우)
- A hasRelative B : kinship	친족 관계

4. 문중 고문서 시맨틱 데이터베이스

문중 고문서 온톨로지를 기반으로 부안 김씨 고문서 중 및 그들의 가계 기록(족보)의 내용을 디지털 데이터로 편찬하였다.

앞에서 제시한 온톨로지(Ontology) 구조를 따르면, 문중 고문서에 담긴 모든 내용이 노드(Node)와 링크(Link)라는 2 가지 형태의 정보로 기술될 수 있게 된다. 노드(Node)와 링크(Link)는 네트워크(Network)를 구성하는 2 가지 필수 요소이다. 즉, 여러 편의 문중 고문서에 담긴 모든 내용을 노드(Node)와 링크(Link)로 표현하는 것은 그 내용 전체를 하나의 관계망(Network)으로 엮어서 살펴볼 수 있게 하는 것이다.

각각의 데이터가 하나의 커다란 관계망을 형성하고 있고, 안에서 각각의 노드가 서로에 대해 어떠한 관계가 있는지를 드러나는 데이터 네트워크를 ‘시맨틱 데이터’(Semantic Data)라고 한다.

본 연구는 궁극적으로 한국학중앙연구원의 한국 고문서 자료관 디지털 아카이브에 등재된 부안 김씨 고문서 740여 건의 내용을 모두 시맨틱 데이터로 편찬하여 그 내용 전체를 하나의 관계망(Network) 속에서 분석할 수 있게 하는 것을 장기 목표로 한다. 단, 데이터 모델 개발을 위한 1단계 연구에서는 김홍원-김명열-김번·김문-김수중 등 4대에 걸쳐 작성된 고문서 17건을 대상 자료로 선정하였으며, 이 자료로부터 432 건의 개별 정보 개체(individual, Node)와 그 개체 사이의 관계(relation, Link) 정보 1,006 건을 추출하였다. 시맨틱 데이터로 편찬한 부안 김씨 문중 고문서의 노드(Node)와 링크(Link) 데이터 일부를 유형별로 예시한다.

1) 노드(Node) 데이터: 개체(individual)

(1) ‘문서’(Document) 클래스 노드

id	sub class	name	doc type	issued year	issued by	issued to
BUKD001	R2	1837_고생원택노분옥_소지	소지	1837	분옥	부안현
BUKD002	R1	1837_고생원택노부록_토지매매명문	매매명문	1837	부록	김용관
BUKD003	R2	1837_고생원택노부록_배지	배지	1837	고생원	부록
⋮	⋮	⋮	⋮	⋮	⋮	⋮
BUKD017	R3	1675_김번_준호구	준호구	1675	부안현	김번

(2) '내용'(Transaction) 클래스 노트

[재산 분재]

id	sub class	name	owner	distributtee	trans year
BUKC001	B1	1672_김명열_노비분재	김명열	김수종	1672
BUKC002	B2	1672_김명열_토지분재	김명열	김수종	1672
BUKC003	B1	1677_김변5남매_노비분재	NULL	김변 외 4인	1677
⋮	⋮	⋮	⋮	⋮	⋮
BUKC06	B2	1688_김변_토지분재	김변	김수종 외 5인	1688

[토지 매매]

id	sub class	name	seller	buyer	purpose	price	deal year
BUKC007	B4	1837_고생원_토지매매	고생원	김용관	NULL	전문(錢文)_100냥	1837
BUKC008	B4	1838_고우진_토지매매	고우진	김용관	NULL	전문(錢文)_96냥	1838
BUKC009	B4	1840_최재영_토지매매	최재영	고진사	NULL	전문(錢文)_300냥	1840
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
BUKC011	B4	1677_윤이홍_토지매매	윤이홍	김변	NULL	쌀[米]10섬, 말[馬]1필	1677

[호구 조사]

id	sub class	name	reporter	receiver	report year
BUKC012	B5	1855_김용관_호구가족	김용관	부안현	1855
BUKC013	B6	1855_김용관_호구노비	김용관	부안현	1855
BUKC014	B5	1675_김변_호구가족	부안현	김변	1675
BUKC015	B6	1675_김변_호구노비	부안현	김변	1675

[관직 임명]

id	sub class	name	appointee	appointed year
BUKC016	B7	1660년 김명열 임명관직	김명열	1660
BUKC017	B7	1664년 김명열 임명관직	김명열	1664
BUKC018	B7	1662년 김변 임명관직	김변	1662
BUKC019	B7	1666년 김문 임명관직	김문	1666

(3) ‘인물’(Person) 클래스 노드

id	sub class	name	chi	alias	birth_s	death_s	legitimacy	clan
BUKP001	N1	김홍원	金弘遠	NULL	NULL	NULL	NULL	부안김
BUKP002	N1	김기원	金期遠	NULL	NULL	NULL	NULL	부안김
BUKP003	N1	김순공	金舜功	NULL	NULL	NULL	庶	부안김
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
BUKP176	N7	송몽득	宋夢得	NULL	辛丑	NULL	NULL	송

(4) ‘관부’(Office) 클래스 노드

id	sub class	name	chi	district_type
BUKO001	G1	부안현	扶安縣	지방
BUKO002	G2	예조	禮曹	중앙
BUKO003	G1	강진현	康津縣	지방
BUKO004	G1	황주목	黃州牧	지방

(5) ‘토지’(Estate) 클래스 노드

id	sub class	name	code	section	extent	norm
BUKE001	L1	대유천전평(大柳川前坪)_답(畓)	잠(箴)	NULL	8마지기	NULL
BUKE002	L1	우반동평(愚磻東坪)_답(畓)1	부(婦)	NULL	11말_5되지기	19복
BUKE003	L1	우반동평(愚磻東坪)_답(畓)2	부(夫)	2배미	4마지기	7복5속
⋮	⋮	⋮	⋮	⋮	⋮	⋮
BUKE006	L1	건선면(乾先面)장작동 (藏作洞)응장(鷹場)_답(畓)	NULL	NULL	8마지기	NULL

(6) ‘노비’(Slave) 클래스 노드

id	sub class	name	chi	birth	sexagerary	parent
BUKS001	S2	보령덕	保令德	· NULL	NULL ·	· 임춘(林春)
BUKS002	S1	경룡	慶龍	· NULL	NULL ·	· 보령덕(保令德)
BUKS003	S2	경춘	慶春	· NULL	NULL ·	· 보령덕(保令德)
⋮	⋮	⋮	⋮	⋮	⋮	⋮
BUKS206	S0	순화	NULL	NULL	NULL	NULL

(7) ‘관직’(Position) 클래스 노드

id	sub class	name	chi	pos type	rank
BUKP001	P1	현감	縣監	문반	통훈대부(通訓大夫)
BUKP002	P1	판관	判官	문반	통훈대부(通訓大夫)
BUKP003	P1	장사랑	NULL	문반	장사랑(將仕郎)
BUKP004	P1	종사랑	NULL	문반	종사랑(從仕郎)

2) 링크(Link) 데이터: 관계(relation)

(1) ‘문서’(Document) 클래스 개체 사이의 링크

node1		node2		relation
BUKD002	1837_고생원택노부록_토지 매매명문	BUKD001	1837_고생원택노분옥_ 소지	isRelatedTo
BUKD002	1837_고생원택노부록_토지 매매명문	BUKD003	1837_고생원택노부록_ 배지	hasRelation
BUKD010	1662_김번_고신	BUKD008	1660_김명열_고신	hasRelation
BUKD011	1666_김문_고신	BUKD009	1664_김명열_고신	hasRelation

(2) ‘인물’(Person) 클래스 개체 사이의 링크

node1		node2		relation
BUKP001	김홍원	BUKP004	김명열	hasSon
BUKP001	김홍원	BUKP005	김용열	hasSon
BUKP001	김홍원	BUKP006	김유열	hasSon
⋮	⋮	⋮	⋮	⋮
BUKP164	고제번	BUKP169	고희주	hasSon

(3) ‘노비’(Slave) 클래스 개체 사이의 링크

node1		node2		relation
BUKS001	보령덕	BUKS169	임춘	isOffspringOf
BUKS002	경룡	BUKS001	보령덕	isOffspringOf
BUKS003	경춘	BUKS001	보령덕	isOffspringOf
⋮	⋮	⋮	⋮	⋮
BUKS164	옥진	BUKS156	점진	isOffspringOf

(4) ‘노비’(Slave) - ‘인물’(Person) 클래스 개체 사이의 링크

node1		node2		relation
BUKS001	보령덕	BUKP004	김명열	ownedBy
BUKS002	경릉	BUKP004	김명열	ownedBy
BUKS003	경춘	BUKP004	김명열	ownedBy
⋮	⋮	⋮	⋮	⋮
BUKS164	옥진	BUKP026	김용관	ownedBy

(5) ‘토지’(Estate) - ‘인물’(Person) 클래스 개체 사이의 링크

node1		node2		relation
BUKE001	대유천전평(大柳川前坪)_답(畓)	BUKP172	고생원	ownedBy
BUKE001	대유천전평(大柳川前坪)_답(畓)	BUKP026	김용관	ownedBy
BUKE001	대유천전평(大柳川前坪)_답(畓)	BUKP174	고우진	ownedBy
⋮	⋮	⋮	⋮	⋮
BUKE006	건선면(乾先面)장작동(藏作洞)웅장(鷹場)_답(畓)	BUKP012	김수중	ownedBy

(6) ‘문서’(Document) - ‘인물’(Person), 노비’(Slave), ‘관부’(Office), ‘
클래스 개체 사이의 링크

node1		node2		relation
BUKD001	1837_고생원택노분옥_소지	BUKS165	분옥	IssuedBy
BUKD001	1837_고생원택노분옥_소지	BUKO001	부안현	IssuedTo
BUKD002	1837_고생원택노부록_토지매매명문	BUKS166	부록	IssuedBy
⋮	⋮	⋮	⋮	⋮
BUKD017	1675_김번_준호구	BUKP010	김번	issuedTo

(7) ‘문서’(Document) - ‘내용’(Transaction) 클래스 개체 사이의 링크

node1		node2		relation
BUKD002	1837_고생원택노부록_토지매매 명문	BUKC007	1837_고생원_토지매매	isProofOf
BUKD004	1838_고우진_토지매매명문	BUKC008	1838_고우진_토지매매	isProofOf
BUKD005	1840_최재영_토지매매명문	BUKC009	1840_최재영_토지매매	isProofOf
⋮	⋮	⋮	⋮	⋮
BUKD011	1662_김문_고신	BUKC019	1666_김문_관직임명	isProofOf

(8) ‘토지’(Estate) - ‘내용’(Transaction) 클래스 개체 사이의 링크

node1		node2		relation
BUKE001	대유천전평(大柳川前坪)_답(畓)	BUKC007	1837_고생원_토지매매	isUnitOf
BUKE001	대유천전평(大柳川前坪)_답(畓)	BUKC008	1838_고우진_토지매매	isUnitOf
BUKE002	우반동평(愚磻東坪)_답(畓)1	BUKC009	1840_최재영_토지매매	isUnitOf
⋮	⋮	⋮	⋮	⋮
BUKE006	감동교(甘洞橋)_답(畓)	BUKC002	1677_김번5남매_토지분재	isUnitOf

(9) ‘노비’(Slave) - ‘내용’(Transaction) 클래스 개체 사이의 링크

node1		node2		relation
BUKS001	보령덕	BUKC001	1672_김명열_노비분재	isUnitOf
BUKS002	경룡	BUKC001	1672_김명열_노비분재	isUnitOf
BUKS003	경춘	BUKC001	1672_김명열_노비분재	isUnitOf
⋮	⋮	⋮	⋮	⋮
BUKS164	옥진	BUKC013	1855_김용관_호구노비	isUnitOf

(10) ‘인물’(Person) – ‘내용’(Transaction) 클래스 개체 사이의 링크

node1		node2		relation
BUKP004	김명열	BUKC001	1672_김명열_노비분재	isGiverOf
BUKP004	김명열	BUKC002	1672_김명열_토지분재	isGiverOf
BUKP012	김수중	BUKC001	1672_김명열_노비분재	isInheritorOf
⋮	⋮	⋮	⋮	⋮
BUKP176	송몽득	BUKC014	1675_김번_호구가족	isUnitOf

(11) ‘내용’(Transaction) – ‘관부’(Office) 클래스 개체 사이의 링크

node1		node2		relation
BUKC016	1660_김명열_관직임명	BUKO003	강진현	placesIn
BUKC017	1664_김명열_관직임명	BUKO004	황주목	placesIn

(12) ‘내용’(Transaction) – ‘관직’(Position) 클래스 개체 사이의 링크

node1		node2		relation
BUKC016	1660_김명열_관직임명	BUKP001	현감	appointsAs
BUKC017	1664_김명열_관직임명	BUKP002	판관	appointsAs
BUKC018	1662_김번_관직임명	BUKP003	허직1	appointsAs
BUKC019	1666_김문_관직임명	BUKP004	허직2	appointsAs

5. 문중 고문서 데이터 네트워크

위에서 보인 바와 같은 노드(Node) 데이터와 링크(Link) 데이터는 각각 독립적으로 존재하는 것이 아니다. 그 데이터는 하나의 커다란 그물과도 같은 관계망을 형성하고 있으며, 그 안에서 각각의 노드가 서로에 대해 어떠한 의미 또는 관계가 있는지를 명시적으로 드러낸다. 이와 같이 ‘의미상으로 관련이 있는 데이터의 네트워크’(Semantic Data Network) 형태로 가공된 고문서 데이터베이스는 문서의 디지털 사본을 제공하는 데 머물던 종래의 데이터베이스와는 다른 방식으로 활용하게 된다.

문서 중심 데이터베이스에서는 개별 문서 한 건, 한 건이 정보 이용의 최종 목적지였다고 할 수 있다. 시맨틱 데이터베이스 상에서는 그 문서 속에 기록된 인물, 사건 등이 서로에 대해 어떠한 관계를 맺고 있는지 알려주는 데이터 관계망을 추출할 수 있다.

온톨로지 설계를 통해 만들어진 시맨틱 데이터베이스 상에서 대상 세계를 탐색하고 분석하는 방법은 매우 다양하게 열려 있다. 가장 일반적인 방법은 ‘스파클’(SPARQL, Simple Protocol and RDF Query Language)¹³⁾과 같은 표준화된 질의어를 사용하여 데이터 사이의 관계성을 탐색하는 것이지만, 이 글에서는 ‘데이터 관계망’에 대한 독자의 이해를 돕기 위해, 데이터 탐색 기능과 탐색 결과의 시각화 기능을 함께 제공하는 ‘그래프 데이터베이스’(Graph Database)를 사용하여 고문서 시맨틱 데이터의 실제적인 활용 방법을 보이고자 한다.

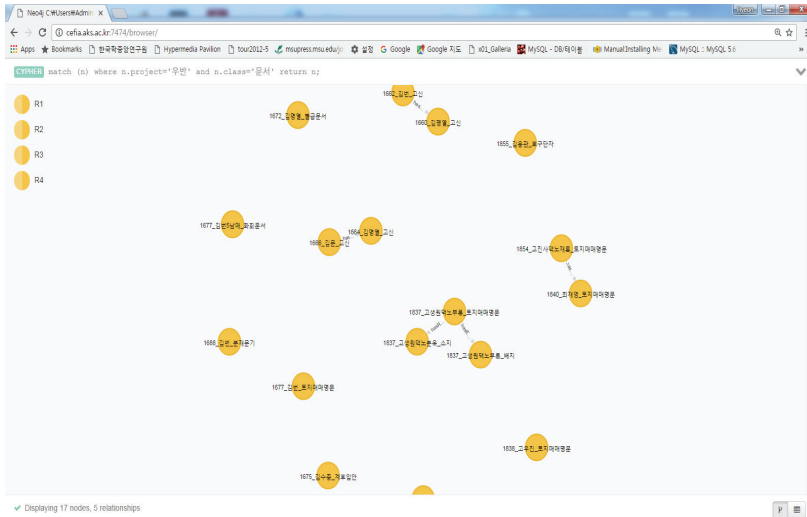
아래의 예시는 부안 김씨 고문서 내용 가운데 몇 가지 관심 사안에 대해 질의하고, 그 결과를 시각적 데이터로 표현한 것이다. 그래프 데이터베이스 관리 시스템 Neo4j¹⁴⁾를 데이터 질의 및 시각화 플랫폼으로 활용하였다.

13) SPARQL(Simple Protocol and RDF Query Language): 온톨로지 기반의 시맨틱 데이터를 탐색할 수 있는 데이터베이스 질의어. W3C(World Wide Web Consortium)에서 2008년에 처음 제안하였다. 2013년 이후 SPARQL 1.1이 W3C의 공식 권장안이다. <https://www.w3.org/TR/sparql11-query/>

14) Neo4j: Neo Technology사에서 개발한 그래프 데이터베이스 관리 시스템. 모든 데이터를 노드,

1) 기본 노드 탐색

(1) 문서: 부안 김씨 문중 고문서



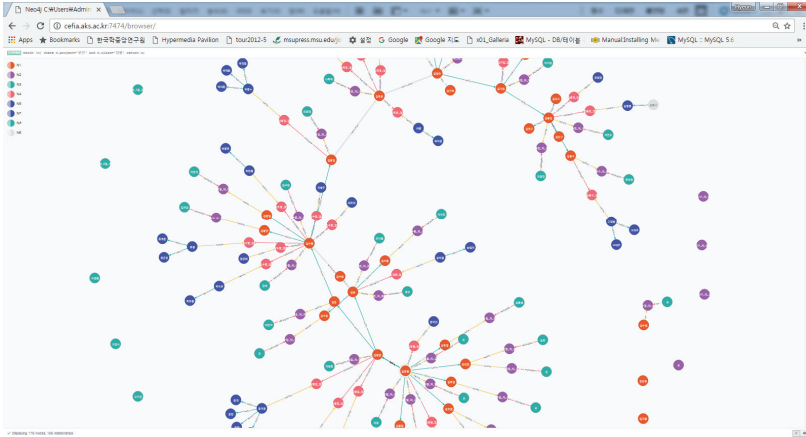
대상 문서 탐색:

Database URL	http://cefia.aks.ac.kr:7474
Cypher Query	<code>match (n) where n.project='우빈' and n.class='문서' return n;</code>

☞ 데이터베이스에 등재된 부안 김씨 문중 고문서를 찾는 질의와 그래프로 표현된 탐색 결과이다. 대상 문서의 목록뿐 아니라 그 문서들이 서로에 대해 어떠한 관계(실행 문서와 공중 문서의 관계 등)를 갖는지를 함께 보일 수 있다.

링크, 속성의 형태로 저장하기 때문에 데이터의 관계성에 대한 질의가 가능하고, 검색 결과를 그래프 형태로 표현할 수 있다. <http://neo4j.com/>

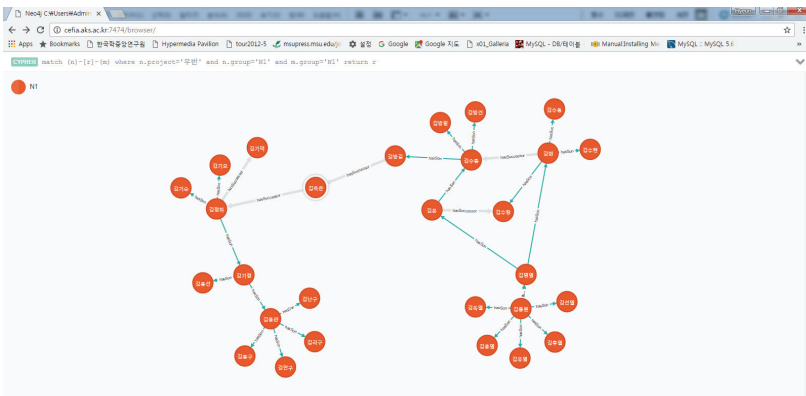
(2) 인물: 부안 김씨 족보



대상 인물 탐색 - 족보 및 분재기에 나타난 부안 김씨 집안 사람들:

Database URL	http://cefia.aks.ac.kr:7474
Cypher Query	match (n) where n.project='우반' and n.class='인물' return n;

☞ 데이터베이스에 등재된 부안 김씨 집안 사람들을 찾는 질의와 결과이다. 각각의 인물이 서로에 대해 어떠한 관계(친인척 관계 및 지인)에 있는지를 보인다.



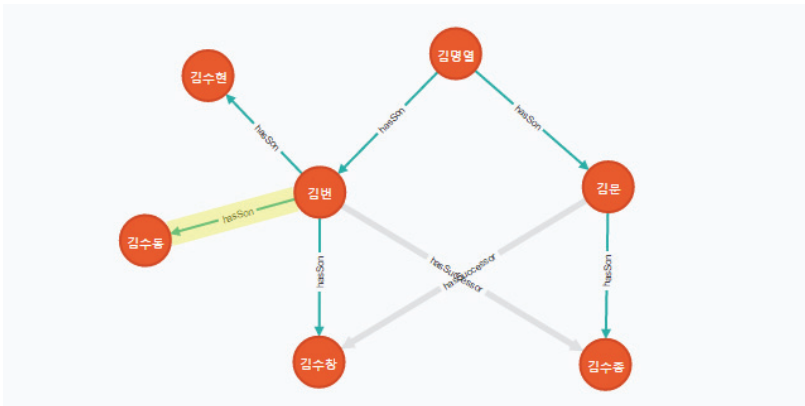
김홍원으로부터 김봉구 형제까지 남자 자손들:

Database URL	http://cefia.aks.ac.kr:7474
Cypher Query	match (n)-[r]-(m) where n.project='우반' and n.group='N1' and m.group='N1' return r

☞ 데이터베이스에 등재된 부안 김씨 집안 남자들의 계대(繼代)를 보이는 질의 결과이다. 친자 혹은 양자로 이어지는 10대의 가계 계승을 살필 수 있다.

2) 관심 사안에 관한 데이터 탐색 예시

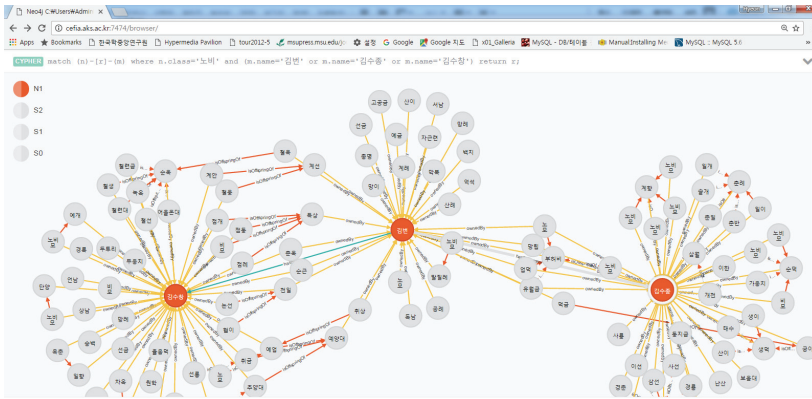
(1) 김번의 후사, 그들에 대한 재산 상속



김번, 김문 형제의 친자와 양자:

Database URL	http://cefia.aks.ac.kr:7474
Cypher Query	match (n)-[r]-(m) where (n.name='김번' or n.name='김문') and m.group='N1' return r;

☞ 김번은 처음에 아들이 없어 동생 김문의 아들 김수종을 양자로 들였다. 그 후에 친자 김수창을 낳게 되자 이번에는 김수창을 동생 김문의 양자로 보냈다. 다소 복잡한 이들의 부자 관계는 위의 그래프와 같이 탐색된다.

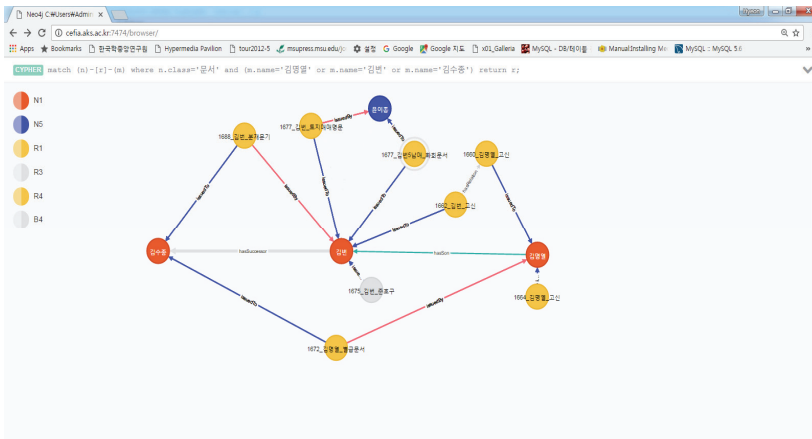


김변, 김수종, 김수창의 노비 소유와 상속:

Database URL	http://cefia.aks.ac.kr:7474
Cypher Query	match (n)-[r]-(m) where n.class='노비' and (m.name='김변' or m.name='김수종' or m.name='김수창') return r;

☞ 김변이 자신의 계후자인 김수종과 친자인 김수창(김변의 동생 김문의 계후자)에게 분배한 재산(노비)을 비교할 수 있는 데이터.

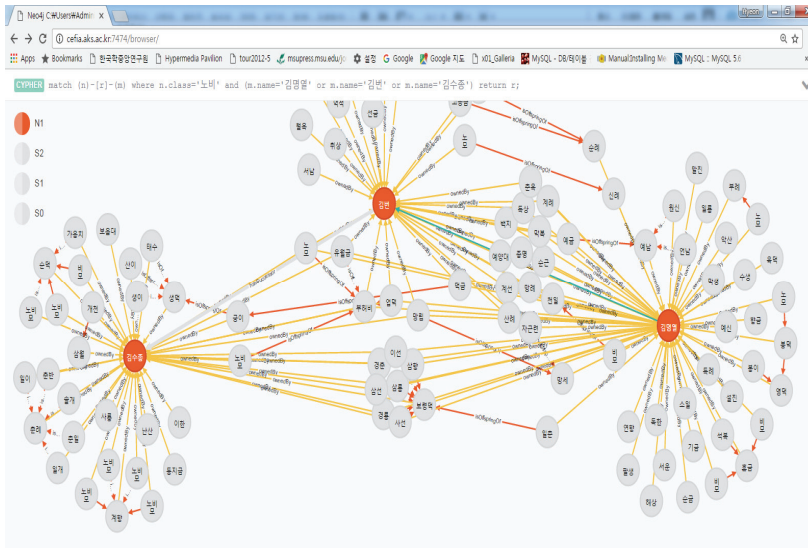
(2) 김명열 - 김변 - 김수종 3대의 재산 이동



김명열, 김변, 김수종 3대에 발생한 문서:

Database URL	http://cefia.aks.ac.kr:7474
Cypher Query	match (n)-[r]-(m) where n.class='문서' and (m.name='김명열' or m.name='김변' or m.name='김수종') return n, m;

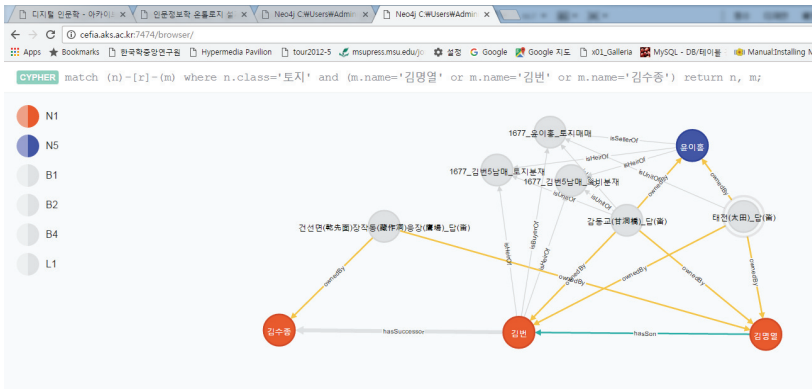
☞ 김명열로부터 그의 아들 김변, 김변의 계후자 김수종까지의 재산 이동을 추적할 수 있는 문서를 탐색.



김명열, 김변, 김수종 3대의 노비 소유권 이동:

Database URL	http://cefia.aks.ac.kr:7474
Cypher Query	match (n)-[r]-(m) where n.class='노비' and (m.name='김명열' or m.name='김변' or m.name='김수종') return n, m;

☞ 김명열, 김변, 김수종이 소유하였던 노비들과 그 노비들의 혈연관계를 보이는 데이터 탐색.

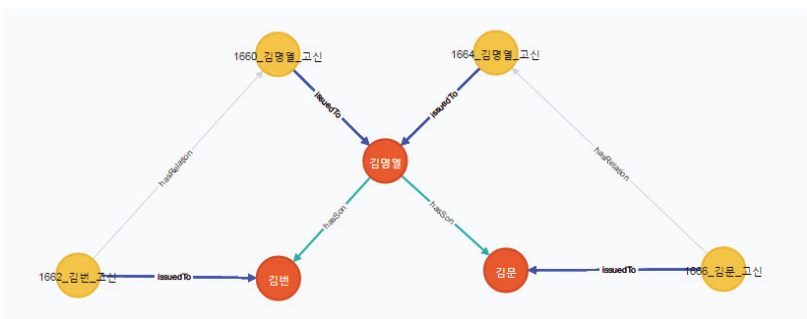


김명열, 김변, 김수중 3대의 토지 소유권 이동:

Database URL	http://cefia.aks.ac.kr:7474
Cypher Query	match (n)-[r]-(m) where n.class='토지' and (m.name='김명열' or m.name='김변' or m.name='김수중') return n, m;

☞ 김명열, 김변, 김수중이 소유하였던 토지와 그 토지의 매수 및 소유권 이동에 관한 데이터 탐색.

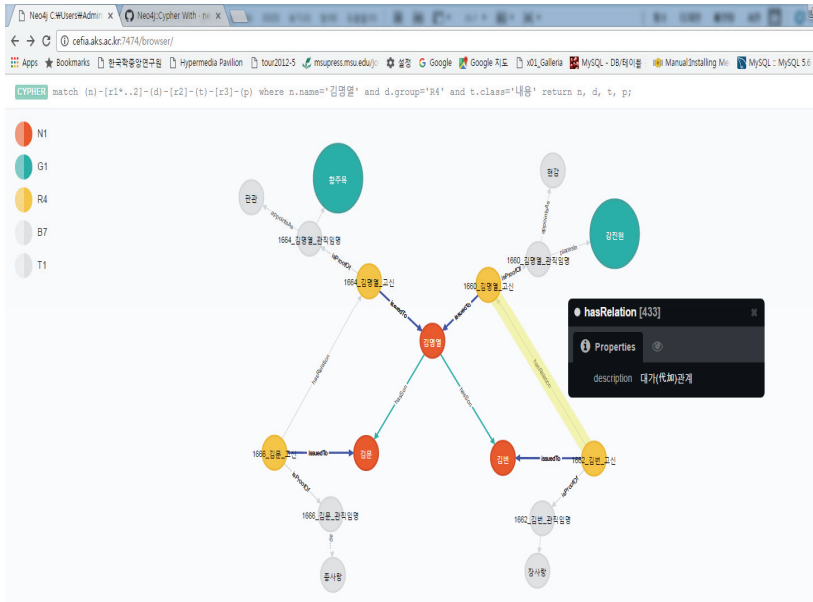
(3) 김명열과 그의 두 아들 김변, 김문의 고신



김명열과 그의 두 아들이 받은 4 건의 고신 문서(관직 임명장):

Database URL	http://cefia.aks.ac.kr:7474
Cypher Query	match (n)-[r*..2]-(d) where n.name='김명열' and d.group='R4' return r

☞ 김명열과 김변, 김문의 가족관계 및 그들에게 발급된 고신 문서(관직 임명장)의 관계를 보이는 데이터.

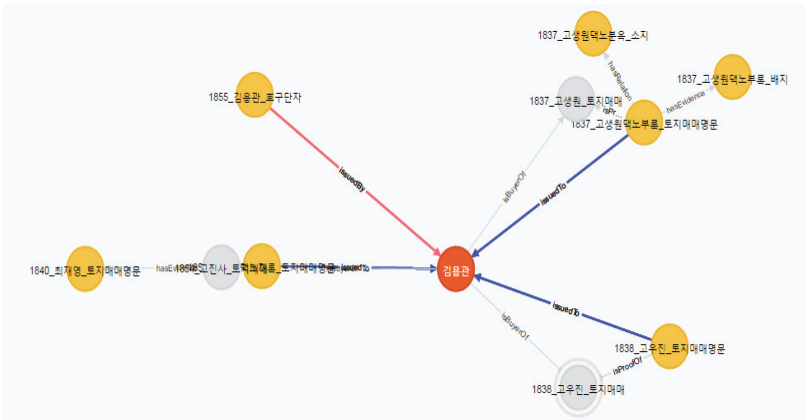


김명열과 그의 두 아들들이 받은 고신의 내용:

Database URL	http://cefia.aks.ac.kr:7474
Cypher Query	match (n)-[r1*..2]-[d]-[r2]-[t]-[r3]-[p] where n.name='김명열' and d.group='R4' and t.class='내용' return n, d, t, p;

☞ 김명열은 1660년에 강진현 현감, 1664년에 황주목 판관으로 임명된다. 이때 자신의 품계를 더하는 대신 (당하관의 최고 품계에 이미 도달했기 때문에) 두 아들이 차례로 장사랑(종9품 품계)과 종사랑(정9품 품계)의 품계를 대신 받도록 했다(代加). 위의 그래프는 김변과 김문의 고신이 김명열의 2차레에 걸친 고신과 대가(代加) 관계에 있음을 보이고 있다.

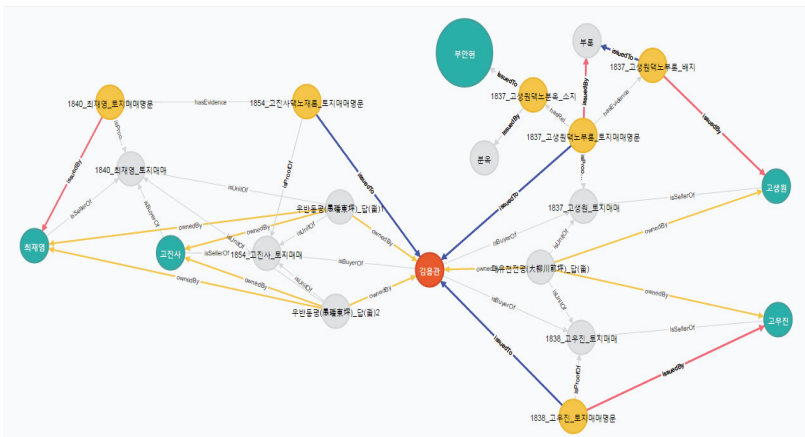
(4) 김용관과 고생원(고진사) 댁 사이의 토지 거래



김용관의 토지 매입 문서:

Database URL	http://cefia.aks.ac.kr:7474
Cypher Query	match (n)-[r*..2]-(e) where n.name='김용관' and e.class='문서' return r;

☞ 김용관에게 또는 김용관에 의해 발급된 문서들. 이 가운데 김용관을 수급자로 하는 3건은 그가 토지를 매입한 사실을 보이는 문서이다.



김용관의 토지 매입 문서에 담긴 내용 및 관련 인물:

Database URL	http://cefia.aks.ac.kr:7474
Cypher Query	match (n)-[r1*.2]-(e)-[r2]-[m] where n.name='김용관' and e.class='문서' and m.class='인물' return r1, r2;

☞ 김용관은 ‘고생원 댁’ 또는 ‘고진사 댁’이라고 불린 인근 사족 집안과 1837년, 1838년, 1854년에 세 차례 토지 매매 거래를 했다. 그런데, 1837년과 1838년의 거래는 한 건의 동일한 토지에 대해 두 차례 매매 계약을 체결한 것이다. 땅을 팔면서 언젠가 그 땅을 돌려받기를 원한 고생원 댁이 땅을 되살 수 있다는 조건을 덧붙여 수정 계약을 한 것으로 보인다. 두 번째 토지매매명문에는 매도자의 이름이 ‘고우진’으로 기록되어 있어서 이 시점에는 고생원이 고우진이었음을 알 수 있다.

1854년에는 김용관이 ‘고진사 댁’으로 부터 2 곳의 토지를 더 사들인다. 부안 김씨 고문서 중에는 이 2건과 동일한 지목의 토지를 1840년에 최재영이라는 인물이 매도한 기록이 있는데, 이는 고진사 댁이 이 토지를 김용관에게 팔 때 소유권 증빙 목적으로 팔려온 문서로 보인다.

6. 맺음말: 데이터 기반 인문학 연구의 전망

‘소통’과 ‘융합’이 ‘4차 산업혁명 시대’의 키워드로 운위되는 것은 데이터의 도메인이 다들지라도 디지털이기 때문에 하나로 묶일 수 있고, 묶이면 더 큰 가치를 발휘한다는 사실을 알게 된 데 기인한다. 인문지식 탐구의 과정에서 그 지식을 ‘데이터’로 기술하여 소통시키는 일, 그렇게 해서 그 지식이 인간사회의 다양한 영역에서 활용되고 가치를 발휘할 수 있게 하는 것이 디지털 인문학이 인식하는 미래 인문학의 과제이다.

종래의 고문서 디지털 아카이브는 ‘인문학 자료 디지털화’의 결과물을 전자문서로 축적하고, 그것을 온라인 상에서 서비스하는 기능만을 제공했다. 그 자료 속

에서 의미 있는 정보를 추출하여 해석하고, 사실을 발견하고 그것을 새로운 문맥으로 재구성하는 일은 디지털 아카이브에서 선별해낸 자료를 인쇄물로 출력한 후 책상 위에서 아날로그적인 방법으로 수행해야 했고, 또 그렇게 해서 찾아낸 ‘연구 성과’들은 별도의 저작물로 간행되기 때문에 ‘고문서 아카이브’는 언제까지나 기초 자료의 서비스 기능에 머물 수밖에 없었다.

디지털 인문학과 동행하는 새로운 형식의 고문서 디지털 아카이브는 연구자들이 직접 그 디지털화된 문서의 내용을 데이터로 취급할 수 있게 한다. 그리고 데이터의 분석, 정리, 해석을 통해 찾아낸 ‘사실’과 ‘이야기’를 확장된 데이터로 축적해 나아갈 수 있게 함으로써, ‘자료’의 아카이브에서 ‘지식’의 아카이브로의 진화를 지향한다.

고문서의 내용을 디지털 정보로 남아낼 데이터 형식은 향후 대상 고문서의 범위 및 연구 주제의 확장에 따라 더 포괄적으로 확대되거나, 더 정교하게 구체화되어갈 것이다. 그리고 그와 같은 데이터 모델의 확장 연구는 이 연구의 연구자들뿐 아니라 유사한 성격의 고문서 연구를 수행하는 여러 연구자들의 참여를 통해 고문서 데이터의 공동 활용을 가능하게 하는 데이터 모델 표준화 연구로 발전해갈 것으로 기대한다.

디지털적인 방법에 의한 고문서 연구에서 가장 의미 있게 보아야 할 점은 그 방법이 한 종류의 문서, 또는 한 집안 문중에 대한 연구에만 적용되는 것이 아니라라는 점이다. 각각 다른 연구자와 연구 팀에 의해서 다른 지역의 고문서가 연구되는 경우라 하더라도, 그 관계망의 노드가 되는 문서와 인물의 식별자만 공유한다면 그 모든 해석과 분석의 결과들이 서로 의미가 통하는 하나의 지식 체계로 통합될 수 있다. 더 나아가, 인물과 문서뿐 아니라 동시대의 역사적 사건이나 그 시대의 제도, 지역적 배경, 언어적인 것에 대한 지식 등 다른 분과 학문에서 탐구되어진 지식들도 여기서 추구하는 고문서 관련 지식과 연관을 맺을 수 있고, 서로에게 의미 있는 지식의 확장을 이룰 수 있다.¹⁵⁾

15) 김현, 「디지털 인문학과 고문헌 자료 연구」, 『열상고전연구』 제50집 (2016. 4. 열상고전연구회) p. 35.

‘인문학 자료의 디지털화’와 ‘디지털 인문학’, 하나의 연장선상에 있다고도 할 수 있는 그 두 가지 사이에 존재하는 가장 큰 차이점은 ‘누가 하느냐’이다. ‘인문학 자료 전산화’는 정보처리 기술을 운용할 수 있는 사람들이 인문학 연구자와 교육자, 피교육자들을 위해 인문학 자료를 디지털 데이터로 전환하는 것을 위주로 하였으며, 그렇게 함으로써 자료 이용의 편의성을 증진시키는 것을 목적으로 하였다. 이에 반해, ‘디지털 인문학’은 인문학 연구자와 교사, 학생들이 디지털 환경에서 자기주도적으로 연구와 교육 활동을 수행하는 것이다. 디지털 기술의 활용을 통해 예전에는 가능하지 않았던 새로운 연구·교육 성과를 도출하고 이로써 인문학의 사회적 기여를 촉진하는 것을 목표로 한다.

고문서 내용을 데이터로 전환하여 기계적 가독성을 부여하는 것은 일차적으로 과거의 생활사에 대한 인문학적 연구에 기여하겠지만, 향후 이러한 데이터의 광범위한 축적으로 만들어질 인문학 빅데이터는 현대에 발생하는 사회과학적, 산업적 빅데이터와 엮여져서 현대사회에 대한 이해와 과거 역사의 탐구 사이의 연결 고리를 만들어내는 것도 가능할 것이다. 인문지식이 데이터로 소통될 수 있도록 하는 노력을 통해 우리의 인문지식은 전통적인 인문학의 경계를 넘어서는 세계와도 교섭할 것이며, 인간과 사회에 대한 이해로서의 인문학의 외연을 더욱 의미 있게 확장해 갈 수 있을 것이다.

〈참고문헌〉

- 김하영, 「문중고문서 디지털 아카이브 구현 연구」, 한국학중앙연구원 한국학대학원 석사학위 논문, 2015.
- 김현, 「한국 고전적 전산화의 발전 방향 -고전 문집 지식 정보 시스템 개발 전략」, 『민족문화』 제28호, 민족문화추진회, 2005, 168~169쪽.
- 김현, 「고문헌 자료 XML 전자문서 편찬 기술에 관한 연구」, 『고문서연구』 제29호, 한국고문서학회, 2006, 183~230쪽.
- 김현, 『인문정보학의 모색』, 북코리아, 2012.

- 김현, 류인태, 「관계의 발견을 위한 디지털 스토리텔링 데이터 모델」 『한국고문서 정서·역주 및 스토리텔링 연구-2015년 연구보고서』, 한국학중앙연구원 장서각, 2015.
- 김현, 「디지털 인문학과 고문헌 자료 연구」, 『열상고전연구』 제50호, 열상고전연구회, 2016, 13~38쪽.
- 김현 외 2인, 『디지털 인문학 입문』, HUEBOOKS, 2016.
- 김현영, 「고문서 개념의 수용과 전근대 아카이브즈 자료의 정리」, 『사학연구』 제95호, 한국사학회, 2009, 107~133쪽.
- 류인태, 「조선시대 고문서 시맨틱 웹 DB 설계 기초 연구」, 2016 세계한국학대회 발표논문, 2016.
- 이희환, 「부안 김씨를 통해서 본 조선말기 相尙의 풍조」, 『역사학보』 제176호, 역사학회, 2002, 111~142쪽.
- 전경목, 『고문서를 통해서 본 우반동과 우반동 김씨의 역사』, 신아출판사, 2001.
- 전경목, 「분재기를 통해서 본 분재와 봉사 관행의 변천 : 부안김씨 고문서를 중심으로」, 『고문서연구』 제22호, 한국고문서학회, 2003, 249~270쪽.
- 전경목, 「우반동 김씨 가문 내에서 서자의 역할과 활동」, 『장서각』 제34호, 한국학중앙연구원, 2015, 198~229쪽.
- 정연식, 「한국 생활사 연구의 현황과 과제 : 조선시대 생활사 연구를 중심으로」 『역사와 현실』 72, 한국역사학회, 2009, 289~314쪽.
- 최승희, 『한국 고문서 연구』, 지식산업사, 1989.
- 한국고문서자료관, 한국학중앙연구원 한국학자료센터, <http://archive.kostma.net>
- 한국역대인물 종합정보시스템, 한국학중앙연구원. <http://people.aks.ac.kr>
- 인문정보학 온톨로지 설계 가이드라인, 한국학중앙연구원 인문정보학교실. http://dh.aks.ac.kr/Edu/wiki/index.php/온톨로지_설계_가이드라인
- OWL Web Ontology Language Overview, W3C Recommendation, <http://www.w3.org/TR/owl-features/>
- London Lives, <https://www.londonlives.org/>

<Abstract>

The Pursuit of Data-based Humanities Research Methodologies: Where Digital Humanities and Archives of Old Documents Meet

김 현·안승준·류인태 (The Academy of Korean Studies)

Rather than simply digitizing humanities materials or providing humanities research results in a digital format, “digital humanities” encompasses ever more creative humanities endeavors in an information technology environment. There is still no research model which can clearly answer the question of what “digital humanities research” looks like. However, if we look at the similarities among humanities research which has been deemed to fall under the umbrella of “digital humanities research,” we can see the transformation of humanities knowledge elements from human-composed language to machine-readable data, which allows computers to aid in the interpretation of said information.

The author will go a step beyond the fundamental discussion of “what digital humanities is,” to explain the nature of the work of transforming humanities knowledge into data, and what kind of scholarly outputs we can expect from such work. To this end, the author presents research led by the Academy of Korean Studies’ Jangseogak Archives and Cultural Informatics Lab which is based on old documents from paternalistic lineage organizations (*munjung*).

A new form of digital archive of old documents made for a digital humanities framework allows researchers to turn the content of digitized old documents into data. In addition, through the accumulation of data which

encompasses “truths” and “stories” discovered via data analysis, organization, and interpretation, an archive of “source materials” can evolve into an archive of “knowledge.”

On preliminary level, transforming the content of old documents into data and making it machine readable contributes to humanities research on the lives of people of the past. However, looking forward, the widespread accumulation of such data can create big data on the humanities. And if this humanities big data is connected to big data dealing with contemporary social, academic, and business phenomena, it can serve as a link between the understanding of modern society and the history of the past. Through efforts to allow humanities knowledge to communicate as data, our humanities knowledge can go beyond the boundaries of traditional humanities and negotiate with the larger world, and it can more meaningfully expand the full realm of the humanities as an understanding of humans and society.

Key words : digital humanities, old documents, paternalistic lineage organizations (*munjung*), ontology, digital storytelling, digital archives, digital curation, humanities big data