

한국역사인물데이터베이스 설계 시론

『홍문록(弘文錄)』을 예시로 하여

김바로

한국학중앙연구원 한국학대학원 조교수, 디지털인문학 전공

ddokbaro@gmail.com

- I. 머리말
 - II. 한국역사인물데이터베이스의 선행 모델: CBDB
 - III. 한국역사인물 관련 선행 정보시스템
 - IV. 『홍문록』 온톨로지 설계
 - V. 맺음말
-

I. 머리말

인문지식을 토대로 하는 데이터베이스(DB: Database) 혹은 아카이브의 설계와 구축은 인문학 프로젝트에서 중요한 부분을 차지한다. 1995년 ‘국역 조선왕조실록 CD-ROM’ 간행을 계기로 인문 데이터의 가치가 인정받기 시작했고, IMF체계를 극복하고자하는 정책의 일환으로 시행된 공공 근로 정보화 사업을 통해서 대량의 인문 데이터가 정부 주도하에 만들어졌다.¹ 그 이후 인문학자들은 각자의 영역에서 본격적으로 다양한 인문 데이터를 구축해왔다.

그러나 인문학자들은 종이를 기반으로 하는 인문지식의 편찬 방법에는 능숙하지만, 디지털을 기반으로 하는 인문지식의 편찬 방법에는 익숙하지 않다. 그래서 인문데이터 구축 작업에서 종이 매체와 디지털 매체라는 매체적 차이에 따른 다양한 가능성들을 온전히 발현하지 못하고, 기존 종이 매체에서 수행했던 설계와 구축 방법을 그대로 디지털 매체로 가져왔다. 그리고는 실제 인문 지식의 설계와 구축의 실무를 인문학자가 아닌 정보공학자에게 ‘외주’로 맡겼다.

문제는 정보화의 과정에서 정보공학자들이 인문학자의 지시대로만 행동했고, 인문학자의 참여 또한 적극적이지 않았기에, 다층적이고 유기적인 인문학의 세계가 디지털에서 온전히 구현되지 못했다는 점이다. 그 결과, 기존에 각각의 대상에 대해서 독자적으로 책을 출판하던 관행이 그대로

※ 이 논문은 2022년 한국학중앙연구원 공동연구과제 ‘조선시대 『홍문록(弘文錄)』과 집권세력의 네트워크’의 일환으로 수행됨(AKSR2022-C14).

※ 본 논문에 꼼꼼한 검토와 함께 유익한 도움말을 해주신 익명의 심사자분들께 감사의 말씀을 드립니다.

1 김현, 『인문정보학의 모색』(성남: 북코리아, 2012), 448~449쪽.

이어졌고, 서로 연결 가능한 다양한 정보들이 독립적인 정보 시스템으로 분리되었다. 그래서 한국학중앙연구원 역대인물정보시스템에 수록된 “이순신”과 국사편찬위원회의 정보시스템에 입력된 “이순신”은 서로 정보가 연결되어 있지 않았다. 이에 따라 “이순신”에 대해 연구하기 위해 인문학 연구자들은 각각의 정보시스템을 따로 방문하여 확인하는 번거로움을 감수해야만 했다. 비록 한국역사정보통합시스템²을 통해서 인문학 기관들의 정보시스템을 상호 연계하려는 시도는 있었으나, 키워드 통합 검색의 수준이었으며, 본질적으로 높은 수준의 기계가독형데이터를 토대로 정보시스템을 상호 연계한 것이 아니었기에 결국 연구자가 개별 사료를 따로 검토해야만 하는 한계가 있었다.

이에 우선은 중국 역사 인물에 관한 정보를 종합하여 디지털에서 구현한 ‘중국인물 데이터베이스 프로젝트(CBDB: China Biographical Database Project, 中國歷代人物傳記資料庫, 이하 CBDB로 통칭)’를 살펴본다. 이를 바탕으로 『홍문록(弘文錄)』³을 예시로 하여, 한국학중앙연구원 한국역대인물정보시스템, 국사편찬위원회 조선왕조실록 부가열람, 한국고전번역원 한국고전종합DB를 상호 연결할 수 있는 한국역사인물데이터베이스(KBDB: Korea Biographical Database Project) 온톨로지⁴ 설계 방안을 탐색하고자 한다.

2 국사편찬위원회 한국역사정보통합시스템(<https://www.koreanhistory.or.kr>).

3 이 글에서는 『홍문록』을 홍문관원을 선발하는 후보 명단으로 개념 정의했다. 『홍문록』은 「본관록」과 「도당록」으로 구성되어 있고, 이를 합본한 책에 『등영록』이라고 명명하기도 하였다. 따라서 특별히 분석 대상으로서의 자료를 언급할 때에는 『등영록』을 사용했다.

4 온톨로지란 정보화의 대상이 되는 세계를 전자적으로 표현할 수 있도록 구성한 데이터 기술 체계를 말한다. 이에 대한 상세한 내용은 다음 책을 참조할 수 있다. 김현·임영상·김바로, 『디지털 인문학 입문』(서울: 한국외국어대학교 지식출판원, 2016), 163~183쪽.

II. 한국역사인물데이터베이스의 선행 모델: CBDB

중국 역사 인물에 관한 정보를 디지털에서 구현한 CBDB는 2022년 8월 기준으로 7세기부터 19세기까지 52만 1,442명의 인물의 정보를 정리한 데이터베이스이다.⁵ CBDB에는 인물의 기본 정보(생몰시간, 생몰공간 등)는 물론이고, 인맥관계(혈연관계와 다양한 사회관계), 입사와 관직 임명 정보 및 관련 공간 정보 등이 포함되어 있다. CBDB는 1980년대 미국 송대사 연구자 로버트 하트웰(Robert Hartwell)이 개인적으로 정리한 송대 인물 전기 데이터를 기반으로 하고 있다. 중국 문학을 전공한 캘리포니아대학교 어바인캠퍼스(University of California, Irvine)의 마이클 풀러(Michael Fuller) 교수가 2005~2006년에 MS Access에 적재했고, 이후 미국 하버드 페어뱅크 중국학 연구센터와 타이완 중앙연구원 역사언어연구소 및 중국 베이징대학교 중국고대사연구센터가 공동으로 편찬하고 있다. CBDB는 수십만 명 단위의 인물에 대한 종합적인 정보를 제공하기에, 전통적인 개념의 역사 인물 공구서로서의 역할을 수행할 뿐만 아니라, 방대한 데이터를 토대로 기존에는 수행하기 어려웠거나 불가능했던 사회네트워크분석, 지리정보분석 등의 다양한 연구 및 프로젝트가 시도되고 있다.⁶ 또한 자신들의

5 CBDB에 대한 상세 정보는 CBDB 홈페이지(<https://projects.iq.harvard.edu/cbdb>)에서 찾아볼 수 있다. 특히 다음 URL에 동영상, 영문과 중문 가이드북, CBDB 소개 등의 정보가 정리되어 있다. <https://projects.iq.harvard.edu/chinesecbdb/下載教學與輔助文件>. 또한 다음의 논문을 참조할 수도 있다. 包弼德·王宏蘇·傅君勳·陳松·柳舟·朱厚權, 「中國歷代人物傳記資料庫(CBDB)의 歷史, 方法與未來」, 『數字人文研究』 2021-1 (2021), 21~33쪽.

6 CBDB 데이터를 토대로 하는 영어 논문은 다음 주소에 정리가 되어 있다. <https://projects.iq.harvard.edu/cbdb/publications-use-cbdb-data>. 주요 중국어 논문은 다음과 같다. 包弼德, 「群體, 地理與中國歷史: 基於CBDB和CHGIS」, 『量化歷史研究』 2017-Z1(2017), 213~246쪽; 李宗翰·鄭莉, 「家族, 婚姻與道學: 《仙溪志·人物傳》中的社會關係」,

데이터를 사용하는 방법에 대한 교육을 학술대회 발표, 워크숍 개최 등의 다양한 방법을 통해서 수행하고 있다.

1. CBDB 데이터 모델

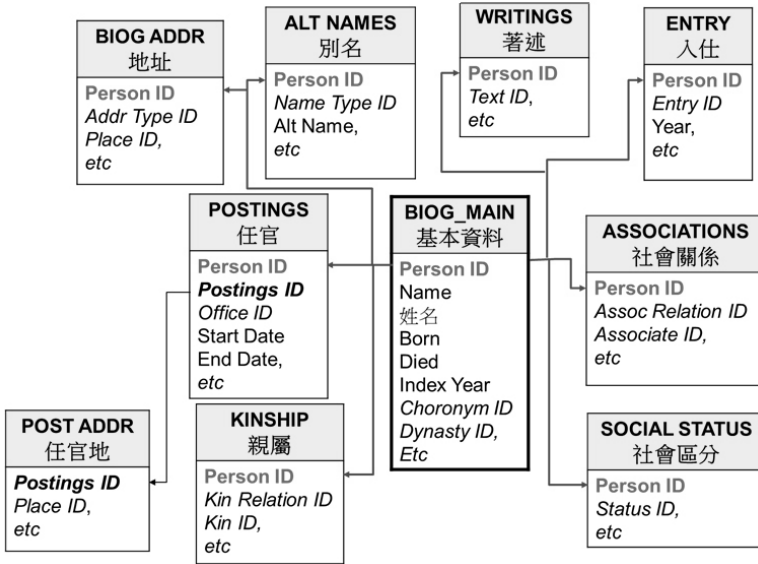


그림1-CBDB 핵심 스키마⁷

⁷『唐宋曆史評論』 2017-0(2017), 33~45쪽; 錢超峰·杜德斌, 「北宋官僚家族網絡的空間結構及其演化: 基於CBDB和CHGIS的考察」, 『曆史地理研究』 2019-2(2019), 83~94쪽; 鄧君·常嚴予·孫紹丹·宋先智·鍾楚依, 「數字人文視闕下清代科舉進士群體時空網絡結構分析」, 『現代情報』 2022-2, 2022, 4~13쪽.

주요 한국어 논문은 다음과 같다. 조원희, 「시론: 네트워크 분석을 통해 본 아우르바르와다 카안(元仁宗, 재위 1311-1320) 시기의 元 朝廷」, 『중앙아시아연구』 제22권 제2호(2017), 73~96쪽; 배숙희, 「디지털 인문학과 송대사 연구: 1163년의 科擧와 樓鑰를 중심으로 본 사회적 네트워크」, 『동양사학연구』 제146호(2019), 157~191쪽; 조원희, 「네트워크 분석을 통해 본 元帝國 성종-무종-인종 시기 朝廷: 몽골의 '한화'를 반박하는 또 하나의 방법론」, 『중국학보』 제99호(2022), 165~192쪽.

CBDB에서 우리가 주의 깊게 봐야 할 지점은 방대한 정보의 양이 아닌 세밀한 구조이다. CBDB는 50만 건에 이르는 방대한 인물에 대한 정보를 편찬했다. 그런데 이보다 중요한 것은 CBDB에서는 기존 인물사전과 공구서에서 다루었던 정보는 물론이고, 기존 인물사전과 공구서에서는 다루기 어려웠던 정보까지 설계에 반영했다는 점이다.

CBDB는 역사 인물 정보를 크게 기본 정보, 관직 정보, 관계 정보로 구분하여 편찬하고 있다. 기본 정보에는 성명, 생년, 몰년은 물론이고, 다양한 이칭 정보⁸, 다양한 관련 공간 정보⁹, 다양한 신분 정보¹⁰, 저작 참여

-
- 7 Harvard University General Introduction to CBDB 201404(https://projects.iq.harvard.edu/files/cbdb/files/general_introduction_to_cbdb_2014.pptx).
- 8 현재 CBDB에서는 18개의 이칭 분류를 사용하고 있다. 이칭 분류의 세부 내용은 다음과 같다. Alternate Personal Name, Previously Used Name(別名, 曾用名), Courtesy name(字), Studio name, Style name(室名, 別號), Posthumous Name(諡號), Birth-order name(行第), Enfeoffment Title(封爵), Childhood Name(小名), Childhood Courtesy Name(小字), Bestowed name or title(賜號), Secular surname(俗姓), Secular personal name(俗名), Temple name(廟號), Honorific name(尊號), Temple title(廟額), Other transliterated name(其他譯名), Original surname(本姓), Dharma name(法號), Daoist name(道號).
- 9 현재 CBDB에서는 20개의 인물 관련 공간 정보를 사용하고 있다. 인물 관련 공간 정보의 세부 내용은 다음과 같다. Basic Affiliation(籍貫(基本地址)), Moved to(遷住地), Former Address(前住地), Last Known Address(最後所知地), Ancestral Address(祖籍), Actual Residence(落籍(實際居住地)), Household Registration Address(本貫), Birth Address(出生地), Burial Address(葬地), Death Address(死所), Migration Route of Branch(本族支裔遷徙路線), visited or went to(遊歷或曾經到過), Eight Banner Qing Dynasty(八旗清代), alternate basic affiliation(另一籍貫(基本地址)), place of refuge(避兵之地), Household address(戶籍地), Place of exile(流放之地), Chose to reside at(卜居), Sojourned at(僑居), Exiled to(坐徙).
- 10 현재 CBDB에서는 9개의 신분 대분류와 271개의 신분 소분류를 사용하고 있다. 신분 대분류의 세부 내용은 다음과 같다. Occupation(事業), Scholarship(學術), Neo-Confucian(理學), Military Distinction(武功), Imperial Clan(宗社), Artistic Distinction(藝術), Religious Distinction(宗教), Life Events(時事), Commoner Activity(布衣事); 신분 소분류는 다음 URL에서 확인할 수 있다. <https://docs.google.com/spreadsheets/d/1m0BYiFq0ZufdQGO93c4PwSM1N9T-Hc0bCt6unoxL14w/edit#gid=723153616>.

정보¹¹를 연계하고 있다. 관직 정보는 인물의 입사(入仕) 정보¹²와 임관(任官) 정보로 세분화되어 있다. 임관 정보는 관직 정보¹³, 관직 임명 방식 정보¹⁴, 관직 부임 정보¹⁵ 및 관직 부임 공간 정보로 다시 세분화 되어 있다. 마지막으로

-
- 11 현재 CBDB에서는 11개의 저작 참여 분류를 사용하고 있다. 저작 참여 분류의 세부 내용은 다음과 같다. author(撰著者), editor(編輯者), compiler(編纂者), publisher(出版者), donor(捐助者), translator(翻譯者), annotator(註疏者), commentator(註釋者(含評點者)), proofreader(校對者), work included in(收入彙集), Editorial Associate(編輯助理)
 - 12 현재 CBDB에서는 20개의 입사 대분류와 267개의 입사 소분류를 사용하고 있다. 입사 대분류의 세부 내용은 다음과 같다. Palace(宮廷門), Kinship(血親門), Marriage(姻親門), Examination(科舉門), Regular Examination(正常科舉), Jinshi Examinations(進士類), Juren Examinations(舉人科), Licentiate(生員), Occasional Examination(非經常科舉), Decree Examination(制舉), Other Examination(其他舉), Schools(學校門), Students in Schools(學生門), Imperial Academy Students(監生門), Tribute Students(貢生門), Yin Privilege(恩蔭門), Recruitment(徵召門), Recommendation(薦舉門), Military Merit(軍功補授門), Specialized Talents(技進門), Surrender(歸降門), Purchase(進納門), Decree of Special Grace(特旨門), Religion(宗教門), Failed Pursuit of Office(求官不得門), Seven Specials(七色補官門), Jinshi Examinations(進士類), Juren Examinations(舉人科), Licentiate(生員), Other Method of Entry(其他入仕途徑); 입사 소분류는 다음 URL에서 확인할 수 있다. <https://docs.google.com/spreadsheets/d/1m0BYiFq0ZufdQGO93c4PwSM1N9T-Hc0bCt6unoxL14w/edit#gid=1876239808>.
 - 13 현재 CBDB에서는 14개의 관직 분류와 3만 2,225개의 관직명을 사용하고 있다. 관직 분류의 세부 내용은 다음과 같다. Classification title(階官), Commission(差遣), Complimentary censorial title(憲官), Dignitary(勳), Honorary office(檢校官), Office(職事官), Prestige title(散官), Probationary title(試秩), Nobility(爵), Titular office(本官), Salary rank(寄祿官), Sinecure post(祠祿官), Nominal supernumerary appointment(帖職), Concurrent appointment(兼職差遣); 관직명의 세부 내용은 다음 URL에서 확인할 수 있다. <https://docs.google.com/spreadsheets/d/1m0BYiFq0ZufdQGO93c4PwSM1N9T-Hc0bCt6unoxL14w/edit#gid=1766441073>.
 - 14 현재 CBDB에서는 112개의 관직 임명 방식 분류를 사용하고 있다. 관직 임명 방식 분류의 세부 내용은 다음 URL에서 확인할 수 있다. <https://docs.google.com/spreadsheets/d/1m0BYiFq0ZufdQGO93c4PwSM1N9T-Hc0bCt6unoxL14w/edit#gid=700932906>.
 - 15 현재 CBDB에서는 5개의 관직 부임 분류를 사용하고 있다. 관직 부임 분류의 세부 내용은 다음과 같다. 赴任(Assumed the office), 辭不就(Declined the office), 未赴任而卒(Died before assuming the office), 未赴任而改命(Appointment changed before assuming the office), 未赴任(Did not assume the office).

관계 정보는 혈연관계¹⁶와 사회관계¹⁷로 세분화되어 있다.

- 16 현재 CBDB에서는 471건의 혈연관계 정보를 제공하고 있다. 혈연관계 정보는 독자적인 기호체계를 도입한 코드를 통해서 471건의 분류를 제공하고 있다. 예를 들어서, 아들을 상징하는 기호는 “S”이다. 손자(孫)는 아들의 아들이기에 “SS”로, 증손자(曾孫)는 아들의 아들의 아들이기에 “SSS”를 부여한다. 남편은 “H”, 부인은 “W”, 첩은 “C”를 부여하여, 손자의 부인을 “SSW”와 같은 식으로 조합하여 표현할 수 있다. 또한 남자형제를 의미하는 기호는 “B”이고, 여자형제를 의미하는 기호는 “Z”인데, 이를 활용하여, 남자형제의 아들의 부인(姪媳婦)을 “BSW”로, 부인의 여자형제의 아들(妻之外甥)을 “WZS”를 부여한다. 그리고 숫자를 순서에 활용하여, 아들의 태어난 순서가 명확하다면, 장자(長子)는 S1, 차자(次子) S2와 같은 식으로 번호를 부여하고, 첫 번째 부인은 “W1”, 두 번째 부인은 “W2”와 같은 식으로 표현할 수 있다. 순서가 명확하지 않을 때에는 “+”와 “-” 기호를 사용하여, 나이가 많은 남자형제를 “B+”로, 나이가 어린 남자형제를 “B-”로, 부인의 나이가 많은 남자형제를 “WB+”로, 나이가 어린 여자형제의 남편을 “Z-H”로 서술한다. 혈연관계 정보의 세부 내용은 다음 URL에서 확인할 수 있다. <https://docs.google.com/spreadsheets/d/1m0BYiFq0ZufdQGO93c4PwSM1N9T-Hc0bCt6unoxL14w/edit#gid=84034231>.
- 17 현재 CBDB에서는 10개의 사회관계 대분류, 35개의 사관계 중분류, 475개의 사회관계 소분류를 사용하고 있다. 10개의 사회관계 대분류의 세부 내용은 다음과 같다. Associations (General)(社會關係(籠統)), Scholarship(學術關係類), Friendship(朋友關係類), Politics(政治關係類), Writings(著述關係類), Military(軍事關係類), Medicine(醫療關係類), Religion(宗教關係類), Family(家庭關係類), Finance(財務關係類); 35개의 사회관계 중분류의 세부 내용은 다음과 같다. Associations (General)(社會關係(籠統)), Association through common membership(同爲 … 之成員), Social Interactions(社會交際), Teacher-Student(師生關係), Intellectual Affiliations(學術交往), Association by Scholarly Topic(學術主題相近), Association through common membership(同爲 … 之成員), Academic Patronage(學術襄助), Literary and Artistic Affiliations(文學藝術交往), Intellectual Attacks(學術攻訐), Friendship(General)(朋友關係(籠統)), Politics (General)(政治關係(籠統)), Connection via office (equal)(官場關係(平級)), Connection via office (subordinate)(官場關係(下屬)), Connection via office (superior)(官場關係(上司)), Supportive political association(政治奧援), Recommendation and sponsorship(薦舉保任), Oppositional political association(政治對抗), Writings (General)(著述關係(籠統)), Commemorative Texts(記詠文字), Epitaphs(墓誌文字), Prefaces/Postfaces(序跋文字), Ritual Texts(禮儀文字), Biographical Texts(傳記文字), Explanatory Texts(論說文字), Mottos(箴銘文字), Correspondences(書札文字), Occasional Texts(應酬文字), Military (General)(軍事關係(籠統)), Supportive military association(軍事支持), Oppositional

CBDB의 기본 정보와 관직 정보는 종이 인물사전과 공구서의 설계와 크게 다르지 않다. 물론 기존의 인물사전이나 공구서에서는 종이 매체의 공간 제약으로 수십 개가 될 수도 있는 이칭 정보를 편집자의 선택에 따라서 중요하다고 판단하는 것을 선별했지만, 디지털 매체에서는 가능한 모든 정보를 입력하여 편찬한다는 구축 방법론상의 차이가 존재한다. 그러나 대다수의 ‘기본 정보’와 ‘관직 정보’는 해당 인물에 대한 핵심 내용을 중심으로 서술하였기에 종이 매체의 제약이 크게 느껴지지 않았다.

그에 비하여 관계 정보는 디지털 매체에서만 온전히 구현할 수 있는 내용이라고 볼 수 있다. 혈연관계 정보는 종이 매체의 인물사전에서는 지면의 한계상 온전히 표현하지 못하는 정보였다. 대부분의 종이 매체 인물사전에서는 혈연관계 정보를 생략하거나, 아버지, 할아버지, 외할아버지와 같은 중요 정보라고 판단되는 내용만 기술했다. 물론 인문학 연구자들은 족보와 같은 사료를 통해서 특정 인물에 대한 혈연 정보를 탐색했다. 그런데 족보조차도 대부분 직계혈통만을 기록하여, 혼인을 통해서 이루어지는 모계를 기준으로 하는 혈연 정보는 각기 다른 가문의 족보를 비교하여 검증할 수밖에 없었다.

혈연관계 정보는 족보를 대표로 하는 혈연관계 사료가 존재하지만, 사회관계 정보는 기본적으로 종이 매체 인물사전에서는 서술되지 않는 항목이다. 그러나 특정 인물의 스승-제자 관계, 편지를 주고받은 관계, 탄핵(彈劾) 관련 관계, 동일 당파 소속 관계, 상사-부하 관계 등의 다양한 사회관계는 인문학 연구에서 중요한 정보이다. 그리고 디지털 매체에서는 이러한 복잡

military association(軍事對抗), Medicine (General)(醫療關係(籠統)), Religion (General)(宗教關係(籠統)), Family (General)(家庭關係(籠統)), Finance (General)(財務關係(籠統)); 사회 관계 소분류의 세부 내용은 다음 URL에서 확인할 수 있다. <https://docs.google.com/spreadsheets/d/1m0BYiFq0ZufdQGO93c4PwSM1N9T-Hc0bCt6unoxL14w/edit#gid=1930353398>.

하고 방대한 사회관계를 서술할 수 있다.

CBDB에는 이백(李白)¹⁸의 사회관계 정보가 37 분류를 토대로 하는 299건이 등록되어 있다. 예를 들어서, 시문을 주고받은 관계¹⁹에 대해서 총 162건이 기록되어 있다. 해당 내용 편찬에서는 이백이 송약사(宋若思)에게 보낸 관계에 대해서 구체적인 시문 제목²⁰을 기록하기도 하고, 한유(韓愈)가 이백에게 시문을 보낸 정보에 대해서 “唐五代人交往詩索引”을 근거로 제시할 뿐, 시문 제목을 서술하지 않기도 한다. 이는 CBDB가 현재 교류 시문을 1개만 입력할 수 있게 설계되어 있기에 발생하는 현상으로 보이며, 추후 보완이 필요한 항목으로 보인다.

이러한 사회관계조차도 비록 번거롭고 많은 시간을 소요하게 되지만, 다양한 공구서와 사료 및 선행 연구를 통해서 연구자들은 당대 시문 교류에 대한 정보를 얻을 수는 있다. 그런데 CBDB의 최대 강점은 컴퓨터가 이해할 수 있는 방식으로 인물에 대한 기본 정보, 관계 정보, 관직 정보가 종합되어 있다는 점이다. 인간이 검색을 통해서 특정 인물에 대한 정보를 종합적으로 살펴볼 수 있다는 점도 유의미한 점이지만, 컴퓨터가 데이터를 통해서 인간으로서는 수행하기 어려운 일을 대신할 수 있다는 점이 보다 중요하다.

인간 사회는 수많은 사람들의 여러 다양한 관계들의 집합체이다. 인간 사회를 파악하기 위해서는 복합적인 관계들을 살펴보아야 한다. 그러나 모든 복합적인 관계들을 살펴보기에는 인간의 인식, 기억, 사유 능력에는 일정한 한계가 있다. 그러나 컴퓨터는 컴퓨터가 이해할 수 있는 방식으로

18 CBDB 온라인 입력 시스템, “이백(李白)”(<https://input.cbdb.fas.harvard.edu/basicinformation/32540/assoc>).

19 해당 인물을 중심으로 시문을 보냈을 때에는 “贈詩, 文” 관계, 시문을 받았을 “收到Y的贈詩, 文”으로 표현한다. 해당 관계들은 서로 역관계이고, 사회관계 대분류인 應酬文字(Occasional Texts)의 하위분류이다.

20 『全唐詩』(170) 「中丞宋公以吳兵三千赴河南 軍次尋陽 脫餘之囚 參謀幕府 因贈之」.

편찬된 데이터를 토대로 아무리 복잡하고 다양한 관계라도 ‘연산’을 할 수 있다.

예를 들어, CBDB 데이터를 토대로, 중국 역사 인물들의 평균 수명이나 12세기 중반의 453명의 성리학자가 서로 주고받은 2,717개의 편지를 근거로 하는 당대 성리학자들의 관계를 컴퓨터에게 질의할 수 있다. 만약 인간이 혼자 해당 과업을 수행하려고 했다면, 얼마의 시간이 소요될지 예측하기 어려울 정도로 지난한 작업이 될 것이다. 혹은 스승-제자 관계를 통해서, 수 대에 이르는 주자의 제자 계보도를 살펴볼 수도 있다.

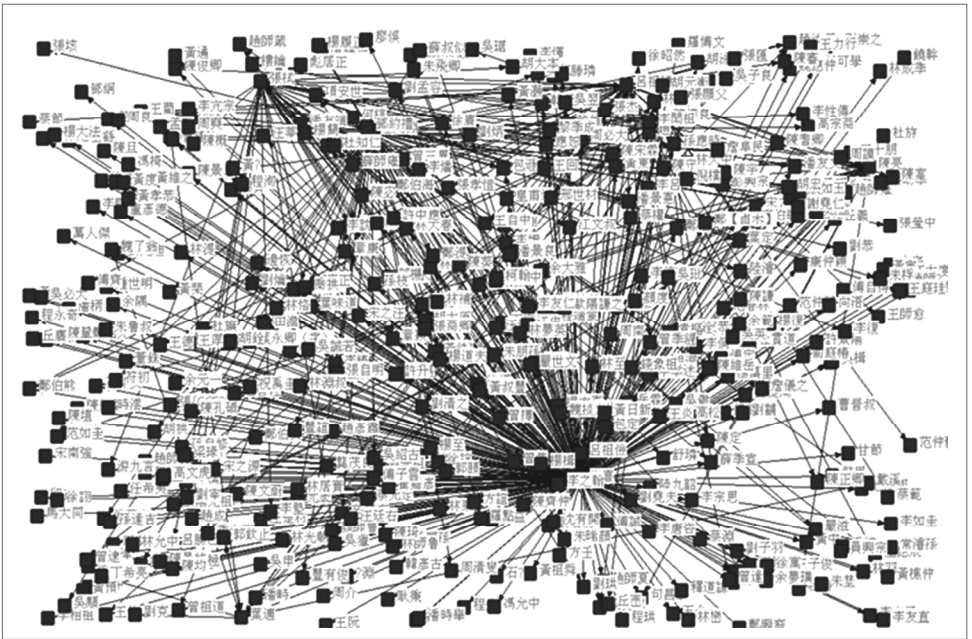


그림2- 성리학자 편지 교류 관계망²¹

21 Peter Bol, Jieh Hsiang, & Grace Fong, CBDB Abstract for Digital Humanities Hamburg 2012(http://projects.iq.harvard.edu/files/cbdb/files/cbdb_for_digital_humanities_hamburg_2012.pdf).

물론 인간도 동일한 작업을 수행할 수는 있다. 다만, 컴퓨터로는 1시간도 소요되지 않는 연산을 인간은 그 수백 배 이상의 시간을 들여야 동일하게 처리할 수 있을 것이다. 그런데 컴퓨터는 연산이 빠를 뿐 대상 관계 설정 및 해당 관계 연산 방법 및 그 의미에 대해서 스스로 판단하지 못한다. 그렇기에 컴퓨터는 분석을 위한 기획 자체가 불가능할 뿐만이 아니라, 분석을 위한 토대 데이터의 설계도 당연히 불가능하다. 따라서 연구자가 집중해야 할 부분은 컴퓨터의 빠른 연산을 활용하기 위한 데이터 설계와 구축이다. 그런 의미에서 CBDB는 다양한 인물 정보가 최대한 세밀하게 컴퓨터가 이해할 수 있는 방법으로 편찬되었다는 데 의의가 있다.

그런데 디지털 인물사전의 가장 큰 학문적 효용은 컴퓨터의 빠른 연산 활용보다 다양한 인물 정보를 판별하는 근거와 출처 및 연구 성과를 집대성한 것에 있다.

중이 인물사전의 인물 관련 정보들도 다양한 근거 자료를 통한 고증의 결과물이지만, 해당 고증 근거와 내용을 편찬 내용에 모두 수록할 수는 없었다. 그래서 인문학 연구 과정에서 해당 인물 관련 정보의 정확성을 판별하기 위하여, 관련 사료를 다시 살펴보아야만 했다. 그러나 CBDB에서는 정보 판단의 출처와 해당 내용을 서술하고 있다. 예를 들어, 이백(李白)²²은 송약사(宋若思)의 추천(推薦)을 받았는데, 이에 대해서 증공(曾鞏)의 『이태백 문집후서(李太白文集後序)』²³를 근거 자료로 제시하고 있다. 또한 이조(李昭)에게 선물을 받은 관계²⁴에 대한 근거²⁵나 두보(杜甫)와 같이 여행한 관계²⁶의

22 CBDB 온라인 입력 시스템, “이백(李白)”(<https://input.cbdb.fas.harvard.edu/basicinformation/32540/assoc>).

23 『李太白文集後序』 「明年 明皇在蜀 永王璘節度東南. 白時臥廬山 璘迫致之. 璘軍敗丹陽 白奔亡至宿松 坐繫尋陽獄. 宣撫大使崔渙與禦史中丞宋若思驗治白 以爲罪薄宜寬 而若思軍赴河南 遂釋白囚 使謀其軍事 上書肅宗 薦白才可用 不報」.

24 CBDB에서는 “受Y之贈物”로 해당 관계를 표현한다.

25 『李太白全集』(30) 「宣城長史弟昭贈餘琴溪中雙舞鶴 詩以見志」, “令弟佐宣城 贈餘琴溪鶴.

근거²⁷ 등을 모두 서술하고 있다. 또한 해당 관계에 대한 판단자와 판단 이력도 관리하고 있다.

이백에 대한 정보는 기본적으로 시앙쉬엔(Xiang Xuan, 相璇)에 의해서 구축되었는데, 최초의 만든 시점과 업데이트 시점을 기록하고 있다. 예를 들어, 이백의 혈연관계 정보는 기본적으로 시앙쉬엔이 직접 입력하고 수정한 내용이다. 그런데 이백 부인의 조부[妻之祖父]²⁸인 허어사(許圜師)와의 관계 정보는 교토대학교인문과학연구소(京都大學人文科學研究所)의 당대인물지식베이스(唐代人物知識ベース)²⁹의 내용을 2010년 12월 8일에 가져왔기에, 그에 대한 정보를 기록하고 있다.

다만 모든 역사 인물의 정보에 때때로 서로 다른 근거로 인한 모순이 발생한다. 그리고 연구자는 서로 다른 근거를 종합적으로 고려하여 ‘판단’한다. 때문에 연구자 간에 상이한 판단이 존재할 때가 종종 발생한다. 예를 들어, 두목(杜牧)³⁰의 몰년에 대해서는 852년, 853년, 857년과 같이 서로 다른 의견이 존재하고 있다. 그런데 CBDB에서는 853년 설을 따르고 있다. 다시 말해서, CBDB도 종이 사전의 전통에 따라서 단일한 최종 판단을 내리고 있다. 그런데 디지털 사전의 특징을 활용하게 되면, 서로 다른 연구자의 서로 다른 판단이 공존할 수 있기에 아쉬움이 남는 지점이라고 할 수 있다.

謂言天涯雪 忽向窗前落。白玉爲毛衣 黃金不肯博。當風振六翮 對舞臨山閣。顧我如有情 長鳴似相託。何當駕此物 與爾騰寥廓”。

26 CBDB에서는 “與Y旅遊”로 해당 관계를 표현한다.

27 『杜詩詳註』(16) 「昔遊」; 『杜詩詳註』(16) 「遺懷」.

28 CBDB는 해당 관계를 부인을 상징하는 “W”와 아버지를 상징하는 “F”를 조합하여 “WFF”로 부호화하였다.

29 Pers DB 唐代人物知識ベース(<http://tkb.zinbun.kyoto-u.ac.jp/pers-db>): 현재는 온라인 접속이 불가능한 상태임).

30 中國歷代人物傳記錄入系統, “두목(杜牧)”(<https://input.cbdb.fas.harvard.edu/basicinformation/21194/edit>).

특정 인물의 삶을 온전히 살펴보기 위해서는 방대한 사료와 저술의 바다에서 헤엄쳐야만 한다. 이는 연구자로서 마땅히 가져야할 마음가짐이라고 할 수도 있지만, 이제 더 이상 원고지에 펜으로 논문을 쓰지 않듯이 디지털 시대에는 디지털 시대의 새로운 방법이 필요하다. 디지털 매체에서는 서로 다른 학문적인 의견을 수용하여 편찬할 수 있다. 따라서 중요한 것은 다양한 전공의 연구자들의 생각을 수용하는 세밀한 데이터 설계 방법이며, 다양한 전공의 연구자들이 직간접적으로 참여하는 데이터 구축 방법이다. 그런데 다양한 전공의 연구자들이 참여하기 위해서는 사실상 구축 중인 데이터가 위키백과와 같이 실시간으로 모든 이에게 공유되어야 한다.

2. CBDB 데이터 공유

CBDB는 Rawdata(Access, SQLite), 웹서비스, API, LOD(Linked Open Data) 등의 다양한 형식으로 자신들의 모든 데이터를 공개하고 있다. 그 결과 해당 프로젝트를 직접적으로 언급하고 있는 논문만 360개에 이른다. Donald Sturgeon(德龍, 도널드 스테전) 개인이 구축한 CTEXT(Chinese Text Project)를 중국학 관련자들이 모두 사용하고 있는 것처럼, CBDB도 계속해서 방대한 데이터 집성을 근간으로 사용자층을 지속적으로 늘려나가고 있다.

CBDB의 RAWDATA는 RDB(Relational Database, 관계형 데이터베이스) 형식으로 구성하여 각각 마이크로소프트 Access 형식과 퍼블릭도메인³¹인

31 퍼블릭도메인(public domain)은 저작권(저작권재산권)이 소멸되었거나 저작자가 저작권을 포기한 저작물을 말하며, 저작권표시, 공유(복제, 배포, 전시 등), 변경, 상업적 사용 등의 모든 저작물 사용에 따른 조건이 없다.

SQLite 형식으로 홈페이지를 통해서 주정기적으로 배포되고 있다.³²

웹서비스는 ‘온라인 입력 시스템(Online Inputting System, 線上錄入系統)³³과 ‘온라인 질의 시스템(Online Query system(by Inindex project), 線上查詢系統(由引得項目提供))³⁴을 제공하고 있다. ‘온라인 입력 시스템’은 GUEST로 접속해서 기본적인 인물 검색을 통해서 인물의 종합 정보를 제공받을 수 있으며, 그 외에도 현행 분류 코드, 공간 코드, 저작 코드 등 다양한 부가 정보들을 제공한다. 그리고 ‘온라인 질의 시스템’은 기존의 CBDB 협력체가 중원짜이시엔지투안(中文在線集團)³⁵과 같이 구축한 서비스로, ‘온라인 입력 시스템’보다 인물에 대한 정보의 탐색이 사용자 친화적으로 만들어져 있다. 그뿐만 아니라 데이터를 토대로 공간 정보(GIS시각화), 시간 정보(타임라인), 언어 정보(워드클라우드) 및 다양한 인터랙티브 인포그래픽을 제공하고 있다.

또한 CBDB는 실시간으로 업데이트하는 CBDB의 정보를 온전히 전달하기 위해서 API³⁶를 제공³⁷한다. 비록 부정기적으로 RAWDATA를 제공하고는 있지만, CBDB를 기반으로 운영되는 서비스들에게 온전한 CBDB 데이터를 전달하기 위해서는 실시간으로 변경되는 데이터에 접근할 수 있는 방법이 필요하기 때문이다. 현재 CBDB API는 캐나다 맥길대학교(McGill University)의 ‘명청부녀저작(明清婦女著作, Ming Qing Women’s Writings)³⁸, 타이완 중앙

32 CBDB RAWDATA 다운로드 페이지(<https://projects.iq.harvard.edu/cbdb/download-cbdb-standalone-database>).

33 CBDB 온라인 입력 시스템(<https://input.cbdb.fas.harvard.edu>).

34 CBDB 온라인 질의 시스템(<https://inindex.com/biog>).

35 중원짜이시엔지투안(中文在線集團)은 2000년 칭화(清華)대학교에서 설립한 디지털 출판업체이다. 대표적인 서비스로는 “17K小說網”, “書香中國” 등이 있다.

36 API는 Application Programming Interface(애플리케이션 프로그램 인터페이스)의 약자로서 컴퓨터와 컴퓨터 프로그램 사이를 연결하는 기술적인 방법이다.

37 CBDB API에 대한 상세 설명은 다음 URL에 있다. <https://projects.iq.harvard.edu/cbdb/cbdb-api>.

연구원 역사언어연구소(中央研究院歷史語言研究所, Institute of History and Philology, Academia Sinica)의 ‘인명권위-인물전기자료고(人名權威-人物傳記資料庫)’³⁹, 네덜란드 라이덴대학교(Universiteit Leiden)의 ‘통신과 제국: 비교 관점에서 본 중국 제국(Communication and Empire: Chinese Empires in Comparative Perspective)’⁴⁰ 등에서 활용하고 있다.



그림3-MARKUS 실행 샘플⁴¹

예를 들어, 네덜란드 라이덴대학교에서 제공하는 고문 내에 등장하는 객체명(인물, 관직, 지명 등)에 대한 자동 식별 플랫폼인 MARKUS⁴²는 CBDB API를 통해서 자동으로 인물명을 식별하고, 동시에 오른쪽의 보조창에서

38 McGill University 명칭부녀저작(<https://digital.library.mcgill.ca/mingqing>).

39 Academia Sinica 인명권위-인물전기자료고(http://archive.ihp.sinica.edu.tw/ttsweb/html_name).

40 Universiteit Leiden 통신과 제국: 비교 관점에서 본 중국 제국(<https://chinese-empires.eu>).

41 『明史』(15) 本紀第十五 孝宗.

42 Universiteit Leiden MARKUS(<https://dh.chinese-empires.eu/markus>).

사용자들의 추가적인 인물 검색을 지원한다.⁴³

그런데 API는 단방향으로 정보를 제공하지만 한다는 한계가 있다. CBDB API는 비록 실시간으로 업데이트되고 있는 CBDB의 모든 인물 정보를 제공하지만, 반대로 다른 시스템의 데이터를 CBDB로 전달하지는 못한다. 예를 들어, 신라 최치원(崔致遠)은 CBDB에서도 기년편고(紀年便考)를 근거로 등록되어 있다.⁴⁴ 하지만 최치원에 대한 정보는 성명(중국어, 영어, 한국어), 성별, 신라 소속이라는 정보 이외에 다른 정보가 존재하지 않고 있다. 그에 반하여 한국에서는 당연히 최치원의 자인 고운(孤雲), 해운(海雲)을 비롯한 다양한 인물 정보가 편찬되어 있다. 그러나 CBDB API에서는 자신의 정보를 전달할 뿐, 반대로 정보를 전달 받을 수는 없다.

이에 양방향으로 정보를 전달하기 위해서 나온 기술이 시맨틱 데이터를 토대로 하는 LOD(Linked Open Data) 개념이다. 시맨틱 데이터는 컴퓨터가 이해할 수 있는 의미를 내재한 데이터라는 의미로, 현재에는 주어-술어-목적어(SVO) 형식을 기본으로 하는 RDF를 근간으로 하고 있고, 이를 토대로 구축한 양방향으로 정보를 제공하는 온라인 서비스가 LOD라고 할 수 있다.⁴⁵ 그리고 CBDB LOD⁴⁶는 2017년부터 상하이도서관에서 서비스하고 있다. 다만 CBDB LOD의 서비스가 불안정한 측면이 있다.

하지만 상하이도서관의 CBDB LOD 서비스나 중원짜이시엔지투안의

43 MARKUS에 대한 보다 상세한 정보는 다음 논문을 참고할 수 있다. Hilde De Weerd, "Creating, Linking, and Analyzing Chinese and Korean Datasets: Digital Text Annotation in MARKUS and COMPARATIVUS," *Journal of Chinese History 中國歷史學刊*, Vol. 4, Special Issue. 2(2020) pp. 519~527.

44 CBDB 온라인 입력 시스템, "최치원(崔致遠)"(<https://input.cbdb.fas.harvard.edu/basicinformation/551626/edit>).

45 시맨틱 데이터에 대한 상세한 내용은 다음 논문을 참조할 수 있다. 김바로, 「인공지능 시대의 사전과 데이터: 한국역사인물사전(데이터)을 위한 탐색」, 『동양학』 제84호(2021), 226~229쪽.

46 Shanghai Library CBDB LOD(<https://cbdb.library.sh.cn>).

CBDB ‘온라인 질의 시스템’은 일종의 CBDB의 주도권을 중국 대륙으로 가져오기 위해서 노력한 결과물이라고 할 수 있다. 디지털 시대에 데이터의 공개는 자신의 것을 빼앗기는 것이 아니라, 오히려 독점적인 지위를 얻기 위한 발판이 되기 때문이다.

예를 들어, 현재 국사편찬위원회에서 서비스 중인 ‘디지털 조선왕조실록’은 이미 한국학 연구자들에게는 없어서는 안 될 서비스이다. 그리고 공공데이터법에 따라서 국민의 세금이 들어가는 공공기관인 국사편찬위원회가 구축하고 유지 보수하는 ‘디지털 조선왕조실록’은 공공데이터포털을 통해서 그 토대 데이터를 모두 공개하고 있다. 그리고 개인 혹은 집단 누구나 ‘디지털 조선왕조실록’의 토대 데이터를 바탕으로 연구는 물론 상업적인 서비스도 구축할 수 있다.⁴⁷ 이는 국사편찬위원회가 ‘디지털 조선왕조실록’을 빼기는 것이 아니다. ‘디지털 조선왕조실록’을 가장 잘 유지-보수할 수 있는 집단은 어디까지나 국사편찬위원회이며, 따라서 ‘디지털 조선왕조실록’에 대한 수요가 많으면 많을수록 국사편찬위원회는 ‘디지털 조선왕조실록’을 유지-보수해야 하는 의무를 가짐과 동시에 ‘디지털 조선왕조실록’에 대한 권리를 가지게 된다. 이처럼 인문지식 데이터를 공유하는 것은 자신의 노력을 타인에게 빼기는 것이 아니라, 자신의 피와 땀이 온전하게 인정받는 가장 빠른 길이 된다는 점을 인지해야 할 것이다.

47 공공데이터법과 인문데이터 공개에 대한 자세한 내용은 다음 논문을 참조할 수 있다. 김바로, 「〈공공데이터법〉과 인문데이터: 공공기관 보유 인문데이터 공개 신청 사례를 중심으로」, 『한국고전연구』 제57호(2022), 167~192쪽.

Ⅲ. 한국역사인물 관련 선행 정보시스템

CBDB는 단일 정보 시스템 내에서 중국 역사 인물에 대한 인명 정보, 생몰년 정보, 지리 정보, 관력 정보, 친족 정보, 사회관계 정보 등을 모두 제공하고 있다. 그런데 한국에는 아직 CBDB에 견줄 만한 종합적인 한국역사인물 정보 시스템은 존재하지 않는다. 다만 대다수의 역사인물 정보 시스템이 국가 지원 사업으로 구축되어 유지되고 있기에, 2013년에 제정된 ‘공공데이터의 제공 및 이용 활성화에 관한 법률(공공데이터법)’에 의해서 RAWDATA가 공개되어 있다. 하지만 각각의 역사인물 정보 시스템은 여전히 분리되어 있으며, 인문학 연구자들은 여전히 개별 역사인물 정보 시스템의 웹서비스를 통해서 역사인물 정보를 파악하고 있다.

1. 한국학중앙연구원 한국역대인물종합정보시스템

한국역대인물종합정보시스템(Korean Historic Figures Database of the Academy of Korean Studies)은 2005년부터 한국학중앙연구원에서 『한국인물대사전』과 기존에 한국학중앙연구원 자체적으로 제작한 방목데이터를 수정·보완하여 13만 8,000여 명의 인물 및 『민족문화대백과사전』의 인물을 포함하여 XML을 기반으로 구축한 한국의 역사 인물에 대한 데이터베이스이다.⁴⁸ 한국역대인물종합정보시스템의 데이터 스키마는 기본적으로 시험 정보와 인물 정보로 나뉘어 있다. 인물 정보는 인명 정보, 생·몰년 정보, 관력 정보, 지리 정보, 친족 정보로 이루어져 있다. 인명 정보는 성, 성명,

48 한국학중앙연구원 한국역대인물종합정보시스템(<https://people.aks.ac.kr>).

본명, 초명, 개명, 일명, 자, 호, 시호, 봉호, 군호, 묘호, 법명, 기타로 속성을 분류했다. 생·몰년 정보는 생년, 연호, 세기, 몰년으로 하위 엘리먼트를 구성했다. 관력 정보는 전력, 증직, 품계, 관서, 관직, 직력, 봉호, 과거, 타과로 세부 정보를 나누었다. 지리 정보로는 본관과 거주지 정보를 기술했다. 친족 정보는 방목자료의 특성상 부계자료를 중심으로 속성을 분류했다.⁴⁹

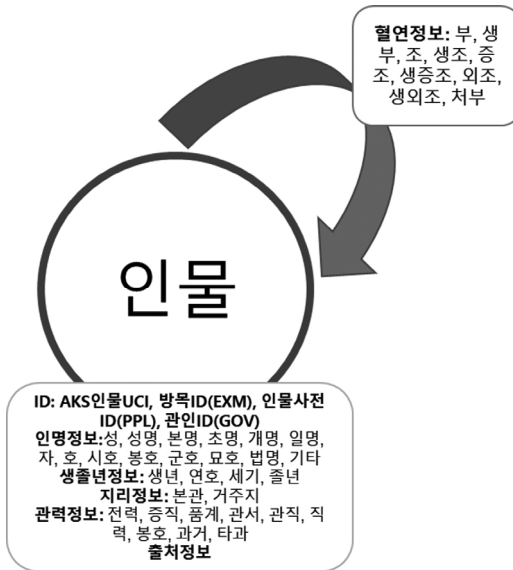


그림4- 한국학중앙연구원 한국역대인물종합정보시스템 온톨로지 개념도

2. 국사편찬위원회 조선왕조실록 부가열람

조선왕조실록 부가열람은 국사편찬위원회에서 『조선왕조실록』에 등장하는 인물 정보 11만 5,871명과 관직 재직 정보 33만 6,268건을 대상으로

49 김바로, 『시맨틱 데이터 아카이브의 구축과 활용』(파주: 보고사, 2018), 96~97쪽.

구축한 데이터이다. 인물 정보는 『조선왕조실록』에 등장하는 거의 모든 인물의 본명 및 이칭 정보 등으로 구성했다. 관직재직 정보는 현직, 전직1, 전직2(사망상태), 증직으로 구분하여 입력하였다. 『조선왕조실록』 인물부 가정보는 각각의 인물을 식별하고, 식별된 인물에 대한 한국학중앙연구원의 인물 UCI가 있을 경우 UCI 정보를 기술했다. 이에 따라서 대상 인물에 대한 확장적인 정보를 제공할 수 있는 토대를 만들었다.⁵⁰

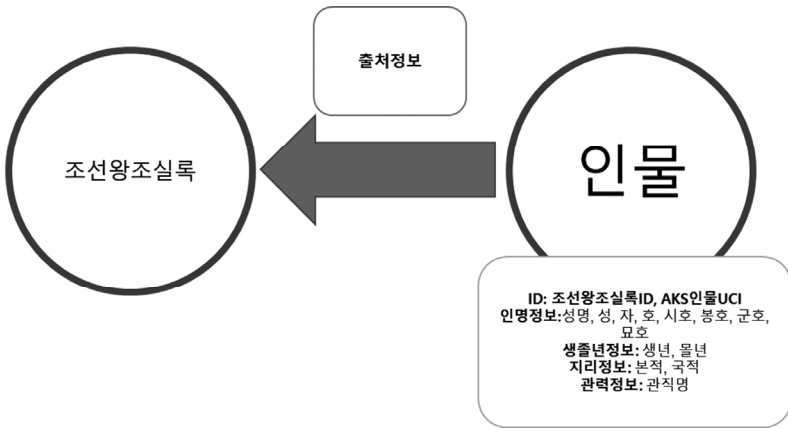


그림5- 국사편찬위원회 조선왕조실록 부가열람 온톨로지 개념도

3. 한국고전번역원 한국고전종합DB

한국고전종합DB는 한국고전번역원에서 진행한 한국문집총간 전산화사업(1997년), 고전문헌 CD-ROM 간행사업(1998~2000년), 지식정보자원관리사업(국가DB사업)(1999~2014년), 고전번역서 생성체계 구축사업(2006~

50 김바로, 앞의 책(2018), 104쪽.

2012년), 한국고전종합DB구축 사업(2013년~), 한국고전종합DB 서비스 개선 및 DB콘텐츠 개발(2016년~)의 성과가 집대성되어 있는 한국고전 종합 시스템이다. 그중에서 2018년부터 기획하여 2021년 1월부터 서비스를 시작한 ‘한국고전종합DB 인물관계정보서비스’에서는 기존 한국고전종합DB의 한국문집충간 해제, 한국문집충간 편목색인, 고전용어 시소러스, 고전번역서 각주 정보 등에 산재되어 있던 인물의 기본정보와 혈연, 혼맥, 문인, 교유 관계 등 인물 정보를 추출하고 정리하여, 사회네트워크시각화 방식을 활용하여 서비스하고 있다.⁵¹

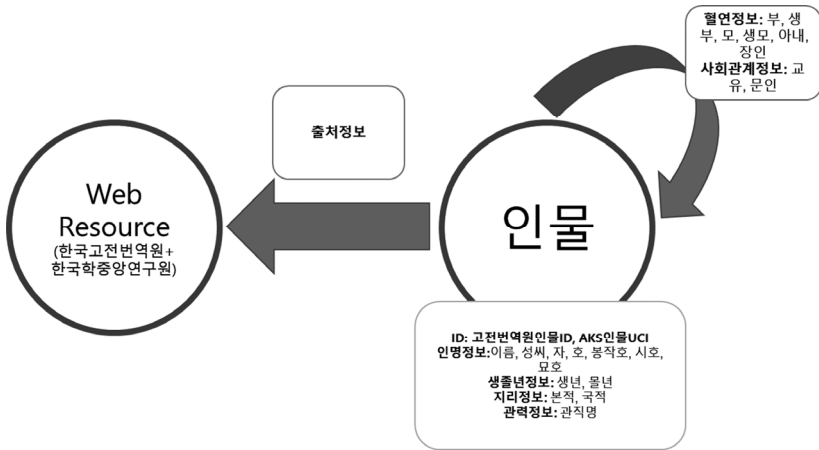


그림6- 한국고전번역원 한국고전종합DB 인물관계정보서비스 온톨로지 개념도

51 한국고전번역원 고전정보센터, 「한국고전종합DB 구축의 성과와 과제」, 『민족문화』, 제59권(2021), 10~15쪽.

4. 선행 역사인물 정보시스템의 스키마 비교

표1-선행 역사인물 정보시스템 스키마 비교표

	한국역대인물 종합정보시스템	조선왕조실록 부가열람	한국고전종합DB
ID	한국학중앙연구원 역사인물UCI	한국학중앙연구원 역사인물UCI	한국학중앙연구원 역사인물UCI
인명 정보	성, 성명, 본명, 초명, 개명, 일명, 자, 호, 시호, 봉호, 군호, 묘호, 법명, 기타	성명, 성, 자, 호, 시호, 봉호, 군호, 묘호	이름, 성씨, 자, 호, 봉작호, 시호, 묘호
생몰년 정보	생년, 연호, 세기, 몰년	생년, 몰년	생년, 몰년
지리 정보	본관, 거주지	본적, 국적	본관, 국적
관력 정보	전력, 증직, 품계, 관서, 관직, 직력, 봉호, 과거, 타과	관직(실록출현일)	
친족 정보	부, 생부, 조, 생조, 증조, 생증조, 외조, 생외조, 처부		부, 생부, 모, 생모, 아내, 장인
사회관계 정보			교유, 문인
기타 정보	출처 정보	직업, 출처 정보(실록)	출처 정보

한국역대인물종합정보시스템은 방목을 근간으로 구축되었기에 기본정보를 중심으로 입사 정보와 주요 관직 정보 및 친족 정보를 제공하고 있다. 그러나 교유, 학연과 같은 사회관계 정보는 제공하지 못하고 있다. 또한 데이터의 근간이 되는 출처 정보 자체는 제공하고 있으나, 해당 출처의 내용을 직접 볼 수 있는 서비스를 제공하지는 못하고 있다.

조선왕조실록 부가열람은 『조선왕조실록』에 등장하는 인물들의 관직 정보를 정리한 것이다. 비록 『조선왕조실록』에 등장하는 인물의 관직 정보가 해당 인물의 모든 관직 정보를 포괄하지 못하는 한계가 있으나, 현재 공개된 역사인물 데이터 중에서 가장 완전한 관력 정보를 제공하고 있다고 할 수 있다. 또한 디지털 『조선왕조실록』을 토대로 하였기에 해당 관력

정보의 출처에 직접적으로 접근할 수 있는 장점이 있다.

한국고전종합DB 인물관계정보서비스의 문집을 중심으로 구축되어 있기에 인물들의 교유와 문인 관계와 같은 사회관계 정보를 포함하고 있으며, 개별 정보들의 근거로서 한국고전번역원 자체적으로 구축한 시소러스, 고전번역서, 편목색인, 한국문집총간해제집, 한국학중앙연구원 한국역대인물 종합정보시스템의 URL을 활용하고 있다. 다만 인물들의 관력 정보는 ‘한국문집총간 저자행력정보’ 서비스를 통해서 일부 제공하고는 있으나, 기계처리가 가능한 형태의 데이터로 이루어져 있지는 않다.

한국역대인물종합정보시스템과 조선왕조실록 부가열람 서비스 및 한국고전종합DB 인물관계정보서비스는 자체적인 인물ID를 부여하고 있는 동시에 해당 인물이 한국역대인물종합정보시스템에 등록되어 있는 경우에는 해당 한국학중앙연구원 역사인물UCI를 기술해주고 있어, 서로 다른 성격의 세 가지 데이터가 상호 연결될 수 있는 단초가 된다.

하지만 기존의 역사인물 정보시스템은 각 기관에서 서비스를 진행하고 있기에 새로운 시도를 하기에 적절하지 않은 측면이 있다. 한편 아직도 많은 역사인물 정보가 디지털로 이식조차 되지 못하고 있기에, 새로운 역사인물 데이터의 구축도 필요한 실정이다. 그런데『홍문록』은 조선시대 주요 인물의 정보가 있기에 각 기관에 산재해 있는 인물 정보와의 연결이 보장되어 있지만, 내용 자체는 비교적 단순한 인물 명단이기 때문에 학계에 큰 관심을 받지 못하고 있었다. 이에, 역사인물 데이터를 축적하고, 나아가 서로 흩어져 있는 역사인물 정보를 통합하기 위한 연구⁵²의 시작점으로『홍문록』을 선정했다.

52 본 논문은 개념적 탐색을 수행한 한국학중앙연구원 2022년 공동연구과제 “조선시대 『홍문록(弘文錄)』과 집권세력의 네트워크(AKSR2022-C14)”의 결과물이며, 데이터 상호 연결의 실증을 위하여 2023년 “조선후기 『정록(政錄)』의 자료적 가치와 음관(陰官)의 인사 실태(AKSR2023-C06)”를 진행하고 있다.

IV. 『홍문록』 온톨로지 설계

1. 『홍문록』의 정보 구조

조선시대 자문기구의 하나였던 홍문관(弘文館)의 관원 후보자의 명부인 『홍문록』은 홍문관 참서관 이상의 관원이 이른바 ‘언론 삼사(三司)’로서 사헌부, 사간원 관원과 함께 ‘청요직’으로 통칭되면서 정치적으로 중요한 인물의 선별에 관여하는 당대의 정치 엘리트 집단이라는 점에서 중요한 사료라고 할 수 있다.⁵³

그럼에도 불구하고, 『홍문록』에 관한 연구는 『홍문록』 자체에 서술되어 있는 정보의 부족으로 큰 주목을 받지 못하고 있다. 따라서 발전한 정보 기술을 활용하여, 『홍문록』에 기재된 인물의 관력 정보, 친족 정보, 사회관계 정보들을 온전히 살펴볼 수 있게 된다면 조선시대 정치 엘리트 집단을 연구하는 데 유용할 것이다.

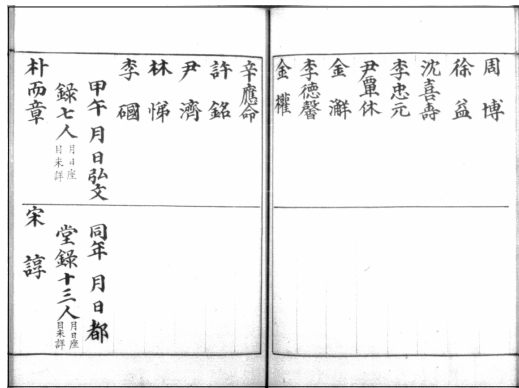


그림7- 『홍문록』 샘플 이미지(규장각 登瀛錄(997))

53 보다 구체적인 『홍문록』에 관한 역사학적 연구는 『한국학』 이번 호(46-3)에 게재된 『홍문록』 기획논문들을 참고해 주시기 바란다.

『홍문록』의 정보 요소는 크게 기본 정보, 명단 정보 및 좌목 정보로 구분한다. 기본 정보는 권수, 쪽수, 내용 구분으로, 권호는 『등영록』의 1, 2, 3, 4 권수이며, 쪽수는 해당 정보의 페이지 수, 내용 구분은 집현전학사록(集賢殿學士錄), 홍문록, 홍문록 좌목, 홍문록 명단, 도당록(都堂錄) 좌목, 도당록 명단, 본관록(本館錄) 좌목, 본관록 명단이다. 「본관록」은 홍문관 관원들의 자체 추천 명단이며, 「도당록」은 「본관록」을 대상으로 의정부와 이조의 당상이 모여서 작성한 최종 후보자 명단이다. 명단 정보는 한자이름, 평가점수로, 평가점수는 권점(圈點) 혹은 3~10의 점수로 기술되어 있다. 좌목 정보는 관직명, 한자이름으로 구성되어 있다.

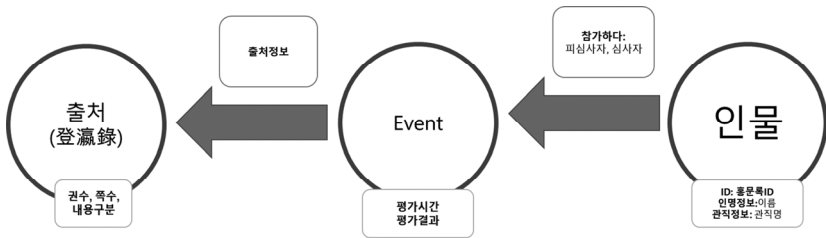


그림8-『홍문록』의 정보구조 개념도

『홍문록』에는 당대 정치엘리트들의 명단과 그 엘리트를 평가했던 심사자에 대한 정보가 존재하지만, 해당 인물의 혈연 정보나 사회관계 정보가 존재하지 않는다. 이에, 선행 역사인물 정보시스템과의 연결을 통해서, 당대 인물의 종합적인 정보를 연구자들이 온전히 획득할 수 있도록 할 필요가 있다.

2. 『홍문록』을 중심으로 하는 역사인물정보 온톨로지 설계

1) 네임스페이스

Prefix	Namespace	비고
kdbd	http://dh.aks.ac.kr/ontology/kdbd	한국역사인물데이터베이스 네임스페이스
aks	http://aks.ac.kr/ontology/	(가정) 한국학중앙연구원 네임스페이스
nikh	http://lod.koreanhistory.or.kr/	국사편찬위원회 네임스페이스
itkc	https://www.itkc.or.kr/ontology/	(가정) 한국고전번역원 네임스페이스
dcterms	http://purl.org/dc/terms/	기본 용어 사용을 위한 더블링크어
foaf	http://xmlns.com/foaf/0.1/	행위자(사람, 그룹) 설정을 위한 FOAF
skos	http://www.w3.org/2004/02/skos/core#	개념 클래스 정의를 위한 SKOS

2) 클래스 설계

Class_1단계	Class_2단계	명칭(rdfs:label)	description
foaf:Agent		행위자	모든 행위자를 의미
	foaf:Organization	조직	조직을 대표하는 개념
	foaf:Person	사람	사람, 인물을 표현
kdbd:Event		사건	행위자의 특정 시공간상에서의 행위
kdbd:Resource		자원	온오프라인의 모든 출처 정보
	bibo:Book	책	문헌자원의 출처정보
	kdbd:Webresource	웹자원	웹자원의 출처 정보
kdbd:Concept		개념	
	kdbd:OfficeName	관직명	관직명 정보
	kdbd:EventType	사건분류	사건분류

3) 데이터 프로퍼티 설계

Property_1단계	Property_2단계	Property_3단계	Property 명칭	Domain	Range
dcterms:identifier				owl:Thing	xsd:string
	kbdb:hasAKSUCI		한국학중앙연구원 인물UCI	owl:Thing	xsd:string
		kbdb:hasEXMID	한국학중앙연구원 방목ID	owl:Thing	xsd:string
		kbdb:hasPPLID	한국학중앙연구원 인물사전ID	owl:Thing	xsd:string
		kbdb:hasGOVID	한국학중앙연구원 관인ID	owl:Thing	xsd:string
	kbdb:hasNIKHID		국사편찬위원회ID	owl:Thing	xsd:string
	kbdb:hasITKCID		한국고전번역원ID	owl:Thing	xsd:string
dcterms:title			이름	owl:Thing	xsd:string
	foaf:familyName		성	foaf:Person	xsd:string
	kbdb:hasGivenName		명	foaf:Person	xsd:string
dcterms:alternative			이칭	owl:Thing	xsd:string
	kbdb:hasChildhoodName		초명	foaf:Person	xsd:string
	kbdb:hasAliasName		표자	foaf:Person	xsd:string
	kbdb:hasNickname		호	foaf:Person	xsd:string
		kbdb:hasPosthumousEpithet	시호	foaf:Person	xsd:string
		kbdb:hasTitleByEmperor	봉호	foaf:Person	xsd:string

Property_1단계	Property_2단계	Property_3단계	Property 명칭	Domain	Range
		kdbb:hasPosthumousTitleOfEmperor	묘호	foaf:Person	xsd:string
		kdbb:hasTitleByEmperorForRoyalFamily	군호	foaf:Person	xsd:string
	kdbb:hasBudhistName		법명	foaf:Person	xsd:string
kdbb:hasBirthTime			생년	foaf:Person	xsd:date
kdbb:hasDeathTime			몰년	foaf:Person	xsd:date
kdbb:hasBirthAddress			출생지	foaf:Person	xsd:string
kdbb:hasBurialAddress			사망지	foaf:Person	xsd:string
kdbb:hasBasicAffiliation			본관	foaf:Person	xsd:string
kdbb:hasNationality			국적	foaf:Person	xsd:string
kdbb:hasBibliography			문헌	owl:Thing	xsd:string
kdbb:hasWebResource			웹자원	owl:Thing	xsd:string
kdbb:hasEventTime				kdbb:Event	xsd:date
kdbb:hasEventPlace				kdbb:Event	xsd:string

4) 오브젝트 프로퍼티 설계

Property_1단계	Property_2단계	Property_3단계	Property 명칭	Domain	Range
owl:sameAs			동일하다	owl:Thing	owl:Thing
dcterms:hasPart			포함 부분	owl:Thing	owl:Thing
dcterms:isPartOf			전체 부분	owl:Thing	owl:Thing
kdbb:isLineageKinOf			혈족	foaf:Person	foaf:Person
	kdbb:hasFather		아버지	foaf:Person	foaf:Person
		kdbb:hasBiologicalFather	생부	foaf:Person	foaf:Person
	kdbb:hasMother		어머니	foaf:Person	foaf:Person
		kdbb:hasBiologicalMother	생모	foaf:Person	foaf:Person
	kdbb:hasSon		아들	foaf:Person	foaf:Person
		kdbb:hasBiologicalSon	친아들	foaf:Person	foaf:Person
		kdbb:hasAdoptedHeir	양자	foaf:Person	foaf:Person
	kdbb:hasDaughter		딸	foaf:Person	foaf:Person
	kdbb:hasGrandfather		조부	foaf:Person	foaf:Person
		kdbb:hasBiologicalGrandfather	생조부	foaf:Person	foaf:Person
	kdbb:hasGreat-grandfather		증조부	foaf:Person	foaf:Person
		kdbb:HasBiologicalGreat-grandfather	생증조부	foaf:Person	foaf:Person

Property_1단계	Property_2단계	Property_3단계	Property 명칭	Domain	Range
	kdbb:hasMaternalGrandfather		외조부	foaf:Person	foaf:Person
		kdbb:hasBiologicalMaternalGrandfather	생외조부	foaf:Person	foaf:Person
	kdbb:hasWife		처	foaf:Person	foaf:Person
	kdbb:hasConcubine		첩	foaf:Person	foaf:Person
kdbb:isAffinalKinOf			인척	foaf:Person	foaf:Person
	kdbb:hasFatherInLaw		장인	foaf:Person	foaf:Person
	kdbb:hasMotherInLaw		장모	foaf:Person	foaf:Person
	kdbb:hasSonInLaw		사위	foaf:Person	foaf:Person
	kdbb:hasDaughterInLaw		며느리	foaf:Person	foaf:Person
kdbb:hasSocialRelations			사회관계	foaf:Person	foaf:Person
	foaf:knows		교우	foaf:Person	foaf:Person
	kdbb:hasScholarship		학문관계	foaf:Person	foaf:Person
		kdbb:hasMaster	스승	foaf:Person	foaf:Person
		kdbb:hasDisciple	제자	foaf:Person	foaf:Person
		kdbb:hasScholarlyMember	동종학문(동학, 동문)	foaf:Person	foaf:Person
hasEventObject			사건주체	kdbb:Event	owl:Thing
hasEventPreObject			사건전대상	kdbb:Event	owl:Thing

Property_1단계	Property_2단계	Property_3단계	Property 명칭	Domain	Range
hasEvnetPost Object			사건후대상	kdbb:Event	owl:Thing
hasEventType			사건분류	kdbb:Event	kdbb:Event Type
hasPreEvent			선행사건	kdbb:Event	kdbb:Event
hasPostEvent			후행사건	kdbb:Event	kdbb:Event

V. 맺음말

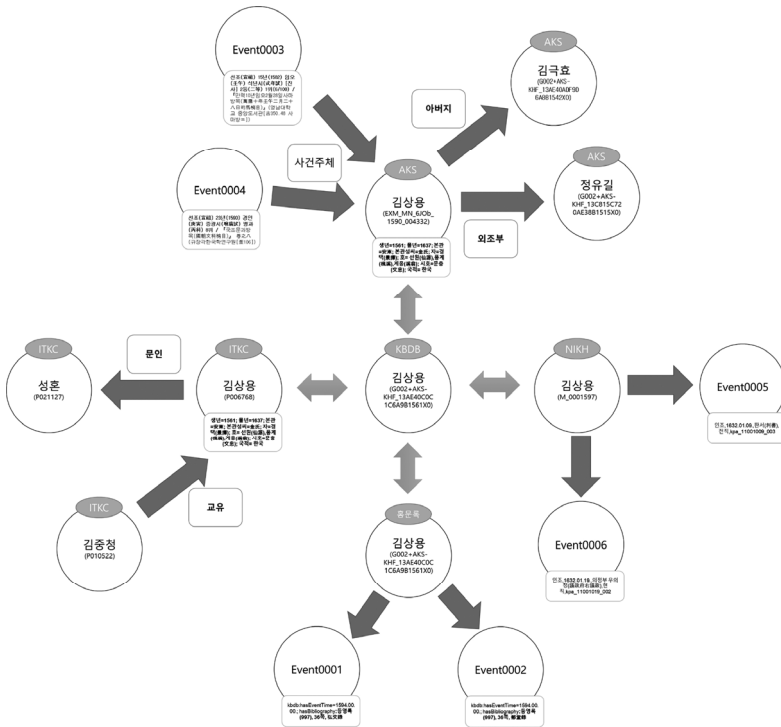


그림9-KBDB 한국 역사인물정보 네트워크 샘플:
김상용(金尙容, G002+AKS-KHF_13AE40C0C1C6A9B1561X0).

〈그림9〉는 본문에서 설계한 온톨로지를 토대로 “김상용(金尙容)”을 중심으로 『홍문록』, 한국역대인물종합정보시스템, 조선왕조실록 부가열람 서비스 및 한국고전종합DB 인물관계정보서비스의 정보를 상호 연계한 개념도이다. 『홍문록』 “김상용”의 수록정보와 채집 정보, 한국역대인물종합정보시스템 “김상용”의 과거시험과 친족 정보, 조선왕조실록 부가열람 서비스 “김상용”의 관직 정보, 한국고전종합DB 인물관계정보서비스 “김상용”의 사회관계 정보가 디지털에서 상호 연결되어 컴퓨터가 연산 가능하다.

물론 이미 현존하는 각각의 서비스를 사용하여 수많은 연구자들이 특정 인물에 대한 정보를 취득하고 상호 비교하여 인물의 생애를 재구하고 있다. 그렇기에 컴퓨터가 처리 가능한 방식으로 정보를 연결하는 것에 대한 효용에 대해서 회의적인 시각이 있을 수도 있다. 하지만 『조선왕조실록』의 디지털화를 통해서 방대한 『조선왕조실록』을 기존과는 다른 방법으로 효율적으로 탐색할 수 있게 되었다. 이러한 점을 상기하면, 컴퓨터가 처리 가능한 형태로 이루어지는 역사인물 데이터베이스의 상호 연계를 통해서 한국학 연구의 효율성이 증대될 것으로 예상할 수 있다.

무엇보다 각기 다른 기관에서 구축한 데이터가 상호 연계되어 방대한 정보망을 구성할 수 있기에, 그동안 각각의 학문 세부 분과 영역에서 탐색해 왔던 전문적인 인물 정보들이 각각의 독자성을 가지면서도 상호 연계될 수 있다. 이와 같은 차원에서 미래 세대에게 온전한 당대의 연구 결과를 전달해야 한다는 사명의식을 가지고 깊이 생각해보아야 할 것이다.

참고문헌

1. 1차 자료

- 『登瀛錄』, 『弘文錄』.
『李太白文集後序』, 『李太白文集後序』, 『杜詩詳註』.
『明史』.

2. 단행본

- 김바로, 『시맨틱 데이터 아카이브의 구축과 활용』. 파주: 보고사, 2018.
김현, 『인문정보학의 모색』. 성남: 북코리아, 2012.
김현·임영상·김바로, 『디지털 인문학 입문』. 서울: 한국외국어대학교 지식출판원, 2016.

3. 논문

- 김바로, 「인공지능 시대의 사전과 데이터: 한국역사인물사전(데이터)을 위한 탐색」, 『동양학』 제84호, 2021, 219~240쪽.
_____, 「〈공공데이터법〉과 인문데이터: 공공기관 보유 인문데이터 공개 신청 사례를 중심으로」, 『한국고전연구』 제57호, 2022, 167~192쪽.
배숙희, 「디지털 인문학과 송대사 연구: 1163년의 科學와 樓鑰를 중심으로 본 사회적 네트워크」, 『동양사학연구』 제146호, 2019, 157~191쪽.
조원희, 「시론: 네트워크 분석을 통해 본 아유르바르와다 카안(元仁宗, 재위 1311-1320) 시기의 元 朝廷」, 『중앙아시아연구』 제22권 제2호, 2017, 73~96쪽.
_____, 「네트워크 분석을 통해 본 元帝國 성종-무종-인종 시기 朝廷: 몽골의 '한화'를 반박하는 또 하나의 방법론」, 『중국학보』 제99호, 2022, 165~192쪽.
한국고전번역원 고전정보센터, 「한국고전종합DB 구축의 성과와 과제」, 『민족문화』 제59권, 2021, 7~57쪽.
鄧君·常嚴予·孫紹丹·宋先智·鍾楚依, 「數字人文視閔下清代科舉進士群體時空網絡結構分析」, 『現代情報』 2022-2, 2022, 4~13쪽.
李宗翰·鄭莉, 「家族, 婚姻與道學: 《仙溪志·人物傳》中的社會關係」, 『唐宋歷史評論』 2017-0, 2017, 33~45쪽.
錢超峰·杜德斌, 「北宋官僚家族網絡的空間結構及其演化: 基於CBDB和CHGIS的考察」.

『歷史地理研究』 2019-2, 2019, 83~94쪽.

包弼德, 「群體, 地理與中國曆史:基於CBDB和CHGIS」. 『量化歷史研究』 2017-Z1, 2017, 213~246쪽.

包弼德·王宏蘇·傅君勳·陳松·柳舟·朱厚權, 「中國曆代人物傳記資料庫'(CBDB)的歷史, 方法與未來」. 『數字人文研究』 2021-01, 2021, 21~33쪽.

De Weerdt, Hilde, "Creating, Linking, and Analyzing Chinese and Korean Datasets: Digital Text Annotation in MARKUS and COMPARATIVUS." *Journal of Chinese History* 中國歷史學刊, Vol. 4, Special Issue. 2, 2020, 519~527쪽.

4. 웹사이트

국사편찬위원회 한국역사정보통합시스템(<https://www.koreanhistory.or.kr>)

한국학중앙연구원 한국역대인물종합정보시스템(<https://people.aks.ac.kr>)

Academia Sinica 인명권위: 인물전기자료고(http://archive.ihp.sinica.edu.tw/ttsweb/html_name).

Bol, Peter, Hsiang, Jieh, Fong, Grace, CBDB Abstract for Digital Humanities Hamburg 2012(http://projects.iq.harvard.edu/files/cbdb/files/cbdb_for_digital_humanities_hamburg_2012.pdf)

CBDB 온라인 입력 시스템(<https://input.cbdb.fas.harvard.edu>).

CBDB 온라인 질의 시스템(<https://inindex.com/biog>)

CBDB 핵심 스키마 정리표 모음(<https://docs.google.com/spreadsheets/d/1m0BYiFq0ZufdQGO93c4PwSM1N9T-Hc0bCt6unoxL14w/edit?usp=sharing>).

Harvard University CBDB 홈페이지(<https://projects.iq.harvard.edu/cbdb>)

Harvard University General Introduction to CBDB 201404(https://projects.iq.harvard.edu/files/cbdb/files/general_introduction_to_cbdb_2014.pptx)

McGill University 명칭부너저작(<https://digital.library.mcgill.ca/mingqing>)

Pers DB 唐代人物知識ベース(<http://tkb.zinbun.kyoto-u.ac.jp/pers-db>): 현재는 온라인 접속이 불가능한 상태.

Shanghai Library CBDB LOD(<https://cbdb.library.sh.cn>).

Universiteit Leiden 통신과 제국: 비교 관점에서 본 중국 제국(<https://chinese-empires.eu>).

Universiteit Leiden MARKUS(<https://dh.chinese-empires.eu/markus>).

국문초록

본 논문은 한국역사인물데이터베이스(KBDB: Korea Biographical Database Project)를 구현하기 위하여, 중국 역사 인물에 관한 정보를 종합하여 디지털에서 구현한 '중국인물 데이터베이스 프로젝트(CBDB: China Biographical Database Project, 中國歷代人物傳記資料庫)'를 탐색했다. CBDB는 복합적인 역사 인물 정보를 구현한 데이터 설계를 제시하였을 뿐만 아니라, 데이터 공유를 통해서 역사학 연구 방법론 발전에 공헌하고 있다. 이를 바탕으로 『홍문록(弘文錄)』을 예시로 하여, 한국학중앙연구원 한국역대인물정보시스템, 국사편찬위원회 조선왕조실록 부가열람, 한국고전번역원 한국고전종합DB를 상호 연결할 수 있는 한국역사인물데이터베이스의 설계 방안을 제시했다.

투고일 2023. 3. 24.

심사일 2023. 4. 14.

게재 확정일 2023. 5. 11.

주제어(keywords) 홍문록(弘文錄, *Hongmunrok*), 인물(figure), 온톨로지(ontology), 디지털 인문학(digital humanities)

Abstract

Designing the Korea Biographical Database Project: Using *Hongmunrok* (弘文錄) as an Example

Kim, Ba-ro

To implement the Korea Biographical Database Project (KBDB), this study examined the China Biographical Database Project (CBDB, 中國歷代人物傳記資料庫), which was digitally implemented by synthesizing information on Chinese historical figures. The CBDB not only presents a data design that realizes complex information on historical personalities, but also contributes to the development of a historical research methodology through data sharing. With the CBDB as a reference, using the *Hongmunrok* (弘文錄) as an example, we proposed a design scheme for a KBDB that can interconnect the Korean Historical Personality Information System of the Academy of Korean Studies, the National Institute of Korean History's Supplement to the Records of the Joseon Dynasty, and the Translation of Korean Classics Comprehensive Database.

