

디지털 인문학과 한문학 연구

— 고문헌 자료 대상 국내 디지털 인문학 연구 사례를 중심으로 —

류인태*

〈차 례〉

1. 디지털 인문학과 고문헌 자료 연구
2. 고문헌 자료 대상 국내 디지털 인문학 연구 사례
3. 한문학 연구와 스몰 데이터(small data)

[국문초록]

2,000년대 중후반 이후로 디지털 인문학에 대한 논의가 전 세계적으로 꾸준히 확산됨에 따라 국내에서도 디지털 인문학에 대한 관심이 점차로 증가하고 있는 상황이다. 아쉬운 점은 디지털 인문학에 대한 관심이 증가하고 있는 것과 별개로, 실제 국내에서 이루어졌거나 또는 이루어지고 있는 디지털 인문학 연구 사례가 매우 적다는 사실이다. 최근 국내에서 '디지털 인문학'이라는 개념에 대한 담론은 많이 형성되고 있지만, 상대적으로 디지털 인문학 연구의 구체적 방법론이 무엇인지 그 실제(實際)에 대한 논의가 좀처럼 이루어지지 않고 있는 것은, 바로 이러한 현실적 상황과 관련이 깊다.

특히 고전 자료 또는 고문헌 자료를 대상으로 한 인문학 연구의 경우, 텍스트로부터 데이터를 풍부하게 추출할 수 있다는 측면에서 디지털 인문학 연구 방법론을 활용할 수 있는 여지가 매우 크다. 그럼에도 불구하고 고문헌 자료를 대상으로 하여 국내에서 진행되었거나 또는 진행되고 있는 디지털 인문학 연구 사례 또한 찾아보기가 어려운 현실이다. 이와 같은 맥락에서 고문헌 자료를 대상으로 하여 디지털 미디어를 본격적으로 활용한 국내의 연구 사례들이 구체적으로 무엇인지 그리고 각각의 연구가 안고 있는 학술적 맥락이 무엇인지 정리해볼 필요가 있다.

당장 고문헌 자료를 대상으로 디지털 인문학 연구를 수행하고자 하는 연구자들에게 가

* 한국학중앙연구원 인문정보학 전공 박사과정 수료 / E-mail: ryuintae@gmail.com

장 훌륭한 참고 요소가 될 수 있는 것은, 기존 연구 사례가 무엇인지에 대한 소개와 개별 사례에 대한 귀납적 검토이다. 각각의 연구가 이루어질 수 있었던 기초적 문제의식과 연구 과정에 적용된 디지털 기술 그리고 구현된 결과물이 갖추고 있는 디지털 형식의 의미가 무엇인지 개략적으로 파악된다면, 고문헌 자료를 대상으로 디지털 인문학 연구를 시도하고자 하는 연구자들에게 매우 훌륭한 참고 요소가 될 것이다.

실제로 본고에서 정리한 9가지 국내 디지털 인문학 연구 사례는 각각이 다른 형식의 고문헌 자료를 대상으로 하면서도 유사한 디지털 플랫폼을 활용하였거나 또는 데이터를 체계적으로 정리하기 위한 온톨로지를 공통적으로 적용하는 등 디지털 인문학 방법론의 차원에서 비교·대조하여 참고할만한 지점이 매우 많다.

주제어: 디지털 인문학, 한문학, 고문헌, 스몰 데이터, 인문정보학

1. 디지털 인문학과 고문헌 자료 연구

불과 몇 년 전만 해도 일부 연구자와 언론에 의해서 드물게 언급되던 ‘디지털 인문학’이라는 키워드가 최근에는 다수의 논문과 단행본 그리고 언론 기사에 의해 점차로 많이 논의되고 있다. 해외의 인문학 동향, 즉 미국과 유럽에서 디지털 인문학이 하나의 뚜렷한 학술적 조류로 받아들여지고 있는 상황이기 때문에 그러한 분위기에 힘입어 국내에서도 디지털 인문학이 조금씩 조명받기 시작한 것이라 짐작해 볼 수 있다. 뿐만 아니라 4차 산업 혁명, 인공지능, 빅데이터 등 최근 첨단기술 쪽에서 이슈가 되고 있는 디지털 관련 키워드들 또한 디지털 인문학에 대한 담론이 늘어나는 데에 직접적 영향을 주었음은 부인할 수 없는 사실이다. 이러한 상황에서 실제 디지털 인문학이 향후 인문학이 추구해 나가야 할 뚜렷한 연구방법론으로서의 실효적 가치를 제시할 수 있는지 아닌지에 대한 구체적인 논의는 추후에 이루어져야 할 것으로 보인다. 현 시점에서 디지털 인문학에 대한 담론이 꾸준히 증가하고 있는 것에 비해 실제 국내에서 이루어졌거나 이루어지고 있는 디지털 인문학 연구 사례는 극소수이기 때문에, 인문학의 미래적 방향으로서 디지털 인문학이 내포하고 있는 가능성을 귀납적으로 검증할 수 있는 여지가 현재로선 매우 부족하기 때문이다. 이에 기초하여 생각해 볼 때, 디지털 인문학

이 무엇인지 그것이 내포하고 있는 바를 이야기하고 정리하는 담론이 더욱 늘어나야 할 필요도 있겠지만, 그보다 더욱 우선적으로 요구되는 것은 실제 디지털 인문학 연구를 수행함으로써 디지털 인문학 연구의 외연을 폭넓게 확장해 나가는 시도라 하겠다.

디지털 인문학은, 디지털 미디어를 적극적으로 이용한 인문학 연구 또는 디지털 환경을 능동적으로 활용한 인문학 연구 정도로 거칠게 정의할 수 있다.¹⁾ ‘적극적으로’ 또는 ‘능동적으로’ 라는 표현은, 곧 연구 대상을 디지털화 하거나 연구 결과물을 디지털 형태의 저작물로 간행하는 소극적인 수준을 넘어서서 디지털 환경에서 본격적으로 연구를 수행하거나 디지털 미디어를 연구방법론에 직접 도입하는 것을 의미한다.

예를 들어 그동안 국내에서 이루어져 온 고문헌 자료를 대상으로 한 전산화 프로젝트의 경우, 대부분 연구 대상을 온라인상에서 서비스할 수 있는 디지털 형식의 결과물로 가공하는 데에 초점을 두었다. 국사편찬위원회의 ‘조선왕조실록 DB’나 고전번역원의 ‘한국고전종합 DB’와 같은 결과물들이 그에 해당하는 예이다. 웹서비스 형식의 해당 결과물들을 통해 수많은 한국학 연구자들이 ‘조선왕조실록’과 ‘한국문집총간’ 자료들을 손쉽게 열람하고 그것을 토대로 효과적으로 연구를 수행하고 있을 뿐만 아니라, 대중들 또한 어려운 고문헌 자료들을 간편하고 쉽게 접근할 수 있음으로 인해 역사 정보에 기초한 다양한 교육 자료 및 문화콘텐츠들이 제작되고 있다. 이와 같이 역사 자료나 문집 자료를 웹에서 자유롭게 열람할 수 있는 형식의 디지털 미디어로 옮기는 작업 자체가 지닌 가치는 매우 큰 것이며, 그러한 작업 또한 기초적 성격의 디지털 인문학 연구로 볼 수 있다.²⁾

1) 디지털 인문학이 내포하고 있는 함의에 대해서는 다음의 서술을 참고할 수 있다. “디지털 인문학이란 정보통신기술(Information & Communication Technology)의 도움을 받아 새로운 방식으로 수행하는 인문학 연구와 교육, 그리고 이와 관련된 창조적인 저작 활동을 일컫는 말이다. 이것은 전통적인 인문학의 주제를 계승하면서 연구 방법 면에서 디지털 기술을 활용하는 연구, 그리고 예전에는 가능하지 않았지만 컴퓨터를 사용함으로써 시도할 수 있게 된 새로운 성격의 인문학 연구를 포함한다. 단순히 인문학의 연구 대상이 되는 자료를 디지털화 하거나, 연구 결과물을 디지털 형태로 간행하는 것보다는 정보기술의 환경에서 보다 창조적인 인문학 활동을 전개하는 것, 그리고 그것을 디지털 매체를 통해 소통시킴으로써 보다 혁신적으로 인문지식의 재생산을 촉진하는 노력 등이 ‘디지털 인문학’이라는 새로운 조어의 함의라고 할 수 있다.” -김현·임영상·김바로(2016), 17면.

그러나 그것은 과거 전산인문학으로부터 파생된 결과물로서의 기능과 가치를 현 시점에서 긍정하는 차원에서 유의미한 것이지, 현재의 '조선왕조실록 DB'와 '한국고전종합 DB'가 앞으로도 지금과 동일한 형식으로 웹에서 계속 서비스될 것이라고 생각하는 사람은 없을 것이다. 2,000년대 초반의 정보통신 기술 수준과 현재의 기술 수준이 엄연히 다를뿐더러 기술 수준의 차이뿐만 아니라 사회·문화적 환경 또한 크게 달라졌기 때문이다.

1,990년대 후반 또는 2,000년대 초반과 비교하였을 때 현재 우리 일상에 웹이 미치는 영향력은 비교할 수 없을 정도로 커졌고, 심지어 웹 환경을 제외하고 우리 일상을 온전히 설명하기가 어려울 정도가 되었다. 학술 환경 또한 마찬가지이다. 선형적(線形的, Linear) 방식의 읽기와 쓰기에 익숙하여 그로부터 사유를 전개하고 이해를 도모하던 이전 세대와 달리 현 세대는 웹 환경의 하이퍼텍스트에 장시간 노출됨으로써 비선형적(非線形的, Nonlinear) 방식의 읽기와 쓰기를 추구한다. 그로부터 전개되는 사유와 의미 이해 또한 이전 세대의 그것과는 다른 방식이며, 향후 학술후속세대가 그들임을 고려할 때 웹 환경을 학술적 차원에서 비판적으로 검토하고 활용할 수 있는 이해를 갖추는 것이 현 단계에서는 필수적이라고 하겠다.

디지털 인문학은 이와 같은 정보통신 기술 발전에 의해 초래된 사회, 문화 환경의 급격한 변화에 따라 인문학 연구 방법과 과정을 온전히 디지털 환경에서 시도하고자 하는 것이다. 오프라인 환경에서 접할 수 있는 『조선왕조실록』, 『한국문집총간』이라는 문헌을 디지털 환경에서 재편찬하는 것을 넘어서서, 디지털 미디어를 이용하여 『조선왕조실록』과 『한국문집총간』에 담긴 내용을 새롭게 연구

2) 엄밀히 이야기할 경우 2,000년대 초반 IMF사태의 여파로 인해 발생한 경제적 어려움을 해결하고자 정부에서 공공근로사업 명분으로 추진하였던 고문헌 자료의 대규모 DB화 작업들은 디지털 인문학의 역사를 참고하였을 때, '전산인문학'의 영역에 해당한다 할 수 있다. 전산인문학은 1차 디지털 인문학으로서(현재의 디지털 인문학은 2차 디지털 인문학) 1980~90년대 코퍼스 구축이나 텍스트 인코딩, 데이터베이스 구축 등 인문학 연구에 보조적 역할을 담당했던 당시의 컴퓨터 기술에 기초한 학술적 개념을 가리킨다. 1차 디지털 인문학으로서의 전산인문학과 현재의 디지털 인문학 사이에 놓인 인과론적 고리에 대해서는 다음의 논문을 참고할 수 있다. '홍정욱(2015), 「디지털기술 전환 시대의 인문학 -디지털 인문학 선언문을 통한 고찰」, 『인문콘텐츠』 38, 인문콘텐츠학회, 41~74면.'

하고 편찬하고자 하는 것이다. 고문헌 자료에 담긴 내용을 새롭게 연구한다는 것은 곧 아날로그 환경의 연구에서 인간의 두뇌로 파악하기 어려웠던 요소들을 디지털 환경에서 찾아내고 그것들을 조직화하고 시각화하는 과정을 새롭게 시도할 수 있다는 것을 의미한다. 한편으로 새롭게 쓰는 말이 다소 오해의 여지를 낳을 수 있겠다는 생각도 든다. 디지털 인문학이 마치 전통적 인문학이 추구해왔던 방향과는 전혀 다른 새로운 것을 시도하고자 하는 모양새로 비춰질 수도 있기 때문이다. 그러나 명확하게도 디지털 인문학 연구의 새로움은 대상을 들여다보는 시각과 대상을 분석하고 그로부터 도출된 정보와 지식을 정리해나가는 방법론에 있어서의 새로움을 가리키는 것이지, 인문학 연구를 통해 추구하고자 하는 본질적 방향이 바뀔을 의미하지는 않는다. 고문헌 자료에는 옛 사람들의 사유와 생활이 녹아 있고, 그것을 밝혀내고 그 의미를 추론하고자 하는 것이 인문학의 본질적 목적임을 고려할 때, ‘디지털’이라는 키워드에서 파생되는 첨단 이미지 즉, 새로움만을 강조한다면 오히려 그것이야말로 디지털 인문학을 크게 왜곡하는 것이다.

디지털 인문학 방법론을 통해 고문헌 자료를 연구한다는 것은, 고문헌 자료에 담긴 정보와 지식을 디지털 미디어로 어떻게 온전히 재매개(remediation)³⁾할 것인가의 고민이다. 고문헌 자료가 담고 있는 내용과 형식을 디지털 미디어로 ‘새롭게’ 옮겨내는 것이 목적이 아니라, 디지털 미디어가 담아낼 수 있는 첨단의 형

3) 재매개(remediation) 개념에 대해서는 다음의 인용을 참고할 수 있다. “기원전 8세기 무렵에 그리스인들은 구전 신화와 영웅 전설 공간을 좀 더 정확하면서 선형적인 파피루스 두루마리 공간(그리고 돌이나 나뭇조각)으로 개조했다. 에릭 해블록(Eric Havelock, 1982)에 따르면 이런 과정은 수백 년 동안 지속됐다. 고대 말기에 파피루스 두루마리에서 코덱스로 바뀌면서 글쓰기 공간은 다시 변화했다. 텍스트를 배치하기 위해 2차원 표면을 좀 더 효과적으로 사용하도록 만든 것이다. 서유럽에서는 필사 코덱스에서 인쇄본으로의 변화로 인해 이런 개조 과정이 또다시 생겼다. 전자 글쓰기로서의 이동 역시 이런 개조 과정의 하나다. 이런 변화를 우리는 “재매개(remediation)”라고 부를 수 있을 것이다. 재매개란 말 속에는 새로운 매체들이 앞선 매체를 대체하면서 그 매체의 글쓰기 특징들을 빌려 와서 재조직하며, 그것들의 문화적 공간을 새롭게 형성한다는 의미가 담겨 있다. 파피루스에 글을 쓰는 것은 귀찮아 아니라 눈까지 관여시키면서 단어들이 현실에 대한 다른 요구를 부여해 줌으로써 구술 커뮤니케이션을 재매개했다. 다른 변천 과정 역시 독자인 우리가 텍스트와 그것이 세계를 재현하는 방식에 접근하는 조건을 요란스럽고 철저하게 바꿨다. 재매개에는 존경과 경쟁이 함께 수반된다. 왜냐하면 새로운 매체는 이전 매체의 일부 기능들을 모방할 뿐 아니라 옛 매체를 개선해야만 한다고 주장하기 때문이다.” -제이 데이비드 볼터 저·김익현 역(2010), 35면.

식을 토대로 고문헌 자료에 담긴 내용과 형식을 ‘온전히’ 재구성함으로써 웹에서 자유롭게 유통될 수 있는 정보와 지식을 생산해내는 과정이 곧 디지털 인문학에 기초한 고문헌 자료 연구의 핵심이다. 고문헌 자료에 담긴 내용과 형식을 ‘온전히’ 재구성하기 위해서는, 첨단 디지털 미디어에 대해 이해하고 있어야 할뿐만 아니라 고문헌 자료가 전통적 미디어로서 담고 있는 여러 역사·문화적 맥락 또한 파악할 수 있어야 한다. 이는 곧, 디지털 인문학 연구가 기존 인문학 연구와 동떨어진 별개의 영역이 아니라 기존 인문학 연구에서 수행되어 온 전통적인 연구시각을 계승하되, 그것을 보완하는 방향으로 나아가야 할 인문학 연구임을 의미한다. 그리고 그 보완의 방향은 곧 개별 고문헌 자료를 각기 연구해 온 개인 연구자들의 정보(information) 개방(openness)과 그것을 토대로 한 집단적 차원의 지식(knowledge) 공유(Shareness)라 할 수 있으며, 그것을 가능하게끔 해주는 매개가 바로 디지털 환경에서의 데이터(data)이다.

데이터는 기본적으로 정량적(定量的) 속성을 갖기 때문에, 정성적(定性的) 연구를 추구하는 인문학 연구에 맞지 않는다는 시각이 존재할 수 있다. 일반적인 자연 과학의 연구 대상과 달리 인간 활동이 내포하고 있는 고유함과 모호함으로 인해, 인간의 직접적 활동 또는 그것을 기록한 문헌 자료를 대상으로 한 인문학 연구는 기본적으로 정성적인 것으로 여겨져 왔기 때문이다. 즉, 정성적 대상을 연구하는 인문학에서 정량적 속성을 지닌 데이터를 적극적으로 연구 과정에 도입하는 것 자체가 아이러니한 모습으로 비춰질 수 있는 것이다. 정량적 실험을 통해 대상에 대한 이해를 도출하는 자연과학 연구와 비교할 경우 인문학 연구는 분명 상대적으로 더욱 정성적 경향을 띠는 것이 사실이다. 그러나 연구 대상이 지닌 특성이 정성적이라고 해서 연구 방법론 또한 반드시 정성적 차원의 접근만 이루어져야 한다는 근거는 어디에도 없다. 오히려 인간을 이해하기 위한 본질적 관점이 정성적 차원에서 다루어지고 그러한 정성적 차원의 연구를 더욱 심화할 수 있는 효과적인 도구로서 정량적 매개가 수용될 수 있다면, 인문학 연구에서 디지털 데이터를 적극적으로 활용하는 것이 필요하다고 할 수 있다.



〈그림 1〉 고문헌 자료를 대상으로 한 인문학 연구에 있어 데이터 모델 적용의 과정

인문학 연구에서 디지털 데이터를 적극적으로 활용한다고 할 경우 기본적으로 데이터 개념에 대한 이해가 필요하다. 그리고 그 때의 데이터 개념에 대한 이해는, 데이터로 옮겨낼 아날로그 대상에 대한 이해와 데이터가 유통될 디지털 미디어에 대한 이해를 함께 포괄한다. 즉, 고문헌 자료를 대상으로 한 디지털 인문학 연구에서 첨단 디지털 미디어에 대한 이해를 갖추고 있지만 고문헌 자료를 읽고 분석할 수 있는 안목이 없다면, 데이터에 대한 온전한 이해를 갖춘 것이라 보기 어렵다. 비유하자면 표현할 도구는 가지고 있지만 표현할 대상이 지닌 본질적 특성이 무엇인지 잘 몰라서 이리저리 대상을 스케치만 하다가 결국 그리기를 포기하는 화가와 같다. 그 반대 또한 마찬가지이다. 고문헌 자료를 읽고 분석할 수 있는 안목이 있지만 디지털 미디어에 대한 이해를 갖추지 못한 상황이라면 이는 표현할 대상이 지닌 본질에 대해서 이해하고 있지만 표현할 도구가 없어서 아예 그리기를 포기한 화가와 같다. 데이터에 대한 온전한 이해는 분석 또는 표현하고자 하는 대상에 대한 이해와 분석 또는 표현할 수 있는 도구에 대한 이해가 함께 갖추어 졌을 때 가능한 것이다.

디지털 인문학은 이러한 디지털 데이터(data) 구축을 기초로 이루어지는 인문학 연구이다. 디지털 인문학 연구를 수행하기 위해서는 앞서 언급했던 것과 같이 인문학 데이터에 대한 관점이 필수적으로 요구된다. 아쉬운 점은 ‘개인 연구’에 치중해 온 기존 인문학 연구의 관습적 풍토로 인해 인문학 연구자들이 그러한 데이터에 대한 관점을 갖기가 쉽지 않다는 사실에 있다. 개인 연구자에게 있어서

연구 수행 과정에서 산출된 데이터는 결국 ‘나의 연구결과물로서 온전히 나의 것’이며, 그렇기 때문에 데이터 구축 포맷 또한 ‘내가 이해하는 방식으로 정리할 수 있다면 그만’이라는 인식이 보편적으로 자리 잡고 있기 때문이다. 개인 연구자에게 있어서 그러한 데이터는 하나의 ‘재산’이기에 그것을 디지털 환경에서 공개하고 공유하는 것은 현실적으로 맞지 않는 이야기이며, 오히려 아날로그 환경의 논문이나 단행본에 자신의 연구 데이터를 일부 공개하고 그에 대한 저작권을 확보하여 행사하는 것이 지배적인 형태의 연구 과정으로 굳어져 있다. 그러한 형태의 기존 연구 방식은 사실 개인 연구자의 입장에서 보았을 때 그렇게 해서 온전히 스스로의 연구를 평가받고 그로부터 사회적 지위를 보장받고 경제적 능력 또한 확보할 수 있는 현재의 학계 시스템이 불러온 결과물로 받아들일 여지 또한 크다.

명확한 것은, 현재의 학술 환경이 어떠한 상황에 놓여 있는나의 문제와 별개로 정보와 지식이 공유되고 그로부터 새로운 지식이 창출되는 웹 환경은 끊임없이 변화하고 있다는 사실이다. 웹을 통해 정보와 지식이 개방적으로 공유되는 풍토가 더욱 확산되고 개인의 능력보다는 집단의 역량이 보다 더 강조되어 나가는 현재의 미디어 환경에 인문학 연구 또한 발을 맞출 필요가 있다. 개인 연구에 초점을 둔 폐쇄적이고 권위적인 학술 풍토에서 벗어나, 개방적인 시각으로 연구를 수행하고 그렇게 수행한 연구 결과물을 디지털 환경에서 공유 가능한 형식의 정보와 지식으로 편찬하여 공개할 수 있을 때, 연구자들 개개인의 연구 업적이 보다 온전히 인정받을 수 있는 연구 평가 시스템이 새롭게 마련될 수 있다.

현재까지도 인문학 연구의 여러 분야에서 전통 인문학 연구에서 중요하게 여겨져 오던 정보의 독창성(originality) 즉, 개인 연구자에 의한 권위(authority)와 저작(authorship)의 가치가 뚜렷하게 강조되고 있는 상황이지만, 향후 집단 연구(또는 개인 연구자들의 연대)를 바탕으로 한 디지털 인문학 연구가 보다 더 활발해질 경우 개방(openness)과 참여(participation)에 초점을 둔 정보의 공유가 보편화될 가능성이 크다.⁴⁾ 각 개인 연구자가 지니고 있던 고문헌 자료에 대한 깊이 있는 이해와 지식을 디지털 환경에서 보편적으로 유통될 수 있는 형식의 데이터

로 치환하고, 그러한 각각의 개인 연구자들에 의해 생성된 수많은 연구 데이터를 유의미하게 연결할 수 있는 디지털 인문학 연구가 앞으로 더욱 활발하게 시도되어야 한다고 본다.

2. 고문헌 자료 대상 국내 디지털 인문학 연구 사례

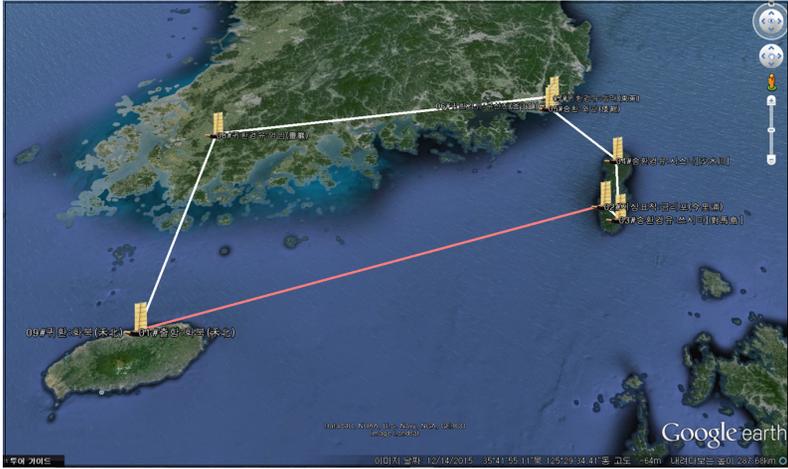
해외의 경우 고문헌 자료 또는 고문헌 자료에 담긴 내용을 대상으로 한 디지털 데이터 구축 관련 프로젝트 또는 디지털 인문학 연구 사례가 점차 증가하고 있음에도 불구하고, 국내에서는 고문헌 자료를 대상으로 한 디지털 인문학 연구가 활발하지 않은 상황이다.⁴⁾ 조금 더 솔직히 이야기한다면, 활발하지 않은 정도가 아니라 현재까지의 연구를 살펴보았을 때 몇몇의 연구 사례를 제외하고서는 뚜렷한 연구 결과물이 거의 없다고 해도 무방하다. 이에 대한 이유를 설명하기 위해서는 여러 측면에서의 논의가 필요하며 다소 장황하면서도 복잡한 논리가 동원되어야 할 것으로 여겨지기에 생략하고자 한다. 오히려 현 시점에서 더욱 필요한 논의는, 지금 당장 고문헌 자료를 대상으로 디지털 인문학 연구를 시도한다면 어떤 형식의 연구들이 가능한지에 대한 구체적 검토가 아닐까 생각된다.

이와 관련하여 필자가 직접 참여하였거나 또는 필자 주변에서 이루어졌던 실제 디지털 인문학 연구 사례들을, 간단하게나마 지면을 할애하여 소개하고자 한다. 아래에 소개하는 9가지 디지털 인문학 연구 사례는 ‘디지털 환경’ 또는 ‘디지털 미디어’를 활용하였다는 측면에서는 유사하지만, 연구 인력, 기간, 대상 그리고 연구 내용 및 결과물의 형식이 모두 다르다.

4) 전통적 인문학과 디지털 인문학이 지닌 상이한 성격에 대해서는 다음의 문헌을 참고할 수 있다. ‘spiro, Lisa. ““This Is Why We Fight”: Defining the Values of the Digital Humanities”, *Debates in the Digital Humanities*. Ed. Matthew K. Gold. MN: University of Minnesota Press. 2012. p.16-34.’

5) 고문헌 자료 또는 그 내용을 대상으로 한 대표적인 해외의 디지털 인문학 사례에 대해서는 다음의 문헌을 참고할 수 있다. ‘김현·임영상·김바로(2016), 『디지털 인문학 입문』, HUEBOOKS, 24~34면.’

1) 조선시대 표류노드 시각망 연구

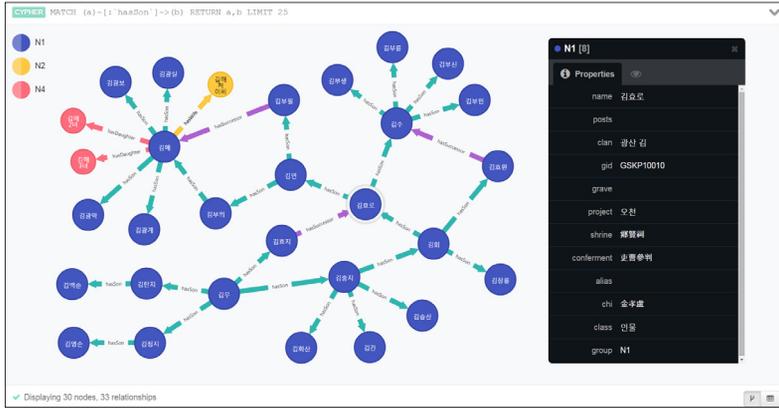


〈그림 2〉 조선시대 표류노드 시각망 연구 시각화 결과물 '표류 경로 지도' 예시)

연구 인력	한문학 / 인문정보학 전공자
연구 기간	2014년 12월 - 2015년 12월
연구 대상	『표인영래등록(漂人領來騰錄)』, 『조선왕조실록(朝鮮王朝實錄)』, 『변례집요(邊例集要)』 등에 남아 있는 조선시대 표류 기록
연구 내용	조선시대 표류에 대한 기록을 대상으로 하여 표류 사건과 관련된 여러 데이터를 추출하여 네트워크 형태의 그래프 데이터베이스와 전자지도 기반의 시각화 콘텐츠를 구현함. 학술 영역과 문화산업 분야에서 자유롭게 활용할 수 있는 조선 시대 표류 사건 관련 디지털 자원을 개발하는 데 주력하였으며, 그 과정에서 이루어졌던 작업 내용과 문제의식을 정리하여 연구 논문을 발표하고 단행본을 출간함.
연구 관련 결과물	디지털 환경에서의 편찬 결과물 ⁷⁾ , 연구 논문 ⁸⁾ , 단행본 ⁹⁾

6) 조선시대 표류노드 시각망 Wiki DB 제공.
 7) 조선시대 표류노드 시각망 Wiki DB (http://www.digerati.kr/mediawiki/index.php/표류_노드_시각망)
 8) 이수진(2015), 『조선시대 표류노드 시각망 구축 과정 -표류 기록의 의미 요소 추출을 중심으로』, 『溫知論叢』 45집, 온지학회, 161~186면.
 9) 허경진·구지현(2016), 『조선시대 표류노드 시각망 연구일지』, 보고서.

3) 고문서(古文書) 연구를 위한 데이터 기술(記述) 모델 연구



〈그림 4〉 문중 고문서 시맨틱 DB 시각화 결과물 예시¹⁴⁾

연구 인력	역사학 / 인문정보학 전공자
연구 기간	2014년부터 현재까지 계속 진행 중
연구 대상	한국학중앙연구원 장서각(藏書閣)의 고문서 자료관 아카이브에 등재되어 있는 광산 김씨 문중 고문서와 부안 김씨 문중 고문서
연구 내용	실물 자료로서 ‘고문서’가 지닌 문헌 정보가 아니라 고문서 안에 기술된 내용을 디지털 환경에서 효과적으로 다룰 수 있는 방법을 강구하고자 하였음. 이러한 맥락에서 디지털 미디어를 이용하여 고문서에 담긴 내용을 자유롭게 검색하고, 검색한 내용을 연결하여 그로부터 새로운 이야기를 찾아 재구성할 수 있는 데이터베이스의 기초 모델에 대한 연구를 꾸준히 진행하고 있음. 고문서 연구자와 인문정보학 연구자들이 함께 협업하여 직접 데이터를 정리하고 분석하는 방향으로 연구를 진행하고 있다는 것이 본 연구의 특징.
연구 관련 결과물	디지털 환경에서의 편찬 결과물 ¹⁵⁾ , 연구 논문 ¹⁶⁾

12) 하영휘·백광렬(2015), 『‘만성대동보(萬姓大同譜)’ 자료를 활용한 ‘친족관계망정보시스템(LNIS, Lineage Network Information System)’ 개발 -개발실무와 활용방안, 이론적 함의』, 진인진.

13) 백광렬(2017), 『조선 후기 ‘양반지베네트워크’의 성격과 구조변동 : 상충양반의 친족연결망 중심으로』, 서울대학교 대학원 박사학위 청구논문.

14) 문중 고문서 시맨틱 DB 제공.

15) 문중 고문서 시맨틱 DB (<http://cefia.aks.ac.kr:7474/browser/>)

16) 김하영(2015), 『광산김씨 문중 고문서 시맨틱 데이터베이스 모델 설계: ‘門中古文書’ 디지털 아카이브

4) 수신사(修信使) 및 조사시찰단 자료 DB 구축 연구



〈그림 5〉 수신사 자료 Wiki DB 예시(7)

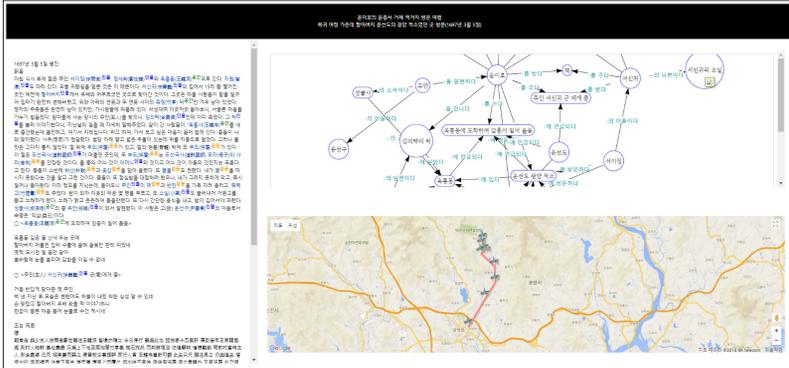
연구 인력	한문학 / 인문정보학 전공자
연구 기간	2015년 9월 - 2018년 8월(종료 예정)
연구 대상	1876년 조일수호조규 체결부터 1884년 갑신정변 사후 처리 시기까지 총 5회에 걸쳐 메이지 일본에 파견되었던 조선 사절단 '수신사(修信使)'와 직·간접적으로 관련된 여러 문헌자료들
연구 내용	수신사와 직·간접적으로 관련된 여러 자료들을 구체적 연구대상으로 삼아 수집하고 번역하며 동시에 그러한 자료들로부터 추출한 유의미한 정보를 종합적으로 제공할 수 있는 온톨로지 및 DB를 고민하고 실제 그것을 구현해 내고자 하는 연구를 진행하고 있음.
연구 관련 결과물	디지털 환경에서의 편찬 결과물 ¹⁸⁾ , 연구 논문 ¹⁹⁾

구현 연구’, 한국학대학원 석사학위 청구논문. 김현(2016), 『디지털 인문학과 고문헌 자료 연구』, 『열상고전연구』 50집, 열상고전연구회, 14~38면. 김현·안승준·류인태(2016), 『고문서 연구를 위한 데이터 기술 모델』, 역사학회, 제59회 전국역사학대회 (주제: 기록의 생성과 역사의 구성) 발표논문, 2016년 10월 28일.

17) 수신사 자료 Wiki DB 제공.

18) 수신사 자료 Wiki DB (<http://susinsa.info/>) 연구가 진행 중인 현재는 개방되어 있지 않으며 연구

5) 『지암일기(支菴日記)』 디지털 아카이빙 연구



〈그림 6〉 지암일기 Wiki DB 토대의 시각화 구현 예시²⁰⁾

연구 인력	한문학 / 역사학 / 미술사학 / 미학 / 인문정보학 전공자
연구 기간	2016년 9월 - 2019년 8월(종료 예정)
연구 대상	조선 후기 해남 양반 윤이후(尹爾厚)가 쓴 『지암일기(支菴日記)』
연구 내용	조선시대 양반인 지암(支菴) 윤이후(尹爾厚)의 『지암일기』를 디지털 환경에서 새롭게 편찬하고, 이 과정에서 산출되는 온톨로지 설계안, Wiki DB, Graph DB, 시각화 결과물 등을 종합적으로 체계화 하여 디지털 기술을 이용한 역사 자료 분석의 새로운 모델을 제시하고자 하는 연구를 진행하고 있음.
연구 관련 결과물	디지털 환경에서의 편찬 결과물 ²¹⁾

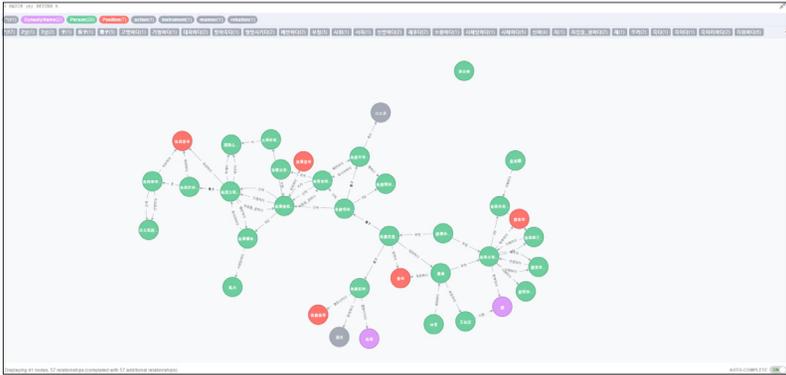
종료 시점인 2018년 8월 이후 공개 예정.

19) 류인태(2016), 『디지털 환경에서의 인문 지식 연구에 관한 小考 -수신사 자료 DB 편찬 프로젝트를 중심으로』, 『열상고전연구』 50집, 열상고전연구회, 101~139면.

20) 지암일기 Wiki DB 제공.

21) 지암일기 Wiki DB (<http://jjiamdiary.info/>) 연구가 진행 중인 현재는 개방되어 있지 않으며 연구 종료 시점인 2019년 8월 이후 공개 예정.

6) 역사 자료 텍스트의 전자적 기술에 의한 지식 관계망 구현 연구



〈그림 7〉 역사 자료 텍스트의 전자적 기술에 의한 지식 관계망 구현 연구의 시각화 결과물 예시²²⁾

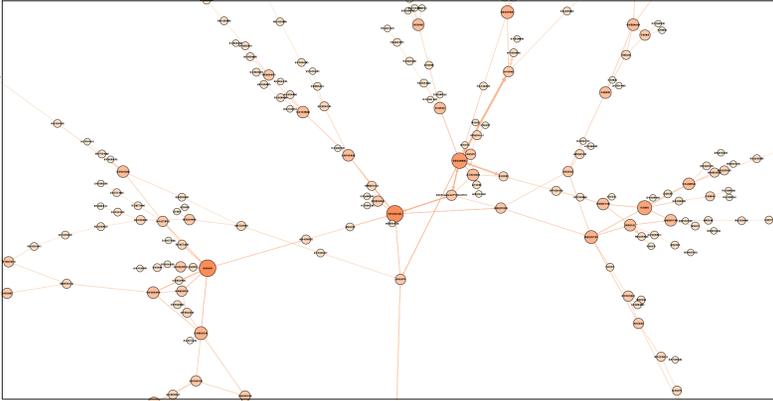
연구 인력	인문정보학 전공자
연구 기간	2009년 3월 - 2017년 2월
연구 대상	『통감절요(通鑑節要)』
연구 내용	조선 시대 사대부들에게 있어서 중국 역사에 대한 지식 제공과 함께 문화적 차원에서 또한 다양한 정보를 제공해 온 기본 교재로서의 역할을 하였던 『통감절요』를 구체적 대상으로 삼아, 디지털 환경에서 전통 시기의 역사 지식을 표현할 수 있는 방안이 무엇인지에 대해 연구하였음.
연구 관련 결과물	디지털 환경에서의 편찬 결과물 ²³⁾ , 연구 논문 ²⁴⁾

22) 역사 자료 텍스트의 전자적 기술에 의한 지식 관계망 구현 연구 Wiki 제공.

23) 역사 자료 텍스트의 전자적 기술에 의한 지식 관계망 구현 연구 Wiki
(<http://dhlabs.aks.ac.kr/~yjw2999/wiki/index.php/>)

24) 윤중웅(2017), 『역사 자료 텍스트의 전자적 기술에 의한 지식 관계망 구현 연구 : ‘通鑑節要’와 譯註 通鑑節要』를 중심으로, 한국학대학원 박사학위 청구논문.

7) 조선시대 과거 합격자의 디지털 아카이브 편찬과 인적 관계망 구현 연구



〈그림 8〉 조선시대 과거 합격자 관계망 구현 연구 시각화 결과물 '문과 합격자 네트워크' 예시²⁵⁾

연구 인력	인문정보학 전공자
연구 기간	2009년 3월 - 2017년 2월
연구 대상	현전하는 여러 방목(榜目) 자료들과 족보(族譜) 자료들
연구 내용	현전하는 방목 자료에 대한 전반적인 조사 및 분석을 통해 방목 자료 속에 담긴 과거 합격자 정보를 종합하고, 과거 합격자와 그가 속한 문중 인물과의 관계를 분석한 과거 합격자 인적 관계 시맨틱 데이터 및 이를 시각화한 과거 합격자 인적 관계 네트워크 그래프를 구현함으로써, 방목 합격자를 중심으로 하는 조선시대 엘리트 인적 관계망에 대한 본격적 연구를 시도하였음.
연구 관련 결과물	디지털 환경에서의 편찬 결과물 ²⁶⁾ , 연구 논문 ²⁷⁾

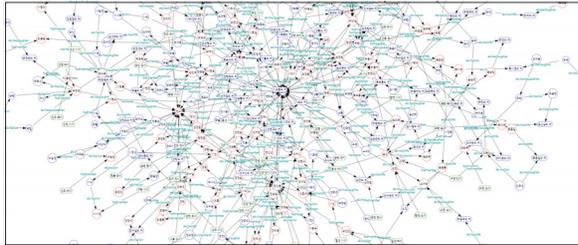
25) 조선시대 과거 합격자 정보 디지털 아카이브 Wiki 제공.

26) 조선시대 과거 합격자 정보 디지털 아카이브 Wiki

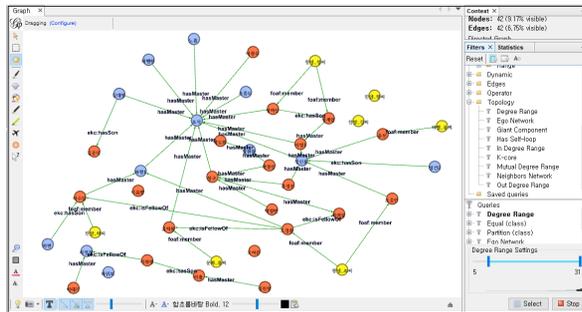
(http://dh.aks.ac.kr/~sonamu5/wiki/index.php/SEDB:과거_합격자_정보_디지털_아카이브)

27) 이재욱(2017), 『조선시대 科擧 合格者の 디지털 아카이브 편찬과 인적 관계망 구현』, 한국학 대학원 박사학위 청구논문.

9) 『덕천원생록(德川院生錄)』과 시맨틱 웹 DB 연구



〈그림 11〉 1609년 『덕천원생록』 수록 인물 관계망³²⁾



〈그림 12〉 Gephi를 활용한 1609년 『덕천원생록』 수록 인물 관계망 분석 예시 화면

연구 인력	한문학 / 인문정보학 전공자
연구 기간	2017년 9월 - 2017년 12월
연구 대상	덕천 서원의 출입 유생 명단인 『덕천원생록(德川院生錄)』
연구 내용	덕천 서원에 출입한 유생들을 기록한 『덕천원생록』에 담긴 인물 정보 및 인물 간의 혈연 및 혼맥 관계를 데이터로 정리하여 관계망으로 나타내고 해당 내용을 분석함으로써, 서원의 정체성 및 운영에 영향을 미쳤을 중심 인물들 간의 역할관계를 실증적으로 제시함.
연구 관련 결과물	디지털 환경에서의 편찬 결과물 ³³⁾ , 연구 논문 ³⁴⁾

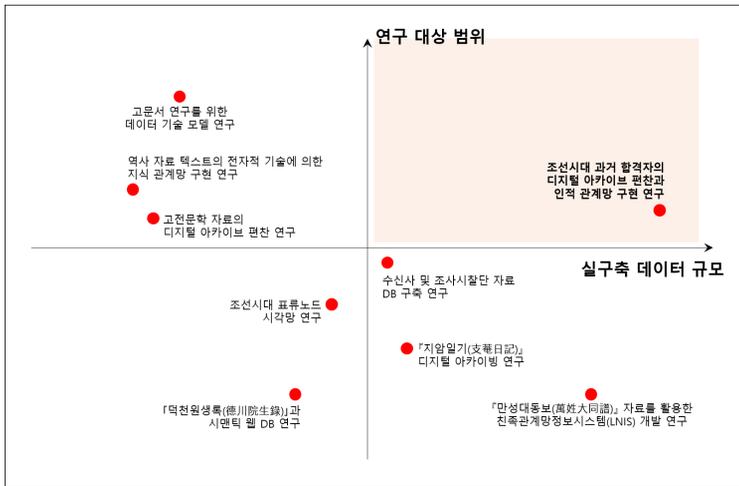
30) 박순(2017a), 『고전문학 자료의 디지털 아카이브 편찬 연구 : 누정기(樓亭記) 자료를 중심으로』, 연세대학교 대학원 박사학위 청구논문.

31) 박순(2017b), 『고전문학 자료의 디지털 아카이브 편찬 연구』, 보고서.

32) 1609년 『덕천원생록』 수록 인물 관계망 (http://www.bigstamp.co.kr/graph/dc_graph.htm)

10) 선행 연구사례들의 의미와 가치

9가지 연구 사례는 공통적으로 고문헌 자료를 연구 대상으로 삼아 수행되기는 하였으나 그 규모와 연구 주체 그리고 연구 목적이 모두 다르다. 연구를 통해 구현해 내고자 한 결과물의 형식과 밝혀내고자 한 학술적 의미가 상이하며, 연구 과정에 적용된 기술적 요소 즉 디지털 환경에서의 연구 방법론 또한 섬세하게 살펴볼 경우 각기 다른 지점들이 있음을 발견할 수 있다.



〈그림 13〉 데이터 규모와 연구 대상 범위를 기준으로 배치해 본 9가지 연구 사례

위의 시각자료는 앞서의 9가지 연구 사례를 ‘데이터의 규모’(가로축)와 ‘연구 대상 범위(세로축)’에 맞추어 배치해 본 결과이다. 데이터 구축을 기초 토대로 이루어지는 연구가 디지털 인문학인 만큼 ‘데이터 규모’는 디지털 인문학으로서 해

- 33) 「덕천원생록」과 시멘틱 웹 DB 연구 Wiki (http://dh.aks.ac.kr/~red/wiki/index.php/덕천원생록과_시멘틱웹_DB_연구)
- 34) 류인태(2017), 「「덕천원생록(德川院生錄)」과 시멘틱 웹 DB 연구」, 한국선비문화연구원, 2017년도 한국선비문화연구원 학술대회 (주제: 한국선비문화연구원의 장기 연구 방향 모색) 발표논문, 2017년 12월 15일.

당 연구가 지닌 성격을 살피는 데 있어서 매우 중요한 기준 요소가 된다.

연구 대상의 범위가 얼마나 넓은지를 기준으로 하여 실제 개별 연구에서 구축한 데이터의 규모를 정리해볼 경우, 해당 연구가 지향하고자 한 목적이 무엇인지 가늠해볼 수 있는 것이다.

가로축과 세로축의 위치 값은 절대적으로 정량화된 기준에 의한 것이라기보다는 어디까지나 다른 연구에 비해 상대적으로 자리매김할 수 있는 위치로 보아야 한다. 예컨대 위의 시각자료에서 우측에 위치한 연구는 그보다 좌측에 위치한 연구에 비해 상대적으로 더 큰 규모의 데이터를 다루고 있다는 것을 의미할 뿐 정확히 몇 배 더 많은 데이터를 다루고 있는지에 대해서까지 전달해주는 것은 아니다. 현재 진행 중인 연구들의 경우 데이터 량이 유동적이며 종료된 연구 가운데서도 데이터 보완이 이루어지고 있는 사례들이 있음을 고려할 때 개별 연구의 데이터 량을 정확히 파악하는 것은 현실적으로 어렵다. 어디까지나 현 시점에서 구축되어 있는 DB 또는 기존 연구논문에 부기된 데이터 규모를 통해 개략적으로 짐작하여 각 연구의 위치를 안배한 결과라 하겠다.

마찬가지로 시각자료의 상단에 위치한 연구는 하단에 위치한 연구보다 상대적으로 조금 더 구체적 연구 대상의 규모가 크거나 또는 연구하고자 하는 대상의 범위가 더 넓다는 것을 의미하지, 개별 연구가 대상으로 삼은 자료 또는 연구 범위 사이에 얼마나 큰 차이가 존재하는 것인지 그 정량적 수치에 대해서까지 알려주는 것은 아니다.

예컨대 〈그림 13〉을 볼 경우, 좌측 상단의 연구 3가지와 하단의 연구 5가지는 그 경계가 명확히 다름을 확인할 수 있다. 좌측 상단에 자리한 3가지 연구는 모두 ‘데이터 모델 설계’를 목적으로 이루어진 것들이다. ‘고문서 연구를 위한 데이터 기술 모델 연구’는 특정 주제나 단일 문헌에 주목한 것이 아니라 향후 디지털 미디어를 이용하여 이루어질 고문서 연구를 위해 필요한 기초적 데이터 모델이 무엇인지에 대해 고민한 메타적 성격의 연구라 할 수 있다. 고문서의 형식적 유형이 다양하고 그 내용 양상 또한 천차만별이기에 굉장히 폭넓은 범위의 대상을 아울러 정리하고자 한 의도를 갖는다.

‘역사 자료 텍스트의 전자적 기술에 의한 지식 관계망 구현 연구’ 또한 걸로
는 『통감절요(通鑑節要)』라는 단일 문헌을 구체적 대상으로 삼아 이루어진 연구
이긴 하지만, 그것을 토대로 역사 자료 일반을 시맨틱 데이터베이스로 편찬해낼
수 있는 방법이 무엇인지 고민하였다는 측면에서, 해당 주제의 데이터 모델을 정
립하고자 한 연구로 볼 수 있다.

‘고전문학 자료의 디지털 아카이브 편찬 연구’ 또한 특정 자료군의 DB 편찬과
관련된 기초 데이터 모델 정립을 고민한 연구이다. 한문학의 전통적인 문체 분류
에 입각하였을 때의 ‘기문(記文)’ 그 가운데서도 ‘누정기(樓亭記)’에 해당하는 작
품들을 표본으로 삼아, 고전문학 자료를 디지털 환경에서 구축하고 그 안에 담긴
의미를 재현해낼 수 있는 효과적 방안이 무엇인지에 대해 본격적으로 고민하였
다.

디지털 환경에서 특정 연구를 수행하기 위한 기초 데이터 모델 정립 또는 특
정 자료군의 DB편찬을 위한 기초 데이터 모델 설계의 맥락에서 3가지 연구는 유
사한 성격을 지니며, 이 때문에 3가지 사례 모두 실제 구축한 데이터의 규모 또
한 크지 않다. 특정한 성격의 텍스트 또는 도메인을 대상으로 한 데이터 모델 설
계는 대규모 데이터를 정리해서 이루어지기 보다는 대부분 샘플 자료라 할 수 있
는 소규모 데이터를 수집하고 정리한 것에 기초하여 진행된다. 데이터베이스를
편찬하는 것이 아니라 데이터 모델을 정립하는 것이 연구의 목적이기 때문에 굳
이 방대한 자료를 모두 입력하고 데이터로 정리할 필요가 없기 때문이다.

그러한 측면에서 ‘고문서 연구를 위한 데이터 기술 모델 연구’, ‘역사 자료 텍
스트의 전자적 기술에 의한 지식 관계망 구현 연구’, ‘고전문학 자료의 디지털 아
카이브 편찬 연구’ 3가지 연구 사례는 모두 특정한 영역에 해당하는 소규모 데이
터를 수집하고 정리하여 그것을 기초로 구체적 데이터 모델을 설계하고자 한 메
타적 성격의 연구로 묶어서 볼 수 있을 것이다. 단일 유형의 텍스트 또는 특정
성격의 도메인을 대상으로 데이터 모델을 설계하는 연구를 수행하고자 한다면
해당 세 연구가 훌륭한 참고자료가 될 수 있을 것이다. 그에 비해 하단의 연구
5가지는 구체적 DB 편찬을 목적으로 하였기 때문에 상대적으로 구축한 데이터

의 규모가 더 크다는 것을 알 수 있다.

‘수신사 및 조사시찰단 자료 DB 구축 연구’는 수신사(修信使)라는 특정 주제 영역(domain)에 해당하는 문헌을 대규모로 수집(collecting)하여 번역한 다음 해당 자료들에 담긴 내용이 각기 어떻게 연결되는지를 보여주는 DB 편찬을 목적으로 삼은 연구이다. ‘수신사’라는 한정된 주제를 다루고 있어 관련 자료의 규모 및 형식이 매우 크거나 다양하지는 않지만, 그럼에도 기초 데이터가 1,000건을 넘어가기 때문에 소규모 연구라고 하기는 어렵다.

이에 비해 ‘지압일기’ 디지털 아카이빙 연구’는 여러 유형의 문헌을 대상으로 하지 않고 단일 문헌을 대상으로 삼은 연구이다. 조선시대 생활일기를 탈초 번역하고 개별 일자의 내용에서 드러나는 미세한 요소들까지 고려한 섬세한 데이터 모델을 정립함으로써, 편찬된 DB 및 시각화 결과물을 통해 가치 있는 인문학적 의미를 발견하여 디지털 미디어로 재현하고자 하는 실험적 연구에 가깝다. 기초 데이터가 최소 5,000여건을 넘어감을 고려할 때 이 연구 또한 소규모 연구로 바라보기에는 어려운 점이 있다.

앞서의 두 연구와 비교하였을 때, ‘조선시대 표류노드 시각망 연구’와 ‘『만성대동보』 자료를 활용한 친족관계망정보시스템(LNIS) 개발 연구’는 상대적으로 DB 편찬에 집중하기보다는 시각화 구현에 더욱 치중한 연구라 할 수 있다. ‘조선시대 표류노드 시각망 연구’는 조선 시대 표류 사건이라는 특정 주제의 시각화를 목표로 하였다. 표류사건과 관련된 고문헌 자료를 수집하여 해당되는 내용을 데이터로 정리한 뒤 그것을 효과적으로 보여줄 수 있는 시각화 결과물을 구현해 내는 데 집중하였다. ‘『만성대동보』 자료를 활용한 친족관계망정보시스템(LNIS) 개발 연구’ 또한 고문헌 자료에 담긴 구체적 내용을 디지털 미디어로 어떻게 효과적으로 보여줄 수 있을 것인지 고민하고, 실제 그것을 디지털 기술로 구현하였다는 측면에서 ‘조선시대 표류노드 시각망 연구’와 유사한 성격을 지닌다. 참고로 시각화 구현을 목표로 하였다는 유사성에도 불구하고 두 연구의 구축 데이터 규모가 다른 것은, 어디까지나 연구 대상으로서 『만성대동보』가 지닌 정보가 방대하기 때문에 그로부터 기인한 것이라 할 수 있다.

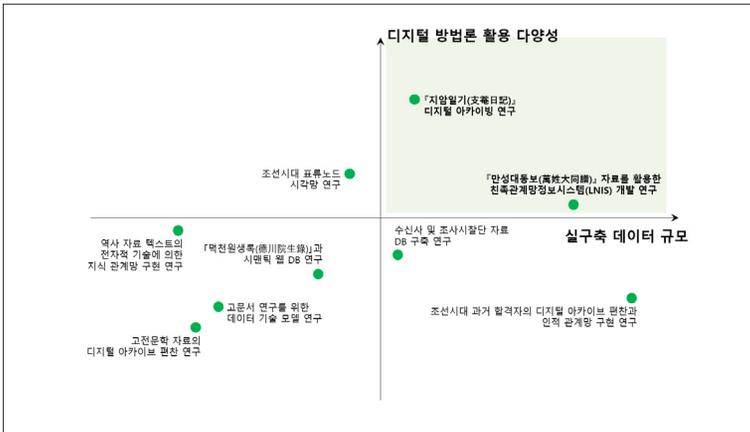
앞서의 연구들이 여러 연구자들의 협업을 통해 이루어진 것과 달리 『덕천원생록(德川院生錄)』과 시맨틱 웹 DB 연구'는 단일 연구자에 의해 수행되었다. 그러한 현실적 제약으로 인해 상대적으로 협업을 통해 이루어진 연구들에 비해 데이터 규모가 작다고 하겠다. 조선시대 지역의 서원이 지니고 있던 정치적·학술적·문화적 성격을 인물 정보에 기초하여 분석하려 하였다는 점에서, 특정 학파나 학단의 학맥을 정리하고 재구성함으로써 그로부터 학파나 학단의 정체성을 분석하고자 하는 연구의 경우 본 연구에서 차용한 방법론을 적극적으로 활용할 수 있을 것이라 판단된다.

이와 같이 하단의 5가지 연구는 데이터 모델 정립을 고민한 좌측 상단의 3가지 연구와 달리 본격적으로 DB를 편찬하고 이를 토대로 시각화 구현 또는 데이터 분석을 시도한 사례들에 해당한다. 깊어보아야 할 점은, 해당 5가지 사례가 상대적으로 매우 큰 규모의 데이터를 대상으로 수행한 연구가 아니라는 사실이다. 다수의 연구자들이 오랜 기간 참여하지 않아도 결과물 편찬까지 도달할 수 있는 디지털 인문학 연구의 경계를 직접 가늠해볼 수 있는 사례들이라는 것이다. 규모가 크지 않은 단일 영역의 대상을 탐구하여 그로부터 유효한 결과물을 만들어내고자 하는 연구자의 입장에서 참고하기에 가장 적절한 사례들이라 하겠다.

좌측 상단의 3가지 연구와 하단의 5가지 연구 사례를 비교하여 볼 때 연구 대상의 범위가 넓어지면 실제 데이터를 풍부하게 구축하기 어렵고, 반대로 데이터를 넉넉하게 갖추고 연구를 시도하고자 할 경우 연구 대상의 범위를 넓게 가져가기 어렵다는 사실을 확인할 수 있다. 그러한 측면에서 '조선시대 과거 합격자의 디지털 아카이브 편찬과 인적 관계망 구현 연구'는 매우 이례적인 연구 사례에 해당한다고 하겠다. 해당 연구는 개인 연구자에 의해 수행되었음에도, 각종 방목과 족보 자료 등 연구 대상을 폭넓게 가져가는 가운데 실제 데이터 또한 여타 연구에 비해 상대적으로 크게 구축하였다. 이것이 가능하였던 이유는 선행 연구의 영향이 크다. 와그너·송준호 교수의 문과 프로젝트, 한국학중앙연구원의 방목 전산화사업, 한국역대인물 종합정보시스템 구축 사업 등 기존에 확보된 방목 자료와 관련된 폭넓은 데이터베이스를 활용할 수 있었기 때문에 연구가 효과적

으로 진행될 수 있었던 것이다. 1차적으로는 개인 연구자의 관심과 노력에 의해 수행된 연구지만, 그 내실을 들여다볼 경우 그동안 해당 분야에서 꾸준히 데이터를 쌓아 온 여러 연구자들의 노력이 합쳐져 탄생한 연구 결과로 볼 수 있는 여지 또한 있겠다.

연구 대상의 범위가 얼마나 넓은지를 기준으로 하여 실제 개별 연구에서 구축한 데이터의 규모를 정리해 보았을 때 각 연구 사이에 드러나는 변별적 지점과 별개로, 디지털 인문학 연구로서 구축 데이터 규모 대비 활용한 디지털 방법론의 다양성을 검토하였을 때 나타나는 여러 연구 사이의 차별적 지점은 또 다른 시사점을 던져준다.



〈그림 14〉 데이터 규모와 디지털 방법론의 활용 다양성을 기준으로 배치해 본 9가지 연구 사례

〈그림 14〉에서 상단에 위치한 연구는 하단에 위치한 연구보다 상대적으로 조금 더 다양한 디지털 방법론을 활용한 연구임을 의미한다. 즉, 하단에서 상단으로 갈수록 ‘Data Model 설계’→‘DB 편찬’→‘Visualization 구현’의 맥락이 점점 강해지는 순서로 개별 연구들이 배치되어 있다고 할 수 있다. 물론 앞서 〈그림 13〉과 마찬가지로, 가로축과 세로축은 상대적 기준일 뿐 개별 연구들 사이에 얼마나 큰 차이가 존재하는 것인지 그 정량적 수치에 대해서 명확히 제시하는 것은

아니다.

예컨대 9가지 사례는 대부분 연구 기초 단계에서 데이터를 효과적으로 정리하고 공유하기 위해 프로테제(Protégé™)와 같은 에디터를 이용하여 온톨로지(Ontology)를 설계하였으며, 이에 근거하여 Media Wiki™ 플랫폼의 DB를 구축하였다. 그리고 거의 모든 연구가 XML, RDF, OWL 등의 기술을 활용하여 대상 자료에 담긴 내용을 시맨틱 데이터(Semantic Data)로 정리하였다. 데이터 모델을 설계하고 데이터베이스를 구축하는 큰 틀에서는 유사한 방법론을 취했다고 하겠다.

9가지 사례가 모두 데이터베이스 구축의 큰 맥락에서 활용한 디지털 방법론이 유사하다고 하더라도, 유추할 수 있다시피 하단의 5가지 사례와 상단의 3가지 사례는 다소 차이가 있다. 하단의 5가지 사례는 ‘데이터 모델 설계’ 또는 ‘데이터베이스 편찬’ 또는 ‘데이터의 효과적 공유’에 집중한 연구이기 때문에 기술적으로 다양한 디지털 방법론을 시도하고 활용하였다고 보기는 어렵다. 데이터를 어떻게 효과적으로 정리하여 공유할 것인가에 초점을 두고 그에 최적화된 디지털 방법론 활용을 궁구하는 것과 구축된 데이터를 토대로 데이터 간의 관계를 어떻게 효과적으로 보여줄 것인가에 집중하여 디지털 방법론을 적용하고자 하는 것은 근본적으로 다른 맥락의 사유에 기초하기 때문이다.

거칠게 비유하자면, 완성된 그림을 사람들이 많이 볼 수 있게끔 하기 위해 어디에 걸어놓을 것인가에 대해 고민하는 것과 그림이 지닌 특징을 효과적으로 드러내기 위해 어떠한 액자를 활용할 것인가의 고민이 다른 지점에 있는 것과 유사하다. 그림을 걸어놓는 위치는 공간의 전체적 구조나 특징에 영향을 받지만, 액자의 활용은 그림 자체의 색감이라든지 성격 그리고 시장에 유통되는 액자의 다양성 등에 보다 많은 측면에서 영향을 받는다. 전자의 경우에 비해 후자가 훨씬 더 복잡한 맥락에 있으며, 이는 곧 후자의 디지털 방법론 적용이 ‘표현’과 보다 더 밀접한 관계를 맺고 있음을 의미한다. 그리고 그 ‘표현’은 웹 환경에서 가능한 여러 형식의 디지털 리터러시에 대한 경험과도 상응하며, 이는 곧 웹 환경의 여러 콘텐츠를 통해 커뮤니케이션을 추구하는 현대인들의 일상과 보다 더 유효하

게 닿아 있음을 의미한다.

예컨대 우측 상단의 2가지 사례 ‘『지암일기』 디지털 아카이빙 연구’와 ‘『만성 대동보』 자료를 활용한 친족관계망정보시스템(LNIS) 개발 연구’는 구축된 데이터를 토대로 데이터 간의 관계를 어떻게 효과적으로 보여줄 것인가에 집중한 연구로서, 시맨틱 데이터를 표현하기 위해 Neo4jTM와 같이 Graph Database를 활용하거나 GephiTM와 같은 데이터 시각화 도구를 이용한 여타 연구와 달리 D3.js 기술을 적용하였다. D3.js는 Neo4jTM Database와 같이 쿼리를 이용하여 즉각적으로 그래프를 구현할 수 있는 도구도 아니며, R과 같이 데이터에 기초하여 그래프를 그릴 수 있는 기본 형식을 제공하는 것도 아니다. 오히려 코드를 작성하여 시각화 결과물을 구현하기 때문에 웹 환경 또는 웹 프로그래밍에 대한 어느 정도 숙련된 이해가 전제되어야 한다. 대신 php나 파이썬(Python)과 연계하여 RDB나 Graph Database에 저장된 데이터를 직접 불러와 매우 다양한 형태로 시각화 결과물을 구현하는 것이 가능하기 때문에 원하는 방식대로 데이터를 디자인하고 그것을 토대로 효과적인 분석을 시도할 수 있다. 데이터 설계와 데이터베이스 편찬 단계에서 효과적으로 적용할 수 있는 디지털 방법론이 무엇인지 고민하는 데 그치지 않고, 더 나아가 시각화 구현의 측면에서 데이터를 보다 폭넓게 표현할 수 있는 방안이 무엇인지 적극적으로 고민한 연구들에 해당한다.

디지털 방법론을 상대적으로 연구에 더 많이 활용하였다고 해서 해당 연구가 더 우수한 연구라고 할 수는 없다. 디지털 방법론의 적용은 분명 디지털 리터러시의 경험 즉, 그것 자체가 ‘목적’으로서의 가치도 지니지만 특정 연구를 위한 ‘도구’의 차원에서 이해할 수 있는 여지가 아직까지는 강하기 때문이다. 그리고 기술은 항상 변화하고 그에 따라 새로운 방법론이 등장하기 때문에 현 시점에서 특정한 디지털 방법론 자체가 연구에 있어서 절대적 가치를 갖는 것 또한 아니다. 다만 디지털 인문학 연구에서 ‘디지털 방법론의 활용’을 통해 기대할 수 있는 가장 큰 효과 가운데 하나가, 대중들로 하여금 접근하기 어려운 학술 자료를 웹 환경의 여러 다채로운 미디어를 통해 손쉽게 공유할 수 있게 해준다는 사실임을 놓쳐서는 안 된다. 그러한 측면에서 웹에서 활용 가능한 다양한 디지털 기술을

연구 방법론에 능동적으로 적용해 나가는 시도는 앞으로도 디지털 인문학 연구에서 추구해야 할 가장 큰 방향성 가운데 하나이지 않을까 판단된다.

9가지 연구 사례를 ‘데이터의 규모’와 ‘연구 대상 범위’ 그리고 ‘디지털 방법론 활용 다양성’에 맞추어 배치해본 의도는 각각의 연구를 평가하기 위한 것이 아니다. 현 단계에서 완료되지 않은 연구들이 있을 뿐만 아니라 현 시점이 국내 디지털 인문학 연구의 초기 단계임을 고려할 때, 평가 기준을 마련하여 그에 따라 개별 연구 사례의 장단점을 나열하는 것은 적절하지 않다 여겨진다. 실제 개별 연구의 맥락이 다르기 때문에 연구 목적 대비 결과의 유효함, 연구 방법론의 적절성, 첨단 기술 활용 여부 등 그와 관련된 요소들을 일률적으로 평가할 수 있는 기준을 설정하는 것도 매우 어렵다. 여차저차 그러한 평가기준을 마련한다 하더라도 평가를 통해 장단점을 이야기할 경우 향후 디지털 인문학 연구의 외연을 넓힐 수 있는 여지를 오히려 좁혀 버리는 결과를 낳을 가능성이 크다. 지금은 ‘유효한’ 연구가 선택적으로 적절하게 이루어지기를 요구할 시점이 아니라 ‘명확하게 유효하지 않을 지라도’ 오히려 다양한 연구가 가급적 많이 시도되어 연구 맥락이 풍부해지기를 바라야 하는 시기이기 때문이다. 앞서 9가지 연구 사례가 모두 고문헌 자료를 활용한 디지털 인문학 연구 또는 한문학 영역에서 시도할 수 있는 디지털 인문학 연구의 훌륭한 참고 요소가 될 수 있음을 인지할 필요가 있다. ‘데이터의 규모’와 ‘연구 대상 범위’ 그리고 ‘디지털 방법론 활용 다양성’에 맞추어 9가지 연구 사례를 살펴본 것은 바로 그러한 측면에서 비롯된 의도임을 밝히는 바이다. 향후 이러한 연구사례들로부터 정리된 디지털 인문학 연구방법론을 발판으로 하여 더욱 다채롭고 풍부한 디지털 인문학 연구가 시도되기를 바란다. 그 때가 되면 개별 연구 사례들에 대한 구체적 분석 및 장단점 개괄을 통해 특정한 맥락의 연구를 수행하기 위한 보다 효과적이고도 적합한 방식의 디지털 인문학 연구 방법론을 이야기할 수 있을 것이다.

3. 한문학 연구와 스몰 데이터(Small Data)

디지털 인문학은 인문학 연구를 통해 정리되는 여러 단위의 정보와 지식을 디지털 데이터로 변환하여 수행하는 연구이기 때문에, 기본적으로 ‘데이터(Data)’라는 키워드와 밀접한 관련성을 갖는다. 이와 같은 태생적 특성으로 인해, ‘데이터’와 관련하여 자주 언급되는 몇몇 키워드들이 학술적으로 디지털 인문학과 어떠한 직접적 관련이 있는지를 논의하는 경우와 종종 마주칠 때가 있다. 그 가운데서도 근래에 가장 많이 접할 수 있는 키워드는 ‘빅데이터(Big Data)’이다.

다룰 수 있는 데이터의 규모가 커진다면 상대적으로 데이터를 가지고 시도할 수 있는 연구의 범위와 데이터를 통해 발휘할 수 있는 직접적인 연구 결과의 외연이 더욱 확장될 것이다. 그러한 차원에서 여러 영역의 인문학 연구를 통해 산출되는 수많은 정보를 종합하여 빅데이터로 구축·활용하는 방안에 대한 논의가, 현 단계에서는 모호하게 받아들여질지 모르겠지만 향후 매우 중요한 학술적 과업이 될 가능성이 크다.

한편으로는 빅데이터가 데이터의 ‘규모’에 기준을 둔 키워드임을 고려할 때, 그보다 더 중요할 수 있는 데이터의 질적인 측면에 대한 논의는 상대적으로 많이 부족한 상황이다. 데이터를 통해 얻고자 하는 지식이나 이해의 수준이 정교하기 위해서는, 데이터의 ‘규모’ 뿐만 아니라 데이터가 얼마나 질적으로 정교하게 구축되었는가의 맥락 또한 중요하다.

예컨대 고전종합DB는 2015년 기준 구축된 한국문집 양만 89종 364책에 이르며, 2015년 한 해 페이지뷰만 22,519,890건을 기록한 ‘빅데이터’에 가까운 고전자료 데이터베이스라고 할 수 있다.³⁵⁾ 아쉬운 점은, 고전종합DB를 통해 확보할 수 있는 데이터의 경우 연구자가 바로 스스로의 연구에 적용하기 어렵다는 사실이다. 현 시점에서 제공되는 한국문집총간 XML 데이터의 경우³⁶⁾, 데이터를 정

35) <http://db.itkc.or.kr/etc/desc?type=dbStat>

36) 현재 고전번역원이 구축해 놓은 고전종합DB 상의 ‘한국문집총간’ 원문과 번역문 데이터는, 공공데이터 포털(<https://www.data.go.kr/dataset/3074298/fileData.do>)을 통해 XML 포맷으로 다운받을 수 있다.

교하게 정리하고 들여다볼 수 있는 섬세한 스키마(Schema) 또는 온톨로지(Ontology)가 적용되어 있지 않다. 그렇기 때문에 한국문집총간 XML 데이터를 활용하여 고문헌 자료에 담긴 내용을 연구하고자 할 경우, 해당 연구의 성격과 규모에 맞게 온톨로지를 새롭게 설계하고 Wiki와 같이 텍스트 정보를 입력·정리할 수 있는 플랫폼을 통해 데이터를 섬세하게 정리하는 단계를 거쳐야 한다.

현 시점의 고전종합DB가 '빅데이터'에 가까운 데이터베이스로서 제공할 수 있는 명확한 이점은, 방대한 양의 문집총간 자료를 넘나들며 쉽고 간단하게 원하는 데이터(B)에 접근할 수 있게끔 해준다는 점이다. 그러나 그 데이터(B)를 가지고 특정한 목적의 연구를 수행하기 위해서는 보다 섬세한 형태의 새로운 데이터(A)로 가공해야 한다. 그리고 그 데이터(A)는 빅데이터에서 제공받은 데이터(B)로부터 비롯된 것이긴 하지만, 빅데이터의 형식과는 다른 스몰 데이터(Small Data)에 해당한다.

데이터는 어떤 주체가 특정한 목표나 의도를 갖고 만들어낸 것으로, 주어진 문제를 해결하기 위해 적절히 활용될 때 그 가치가 발현된다. 물론 주어진 문제가 다차원의 복잡성을 갖고 있다면 충분히 규모 있는 데이터가 효과적이다. 그러나 주어진 문제를 해결하는 데 필요한 최소의 데이터만으로 효과적 결과를 얻는 것도 가능하다. 스몰 데이터(small data)는 규모 중심의 빅데이터 현상과 구별하여 품질이 높은 최소의 데이터를 말한다. 일반적으로 빅데이터는 데이터의 수집과 분석을 위해 시간과 비용이 상당히 소요된다. 스몰 데이터는 특정 문제에 대한 상황을 이해하기 위해 필요한 데이터만을 수집하고, 신속한 분석을 통해 문제를 해결할 수 있는 기회를 제공한다.³⁷⁾

위의 인용에서 강조하듯이, 스몰 데이터는 빅데이터와 반대되는 말로 그 규모가 작은 데이터를 가리킨다. 규모가 작은 데이터가 왜 필요한가에 대한 의문을 제기한다면, '필요한 최소의 데이터만으로 효과적 결과를 얻는 것도 가능'하다는

37) 김학래(2017), 『지식그래프』, 커뮤니케이션북스, 5~6면.

답을 이야기할 수 있을 것이다. 예컨대 앞서 제시했던 디지털 인문학 연구 사례 가운데 ‘고전문학 자료의 디지털 아카이브 편찬 연구’에서 구축하였던 12편의 누정기 데이터 또한 스몰 데이터에 해당한다. 한국문집총간 전체 데이터 또는 한국문집총간에서 추출할 수 있는 누정기 전체 데이터 또는 조선시대 누정기 전체 데이터와 비교할 시 매우 작은 규모의 데이터이지만, 한편으로 ‘한문 산문의 한 장르로서 누정기가 지닌 여러 정보를 체계화하여 시각화할 수 있고 이를 통해 분석까지 가능한 데이터 모델은 무엇일까?’에 대한 의문을 해결하기 위한 목적에서는 충분히 기초 데이터로 기능할 수 있는 규모라고 할 수 있다.

빅데이터 개념을 인문학 연구와 연계하여 그 생산성을 가능하는 고민도 꾸준히 이루어져야 하겠지만, 상대적으로 소규모 인문학 연구에 적용하여 유의미한 학술적 결과를 얻어낼 수 있는 스몰 데이터 개념에 대한 접근 또한 매우 필요하다고 생각된다. 특히 한문학의 경우 작가론, 작품론, 문체론 그리고 특정시기의 문학풍조나 특수한 성격을 지닌 문학적 개념과 소재를 대상으로 연구하는 데 있어서 빅데이터 개념에 기초한 ‘규모’의 접근보다는, 오히려 스몰 데이터 개념에 근거한 정교한 데이터 모델을 통해 의미 있는 지식을 생산해낼 수 있는 가능성이 더욱 크다. 그리고 한문학 연구에서 스몰 데이터를 이용한 연구가 활발해지고 그로부터 여러 유의미한 연구 결과들이 누적될 경우, 그것은 향후 빅데이터에 가까운 고전번역원의 고전종합DB를 보다 미래적인 형식의 데이터베이스로 전환하는 데 있어서 훌륭한 기초자원으로 활용될 수 있다. 스몰 데이터를 활용한 연구가 양적으로 많아질 경우, 향후 구축되거나 개선될 빅데이터의 질적인 수준을 끌어올리는 데 있어서 중요한 영향을 미칠 수 있는 것이다. 막연하면서도 선불리 접근하기 어려운 빅데이터 한문학 연구가 아니라 구체적이고도 현실적인 연구 방법으로서의 스몰 데이터를 활용한 한문학 연구가 앞으로 활발하게 시도되기를 기대해본다.

■ 참고문헌

1. 연구 논저

- 김하영(2015), 『광산김씨 문중 고문서 시맨틱 데이터베이스 모델 설계: '門中古文書 디지털 아카이브 구현 연구』, 한국학대학원 석사학위 청구논문.
- 김학래(2017), 『지식그래프』, 커뮤니케이션북스.
- 김현(2016), 『디지털 인문학과 고문헌 자료 연구』, 『열상고전연구』 50집, 열상고전연구회, 14~38면.
- 김현·임영상·김바로(2016), 『디지털 인문학 입문』, HUEBOOKS.
- 김현·안승준·류인태(2016), 『고문서 연구를 위한 데이터 기술 모델』, 역사학회, 제59회 전국역사학대회 (주제: 기록의 생성과 역사의 구성) 발표논문, 2016년 10월 28일.
- 류인태(2016), 『디지털 환경에서의 인문 지식 연구에 관한 小考 -수신사 자료 DB 편찬 프로젝트를 중심으로』, 『열상고전연구』 50집, 열상고전연구회, 101~139면.
- 류인태(2017), 『덕천원생록(德川院生錄)과 시맨틱 웹 DB 연구』, 한국선비문화연구원, 2017년도 한국선비문화연구원 학술대회 (주제: 한국선비문화연구원의 장기 연구 방향 모색) 발표논문, 2017년 12월 15일.
- 박순(2017a), 『고전문학 자료의 디지털 아카이브 편찬 연구 : 누정기(樓亭記) 자료를 중심으로』, 연세대학교 대학원 박사학위 청구논문.
- 박순(2017b), 『고전문학 자료의 디지털 아카이브 편찬 연구』, 보고서.
- 백광렬(2017), 『조선 후기 '양반지베네트워크'의 성격과 구조변동 : 상층양반의 친족연결망을 중심으로』, 서울대학교 대학원 박사학위 청구논문.
- 윤종웅(2017), 『역사 자료 텍스트의 전자적 기술에 의한 지식 관계망 구현 연구 : '通鑑節要'와 '譯註通鑑節要'를 중심으로』, 한국학대학원 박사학위 청구논문.
- 이수진(2015), 『조선시대 표류노드 시각망 구축 과정 -표류 기록의 의미 요소 추출을 중심으로』, 『溫知論叢』 45집, 온지학회, 161~186면.
- 이재욱(2017), 『조선시대 科擧 合格者의 디지털 아카이브 편찬과 인적 관계망 구현』, 한국학대학원 박사학위 청구논문.
- 제이 데이비드 볼더 저·김익현 역(2010), 『글쓰기의 공간』, 커뮤니케이션북스.
- 하영휘·백광렬(2015), 『"만성대동보(萬姓大同譜)" 자료를 활용한 '친족관계망정보시스템(LNIS, Lineage Network Information System)' 개발 -개발실무와 활용방안, 이론적 함의』, 진인진.
- 허경진·구지현(2016), 『조선시대 표류노드 시각망 연구일지』, 보고서.
- 홍정욱(2015), 『디지털기술 전환 시대의 인문학 -디지털 인문학 선언문을 통한 고찰』, 『인문콘텐츠』 38, 인문콘텐츠학회, 41~74면.
- spiro, Lisa, "This Is Why We Fight": Defining the Values of the Digital Humanities", Debates

in the Digital Humanities. Ed. Matthew K. Gold, MN: University of Minnesota Press, 2012.

2. 참고 URL

- 조선시대 표류노드 시각망 디지털 아카이브 :
http://www.digerati.kr/mediawiki/index.php/표류_노드_시각망
- 친족관계망정보시스템(LNIS) :
<http://www.lnis.kr/>
- 문중 고문서 시맨틱 DB :
<http://cefia.aks.ac.kr:7474/browser/>
- 수신사 및 조사시찰단 자료 디지털 아카이브 :
<http://susinsa.info/>
- 『지암일기』 디지털 아카이브 :
<http://jiamdairy.info/>
- 역사 자료 텍스트의 전자적 기술에 의한 지식 관계망 구현 연구 위키 :
<http://dhlab.aks.ac.kr/~yiw2999/wiki/index.php/대문>
- 조선시대 과거 합격자의 디지털 아카이브 편찬과 인적 관계망 구현 연구 위키 :
http://dh.aks.ac.kr/~sonamu5/wiki/index.php/SEDB:과거_합격자_정보_디지털_아카이브
- 덕천원생록과 시맨틱웹 DB 연구 :
http://dh.aks.ac.kr/~red/wiki/index.php/덕천원생록과_시맨틱웹_DB_연구
- 고전번역원 고전종합DB 통계 :
<http://db.itkc.or.kr/etc/desc?type=dbStat>
- 공공데이터 포털의 한국문집총간 자료 XML 제공 웹페이지 :
<https://www.data.go.kr/dataset/3074298/fileData.do>

Digital Humanities and the study of Korean-Chinese Literature

– An introduction of digital humanities researches for the
medieval books and documents of Korea –

Ryu In-tae*

As debates in the digital humanities has continued to grow largely across the world since the mid 2000s, there also has been a growing interest in the digital humanities in Korea. In spite of the growing interest in the digital humanities, it is disappointing that few researches have been done in Korea. Recently, there has been a lot of discourse about the concept of 'digital humanities' in Korea, but the fact that discussions on what is the specific methodology of digital humanities research have not been active is deeply related to this realistic situation. In fact, digital humanities in Korea is still at an early stage. Therefore, in order to have a complete understanding of digital humanities and to avoid embracing imprudent opinions, we have to approach cautiously so as to remove abstractness from discussions on digital humanities. With this as consideration, it is necessary to epitomize humanities researches in Korea that uses digital platforms in earnest. There are nine cases of digital humanities researches in this paper. Though each case has its own object that distinguishes one case from another cases, digital platforms that they are based on are very similar to one another's. Web Ontology Language, Network Graph Visualization and Media

* Academy of Korean studies / E-mail: ryuintae@gmail.com

WikiTM are the examples. It's likely that scholars and researchers who want to study digital humanities can reach a better understanding by reading the cases presented in this paper.

Key Words: Digital Humanities, Korean–Chinese Literature, Medieval Books and Documents of Korea, Small Data, Cultural Informatics

본 논문은 2018년 1월 15일 투고되어 2018년 2월 22일 심사를 완료하여
2018년 2월 26일 게재를 확정하였음

www.kci.go.kr