

인문학술 데이터 프로세싱에 관한 시론

류인태

성균관대학교 국어국문학과 박사후연구원, 인문정보학 전공

ryuintae@gmail.com

I. 머리말

II. 인문학술 데이터 편찬 과정의 실험적 모색

III. 참여형 인문데이터 편찬 모델 사례: Linked Jazz

IV. 맺음말

이 논문은 '천정환 외 4인, 『(경제·인문사회연구회 인문정책연구총서 2021-02) 인문학과 지식공유: 시민 참여 모델의 개발』, 경제·인문사회연구회, 2022.'에 수록된 원고 가운데, 「5.4. 디지털 환경에서의 인문지식 공공성 확대 방안 모색: 데이터 기반 인문지식 공유 모델 구현 연구」(97~124쪽)의 내용을 수정·보완한 것이다. 해당 연구에 공동으로 참여해 여러 조언을 아끼지 않으신 천정환, 박서현, 박지훈, 안혜연 네 분 선생님께 작게나마 감사의 인사를 드리며, 연구 과정에서 구체적 기여를 해주신 장문석, 김병준, 최동혁, 김지선 네 분 선생님께도 늦었지만 감사의 인사를 올린다.

I. 머리말

상식(common sense)으로 통용되는 일반적 성격의 지식과 달리, 인문학 전공 연구자에 의해 편찬되는 고급 학술 지식은 생산·유통되는 채널이 한정되어 있다. 웹 리소스만 가지고서는 탐구할 수 없는 층위의 정보가 대부분이기에, 아날로그 환경에서 자료를 찾아 모으고 다시 그것들을 읽어 나가면서 그 사이의 관계를 변별하고 그로부터 의미를 발견하는 과정이 필수적이다. 그리고 그에 대한 내용을 한 편의 ‘학술논문’으로 정리해서 발표하는 프로세스는, 오랜 기간 학술 행위의 공인된 인준 형식으로 기능해 왔다. 학부나 대학원의 인문학 교과 과정에서 아날로그 환경의 문헌을 매개로 이루어지는 읽기와 쓰기 훈련은 바로 그러한 과정을 자연스레 익혀나가는 절차에 해당한다.

문제는, 그러한 훈련을 거쳐 고급 학술 지식으로 세상에 나온 인문학 분야의 ‘학술논문’이 읽히지 않는 경향이 점점 더 확대되고 있다는 것이다. 심지어 연구자들 사이에서조차도 ‘내 연구 분야와 직접적 관련이 없는 논문은 읽지 않는다’, ‘아무도 읽지 않을 논문을 쓰고 있다’와 같은 푸념 섞인 이야기를 듣는 것이 어렵지 않은 세태이다. 그 이유에 대해서는 여러 맥락에서 살펴볼 수 있겠으나, 디지털·데이터 기술이 보편화됨에 따라 지식 커뮤니케이션 환경에서의 리터러시 형식이 급변하고 있는 정황을 가장 먼저 거론할 수 있다.

지식을 배포하는 미디어 자원이 다양해지고 그에 따라 읽기 문화가 조금씩 축소되어 가는 가운데, 논문을 통해 생산되는 전문 지식이 대중들 사이에서 직접적으로 유통·소비될 수 있는 가능성은 점차 줄어들고 있다. 스마트 디바이스가 널리 보급되면서 멀티스크린 환경이 보편화되고, 그 연장선상에서 웹 기반의 영상 미디어를 중심으로 한 디지털 콘텐츠의

영향력이 증대함에 따라, 인문적 성격의 학술 지식 또한 영상을 매개로 소비되는 경향이 확산하는 추세이다.

소위 ‘텍스트’ 미디어로서 논문을 직접 읽는 경험은 축소되고, 상대적으로 그러한 고급 지식을 영상 미디어로 쉽게 잘 풀어 전달할 수 있는 ‘지식 큐레이터’로서의 전달자들이 지상파 채널이나 유튜브와 같은 미디어를 통해 등장하고 주목받는 최근의 추세 또한 그러한 징후 가운데 하나이다. 이러한 풍조는 웹 환경 그리고 스마트 디바이스와 멀티스크린에 상대적으로 익숙한 젊은 세대로 갈수록 더욱 두드러진다. “검색하면 다 나오는데 뭐하러 책을 읽어요?”라는 다소 심드렁한 태도는, 웹상에 마련된 키워드 중심의 검색 환경을 채널로 삼아 멀티스크린을 매개로 소비되는 지식의 파편적 교환 양상이 우리 사회에서 일반적 양태로 자리 잡아 가고 있음을 넉넉히 알려준다. 이러한 경향이 지속된다면, 웹에서 검색 결과가 없는 정보나 영상으로 만들어져 유통되지 않는 지식은 세상에 없는 것으로 여겨지는 극단적 풍조가 언젠가 생겨날지도 모를 일이다.

스크린을 매개로 한 멀티미디어의 보편화는 텍스트를 대상으로 한 ‘읽기’ 경험과 그 문화적 외연을 지속적으로 축소하고 있으며, 디지털 포맷으로 출력되어 나오는 다량의 자료는 ‘개인이 절대 읽을 수 없는’ 분량임을 강조하며 꼼꼼히 읽기(close-reading)로서의 전통적 ‘읽기’의 한계를 끊임 없이 지적하고 있다. ‘고급 지식을 다루는 영상미디어에 어떻게 비판적으로 대응할 것인가’, ‘웹 환경에서 끝없이 늘어나고 있는 지식 자원을 어떻게 다룰 것인가’에 대한 구체적 고민이 반드시 있어야 하는 상황이다. 그러한 적극적 고민 없이 그동안 지속되어온 아날로그 환경에서의 학술 지식 유통 체제를 지속하기란 쉽지 않을 것이다.

1 김성우·엄기호, 『유튜브는 책을 집어삼킬 것인가』(따비, 2020), 163쪽.

한편으로 전문 지식을 담고 있는 학술논문을 읽는 행위 또는 그러한 맥락의 읽기 문화가 축소되는 데 있어서 '디지털·데이터 환경의 확대라는 기술적 요인만 영향이 있는 것이냐'라고 질문을 한다면 반드시 그렇지만도 않다. 분과 전공을 매개로 설립된 학회와 연구소가 오랜 기간 대학을 중심으로 운영되어 오면서, 학회와 연구소에서 간행하는 학회지를 '등재지'로 평가하고 그에 수록된 학술논문을 연구 업적으로 삼는 평가시스템이 2,000년대 이후로 지금까지 줄곧 유지되어 왔다. 그로 인해 대학의 분과 전공을 중심으로 한 학술 시스템은 더욱 공고해졌다.

'전공(major)'은 특정 분야에 대한 '전문성'을 전제하는 요인이기에, 해당 분야에서 다루어지는 깊은 이해를 발견·전달하는 데 초점을 둔다. 다만 이러한 지향이 오랜 기간 지속됨에 따라 유관 분야의 연구자만 이해할 수 있는 전문적 성격의 어휘와 특정 전공 분야에서만 다루어지는 개념어를 기반으로 작성된 연구 결과물로서의 '학술논문'이 쉽 없이 쏟아져 나오고 있는 것 또한 사실이다. 전문성의 장벽이 높기 때문에 대중을 거론할 필요도 없이 인접 분야의 연구자들조차도 해당 전공에 관한 일정 이상의 조예가 없다면, 논문을 온전히 읽어내는 것이 쉽지 않다.

이와 관련해 굳이 전공자가 아닌 이들에게 특정 분야의 전문 지식을 직접 다루는 학술논문을 읽도록 요구하는 것이 부적절하다는 반문이 튀어나올 수도 있다. 학술논문과는 다른 포지션으로 '대중서'의 기능과 역할을 고려하는 시각이 그러한 입장에 가깝다. 학술 행위 또는 학술 지식이 내포한 일반적 속성을 전제할 경우 그러한 입장이 일견 타당해 보일 수도 있겠으나, 여타 분과 학문과는 다른 인문학의 특수성을 고려한다면 꼭 그렇게만 볼 일도 아니다.

예컨대 이공계열(STEM) 연구자에게 인간과 사회에 관한 지식을 바탕으로 철학·사회적 담론과 가치를 활발하게 생산하고 유통해 줄 것을 기대하는

경우는 거의 없다. 인간과 사회에 대한 가치와 담론을 표명하고 그와 관련된 지식을 정리·체계화하는 것이 이공계열 연구의 목적과 역할이 아니기 때문이다. 그렇기 때문에 엄밀하게 이야기할 경우, 이공계열 연구에서 연구 과정과 결과물을 정리하는 절차로서의 논문 쓰기는, 해당 분야에서의 고도화된 ‘전문성’을 표준적인 지식 체계하에 합당하게 드러내고 그것을 인준 받는 기능적 차원에서의 과업에 가깝다.

그에 반해 인문학 연구자에게 있어서 논문 쓰기는 자신이 소속된 분과 학문의 공인된 인준 과정을 통과하는 기능적 의례 그 이상의 의미가 있다. ‘논문’이란 자신이 발견한 지식을 정리·체계화함으로써 소속된 분과 전공에서 통용되는 전문적 학술 역량을 인준 받을 수 있는 매개물일 뿐만 아니라, 거기서 더 나아가 인문학 연구자로서 오랜 기간 탐구해온 인문학적 가치와 담론을 섬세하게 재현하는 무대(場, stage)이기 때문이다. 새롭게 발견한 지식을 인준 받는 기능적 채널과 별개로, 인간과 사회에 대한 가치와 담론을 재현하는 속성은 ‘글쓰기’라는 매체 행위와 직결되기 때문에, 여타 분과의 연구자가 아닌 소위 ‘글쟁이’로서 인문학 연구자들에게만 부여되는 일종의 특수한 역할에 가깝다.

이러한 ‘글쓰기’란 지식을 전달하는 기능적 행위가 아니라 자본으로 환원할 수 없는 가치와 담론에 관한 소통 행위로 이해할 수 있다. 읽고 쓰는 문화는 곧 가치와 담론으로 표방되는 사회적 생각의 토대이며, 그러한 사회적 생각을 이끌어가는 가시적 매개로서 인문학자들이 읽고 쓴 글은 해당 사회의 지적 수준과 사회적 문제에 관한 의식을 반영하기 때문이다.²

2 이와 관련해 최근 몇 년 사이 논란이 된 철학 연구자 윤지선의 논문 사태를 거론할 수 있다. ‘윤지선, 「‘관음충’의 발생학: 한국남성성의 불완전변태과정(homomorphism)의 추이에 대한 신물질주의적 분석」, 『철학연구』 127집(2019), 259~288쪽.’ 해당 논문은 인터넷 방송인 보검의 특정 표현에 대한 서술 및 해석과 관련해 논란을 낳았을 뿐만 아니라 이것이 법적 공방으로 확대되면서, 해당 논문과 관련된 비판

이와 관련해 ‘굳이 논문을 통해서 그런 이야기를 할 필요가 있느냐. 논문 작성은 전문적 학술 지식을 인준 받는 채널로 삼고, 가치와 담론에 대한 제시는 대중적 저술을 통해서도 얼마든지 가능하다’와 같은 입장이 있을 수 있다. 그러나 ‘대중서’의 경우 소비적 입장이 전제된 대중의 읽기에 초점을 둘 수밖에 없다는 측면에서 기본적으로 교환의 논리가 전제된다. ‘지금 쓰고 있는 글이 팔릴 수 있는지 없는지’의 관점에 휘둘린다면, 애초 글쓰기의 성격이 달라질 수밖에 없는 것이다.

한편으로 그와 같은 인문학 글쓰기의 특수성을 고려한다고 하더라도, 디지털·데이터 기술의 발달에 따라 미디어 커뮤니케이션 환경이 급변하는 현실적 상황에 대한 검토 없이, 그저 아날로그 환경 기반의 ‘강단’과 ‘전공’ 중심의 전통적 인문학 지식 유통 채널을 고수하고만 있는 인문학 연구자들의 태도를 마냥 긍정적으로 보기는 어렵다. ‘그동안 줄곧 강단에서 전공 분과의 전문성을 추구해왔고, 앞으로도 계속 순수학문으로서의 위상에 주력하는 것이 인문학을 살리는 길이다’와 같은 보수적 태도는, 인문학의 본령을 지키고자 하는 자주(自主)의 표명이라기보다는, 오히려 그동안의 구색(具色)을 유지하기 위한 변명에 가깝다는 비판이 적지 않다.

학령인구 급감과 그에 따른 대학 정원 감축으로 인해 대학 구조조정의 최우선 대상으로 인문학 전공 학과들이 도마 위에 오르고 있음에도, 그에 능동적으로 대응할 수 있는 자생적 역량이 없는 현실은, 바늘구멍 취업문을 통과하기 위한 실적 쌓기를 핑계로, 연구자들에게 오로지 논문 쓰기만을 강제하고 있다. 질이 아니라 양으로 대변되는 현재의 업적평가 체제는 소위 양산형 논문을 기계적으로 쓸 수밖에 없는 상황으로 연구자를 몰아가고 있으며, 이러한 상황은 개별 학술논문의 질뿐만 아니라 학계 전체의

적 논쟁이 대중과 전문연구자들 사이에서 현재까지도 계속 이어지고 있는 상황이다.

학술 활동을 건강하지 않은 상태로 유도하는 결과를 낳고 있다.

이러한 상황을 종합적으로 살펴보았을 때, 학술논문을 읽는 행위 또는 고급학술 지식을 다루는 문화가 축소되어가는 양상을, 단순히 ‘글쓰기’로 대변되는 인문학 연구 방법론이 지닌 한계적 문제로만 환원해서는 곤란하다. 오랜 기간 인문학계에서 유지되어온 전공 간 장벽의 문제 그리고 그에 입각한 글쓰기가 업적평가 제도와 맞물려 다시 취업 채널까지 연결되는 인문 학술 환경의 현실적 여건 전반과 관련이 있다고 보아야 할 것이다. 게다가 디지털 기술의 발달과 영향력 확대에 따른 리터러시 형식의 급격한 변화는, 학술환경의 현실적 문제와는 다른 맥락에서 ‘학술적 글쓰기’의 변모를 끊임없이 요청하고 있다.

이와 같은 현실에 착안하여 이 글은 본론을 두 부분으로 나누어 크게 두 가지 문제의식을 풀어내는 데 주력하였다. 첫 번째는 ‘데이터’를 매개로 각각의 학술논문을 아카이빙하고 그것들에 담긴 지식을 연결하는 방안을 실험적으로 모색해보는 것이다. 인문학 방면의 학술 활동은 전통적으로 연구자 개인의 논문 작성 및 발표를 매개로 ‘지식’을 직접 전달하는 데 주력해왔다. ‘글쓰기’는 연구자 개인의 문제의식과 학술적 지향을 전제로 지식을 다루는 단독적 채널로 여겨져 왔으며, 이에 따라 ‘쓰기’와 ‘읽기’는 인문학 지식을 인코딩하고 디코딩하는 유일한 경험으로 수용되어 왔다. 논문을 ‘읽고 쓰지’ 않으면 학술 지식을 얻지 못하는 상황에서 분과 전공을 기준으로 분기마다 수백 편씩 쏟아져 나오는 학술논문은 개별 전공의 장벽을 더욱 공고히 하는 데 일조하고 있으며, 그로 인해 전공 간 지식의 교류와 생산적 토론은 점점 더 요원한 일이 되어가고 있다.

전공 지식이 연구자들이 추구하고자 하는 궁극적 삶에 가까운 것인가에 대해 질문을 던진다면 대부분 ‘아니’라는 답을 할 것이다. 특수한 것(the special)이 아니라 보편적인 것(the general)을 찾고자 하는 태도야말로 학술

활동의 본질에 가깝기 때문이다. 그런데도 현실에서의 학술 글쓰기가 전공 지식을 표방할 수밖에 없는 것은, 앞서 서술했듯이 현행 학술 환경이 그것을 강제하고 있기 때문이다. 이와 같은 상황에서 전공 간 활발한 지식 교류와 생산적 논의를 촉진하기 위해서는 그것을 가능하게끔 하는 연결고리가 필요하다. 그리고 ‘데이터’는 ‘전공’과 ‘글쓰기’ 그리고 ‘지식 전달’로 표방되는 기존 인문 학술 환경의 좁은 틈을 비집고 들어가, 글쓰기를 통해 축적된 지식(knowledge)을 모듈화해서 정보(information) 단위로 공유하고 그로부터 각각 분리된 전공에 관한 이해를 연결하게끔 하는 역할을 수행할 수 있다.

최근 몇 년 사이 정부의 뚜렷한 정책 기조로 강조되어 온 ‘디지털 뉴딜’, ‘데이터 댐’과 같은 미래 과제는 디지털 환경에서 미분화된 데이터를 다루는 경험이 산업으로부터 생활의 영역에까지 확대되고 있으며, 이러한 가운데 학술 영역 또한 그 영향에서 벗어날 수 없음을 의미한다. ‘데이터’ 기반의 커뮤니케이션이 범사회적 차원에서 보편화될 경우, 글쓰기를 통해 특정한 지식을 논문으로 발표·유통하는 학술 활동은 점차 그 자리를 잃어갈 소지가 다분하다. 현시점에서 학술논문 또는 학술 지식 편찬 과정에 ‘데이터’ 기술의 활용 가능성을 검토하는 것은 필연적 문제의식이라 해야 할 것이다. 학술논문을 표준적 형식으로 아카이빙하고 그 과정에서 개별 논문에 담긴 각각의 지식을 유기적인 형식의 데이터로 정리해서 종합하는 방안을 모색할 수 있다면, 현재 인문 학술 환경이 봉착한 어려움을 개선할 단서로 삼기에 충분하리라 생각된다.

두 번째 문제의식은 디지털 환경에서 학술 데이터를 능동적으로 ‘활용’하고 더 나아가 학술 지식을 편찬하는 과정에 적극적으로 ‘참여’할 수 있는 방안이 무엇인지 그에 대한 사례를 검토해보는 것이다. 아날로그 환경에서 간행된 학술논문이 폭넓게 읽히지 않는 이유 가운데는, 논문이 다루는

지식이 만들어지고 그것을 검증하는 과정에 타 전공자 또는 대중이나 시민이 '참여할 수 있는' 형식이 없다는 사실이 존재한다. 연구자는 지식을 일방적으로 생산하고, 잠재적 독자로서의 타 전공자 또는 대중은 오로지 생산된 지식을 소비할 뿐이다. 독자로서 텍스트를 소비하는 형식이 아니라면 대중이 학술 지식을 접할 기회가 거의 없다고 해도 과언이 아니다.

아날로그 환경의 학술 유통 체계와 달리 디지털 환경에서 이루어지는 지식 커뮤니케이션의 양태는, 이미 '참여'의 형식이 보편적으로 수용되어 가는 분위기이다. Web 1.0에서 2.0으로 다시 3.0으로 진화해 나가고 있는 웹 환경의 경우 A가 B에 대한 지식을 C에게 일방적으로 전달하는 커뮤니케이션 방식이 상대적으로 줄어들고 있으며, 오히려 A가 B에 대한 지식을 다루는 과정에 C를 참여하게끔 함으로써 그에 대한 이해를 자연스럽게 유도하는 확장적 커뮤니케이션 방식이 확대되고 있다. 이와 같은 디지털 환경에서의 지식 유통 방식의 변화를 참고할 때, '전문가'와 '전문성'에 기댄 일방적 지식 전달 형식은 향후 유지되기 어려울 가능성이 크다. 최소한의 수준에서나마 유지된다고 하더라도 지식 편찬 및 유통 과정에서 폭넓은 참여가 전제되지 않는다면, 학술의 사회적 확산을 도모하지 못하는 소위 '그들만의 리그'로 전락할 가능성이 크다.

지식을 일방적으로 전달하는 '글쓰기' 형식에서 지식 구성 요소로서의 '데이터' 공유와 그것을 매개로 참여를 유도하는 방향으로 학술 활동이 전환되어야 하는 이유가 여기에 있다. 학술 지식이 생산되는 환경과 그러한 지식이 다시 연구자와 대중 사이에서 소비되는 환경이 선순환 체계를 갖추고 있지 않은 현재의 지식 유통 체계는 진단과 혁신이 필요한 상황이다. 본론의 두 번째 장에서, 디지털 환경을 매개로 학술 지식 편찬 과정의 참여 형식을 고민한 구체적 사례를 검토해보고자 하는 것은, 단일 연구자에 의해 지식이 일방적으로 전달되는 아날로그 기반의 학술 커뮤니케이션

형식에서 벗어나, 디지털 환경이 제공하는 상호적(interactive) 특성을 학술 활동에 적극적으로 응용할 수 있는 방법이 무엇인지 구체적으로 살펴보고자 하는 노력에 해당한다.

‘참여’로 표방되는 집단지성(Collective Intelligence)으로서의 인문 학술 지식 생산 및 유통이 디지털 환경에서 어떻게 가능한지 그 구체적 방안을 실제 구현하고 그와 관련된 구체적 사례를 검토하고자 하는 것은 곧 학술 공공성에 대한 인식을 확대하고자 하는 문제의식에 기인한다. 이러한 작업이 향후 인문 학술이 나아갈 장기간의 방향을 모색하는 데도 구체적인 도움이 될 것으로 생각하며, 실험적 모색이자 미래를 위한 검토로서 본 연구가 조금이라도 그 역할이 있기를 바란다.

II. 인문학술 데이터 편찬 과정의 실험적 모색

학술논문을 대상으로 한 기초 학술 데이터를 온전히 편찬하기 위해서는 학술 지식을 직접 다루는 연구자들이 그 과정에 능동적으로 참여할 수 있어야 한다. 글쓰기 과정에서 이루어지는 연구자들의 지식 활동이 데이터 모델링 과정에 섬세하게 반영되지 않은 상태에서 소위 소프트웨어 기술을 다루는 용역업자들에 의해 구축된 표면적 성격의 학술 데이터는, 그 형식이 단순하며 활용 가능성 또한 제한적일 수밖에 없다.

다양한 연결고리를 매개로 측면을 향해 끊임없이 확장하고 그러한 확장 가운데 수많은 담론을 만나면서 다채롭게 변용되는 것이 지식의 본질임을 고려할 때, 학술 데이터 편찬은 연결(linking)과 확장(expansion)으로 대표되는 지식의 속성을 극대화하는 방향으로 진행될 필요가 있다. 그것은 곧

학술 지식을 다루는 연구자들의 관심과 흥미 그리고 참여를 유도하는 길이며, 내가 다루는 나의 인문 지식이 생생하게 살아 있다는 것을 증명할 기회가 연구자에게 주어짐을 의미한다.

연구자에 의해 직접 구축된 기초 데이터는 그것의 확장성을 반드시 내포할 수밖에 없다. 죽은 지식이 아니라 살아 있는 지식으로서의 생생함은 여타 다른 지식과의 연결고리가 풍부해질수록 더욱 크게 드러날 수밖에 없으며, 그것을 위해서는 반드시 표준적 형식의 지식 디자인에 대한 고민이 선행되어야 한다. 이를 고려할 때 웹에서 보편적으로 활용되는 포맷을 준용한 학술 데이터를 편찬하는 것은 선택이 아니라 필수에 가깝다. 현시점에서 웹에서 보편적으로 활용되는 표준적 성격의 데이터 포맷은 XML, RDF, JSON, CSV 등이 있다.

XML과 RDF는 시맨틱 웹 아키텍처를 구성하는 표준적인 웹 언어이자 데이터 포맷이다. XML과 RDF로 기술된 데이터를 편찬해 웹에서 공유한다는 것은 곧 시맨틱웹 철학에 입각한 데이터 활용을 전제로 함을 의미한다. JSON은 데이터 객체를 속성·값의 쌍 형태로 표현하는 형식이다. 여러 프로그래밍 언어와 연계해서 사용할 수 있는 데이터 포맷이며, 텍스트로 기술하여 사람도 쉽게 읽고 작성할 수 있다는 것이 특징이다. CSV는 쉼표(comma)를 기준으로 항목을 구분해 저장한 데이터를 말한다. 엑셀과 같은 데이터 처리 소프트웨어를 통해 편집·저장하는 것이 가능한 보편적 데이터

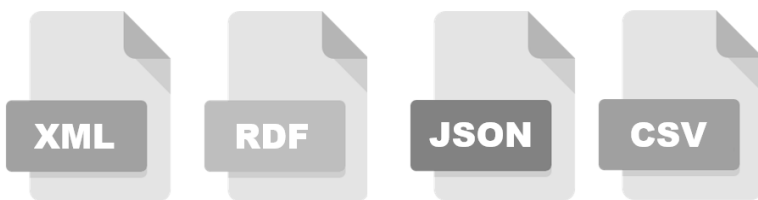


그림1-웹에서 표준적으로 활용되는 대표적 데이터 포맷: XML, RDF, JSON, CSV

포맷이다.

학술 지식을 XML, RDF, JSON, CSV와 같은 웹 표준 형식 또는 웹에서 보편적으로 활용되는 포맷의 데이터로 정리·가공할 수 있는 방안을 강구한다는 것은 곧, 디지털 환경에서 인문 지식 공유 환경을 더욱 확대해 나가고자 하는 적극적 고민에 해당한다. 한편으로 인문학 연구자가 웹에서 보편적으로 활용되는 포맷의 데이터를 자유롭게 다룰 수 있는 역량을 갖추고 있기란 어려운 일이다. 디지털 리터러시와 데이터 리터러시 방면의 교육을 통해 그러한 내용을 전달하는 것이 가능할 수도 있겠으나, 모든 인문학 연구자들을 대상으로 그러한 교육을 진행한다는 것은 현실적으로 한계가 뚜렷하다.

이러한 점을 고려할 때 ①연구자들이 상대적으로 쉽게 쓸 수 있는 텍스트 데이터 편찬 도구이자 ②해당 텍스트 데이터를 표준적인 포맷의 데이터로 변환하기에 용이한 플랫폼을 활용하는 방안을 생각해볼 수 있다. 그리고 그러한 맥락에 부합하는 도구로서 위키(Wiki)는, 데이터 설계 및 편찬 과정에 참여하고 결과물을 공유하는 효과적인 디지털 플랫폼이자 활용하기 용이한 도구에 해당한다.³

위키(Wiki)는 위키 소프트웨어 기반의 웹사이트다. 위키 엔진, 위키 애플리케이션으로도 불리는 위키 소프트웨어는 인터넷 접속이 가능한 웹 브라우저에서 마크업 언어인 위키 문법을 사용해 다수의 사용자가 공동으로 문서 작성 및 수정 작업을 수행할 수



그림2-MediaWiki의 로고

3 김지선·장문석·류인태, 「공유와 협업의 글쓰기 플랫폼, 위키」, 『한국학연구』 60집 (2021), 394쪽.

있게 하는 협업적 성격의 소프트웨어이다.

위키 문법(Wiki Syntax)이라고 불리는 마크업 언어는 텍스트에 태그(Tag) 등을 이용해 문서의 형식과 서식을 지정하는 언어이다. 위키 문법은 일반적인 웹 페이지 구현 언어인 HTML보다 간편한 마크업 형식을 바탕으로 하기에, 웹 환경에 대한 깊은 이해 없이도 위키 문서를 어렵지 않게 작성할 수 있다. 인문학 연구자 입장에서 학술 데이터 편찬 작업에 참여하기에 상대적으로 부담이 덜한 협업 플랫폼으로 위키가 다루어질 수 있는 이유가 여기에 있다. 또한 미디어위키(MediaWiki)의 경우 작성된 페이지를 XML 데이터로 내보내기 기능을 갖추고 있기에, 편찬된 위키 페이지를 웹 표준 형식의 데이터로 변환하기에도 용이하다.

본 연구에서 실험적으로 수행한 학술 데이터 편찬 작업은, 미디어위키를 기초 플랫폼으로 삼아 학술 논문을 대상으로 텍스트 데이터를 구축한 다음 해당 텍스트 데이터를 XML, CSV, JSON 등 표준적 성격의 데이터로 가공하는 프로세스로 진행되었다. 해당 프로세스를 거치면서 정교한 결과물을 만들어내기 위해서는, 오랜 기간 연구자들 사이의 논의를 거쳐 정리된 인문 학술 온톨로지(Ontology)를 데이터 모델링 단계에서 적용하는 것이 필수적이다. 학술 지식을 직접 다루는 연구자들이 학술 데이터 편찬 과정에 능동적으로 참여하는 핵심적인 채널이 곧 온톨로지 설계 과정이기 때문이다.

이와 관련해 본 연구는, 데이터 처리 프로세스의 가능성 유무와 기술적 공정 절차를 정리하는 데 초점을 두었기 때문에, 문제의식의 외연과 지면의 한계를 고려해 그러한 인문 학술 온톨로지 설계 과정에 대한 구체적 고민까지는 다루지 못했다. 이 부분에 대해서는 ‘연구자들의 협업이 이루어진다’는 사실을 전제로 삼아 진행하였으며, 추후 유관 연구가 독립적으로 이루어져야 할 것이라는 판단이다.

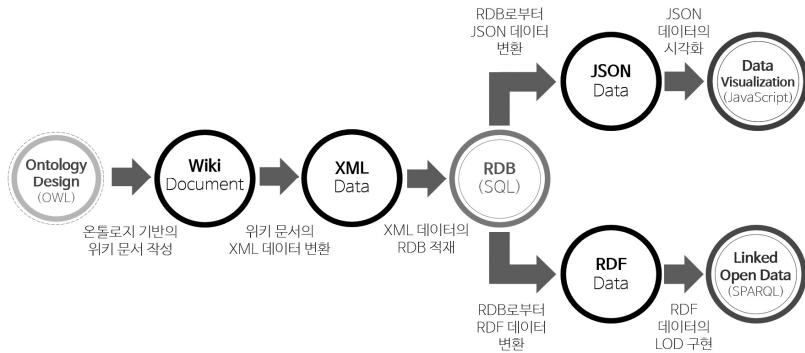


그림3- 위키 텍스트에서 웹에서 통용되는 표준적 포맷의 데이터로 변환 절차

한편으로 인문 학술 온톨로지 설계가 구체적으로 전제되지 않은 상황이라 데이터 스키마의 정교함이 다소 부족하다고 하더라도, 이러한 프로세스가 가능하다는 것이 증명된다면 인문 학술 지식의 공유 가능성을 디지털 환경에서 확대하고, 그것을 매개로 연구자 집단 더 나아가 시민들이 해당 지식을 다룰 수 있는 채널을 더욱 다양화하는 것이 가능해질 것이다. 그림3은 그러한 의도를 바탕으로 본 연구에서 시도한 학술 데이터 편찬 프로세스를 압축한 것으로서, 위키 기반의 학술 데이터 편찬 결과물이 웹에서 보편적으로 통용되는 표준적 포맷의 데이터 변환을 거쳐 데이터 시각화 결과물 또는 링크드 오픈 데이터로 변환되는 절차를 도식화한 것이다. 전체 데이터 처리 프로세스에 적용된 개별 절차로서 각 과업의 내용과 의미를 아래에 정리해보았다.

1. 대상 자료 수집

본 연구가 실험적 성격의 학술 데이터 프로세싱을 진행하기 위해 대상으로 삼은 것은 국어국문학 계열의 여러 학회가 간행한 학술지 수록 논문

40여 편이다. 상허학회의 『상허학보』, 구보학회의 『구보학보』, 민족문학사연구소의 『민족문학사연구』, 반교어문학회의 『반교어문연구』, 서강대 인문과학연구소의 『서강인문논총』, 국제어문학회의 『국제어문』, 한국여성문학회의 『여성문학연구』, 인하대 한국학연구소의 『한국학연구』, 성균관대 대동문화연구원의 『대동문화연구』 등 10여 개 학회 및 학술지가 그 구체적 대상에 해당한다.⁴

다양한 학회지에 수록된 논문을 대상으로 삼은 것은, 여러 학회와 연구소가 저마다 다른 학술 방향성을 추구하는 한편 ‘국어국문학’이라는 동종의 전공을 기반으로 하고 있기 때문에 지식이 연결되고 한편으로 단절되는 양상이 데이터로 충실하게 드러날 것으로 판단했기 때문이다. 대규모 학술 지식을 본격적으로 다루는 것이 아니라, 데이터 편찬 프로세스를 진행하는 가운데 논리적 오류와 기술적 난항을 확인하는 것이 본 연구의 주된 목적임을 고려할 때, 샘플 데이터로서 적절하다는 판단이 전제되었다.

2. Wiki 텍스트 데이터 편찬

40여 편의 학술논문을 대상으로 선정한 이후 미디어위키를 기반으로 한 텍스트 데이터 편찬 작업이 진행되었다. 해당 작업은 성균관대학교 국어국문학과를 중심으로 여타 국어국문학 유관 학회가 참여하고 있는 한국어문학 지식공유 위키⁵를 활용해 진행되었으며, 학술논문을 대상으로

4 대상으로 삼은 학술논문의 구체적 목록은 ‘인문학과 지식공유: 시민 참여 모델의 개발’ 연구 페이지(<http://www.klbksk.com/wiki/index.php/HumanitiesDataArchive>)의 ‘연구데이터’ 항목을 통해 확인할 수 있다. 대상 논문 선정은 개별 학회 및 연구소의 동의를 전제로, 경희대학교 국어국문학과 장문석 교수의 자문을 더해 이루어졌다.

5 한국어문학 지식공유 위키(<http://www.klbksk.com/wiki>).

한 위키 페이지 구축은 논문의 기본적 형식을 준용하되 유관 정보를 함께 정리해나가는 방식으로 이루어졌다. 소위 메타데이터로 통칭하는 논문의 상위정보와 목차 및 본문 그리고 참고문헌 및 주석을 모두 위키 문법으로 기술하였으며, 해당 기술(description) 형식은 추후 XML로 데이터를 변환하는 과정에서 XML Schema가 연계될 수 있게끔 안내하였다.

아래 그림4는 학술논문 ‘천정환, 『문화론적 연구』의 현실 인식과 전망’, 『상허학보』 19집(2007), 상허학회.’를 대상으로 한 위키 페이지⁶ 구축 예시로서, ‘상위정보’ 부분을 확대한 것이다. ‘상위정보’는 논문의 제목과 영문제목, 게재연월, 학술지, 게재권호, 쪽수, 발행기관, 저자, 참여저자수, 참여형식, ISSN, DOI, RISS url, 주제어, 초록 등 학술논문의 메타데이터에 해당하는 여러 정보를 디자인하고 개별 데이터값을 입력한 것이다. 초록을 ‘상위정보’에 포함한 것은, 대부분의 학술논문이 초록(abstract)을 기본 정보로 가지고

상위정보 [편집]

'문화론적 연구'의 현실 인식과 전망 :
Recognition and Perspective of 'the Cultural Studies'

게재연월	2007.02	학술지	상허학보 ^{학제}	게재권호	19	쪽수	011-048	발행기관	상허학회 ^{학제}	
저자	천정환 ^{학제}	참여저자수	1인	참여형식	주제어	ISSN	DOI	10.22936/sh.19.200702.001	식별자	SP0001
RISSurl	http://www.riss.kr/link?id=A104965467				Wikiurl	http://www.klbksk.com/wiki/index.php/HDAA0001				
주제어	문화론: 근대문학의 종언; 인문학의 위기; 정상성: 국문학; 대중문화									

(초록) 이 글은 '문화론적 연구'가 처한 현재의 자리를 점검하고 '국' '문학'이 낳은 새로운 연구와 글쓰기를 더욱 급진화하는 제언을 목표로 한다. 본질적으로 '문화론적 연구'는 문학이 지식체계와 문화권 안에서 지니는 지위 및 역사에 대해 민감한 태도를 취하며 '국' '문학'의 정상성 규범에 대해 성찰한다. 한국 근대문학은 크게 보아 세 속의 상호작용에 의해 이루어져 왔다: 정치적 계몽주의 문학, 내면성의 문학과 모더니즘소설, 그리고 대중적 낭만주의문학. 이 각각의 위상과 역할이 달라지고 중단된 상황이 소위 '근대문학의 종언'이다. 지금의 시점에서 필요한 것은 '근대문학'이 생산한 '문학적인 것'을 재고하고 미학주의에 대해 반성하여 새로운 사회적 상황에서 문학이 인간과 삶을 위해 기능하도록 재-규범화하는 것이다. '문화론적 연구'는 문학의 사회문화적 맥락을 되살리는 데 효과적인 방법의 하나다. 오늘날 문학연구와 문학은 '양극화'라는 한국의 사회-문화적 상황에 대해 윤리적으로 이해하고 교육과 문화의 불평등에 대해 성찰하게 해야한다. 문화의 큰 위력은 양극화에 대한 바라케이드로서, 여기에 개입하고 연대할 발도를 찾고 미와 윤리의 긴장과 융합을 이끌어내는 것이 문화론적 연구의 임무이다. 문화론적 연구의 제1국면은 근대성 연구에 정한된 '국' '문학'으로부터 출발하여 문화사로서의 문화론으로의 진입까지이며, 제2국면은 제1국면 연구의 안착-확장과 문학사의 재구, 그리고 문화적-담론적 실천으로서 구성되는 단계이다. 즉 '문화론적 연구'는 새로운 '국문학사'나 '문학사' 연구에 한정되지 않고 '지금-여기'의 문화-정치의 현실과 접속할 것이다.

그림4- 학술논문 위키(Wiki) 페이지의 ‘상위정보’ 예시

6 학술논문 위키(<http://www.klbksk.com/wiki/index.php/HDAA0001>).

있으며 해당 내용이 논문의 요지를 파악하는 데 단서가 되기 때문에, 추후 대규모 데이터가 확보되었을 때 텍스트 마이닝을 적용할 가능성을 염두에 둔 결과라 하겠다.

‘목차’는 논문 본문의 개별 장절과 하이퍼링크로 연결하였으며, 추후 XML 데이터에서 요소(Element) 간 연계가 가능하게끔 식별자(ID) 값을 공유하였다. ‘본문’은 논문이 구성된 장절과 단락을 그대로 살려 텍스트 데이터로 구축하였으며, 위키의 틀(Template) 기능을 활용해 논문의 쪽수와 주석 정보를 담을 수 있는 인덱스를 디자인한 다음 적용해야 하는 위치마다 모두 부기하였다. 본문 내용 가운데 추후 개별 논문들을 연결할 수 있는 문맥 요소(Contextual Element)로서 ‘인물(Person)’과 ‘용어(Term)’ 두 가지를 임의로 설정하였다. 두 정보 항목은 인문 학술 온톨로지 설계가 이루어졌다

목차 [편집] 1. 서론 2. 문학사의 경회와 '문화론적 연구' 3. 근대 문학주의와 문화적 현실에 대한 성찰 4. 문화사로서의 '문화론적 연구'와 문화연구로서의 '문화론적 연구' 5. 결론에 대신하여
본문 [편집] 본문1: 서론 [편집] 〈목>▶^{P11}이 글은 〈상학화〉 2006년 가을 심포지움 "한국 근대문학 연구의 역사적 전환과 창조적 모색"에 제출되었던 것으로서, 심포지움에서 제기된 논의와 문제제기를 바탕으로 개고한 것이다. 원래 필자의 발표는 심포지움의 '총론'이 아니었음에도, '문화론적 연구'라는 명칭이 가진 포괄성¹⁾과 제1발표로 제출된 사정 때문에 총론인 것처럼 읽히고 관심과 ▶^{P12}비판의 대상이 되었다. 〈상학화〉 심포지움에서 제기된 비판 가운데에는 오독과 편견에 기초한 부당한 것도 있었지만, 대부분의 문제제기와 심포지움 자체는 유익했다. 특히 이번 〈상학화〉의 심포지움의 '성과'의 하나는 '문화론적 연구' 내부의 성세하지만 중요한 차이가 무엇인지 공식적으로(?) 드러낸 데 있다고도 볼 수 있다. 입장 차이를 드러냄으로써 오히려 '문화론적 연구'의 지향점은 분명해졌다고 할 수 있다. 여러 조건 때문에 심포지움 자리에서 제기된 문제에 대한 필자의 입장과 세론을 충분히 말하지 못했기 때문에, 원래의 원고를 개고하여 '문화론적 연구'에 대한 물음에 답하고자 한다. 1) 정상성과 '국문학' 학문의 '정상성'에¹⁾대해 지적²⁾은 흥미로운 예를 들어 말한 적 있다. 그에 의하면 오늘날 미국 학계에서 '당신이 뒤를 분석한다면 당신은 철학을 하고 있는 것'이고 '해결을 분석한다면 비교문화에 속한 것'이라 한다. 오늘날 미국의 철학과는 인지주의와 뇌과학이 주도하고 있고, 전통적인 의미의 철학은 비교문학·문화연구학과와 영문학·불문학 등에서 다루고 있다는 것이다. 지적은 논의를 확대하여 비단 미국 철학계 뿐 아니라 철학사 전체를 볼 때, 늘 다른 분과학문들이 철학의 '정상적(규범적)' 역할을 떠맡거나, 또는 철학 자신이 여타의 학문적 때로는 심지어 비학문적 실천과 분야의 과제를 맡아왔다는 점을 말한다. 철학 '고유'의 공간이란 원래 없다. 지적은 이 논의 끝에 마지막▶^{P13}으로 다음과 같이 뼈아프게 한 마디를 덧붙인다. '가장 캐케묵고 강단적이고 적실성이 없고 '죽어 있는' 철학'²⁾만이 온전히 철학의 '정상적 역할'을 했다.

그림5- 학술논문 위키(Wiki) 페이지의 ‘목차’와 ‘본문’ 예시

주석 [편집]

1. 이 글에서는 '문화연구' (문화론)는 영미의 cultural studies를, '문화론적 연구'는 상허학회 심포지움 논제의 하나인 '문화론적 (한국문화) 연구'를 가리킬 때 쓴다. 양자 와 내용의 관계에 대해서는 후술. '문화론적 연구'는 흔히 '문화 (론적) 연구'와 병기되거나 동일시되기도 한다. 그러나 '문화'이라는 말이 지닌 소재론적 부담스 때문에 필자는 가끔씩 이 용어를 피하고 보다 포괄적인 함의를 지닌 '문화'를 택한다. '문화론적 연구'는 '문화연구'와 친연성을 갖기 때문에 '절대'나 '탈식민주의적 연구'와도 관련이 있다. 그런데 이경홍·권보드라와 같은 연구자는 '문화'를 '문화'와 병렬될 수 있는 의미로 '넓게' 해석하여 자신들의 연구방법을 옹호한 바도 있다. (이경홍, "오딧세우스의 변명", 한국현대소설학회, 『현대소설연구』 27, 2005; 권보드라, "동숙사 와 문화의 질서 재', 한국현대소설학회, 『현대소설연구』 27, 2005 참조).
2. ↑ 술라보에 지역, 필자론 외 역, "신체 없는 기관-폴리츠와 결과물들", 도서출판 b, 2006, 8-10쪽.
3. ↑ 비군한 예를 들어 1990년대 이후 한국 근현대사 연구의 성과를 나름의 관점에서 종합한 책들인 『해방전후사의 재인식』^[16],이나 『근대를 다시 읽는다』에 포함된 '국' '문학' 연구자들의 논문 연수를 생각해보라. 어떤 문학연구자들은 자신과 다른 지향성을 가진 문학연구자들이 '타 분야'에 출입하는 것에 대해 비판적으로 말하며 심지어 그들이 이도저도 아닌 '2류' 연구자가 되는 게 아니냐는 걱정(?)을 하곤 한다. 그러나 이는 국사학·언문학·출판학·영화학 등 인문학계와 인문학 전반의 현황을 잘 모르는 기우이거나, 인문학의 '본연' 과 무관한 영역 지킴이 의식 또는 질의의 소신일 뿐이다.
13. ↑ 문학이 정치적 반혁명과 인간행방과 같은 큰 목적을 위해 기여해야 하며 기여할 수 있다는 식의 근대적 문학주의가 '큰 문학주의'라면, 문학이 온전히 자율적인 체계이며 무목적적 가치를 지닌다는 식의 미학주의가 '작은 문학주의'이다. '작은 문학주의'는 텍스트에 나타나는 미시경지의 징후들을 긴장한 '정치적인 것'으로 과장한다. 이는 때로 미시경지와 탈정치화를 중립화하고, 탈정치화를 변명하는 구실로 쓰인다.
14. ↑ 이는 가라타니 고진의 소론에 대한 황종연의 비판의 핵심적 지경이기도 하다. 황종연, 『문학의 육시록 이후편』, 『현대문학』, 2006년 8월 참조.
15. ↑ 『FOCUS '문학의 시대' 이후의 문학비평: 이광호·류보선·김형중·김영찬』, 『문학동네』, 2006 가을호 등을 참조.
16. ↑ 황종연, 앞의 글, 211쪽. "근대문학이 끝났다고 해서 문학적 글쓰기에 의미와 가치를 부여할 모든 이유마저 사라지는 것은 아니다. 영미권의 포스트모던문학을 굳이 따올라 지더라도 근대문학 이후의 문학을 인정하는 것은 이롭소로 가능하다."
17. ↑ 황종연, 같은 곳. "대학원과 교육의 전반적인 부실화, 대용업체의 문화적 저열성과 기술적 낙후성, 권위주의적 경로의 삼엄한 언론 통제, 글에 능통한 사람들과 열의의 폭탄을 부여한 유교문화의 낡은 유산, 학력귀족이 입신출세를 독점하는 사회계제-이런 요인들을 떠나서 과거 한국문학의 위세를 설명하기란 불가능하다. 그렇게 저발전 사회의 동토 속에 자리한 문학이 과연 그렇게 훌륭한 것이었는지도 모른다."
18. ↑ 권리의 소설 수용의 변화에 대한 일련의 글들과 글고(『파라 21』, 2004년 봄호; 『세계의 문학』, 2007년 봄호)를 참조 바람.

그림6- 학술논문 위키(Wiki) 페이지의 '주석' 예시

는 시나리오를 전제로 한, 일종의 샘플 클래스(Class)에 해당한다.

‘주석’은 미디어위키에서 제공하는 <ref></ref> 태그와 <references/> 태그를 활용해 정리하였다. 그 과정에서 본문에서 주석이 적용되는 위치와 그에 해당하는 주석의 내용을 구분해서 두 가지 정보가 동기화될 수 있게끔 식별 목적의 번호를 부여하였다. 해당 번호는 XML 데이터 변환 과정에서 개별 주석(위치-내용)의 식별자(ID) 정보로 처리되어, 추후 XML 데이터가 RDB에 적재되었을 때 주석 테이블을 구성하기 위한 매개로 활용될 수 있도록 하였다.

그리고 주석 내용 가운데 출처로서 참고문헌 정보가 함께 표기된 것들을 대상으로, 추후 RDB를 구축하는 과정에서 논문-참고문헌 간 인용 데이터를 구축할 수 있게끔 위키의 틀(Template) 기능을 활용해 문헌 정보를 담을 수 있는 인덱스를 디자인해서 적용하였다. ‘참고문헌’에 관한 서식은 개별 학술지마다 형식이 모두 다르기에 임의로 표준적 형식을 마련하고, 그에 맞추어 정리하는 방식을 적용하였다. 학위논문과 학술논문의 경우 ‘연구논문’ 항목으로, 단행본이나 전집, 총서, 문고의 일환으로 출간된 서적의

참고문헌 [편집]

연구논문 [편집]

- 'FOCUS' 문학의 시대 이후의 문학비평: 이광호·류보성·김형중·김영찬, 『문학동네』, 2006 가을호.
- 고훈준민, 『근대문학의 종언, 그리고 '소설' 이라고 불리는 대략난감한 글쓰기들』, 『작가와 비평』, 2006년 상반기.
- 권보드래, 『"풍속사"와 문학의 결서』, 『현대소설연구』, 27집, 한국현대소설학회, 2005.
- 권성우, 『문학을 넘어서는 문학의 길』, 『문학수필』, 2005년 봄.
- 박천호, 『문학 '소' 없는 시대의 문학연구』, 『역사비평』, 75호, 역사문제연구소, 2006년 여름.
- 유종호, 『문학의 전락—무라카미 [村上] 현상을 놓고』, 『현대문학』, 2006. 6.
- 이경준, 『오딧세우스의 변명』, 『현대소설연구』, 27, 한국현대소설학회, 2005.
- 천정환, 『새로운 글쓰기와 문학연구에 대한 시론』, 『민족문화사연구』, 민족문화학회, 2004. 11.
- 하정일, 『개인의 이데올로기를 넘어서』, 『비평과 전망』, 8호, 2004. 6.
- 황종연, 『무엇이 한국문학의 보람인가—문학평론가 백낙청과의 대화』, 『창작과 비평』, 131, 2006년 봄.
- 황종연, 『문학의 특사록 이후』, 『현대문학』, 2006년 8월.

연구서적 [편집]

- 가라타니 고진, 『조영일』, 『근대문학의 종언』, 도서출판, 2006.
- 김동식, 『풍속·문학·문화사』, 『가을과 미로』, 청년의사, 동국대 한국문화연구소, 2006.
- 박지향, 『외 판, 『해방전후사의 재인식』, 책세상, 2005.
- 슬라보예 지젝, 김지훈 외 역, 『신체 없는 기관—돌피조와 결과들』, 도서출판, 2006.
- 와타나베 히로시, 윤대석 역, 『창공의 탄생』, 강, 2006.
- 조성연, 『대중문화와 정전에 대한 반역』, 소명출판, 2002.
- 토트 가를린, 남재일 역, 『무한 미디어』, 휴먼 앤 북스, 2006.
- 토머스 L. 프리드먼, 김상철 역, 『세계는 평평하다』, 창, 2005.

그림7- 학술논문 위키(Wiki) 페이지의 '참고문헌' 예시

경우 '연구서적' 항목으로, 그 외에 신문이나 잡지와 같은 일·주간 출간물과 웹 리소스의 경우 '기타' 항목으로 처리하였다.

3. Wiki 텍스트 데이터의 XML 데이터 변환

앞서와 같이 표준적 형식의 위키 페이지로 편찬된 학술 데이터는 미디어 위키의 '내보내기' 기능을 통해 XML 포맷의 데이터로 변환할 수 있다. 아래 그림8은 그와 같은 방식으로 학술논문 '천정환, 『문화론적 연구』의 현실 인식과 전망, 『상어학보』 19집(2007), 상허학회.' 위키 페이지 정보를 XML 문서로 내보낸 예시이다.

```
{{논문정보-  
|식별자=SP0001-  
|제목='문화론적 연구'의 현실 인식과 전망-  
|부제=-  
|영문제목=Recognition and Perspective of 'the Cultural Studies'-  
|개제연월=2007.02-  
|학술지=상허학보-  
|개제권호=19-  
|쪽수=011-048-  
|발행기관=상허학회-  
|저자=정정환-  
|참여자자수=1-  
|참여형식=주저자-  
|ISSN=-  
|DOI=10.22936/sh.19..200702.001-  
|RISUrl=http://www.riss.kr/link?id=A104965467-  
|WIKIUrl=http://www.kibsis.com/wiki/index.php/HDAA0001-  
|주제어=[[문화론]] [[근대문학의 종언]] [[인문학의 위기]] [[정상성]] [[국문학]] [[대중문화]]-  
|초록=이 글은 '문화론적 연구'가 처한 현재의 자리를 점검하고 '국' '문학'이 낳은 새로운 연구와 글쓰기를 더욱 급진화하자는 제안을 목표로 한다. 본질적으로 '문화론적 연구'는 문학이 지식체계와 문화전반에서 지니는 지위 및 역사에 대해 민감한 태도를 취하며 '국' '문학'의 정상성 규범에 대해 성찰한다. 한국 근대문학은 크게 보아 세 축의 상호작용에 의해 이루어져 왔다: 정치적 계몽주의 문학, 내면성의 문학과 모더니즘소설, 그리고 대중적 낭만주의문학. 이 각각의 위상과 역할이 달라지고 중단된 상황이 소위 '근대문학의 종언'이다. 지금의 시점에서 필요한 것은 '근대문학'이 생산한 '문학적인 것'을 재고하고 미학주의에 대해 반성하여 새로운 사회적 상황에서 문학이 인간과 삶을 위해 기능하도록 재 - 규범화하는 것이다. '문화론적 연구'는 문학의 사회문화적 맥락을 되살리는 데 효과적인 방법의 하나다. 오늘날 문학연구와 문학은 '양국화'라는 한국의 사회 - 문화적 상황에 대해 윤리적으로 이해하고 교육과 문화의 불평등에 대해 성찰하게 해야한다. 문화의 큰 위력은 양국화에 대한 바리케이드로서, 여기에 개입하고 연대할 방도를 찾고 미와 윤리의 긴장과 융합을 이끌어내는 것이 문화론적 연구의 임무이다. 문화론적 연구의 제1국면은 근대성 연구에 정향된 '국' '문학'으로부터 출발하여 문학사로서의 문화론으로의 진입까지이며, 제2국면은 제1국면 연구의 안착·확장과 문학사의 재구, 그리고 문화적·담론적 실천으로서 구성되는 단계이다. 즉 '문화론적 연구'는 새로운 '국문학사'나 '문화사' '연구'에 한정되지 않고 '지금 - 여기'의 문화 - 정치의 현실과 접촉할 것이다.-  
}}-  
-  
===''목차''===-
```

그림8- 미디어위키의 ‘내보내기’ 기능을 통해 확보한 위키(Wiki) XML 문서 예시

해당 XML 포맷의 데이터는 정제되지 않은 전자문서로서, 사전에 설계해둔 표준 Schema 기반의 Valid XML 데이터로 변환하는 작업이 필요하다. 이를 자동으로 처리하기 위해 사전에 설계해둔 표준 Schema 기반의 Valid XML 데이터로 변환해주는 Python 코드를 구현하였다.⁷ 해당 Python 코드를 활용해서 위키에서 내보낸 정제되지 않은 XML 문서를 사전에 설계해둔 Schema를 반영한 Valid XML 데이터로 변환하는 작업을 진행하였다.

해당 과정은 위키 페이지로 저장된 텍스트 데이터를 사전에 설계한 표준 Schema 기반의 Valid XML 데이터로 변환하는 절차 정도로 이해할

7 XML 데이터 자동 변환을 위한 Python 코드 구현 작업과 Valid XML 데이터 대상 RDB 구축을 위한 SQL 구문 작성 및 정리 그리고 RDB 기반의 데이터베이스를 활용해 JSON 데이터를 자동으로 생성하기 위한 SQL 구문 작성 및 정리는 한국학중앙연구원 디지털인문학연구소 소속 김지선 연구원에 의해 수행되었다.

```
<?xml version="1.0" encoding="UTF-8"?>↓
<?xml-stYLESHEET type="text/xsl" href=""?>↓
↓
<학술정보 id="SP0001">↓
↓
  <상위정보 id="MD0001">↓
    <식별자>SP0001</식별자>↓
    <제목>문화론적 연구의 현실 인식과 전망</제목>↓
    <부제></부제>↓
    <영문제목>Recognition and Perspective of 'the Cutral Studies'</영문제목>↓
    <게재연월>2007.02</게재연월>↓
    <학술지><journal id="">상학학보</journal></학술지>↓
    <게재권호>19</게재권호>↓
    <쪽수>011-048</쪽수>↓
    <발행기관><group id="">상학학회</group></발행기관>↓
    <저자><person id="">한정환</person></저자>↓
    <참여저자수>1</참여저자수>↓
    <참여형식>주저자</참여형식>↓
    <ISSN></ISSN>↓
    <DOI>10.22936/sh.19.200702.001</DOI>↓
    <Rissurl>http://www.riss.kr/link?id=A104965467</Rissurl>↓
    <Wikiurl>http://www.klibsit.com/wiki/index.php/HDAA0001</Wikiurl>↓
    <주제어>문화론: 근대문화의 종언: 인문학의 위기: 정상성: 국문학: 대중문화</주제어>↓
    <초록>이 글은 '문화론적 연구'가 처한 현재의 자리를 점검하고 '국' '문학'이 낳은 새로운 연구와 글쓰기를 더욱 급진화시키는
    제안을 목표로 한다. 본질적으로 '문화론적 연구'는 문학이 지식체계와 문화전반에서 지니는 지위 및 역사에 대해 민감한 태도를 취하며 '국' '문
    학'의 정상성 규범에 대해 성찰한다. 한국 근대문학은 크게 보아 세 속의 상호작용에 의해 이루어져 왔다: 정치적 계몽주의 문학, 내면성의 문학
    과 모더니즘소설, 그리고 대중적 낭만주의문학. 이 각각의 위상과 역할이 달라지고 중단된 상황이 소위 '근대문학의 종언'이다. 지금의 시점에서
    필요한 것은 '근대문학'이 생산한 '문학적인 것'을 재고하고 미학주의에 대해 반성하여 새로운 사회적 상황에서 문학이 인간과 삶을 위해 기능하
    도록 재 - 규범화하는 것이다. '문화론적 연구'는 문학의 사회문화적 맥락을 되살리는 데 효과적인 방법의 하나다. 오늘날 문학연구와 문학은 '양
    극화'라는 한국의 사회 - 문화적 상황에 대해 윤리적으로 이해하고 교육과 문화의 불평등에 대해 성찰하게 해야한다. 문화의 큰 위력은 양극화에
```

그림9- Python 코드를 통해 자동으로 변환된 Valid XML 데이터

수 있다. 이러한 절차가 기술적으로 가능한 이유는, 본 연구에서 기초 데이터 구축 플랫폼으로 활용한 미디어위키가 ‘XML 데이터 내보내기’ 기능을 가지고 있고, 그것을 전제로 XML Schema를 사전에 설계하였으며, 애초 설계한 Schema를 기준으로 XML 데이터 변환을 염두에 두고 학술논문 위키 페이지를 작성했기 때문이다.

4. XML 데이터 RDB 적재 및 기초 DB 구축

미디어위키에서 내보내기 기능으로 확보한 XML 데이터를 대상으로 Python 코드를 활용해 변환한 Valid XML 데이터를 RDB에 적재하였다. 그리고 RDB에 적재한 Valid XML 데이터를 매개로 테이블 형식의 기초

데이터베이스를 구축하는 작업을 진행하였다. 아래에 나열된 질의어 (Query)는 해당 작업 과정에서 활용한 SQL 구문 가운데 일부 예시이다.

● 예시1: XML 문서의 RDB 적재를 위한 기본 Table 생성

```
CREATE TABLE EAS_xml (  
  id nvarchar(10) NOT NULL,  
  xmltxt xml  
  primary key(id)  
)
```

● 예시2: <학술정보> ID와 content의 자동 매핑을 위한 Trigger 생성

```
CREATE TRIGGER EAS_getid ON EAS_xml INSTEAD OF INSERT AS  
INSERT INTO EAS_xml(id, xmltxt)  
SELECT  
  t.xmltxt.value('/학술정보/@id[1]', 'nvarchar(10)') AS id,  
  t.xmltxt AS xmltxt  
FROM inserted t
```

● 예시3: 문맥요소 <term> 데이터 적재를 위한 View 생성

```
CREATE VIEW EAS_xml_term AS  
SELECT id, 본문id, 본문장절id,  
  term.value('/@id', 'nvarchar(40)') AS 용어id,  
  term.value('.', 'varchar(max)') AS 용어name  
FROM EAS_xml_contents CROSS APPLY 본문장절_내용.nodes('/본문장절/page/  
term') AS R(term)
```

5. RDB에서 SQL을 활용한 JSON 데이터 생성

RDB로 구축된 기초 데이터베이스를 활용해서 CSV 데이터를 생성할

수 있게 되었다. 추가적으로 SQL을 활용해 JSON 데이터를 자동적으로 생성하기 위한 작업을 진행하였다. 아래에 나열된 질의어(Query)는 해당 작업 과정에서 활용한 SQL 구문 가운데 일부 예시이다.

- 예시1: EAS_Node_term Table에 class, uri column 추가 및 term Class ID를 RDB에서 TR--- 형식으로 자동 부여하기 위해 ID의 문자, ID의 숫자 column 추가 및 자동증가함수(IDENTITY) 사용

```
ALTER TABLE EAS_Node_term  
ADD class nvarchar(50), uri nvarchar(126), id용숫자 int IDENTITY(001, 001) NOT  
NULL, id용문자 AS 'TR' + RIGHT(REPLICATE('0', 2) + CAST(id용숫자 AS VARCHAR), 3)
```

- 예시2: 'class' column에 'Term', 'relationEng' column에 'isMentionedIn', 'relationKor' column에 '언급되다' 데이터값 추가

```
UPDATE EAS_Node_text  
SET class = 'Text', relationEng = 'isPartOf', relationKor = '포함되다'
```

- 예시3: EAS_Node_term Table을 활용해 Node_term.json 생성

```
SELECT id용문자 AS [id], class AS [class], 용어name AS [name], uri AS [uri]  
FROM EAS_Node_term  
FOR JSON PATH, ROOT('Node_term'), INCLUDE_NULL_VALUES
```

- 예시4: EAS_Link_term2text Table 활용 Link_term2text.json 생성

```
SELECT sourceid AS [sourceid], targetid AS [targetid], relationEng AS [relationEng],  
relationKor AS [relationKor]  
FROM EAS_Link_term2text  
FOR JSON PATH, ROOT('Link_term2text'), INCLUDE_NULL_VALUES
```

```

... "Node_group": [
  { "id": "G001", "class_field": "Group", "name": "상허학회", "chiname": "" },
  { "id": "G002", "class_field": "Group", "name": "한국현대소설학회", "chiname": "" },
  { "id": "G003", "class_field": "Group", "name": "동국대 한국문화연구소", "chiname": "" },
  { "id": "G004", "class_field": "Group", "name": "역사문제연구소", "chiname": "" },
  { "id": "G005", "class_field": "Group", "name": "민족문화사학회", "chiname": "" },
  { "id": "G006", "class_field": "Group", "name": "작가와비평 편집동인", "chiname": "" },
  { "id": "G007", "class_field": "Group", "name": "서강대 인문학연구소", "chiname": "" }
],

... "Node_term": [
  { "id": "TM001", "class_field": "Term", "name": "정상성", "chiname": "" },
  { "id": "TM002", "class_field": "Term", "name": "군대성", "chiname": "" },
  { "id": "TM003", "class_field": "Term", "name": "대중문화", "chiname": "" },
  { "id": "TM004", "class_field": "Term", "name": "시민성", "chiname": "" }
],

... "Link_person2paper": [
  { "sourceid": "P001", "targetid": "SP001", "relationEng": "writes", "relationKor": "저술하다" },
  { "sourceid": "P001", "targetid": "SP003", "relationEng": "writes", "relationKor": "저술하다" },
  { "sourceid": "P001", "targetid": "SP004", "relationEng": "writes", "relationKor": "저술하다" },
  { "sourceid": "P021", "targetid": "SP012", "relationEng": "writes", "relationKor": "저술하다" },
  { "sourceid": "P022", "targetid": "SP014", "relationEng": "writes", "relationKor": "저술하다" },
  { "sourceid": "P002", "targetid": "SP101", "relationEng": "writes", "relationKor": "저술하다" },
  { "sourceid": "P003", "targetid": "SP102", "relationEng": "writes", "relationKor": "저술하다" },
  { "sourceid": "P004", "targetid": "SP103", "relationEng": "writes", "relationKor": "저술하다" },
  { "sourceid": "P006", "targetid": "SP104", "relationEng": "writes", "relationKor": "저술하다" },
  { "sourceid": "P007", "targetid": "SP105", "relationEng": "writes", "relationKor": "저술하다" },
  { "sourceid": "P008", "targetid": "SP106", "relationEng": "writes", "relationKor": "저술하다" },
  { "sourceid": "P001", "targetid": "SP107", "relationEng": "writes", "relationKor": "저술하다" },
  { "sourceid": "P009", "targetid": "SP108", "relationEng": "writes", "relationKor": "저술하다" },
  { "sourceid": "P010", "targetid": "SP109", "relationEng": "writes", "relationKor": "저술하다" },
  { "sourceid": "P010", "targetid": "SP110", "relationEng": "writes", "relationKor": "저술하다" },
  { "sourceid": "P020", "targetid": "SP111", "relationEng": "writes", "relationKor": "저술하다" }
],

```

그림10-SQL을 활용해 자동적으로 변환한 JSON 데이터 예시

6. JSON 데이터의 시각화

RDB 기반의 기초 데이터베이스를 대상으로 SQL을 활용해 JSON 포맷의 데이터를 자동적으로 생성하는 프로세스를 구현하였으며, 그로부터 JSON 데이터 샘플을 확보하였다. 그리고 해당 JSON 데이터를 대상으로 삼아 자바스크립트 라이브러리 Sigma.js⁸를 활용해 네트워크 그래프 형태의 시각

화 구현 작업을 시도하였다.⁹

학술 데이터를 네트워크 그래프 형태로 시각화하는 작업을 시도한 목적은, JSON 데이터로 변환한 학술 데이터를 대상으로 자바스크립트 라이브러리의 활용 가능성을 검토하고 동시에 해당 데이터를 활용해 학술 지형도에 해당하는 지식그래프를 그리는 것이 가능한지 가시적으로 확인하기 위함이다.

하단의 그림11은 구현한 시각화 결과물의 예시이다. 미디어위키 문서

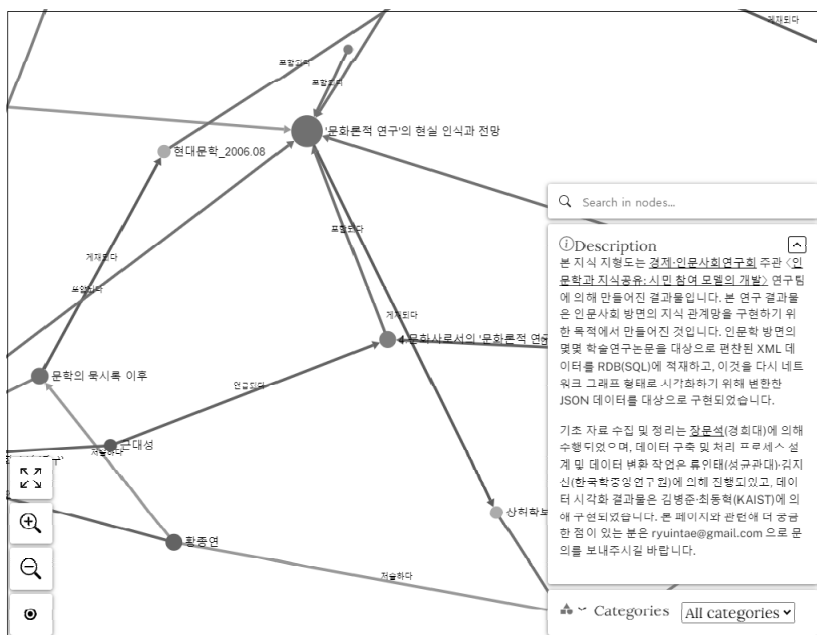


그림11-JSON 데이터(샘플)를 대상으로 한 시각화 결과물¹⁰ 예시

8 Sigma.js(<http://sigmajs.org>).

9 Sigma.js를 활용한 JSON 데이터 시각화 구현 코드 작성 작업은 한국과학기술원(KAIST) 디지털인문사회과학센터 소속 김병준, 최동혁 연구원에 의해 수행되었다.

10 <https://byungjunkim.github.io/PaperKnowledgeGraph/>

작성 단계에서부터 편찬된 여러 정보 항목이 Triple 형식(S-V-O)의 데이터로 분절되어 네트워크 그래프로 연결되어 있음이 확인된다. 현시점에서 구현된 데이터 시각화 결과물에서 섬세한 UI나 정교한 데이터 검색 기능을 확인하기는 어렵다. 그 이유는, 본 연구의 목적이 완성도 높은 결과물을 구현하는 것이 아니라 학술 데이터 처리 과정을 구현하는 데 초점을 두었기 때문이다. 추후 유관 학술논문이 이와 같은 방식으로 모두 데이터로 정리될 경우 빅데이터로서 인문 학술 지형도를 그리는 것도 불가능한 일이 아님을 짐작할 수 있다. 향후 더욱더 큰 규모의 학술 데이터 구축을 도모하기 위한 초기 단계의 결과물 정도로만 의미를 부여해도 충분할 것이다.

7. 학술 데이터 편찬 과정 구현과 그 시사점

학술논문을 대상으로 미디어위키 기반의 텍스트 데이터를 편찬하고, Python 코드를 작성해 해당 데이터를 사전에 설계된 XML Schema 기반의 Valid XML 데이터로 변환하였다. 그렇게 변환된 Valid XML 데이터를 RDB에 적재하고 적재된 XML 데이터를 매개로 SQL을 활용해 기초 DB를 구축하였다. 구축된 기초 DB를 바탕으로 SQL을 활용해 JSON 데이터를 생성하였으며, 해당 JSON 데이터를 대상으로 자바스크립트 라이브러리를 활용해 지식그래프 형식의 시각화 결과물을 구현하였다. 최종적으로 구현된 시각화 결과물은 곧 최초 대상이 된 복수의 학술논문에 담긴 여러 지식정보 사이의 의미적 연결망이라 할 수 있다.

본 연구에서 시도한 일련의 데이터 처리 과정은 일견 ‘프로그래밍’과 ‘코딩’으로 언급되는 공학 방면의 작업처럼 보이지만, 그 이면에는 인문 학술 정보를 효과적으로 다루기 위한 고민이 다층적으로 반영되어 있다. Valid XML 데이터를 편찬하기 위한 표준 형식의 XML Schema 설계와 그것을

반영한 미디어위키 문서 작성이 그러한 고민의 대표적 흔적에 해당한다. 인문 학술 논문에 담긴 다양한 정보를 디지털 데이터로 정교하게 옮겨내기 위한 방법을 그동안 적극적으로 고민한 사례가 없었다는 측면에서, 그리고 해당 고민을 문제의식으로 이러한 실험을 수행한 주체가 이공계열(STEM)이 아니라 인문학 전공 연구자들이라는 사실이 그 나름의 시사점이라 할 것이다.

파일럿 성격의 융합적 시도로서 본 연구를 통해 구현된 데이터 처리 과정은 향후 디지털 환경에서 인문 학술 데이터를 효과적으로 편찬·유통·활용하기 위한 기초 자료로 활용될 것이다. Triple(S-V-O) 형식으로 가공된 데이터는 학술 LOD와 학술 지식그래프를 구현하고, 각종 학술 데이터 분석과 시각화 작업을 위한 기초 리소스가 될 뿐만 아니라 인문 학술 인공지능을 개발하기 위한 학습 데이터로도 활용할 수 있다.

본 연구에서 구현한 내용을 토대로 추후 더욱 정교한 Valid XML 데이터를 편찬하기 위한 표준 형식의 XML Schema를 보완하는 과정에서, 인문 학술 지식그래프를 구성하는 여러 문맥 요소(contextual element)를 마련하고 그 연장선상에서 인문 학술 온톨로지(Ontology)를 설계하기 위한 연구에, 다양한 전공의 인문학 연구자들이 참여할 기회를 마련하는 것이 다음 단계의 과제가 될 것이다.

Ⅲ. 참여형 인문데이터 편찬 모델 사례: Linked Jazz

앞 장에서 정리한, 학술논문을 대상으로 한 표준적 형식의 데이터 처리 과정에 대한 실험적 모색과 별개로, 추후 그와 같은 방식으로 구축된 데이터를 디지털 환경에서 어떻게 활용할 것인지, 데이터의 활용 맥락에서 대중이

나 시민이 학술 과업에 능동적으로 참여할 방안이 가능한지, 가능하다면 그것은 구체적으로 어떠한 형식인지에 관한 탐구는 기초적 수준의 학술 데이터 프로세싱에서 더 나아가 보다 확장된 학술 데이터 프로세싱에 해당한다. 앞 장에서 수행한 파일럿 형식의 연구와 같이 기초 작업을 수행하는 것 정도로 본 연구의 의미를 한정할 수도 있겠으나, 본 연구자는 기초 연구에 더해 조금 더 확장된 학술 데이터 프로세싱의 외연을 탐색하는 것이 향후 본 연구를 심화해나가는 과정에서 장기적 발전을 도모하는 단서가 될 것으로 판단하였다.

이에 따라 본 연구는 디지털 환경에서 인문 학술 데이터를 처리하는 과정을 구현한 데서 더 나아가, 그러한 기초 데이터를 매개로 대중이나 시민이 폭넓게 학술 연구에 참여할 수 있는 방안을 고민해보고자 하였다. 당장 특정한 학술 모형을 전제로 구체적인 연구 참여플랫폼을 구현하는 것은 현실적으로 어렵기에, 실제 유관 사례를 검토하고 그 시사점을 정리하는 방향으로 연구를 진행하였다. 이번 장에서 대상으로 삼은 연구 사례는 연구 데이터를 편찬하는 과정에 대중이 직접 참여할 수 있는 형식을 반영한 Linked Jazz¹¹ 프로젝트이다.¹²

해당 프로젝트는 재즈 전문가로서의 연구자들뿐만 아니라 재즈를 애호하거나 대중음악에 관심이 있는 일반인들에게 데이터를 매개로 연구 프로젝트에 참여할 수 있는 채널을 열었다는 측면에서, 인문 데이터 편찬과 공유에 있어 디지털 환경의 개방성을 적극적으로 활용할 수 있는 방법론적 모델을 선도적으로 보여주는 사례에 해당한다.

11 Linked Jazz 프로젝트 홈페이지(<https://linkedjazz.org/>, 확인: 2022년 1월10일).

12 본 장절의 Linked Jazz 프로젝트에 관한 내용은, redInt Wiki의 Linked Jazz 항목 페이지(http://dh.aks.ac.kr/~red/wiki/index.php/Linked_Jazz)에 정리된 여러 정보를 참고하여 서술되었다(확인: 2022년 1월 10일).

1. 연구 목적과 대상

Linked Jazz 프로젝트는 디지털 포맷의 문화유산 데이터 구축과 그것을 대상으로 한 시각적 구현 도모 그리고 그 과정에서의 새로운 지식 발견을 목적으로 진행되었다. 미국을 대표하는 여러 재즈 뮤지션 사이의 관계를 디지털 아카이브로 구축하는 가운데, 그들을 매개로 한 지역 사회와 문화 네트워크를 효과적으로 드러내기 위한 방법론으로서 Linked Open Data (LOD) 기술의 유효성을 검토하고자 하였다. 예컨대 여러 아카이브와 라이브러리에서 추출한 다채로운 데이터를 연결하기 위해 LOD 기반 데이터 처리 플랫폼 구축을 중요한 목표로 삼았으며, 그렇게 구축한 데이터를



그림12-Linked Jazz 메인페이지 모습

대상으로 네트워크 시각화와 분석까지 시도하였다.

Linked Jazz 프로젝트는 재즈를 연구 대상으로 하였다. 재즈라는 장르의 역사와 문화적 특질을 연구하기 위한 지름길은, 재즈 뮤지션들에 관한 기록을 토대로 그들에 관한 정보를 정리하는 것이다. 재즈 커뮤니티는 뮤지션, 멘토, 라이벌, 연인 및 친구 사이 등 그들 사이에 존재하는 다양한 관계를 통해 그 의미가 드러난다. 뮤지션들 사이의 네트워크를 정교하게 구현하고 그로부터 재즈 커뮤니티의 풍부함을 확인할 수 있다면, 재즈라는 장르를 이해하는 새로운 채널을 구축하게 되는 것이다.

이와 같은 문제의식을 바탕으로 Linked Jazz 프로젝트는, 재즈 뮤지션들의 음악 활동과 그것을 매개로 그들 각자의 삶이 반영된 사적 기록의 유의미한 연결점을 찾아 폭넓게 드러내는 결과물 구현을 목표로 하였다. 여러 도서관, 박물관, 아카이브 등에서 소장한 다양한 자료를 대상으로 데이터를 구축하였으며, 이로부터 재즈 뮤지션들이 이루었던 재즈 커뮤니티의 네트워크를 구현함으로써 디지털 환경에서 재즈의 역사를 본격적으로 재현하고자 하였다.

2. 연구 주체와 기간

Linked Jazz 프로젝트는 뉴욕 맨해튼에 위치한 Pratt Institute: School of Information의 Semantic Lab¹³에서 시작되었다. 해당 Lab은 도서관, 박물관, 아카이브 등에 적용하기 위한 새로운 시맨틱 기술 방법론 및 도구 개발을 위한 연구에 주력하고 있다. Lab을 대표하는 연구자(Co-director)로 크리스티나 파투엘리(Cristina Pattuelli) 교수와 매트 밀러(Matt Miller) 교수

13 Semantic Lab 홈페이지(<https://semlab.io/>, 확인: 2022년 1월 10일).

가 있다.¹⁴

크리스티나 파투엘리 교수 중심으로 시작된 본 프로젝트는 2011년 9월 Dublin Core and Metadata Applications 국제 컨퍼런스에서 파일럿 버전의 연구 내용이 소개되었다.¹⁵ 이후로 데이터 모델링 작업과 데이터베이스 편찬 작업을 효율적으로 진행하기 위한 디지털 도구들을 매년 지속적으로 개발해왔다. 대표적인 결과물은 Network Visualization Tool(2012), Linked Jazz 52nd street(2013), ECCO!(2014) 등이 있다.

프로젝트 적용 목적의 기술 개발뿐만 아니라 링크드오픈데이터(LOD)나 디지털 인문학 또는 잊힌 재즈아티스트 복원과 관련된 출판 및 학술 활동 또한 2011년 이후로 현재까지 꾸준히 지속하고 있다. 특히 2018년 이후의 활동은 주로 그동안 구축한 전체 데이터를 개방하고 공유하는 데 초점을 두고 있는 것처럼 보인다.¹⁶

3. 연구 절차와 방법

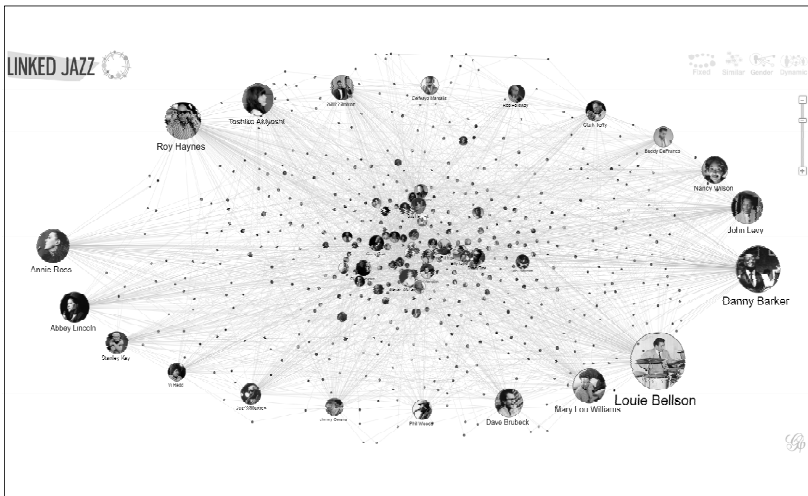
Linked Jazz 프로젝트는 미국 럿거스 대학교(Rutgers University) 부설

14 크리스티나 파투엘리 교수는 Linked Jazz 프로젝트의 실질적인 연구 책임자이다. 정보 및 도서관학 전공자이며 철학 및 문화유산 연구를 함께 전공하였다. 현재는 문화유산 자원에 적용되는 시맨틱 웹 기술을 연구하고 있다. 매트 밀러 교수는 Linked Jazz 프로젝트의 기술 책임자이다. 문화유산 자료를 다루는 기술 연구자로서, 문화 및 역사와 관련된 지식에 보다 효율적으로 접근할 수 있는 새로운 디지털 방법론을 개발하는 데 주력하고 있다.

15 'Linked Jazz: An Exploratory Pilot'이라는 제목의 논문으로 발표되었다. 관련 발표문 페이지(<https://dcpapers.dublincore.org/pubs/article/view/3637.html>, 확인: 2022년 1월 10일).

16 Linked Jazz 프로젝트 홈페이지(<https://linkedjazz.org/>)에서 상단의 'NEWS' 항목을 선택하면 2011년 이후로 프로젝트 연구팀에서 주로 활동한 내용을 모두 확인할 수 있다(확인: 2022년 1월 10일).

재즈학 연구소(Institute of Jazz Studies)¹⁷에서 소장한 재즈에 관한 방대한 자료에서부터 시작되었다. 재즈 뮤지션 57인의 모습이 담긴 에스콰이어의 흑백 사진 'A Great Day in Harlem(1958)¹⁸'을 매개로, 재즈 뮤지션 개개인의 위키피디아 문서는 있지만, 재즈 뮤지션들 사이의 관계를 다룬 데이터베이스는 없다는 점에 주목하였다.



17 러거스 대학교(Rugers University) 부설 재즈학 연구소홈페이지(<https://www.libraries.rutgers.edu/newark/visit-study/institute-jazz-studies>, 확인: 2022년 1월 10일).

18 'A Great Day in Harlem(1958)' Wikiedia([https://en.wikipedia.org/wiki/A_Great_Day_in_Harlem_\(photograph\)](https://en.wikipedia.org/wiki/A_Great_Day_in_Harlem_(photograph)), 확인: 2022년 1월 10일).

19 <https://linkedjazz.org/network/>, 확인: 2022년 1월 10일.

프로젝트 팀에 합류함으로써 두 기관의 본격적인 협업이 시작되었다. 이후 Linked Jazz 프로젝트 팀은 자료 수집 범위를 확대해서, 미국 전역의 재즈 관련 구술 자료를 확보하는 작업을 진행하였다.

여러 대학과 아카이브 기관에서 소장한 자료를 대상으로 체계적인 수집이 진행되면서, 그것들을 종합한 재즈에 관한 구술 역사 기록(Oral History Transcripts)을 매개로 개개의 재즈 뮤지션 이름을 정리하는 작업이 이루어졌다. 다양한 자료에 중복해서 담긴 수많은 재즈 뮤지션의 이름을 매핑하는 작업을 통해 9,000명 이상의 뮤지션 이름 및 관련 URI의 dataset을 구축하였다.

해당 dataset은, 대상과 객체 그리고 술어(S-V-O)로 구성되는 Triple 형식의 데이터이다. LOD 기반의 Triple 데이터는 URI를 연결하는 동시에 노드 사이의 관계에 의미를 정의하는 온톨로지를 사용하는데, Linked Jazz 프로젝트는 기본적으로 FOAF(Friend of a Friend)²⁰ 온톨로지를 사용함으로써, 재즈 뮤지션들 사이의 인연을 가까운 친구, 자녀 그리고 멘토와 같은 유의미한 관계로 정의하였다. 예컨대 해당 작업에 적용된 온톨로지는, 어느 뮤지션이 인터뷰에서 다른 뮤지션을 언급하면 두 뮤지션이 서로 알고 있다고 가정해 뮤지션들 사이의 관계를 의미적으로 판별할 수 있게끔 한다. 이러한 방식으로 재즈 뮤지션들을 연결해 그들 사이의 소셜 네트워크를 구현하는 것이다. 이와 같은 작업은, 단절된 것처럼 보이던 재즈 뮤지션들 사이의 관계를 의미적 데이터로 복원해 실제 연결되어 있던 재즈 커뮤니티를 디지털 환경에서 복원하는 연구에 해당한다.

4. 인문 데이터 편찬 참여 모델 구현과 그 의의

디지털 인문학(Digital Humanities) 연구에 해당하는 여타 프로젝트와는

20 FOAF Vocabulary Specification(<http://xmlns.com/foaf/spec/>, 확인: 2022년 1월 10일).

차별되는 Linked Jazz만의 독특한 데이터 처리 과정이 있다. 데이터 편찬 과정에 적용되는 크라우드소싱(Crowd Sourcing) 방식이 그것인데, 이는 인문학 데이터를 처리하는 과정에서 데이터를 매개로 대중이나 시민이 인문학 연구에 폭넓게 참여할 수 있는 방안을 현실에서 구현한 것으로서의 의미가 있다.

재즈를 전공으로 하는 인문학 연구자에 의해 생산된 전문적 수준의 재즈 지식을 데이터로 구축해서 그것을 디지털 환경에서 일방적으로 제공하는 방식이 아니라, 재즈에 관심이 있는 대중이나 시민이 자발적으로 재즈에 관한 데이터를 구축하고, 그렇게 구축된 데이터를 전문가들이 다시 검토하는 쌍방향적 과정을 거쳐 최종적인 연구 데이터를 공개하는 방식을



그림14-〈Linked Jazz 52nd Street〉의 메인 화면 예시

도입한 것이다. 이러한 쌍방향적 모델의 데이터 처리를 가능하게 하기 위해 Linked Jazz 연구팀은 특별한 데이터 처리 플랫폼을 구현하였다. 아래 <Linked Jazz 52nd Street>²¹가 그것이다.

재즈 뮤지션들 사이의 네트워크는 뮤지션들 사이를 선으로 연결하는 형태로 나타낼 수 있겠으나, 특정 뮤지션이 누구와 어떠한 성격의 관계를 얼마나 오랫동안 유지했는가를 구체적으로 표현하기 위해서는 더욱 발전된 도구가 필요하다. 이러한 문제의식은, 섬세한 성격의 인문학 데이터를 어떻게 하면 대규모로 그리고 지속해서 편찬할 수 있을까에 대한 고민과 밀접하게 닿아 있다.

이와 관련해 재즈를 전문적으로 연구하는 소수의 인문학 전문가들의 견해에만 의존해서 연구 데이터를 편찬하는 작업은 한계가 있으며, 그렇기 때문에 재즈를 잘 아는 대중의 관심과 이해를 본 연구에 끌어들이고 그들이 데이터를 편찬하는 데 있어서 직접적으로 참여하게끔 유도하는 것이 매우 중요하다는 사실을 알 수 있다. 이러한 고민을 바탕으로 Linked Jazz 연구팀은 클라우드 소싱 플랫폼 개발을 진행하였다.

클라우드 소싱은 ‘대중(crowd)’과 ‘외부 자원 활용(outsourcing)’의 합성어로서, 전문가 대신 비전문가인 대중에게 문제의 해결책을 아웃소싱하는 것을 의미한다. 대중음악으로서 재즈에 대한 정보나 그것을 대상으로 한 이해를 고려할 때, 여기에 관심을 둔 대중들의 지식과 이해를 모을 경우 더욱 풍부한 결과물을 구축할 수 있겠다는 생각에서 비롯된 것이다. 매트 밀러 교수는 클라우드 소싱 데이터 편찬 플랫폼으로 <Linked Jazz 52nd Street>를 개발하였으며, 이것을 활용해 재즈 뮤지션들 사이의 관계에 대한 더욱 섬세한 조망을 도모하고자 하였다.

21 <Linked Jazz 52nd Street>(https://linkedjazz.org/52ndStreet/, 확인: 2022년 1월 10일).

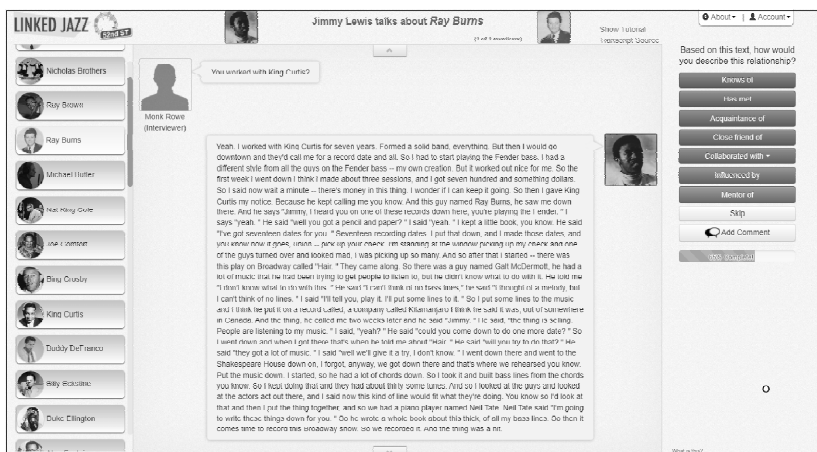


그림15-〈Linked Jazz 52nd Street〉의 데이터 구축 참여 페이지 화면 예시

〈Linked Jazz 52nd Street〉는 아카이브에 저장된 재즈 관련 녹취록을 사용자에게 공개함으로써, 그것을 읽은 사용자가 뮤지션들 사이의 관계를 직접 정의할 수 있도록 설계된 클라우드 소싱 성격의 데이터 편찬 도구다. 특정 재즈 뮤지션이 다른 재즈 뮤지션에 관해 언급한 녹취록을 텍스트로 읽은 사용자는 자신만의 이해와 시각을 토대로 해당 내용을 해석해서 그들 사이의 관계에 대해 직접 정의를 내리고 그에 대한 정보를 Triple 형식의 데이터로 추가할 수 있다. 연구 데이터의 완성 버전은 연구팀의 최종 검토를 거쳐 정의되지만, 해당 데이터를 구축하는 과정에서 중요한 단서를 제공하는 주체로서 대중이 직접 참여하고 기여할 수 있다는 사실은, 디지털 환경에서 대중 참여 기반의 인문 데이터 편찬 및 유통 모델이 어떤 형식으로 설계되어야 하는지에 대해 새로운 통찰을 제공한다.

연구자의 전문적 시각에만 의지하기에 일방적으로 전달될 수밖에 없는 폐쇄적 성격의 인문 학술 지식이 아니라, 시민과 대중이 지식을 생산하는 과정에 능동적으로 참여함과 동시에 전문연구자들이 그들과 함께 호흡할

수 있는 학술 활동이 디지털 환경에서 가능하며, 그것을 구체적으로 매개하는 역할로서 데이터의 가능성을 생각해보게끔 만든다.

〈Linked Jazz 52nd Street〉를 통해 드러나는 Linked Jazz 프로젝트의 지향점은, 곧 데이터 기반의 참여형 인문 학술 연구플랫폼으로서 미래의 인문학 연구 환경을 적극적으로 고민하는 데 있다고 해도 과언이 아닐 것이다. 적극적 ‘개방’과 ‘참여’ 그리고 ‘공유’의 실천으로서 인문학 연구 데이터 프로세싱의 확장된 외연을 추후 더 능동적으로 모색할 필요가 있을 것이다.

IV. 맺음말

아날로그 환경에서 간행되는 학술지 중심의 지식 편찬 및 유통 시스템은 최초의 지식 생산자(연구자 및 학회)가 해당 지식을 소유하고 그로 인해 특정 지식이 고정 불변적 형태로 유통되는 방식이 보편적이다. 지식의 내용은 곧 해당 지식을 다루는 학술 영역(field)의 방법론적 형식과 불가분의 것으로 취급되어 왔으며, 이에 따라 개별 학술 분과에서 보편적으로 다루는 여러 방법론적 형식을 익히는 것이 지식을 다루는 핵심이자 전문성을 확보하는 근거로 수용되어 왔다. 이러한 학술 환경의 지속은 연구자들 사이에서 그 나름의 분과 전문성을 인준하는 안정적 시스템으로 자리 잡아 왔으나, 한편으로 분과 전공을 중심으로 한 학술 활동의 진입 장벽이 높아짐에 따라 분과 간 지식 유통 가능성에 있어서 부정적 영향을 끼쳐 온 것도 사실이다.

디지털 환경에서 이루어지는 지식 공유 문화의 확대는 그러한 아날로그 환경의 학술 풍토를 비판적으로 바라보게끔 유도한다. 디지털 환경에서

유통되는 지식은 특정한 주체(연구자, 학회)에 의해 소유된 것으로서 ‘고정 불변’한 것이 아니라, 해당 지식을 다루고자 하는 여러 사용자에게 의해 점유되어 자유롭게 활용되는 형식으로 나타난다. 확정되고 고정된 것으로서의 지식이 아니라 조립 가능한 데이터의 형식으로 분절되어 다루어지기에, 특정 지식이 다루어지는 맥락에서 오류가 발생할 가능성은 항상 존재하며, 동시에 그 오류가 수정될 수 있는 가능성 또한 언제나 상정된다. 이러한 점은 향후 디지털 환경에서의 지식공유 형식이 아날로그 환경의 학술 문화에 끼칠 영향력의 성격이 무엇인지에 대해 짐작해볼 수 있게끔 한다.

디지털 환경에서 다루어지는 지식의 ‘모듈화’ 경향은 오픈소스(open source)에서 두드러진다. 단일 전문가에 의해 일방적으로 구성되는 지식이 아니라, 여러 사용자에게 의해 개량되고 재배포되는 소스 코드의 양상은, 고정된 중심 지식의 ‘부속’으로서의 모듈이 아니라 ‘조립’을 통해 점차 그 모습을 구체화해 나가는 새로운 형식의 지식에 대한 간접적 인상을 드러낸다. 불변하는 정답이자 고정된 사유로서의 인문 지식이 아니라 변화하는 가운데 유연한 질문으로 연결될 수 있는 인문 지식 특유의 외연을 재현할 수 있는 공간으로서 디지털 환경을 이해할 필요가 있다.

대중이 지식을 ‘소비’하는 형식은 점차 선택과 조립에 기초한 모듈화 형태로 이행해 갈 가능성이 크다. 전문가로서 인문학 연구자가 지식을 생산하고 생산된 지식이 디지털 환경에서 효과적으로 유통되고, 유통된 지식이 대중에게 자유롭게 소비되는 선순환 시스템에 대한 고민이 그 어느 때보다 필요한 시점이다. 그러나 여전히 인문 지식을 전문적으로 ‘생산’하는 입장에 서 있는 대부분의 연구자는, 글쓰기 형식이 익숙하다.

디지털 환경에서 인문 지식을 효과적으로 ‘유통’하기 위해서는 글쓰기를 통해 편찬된 지식을 데이터화하는 프로세스를 반드시 고민해야 한다. 연구자들이 글쓰기를 통해 생산한 지식으로서의 플레인 텍스트(plain text) 데이

터는 대중의 참여를 유도할 수 있는 모듈화된 지식으로서의 시맨틱 데이터(semantic data), 즉 Triple(S-V-O) 형식으로 연계할 수 있게끔 구축되어야 하며, 그에 최적화된 여러 가지 디지털 기술이 도입되어야 한다. 당연히 그것을 다루는 여러 주체(연구자, 일반대중)를 대상으로 한 다양한 교육 또한 이루어질 필요가 있다.

이 글에서 인문학 데이터 편찬 프로세스에 관한 문제의식을, ‘구현 연구’와 ‘사례 연구’ 두 가지 맥락으로 나누어 각기 정리한 것은, 바로 그러한 문제의식에 기초한 실험적 작업에 해당한다. 인문 학술 지식 생산 및 유통이 디지털 환경에서 가능한지 가능하다면 그 구체적 형식은 무엇인지 실제 웹 환경에서 컴퓨터 기술을 활용해 구현하고 그 내용을 정리한 것이 II장이며, 그러한 과정에서 전문 연구자에 의한 일방적 지식 전달이 아니라 디지털 환경에서 시민이나 대중이 참여할 수 있는 새로운 학술 모델에 관한 구상을, 실제 사례를 매개로 살펴본 것이 III장이다.

II장과 III장을 종합해서 본 연구자가 논증하고자 한 핵심 지점은, 웹에서 보편적으로 통용되는 포맷의 인문 학술 데이터 편찬이 불가능한 것이 아니며 동시에, 데이터 기반 학술 공공성에 대한 인식이 지금이라도 확대되어야 한다는 사실이다.

예컨대 기존의 공공(公共, public) 개념을 구성하는 각각의 요소로서 ‘공동적(共同的, common)’, ‘공개적(公開的, open)’, ‘공유적(共有的, sharing)’ 속성이 디지털 커뮤니케이션을 매개로 더욱더 유기적으로 얹혀 들어가는 양상은, 기술 발달에 의해 산업과 사회가 점차 모듈화되어 가는 일련의 과정으로 이해할 수 있다. 넷(Net) 또는 웹(Web)으로 대표되는 디지털 환경은 사용자들이 공동으로 참여하는 공간이자, 무엇이든 자유롭게 공개할 수 있는 공간이자, 소통을 위해 공유하는 공간이다. ‘민중(民衆, People)’이나 ‘대중(大衆, Mass)’과는 다른 공공의 주체로서 ‘다중(多衆, Multitude)’의 등장은

기본적으로 ‘개체(the individual)’이지만, 디지털 환경에 ‘접속(access)’함으로써 동시에 ‘공동체(communitiy)’로 기능하는 새로운 형태의 사회적 주체의 탄생을 의미한다.

디지털 환경에서 이루어지는 다중의 참여는, 민중이나 대중의 참여 양상과 달리 참여 그 자체가 목적이자 놀이가 된다. 무언가를 얻거나 소비하기 위해 참여하는 것이 아니라, 참여가 자연스럽게 놀이로 인지되는 맥락은, 다양한 모듈이 종합된 사회 내에서 개체에 의한 역할 수행의 가치가 점차 커지고 있음을 넉넉히 알려준다. 학술 지식 생산과 유통 또한 마찬가지로, 수용자가 능동적으로 참여할 기회 없이 소수의 생산자에 의한 일방적 지식 전달 방식이 지속된다면, 그로부터 발생하는 정보와 지식은 교환되지 못한 채 대부분 폐기될 가능성이 크다. 이 글에서 제안한 인문 학술 데이터 프로세싱에 관한 시론적 내용이 향후 데이터 기반 인문학 연구 환경에 관한 논의를 더욱 풍부하게 확대해 나가는 데 있어서 기여하는 바가 있기를 바란다.

참고문헌

1. 단행본

김성우·엄기호, 『유튜브는 책을 집어삼킬 것인가』. 서울: 따비, 2020.

2. 논문

김지선·장문석·류인태, 「공유와 협업의 글쓰기 플랫폼, 위키」. 『한국학연구』 60집, 2021, 371~419쪽.

윤지선, 「‘관음충’의 발생학: 한국남성성의 불완전변태과정(homomorphism)의 추이에 대한 신물질주의적 분석」. 『철학연구』 127집, 2019, 259~288쪽.

천정환, 「‘문화론적 연구’의 현실 인식과 전망」. 『상어학보』 19집, 2007, 11~48쪽.

3. 웹리소스

인문학과 지식공유: 시민 참여 모델의 개발 연구 페이지(<http://www.klbksk.com/wiki/index.php/HumanitiesDataArchive>).

천정환(2007), 「‘문화론적 연구’의 현실 인식과 전망」 학술논문 위키(<http://www.klbksk.com/wiki/index.php/HDAA0001>).

한국어문학 지식공유 위키(<http://www.klbksk.com/wiki>).

‘A Great Day in Harlem(1958)’ Wikiledia([https://en.wikipedia.org/wiki/A_Great_Day_in_Harlem_\(photograph\)](https://en.wikipedia.org/wiki/A_Great_Day_in_Harlem_(photograph))).

FOAF Vocabulary Specification(<http://xmlns.com/foaf/spec/>).

Linked Jazz 프로젝트 웹사이트(<https://linkedjazz.org/>).

Linked Jazz network(<https://linkedjazz.org/network/>).

Linked Jazz 52nd Street(<https://linkedjazz.org/52ndStreet/>).

‘Linked Jazz: An Exploratory Pilot’ 발표문(<https://dcpapers.dublincore.org/pubs/article/view/3637.html>).

Sigma.js 라이브러리(<http://sigmaj.js.org>).

JASON 데이터 네트워크 시각화 결과물(<https://byungjunkim.github.io/PaperKnowledgeGraph/>).

redInt Wiki의 Linked Jazz 정보 페이지(<http://dh.aks.ac.kr/~red/wiki/index.php/>)

Linked_Jazz).

Rutgers University 부설 Institute of Jazz Studies(<https://www.libraries.rutgers.edu/newark/visit-study/institute-jazz-studies>).

Semantic Lab 홈페이지(<https://semlab.io/>).

국문초록

학술논문을 읽는 행위 또는 고급학술 지식을 다루는 문화가 축소되어가는 양상을, 단순히 ‘글쓰기’로 대변되는 인문학 연구 방법론이 지닌 한계적 문제로만 환원해서는 곤란하다. 그 원인은, 오랜 기간 인문학계에서 유지되어온 전공 간 장벽의 문제 그리고 그에 입각한 글쓰기가 업적평가 제도와 맞물려 다시 취업 채널까지 연결되는 인문 학술 환경의 현실적 여건 전반과 관련이 있다고 해야 할 것이다. 뿐만 아니라 디지털 기술의 발달과 영향력 확대에 따른 리터러시 형식의 급격한 변화는, 학술환경의 현실적 문제와는 다른 맥락에서 ‘학술적 글쓰기’의 변모를 끊임없이 요청하고 있다. 이와 같은 현실에 착안해 이 글은 본론을 두 부분으로 나누어 크게 두 가지 문제의식을 풀어내는 데 주력하였다. 첫 번째는 ‘데이터’를 매개로 각각의 학술논문을 아카이빙하고 그것들에 담긴 지식을 연결하는 방안을 실험적으로 모색해보는 것이다. 또 하나는 디지털 환경에서 학술 데이터를 능동적으로 ‘활용’하고 더 나아가 학술 지식을 편찬하는 과정에 적극적으로 ‘참여’할 수 있는 방안이 무엇인지, 유관 사례를 통해 검토해보는 것이다. ‘데이터’는 ‘전공’, ‘글쓰기’, ‘지식’으로 표방되는 기존 인문 학술 환경의 좁은 틈을 비집고 들어가, 학술연구를 통해 축적된 다양한 정보를 공유하고 그로부터 각기 분리된 전공 영역의 이해를 연결하는 매개가 될 수 있다. 인문학술 데이터 편찬 및 활용 방안에 대한 모색은 곧 현재 상황을 개선하기 위한 적극적 고민이자 미래 학술지형을 준비하는 능동적 대응이라 할 것이다.

투고일 2022. 1. 17.

심사일 2022. 4. 12.

게재 확정일 2022. 5. 3.

주제어(keyword) 인문지식(Humanities Knowledge), 학술정보(Academic Information), 웹표준(Web-Standard), 디지털 인문학(Digital Humanities), 링크드 재즈(Linked Jazz)

Abstract

An Essay on a Humanities Data Processing

Ryu, Intae

The humanities research methodology represented by 'writing' is not wholly to blame for the atrophy of the culture dealing with sophisticated academic knowledge. It has to do with long-standing barriers between humanities majors, as well as the practical realities of the academic environment, where writing based on it is entwined with the research performance evaluation system and links to job channels. Furthermore, the rapid changes in literacy brought about by the advent of digital technology necessitates the ongoing modification of 'academic writing', regardless of the practical issues that face academics. This paper provides examples for academic knowledge data processing in the humanities in this context by focusing on two points of view. The first is to see how 'data' can be used to archive academic papers in the humanities and connect the knowledge contained within them. The second is to look at ways to actively 'participate' in the process of gathering academic knowledge, rather than just using academic material in a digital environment through case review. By exchanging diverse information collected via academic research, data can connect understanding of different fields amongst majors, bypassing the existing academic environment's closed structure, which is characterized by 'major', 'writing', and 'knowledge'. Finding new ways to create and use data in the academic environment of humanities is a proactive concern for improving the existing situation and preparing for the future academic environment.