

경제·인문사회연구회
인문정책연구총서 2019-10

디지털-빅데이터의 시대 인문학의 새로운 역할에 대한 고찰

이주영 UNIST



연구수행자

연구기관	연구책임자	참여연구진
UNIST	이주영	
중앙대학교		김바로
충남대학교		노서경
이화여자대학교		노경덕
아주대학교		박구병
서울대학교		박재욱
서울대학교		조미혜
연세대학교		김지윤

이 보고서는 『경제·인문사회연구회 2019년도 인문정책연구사업』의 일환으로 수행된 연구 과제 중 하나입니다.

이 보고서에 수록된 내용은 집필자의 개인적인 견해이며 경제·인문사회연구회의 공식적인 입장이 아님을 밝힙니다.

연구요약

인간 사회의 전 분야를 쓰나미처럼 휩쓸고 들어오는 디지털화는 개인의 생활 방식, 개인 들기리의 교류 방식, 더 나아가 사회가 작동하는 방식을 총체적이고 근본적으로 변화시키고 있는 상황이다. 우리는 현재 디지털을 사용하는 선을 넘어서서 디지털 가운데서 살고 있다. 디지털이라는 매체가 인간 삶의 전반적 틀을 크게 좌우하는 이러한 상황은 여러 가지 전례 없는 현상들을 낳고 있으며, 이러한 변화가 긍정적일지 부정적일지는 현재의 우리가 어떻게 대처하느냐에 달려있을 수밖에 없다. 평자에 따라서는 농업혁명, 산업혁명 이상의 역사적인 전환점이라고도 평가되고 있는 현재 시점은 곧 새로운 사회구조의 원리와 더불어 새로운 방식의 학문이 출현하는 시점인 듯하다.

이 글에서는 먼저 개인정보의 문제와 젠더의 문제에 집중하여 디지털-빅데이터 시대가 낳은 사회적 변화의 양상을 조망하고자 했다. 인간의 생활 전반이 디지털에서 흔적으로 남고 있는 현실은 그 흔적, 즉 “trace data”를 그 자체로 “상품”으로 간주하려는 경향과 결부되고 있다. 디지털 사회로의 진입이 이미 이루어졌고 후퇴가 불가능할 뿐만 아니라 디지털-빅데이터 사회가 갖는 장점이 매우 큰 것을 고려할 때 “데이터 공동체”, 혹은 “연대적인 데이터 저장” 등 현재 사회학에서 제시되고 있는 방법들에 대해 성찰함으로써 개인데이터가 “상품”으로 직진하지 않도록 하기 위한 구체적인 방법을 찾아내기 위해 노력할 필요가 절실히 보인다.

젠더의 경우 디지털-빅데이터 시대가 사회 진보의 측면에서 가능성과 한계를 동시에 안고 있음을 선명하게 보여주고 있다. 한편으로 디지털 전환 자체가 젠더에 따라 다르게 경험되고 있고, 디지털 세계에서도 기존의 남성중심적인 구조가 재생산되는 경향이 분명한 것이 현실이다. 그러나 다른 한편으로 빅데이터는 2000년대 이후 정체상태에 머물러 있는 양성평등에 한걸음 더 가까이 갈 수 있는 가능성을 내포하고 있기도 하다. 젠더 격차의 문제를 빅데이터를 통해서 입체적으로 보여줄 수가 있고, 그 결과로 구체적이고 적극적인 개입이 가능하다는 것이 여러 가지 경로를 통해서 드러나고 있는 것이다. 가령 유엔이 클린턴 재단 등 거대 민간 재단들과 공동설립한 Data2X의 경우 기존의 데이터들을 모으는 데에서 더 나아가 이 데이터들을 해체함으로써 결핍된 데이터, 즉 여성에 대한 데이터를 채우는 것을 목표로 삼고 있다. 이와 같이 빅데이터의 입체성은 인간의 사회가 젠더의 측면에서 얼마나 기울어진 운동장인지를 보다 더 명징하게 보여주고 있고, 디지털을 통해 가능해진 과거와 비교할 수 없는 정도로 진전된 국제협력을 통해서 과거보다 더 적극적이고 구체적인 방식의 대응을 가능케 할 가능성을 내포하고 있는 것이다.

디지털화가 내포하고 있는 사회 전반의 변화는 분과학문 단위에서 접근할 수 있는 범위를 넘어서 있다. 이러한 문제적 상황에 대해 인문학과 사회과학 분야가 각각 어떻게 대응하고 있는지를 II부와 III부에서 살펴보았다. II부에서는 먼저 디지털인문학이 출현함으로써 새

롭게 제기되고 여러 쟁점들, 즉 지식 생산에 있어 학자와 대중의 경계 허물기의 가능성, 디지털 인문학이 기존 위계질서의 전복에 기여하는지 혹은 재생산에 기여하는지, 디지털 인문학이 과거 반짝 인기를 누리다가 사라진 계량경제사(Cliometrics)와 같은 전철을 밟아가게 될 것인지, 디지털 인문학 연구에 대한 평가 기준 마련 등 디지털 인문학 분야 특수성을 어떻게 다룰지 등의 문제점에 대해 검토하였다.

다음으로 디지털 인문학 연구에서 앞서가고 있는 미국사, 영국사, 그리고 고전학에서 보여준 최근 연구 성과물에 대해 검토하였다. 디지털화의 토대가 되는 데이터가 매우 분명하고 동시에 대량인 고전학 분야의 경우 일반 역사학 분야보다 연구와 교육 양 측면에서 모두 디지털화에 앞서 나가고 있어서 기존의 모든 프로젝트들이 디지털 연구로 전환되는 상황이라고 평가될 정도이다. 자료의 디지털화는 연구방법의 디지털화로 이어져서 지리정보, 네트워크 분석 등 다양한 방식으로 고대 세계를 입체적으로 재현하는 연구가 다양하게 진행 중이다. 이 디지털 연구의 성과는 대중과 공유되기에 용이하기 때문에 디지털화의 진전은 곧 일반 사회와 학계의 소통이 증대되는 방식으로 나아가고 있다.

미국과 영국의 디지털인문학 연구 성과물들에 비추어볼 때 디지털 인문학 연구는 과거에 대한 시뮬레이션을 가능케 하는 입체성을 내포하고 있다. 그 결과 젠더와 빅데이터의 관련성에서 드러났던 것과 마찬가지로, 2차원에 그치는 기존의 문서자료를 통해서는 드러나지 못하는 “결핍”에 대한 문제제기를 가능케 할 수 있다는 것을 알 수 있었다. 디지털 인문학에서 전통적인 역사학 방법을 대체하거나 능가할 혁신을 기대하는 비평가들에게는 실망스러운 결론일 수도 있겠으나, 학자의 임무 및 가능성이 결국 “가능한 최선의 방법”으로 역사적 실체에 접근하기 위해 노력하는 것이라면 충분한 존재의의를 가진다고 생각할 수 있을 것이다.

다음으로 영국과 독일을 중심으로 각국 정부 및 역사가 단체에서 디지털 인문학 연구에 대해 어떻게 대응하고 있는지를 살펴보았다. 영국과 독일 정부가 디지털 인문학을 대하는 태도는 회의와 호기심이 공존하고 있는 학계의 상황과 달리, 매우 과감한 투자를 하고 있음을 볼 수 있었다. 학문 생산을 뒷받침하는 조직과 펀딩을 통해 양국 정부 공히 디지털 인문학 분야를 육성하기 위해 매우 포괄적인 노력을 기울이고 있음을 알 수 있다. 60~80개에 달하는 교수직의 수는 그 대표적인 예라고 할 것이다. 또한 양국의 역사학계 역시도 매우 적극적으로 대응하고 있다. 영국 왕립 역사학회의 경우 디지털 분야에 “대중역사학상”을 수여하고 있으며, 정기적으로 관련 컨퍼런스를 진행하고 있다. 독일역사학회의 경우 “디지털인문학연구그룹”이 산하조직으로 자리잡아서 2년마다 디지털역사학 학회를 개최함으로써 연구성과들을 학계 및 대중과 공유하고 있다.

사회과학 분야의 논의를 분석한 III부에서는 사회과학에서의 디지털 흡수 노력이 훨씬 더 포괄적이고 근본적인 논의의 차원에 이르러 있음을 보여주고 있다. 포괄적이라 함은 디지털 사회학이 학계와 국가라는 전통적인 틀을 넘어서, 한편으로 마이크로소프트 등 기업이 직접 참여하는 프로젝트가 진행되고 있는가 하면, 다른 한편으로 글로벌한 차원에서의 연

구 프로젝트들이 다수 진행되고 있다는 점이다. 근본적이라 함은 디지털 데이터의 성격에 대한 논쟁을 통해서 잘 드러나고 있다. 디지털 플랫폼을 통해서 매개되는 디지털 데이터에 남는 기록들, 즉 “trace data”에서 사회적인 행위의 기록으로 살펴볼 수 있을 것인가, 그렇다고 한다면 디지털 데이터의 어떤 측면을 “사회적인 것”으로 볼 수 있을 것인가, 즉 보다 근본적으로는 기술과 사회성, 지식의 상호 작용을 어떻게 평가할 수 있을 것인가에 대한 성찰을 시도하고 있다는 점이다.

독일의 매체학자로 주목받고 있는 프리드리히 키틀러(Friedrich Kittler)는 인간의 사유는 구체적으로는 “철자, 신호, 정보처리작업, 수학적 계산과 연산자 등 기술적 표준들이 만드는 종합적인 효과”에 불과하다고 말할 정도로 기술에 종속적인 존재로 인간을 그려냄으로 반휴머니즘, 혹은 포스트휴머니즘적인 사유의 전범을 보여주었다. 매우 극단적이기는 하지만 21세기의 호모 사피엔스들 다수가 “로그아웃”이 불가능한 방식으로 생활하고 있음을 고려할 때 그 사유의 조건에 대한 근본적인 성찰과 그에 대한 인적, 물질 투자가 매우 절실한 국면에 있다는 점은 부인할 수 없다.

목 차

1. 연구 배경	1
2. 연구 목표	1
3. 연구 진행	3

Part I. 디지털-빅데이터 시대 인문학과 사회과학

1. 디지털-빅데이터 사회가 제기하는 문제들	4
1.1. 개인정보 보호의 문제	4
1.2. 빅데이터와 젠더	7
1.2.1. 빅데이터와 “젠더 데이터 혁명(Gender Data Revolution)”의 요구	7
1.2.2. 각 분야의 젠더 불평등과 빅데이터	9
1.2.3. 젠더 문제와 빅데이터의 새로운 가능성	19
2. 디지털 시대의 인문학	21
2.1. 디지털 인문학의 여러 쟁점들	23
2.2. 디지털 인문학 연구의 성과들	29
2.2.1. 미국사 분야	29
2.2.2. 영국사 분야	33
2.2.3. Digital Classics: 고전학의 재탄생?	41
2.3. 각국 정부의 디지털 인문학 지원	45
2.3.1. 미국	45

2.3.2. 영국	49
2.3.3. 독일	57
2.3.4. 지원의 성과들	61
2.4. 역사학자들의 대응 양상	72
2.4.1. 미국역사학회의 대응	72
2.4.2. 영국역사학회의 대응	73
2.4.3. 독일 역사학회의 대응	75
2.5. 디지털 인문학 연구의 미래	76
3. 디지털 시대의 사회과학	78
4. 결론: 종합 분석	84
5. 정책 제언	87

Part II. 윈터스쿨

1. 해외 프로그램 사례	93
2. 윈터스쿨 개요 및 교육내용	97
3. 결론: 윈터스쿨에 대한 평가와 종합	98
4. 디지털 인문학 교육의 정책적 제언	101

첨부자료	107
------------	-----

참고문헌	121
------------	-----

Abstract	127
----------------	-----

1. 연구 배경

최근 몇 년간 점점 더 많은 사회의 영역들이 디지털화되어 가고 있으며, 그 결과 빅데이터, 디지털 등의 용어는 이미 “유비쿼터스”한 상태가 되었다. 오스트리아의 역사가인 볼프강 슈말레(Wolfgang Schmale)가 “디지털 시대는 이미 우리 시대의 명칭이 되었다”고 주장했을 때, 이에 대해 선뜻 반박하기는 어려워 보인다. 위키피디아에 따르면 이미 2007년에 전지구적 정보 가운데 94%가 디지털로 저장되었다.¹⁾

빅데이터는 “전통적인 데이터 가공 방식으로 평가되기에 그 양이 매우 크고 복잡하며 단기적이거나 혹은 구조화되기 어려운 데이터”를 지칭하는 말이다. 이 빅데이터는 어디에서 오는가? 매우 다양하고 포괄적인 감시 체계들, 은행의 고객 정보들, 스마트 폰 등을 통한 디지털 커뮤니케이션, 스마트 홈 등 가정 내 디지털 기술, 정부와 기업, 도서관과 대학의 데이터들 등이 그 근원이다. 이처럼 빅데이터의 근원은 포괄적이며, 2011년 통계에 따르면 세계적으로 생산되는 데이터 양은 2년마다 두 배로 증가하게 된다고 한다.²⁾ 가히 폭발적이라 할 수밖에 없다.

이러한 디지털 빅데이터 시대의 변화에 직면하여, 혹자는 농업혁명, 산업혁명 이상의 역사적 전환점이라고 평가하고 있기도 하다. 이러한 변화가 긍정적일지 부정적일지는 현재의 우리가 어떻게 대처하느냐에 달려 있을 수밖에 없다. 새로운 사회 구조의 원리와 더불어 새로운 방식의 학문이 요구되고 있는 작금의 현실은 한국 지식 사회 전반의 지혜를 필요로 하고 있다. 인문, 사회과학 분야 여러 전문가들이 모여 디지털-빅데이터 시대가 제기하는 도전에 대해 성찰하고 논의하는 장이 마련될 필요가 절실한 것이다.

2. 연구 목표

이러한 디지털 전환의 상황이 인류의 삶과 학문 발전에 있어 어떤 함의를 가지는지를 포괄적이고도 구체적으로 살펴보는 것이 이 글의 목표이다. 이 보고서는 크게 두 부분으로 구성되어 있다. Part I에서는 디지털 빅 데이터 시대의 변화가 넓게는 사회 전체, 좁게는 인문사회 과학 분야에 미치는 영향에 대해서, 즉 디지털 빅 데이터 시대가 제시하고 있는 도전과 그에 대한 학계의 응전 양상에 대해 성찰하고자 했다.

한국의 학계가 그동안 이 주제에 대해 많은 관심을 가져왔음은 물론이다. 인문학 분야에서는 기존의 고전 역사서, 문학 작품, 신문기사 등의 디지털 자료화가 꾸준히 진행되어 연

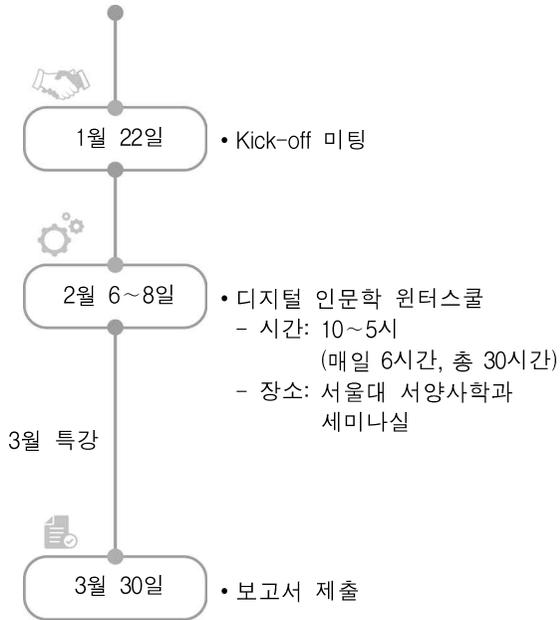
1) https://en.wikipedia.org/wiki/Big_data#/media/File:Hilbert_InfoGrowth.png

2) <https://archive.ph/20131202232836/http://www.cio.de/dynamicit/bestpractice/2281581/index.html>

구자들에게 공개되었고 이러한 디지털 데이터를 활용한 연구들이 다수 축적된 상황이다. 사회과학 분야에서는 기존의 사회조사 분석, 연결망 분석 등을 통해 양적 연구를 수행하던 양적 사회학 및 언론학 분야가 빅데이터에 대한 친화성을 발판으로 초기 빅데이터 연구에 적극 참여해 왔다. 뉴스 댓글을 통한 여론 읽기, 소셜미디어의 인터넷 담론 분석, 오프라인에서 온라인으로 중심축이 이동한 정치 및 사회 이슈에 대한 온라인 토론, 가짜 뉴스의 생성과 전파가 사회·문화에 미치는 영향에 대한 연구들이 다수 진행되고 있다. 디지털화된 공적 데이터에 대한 접근이 용이해지면서 도시의 문제를 실시간으로 감지하고 해결하기 위한 공간지리정보 활용 방안 및 스마트 도시론이 주요 이슈로 등장하기도 하였다. 인공지능 관련 기술 발전이 급진전되면서 최근에는 인공지능 기술이 노동의 구조와 임금 체계 등 사회의 거시 구조 변동에 미칠 영향, 기계가 인간의 신체를 상당 부분 대체하거나 변형시킬 포스트휴먼 시대에 인간을 가장 인간답게 하는 것이 무엇일지에 대한 존재론적 질문에 대한 연구자들의 관심도 증가하고 있다. 이러한 개별 연구들은 물론 제도적인 지원을 통해서 이루어져왔다. 이미 2002년 학술진흥재단의 과제로 디지털 인문학 관련 과목이 개발되었고 20005년 이후 한국학 중앙연구원, KAIST 문화기술대학원, 한국외국어대학교 인문대학 지식콘텐츠 학부, 연세대학교 연계전공 디지털 예술학, 한림대 디지털인문예술 전공 등이 대학의 일부분으로 자리잡았다. 또한 2013년부터 정보통신정책연구원(KISDI)이 주최한 ICT-인문사회 연구 포럼 등 대중에게 열린 학술회 등도 일반 대중에게 디지털인문학의 필요성을 각인시키고 있다.

이러한 한국 학계의 논의 및 제도적 발전을 더욱 풍성하게 하기 위해서는 세계 각국의 상황에 대한 면밀한 검토가 필요하다. 이 글에서는 서양각국이 디지털 빅데이터 시대와 관련하여 어떤 사회적이고 학문적인 논의를 해 나가고 있는지를 검토하고자 한다. 이를 위해 먼저 I부에서는 디지털-빅데이터 시대가 가져올 사회적 변화를 개인정보보호와 젠더라는 두 가지 측면에 집중하여 살펴보았다. II부에서는 이처럼 심대한 변화를 동반하고 있는 디지털-빅데이터 시대가 인문학에 어떤 도전을 제기하고 있는지를 검토하고, 다음으로 미국, 영국과 독일 등 서양 각국에서는 이러한 도전에 대해 어떻게 대응하고 있는지를 구체적으로 살펴보았다. 이어서 III부에서는 사회과학자들 사이에서 이루어지고 있는 논의들을 분석했다. IV부에서는 마지막으로 이러한 경향들을 종합하고 몇 가지 정책적인 제언을 시도했다. Part II는 실습 부분으로서 인문사회학자들이 각각의 분야에서 필요한 디지털 테크놀로지들을 효율적으로 익히기 위한 방법을 모색하고자 2020년 2월 6~8일 사흘간 실시된 워크숍 프로그램을 종합적으로 평가하고, 그에 근거하여 향후에 있을 수 있는 교육 프로그램의 방향성을 제시하고자 했다.

3. 연구 진행



Part I

디지털-빅데이터 시대 인문학과 사회과학

1. 디지털-빅데이터 사회가 제기하는 문제들

1.1. 개인정보 보호의 문제

■ 디지털-빅데이터에 대한 기대와 우려

빅데이터는 데이터를 가공하고, 보관하고, 분석하고, 공유하는 등 그 처리 과정 전반에 걸쳐서 사회적 논의를 필요로 하고 있다. 빅데이터와 관련한 사회적 논의에서는 희망과 절망이 극단적으로 공존하고 있는 듯하다. 한편으로, 여러 기업들과 정부기구들이 빅데이터들을 잘 분석함으로써 인간의 삶 전반에 걸쳐 효율성을 제고할 수 있으리라는 기대가 높다. 기업의 경우 경제 전 영역에서 비즈니스 과정을 개선할 수 있으리라는 기대가 높고, 특히 정보 기술 및 마케팅 분야에서 그러하다. 또한 정치 예측 분야, 사회적 신용 평가, 교육, 연구 등 공공부문에서 활용될 여지가 많을 것으로 관측되고 있다. 학문 분야의 경우, 특히 지리학, 유전학, 기호연구, 핵물리학 등에서 현재 빅데이터가 널리 활용되고 있다. 넓게 인문학으로 분류되는 분야에서는 미술관, 공연 예술 및 전시, 박물관의 이용과 효과를 촉진하기 위한 빅데이터 활용이 눈에 띈다.

반대로 비판자들은 구글 및 페이스북과 같은 대규모 데이터 회사들이 개인과 관련된 정보들을 무제한 활용할 수 있게 됨으로써 “감시자본주의”가 태동하게 될 것에 대해 우려하고 있다. “감시자본주의”를 저술한 미국학자 쇼사나 주보프(Shoshana Zuboff)에 따르면, 20세기 자본주의가 산업자본과 노동자 간의 투쟁이었다면 21세기 감시자본주의에서는 사회 전체가 개인에 맞서게 될 우려가 있다. 이 전쟁 하에서 우리의 육체, 우리의 자식들, 가정 생활, 도시 등 모든 것들이 감시장비의 목표물이 되고, 인간적인 자율성과 민주적인 주권이 권력과 이익을 향한 투쟁의 과정에서 희생될 수 있다는 것이다.

특히 결정적인 문제 가운데 하나는 개인과 관련되는 데이터의 소유권 문제이다. 실제로, 현재 상품구매, 은행 업무, 여행 계획 등 개인 일상생활의 많은 부분들이 온라인에서 이루어지고 있고, 이 데이터들은 다국적 기업들에 의해 수집되고 있다. 이 과정에서 개인들은 “제3자에 개인정보제공”에 동의하지 않을 경우 온라인 구매 활동 자체가 불가능한 경험들을 종종 하게 된다. 경우에 따라 이러한 데이터들이 판매되고 있기는 하지만, 무기명 데이터로 만들어지기 때문에 개인 사용자들에게는 해가 가지 않는다는 것이 그동안의 설명이었다.

그러나 독일의 데이터 전문가인 안드레아스 데베스(Andreas Dewes)의 연구에 따르면 이는 사실이 아니다. 그는 이처럼 무기명으로 된 데이터들을 사들여서 결국 사용자의 신원을 밝혀내는데 성공했던 것이다. 그가 분석한 자료는 독일의 “NDR” 방송국이 가지고 있던 삼백만 독일인에 대한 정보로서, URL을 오픈 횟수 300억 번에 해당하는 URL 자료를 확보하였고, 특정한 사용자가 어떤 소식을 클릭하는지, 어디에서 사는지, 무엇에 관심이 있는지, 새로운 직업을 구하는지, 혹은 포르노 사이트를 열어보는지, 혹은 약을 찾는지 등에 대한 간단한 정보를 알아보는 것으로 시작하였다고 한다. 한꺼번에 사용자 전체의 신원을 확인하는 것은 불가능하지만, 개인 사용자들의 경우 이들의 ID가 페이스북이나 트위터 등의 플랫폼에서 사용되는 것을 확인하는 방식으로 신원을 확인해내었다고 한다. 특히 신원확인이 용이한 트위터가 주로 활용되었다. 데베스 자신은 인터넷상에서 자신에 관한 정보가 생성되는 것을 막기 위한 방안으로 다양한 브라우저들을 활용하고, 자신의 IP 주소를 가능한 감추는 등의 방법을 제시하고 있다. 그러나 이와 같은 개인적인 노력이 문제의 해결책이 될 수 없음은 물론이다.

- 해결을 위한 다양한 접근법들: “연대적인 데이터저장”, 영업기밀보다 소비자 이해 우선, etc.

이러한 문제 상황에 대해서라면 현재 다양한 스펙트럼의 의견들이 개진되고 있는 상황이다. 독일의 사회학자 닐스 주라프크스키(Nils Zurawski)는 “연대적인 데이터 저장(Solidarische Datenspeicherung)”의 필요성을 역설하고 있다.

그는 빅데이터가 감시 사회를 만들어낼 것에 대한 우려와 관련하여 기본적으로는 “감내할 것”을 주장하고 있다. 환경을 개선하고 교통문제를 해결하는 등의 방식으로 인류의 생활 전반을 보다 나은 방식으로 조직하는 데 매우 큰 도움이 되는 등, 빅데이터 활용과 관련하여 단점보다 장점이 크기 때문이라는 것이다. 그가 대표적인 예로 지적하고 있는 것은 에너지 조달의 문제이다. 빅데이터 분석을 통해 인류의 에너지 소비 형태를 보다 잘 파악함으로써 태양열 발전, 풍력 발전 등 여러 가지 방식의 에너지 생산을 통합하고 보다 나은 경영을 통해 에너지 흐름을 효율화할 수 있다는 것이다. 이는 교통의 경우에도 마찬가지로 개인적인 동선을 반영한 효율적인 내비게이션이 만들어질 수 있다면 교통문제 역시도 완화될 수 있으리라는 것이다.

그는 디지털 세계로 들어설 지와 관련된 주사위는 이미 던져진 상태이기 때문에, 논의의 초점이 되어야 할 것은 이러한 데이터를 누가 통제하느냐일 수밖에 없다는 현실인식을 가질 필요를 강조하고 있다. 결국 그가 강조하는 것은 “연대적인 데이터 저장”으로서 일종의 “데이터 공동체(Daten-Genossenschaft)”를 만들어서 디지털 시대의 새로운 사회적 공생의 모델을 생각해 낼 필요가 있다는 것이다.

중국정부가 개인에 대한 정보들을 대규모로 축적하여 독재 체제에 버금가는 통제 시스템을 갖춤으로써 “빅브라더”를 체현하고 있기는 하지만, 민주사회의 경우 정부가 “빅브라더”로 직접 나설 가능성은 낮아 보인다. 그러나 “데이터 브로커(Data Broker)”들이 관련 데이터를 모으고 기술력을 갖추어서 이를 상업화하게 될 경우에 대비하기 위한 법규를 만들 필요성은 강조되고 있다.

가령, 현재 온라인 상으로 다양한 “지불능력평가”가 이루어지고 있는 상황이지만, 사용자들은 이를 인지하지도 못하면서도 타인에 대해서는 지속적으로 평가할 준비가 되어 있는 것이 현실이다. 에어비엔비, 우버, 레스토랑 평가 등이 그 대표적인 예이다. 이러한 평가 결과를 관리 감독 기관에만 공개할 것이 아니라 소비자들에게도 밝히는 것이 필요하며, 특히 정부가 법을 만들어서 무엇이 중요하고 무엇이 중요하지 않은 지에 대한 평가결과, 더 나아가 평가 과정을 전반을 소비자들에게 공개할 필요성이 있다는 것이 전문가들의 견해이다. 이들은

민주 사회에서 소비자의 이해가 기업의 영업비밀보다 우선시 되어야 한다는 점이 강조되어야 한다고 본다. 결국, 유럽의 데이터 보호 감독관이던 지오반니 보타렐리(Giovanni Buttarelli)가 말한 것처럼, 개인정보는 상품일 수 없다는 사회적 합의가 절실한 상황인 듯하다.

1.2. 빅데이터와 젠더

1.2.1. 빅데이터와 “젠더 데이터 혁명(Gender Data Revolution)”의 요구

오늘날 빅데이터는 전세계적으로 젠더 문제와 관련해서도 중요한 문제들을 제기하고 있다. 첫째로, 그리고 가장 근본적인 문제는 디지털 전환 자체가 젠더에 따라 다르게 경험된다는 점이다. 전세계적으로 여성들은 디지털 세계에서 남성들보다 뒤처져 있다. 이는 흔히 말하는 ‘남자 뇌와 여자 뇌’의 차이 때문이 아니라 여러 사회문화적 차별과 편견 때문에 여성들이 디지털 기술이나 과학에 접근하기 더 힘들기 때문이다. 다음으로 주목할 지점은 인터넷 환경에서 학습한 시가 성차별적 언사와 편견을 그대로 습득한 최근 사례들에서 보듯이, 기존의 데이터가 젠더 불평등에 따라 왜곡되었을 가능성이 크다는 사실이다. 그동안 보편적으로 생각해왔던 ‘인간’에 대한 자료는 실상 남성의 것이었으며, 여성은 데이터 속에 과소대표되거나 아주 배제된 경우가 많다. 이에 따라 데이터 세계에서 소위 ‘젠더 격차(gender gap)’가 심대하다는 인식이 퍼지고 있다.

하지만 젠더 불평등 해소에 새로운 돌파구를 열 가능성 역시 빅데이터에서 찾을 수 있다. 빅데이터는 젠더 불평등의 양상을 전지구적인 통계 자료를 통해 훨씬 명확하고 극적으로 보여준다. 하나의 예로 UN은 2018년 ‘지속가능한 발전 목표(Sustainable Development Goals: SDGs)’를 위한 빅데이터의 활용에 대해 보고서를 발표하면서, 주요 전략 가운데 하나가 ‘젠더 데이터를 보이게(visible) 하는 것’이라고 천명한 바 있다.³⁾ ‘젠더지수(SDGGender Index)’와 같이 현재 매년 업데이트되는 글로벌한 젠더 지표들은 충격적이다. 20세기 후반 이후 전반적으로 젠더 격차가 완화되고 있는 추세에도 불구하고, UN 보고서에

3) <https://www.unglobalpulse.org/sites/default/files/Gender-equality-and-big-data-en-2018.pdf>

따르면 아직 갈 길은 멀다. 2010년대 이후 정체 상태에 머물고 있는 미국처럼 젠더 평등의 경로가 점진적 발전이 아니라 정체나 퇴보를 보이는 선진국들도 있다. 특히 신체적이고 사회적인 성폭력을 경험했다고 하는 여성이 35%에 달하는 것이 우려스럽다. 선진국들에서는 실감하기 어렵겠지만 전세계적으로 보면 여전히 매년 7억 5천만명이 18세 이하의 나이에 결혼한다.

여기서 빅데이터의 장점은, 한 기사가 표제어로 정리했듯이, “Big Data Reveals Big Gender Inequality(빅데이터는 젠더 불평등이 얼마나 큰지 보여준다).”는 것이다.⁴⁾ 빅데이터에서 드러나는 전지구적인 남녀 간 임금, 교육, 기회의 격차나 여성살해(femicide)의 현실은 젠더 불평등이 특정 국가나 특정 집단의 문제가 아니라 전지구적인 공동의 문제임을 깨닫게 함으로써 새로운 사회운동을 낳고 있는 상황이다. 오늘날 젠더 문제가 우리나라뿐만 아니라 대부분의 나라에서 첨예한 쟁점으로 부상한 데에는 인터넷과 초연결의 시대가 가져온 새로운 인식과 연대가 바탕에 깔려 있는 것이다. 또한 빅데이터는 젠더 불평등을 ‘hard facts’로 보여줌으로써 젠더 불평등을 인지하지 못하거나 부정하는 이들에게도 좋은 설득의 자료가 된다. 더 나아가 ‘데이터에 자극받은(data-driven)’ 문제제기와 ‘데이터에 기반한(data-based)’ 정책 결정은 추상적 논의보다 젠더 불평등에 접근하는 데에 훨씬 구체적이고 효율적인 것으로 점차 입증되고 있다. 젠더 불평등에 대한 논의가 끊어오르는 우리나라도 여기에 좀 더 적극적으로 보조를 맞출 필요가 있다. 한국의 젠더 격차는 세계 젠더 격차 지수(Gender Gap Index: GGI)에서 149개국 중 115위를 차지할 정도로 심각한 상황이다.⁵⁾

그렇다면 우선 중요한 것은 젠더문제를 데이터로 가시화하는 것이다. 이를 통해 문제의식의 확산과 파급 효과를 꾀하는 국제기구나 민간조직들이 최근 늘어나고 있다. 좋은 예로 국내에 자세히 소개된 UNDP(United Nations Development Programme: 유엔개발계획)의 젠더 데이터와 세계경제포럼(World Economic Forum)의 전지구적 젠더 격차 자료는 거의 전 세계를 포함하는 젠더 현황과 연도별 추이를 집약하여 젠더 데이터를 “보이게 하는 데” 효과적이다.⁶⁾ 넓은

4) <https://dataconomy.com/2016/03/big-data-reveals-big-gender-inequality/>

5) http://www.mogef.go.kr/eng/lw/eng_lw_f002.do

6) UNDP Goal 5: Gender equality

지리적 공간과 연도별 시간을 합성하는 정밀한 수치 앞에서 각 사회는 일반적인 추이와 비교한 자신의 젠더 수준을 가늠할 수 있다.⁷⁾

민간 조직 중 가장 유명한 예는 UN이 클린턴 재단을 비롯한 거대 민간 재단들과 공동설립한 Data2X일 것이다.⁸⁾ Data2X가 천명한 사명은 “젠더 데이터 혁명(Gender Data Revolution)”이다. 여기에는 단지 빅데이터를 모으는 것(aggregate)뿐만 아니라 해체하여(disaggregate) 그 속의 ‘결핍된 데이터(missing data)’를 찾아내고 채우는 것까지 포함된다. 이 부재하는 데이터가 바로 여성에 대한 것이다. 예컨대 UN과 구호단체들이 아프리카에서 활용하는 지리정보 빅데이터는 어디가 취약 지역이고 구호물자와 의료 자원 분배가 지역 단위로 어떻게 이뤄지는지 보여주지만, 정작 그 속에서 여성과 소녀들이 어떻게 배제되고 소외되는지는 말하지 않는다. Data2X는 젠더에 입각한 빅데이터를 새롭게 수집함으로써 젠더 불평등을 촉발하는 “유리 문(glass door)”의 존재를 파악하고 새로운 빅데이터를 통해 이를 교정하기 위한 구체적 해결책을 찾자 한다. 여기에는 빅데이터를 공개하고 여성들이 쉽게 접근할 수 있도록 함으로써 소외된 여성들에게 “힘을 부여하는(empowering)”하는 전략도 포함된다. 즉, “젠더 데이터 혁명”은 빅데이터에 내재한 젠더 격차를 메움으로써 젠더 불평등을 감소시킬 길을 찾는다.⁹⁾

다음에서는 사회 여러 부문들에서 빅데이터가 보여주는 젠더 불평등의 양상을 조망하고, 이를 교정하고자 하는 의도로 나타나고 있는 빅데이터에 입각한 새로운 노력들을 실례로서 보여주하고자 한다.

1.2.2. 각 분야의 젠더 불평등과 빅데이터

1.2.2.1. 디지털 세계와 과학

<https://www.undp.org/content/undp/en/home/sustainable-development-goals/goal-5-gender-equality.html>: World Economic Forum, “The Global Gender Gap Report, 2018”에서 Figure 2: Global performance, 2018 http://www3.weforum.org/docs/WEF_GGGR_2018.pdf

7) Gender-related Development Index (GDI) and Gender Empowerment Measure (GEM).

8) <https://data2x.org/>

9) 2017년 Data2X가 공개한 보고서 참조.

<https://data2x.org/resource-center/big-data-and-the-wellbeing-of-women-and-girls/>

디지털화가 향후 미래 세계의 관건이라면 여기서 남녀격차가 벌어지는 상황은 젠더 불평등을 더 강고하게 만들 가능성이 크다. 여성들이 디지털 세계에서 남성들에게 뒤처져 있다는 것은 가정이 아니라 빅데이터가 보여주고 있는 엄연한 현실이다. 단순히 스마트폰 사용자만 보아도 여성이 남성보다 3억 2,700만 명이나 적다.¹⁰⁾ 여성들은 디지털 기술이 필요없는 단순직무나 보조직에 종사하는 비율이 훨씬 높고, 디지털 전문가가 되기 위한 교육과 투자에서도 소외되기 쉽다.

디지털 전문가 집단의 경우 젠더 격차는 더 벌어진다. ICT(Information and communications technology, 이하 ICT) 직업군에 여성은 남성보다 훨씬 적고, ICT 전문의 고위 관리직과 학계에는 남성이 여성보다 4배 많다. 미래 과학과 기술에 여성이 부분적으로만 참여하고 있다는 것은 인공지능(AI) 분야에서 더욱 분명하게 드러난다.¹¹⁾ 링크드인(Linked In)과 협동으로 조사한 바에 따르면 전 세계적으로 AI 전문직 여성은 22퍼센트이고, 나머지 78퍼센트는 남성이다. 조사자들은 이 추세는 앞으로도 별반 변화하지 않으리라 본다. OECD 국가들의 경우만 보더라도 정보학과 엔지니어 고등교육에서 여성의 비중은 20퍼센트 미만이며, 정보기술과 커뮤니케이션 분야의 여성 학위 수여자는 3퍼센트일 뿐이다.

이를 비롯한 미래 과학 세계에 여성의 참여가 미약한 상황이 반드시 개선되어야 한다는 것이 세계적인 인식이며, OECD를 비롯한 여러 기구에서 개선방안을 제시하고 있다.¹²⁾ 2008년에 시작된 독일의 Go Mint(National Pact for Women in MINT Careers) 프로그램은 전문과학 분야 여성참여를 촉진시키기 위해 250명 이상의 정치 경제 과학 미디어 종사자들과 협력하는 국민협약이다.¹³⁾ 또한 말레이시아는 과학 학위 수여자의 57퍼센트를 여성으로, 정보학에서는 50퍼센트를 여성으로 책정했으며 이 계획은 캄보디아, 케냐, 나이지리아, 베트남에서도 관심을 끌고 있다.¹⁴⁾ 디지털 미래(Digital Future)를 위한 디지털

10) <http://www.oecd.org/internet/bridging-the-digital-gender-divide.pdf>

11) <https://cursus.edu/articles/37738/les-femmes-et-le-disciplines-scientifiques-rapport-unesco>

12) Chetty, Krish; Aneja, Urvashi; Mishra, Vidisha; Gcora, Nozibele; Josie, Jaya, "Bridging the digital divide in the G20: Skills for the new age," Economics, Vol. 12, Iss. 2018-24, pp. 1-20, <http://dx.doi.org/10.5018/economics-ejournal.ja.2018-24>; Federica Minichiello, "Disciplines scientifiques et inégalités de genre : une préoccupation mondiale," <https://journals.openedition.org/ries/6787>

13) <https://www.komm-mach-mint.de/Komm-mach-MINT/English-Information>

정책, 특히 디지털 경제의 여성참여를 동등하게 끌어올리려는 것은 아르헨티나 정부의 목표이다.¹⁴⁾ 멕시코 역시 2017년부터 교육부 주관으로 ‘여성의 과학적 역량 신뢰’라는 과제에 착수했다. 2009년 20개국의 교육부가 모인 교육자들의 유럽 공동체 “Scientix”는 과학 분야에 중점을 두고 유럽스쿨넷의 조정 아래 재정 지원을 받아 소위 STEM 부문에서의 여성 참여를 늘리려 한다.¹⁶⁾

디지털과 여성 사이를 긴밀하게 하려는 것은 단순히 젠더에 기초한 적극적 우대 조치(affirmative action)만 뜻하지 않는다. 이는 디지털 세계의 다양성을 확보하고 더 나아가 이를 통해 디지털에 더욱 의존하게 될 미래사회의 민주화를 보장하기 위한 것이다. 과학이 진정한 보편성을 확보할 수 있을 것인가 하는 문제가 디지털화를 통해서 다시 한 번 시험대에 올라온 셈이다.

1.2.2.2. 경제와 산업 부문의 젠더

전 세계적으로 남녀간 임금 격차를 보면 생산노동에서 평균적으로 남성이 1달러를 벌 때 여성은 77센트를 버는 것으로 나타난다. 이 격차는 남성과 여성이 비슷한 수준의 교육을 받았어도 크게 다르지 않다. 여성 전일 노동의 연간 임금은 2008년 기준 남성 임금의 77% 정도였다.¹⁷⁾ 남녀간 불균형한 취업과 승진 기회로 격차가 벌어지는 일이 가장 흔하지만, 동일노동 동일임금의 원칙 역시 제대로 준수되지 않고 있는 것이 현실이다.

농업이 발전한 아시아와 아프리카 대륙에서 각종 농사에 종사하는 인구는 절대수로 보자면 여성이 더 많다. 그러나 IMF 자료 상에서는 66%의 여성이 일하지만 공식 경제에서는 여성의 기여도가 25%로 줄어든다는 사실이 보여주듯이 농업 여성을 데이터로 파악하는 것 자체가 힘들다. 게다가 다수 여성노동력의 투입에 비해 여성의 농업용 토지 소유 비율은 전체의 13%에 불과하다. 더구나 토지가 아니라도 농업 생산의 직간접 이윤이 여성 생산자에게 돌아오는 몫은

14) <https://www.moe.gov.my/>

15) <http://ninastem.aprende.sep.gob.mx/>

16) <http://www.scientix.eu/news> ; <http://www.gender4stem-project.eu/>

17) Institute for Women's Policy Research 2010.

상당히 적다. 경제적으로 낙후된 지역일수록 여성이 가장인 가구가 많다는 점을 생각하면 통계에 잡히지 않는 여성 노동의 비중은 더욱 커진다.¹⁸⁾

이 문제를 빅데이터의 관점에서 접근한 국제적인 프로젝트가 진행되고 있다. 국제수자원협회(International Water Institute)가 여성에게 수자원에 대한 접근권을 주려는 조직들(Network of Women Water Professionals & Women for Water Partnership)과 협업하여 진행 중인 사업들이 그것이다. 그중 '4개 강 유역 젠더 프로파일(4 Basin Gender Profiles)' 프로젝트는 갠지스 강, 메콩 강, 나일 강, 볼타 강 유역의 수자원에 대한 접근권 및 분포도가 어떻게 인구, 영양 상태, 사망률, 위생 상태와 연결되는지, 이에 대한 빅데이터를 젠더에 따라 나누면 그림이 어떻게 달라지는지 보여준다. 이에 대한 인터랙티브 지도는 모두 웹에 공개되어 있다.¹⁹⁾ 이들의 연구목표를 들어보면 젠더 기반 데이터의 효용을 알 수 있다.

젠더 비형평성(gender inequity)은 수자원 관리와 농업에 넓게 퍼져 있지만, 이 분야의 젠더화된 측면에 대한 정확한 데이터는 거의 없다. (자원에 대한) 접근성과 의사결정에서 젠더 형평성을 향상하기 위하여 연구자, 정책 결정자, 개발 투자자들에게는 수자원과 농업 체제 내의 젠더화된 관계에 대한 더 명확한 이해와 더 나은 정보가 필요하다. 우리 웹사이트의 4개 강 유역 지도는 젠더 불균형에 대한 분석과 새로운 개발 기회를 위한 기반 증거를 제공하고자 한다.²⁰⁾

산업 부문의 고용 격차 역시 심각하다. 많은 경우 이 문제는 직업군 자체가 젠더화되어 고임금직과 관리직을 남성이 독차지했던 상황에 연유했으나, 남성 직군에 여성이 진출한 이후에도 젠더 불평등은 좀처럼 사라지지 않았다. 미국의 경우 남성 직군에 여성이 진출하던 추세가 계속 상승하다가 1990년대부터 둔화되었다. 미국의 통계에 의하면 2000년대가 여성 노동력 증대의 한계선처

18) <http://hdr.undp.org/en/content/table-4-gender-development-index>

19) <http://maps.vista-info.net/gis/htm/IWMIBasinMaps/index.html>.

20) <https://www.iwmi.cgiar.org/2018/11/indus-basin-gender-profile-mapper/> 인더스 강 유역 지도는 http://www.indusbasin.org/applications/indusbasingp/gender_map.php.

럼 보인다. 25세에서 54세 사이의 여성, 즉 주된 노동연령층 여성의 74%가 노동에 종사하게 되었는데 이 수치는 2000년부터 더는 늘지 않았다. 반면 남성은 같은 시기에 86%의 참여율을 보였다.²¹⁾ 데이터 수집이 가능한 한, 기업의 임원 직위에 오른 여성은 34퍼센트이다. 이 분야는 국가별 차이가 크다. 이집트, 사우디아라비아, 예멘 그리고 파키스탄에서는 7% 이내이고 바하마, 컬럼비아, 자메이카, 필리핀에서는 남녀가 완전 동수를 보이고 있다. 다른 19개국에서도 관리직 여성이 적어도 40%이다. 한국의 경우 기업의 임원직에서 여성이 차지하는 비율은 2.1%로서 전세계 최하위권이다.²²⁾

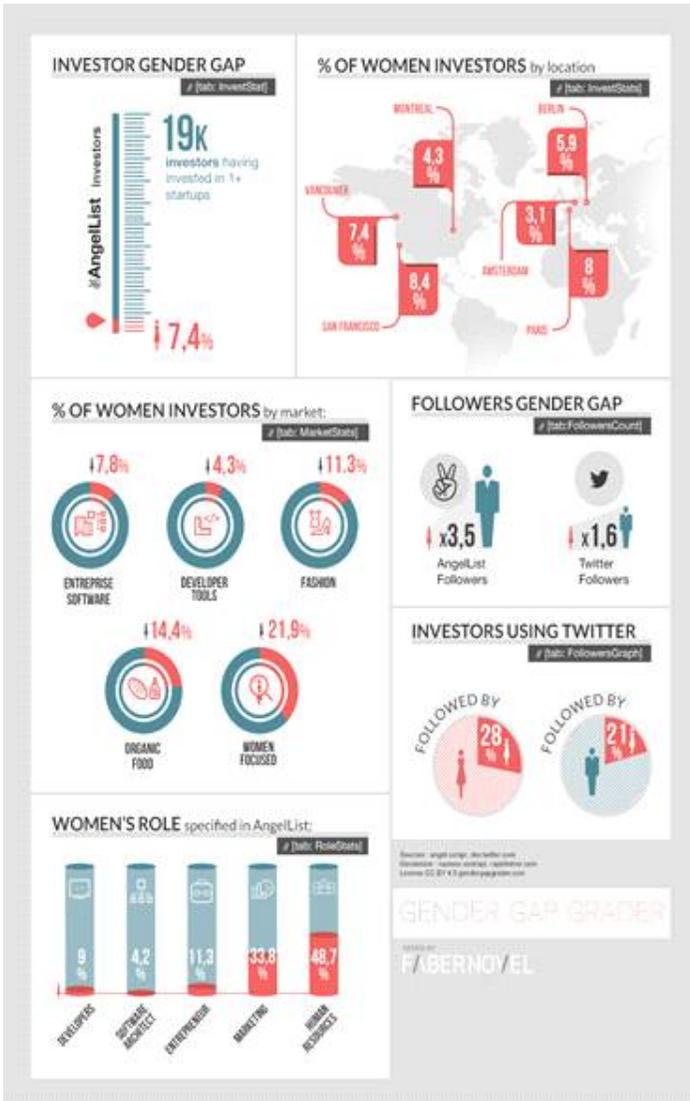
“Gender Gap Grader”라는 프로젝트는 여러 전문 분야 중 과학, 창업 투자, 항공업에서 젠더 불평등이 얼마나 심한지 데이터로 보여준다.²³⁾ 항공업의 경우 직군 자체가 젠더화된 곳으로 악명이 높는데, 이들이 전세계(북미 중심이지만)의 항공기 조종사들의 데이터베이스를 기반으로 밝혀낸 바에 따르면 민간 항공기 조종사 중 여성 비율은 5.44%에 불과하다.²⁴⁾ 또한 이들은 창업 투자(startup)에서 젠더 불균형이 어떻게 나타나는지 보기 위하여 창업자와 투자자를 연결해 주는 AngelList의 65만 명을 분석했다. 그 결과를 인포그래픽으로 나타낸 것이 아래의 그림이다.

21) U.S. Bureau of Labor Statistics 2009

22) <https://www.weforum.org/reports/gender-gap-2020-report-100-years-pay-equality>

23) <http://gendergapgrader.com/>

24) <http://gendergapgrader.com/studies/airline-pilots/>



<그림 1> <http://gendergapgrader.com/studies/angel-investing/>

이 프로젝트는 기존의 빅데이터에서 이와 같은 젠더 기반 빅데이터를 새롭게 만들어 냈으므로 젠더 불평등의 현상을 드러낼 뿐만 아니라, 이에 대한 의견을 교환하는 장을 창출하여 미래에 젠더 형평성을 높이고자 한다.

1.2.2.3. 교육과 젠더 지표들

교육은 여성 개인에 필요할 뿐 아니라 젠더 격차에 변화를 가져올 수 있는 문화적 환경을 조성한다는 점에서 젠더 불평등의 근본적인 해결책이다.²⁵⁾ 유네스코통계국(UNESCO Institute for Statistics)은 매년 전세계의 교육 현황과 이에 대한 젠더 문제를 담은 통계를 내보내고 있다. ‘유네스코의 교육의 젠더 불평등 문제를 담은 e-지도(UNESCO eAtlas of Gender Inequality in Education)’를 보면 매년 업데이트되는 교육에서의 젠더 불평등 문제를 한 눈에 볼 수 있다.²⁶⁾ 이 지도는 기존 데이터를 젠더로 해체(disaggregate)하는 과정을 통해 만들어진다. 전세계적으로 보자면 초등교육 수준의 젠더 평등은 놀라운 수준으로 진행되었으나 아프리카의 경우 아직 가장 취약한 지역으로 남아 있다. 고등교육의 경우 젠더 간의 불균형이 심화된 상태로 남아있다. 학사와 석사까지는 학위 취득자의 비율이 남녀 비슷하거나 여자가 더 많은 나라들이 많지만, 박사 이상의 전문 인력으로 가면 이 비율은 역전된다. 박사 학위 취득자의 54%가 남성이고 연구자의 71%가 남성이다.

더 심각한 것은 젠더별로 전문분야가 눈에 띄게 나뉘어 있는 것이다. 여성은 교육, 인문학과 예술, 사회과학, 비즈니스와 로스쿨, 건강과 복지 분야에 많이 진출하는 반면, 정보, 커뮤니케이션, 엔지니어링, 매뉴팩처링과 건설 등은 남성에게 독점되어 있다.²⁷⁾ 젠더 격차 Gender Gap Grader가 과학 관련 상호참조 사이트에 축적된 빅데이터를 분석한 결과는 전반적인 젠더 불균형에도 불구하고 몇몇 부문에서 최근 점차 완화되는 상황을 보여준다. 수집한 과학자 데이터 중 33%가 여성 과학자였으며, 많이 참조되는 과학 논문 저자 등 여성의 비율은 2001년 7%에서 2014년 13%로 늘어났다. 젠더 격차가 완화되는 추세는 사회과학에서 더 분명하게 나타났으며(2014년 31%), 컴퓨터 과학의 경우 가장 차이가 컸다(9%). 여성의 비율이 매우 낮던 엔지니어링이나 수학에서 빠른 성장

25) Ronald Inglehart and Pippa Norris, *Rising Tide: Gender Equality and Cultural Change Around the World*, Cambridge University Press, 2003, pp. 8-18.

26) <https://tellmaps.com/uis/gender/#!/topic/GENDER>

27) <https://tellmaps.com/uis/gender/#!/tellmap/79054752>

을 보인 반면, 물리학처럼 답보 상태에 머문 분야들도 있었다.²⁸⁾

교육에서 젠더 불평등은 기회의 균등을 보장하는 제도적 평등만으로는 해결될 수 없는 다양한 사회문화적 요인으로부터 배태되며, 이 문제들은 전통적 데이터로는 좀처럼 포착되지 않는다. 세계의 많은 연구소와 시민단체들은 교육에서의 젠더 격차를 만들어내는 다양하고 복합적인 요인을 잡아내고 이를 완화시키기 위해 리서치를 진행하고 있다. 하나의 예로 Data2X의 지원을 받은 “거리의 성희롱과 여성의 교육 진로 선택(Street Harassment and Women’s Educational Choices: A Geospatial Analysis)”이라는 프로젝트는 인도의 뉴델리를 대상으로 빅데이터를 수집하여 여성들의 교육 기회를 제한하는 보이지 않는 장애가 있음을 밝혀냈다.²⁹⁾ 이에 따르면 델리의 여성들은 상위 20% 대학에 갈 수 있는 점수로 하위 50% 대학을 선택하는 경우가 많은데 통학 과정에서 성희롱이나 성추행의 가능성을 낮추기 위해서이다. 여성들은 공공장소나 거리에서의 성희롱을 피해 통학에 더 많은 돈과 시간을 쓰며, 이 과정에서 더 안전하게 통학하기 위해 더 낮은 수준의 대학에 가는 희생을 감내한다. 이를 통해 언뜻 보면 평등해 보이는 남녀간 교육 기회를 불평등하게 제약하는 젠더 폭력의 모습을 포착하고, 이를 정책 결정 과정에 반영하는 것이 이 프로젝트의 목적이다.

1.2.2.4. 여성의 안전, 건강과 여성살해

최근의 빅데이터 통계에 따르면 매 2분 세계 어딘가에서 여성이 임신이나 출산 때문에 목숨을 잃고, 여성의 3분의 1은 일생에 한 번은 젠더 폭력(Gender-Based Violence: GBV)을 경험한다.³⁰⁾ 여성의 의료 자원 접근권에 대한 제약, 사생활 침해, 공공장소에서의 성희롱, 임신과 출산 문제, 직장과 가정 내 성폭력의 위험은 그대로 여성의 사회 활동을 위축시키는 결과를 가져와 사회경제적 젠더 격차로 직결된다. 여성의 안전과 건강은 젠더 평등을 위한 기본 조건이다.

28) <http://gendergapgrader.com/studies/gender-gap-in-science/>

29) https://data2x.org/wp-content/uploads/2019/09/StreetHarassmentWomensEdChoices_WBBrown.pdf

30) <https://www.equalmeasures2030.org/how-we-work/data-and-analysis/>

이 문제를 빅데이터 분석을 통해서 가시화시키는 것이 문제를 해결하는 출발점이 될 수 있다. 좋은 예로 많이 인용되는 것이 2010년 베트남의 사례이다. 베트남의 가정폭력 문제는 오랫동안 개인과 가족의 문제로 여겨져 함구되었다. 다양한 시민단체와 새로운 법들이 지속적으로 이 문제를 제기했음에도 불구하고 대중이 이 문제의 심각성을 이해하기 위해서는, 그리고 정책결정자들이 구체적 증거 위에 적극적 조치를 밀고 나가기 위해서는 견고한 사실이 필요했다. 2009년 베트남 통계국은 세계보건기구 및 여러 단체와 협업을 통해서 전국적인 가정 폭력 실태 조사를 처음으로 실시했고, 그 결과를 2010년 공개했다. 이에 따르면 베트남 여성들의 58%가 남편에 의해 물리적, 성적, 감정적 학대를 경험했고, 그 중 87%는 피해 사실을 숨기느라 어떠한 도움도 구하지 못했다. 이 통계는 가정폭력문제를 수면에 부상시켜 넓은 범위의 사회적 대화와 논쟁을 유발했고, 그 결과 가정폭력을 막기 위한 법적 조치와 피해 여성을 위한 구호 조직들이 생겨났다. 이 사례는 젠더 데이터가 정부의 의지와 여러 조직 간 파트너십과 결합하여 사회를 바꾸는 힘을 발휘한 예로 주변나라들에서도 많이 모방되었다.³¹⁾

하지만 많은 사회에서 가정폭력을 공론화시키는 데에는 여전히 위험부담이 크며, 이 때문에 데이터를 수집하는 것 자체가 힘들다. 모로코 마라케시에 있는 카디 아야드 대학(Cadi Ayyad University) 연구자들이 조직한 '여성에 대한 폭력을 고찰하기 위한 웹 마이닝(Mining the Web for Insights on Violence Against Women)' 프로젝트는 젠더 폭력에 대해 계속 변화하는 사회적 관습과 개인의 태도를 추적한다. 이에 관한 전통적 통계 데이터를 수집하기에는 비용과 시간이 너무 많이 들 뿐만 아니라 답변자에게 가해지는 사회적 압력 때문에 답변 자체가 왜곡될 가능성이 높다. 이 프로젝트에서는 유튜브와 트위터에서 일상적으로 벌어지는 SNS 활동에 대한 데이터를 수집하여 이 문제에 새롭게, 그리고 보다 저비용으로 접근하고자 했다. 여기서 이들은 젠더 관련 콘텐츠를 담은 게시물들과 이에 대한 구독자들의 반응을 분석하여 젠더 관련 중요 쟁점들이 아랍어를 쓰는 이용자들 사이에 어떠한 반응을 일으키는지 알아보려고 했다.

31) <https://data2x.org/wp-content/uploads/2020/02/Impact-Case-Studies-Vietnam-4P.pdf>

사이버 성폭력이 실제 성폭력과 높은 상관관계를 가지고 있음이 판명되고 있기 때문에 이러한 빅데이터는 앞으로 더 중요해질 것이다.³²⁾

더 나아가 일부 연구들은 가정뿐만 아니라 공공장소에서의 가시화되지 않는 젠더 폭력이 여성들의 사회활동을 저해함을 밝혀내고 있다. 뉴욕대의 GovLab에서 진행 중인 “칠레의 젠더와 도시 이동성(Gender and Urban Mobility) 연구는 여러 집단 간의 데이터 협업(data collaboratives)을 통해 도시 거버넌스에 젠더 문제를 결합시키기 위한 노력의 산물이다. 칠레 인구의 절대다수가 수도인 산티아고에 사는 가운데 산티아고의 교통 문제는 더 심각해 지고 있다. 그러나 당국이 생각하는 시민의 디폴트는 남성이고, 여성이 이동할 때 직면하는 특별한 문제들은 그다지 고려하지 않는다. 여성은 남성보다 도보와 대중교통을 이용하는 비율이 훨씬 높으며, 육아나 집안일과 관련된 일을 위해 여러 지역을 경유하는 일이 많다. 특히 문제가 되는 것은 85% 이상의 칠레 여성들이 거리에서 성희롱을 경험한다는 것이다. 이 프로젝트는 데이터 협업으로 통화 자료를 수집하여 만든 여러 맵핑 자료와 지도를 통해 보다 젠더 통합적인(gender inclusive) 도시 계획과 모델을 제공하고자 한다.³³⁾

특히 최근 전세계적으로 쟁점이 된 것은 ‘여성살해(femicide)’의 문제이다. 여성살해는 여성이라는 이유로 주로 남성 파트너나 가족 구성원에게 살해당하는 경우를 말한다. 여성살해야말로 다른 전통적 데이터들 속에 살인의 여러 모습으로 숨어 있다가 새로운 인식과 전통적 데이터의 젠더별 해체 과정을 통해 비로소 드러나게 된 예이다. 새롭게 만들어진 Global Femicide Watch는 전지구적으로 벌어지는 여성살해에 대한 데이터를 계속 수집하고 있다.³⁴⁾ 이로 인해 드러나는 통계는 그 자체로 충격적이다. 2017년 87,000명이 여성이라는 이유로 살해당했고, 그중 5만 명은 남성 가족이나 파트너에게 살해당했다. 환산하면 매 시간마다 6명의 여성이 가까운 이들에게 살해당하고 있는 것이다.³⁵⁾ 여성살해의 정도를 나라별로 알린 자료가 속속 나오고 있는데 특히 순위가 높

32) <https://data2x.org/wp-content/uploads/2019/10/MiningWebViolenceAgainstWomen.pdf>

33) <http://www.thegovlab.org/project-gender-and-urban-mobility-in-chile.html>

34) http://femicide-watch.org/search?f%5B0%5D=field_freehandtags%3A3545

35) <http://femicide-watch.org/sites/default/files/FemicideVolXI.pdf>

은 곳은 라틴 아메리카 지역이다.³⁶⁾ 이러한 상황은 유럽에서도 마찬가지라 2018년 프랑스에서는 여성살해에 반대하는 대대적인 캠페인이 벌어지기도 했다.³⁷⁾ 더 나아가 이탈리아에서는 그동안 함구되던 여성살해에 대한 경각심을 일깨우고 공론장에서의 논의를 활성화하기 위하여 여성살해에 대한 데이터를 담은 디지털 아카이브를 만들고 통계자료를 확충하고 있다.³⁸⁾

1.2.3. 젠더 문제와 빅데이터의 새로운 가능성

UN의 SDGs 2030 Agenda는 2030년까지 달성해야 할 목표로 젠더 평등을 지목했다. 그러나 14년째 발표되고 있는 Global Gender Gap Report의 올해 보고서에 따르면 우리는 살아있는 동안 젠더 평등을 보지 못할 것이다. 완만한 변화가 이루어지고 있지만(그리고 몇몇 나라의 경우 예전과 비교할 때 혁명적 변화가 일어나고 있지만) 지금의 추세를 따르자면 젠더 평등이 이루어지기까지는 99.5년이 걸릴 것이다.³⁹⁾ 현재까지의 데이터가 가져다주는 암담한 전망을 극복하려면 어떻게 해야 할까?

많은 데이터 전문가들은 우리가 빅데이터에 더 윤리적이고 책임감 있게 접근한다면 빅데이터를 통해 젠더 격차를 좁히는 데에 이바지할 수 있다고 말한다. 이를 위해서는 데이터 과학내 성별 격차를 좁혀 데이터 세계의 다양성을 늘리고, 젠더 감수성이 있는 데이터 체계(Gender Sensitive Data System)를 만들 필요가 있다. 또한 빅데이터를 통해 젠더 문제에 접근하는 연구자와 운동가들은 데이터 자체를 넘어서 문제를 보다 폭넓게, 그리고 하나의 프로세스로 조망해야 한다고 조언한다.

1. 먼저 새로운 질문을 제기할 필요가 있다. 창조적인 질문이 없으면 새로운 데이터도 수집할 수 없다. 젠더 문제야말로 관습에서 벗어나 보이지 않는

36) <https://www.balcanicaucaso.org/eng/Areas/Europe/Femicide-the-numbers-in-Europe-184329>

37) <https://information.tv5monde.com/video/femicides-combien-de-femmes-tuees-dans-le-monde>; <https://www.lejdd.fr>

38) <http://www.crit.rai.it/CritPortal/?notizia=1553&lang=en>

39) <https://www.weforum.org/reports/gender-gap-2020-report-100-years-pay-equality>

불평등을 포착할 수 있는 새로운 문제제기가 필요하다.⁴⁰⁾ 뉴욕대 GovLab의 "100가지 질문 이니셔티브(100 Questions Initiatives)"는 해당 분야의 전문가와 데이터 과학자 양쪽을 수십 명씩 모아 지속가능한 발전 및 빅데이터와 관련된 새로운 질문을 만들고 있다. 그중 하나가 바로 젠더 문제로서 이를 통해 다양한 분야의 전문가들이 모이는 포럼이 만들어졌다. 이에 따르면 빅데이터 자체로는 문제를 해결할 수 없으며, 더 새롭고 통찰력 있는 질문을 통해 데이터를 모으고 이용하는 방식을 혁신할 필요가 있다.⁴¹⁾ 이를 통해 문제의 제기부터 데이터수집, 데이터 활용, 정책 수립에 이르기까지 젠더 문제에 대해 통합적이고 전체적인 시각을 확보하는 것이 이 프로젝트들의 목표이다.

2. 더 나아가 빅데이터 혁명은 데이터와 사회 변화를 결합하고자 한다. 젠더와 관련하여 빅데이터 혁명이 추구하는 바는 "데이터가 효용이 있으려면 (useful), 행동을 불러일으켜야(actionable)" 한다는 것이다. 이미 힐러리 클린턴은 2012년 UN에서 열린 회합에서 "데이터는 진보를 측정하는 데 그치는 것이 아니라 이를 촉진한다(Data not only measures progress, it inspires it)."고 말한 바 있다.⁴²⁾ 위에서 제시한 젠더 관련 빅데이터들은 수치로 끝나는 것이 아니라 그 자체로 발언권을 가지고 사회를 바꾸는 힘을 발휘해 왔다. 이러한 측면에서 젠더 관련 빅데이터는 '데이터를 위한 데이터'가 아니라 사회 변화와 처음부터 끝까지 밀접하게 연관된 새로운 데이터 세계의 전망을 제시하고 있다.

3. 마지막으로 빅데이터와 디지털은 사회에 호소할 뿐만 아니라 여성들에게 직접 힘을 불어 넣어(empower) 이러한 변화를 촉진할 수 있다. 좋은 예로 Girl Effect라는 시민단체에서 개발한 TEGA(Technology Enabled Girl

40) <https://gender.the100questions.org/>

41) <https://gender.the100questions.org/>

42) <https://www.clintonfoundation.org/press-releases/clinton-foundation-and-gates-foundation-partner-measure-global-progress-women-and>

Ambassadors)라는 리서치 앱은 소외 계층 소녀들이 스스로 데이터 상의 젠더 격차를 메울 수 있게 돕는다.⁴³⁾ 이 앱과 관련 프로그램은 개발도상국 소녀들에게 디지털 기기를 공급하여 그들이 직접 공동체 내에서 인터뷰를 진행하고 데이터를 수집하도록 훈련시킨다. 이를 통해 전통적 방식으로는 수집하기 힘든 소녀와 여성들에 관한 데이터를 모을 뿐만 아니라 정책 결정자들이 보다 구체적으로 목표를 설정할 수 있도록 돕는다. 동시에 소녀들은 디지털 세계에 접속함으로써 자신들의 삶을 직접 선택할 뿐만 아니라 공동체 자체를 변화시키는 촉매가 될 수 있다. “젠더 평등과 빅데이터” 보고서 작성에 참여한 UN Global Pulse의 정책 담당의 말이야말로 빅데이터와 젠더 문제의 관계를 요약해 준다: “여성과 소녀들에게 힘을 실어주는 것은 지속가능한 발전 목표를 달성하는 가장 영리한 길 중 하나다. 여성과 소녀들이 힘을 가질 때 가족, 공동체, 사회가 번영하고, 경제가 성장하며, 발전은 보다 지속가능한 것이 된다. 우리는 여성에 대한 더 많은 데이터, 여성을 위한 더 많은 데이터를 필요로 한다.”⁴⁴⁾

2. 디지털 시대의 인문학

디지털-빅데이터 시대는 개인의 일상생활뿐만 아니라 학문세계에도 심대한 영향을 미칠 수밖에 없다. 디지털화가 사회 전 영역에 침투하는 현상에서 인문학이라고 예외일 수는 없는 것이다. 인문학 분야도 점차 디지털 기술을 연구에 포함시켜 새로운 연구 대상과 방법론을 만들어냈고, “디지털 인문학”, “eHumanities”, “컴퓨터 언어학”, “고고지리학” 등 이에 상응하는 교과목 및 교과과정을 만들어냈다.

역사학의 경우에 ‘사료의 디지털화’가 이루어지고 있다는 점에서 유럽에서

43) <https://www.girleffect.org/what-we-do/mobile-platforms/tega/>

44) <https://www.unglobalpulse.org/2018/03/new-gender-equality-and-big-data-report-takes-stock-of-innovation-efforts-to-close-gender-gaps/> 여기서 퍼낸 빅데이터와 젠더 평등 보고서는 https://www.unglobalpulse.org/wp-content/uploads/old_site/Gender-equality-and-big-data-en-2018.pdf

인쇄술이 활용된 14세기 이래 가장 근본적인 변화를 목도하고 있다고 해도 과언이 아닐 것이다. 가장 단순하게는 일반 문서를 컴퓨터가 읽을 수 있는 형태로 스캔하는 것을 포함하는 디지털화가 “디지털 턴”이라는 용어로 구체화되기에 이르렀지만, 특정한 논점에 대해 입증하거나 반박하는 것을 중시하는 전통적인 인문학은 이러한 변화에 대해 소극적이거나 부정적으로 대응해 왔다. 엄청난 규모로 이루어지고 있는 자료의 디지털화에 대해, 그리고 디지털 데이터에 대한 접근이 용이해진 점에 대해서는 경외감을 보이면서도, 이러한 연구들이 분석적인 측면에서의 기여가 분명하지 않다며 “디지털”, “디지털 인문학” 등에 대해 회의적인 시선을 보내고 있는 것이 인문학자들 대부분이 보이는 태도인 듯하다.

그러나 현재까지의 각국의 디지털 인문학 관련 논의의 범위와 깊이는 이 분야에 대해 학계 전체의 체계적인 공동대응이 필요함을 보여주고 있다. 먼저 1장에서는 디지털 인문학과 관련되는 여러 쟁점들을 짚어보게 될 것이다. 이 경우 디지털 인문학 연구가 가장 발전한 미국의 최신 연구 및 관련 논의들을 종합적으로 정리한 샤론 블럭(Sharon Block)의 논문을 주로 참고하였다. 이 글은 개인 저자인 블럭의 글이기보다 초기 미국사 연구의 최고 권위자인 “William and Mary Quarterly”가 2018년 10월 개최한 워크숍의 논의 결과들을 정리한 글이기 때문에, 이에 근거하여 미국의 인문학자들이 최근의 디지털인문학 연구 경향에 대해 어느 정도의 깊이와 범위로 논의를 전개하고 있는지를 파악할 수 있었다.⁴⁵⁾ 2장에서는 다시 디지털 인문학 연구의 측면에서 앞서가고 있는 미국과 영국사 분야의 연구 및 고고학 분야의 연구를 살펴봄으로써 그동안의 성과에 대해 가능해보고자 했다. 3장에서는 영국과 독일을 중심으로 하여 각국의 정부가 어떤 방식으로, 그리고 어느 정도로 디지털인문학 연구를 지원하고 있는지 살펴보고, 4장에서는 이에 근거한 종합을 시도하게 될 것이다.

45) Sharon Block, #DigEarlyAm: Reflections on Digital Humanities and Early American Studies, *William and Mary Quarterly*, 3d ser., 76, no. 4, October 2019 DOI: <https://doi.org/10.5309/willmaryquar.76.4.0611>

2.1. 디지털 인문학의 여러 쟁점들

■ 사료 디지털화의 장점

디지털-빅데이터 시대에 역사학과 관련되는 자료가 엄청난 규모로 디지털화 되는 현상은 그 자체만으로도 많은 장점을 갖고 있다고 볼 수 있다. 먼저, 이 자료들의 영구적인 보존이 가능해지게 된다. 가령 도서관이 불타거나 하는 사고 등을 통해 자료가 소실될 위험도 사라지게 된다. 실제로 2019년 남미 최대의 자연사 박물관이던 브라질 국립박물관 화재로 인해 아메리카에서 발견된 가장 오래된 구석기 시대 인간 두개골인 “루지아”를 비롯하여 소장품의 90%가 소실된 바 있다.

두 번째로, 도서관과 문서고의 소장 자료들이 디지털화되는 순간 언제 어디서나 활용가능하게 된다. 따라서 자료를 보기 위해 특정한 장소로 가지 않아도 좋고, 이는 자료들 역시도 이리저리 옮겨지지 않아도 좋다는 의미로서 자료 손상의 위험성도 줄어들게 되는 것이다.

세 번째로, 역사학 연구 분석 데이터의 규모가 커짐으로써 신뢰도를 훨씬 높일 수 있게 되었다. 특정 시기 신문 전체, 혹은 경찰 보고서 전체가 디지털화되어 학문 연구 목적에 활용될 수 있다면, 과거에는 답할 수 없었거나 추측만 하던 역사적인 의문들에 대한 답을 찾을 수 있게 될 것이다. 즉, 특정한 사회와 인간들의 삶 전체, 사고 구조 전체에 대한 조망이 가능해질 수 있게 되는 셈인 것이다.

■ 학문 생성 메커니즘의 변화: 학자와 대중의 경계 허물기

디지털인문학이 지식 생산에 있어 학자와 대중의 경계 허물기에 기여하게 되리라는 기대는 미국에서 특히 높은 듯하다. 유럽국가들에서 디지털 인문학은 전문연구자들을 중심으로 이루어지고 있지만, 미국의 경우 광범위한 대중 참여를 유도하는 것이 하나의 큰 흐름으로 자리잡았다. 따라서 미국 학계에서는 디지털 인문학의 근본적인 특징 가운데 하나로 가상 커뮤니케이션(Virtual communication)과 인터랙션(interaction)을 꼽고 있다. 그 결과 미국 학계에서

는 디지털 인문학을 “사회적인 기획”으로서 파악하는 가운데, 디지털 인문학이 지식이 생산되고 공유되고 정당화되는 메커니즘을 바꾸고 있다고까지 주장하기에 이르렀다.

대표적인 예로 미국 역사가인 로렐 울리히(Laurel Thatcher Ulrich)는 2000년 “Do History”라는 웹사이트를 통해 그가 “산파일기”라는 그의 책 저술과정에서 활용한 마서 벨라드(Martha Ballard)의 일기를 일반에게 공개하였다.⁴⁶⁾ 20년이 지난 현재까지도 이 웹 사이트는 여전히 널리 활용되고 있다. 이 사이트를 통해서 18세기 산파의 세계에 대한 정확한 정보를 알 수 있을 뿐만 아니라, 사용자들 자신이 역사적인 사료를 어떻게 다루는지를 배우게 되며, 한걸음 더 나아가 자기 자신 혹은 가족의 자료에 어떻게 접근하고 이를 근간으로 텍스트를 만들어낼 수 있는지를 배우게 되는 것이다.

그 외에도 역사학 연구 과정 자체를 대중에게 공개하는 유사한 프로젝트들이 다수 진행되고 있다. 대표적으로 미국사의 경우 “The Salem Witch Trials: Documentary Archive and Transcription Project”⁴⁷⁾, “The Geography of Slavery in Virginia”⁴⁸⁾를 꼽을 수 있고, 세계의 공개사진보관소를 만들고자 하는 시도인 “Flickr: Commons”⁴⁹⁾ 역시도 이에 속한다고 볼 수 있을 것이다.

한걸음 더 나아가 미국 역사학회의 경우 홈페이지에서 “디지털 히스토리의 수혜자가 되는 방법(How can I benefit from Digital History?)”라는 제목하에 해쉬태그(: #twitterstorians and #DH (digital humanities) 등 여러 통로들을 공개하고 있다.⁵⁰⁾ 역사학 연구 과정 자체가 대중에게 공개되는 이 현상이 인문학의 저변을 넓히는데 크게 기여할 가능성이 높아 보인다.

■ 디지털 인문학: 기존 위계질서의 전복 혹은 재생산?

이처럼 디지털 테크놀로지는 우리로 하여금 역사를 하는 방식, 우리가 역사

46) <http://dohistory.org/>

47) <http://saalem.lib.virginia.edu/home.html>

48) <http://www2.vcdh.virginia.edu/gos/>

49) <https://www.flickr.com/commons>

50) <https://www.historians.org/teaching-and-learning/digital-history-resources/resources-for-getting-started-in-digital-history>

를 기록하는 방식 등에 대해 재고하도록 하고 있다. 문서고에 어떤 문서를 소장할지, 어느 정도로 문서를 개방할지 등은 언제나 권력 관계를 반영할 수밖에 없었다. 그리고 사료가 디지털화됨으로써 문서고에의 접근에 대한 전면적인 변화가 일어나는 시점은 다시금 이 권력 관계를 고려하게 만드는 중요한 시점이기도 하다.

현재까지 인문학자들의 입장은 사료의 디지털화가 전통적인 계서제에 도전할 수도, 반대로 이를 강화할 수도 있다는 것이다. 가령 역사적으로 중요한 인물에 대한 디지털화된 문서들은 넘쳐나지만 지역 문서고가 소장하고 있는 역사적으로 주변화된 인물에 대한 자료의 경우 디지털화될 가능성이 낮은 것이 현실이다. 설령 주변적인 인물, 혹은 사건에 대한 자료가 디지털화되더라도, 예컨대 식민지 원주민들처럼 식민지 폭력의 오랜 희생자들의 경우에 대해서 디지털 상으로 사료를 접하게 되는 역사가들이 충분히 진지하게 고려하게 될지에 대한 우려, 즉 역사학이 역사적인 공간으로부터 멀어지고 “책상물림 학문(Desk Discipline)”으로 남게 되리라는 우려가 적지 않다.⁵¹⁾

디지털 인문학이 오랫동안 지속되어 온 성차별적이고 인종 차별적인 학문 생산 구조를 변화시키려고 하는 두 가지 구체적인 사례를 살펴보자. 먼저, “블랙 디지털 인문학(Black Digital Humanities project)”의 경우 “흑인의 시선(Black Perspective)”을 가시화시키려는 시도이다.⁵²⁾

“문서고를 탈식민화하기(Decolonizing the Archive)”라는 제목이 붙은 이 사이트는 20세기 이전 카리브해 관련 텍스트, 지도, 이미지 등을 한 곳에 모아둠으로써 플랜테이션 노예제의 실상, 더 나아가 근대 자본주의의 기원을 파악하도록 하는 데 큰 기여를 하고 있다. 관련 자료들을 아카이브(Archive), 클래스룸(Classroom), 전시(Exhibits) 등 세 개의 항목으로 나누어 정리함으로써, 교육 및 문화 전반에서 활용될 수 있는 방향성을 제시하고 있기도 하다.

두번째는 위키피디아에서 여성의 대표성을 높이기 위한 움직임이다. 위키피디아는 기고자의 측면에서나 주제상으로도 여성이 과소대표되고 있고 다양성

51) Lara Putnam, “The transnational and the text-searchable: Digitized sources and the shadows they cast”, *American Historical Review*, Vol. 121-2, 397

52) <https://ecda.northeastern.edu/>

이 부족하다는 지적을 받아왔다. 이에 따라 100여명의 미국 여성들을 위키피디어에 포함시키기 위한 “Mary Maples Dunn” 움직임이 벌어졌는가 하면, #CiteBlackWomen, #WomenAlsoKnowHistory, #Decolonize 등의 움직임이 나타나기도 했다.

결국, 디지털 역사학이 그 자체로 전복의 가능성을 내포하고 있기보다는 이러한 방향으로의 의식적인 노력이 필요하다는 의미로 해석되어야 하는 편이 정확할 것 같다. 이러한 시도들이 성공할지 여부는 결국 표면적으로 중립적인 “디지털화”가 자칫 기존의 권력 관계를 재편하기 보다 강화하는 방향으로 작용하게 될 위험성에 대해 인문학자들이 얼마나 진지하게 인식하느냐에 달려 있을 수밖에 없다.

■ 디지털 인문학: 하나의 분야 혹은 방법론?

디지털 인문학에 대한 논의가 엄청난 기대와 적대감을 동시에 불러일으키고 있는 현재의 상황이 전무후무한 예외적인 상황이라고 보기는 어렵다는 평가가 많다. 보다 좁게는 컴퓨터가 최초로 도입되고 계량경제사(Cliometrics)에 대한 역사가들의 희망이 고조되던 시기와 다를 바 없다는 견해도 한편에서 제기되고 있다. 결국 디지털 인문학이 과거의 양적 연구처럼 일시적인 유행에 불과한 것으로 판명되지 않겠는가 하는 회의적인 시선도 만만찮은 것이 현실인 것이다.

그러나 현재까지의 상황을 볼 때 일단 “양적연구”와 같은 일시적 유행에 그 치기는 어려워 보인다. 양적인 연구의 경우 다른 인문학자들에게 있어 전혀 별도의 영역으로 간주되었지만, 대부분의 학자들이 자신의 연구를 보다 효과적으로 수행하기 위해 이미 다양한 정도와 방식으로 디지털 기술을 활용하고 있는 것이 현실인 것이다. 학자들 사이에서 활용되고 있는 툴의 예를 몇 가지만 들어 보도록 하자.

분야	도구들
Open source tool	Zotero(서지사항 정리 프로그램) Omeka(온라인 디지털 프로젝트 만드는 프로그램) ⁵³⁾ Scripto(클라우드소스 transcription) ⁵⁴⁾ Tropy(리서치 포토 매니지먼트 프로그램) ⁵⁵⁾
시각화 툴	StoryMaps(지도, 이미지, 멀티미디어를 종합해서 스토리텔링을 가능케 함) Vectorwork(2-D, 3-D 디자인 능력) Graph Commons Tableau Public: 쉬운 시각화 기술 palladio: 역사 데이터의 시각화
텍스트 분석	Voyant: 사용자친화적인 텍스트 분석 도구 MALLET: 토픽모델링과 같은 자연언어 처리 도구

이 디지털 인문학의 여러 툴들은 지형에 대한 대규모 분석을 가능케 하고, 엄청난 분량의 텍스트에 대한 분석을 할 수 있으며, 공동체 구성원들 간의 관계를 그려내는 데 유리하고, 문헌들을 읽는 새로운 방식을 제공한다. 결국, 디지털 인문학은 하나의 도구이며 전통적인 역사학을 대체하기보다는 풍부하게 한다는 판단이 유력해 보인다. 전통적인 아날로그적인 토론에서 제기되는 이론적이고 내용에 근거한 연구에 디지털 인문학의 방법론적인 개입이 있을 경우에 시너지 효과를 낼 수 있을 것이다.

그러나 디지털 인문학 툴의 리스트는 현재도 충분히 길고, 앞으로 더 길어질 수밖에 없다. 그리하여 성향상 기술을 가까이 하지 않기 일쑤인 대다수의 인문학자들에게 그림의 떡처럼 느껴지는 것도 사실이다. 결국 이러한 프로그램들을 쫓아가기만도 벅찬 상태가 아닌지 자괴감에 휩싸이게 되기도 한다. 21세기의 인문학자는 이제 전통적인 인문학적인 연구를 수행해야 할 뿐만 아니라 디지털 인문학 프로젝트를 생산해내기 위해서 “슈퍼-인문학자”가 되어야 한다는 자조

53) <https://omeka.org/>

54) <https://scripto.org/>

55) <https://tropy.org/>

의 목소리마저도 나오고 있는 실정이다.

이러한 문제적 상황을 극복하기 위해서는 디지털적인 방식을 학부생 및 석사 과정 프로그램에 넣어 개인적인 트레이닝을 할 수도 있겠으나, 보다 더 큰 학문 생태계를 만들어내는 방안도 고려해볼 수 있을 것이다. 즉, 컴퓨터 공학 분야 등과의 학제간 연계를 제도적으로 활성화하고, 도서관 등 관련 기관에서 정기적으로 교육 프로그램을 제공하는 현재의 방식을 보다 확대해갈 필요가 절실한 듯하다.

■ 디지털 인문학 연구의 특수성

한편으로 디지털 인문학의 툴들을 개인 인문학자들이 필요에 따라 다양한 방식으로 활용함으로써 디지털인문학적인 방법이 보다 보편적으로 활용되기도 하겠지만, 다른 한편으로 디지털 인문학 분야가 가지는 특수성에 대한 고려도 필요할 수밖에 없다. 앞서 언급된 2018년에 열렸던 초기 미국사 분야 워크숍에서 지적된 디지털 인문학 연구와 관련된 문제점들을 표로 만들면 다음과 같다.

	내용
평가자의 문제	디지털 인문학의 연구 성과를 평가하기 위해서는 평가자가 내용뿐만 아니라 활용 도구에 대한 방법론적인 전문성도 가져야 한다는 현실적인 문제점도 내포하고 있다.
소스 공개의 문제	관련 디지털 데이터들, 코드들 등 역시도 공개되어야 하겠으며, 인문학자들은 전통적으로 1차 사료를 공유하지 않는 연구 관행을 보여왔다. 데이터 수집 자체가 역사학적인 연구의 과정이자 성과로 간주되어 왔던 것이다.
전거, 방법론 표기의 문제	디지털인문학 연구 결과의 경우 기존의 연구서들에서 사용되는 전거표기방법인 각주와는 다른 방식의 설명이 필요하다. 현재까지 디지털인문학 연구의 경우 사용된 방법론 부분들이 “내러티브”를 침해하지 않고 자 축소되는 경향성을 보여 왔다. 전통적인 역사 서술에서도 연구에 활용된 자료에 대해 인용을 해왔지만, 이는 연구 과정을 공개하는 것과는 전혀 다른 일이라고 볼 수 있다. 구체적인 예를 따지자면, 원본 사료를 본 경우, 원본의 온라인 이미지를 본 경우 등 사료에 대한 접근 통로도 매우 다양해진 것이 현실이다.

	내용
평가 기준의 문제	인문학자들의 세계에 있어 단독저자가 쓴 단행본이 금본위제 하의 금으로 간주되고 있는 것이 현실이다. 향후 디지털인문학 연구 과제를 평가하기 위해서는, 전통적인 학자 평가 방식과는 전혀 다른 접근법이 개발되어야 한다.
연구 결과 보존의 문제	디지털인문학적인 연구 성과를 지속적으로 유지보존하는 것도 매우 시급하고 어려운 과제이다. “page not found”, “404 Error”라는 메시지를 접한 것은 우리 모두에게 너무 흔한 일이다. 따라서 디지털적인 전거들, 그리고 디지털 연구 성과물들을 보존하는 것이 향후의 과제로 남겨지는 것이다. 미래에도 보존할 수 있도록 웹페이지를 보관하기 위한 사이트로는 Internet Archive(archive.org/about/)에서 “Save page now”를 하기 위한 “archive.org/create/”, Archive-it.org을 들 수 있다.

이 가운데 가장 시급한 것은 디지털 인문학 연구 결과물의 평가 방식을 개발하는 문제인 듯하다. 디지털 인문학에 대한 사회적 수요는 높으나, 이를 전문적으로 연구한 전문가는 매우 드문 것이 한국의 현실이다. 이러한 불일치는 자칫 디지털 인문학에 초기에 입문했다는 이유로 불성실한 연구자가 노력 이상으로 평가받거나 전체적인 디지털 인문학 연구를 오도할 위험성을 내포한다. 혹은 반대로 디지털 인문학 분야 연구에 대한 평가 기준이 미비하고 평가가 가능한 인원이 극히 희소한 나머지 디지털 인문학분야 연구자들이 그 노력이상으로 저평가될 위험성도 동시에 존재한다.

어느 편이건 평가 기준의 미비는 디지털 인문학 분야 전반에 대한 학계의 주저와 우려의 근본적인 원인이 되고 있기도 하다. 따라서 다른 분야와 마찬가지로 엄정한 평가 기준을 마련하고, 이를 평가할 수 있는 전문가 군을 갖추는 것이 매우 시급한 상황인 듯하다.

2.2. 디지털 인문학 연구의 성과들

2.2.1. 미국사 분야

미국이 디지털 인문학 분야에서 가장 앞서가고 있는 나라이니 만큼, 미국의 디지털인문학 연구 성과들은 향후 디지털 인문학 연구의 방향성과 깊이를 가늠

하는 데 있어 매우 유용한 전거가 된다고 볼 수 있을 것이다. 대표적인 미국의 디지털 인문학 연구 결과물들을 표로 만들 경우 다음과 같다.

연구과제명	연구내용
University of Georgia eHistory project ⁵⁶⁾	<p>아메리카의 침공: 미국이 어떻게 세계의 1/8을 차지하게 되었는가 “Invasion of America: How the United States took over an Eighth of the World</p> <p>1900년 센서스에 기록된 237,000명의 원주민들의 위치를 기록하기 위한 프로젝트로 시작되었다. 현재는 이 주제를 연구하는 여러 학자들이 공동으로 연구하는 플랫폼으로도 활용되고 있으며, 이에 근거하여 Invasion of America, CSI:Dixie, Private Voices, Mapping Occupation, USNewsMap.com, Of Methods and Madness, Pox Americana, Fugitive Federals 등 여러 흥미로운 프로젝트가 완성되었다. 그 외에도 “Mapping the American Population Project”, “Freedmen’s Teacher Project”, “Arresting Inequality” 등이 현재 개발중인 프로젝트이다.</p> <p>예컨대 Invasion of America⁵⁷⁾의 경우 1776년부터 현재에 이르기까지 미국이 조약과 행정명령을 통해서 거주가능한 토지의 1/8 정도를 원주민으로부터 획득하는 과정과 관련된다. 이 과정을 interactive map으로 보여줌으로서 그동안 충분히 드러나지 않았던 미국 원주민 수탈의 역사를 극적으로 보여주고 있다. 이 프로젝트는 백악관이 주관한 캠페인인 “ConnectED”에 포함되고, 무수한 언론의 주목을 받을 수 있었다.</p> <p>이 웹사이트에서 미국 지도의 한 부분을 누르면 누가 그 땅을 언제 양도했는지, 그리고 어떤 조약 및 행정명령이 이를 정당화했는지를 보여주고 있다.</p> <p>SmallPox Epidemic⁵⁸⁾ 역시도 마찬가지로 인상적인 프로젝트이다. 1775년부터 1782년까지 7년간 지속된 천연두는 북아메리카 대륙 전체를 휩쓸었다. 치사율이 30%로 알려진 이 병의 확산 과정을 밝히고자 진행된 프로젝트가 “SmallPox Epidemic”이었다. 과제를 진행한 Elisabeth Fenn은 컴퓨터의 경우 우리의 사고방식을 바꾸지는 않지만, GIS의 경우 새로운 질문을 제시하도록 했고, 당대인들이 대륙 내에서 서로 어떻게 상호작용했는지에 대한 통찰력을 얻게 했다고 고백했다. 디지털 인문학과 관련하여 매우 유의미한 논평이 아닐 수 없다.</p>
Emory University project ⁵⁹⁾	<p>유괴된 천만 이상 아프리카 노예들의 행로에 대한 시각화를 통해서 “The Atlantic Slave Trade in Two minutes”라는 Youtube 영상을 제공하였다.⁶⁰⁾</p> <p>이 프로젝트는 9만명 이상의 노예에게 아이디를 부여하였는데, 개별 노예들의 신장, 성, 출신, 승선한 배의 출발지와 도착지를 모두 파악할 수 있다. 또한 관련 기록의 원문, 관련된 장소, 노예들</p>

연구과제명	연구내용
	<p>의 초상을 포함하는 이미지, 배의 모형 등을 모두 시각화하였다. 관련되는 모든 역사적인 상황 및 사건에 대한 상세한 설명들도 담고 있어 미국 노예 무역의 양태에 대한 전체적인 조망이 가능하다. 노예무역과 관련되는 온라인 박물관으로 생각할 수 있을법하다.</p>
Stanford University ⁶¹⁾	<p>Mapping the Republic of Letters 역사적 엘리트들의 커뮤니케이션 네트워크. 학술대회, 협회 등이 존재하기 이전 학자들의 세계는 개인적인 네트워크에 의존할 수 밖에 없었다. 이 네트워크가 어떤 방식으로 작동했으며, 어떻게 진화하였는지 등의 의문에 답함으로써, 국제적인 학문 공동체의 생성 및 작동방식을 매핑을 통해 드러내 보여주고 있다. 구체적으로는 콩도르세, 달랑베르, 볼테르 등 대표적인 학자들 간의 서신, 이들의 여행 경로, 이들의 출판 상황 등을 통해 이들 지식 엘리트들 간의 교류를 잘 보여주고 있다.</p>
UCLA ⁶²⁾	<p>Mapping Indigenous LA: LA는 미국의 어느 도시보다도 토착민의 비율이 높은 도시이지만, 이들의 역사는 충분히 조명받지 못했다. LA에 거주했던 Gabrieleno/Tongva, Tataviam 부족의 민화, 음악, 원주민들과의 인터뷰 등이 소개되는 등 온라인 박물관으로서도 손색이 없는 듯하다.</p>

이 연구의 결과들을 하나하나 클릭해서 살펴보는 것은 매우 흥미로운 경험이다. 1차 사료의 시각적인 이미지를 가지고 작업하거나 혹은 텍스트 자료들을 시각화한 이러한 연구과제들은 역사적 주제, 공간, 혹은 역사적 문헌들끼리의 연결 관계, 혹은 장기적인 변화의 경향성 등을 보여주는데 매우 유용하다. 그러나 이처럼 손에 잡힐 듯 드러나게 되는 사실 자체가 역사적으로 의미있는 “테제”가 되지 못한다는 것 역시도 분명하다. 포괄적인 데이터가 시각화를 통해 손에 잡힐 듯 드러내주는 역사적 현실은 다시금 역사가를 통해서 해석될 필요가 있는 것이다.

56) <https://ehistory.org/about>

57) <http://usg.maps.arcgis.com/apps/webappviewer/index.html?id=eb6ca76e008543a89349ff2517db47e6>

58) <http://usg.maps.arcgis.com/apps/webappviewer/index.html?id=2c94e981a6c6485ab0e26bb680e0b063>

59) <https://www.slavevoyages.org/>

60) http://www.slate.com/articles/life/the_history_of_american_slavery/2015/06/animated_interactive_of_the_history_of_the_atlantic_slave_trade.html

61) <http://republicofletters.stanford.edu/>

62) <https://mila.ss.ucla.edu/>

이러한 해석을 위해 시각화가 기여하는 바는 개별 사료들만을 통해서만 드러나지 않던 연관관계들에 대해 명징하게 인식하도록 하고, 그러한 인식에 근거하여 텍스트에 근거한 연구로는 던질 수 없었던 “새로운 질문”을 던지는 것을 가능케 한다는 점일 듯하다. 결국 디지털 역사학과 전통적인 역사학은 상보적인 관계에 있는 것처럼 보이며, 디지털 인문학이 새로운 아젠다를 제시하지 않는다는 비판은 그런 점에서 논지를 벗어난 지적인 듯하다. 결국 디지털 인문학은 역사학자들의 사고를 날카롭게 해줄 수 있는 유용한 툴인 듯하기 때문이다.

가령 자메이카의 노예 반란을 연구하여 이를 시각화한 빈센트 브라운(Vincent Brown)이 이 시각화를 통해서 반란군들과 진압군 모두의 전략을 알게 되었고, 반란의 경로에 지형이 가지는 중요성을 파악하는 데 있어 지도가 매우 유용한 도구였다고 언급한 것은 대표적인 예이다.⁶³⁾ 그에 따르면 역사적 기록을 작성한 식민주의자들과 제국의 관리들은 반란에 대해 공감하지 않았고, 따라서 기본적으로 반란군에 대해서라면 부정적인 서술로 일관할 수밖에 없었다. 반면, 반란군은 문서를 생산하지도, 생산할 수도 없었다. 그러나 맵핑을 통해 시간이 경과함에 따라 반군이 어느 위치에 있는지를 추적함으로써 노예 반란의 전술적 역학을 관찰할 수 있고, 반란군의 의중을 짐작할 수 있게 한다는 것이 저자인 빈센트 브라운의 설명이다. 예컨대 반군은 독자적인 교역을 원했기 때문에 반란이 진행되는 가운데서도 수출 생산에 맞춘 농장을 유지하고자 했음을 알 수 있다는 것이다.⁶⁴⁾ 이는 매우 흥미롭고도 몽클한 결론이라 하지 않을 수 없으며, 그의 연구는 이처럼 흥미있는 여러 관찰결과들로 가득차 있다. 이처럼 역사적 사실에 지리 정보가 더해지게 됨으로써 당대인들이 느꼈을 위기, 희망, 가능성에 대해 더 입체적으로 접근할 수 있게 된다는 점에 대해서는 논의의 여지가 없어 보인다.

이 디지털 인문학 연구 결과물을 종합해 볼 때, 다음과 같은 지점에 주목할 필요가 있다고 판단된다. 먼저, 기존의 사료가 불충분하게 보여주고 가리던 내용을 시각화해서 보여줌으로써 당대의 역사적 현실에 훨씬 근접할 수 있도록

63) <http://revolt.axismaps.com/>

64) <http://revolt.axismaps.com/project.html>

돕는다는 결론에 이르지 않을 수 없다. 미국 원주민들이 어떻게 자신들의 땅에서 손님이 되어갔는지의 과정을 시각화를 통해서 문자 언어로 표현할 수 있는 것보다 훨씬 더 적나라하게 드러낼 수 있는가하면, 전염병의 전파 과정을 시뮬레이션처럼 보여줌으로써 역사가로 하여금 훨씬 더 당대의 현실에 근접한 질문을 던질 수 있도록 해주고 있는 것이다.

다음으로 인문학자적인 작업의 본질에 대한 성찰이 필요한 듯하다. 고독한 사색을 인문학의 근간으로 여겨온 것이 인문학자들의 세계이니만큼 협업은 여전히 낯선 작업 방식이고, 학자들 개인이 수집한 자료를 공유하는 일은 매우 드물다. 그러나 현재까지 살펴본 의미있는 디지털 인문학 연구 성과들은 대체로 협업이었고, 학자들이 자신들이 모은 자료들을 공개함으로써 그 연구 결과가 더욱 풍요로워질 수 있었다. 조지아 대학(Georgia University)의 “eHistory project” 역시도 클라우디오 산트(Claudio Saunt)와 스티븐 베리(Stephen Berry)라는 두 학자가 자신들의 연구 기록을 공개함으로써 시작되었고, 이후 다른 학자들도 자신들이 저술 과정에서 수집한 자료들을 공개함으로써 후속 연구를 촉발시키고 있음을 볼 수 있다. 이처럼 역사학적인 연구의 규모와 방식이 달라지는 현실은 역사학자들로 하여금 역사학의 본질에 대해, 그리고 역사학자들이 사회에 대해 저야할 책무에 대해 다시금 성찰하게 하고 있는 듯하다.

2.2.2. 영국사 분야

영국 또한 지난 20여년 간 디지털 웹 아카이브 구축을 중심으로 디지털 프로젝트들이 진행되어 왔다. 영국에서 대표적인 성과로 손꼽히는 사례들과 최근의 프로젝트들을 살펴봄으로써 대체적인 경향을 가늠해 볼 수 있을 것이다.

프로젝트	프로젝트 내용
The Proceedings of the Old Bailey, 1674-1913 (by Institute of Historical Research)	2003년에 시작된 올드 베일리 디지털 프로젝트는 1674년부터 300여년에 걸친 런던 사법재판소 기록을 디지털화하였다. 이 데이터베이스는 197,000건의 재판 기록과 타이번에서 처형된 약 2500명의 사람들의 인적 정보를 담고 있다. 디지털화된 신문 자료 및 잉글랜드/웨일스의 인구조사 자료를 함께 참조하면 범

프로젝트	프로젝트 내용
<p>홈페이지: https://www.oldbaileyonline.org</p>	<p>죄인과 사법 관행 등 범죄와 관련된 역사를 세세하게 구성할 수 있다. 더욱이 이 데이터베이스는 무료로 이용 가능하다. 이 데이터베이스를 기반으로 한 디지털 역사학 프로젝트들 역시 다수 출현하였다. ‘디지털 파놉티콘(Digital panopticon)’, ‘런던의 삶(London’s Lives)’, ‘런던의 과거 지도에 옮기기(Locating London’s Past)’ 등이 있다.</p>
<p>Digitized First World War personnel records (National Archives)</p> <p>홈페이지: https://www.nationalarchives.gov.uk/first-world-war/centenary-digitised-records/</p>	<p>남아프리카를 제외한 영국 및 영국의 백인 자치령 출신 제1차 세계대전 참전용사들에 대한 데이터이다. 기본적으로 당시 사람들의 세세한 인적 정보이기 때문에 군사사 자료로뿐만 아니라 아닌 다른 목적을 위한 자료로도 쓰일 수 있다. 예를 들면 다양한 지역에서의 생활 수준을 추적하기 위한 증거로 쓰일 수 있다. 다른 시기와 지역의 군인들 신장을 비교함으로써 사람들의 영양 상태를 추적할 수 있다. 호주인의 체격은 점점 작아졌고 캐나다의 경우에는 그 감소 폭이 더 컸다. 즉 도시화와 경제적 불평등이 영연방에서 영국에서 그랬던 것과 같은 건강상의 악영향을 가져왔음을, 그러나 그 정도는 영국보다 덜 했음을 알 수 있다. 또한 캐나다로 이주한 스코틀랜드인이 평균신장이 더 컸음을 발견하여 영국에 남은 사람이 북아메리카로 이주한 사람들보다 더 작고 허약했으며 더 건강한 사람들이 이주했음을 결론내릴 수 있다. 이 패턴은 다른 자료들로부터도 확인 가능하며, 영국에서 영연방으로 떠나는 이주는 장기적으로 영국의 경제 성장을 저해했다는 결론에 이를 수 있는 것이다.⁶⁵⁾</p>
<p>Political Meetings Mapper (by Katrina Navicka)</p>	<p>2015년에 영국도서관 랩의 디지털 역사학 연구 우승작으로 1789-1848년에 잉글랜드 북부에서 일어난 대중 운동을 공간적으로 시각화한 연구이다. 이 시기 민주주의와 노동자의 권리를 요구하는 대중 집회와 회합이 많이 일어났는데 이 프로젝트를 수행한 연구자 카트리나 나비카는 정치 집회는 공간과 장소의 문제임을 관찰했다. 신문 자료가 주 데이터 소스로 쓰였다. 신문에 나온 회합 시기와 장소를 추적하여 자료를 대규모로 축적했다. 이전에 역사가들은 일일이 신문을 뒤져 몇 차례의 정치적 회합 장소만을 찾아낼 수 있었을 뿐이었다. 나비카는 Political Meeting Mapper라는 툴을 만들어 역사 신문에서 회합에 대한 공지를 추출하고 그 데이터를 영국 도서관이 소장한 역사 지도에 층층이 배치하였다. 1830-40년대의 핵심적 시기에 정치적 이벤트의 장소들을 시각화하였다.⁶⁶⁾</p>

프로젝트	프로젝트 내용
<p>Black Abolitionist Performances and Their Presence in Britain (by Hannah-Rose Murray)</p>	<p>위의 연구에서 자극을 받아 비슷한 디지털 지도를 만들어낸 연구로, 흑인 노예제 폐지론자들의 연설 장소를 주제로 삼았다. 이 프로젝트를 수행한 머레이는 나비카보다 훨씬 더 많은 데이터셋을 사용했다. 1830년부터 1900년까지 영국에 온 흑인 노예제 폐지론자들은 대도시뿐만 아니라 작은 마을에서도 연설과 강의를 했다. 이들에 대한 소식은 전국지와 지방지를 망라하여 신문에 보도되었고 이것이 데이터셋이 되었다. 데이터셋의 규모 자체가 매우 방대하다는 것과 19세기 영국 신문에 대한 OCR 기술이 불완전하다는 것이 이 프로젝트가 해결해야 할 문제였다. 잘못된 결과가 나오기도 하고 필요한 정보를 놓치기도 했다. 따라서 OCR을 ‘클린’하게 하는 것, 그리고 컴퓨터로 하여금 무엇이 ‘긍정적인’ 텍스트이고 무엇이 ‘부정적인’ 정보인지 가르치는 것을 해결해야 했다. 머레이는 몇 년 동안 흑인 노예제폐지론자들의 연설을 컴퓨터상으로 옮겨야 했고 이 텍스트 데이터가 포지티브 텍스트로서 기능하게 되었다. 특정 텍스트의 출현 빈도와 비중을 측정하고 정확성을 높이기 위해 노력해야 했다.⁶⁷⁾</p>
<p>Content Analysis of 150 years of British Periodicals (by Think Big project of Bristol University)</p>	<p>브리스톨 대학에서 인공지능을 연구하는 넬로 크리스티아니(Nello Cristianini) 교수가 이끄는 연구팀은 Think Big이라는 프로젝트를 통하여 빅데이터의 사회적 적용에 관하여 여러 하위 프로젝트를 수행했다. 그 중 지난 150년 동안의 영국 역사에 관한 프로젝트가 브리스톨 대학 역사학과와 영국도서관의 신문 디지털화 작업을 해온 민간업체와 협력하여 수행한 빅데이터를 적용한 역사학 연구를 잘 보여준다. 연구팀은 1800년부터 1950년까지 영국에서 발행된 신문들로부터 3,500만 개 기사(해당 시기에 간행된 영국 신문의 14퍼센트에 해당)를 영국도서관이 소장하고 있는 디지털 신문 데이터베이스로 접근했다. 이들을 인공지능으로 분석하여 지금까지의 전통적 역사 연구가 수행하지 못한 장기간의 경향, 변화, 핵심 시기 등을 파악하는 데 중점을 두었다. 예를 들어 영국에서 언제 전기가 증기를 대체했는지, 언제부터 기차가 마차보다 대중교통수단으로서 더 많이 이용되었는지, 자유당과 보수당 등 정치 정당이 신문에서 각각 얼마나 더 많이 언급되는지, 여성의 정치적 대표성이 언제부터 증가했고 남성과 비교하여 어느 정도로 의회에서 덜 대표되는지 등이다.⁶⁸⁾</p>

65) Richard A. Hawkins, “Chapter 7. Use of Big Data in Historical Research,” in *Big Data in the Arts and Humanities: Theory and Practice*, eds. Giovanni Schiuma and Daniela Carlucci (CRC Press, 2018), pp. 81-82.

영국에서 본격적으로 실물 사료를 디지털로 옮기기 시작한 것은 거의 20년 전이다. 초기부터 실물 사료를 디지털화하는 데에 집중해 온 결과, 위에서 예들 든 ‘올드 베일리’ 사이트 뿐만 아니라 다수의 디지털화된 사료 데이터베이스를 구축하게 되었다. 19세기 이래 영국 의회 논의록을 디지털화한 ‘Hansard 1803–2005’, 신문 더타임스(The Times)를 1785년판부터 제공하는 ‘더타임스 아카이브(The Time Archive)’, 영국도서관이 소장하고 있는 방대한 신문 사료를 웹상으로 옮긴 ‘영국 신문 아카이브(British Newspaper Archive)’, 16세기 이래 환자 진찰 기록을 디지털화한 ‘케이스북(casebooks)’ 등이 널리 쓰이고 있는 대표적 웹 데이터베이스라 할 수 있다.⁶⁹⁾

최근에는 사료 디지털화가 더욱 광범위하게 이루어지고 있는 동시에, 이미 구축된 디지털 데이터를 이용하여 역사적 물음에 대한 답을 구하고, 그것을 시각적으로 표현하는 디지털 역사학 프로젝트들이 진행되고 있다. 올드 베일리 데이터베이스로 인해 디지털상으로 이용가능해진 역사적 데이터가 다수의 디지털 역사학 프로젝트로 이어졌듯이, 최근의 디지털 역사학은 이미 빅데이터가 된 디지털 자료를 전통적인 역사학 연구가 할 수 없었던 방식으로 다루려 한다. 2013년에 출범한 영국도서관 디지털 랩은 이러한 경향을 장려하면서, 매년 우수한 프로젝트에 대해 시상해오고 있다.⁷⁰⁾

위 표에서 제시한 예들 중, 마지막 프로젝트인 ‘150년 동안의 영국 정기간행물 분석’을 좀 더 자세히 살펴봄으로써 빅데이터를 이용한 역사학 연구의 가능성과 역사학자의 역할에 대해 생각해 볼 수 있다. 먼저 이 연구는 빅데이터의 구체적 가능성을 제시하였다. 프로젝트 팀은 이 연구의 목적이 “방대한 뉴스 말뭉치들을 ‘멀리서 읽기’에 토대한 방법으로 분석하여 역사에서의 지속과 변화를 이해하며, 전통적인 역사학의 가까이 읽기를 보충”하는 것이라고 밝혔다.⁷¹⁾

66) Ibid., p. 79.

67) Ibid., pp. 79–80.

68) <https://thinkbig.enm.bris.ac.uk/projects/content-analysis-of-british-periodicals/>

69) ‘올드 베일리’ 이후 언급된 데이터베이스들의 순서대로 각각의 웹사이트 주소는 다음과 같다. <https://api.parliament.uk/historic-hansard/index.html>; <https://www.thetimes.co.uk/archive/>; <https://www.britishnewspaperarchive.co.uk>; <https://casebooks.lib.cam.ac.uk>

70) 다음 웹사이트에서 영국도서관 디지털 랩의 목적, 사업 내용과 수상 후보와 내용 등에 대해 확인할 수 있다. <https://www.bl.uk/projects/british-library-labs>

프로젝트의 시간적 범위인 150년은 한 명의 역사가자가 다루기에 너무나 긴 시간이다. 또한 그 안에서 정치, 경제, 사회적 규범과 가치, 특정한 이벤트, 문화 현상 등에 이르기까지 거의 삶의 전반적인 측면에서의 지속성과 변화를, 해당 시기 동안 발행된 신문 전체의 14%를 분석해서 도출해 낸다는 목표는, 전통적인 방법에 따르면 한 명의 연구자가 할 수 없음은 물론이고 웬만한 협업으로도 도달하기 힘든 지점이다. 프로젝트 결과 19세기 이래 영국인의 삶의 다양한 측면에 대해 다양한 발견이 이루어졌다. 연구 결과와 방법론에 대한 자세한 소개는 온라인상에서 논문으로 확인 가능하다.⁷²⁾

- 정치: 19세기 동안에 자유당의 글래드스톤(Gladstone)이 보수당의 디즈레일리(Disraeli)보다 신문상에서 훨씬 더 많이 언급되었음이 밝혀졌다. 각 정당에 대한 신문의 언급은 비슷한 정도로 이루어졌으나 디즈레일리의 수상 재임 기간에도 글래드스톤이 신문 지상에서 훨씬 더 많은 주목을 받았음은 새롭게 밝혀진 사실이다. 노동당의 등장 이후, 특히 노동당의 첫 소수 정부 구성 시기인 1924년을 즈음하여 자유당과 보수당이라는 두 전통 정당에 대한 언급이 확연히 줄고, 노동당에 대한 언급이 늘어난 것 역시 눈에 띈다. 신문에서의 언급이 정당 지지율을 뜻하는 것은 결코 아니나, 노동당의 출현과 성장이 중앙지와 지방지 모두에 중요한 어젠다였음은 확인할 수 있다.
- 가치와 규범: 빅토리아 시대는 특정한 문화적 성격과 가치가 지배한 시대이기도 하다. 연구팀은 n-gram 자연어 분석을 통해 점차 빅토리아적 가치가 쇠퇴했다는 고전적인 학설을 확인하였다. 몇몇 핵심 가치가 전반적으로 쇠퇴함을 확인한 것이다. “의무”, “용기”, “인내” 등이 전시와 같은 특수한 때에는 일시적으로 많이 언급되었으나 또 다른 핵심적 가치인 “검약” “인내” 등은 쇠퇴하는 경향을 보였다.
- 정체성: 영국인의 정체성은 오랫동안 학자들의 큰 관심 대상이었다. 이 빅데이터 연구는 “영국성(Britishness)”의 부상에 대한 뚜렷한 타임 라인을

71) <https://www.pnas.org/content/early/2017/01/03/1606380114.full>

72) Ibid.

드러났다. 대중적인 관념으로서 영국성은 ‘잉글랜드성(Englishness)’을 19세기 후반부터 대체하기 시작한 것이다. 그 이후 20세기 초반에 ‘영국의(British)’ 라는 용어의 사용이 계속 증가하고 (특히 양차 세계대전 시기에 급증) 같은 기간 ‘잉글랜드적(English)’이라는 용어는 감소했다. 잉글랜드는 심지어 1940년대 후반에 이르면 스코틀랜드의 라는 용어보다도 더 적게 사용되었다. 즉 영국이 기본적인 국민 정체로 잉글랜드를 대체한 것이다. 지금까지의 연구사는 영국성의 발달을 그 이전으로 보고 있으나, 이번 결과가 의미하는 것은 대중적 상상력에서 영국성의 지배력은 20세기의 현상이라는 것이다.

- 기술: 최초의 산업혁명이 시작된 곳인 영국에서는 19세기 이후 급격한 기술 발달을 이루었다. 해당 시기에 증기기관이 점진적으로 쇠락하고 반대로 전기가 꾸준히 상승함을 관찰하였다. 둘의 이용률이 역전된 크로스 포인트는 1898년이었음을 알아낼 수 있었다. 또한 대중 교통 수단으로서 기차가 말을 앞지르기 시작한 것은 1902년으로, 기차의 시대가 1840년대에 시작되었음을 고려할 때 마력이 통념과는 달리 19세기 내내 일상적으로 문화적으로 중요했음을 새롭게 알려주는 것이다.
- 경제: 학문적, 문화적으로 ‘경제’라는 것이 별도의 개념이자 분야로서 빅토리아 시대 후기에 발달하기 시작했다는 경향이 발견되었다. 19세기 초중반의 ‘정치경제’라는 개념이 서서히 자취를 감추고 ‘경제’가 정치와 독립된 하나의 단위로 생각되기 시작했던 것이다. 유의할 점은 경제가 ‘절약, 저축’의 필요와 밀접하게 관련된 것으로 나타난다는 것이며 이는 특히 경제적으로 어려웠던 1922년과 1932년에 뚜렷했다.
- 지방과 중앙의 비교: “패닉”이라는 용어는 금융시장의 폭락에 상응하는 용어로 쓰였고, 도덕성이나 범죄 등에 대한 사회적 우려와는 관련되지 않았다. 1826, 1847, 1856, 1866년의 금융 시장 위기 시기에 대한 상세한 독해를 통한 검토에서도 확인되었다. 재미있게도 지방 신문들을 다수 샘플링하여 런던이 영국 전체를 잘 못 대표할 수도 있는 결과를 방지하려 했지만, 패닉에 대한 논의에서 지방지와 중앙지를 막론하고 런던 금융 시장의 중요

한 존재감이 다시 한 번 입증되었다.

- 젠더: “Suffragette”라는 용어의 등장 빈도는 매우 뚜렷하게 1906~1918년이라는 시기 안에 한정되었다. 당시 여성 참정권 획득을 위한 공공장소에서의 활동 때문인 것으로 보인다. 비록 여성 참정권 운동은 훨씬 더 전에 시작되었지만, 애스콧 경마장에서 달리는 왕의 말에 뛰어들어 죽은 참정권 활동가인 에밀리 데이비슨의 극적인 사망 사건 이후 뚜렷한 증가세를 보였다. 정치적 캠페인에 ‘미디어 이벤트’가 중요하다는 것을 뚜렷하게 보여주는 증거일 수 있다. 또한 전 시기 동안 남성이 여성보다 절대적으로 더 많이 언급되었다. 즉 우리는 역사 사료에서 실제로 남성에 대한 정보를 여성에 대한 뉴스보다 더 많이 접하고 있는 것이다. 또한 오늘날에도 남녀에 대한 언급도의 차이는 달라지지 않았다는 것 역시 흥미롭다. 여성에 대한 언급이 150년에 걸쳐 서서히 증가하기는 한다. 즉 시기 전체를 놓고 보았을 때 공적 영역에서 여성의 역할이 점진적으로 증가했음을 시사한다. 양차 대전 동안 그룹으로서 여성에 대한 인식이 드라마틱하게 증가했음도 흥미로운 현상이다.
- 문화 오락 영역에 대한 관심: 신문지상에서 사람들의 흥미와 관련된 것일 수 있는 “배우”, “가수”, “댄서”등에 대한 언급이 1890년대부터 증가한 반면 “정치인들”에 대한 언급은 20세기 초부터 서서히 감소했음이 관찰되었다. 오늘날 영국의 국민스포츠라 할 수 있는 축구(풋볼)이 1909년부터 크리켓보다 더 두드러지게 나타난 것도 재미있는 결과이다.⁷³⁾

위와 같이 오늘날에는 빅데이터 분석을 통해 장기적인 경향과 변화를 추적할 수 있다는 점이 이 연구의 핵심이라고 연구팀은 밝혔다. 150년, 해당 시기 간행된 신문의 14%라는 방대한 연구 대상은 빅데이터의 가능성을 최대한 살리기 위한 의도적인 선택이었다고 한다. 그렇지만 컴퓨터의 분석 결과에 의미와 해석을 입히는 것은 빅데이터 자체가 아닌 역사적 전문성의 영역이며, 질문과 키워드의 선정, 분석 결과의 해석에 이르기까지 역사학자의 식견이 영향을 미칠

73) Ibid.

수밖에 없다. “영국성”과 “잉글랜드성”, “서프러제트” 등 무엇이 학문적으로 가치있는 분석 결과를 도출하는 데 유용할 키워드인지를 판단할 수 있는 것은 오직 역사적 맥락에 대한 이해를 통해서만 가능한 것이다. 마찬가지로 20세기에 들어서서야 비로소 “영국성”이 “잉글랜드성”을 전사회적으로 대체하기 시작한 것이 왜 중요한지를 포착하는 것 역시 역사학자의 몫이다. 빅데이터 인문학은 컴퓨터에 의한 데이터 분석 그 자체라기보다 전통적 방법으로는 관찰되지 않을 경향과 양상을 드러나게 해 줄, 즉 ‘가까이 읽기’에 기초한 역사학 연구를 보강해 줄 수 있는 가능성을 보여 주었다.

영미권의 디지털 역사학 발전의 양상을 살펴보았다. 여기서 잠시 다음과 같은 아주 근본적인 수준에서의 입장 정리를 하고 넘어갈 필요가 있을 듯하다. 먼저 과거에 관하여 “살아남은 모든 것”을 가지고 연구하던 시대로부터 “모든 것이 살아남은” 시대로의 전환에서 역사학자들은 이전과 같이 사료와의 내밀하고 밀착된 관계를 지속해나가기 힘들 것이다.⁷⁴⁾ 이러한 상황에 대하여 역사가들이 가진 공포는 그 ‘모든 것’을 역사가 개인의 통제를 벗어나 알 수 없는 컴퓨터 프로그래밍 기술이 다루게 될 것이라는 우려이다. 디지털 턴의 역설이라고도 부를 수 있을 것이다. 이러한 공포에서 벗어나기 위해서는 역사가들 자신이 다시 컴퓨터 공학을 전공해 학위를 딸 필요는 없지만 최소한 디지털 방법론을 이용한 연구의 기저에서 작동하고 있는 기술적 측면에 대하여 이해하고 자신의 문제의식을 해결하기 위한 구상을 할 수 있는 능력은 필요할 것으로 보인다. 디지털화와 빅데이터가 가진 가능성은 인문학자들에게 새로운 무기가 되어 줄 수 있기 때문이다.

이에 더하여 빅데이터 시대에 직면하여 기존의 전통적 인문학자들이 지나치게 소심해지거나 자신감을 잃을 필요도 없다. 빅데이터 연구를 설계하고 결과를 해석하는 것은 여전히 ‘내러티브’에 관한 것이며, 이것은 늘 그랬듯이 인문학의 역할이기 때문이다. 폴 무어는 빅데이터에서 한 발짝 나아가 ‘두터운 데이터(Thick data)’라는 개념을 제시하는데, 데이터를 해석하여 의미를 구성하는

74) Ian Milligan and Robert Warren, “Chapter 6. Big Data and the Coming Historical Revolution: From Black Boxed to Models,” in *Big Data in the Arts and Humanities: Theory and Practice*

것, 즉 데이터를 내러티브로 전환하는 것의 중요성을 그렇게 표현한 것이다. 즉 데이터를 분석, 해석, 나아가 활용하는 데에서는 인문학적인 내러티브가 유효할 뿐만 아니라 핵심적이라는 것이다. 데이터 그 자체가 무적이 아니라는 것은 이미 2016년의 미국 대선 결과와 같은 해 영국의 브렉시트 결과를 예측하는데 많은 빅데이터 기반 예측이 실패한 데에서 극명히 드러났다. 브렉시트 투표 결과 데이터로 포착되지 못한 많은 ‘샤이(shy)’ 브렉시트파가 있었음이 드러났고, 미 대선에서 ‘샤이’ 트럼프들도 마찬가지였다. 데이터 기저에 있는 것은 미묘한 감정과 복잡성을 지닌 사람들의 삶과 사회인 것이고, 뉘앙스를 고려하여 분석하는 데에서만 드러나는 내러티브의 구축은 인문학자들이 주된 역할을 할 수 있는 분야이다. 인류학자 클리포드 기어츠가 일찍이 제시한 “두텁게 묘사하기”라는 컨셉은 빅데이터 시대에도 여전히 유효할 것이다.⁷⁵⁾

2.2.3. Digital Classics: 고전학의 재탄생?

■ 19-20세기의 고전학

유럽학계는 19세기와 20세기 초에 이르는 시기에 ‘고전학의 탑’을 쌓았다. 고대의 문헌과 유물, 유적이 대규모로 연구되었고, 그 결과 다양한 사전류(dictionary, lexicon, encyclopedia)와 비문 집성이 출현했다. 아직까지 연구되지 않은 비문이나 필사본이 일부 존재하나, 고대 그리스, 로마의 주요 문헌은 사실상 20세기 초에 총체적으로 종합되었다. 이러한 문헌학적 기반 위에서 고대의 모든 것에 대한 연구(Altertumswissenschaft)가 추구되었다. 20세기 후반에 실증주의에 대한 비판과 극복이 이루어졌지만, 고대 연구자들의 일용할 양식은 여전히 그러한 대규모 종합 작업이었다. 단적으로 말해 19세기 고전학이 20세기 고전학을 먹여 살렸다.

이러한 19-20세기 고전학은 21세기인 현재 돌이켜 볼 때 몇 가지 특징을 가졌다.

75) Paul Moore, “Chapter 5. From Big Data to Thick Data: Theory and Practice,” in *Big Data in the Arts and Humanities: Theory and Practice*, p. 56.

1. 학문적 보편주의: 그리스어-라틴어를 중심으로 하는 유럽 공통의 문화적 기반을 추구했다. 고전학과 고전교육은 제후령이나 왕국을 넘어서서 유럽 전체의 “Republic of Letters”의 회원권이였다.
2. 민족주의: 고전학 연구는 각각의 언어권(영어, 프랑스어, 독일어, 이탈리아어) 내에서 배타적으로 추구되는 경향을 보였다. 즉 고전학은 민족문화의 토대로 간주되었다. 단적으로, 독일의 경우 2015년 현재 약 200개의 고대 관련(문헌학, 고대사, 고전 고고학) 교수직 중, 190개가 독일 박사 출신으로 구성되어 있다.
3. ‘서양 문명’의 근간: 특히 동아시아에서, 고전학의 인상적이고 압도적인 학문적 성과물은 서양의 학문적, 문화적 우월성의 주요한 근거로 간주되었다. 경성제국대학 시절에 유럽의 출판과 거의 동시에 수집된 자료들은 이러한 태도를 반영한다.

■ Digital Classics, 21세기의 양식?

디지털 기술은 인문학 중에서도 특히 고전학 분야에서 두드러지게 활용되고 있다. 디지털화의 토대가 되는 데이터가 분명하며 동시에 대량이기 때문이다. 고전학에서 빅데이터의 출발점은 바로 고전어 어휘들이다. 예컨대 예일대학에서 학부 수준에서 요구하는 고전 그리스, 로마 독서 목록에는 누적적으로 계산하여 약 150,000개의 단어가 나오며, 미국 대학에서 고전학 박사에게 요구하는 독서 목록에는 약 1,000,000개의 단어가, Loeb Classical Library에는 총 20,000,000개의 단어가 쓰인다. 고전기 이후 그리스어와 라틴어로 쓰인 모든 단어는 거의 1,000,000,000개에 달할 것으로 추정된다. 19-20세기의 도구들은 학자와 대학원생의 고된 노동으로 제작되었지만, 디지털 기술은 이러한 ‘빅데이터’의 가능성을 새로운 차원으로 확대했다.

1. 고전 연구 도구의 디지털화

수많은 단어, 항목, 면수를 보유한 대규모 저작물들이 디지털화되어 막대한 도서관의 공간에서 연구자의 랩탑 속으로 들어왔다. 대표적으로,

Thesaurus Linguae Graecae(TLG)는 1972년부터 가동한 프로젝트로서, 호메로스부터 비잔티움(1453년)까지의 모든 그리스어 문헌을 수집하여 디지털화하고 있다. 완전한 그리스 문헌 디지털 도서관을 지향한다. (<http://stephanus.tlg.uci.edu/>)한편 Perseus Projects처럼 TLG와 유사하나 보다 접근성이 좋은 데이터베이스도 널리 활용되고 있다. (<http://www.perseus.tufts.edu/hopper/>)이들은 고전 텍스트의 디지털화를 초기부터 선도하면서, 그리스어와 라틴어의 OCR은 물론 웹 상에서 구현하는 기술적인 문제들에서 많은 경험과 발전을 이루었고, 어형 변화를 구별하는 도구와 텍스트 분석 도구를 제공한다.

2. 고대 세계의 시각화

GIS 기술의 도입은 고대 세계를 이해하는 방식에 혁신을 가져왔다. 예컨대 ORBIS: The Stanford Geospatial Network Model of the Roman World (<http://orbis.stanford.edu/>)는 고대 로마 세계를 단순히 물리적인 지리정보를 통해 제시하는 기존의 방식을 변화시켰다. 이 프로젝트는 자료들을 조정하고 비교하는 데에서는 기존 연구와 유사한 방법을 사용하지만, 역사 지식의 소통에서는 새로운 재현 방법을 보여준다. 이 모델을 이용하여 학자들은 고대의 교통과 통신을 반영한 지리적 시뮬레이션을 시도할 수 있다. 지중해의 물리적 지도 위에 도시와 육지와 바다의 교통로가 표시된다. 계절, 경로 특성(최단 시간, 최소 경비 등), 운송 방식(도보, 말, 우마차, 행군 등) 등 여러 변수를 조정하여, 기존의 지리적 거리를 시간 거리로 시각화한다. 이는 고대 세계에 대한 이해를 근본적으로 변화시킬 수 있다.

이러한 새로운 가능성은 고전학계에 근본적인 변화를 일으켰다. 특히 서양 고전학계는 20세기 후반부터 교수직, 대학원생 수, 고전 교육을 받는 인구 수 등이 꾸준히 감소하면서 위기가 감이 고조되었고 새로운 돌파구를 갈망했다. 디지털 기술은 엄밀한 고전학 연구 자체는 물론, 고전학/고대사 교육의 확대와 일반 사회로의 지식의 확산에 중대한 변화를 일으켰다.

1. 대규모 데이터베이스 구축과 프로젝트 운영: 유럽과 미국의 주요 대학과 기관들에서 고대의 문헌과 유물을 다루는 대규모 데이터베이스 구축 사업이 지속적으로 확대되고 있다. 개별 프로젝트는 일일이 나열하기 어려울 정도이며, 사실상 이전의 모든 연구 프로젝트가 디지털 연구로 전환되고 있다.
2. Digital Classics 세미나: 디지털 고전학에 전념하는 세미나가 영국, 독일, 미국에서 활발히 운영되고 있다. 이 세미나들은 전 세계의 디지털 고전학의 연구와 토론의 핵심적인 소통의 장으로 자리잡았다. 신진 연구자들은 이제 디지털 고전학으로 박사 학위를 받을 수 있다.

이렇게, 서양 고전학계에서는 19세기식 문헌학적 종합을 21세기에 디지털화로 재현하고, 나아가 디지털 기술이 여는 새로운 연구와 교육의 가능성을 확대하고 있다. 대규모 프로젝트에서 쌓인 기술적인 숙련도를 바탕으로, 새로운 세대에 디지털 기술을 교육하고 더 개선된 연구 도구를 개발하며, 대중과 학생들에게 접근이 용이한 새로운 지식 확산 환경을 개발하고 있다. 주요 대학의 고전학과와 연구센터들이 이러한 사업을 제도적, 재정적으로 뒷받침하고 있고, 디지털 고전학 자체에 매혹된 연구자들이 증가하고 있다.

한국의 연구자로서 이러한 서양 고전학계의 흐름을 보며 몇 가지 단상을 가진다.

1. 서양 고전학계의 디지털화는, 마치 20세기 초까지 라틴어가 수행했던 것처럼 지방적, 민족국가적 경계를 뛰어 넘는 국제적인 문화로 발전할 가능성이 보인다. 아직 대부분의 프로젝트가 여전히 개별 국가의 재정 지원에 의존하고 있지만, 디지털 고전학은 학제간 협력뿐 아니라 국제적인 협력이 필수적이며 또 그것을 촉진한다.
2. 서양 고전학계는 연구와 교육에서 디지털화에 매우 앞서가고 있으나, 그러한 성과를 대중적으로 확산시키는 문제에서는 아직 뚜렷한 흐름을 보여

주지 않는다. 연구 인프라의 확충이 우선시되는 것은 이해할 수 있으며, 아마 앞으로는 전문 연구, 교육의 차원과 동시에 일반사회와의 소통이 더욱 강조될 것으로 추측된다.

3. 디지털 고전학을 국내에서 수행하려 할 때, 특히 교육과 확산 차원에서 접근할 때 일차적으로 직면하는 문제는 저작권이다. 서양에서는 고전 번역의 역사가 오래되어서, 이미 저작권에서 자유로운 고전 번역 텍스트가 매우 많다. 그 자체가 중요한 데이터 자원이 된다. 그러나 국내에서는 최근에야 고전 번역이 제대로 이루어지고 있어서 그 텍스트를 교육에서 활용하려면 저작권 문제가 해결되어야 한다. 이것은 그리 쉬운 과제는 아닐 것으로 보인다.

2.3. 각국 정부의 디지털 인문학 지원

각국 정부는 2010년 이후 디지털 인문학에 대한 대대적인 지원을 천명하고 실천하고 있다. 미국, 영국 및 독일을 중심으로 하여 각국의 정부가 디지털 인문학에 대해 어느 정도의 규모와 방식으로 디지털 인문학을 지원하고 있는지를 살펴보고자 한다.

2.3.1. 미국

미국의 디지털 인문학은 초기에 주로 대학 내의 디지털 인문학 센터별로 아카이브 구성, 웹사이트 제작, 교육 콘텐츠 제작 등의 프로젝트 중심으로 발전을 해왔는데 2010년대 들어서 이러한 발전이 교육 분야로 자연스럽게 확대되고 있다. 따라서 디지털 인문학 발전 초기에는 프로젝트에 대한 지원이 중심이었으나 최근 교육에 대한 지원도 증가하고 있는 추세이다. 다양한 기관들이 디지털 인문학을 지원해오고 있고 최근 그 규모가 확대되고 있다. 그 중에서 앤드류 멜론 재단(Andrew W. Mellon Foundation)이 1994년부터 디지털 인문학을 지원해오고 있는데 이 재단에서 최근 주요 대학들의 디지털 인문학 교육 프로그

램에 대규모 지원을 하고 있다는 점이 주목할 만하다. 다양한 디지털 인문학 프로젝트를 진행하면서 쌓인 인력과 지식을 활용할 수 있는 상황에서 디지털 인문학의 확장을 위해서는 관련 교육이 필요하다는 인식이 생겨나면서 디지털 인문학 교육의 중요성에 대한 인식이 점차 확산되고 있다. 따라서 대학들이 디지털 인문학 교육에 관심을 가지면서 많은 지원이 디지털 인문학 교육 프로그램을 개발하고자 하는 대학들에게 제공되기 시작하고 있다. 물론 디지털 인문학 연구에도 여전히 많은 지원이 이루어지고 있다. 다양한 종류의 디지털 인문학 프로젝트를 지원하고 있고 특히 최근에는 디지털 인문학의 출판에도 점차 관심을 확대해나가고 있다. 여러 지원 기관 중에서 가장 중요한 역할을 하고 있는 기관은 앤드류 멜론 재단(Andrew W. Mellon Foundation)과 미국인문학 재단(National Endowment for Humanities)의 디지털인문학단(Office of Digital Humanities)이다.

2.3.1.1. 앤드류 멜론 재단 (Andrew W. Mellon Foundation)⁷⁶⁾

- 1994년부터 현재까지 186개의 프로젝트에 총 9천6백만 달러 (1170억 정도)를 지원해오고 있다.
- 초기부터 2010년대 초반까지 주로 디지털 아카이브 구축이나 디지털 인문학 프로젝트에 대한 지원이 간헐적으로 이루어지다가 2013년부터 지원금 규모와 대상이 확대되었다.
- 특히 2013년부터 메이저 대학들에 디지털 인문학 교육 프로그램 지원을 집중적으로 진행하고 있다. 2013년부터 2019년까지 다음의 대학들에 각각 100만 달러 (12억) ~ 200만 달러 (24억 정도)의 대규모 지원금을 제공하여 디지털 인문학 교육의 틀을 마련하는데 기여하고 있다.
 - ▶ University of Maryland at College Park, University of Pennsylvania, University of California at Berkeley, Carnegie Mellon University, University of Pennsylvania, University of Southern California, Grinnell

76) <https://mellon.org/grants/>

College, Georgia Institute of Technology, Vanderbilt University, Claremont University Consortium, Five Colleges Incorporated at Amherst, MIT, University of Rochester

2.3.1.2. 미국인문학재단(National Endowment for Humanities)의 디지털 인문학단⁷⁷⁾

- 목표와 방향

- (1) 디지털 기술들을 활용하여 인문학 연구, 교육, 학습, 대중 참여, 학문적 소통을 위한 새로운 방법들을 개발하는 프로젝트 팀들을 지원한다.
- (2) 또한 인문학적 관점으로 디지털 문화를 연구하는 학자들과 디지털 출판물을 하고자 하는 인문학자들을 지원한다.
- (3) 궁극적으로 인문학이 디지털 방법들을 적용하는 능력을 키우는 것을 돕는다.

- 방법

- (1) 국경과 학문 분과를 초월하여 미국뿐 아니라 해외의 학문 공동체와 다른 지원 기관들과의 긴밀한 협업을 통해 지원한다.
- (2) 담당 직원들이 학문 공동체와 컨퍼런스나 워크숍에서 만나 현장의 목소리를 듣고 현장에서 필요한 요구를 충족시키기 위해 노력한다.
- (3) 결과물을 백서로 출판해서 널리 공유하여 디지털 인문학 발전에 실질적인 기여를 한다.

- 현재까지 지원한 프로그램

- (1) 2007년에 3만 달러 (3천 6백만원) 정도 규모의 프로젝트 25개를 지원했고 이후 매년 10~20개 사이의 프로젝트를 지원해오고 있다.
- (2) 지원액은 적게는 2만 달러 (2천 4백만원) 정도에서 많게는 40만 달러 (4억 8천만원)까지 다양하고 지원 기간도 6개월에서 3년까지 다양하다.

77) <https://www.neh.gov/divisions/odh>

(3) 구체적인 프로젝트 내용과 지원액은 각주의 웹사이트에서 확인할 수 있다.⁷⁸⁾

- 자체 운영 프로그램

(1) 데이터 발굴 (Digging into Data) 프로그램

* 2009년, 2011년, 2013년, 2016년 총 네 번 소수(10개 전후)의 장기 프로젝트에 대규모 지원

* 2017년에 14개 팀에 총 9.2백만 달러 (110억원 정도)를 지원했음 (지원받은 프로젝트들에 관한 정보는 각주의 웹사이트를 참고하면 됨)⁷⁹⁾

(2) 인문학 오픈북 프로그램 (Humanities Open Book) 프로그램⁸⁰⁾

* 절판된 책들 중에서 뛰어난 도서들을 폭넓은 독자들이 접근 가능하도록 하는 프로젝트 NEH와 앤드류 멜론 재단이 공동 후원함.

(3) 디지털 인문학 발전 지원금 (Digital Humanities Advancement Grants)

* 디지털 인문학 프로젝트를 단계별로 지원. 개인이나 팀이 아닌 기관에 지원⁸¹⁾

* 1단계- 5만 달러

* 2단계 - 10만 달러

* 3단계- 32만 5천 달러 (추가적으로 5만 달러 매칭 펀드)

(4) 디지털 인문학 교육 지원 사업 (Institutes for Advanced Topics in the Digital Humanities)

* 학자, 인문학 전문가, 대학원생들을 대상으로 한 디지털 인문학 교육 사업 지원

* 인문학자들이 디지털 기술을 활용하도록 돕고 인문학과 연관된 기술

78) <https://securegrants.neh.gov/publicquery/main.aspx?q=1&a=0&n=0&o=0&ot=0&k=0&f=0&s=0&cd=0&p=0&d=1&dv=12&y=0&prd=0&cov=0&prz=0&wp=0&ob=year&or=DESC>

79) <https://diggingintodata.org/awards/2016/news/winners-round-four-t-ap-digging-data-challenge>

80) <https://www.neh.gov/divisions/odh/humanities-open-book>

81) <https://www.neh.gov/grants/odh/digital-humanities-advancement-grants>

- 적 도구들과 방법론들에 관한 지식을 널리 전파하는 것이 목표
 * 36개월 동안 25만 달러(3억원 정도)까지 지원

2.3.1.3. 미국학술단체협의회(American Council of Learned Societies, ACLS)의 디지털확대지원금 (Digital Extension Grants)⁸²⁾

- 자체 운영 프로그램

- (1) 인문학 및 관련 사회과학의 모든 분야의 디지털 기반 연구 프로젝트를 지원
- (2) 기존의 디지털 프로젝트를 강화하거나 새로운 커뮤니티로 영역을 넓히고, 다양한 학자들로 구성된 팀을 지원함으로써 인문학적 연구를 발전시키는 것을 목표로 함

- 지원 내용

- (1) 총 5개 프로젝트에 각각 15만 달러(1억 8천만원 정도)까지 지원
- (2) 초기 단계를 지난 프로젝트에 지원 (12-18개월 프로젝트)

- 재원 조달 - 앤드류 멜론 재단 기금 활용

2.3.2. 영국

2.3.2.1. 정책적인 입장

디지털 인문학과 빅데이터에 대한 영국 정부의 입장은 ‘예술 인문 연구회 (Arts and Humanities Research Council, 이하 AHRC)’를 통해서 가늠해 볼 수 있다. AHRC는 인문학과 예술 분야에서 수행되는 여러 프로젝트들에 대한 펀딩을 주관하는 국립 에이전시로, 영국의 인문학계가 집중적으로 다루어야 할 주제, 즉 AHRC가 중점적으로 지원할 테마에 대한 논의를 몇 년 단위로 제시하고 있다.

AHRC가 발표한 최근 리플렛들을 살펴보면 AHRC는 빅데이터 및 디지털을

82) <https://www.acls.org/programs/digitalextension/>

인문학에 적극적으로 포함시키려 하고 있음을 알 수 있다. 2013년-2018년의 5년 전략에서 AHRC는 ‘디지털적 변화(Digital Transformation)’를 5개의 핵심 육성 테마 중 하나로 정했고, 빅데이터는 그 테마의 하위 주제로 설정되었다. 그럼으로써 “인문학과 예술 연구에 변화를 일으킬 디지털 기술의 잠재력을 탐구하고, 디지털 시대의 지적 재산권, 문화적 기억과 정체성, 커뮤니케이션과 창조력과 같은 주제들에 주목”할 것이라고 천명하였다.⁸³⁾

이처럼 AHRC는 빅데이터가 인문학·예술 연구에 가져올 수 있는 가능성을 충분히 인지하고 있을 뿐만 아니라, 반대로 빅데이터에 대한 접근과 사용에서 인문학이 할 수 있는 역할이 있다는 점 역시도 중시하고 있다.⁸⁴⁾ 특히 빅데이터가 사용되는 다양한 맥락을 고려하는 데에는 인문학이 필수적이라는 점을 강조하였다.

이 2013-2018 프로그램의 연장선상에서 2019년에 발표된 향후 5년에 대한 상세 계획서(Delivery Plan)는 한층 더 구체화된 입장을 드러냈다. ‘인공지능(AI)과 데이터에 대한 윤리적 규제’를 오늘날의 사회가 맞닥뜨린 문제들 중 하나로 꼽았으며, 다른 문제들인 미디어와 공적 신뢰성, 정신적 질병, 인구 노령화 등의 이슈들과 마찬가지로 그 해결에는 인문학적 사고가 필수적이고 인문학 안에서의 학제간 연구, 그리고 인문학과 다른 분과와의 협업을 추구해야 한다는 점을 강조했다. 특히 “인문학 및 예술이 과학과 공유하는 공통의 가치와 관심을 지속적으로 추구할 것”이라고 하면서 빅데이터 및 디지털 기술의 포섭이라는 방향을 폭넓게 뒷받침했다.⁸⁵⁾

AHRC는 디지털이 인문예술을 실질경제로 연결시키는 데에서도 핵심임을 강조했다. 대학, 박물관, 갤러리, 문서고 사이의 긴밀한 협업을 통해 인문학·예술 연구와 문화유산 간의 연결고리를 긴밀하게 하고, 또 이를 통해 대중적 참여와 문화유산 관련 경제를 더욱 육성하겠다고 발표했다. 디지털 기술은 문화유산 부문, 디자인, 창조 산업들과 함께 영국 경제에서 중요한 부분을 담당하고

83) Arts & Humanities Research Council, “The Human World: The Arts and Humanities in Our Times,” AHRC Strategy 2013-2018, p. 15.

84) Arts & Humanities Research Council, Delivery Plan 2019, p. 12.

85) Ibid., p. 2.

있는 문화·창조 산업의 핵심이며, 바로 여기에 미래의 인문학과 예술의 방향을 맞추어 세계에서 선두를 차지하겠다는 의지를 표명했다.⁸⁶⁾

정리하자면 영국 정부는 AHRC를 통해, 지금까지 디지털과 인문학의 관계에 대한 논의에서 제기된 두 가지 줄기를 모두 의식하고 포섭하고 있다고 볼 수 있다. 그 두 줄기 중 하나는 빅데이터는 인문학에 새로운 가능성과 혁신, 나아가 경제적 가치 창출을 가져올 수 있다는 전망이고, 다른 하나는 인문학이 빅데이터의 기술적 도전과 가능성을 철학적, 비판적으로 이끄는 역할을 맡아야 한다는 필요성이다.

또 하나 주목할 만한 점은 위의 논의를 통틀어 국가의 수사가 빈번히 등장한다는 것이다. 디지털과 빅데이터를 포함한 현재적 이슈들에 대한 대응에서 영국의 인문학이 “세계에서 지도적 역할”을 맡고 “기준을 제시할 것”이라는 선언이 여러 차례 눈에 띈다.⁸⁷⁾ 이러한 수사는 AHRC가 국립기관이기 때문에 삽입된 것일 수도 있지만 빅데이터와 디지털 인문학에 대한 영국 정부의 지원 의지를 거듭 나타내고 있다고 볼 수 있다.

2.3.2.2. 구체적인 지원 방안

‘디지털적 변화(Digital Transformation)’는 AHRC의 2013-2018년의 5년 전략에서 집중 육성 테마 중 하나였다. (다른 집중 육성 테마들: Care for the Future, Science in Culture, Translating Cultures, Connected Communities) 이 테마 하에서 빅데이터에 관련된 총 36개의 프로젝트들이 다시 크게 3가지 분야로 나뉘어 펀딩을 받았다.⁸⁸⁾

1) Digital R&D Fund for the Arts

이 분야는 예술 및 문화 분야에서 빅데이터의 가능성을 검토하기 위한 것으로

86) Ibid., pp. 2-3.

87) Ibid., p. 3.

88) <https://ahrc.ukri.org/research/fundedthemesandprogrammes/themes/digitaltransformations/bigdata/>

4가지 프로젝트에 펀딩을 하였다. 구체적으로 이 프로젝트들은 예술 및 문화 분야가 성장하고 새로운 비즈니스 전략을 발전시키는데 데이터를 어떻게 이용할 것인가를 탐구했다. 4가지 프로젝트에 대한 세부 사항은 다음과 같다.

프로젝트	내용	금액
ArtsAPI	웹 기반 서비스가 데이터를 모으고 분석하고 제시하도록 함으로써 예술 기관들은 결과물의 가치와 영향력을 가늠해 볼 수 있음. 예술 기관들로 하여금 더 새롭고 정제된 비즈니스 모델을 만들어 내도록 하는 데 도움.	£292,343
Arts Data Impact	ADI 프로젝트는 Barbican, English National Opera, National Theatre 등에 최초로 상주 데이터 과학자를 파견하여 날로 급증하는 데이터를 확보하도록 함. 또한 국립 데이터 저장소를 만들어 예술 및 문화 분야 전반에 걸쳐 포괄적으로 데이터를 수집하고 공유하도록 함.	£299,985
Culture Counts	맨체스터에 소재한 예술 문화 기관들로 하여금 작업물의 질과 영향력을 평가할 수 있도록 하는 온라인 플랫폼 품을 테스트해보는 프로젝트.	£300,000
The Unusual Suspects	뉴캐슬과 게이트헤드에 소재한 예술 문화 기관들이 관객들의 데이터를 수집, 분류하여 사람들의 방문횟수를 끌어올리고 예술 문화 경험을 질적으로 향상시키는 방법을 찾도록 함.	£184,325

2) Digital Transformations in Community Research Co-production in the Arts and Humanities

지역사회 기반 창조 경제(Creative Economy)와 관련된 프로젝트 11개에 펀딩했다.

3) Digital Transformations in the Arts and Humanities: Big Data Research

21개의 프로젝트가 펀딩을 받았는데, 그 중 10개는 소규모, 11개는 대규모

프로젝트였다. 이 펀딩의 목적은 빅데이터 시대에 인문학이 직면한 도전과 빅데이터를 가지고 인문학을 하는 데에서 생겨나는 이슈를 다루는 것, 그럼으로써 데이터에 대한 유의미한 질문을 제기하고 미래의 인문학·예술을 위한 혁신적 창조적 자산을 창출해내는 데에 있다. 21개 프로젝트들은 다음 표와 같다.

[표 1] 소규모 10개 프로젝트 (small projects under £100k)

펀딩 수혜자	프로젝트 제목	연구 수행 기관	금액
Dr. Humphrey Southall	Frameworks and tools for statistical big data in the humanities	University of Portsmouth	£79,936.40
Dr. Matthew Wall	What are the odds? Capturing and exploring data created by online political gambling markets.	Swansea University	£60,800.54
Dr. Gabriel Bodard	Standards for Networking Ancient Prosopographies: Data and relations in Greco-roman Names	King's College London	£81,795.92
Dr. Jo Bates	The Secret Life of a Weather Datum	University of Sheffield	£79,459.66
Dr. Alan Marsden	Optical Music Recognition from Multiple Sources	Lancaster University	£78,575.82
Dr. Sarah Atkinson	DEEP FILM Access Project (DFAP)	University of Brighton	£79,710.75
Dr. Dominique Jeannerod	Visualising European Crime Fiction: New Digital Tools and Approaches to the Study of Transnational Popular Culture	Queen's University of Belfast	£79,896.21
Dr. Robert Villa	Understanding the annotation process: annotation	University of Sheffield	£79,962.37
Dr. Stephen Rose	A Big Data History of Music	Royal Holloway, Univ of London	£79,838.33
Dr. Susan Rennie	A Pilot Historical Thesaurus of Scots	University of Glasgow	£77,678.62

[표 2] 대규모 11개 프로젝트 (large projects under £600k)

펀딩 수혜자	프로젝트 제목	연구 수행 기관	금액
Mr. John Sheridan	Legislation Data Research Infrastructure	The National Archives	£380,000.00
Prof. Julia Thomas	Lost Visions: retrieving the visual element of printed books from the nineteenth century	Cardiff University	£350,118.92
Dr. Sonia Ranade	Traces through Time: Prosopography in practice across Big Data	The National Archives	£380,000.00
Dr. Tillman Weyde	Digital Music Lab - Analysing Big Music Data	City University London	£450,320.90
Dr. Marc Alexander	Semantic Annotation and Mark Up for Enhancing Lexical Searches(SAMUELS)	University of Glasgow	£413,785.33
Prof. James Loxley	Palimpsest: an Edinburgh Literary Cityscape	University of Edinburgh	£471,922.34
Prof. Sophia Ananiadou	Mining the History of Medicine	The University of Manchester	£258,908.64
Dr. Dirk Obbink	Proteus: Capturing the Big Data Problem of Ancient Literary Fragments	University of Oxford	£454,303.78
Dr. Helen Kennedy	Seeing Data: are good big data visualisations possible?	University of Leeds	£224,053.99
Dr. Jane Winters	Big UK Domain Data for the Arts and Humanities (BUDDAH)	University of London	£341,373.39
Prof. Jane Stuart Smith	Dynamic dialects: integrating articulatory video to reveal the complexity of speech	University of Glasgow	£188,587.05

위의 21개 빅데이터 연구 프로젝트들 중 대표적으로 두 가지 프로젝트에 대해 살펴보자. (위 붉은 글씨로 표시된 두 개) 2014~2016년에 진행된 두 프로젝트들은 공통적으로 데이터의 구축과 형성이 아닌 빅데이터를 다루는 분석 도구를 만들고 개선하는 데 목적을 두었다. 역사학 뿐만 아니라 인문학에서 웹 아카이브는 점점 더 중요해지고 있으며 데이터의 규모는 하루가 다르게 증가하

고 있다. 그러나 연구자들은 자신들이 이용하는 데이터의 형성과 수집 절차 모두에 대해 거의 무관심하다. 현재의 디지털 분석방법으로는 너무 일반적이고 광범위한 결론만을 이끌어 낼 수 있을 뿐이다. 따라서 데이터를 효과적으로 분석할 수 있는 분석 도구와 전문성의 개선이 시급한상황이다. 아래 두 프로젝트는 모두 이처럼 빅데이터를 이용한 연구를 개선하는 데 전제조건인 분석 방법에 초점을 두고 수행되었다.

프로젝트명	프로젝트 내용
Traces through Time: Prosopography in practice across Big Data (https://gtr.ukri.org/projects?ref=AH%2FL010186%2F1)	디지털화가 진행됨에 따라 역사 자료에서의 데이터의 증가 뿐만 아니라 컴퓨터 알고리즘의 출현은 역사학 연구를 크게 변화시킬 잠재력을 가져왔다. 빅데이터로 지금까지 시도되지 못했던 전체 인구 규모를 단위로 하는 역사 연구를 수행하기 위해서는 알맞은 톨과 접근법이 전제되어야 한다. 지금까지 자료의 디지털화에 대해서는 엄청난게 투자가 이루어져 왔고 역사적 빅데이터는 그 다양성, 복잡성, 규모에서 현존하는 분석 톨과 테크닉을 넘어서고 있다. 이 프로젝트는 대규모 역사 데이터셋을 효과적으로 분석하기 위한 톨을 가다듬고, 경영과 과학 등에서 앞서 개척된 분석 톨을 적용시키는 것에 초점을 두었다.
	영국사 500년에 나타나는 개인들에 관한 빅데이터를 대상으로 두고, 자연어 처리, 데이터 마이닝 등 지금까지 디지털 인문학에서 축적된 경험을 망라하여 분석 톨을 개발하였다. 프로젝트 결과 대상 데이터셋 안에서 인적 정보를 구별할 수 있는 톨킷을 개발했다. 이 톨킷은 역사적 데이터에서 종종 나타나는 비일관적인 스펠링, 공식 문서에서 등장하는 오류 등을 다루는 데 적합하도록 만들어져, 데이터 속을 헤엄치며 산재된 개인들의 흔적을 발견하고 생애사를 발견하게 된다.
Big UK Domain Data for the Arts and Humanities(BUDDAH) (https://buddah.projects.history.ac.uk/)	웹 아카이브가 인문학 연구에 점점 더 중요해지고 있으나 그것들을 효과적으로 활용할 수 있는 도구에 대해서는 상대적으로 주목하지 않았다. 이 프로젝트는 1996년부터 2013년까지 UK 웹 도메인을 크롤링하여 65테라바이트, 수십억개의 단어들로 이루어진 데이터셋을 가지고 효과적인 이론적, 방법론적 톨을 개발하고자 했다. 프로젝트 결과 개발된 분석 도구를 검토할 하부 프로젝트들(역사, 문학, 젠더 연구, 물질 문화 등에 대한 케이스 스터디)도 진행되

프로젝트명	프로젝트 내용
	었다. 즉 2013년 이후에 만들어지고 있는 웹 공간의 크롤링에도 적용될 수 있는 도구의 개발과 그 도구를 이용하여 1996-2003년까지의 영국 웹공간에 대한 연구(언어, 파일 형식, 멀티미디어 콘텐츠의 발전 등)가 모두 수행되었다.

[표 3] AHRC가 지원한 빅데이터 인문학 프로젝트 예시

프로젝트명	프로젝트 내용
The Digging into Linked Parliamentary Data Project (http://dillipad.history.ac.uk)	국제적인 ‘Digging into Data Challenge’의 2013년 세 번째 라운드에서 수상한 14개 프로젝트 중의 하나이다. AHRC는 JISC와 함께 펀딩 후원자로서 참여하고 있다. 영국, 네덜란드, 캐나다의 사회과학자, 역사학자, 컴퓨터 언어학자가 협업하여 1800년 이후부터 지금까지의 각 3국의 의회 기록을 분석했다. 3국의 200년 동안의 의회 기록이라는 방대한 데이터를 다룰 효과적인 분석 툴들을 개발하고 그것을 테스트하기 위해 기술적인 개발에만 머물지 않았다. 연구팀은 젠더와 정치라는 구체적인 주제에 초점을 맞춰 세 나라에서 여성이 정치무대에서 대표되는 모습을 비교하는 등 정치문화, 정치적 이데올로기, 의회적인 용어 등 케이스 스터디 역시 수행하였다.
Living with Machines (https://www.turing.ac.uk/research/research-projects/living-machines)	영국도서관과 앨런 튜링 연구소의 협업 프로젝트로, 무려 920만 파운드의 예산으로 5년 동안 수행되는 가장 야심찬 규모의 프로젝트이다. 1차 산업혁명 시기 북부 잉글랜드에 대한 연구가 초점으로, 이 주제는 1960-70년대 이래 계량적·전산학적 방법론을 추구했던 신경제사가 쇠퇴한 이후 상당 기간 동안 주목에서 빗겨가 있었고 연구사적 공백이 남겨졌다. 이 프로젝트는 1차 산업혁명 시기 기술의 변화가 사회의 곳곳에 어떤 영향을 미쳤는지 알아보기 위해 역사적 데이터 분석에 적용될 수 있는 인공지능 툴을 개발하고자 하였다.

2.3.3. 독일

독일의 경우 교육연구부(German Federal Ministry of Education and Research, BMBF)와 독일연구재단(Deutsche Forschungsgemeinschaft, DFG)이 누 차례에 걸쳐 디지털 인문학의 중요성을 강조하고 이에 대한 연구

지원을 약속한 바 있다. 먼저 독일교육연구부의 디지털 인문학 관련 사업은 다음의 세 가지로 나뉘고 있었다.

독일 교육 연구부의 디지털 인문학 관련 사업

eHumanities(2011~2013):

<https://www.bmbf.de/foerderungen/bekanntmachung-643.html>

정보인프라구조

<https://www.bmbf.de/de/informationsinfrastrukturen-745.html>

디지털 대학 육성

<https://www.bmbf.de/de/digitale-hochschullehre-2417.html>

눈에 띄는 것은 eHumanities 분야를 독자적인 분야로 설정하고 있다는 점이다. 관련되는 규정은 다음과 같다.

<2011년 독일교육 연구부는 eHumanities 분야의 연구를 촉진하기 위해 다음과 같은 원칙을 선언함>⁸⁹⁾

1. 활용 목표 및 법적 토대

1.1. 활용 목표

연방교육연구부는 eHumanities 분야에서 연구개발 사업을 장려할 계획이다. eHumanities는 현대 정보 기술의 연구, 개발 및 응용을 통해 인문학의 작업을 용이하게 하거나 개선하고자 하는 모든 접근방식을 합친 것으로 이해되고 있다.

인문학적 및 질적인 연구를 하는 사회과학 분야들이 그들의 전문분야에서 새로운 연구 접근 방식을 개발하는 데 있어, 도움이 요구되고 있다.

그 경우 연구 과제들은 일회적인 분과 학문에 특수한 의제들을 다룰 뿐만 아니라 eHumanities의 발전에 중요한 기여를 해야 한다. 그 목표는 디지털 소스 기반의 혁신적인 연구 방법 개발, 인문학과 사회과학 분야의 관련 전문 지식 개발, 그리고 관련 정보 관련 과목의 eHumanities에 대한 전문 지식 개발이다.

89) <https://www.bmbf.de/foerderungen/bekanntmachung-643.html>

인문과학 또는 사회과학 연구 의제가 이 프로젝트의 중심이 되어야 한다.

이 발표를 하게 된 것은 광범위한 디지털 소스 자료를 기반으로 하는 연구에서 정보통신 기술 활용이 증가되고 있는 가운데, 이와 관련된 잠재력을 보다 더 잘 활용하고자 함이다....

이처럼 독일 교육연구부는 eHumanities를 별도의 분야로 지원함으로써 인문학 및 사회과학 분야의 개별 연구를 촉진하는 선을 넘어서서 디지털 자료를 잘 다룰 수 있도록 하기 위한 토대를 마련하고자 한다. 이외에도 정보인프라구조 사업, 디지털대학 육성 사업 등을 통해 디지털-빅데이터 시대의 변화에 본격적으로 대응하고자 한다.

교육연구부 이외에 독일 연구재단인 DFG도 디지털인문학 연구 지원에 적극적이다. 2018년 연구재단 이사장이 보낸 신년 메시지의 경우 “디지털러티(Digitality)가 다른 어떤 것보다도 우리의 존재와 과학과 인문학을 변화시키게 될 기회가 될 것”이라고 모두에서 언급하는 등 신년사 전체가 “우리가 디지털러티라고 불리는 세계의 변화”에 어떻게 대처할 지에 집중되고 있음을 볼 수 있다.⁹⁰⁾

“디지털”을 이토록 중시한 결과 여러 차원에서 매우 적극적인 대응이 이루어지고 있다. 이를 위해 연구재단 최고위원회 (DFG Senate)는 매우 포괄적이고 체계적인 정책을 수립하기로 결정한 바 있다.

	내용
배경 ⁹¹⁾	과학과 인문학에서의 디지털 턴은 디지털 기술이 제공하는 새로운 기회들을 이해하고 활용하는 것뿐만 아니라 아날로그 데이터를 디지털 형태로 변환시키는 것과 관련되는 다양한 과정들에 대한 것이다. 디지털 세계 내의 발전의 속도와 역동적인 속성은 이러한 전환에서 핵심적인 특징이다. 이러한 역동적인 변화는 사회 여타의 영역 뿐만 아니라 과학과 인문학으로 하여금 그 조건을 명백히 하고 발전과 그 잠재력을 평가하도록 한다. 또한 최적의 연구를 보장하기 위해서 법적, 윤리적, 조직적, 재정적 측

90) https://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/2018/170116_rede_strohschneider_neujahrsempfang_en.pdf

	내용
전문가 위원회	<p>면과 관련되는 새로운 규칙들을 세우는 것과 관련될 수도 있다. 독일연구재단은 디지털 턴의 역동성을 검토하기 위해 전문가들로 구성된 위원회를 결성하였으며, 이 위원회는 다음과 같은 네 가지 영역에서 주로 활동하게 된다.</p> <ul style="list-style-type: none"> • 디지털 턴의 결과로서 방법론적이고 과학적인 개념들에서의 변화 • 디지털 턴에서 리서치 과정 • 디지털 방법과 과학적 원칙들의 관련성 • 연구의 디지털 자산(데이터, 출판물, 소프트웨어)과 하나의 사회 체계로서 연구에 미치는 그 중요성
연구재단 Head Office Structuring Project	<p>체계적인 정책 입장을 구성하기 위해서 연구재단 Head Office는 “디지털 시대 과학과 인문학의 현재 상황”에 대해 평가하기 위한 프로젝트를 시작했다.</p>
“인문학에서 디지털러티”라는 제목의 심포지움 지원 ⁹²⁾	<p>인문학 분야 전반에서 디지털적인 변화들에 대한 공통의 이해를 가능케 하고자 다양한 학문 분야의 대표자들로 구성된 심포지움 시리즈들을 지원한다. 그 경우 “<u>디지털 테크놀로지들이 인문학적인 연구와 그 연구대상들을 어떻게 변화시키는가?</u>”, “<u>인문학의 확대가 일어나는가 혹은 인문학의 본질이 달라지게 되는가?</u>”, “<u>고전적인 해석학적인 단초와 디지털적인 방법과의 관계는 어떻게 설정되어야 하는가?</u>”, “<u>그 결과들은 학문후속세대에 어떤 영향을 미치는가?</u>” 등의 문제가 논의된다.</p> <p>디지털적인 맥락에서 인문학 연구와 연구 방법론들에 대한 전반적이고 근본적인 주제들에 대한 논의가 이루어지게 될 계획이다.</p>

그 외에도 독일연구재단은 다음과 같은 다양한 프로그램을 통해서 과학과 인문학 분야의 “디지털 턴”과 관계되는 다양한 펀딩 옵션들을 제공하고 있다.

명칭	내용
NEH/DFG 양자간 디지털인문학 프로그램 ⁹³⁾	<p>NEH/DFG 기금: 독일의 연구재단과 미국의 인문학재단이 함께 프로젝트 펀딩에 나서게 된 것은 물론 양국간의 협력을 촉진하기 위한 것이다.</p> <ul style="list-style-type: none"> • 구체적으로는 인문학 연구에서 활용될 수 있는 주요한 디지털 컬렉션을 만들어내고 병합하기 위한 혁신적인 방법을 개발해내는 것. • 인문학 연구와 관련되는 디지털 리소스를 처리하기 위한 도

91) https://www.dfg.de/en/research_funding/principles_dfg_funding/digital_turn/conference/index.html

92) <http://digitalitaet-geisteswissenschaften.de/veranstaltungen/symposienreihe>

명칭	내용
	구, 방법, 기술을 개발하는 것 • 국제 협력 및 인문학 성과 확산을 용이하게 하는 학문적 소통을 위한 새로운 디지털적인 양식을 만들어내는 것 • 인문학 연구 과제에서 생성된 디지털 데이터를 효과적으로 다루기 위한 도구를 개발하는 것
e-Research-테크 놀로지:94)	
일반 리서치 그랜트	
Excellence Strategy	국제적으로 경쟁력 있는 분야에 대한 펀딩
Priority Programme	참여연구자들 간 전국 단위의 협업을 촉진하고자 함: “Digital Image”, “Computational Literary Studies”, “Robust Argumentation Machine” 등이 이 프로그램에서 지원받은 프로젝트의 예

93) http://www.dfg.de/en/research_funding/programmes/infrastructure/lis/international/dfg_neh_humanities

94) http://www.dfg.de/foerderung/info_wissenschaft/2016/info_wissenschaft_16_36/index.html

2.3.4. 지원의 성과들

2.3.4.1. 미국의 디지털 인문학 발전

■ 연구 분야 (프로젝트)

- 디지털 인문학의 발전을 선도하고 있는 미국은 국가의 정책적으로 디지털 인문학을 체계적으로 지원하고 발전시키고 있는 유럽 대륙의 국가들과 달리 자생적으로 발전해왔다.
- 주로 대학 내의 센터들을 중심으로 인문학자들과 사서들(librarians) 및 프로그래머들(programmers)의 협업으로 진행되었고 주로 프로젝트 단위로 추진되면서 인문학의 각 분야의 연구와는 별개로 진행되는 경우가 많았다.
- 디지털 인문학이 아직 생소한 분야였던 초기에는 학교 내외의 다양한 지원을 받아서 진행했는데 앤드류 멜론 재단이 1994년부터 디지털 인문학에 지원을 하기 시작했고 미국인문학재단이 2007년부터 지원을 시작함으로써 디지털 인문학이 본격적으로 발전할 수 있는 기반을 마련해주었다. 현재 미국 대학들에 다수의 디지털 인문학 센터들이 자리를 잡고 활발하게 활동을 이어나가고 있다.
- 이러한 디지털 인문학 센터들이 발전시켜온 지식과 노하우, 그리고 인력들은 2010년대에 들어와 교육 분야로 확산되기 시작하면서 디지털 인문학교육을 발전시키고 있는데 이러한 발전은 다시 디지털 인문학의 프로젝트와 연구 분야 발전의 중요한 자양분이 되는 선순환이 진행되기 시작했다.

■ 교육 분야

- 2010년대 초부터 앤드류 멜론 재단이 주요 대학들의 디지털 인문학 교육 분야에 대규모 투자를 진행하고 있는 것은 이러한 디지털 인문학의 확산과 맞물려 있다. 즉, 관심있는 일부 센터들 중심으로 발전의 토대가

어느 정도 마련된 상황에서 디지털 인문학이 보다 광범위하게 확산되기 위해서는 주요 대학들의 교육 프로그램과 접목되어 보다 체계적으로 디지털 인문학 교육을 해야 할 필요가 있다. 디지털 인문학 교육 프로그램을 이수한 학생들과 연구자들이 다양한 분야에서 연구와 프로젝트를 진행하게 되면 디지털 인문학의 저변이 더욱 확대될 수 있는 것이다.

- 따라서 2013년부터 앤드류 멜론 재단이 디지털 인문학 교육 프로그램 및 학위 과정을 만드는데 대규모 지원을 했고 이러한 지원이 결실을 맺고 있다. 2015년에 실시된 디지털 인문학 현황 연구에서 조사했을 때 미국에 26개의 디지털 인문학 센터가 있었고 이후 그 수가 지속적으로 증가하고 있다.⁹⁵⁾
- 미국의 디지털 인문학 지원은 그 개념조차 생소했던 초기부터 이루어졌고 최근 확산과 발전의 단계에서 필요한 교육 분야에 대한 적절한 지원이 이루어지고 있어서 미국의 디지털 인문학의 발전에 중요한 역할을 하고 있다.

2.3.4.2. 영국의 인문학-빅데이터 관련 제도적 발전

- 2013년, 영국도서관(British Library)은 앤드류 멜론 재단(Andrew Mellon Foundation)의 펀딩을 받아 영국도서관 랩(British Library Labs)을 설립했다. 목적은 영국도서관이 소장한 방대한 디지털 컬렉션 및 데이터를 더욱 획기적으로 사용하는 것을 촉진하기 위함이었다. 2015년부터 그 디지털 컬렉션과 데이터를 이용한 연구 결과물에 대하여 매년 시상해오고 있다.
- 2014년 3월 영국 정부는 앨런 튜링 연구소(Alan Turing Institute)를 설립할 계획을 선포했다. 목적은 방대한 데이터 셋(large data sets)을 수집하고 분석하는 것으로 2015년에 출범했다.

95) Chris Alen Sula, S. E. Hackney, Phillip Cunningham, "A Survey of Digital Humanities Programs," *The Journal of Interactive Technology & Pedagogy* 11 (May 24, 2017)

- 대학의 교수직

디지털 인문학의 제도적 확산과 정착은 대학의 교수채용에서 관련 포스트가 점차 증가하는 추세에도 반영된다. 현재 영국 대학에 있는 “디지털 00”라는 타이틀의 교수 자리를 연구자가 살펴본 바에 따르면 64 개 이상의 디지털 인문학 교수직들이 있음을 추산한다. 공식적인 통계나 집계 없이 찾아볼 수 없는 경우가 있을 수 있음을 염두에 두어야 한다. 교수자리 중 중대수는 “디지털 인문학 교수”라는 포괄적인 직함이나, 교수들은 각자 역사학, 고고학, 문화유산, 철학, 문학 등 구체적인 전공 분야가 있다. 그 외 정보학 및 지리학을 전공으로 하는 교수도 디지털 인문학 연구소 등에 소속되어 있는 경우가 많다.

염두에 둘 것은 대학이 “디지털 인문학(혹은 기타) 교수”라는 타이틀의 채용을 점차 늘리고 있는 추세를 둘러싸고 있는 더 큰 맥락이다. 각 대학의 디지털 인문학 연구소나 프로그램에 참여하고 있는 교수들 중에는 직함 상으로는 ‘고고학 교수’, ‘근대사 교수’와 같지만 각자의 분야에서 디지털 기술과 방법론을 연구에 이용하는 교수진들이 훨씬 많음을 고려하면 디지털 인문학은 제도권 아카데미 내에 빠르게 확산, 정착되고 있음을 알 수 있다. 영미권 뿐만 아니라 “Professorship in digital humanities”로 구글 검색을 했을 때 독일, 네덜란드, 덴마크, 핀란드 등의 대학들에서도 채용 공고를 발견할 수 있다. 직함에 디지털 인문학이 들어가지 않더라도 디지털 방법론을 조건으로 내걸거나 선호하는 경우도 있었다. 예를 들어 영국 왕립 역사학회에 올라왔던 안트베르펜 대학의 박사 후 연구원 채용 공고는 중세 후기에서 근대 초의 미래상이라는 주제를 제시하며 연구자는 “멀리서 읽기(distant reading)” 방법과 “가까이 읽기(close reading)” 방법을 모두 사용하여 연구를 수행해야 한다는 조건을 적시했다.

위와 같은 양적 팽창을 명백한 추세로 볼 수 있기는 하다. 하지만 이러한 서구의 사례를 통해, 채용 후보자가 뚜렷한 인문학적 전문분야 없이 디지털 기술을 이용할 수 있다는 점에 근거해 우후죽순으로 자리를 늘리는 것은 경계해야

할 것이다. 예를 들어 케임브리지 대학교는 2019년 하반기에 대학 최초의 ‘디지털 인문학 교수’를 채용했다. 채용된 교수의 프로필을 살펴보면 영문학에서 시작하여 과학 소설, 유토피아니즘, 미디어 비평, 인공지능으로 차근차근 연구 분야를 넓히고 서섹스 대학 디지털 인문학 연구소 등을 거쳐 케임브리지 디지털 인문학 교수로 부임하였다.

	철학	고고학	역사학	기타
University of Sussex (5)			디지털 역사학	디지털 인문학
University of London (2)				디지털 인문학 디지털 고전학
University of Cambridge (1)				디지털 인문학
UCL (3)				디지털 인문학 디지털 정보학
University of Edinburgh (1)				디지털 문화유산
Durham University (1)				디지털 인문학
University of Glasgow (5)		디지털 고고학 디지털 공간		디지털 인문학 디지털 문화유산
Queen Mary (1)				디지털 인문학
University of Birmingham (2)	디지털 철학			디지털 문화
Lancaster University (1)				디지털 인문학
University of York (1)		디지털 고고학과 문화유산		
King's College London (41)				*디지털 인문학과에 소속된 교수진은 41명

2.3.4.3. 독일의 경우

■ 기관들끼리의 네트워크의 사례: 독일디지털도서관(Deutsche Digitale Bibliothek)

자료가 디지털화된다는 것은 첫 번째 걸음일 뿐이다. 이 디지털화된 자료가 학문적으로 이용될 수 있도록 하기 위해서는 다양한 연구소들끼리 서로 이어져서 가능한 많은 연구자들이 가능한 많은 자료에 접근할 수 있어야 한다. 한 기관이 자기 자료를 가지는 것으로는 불충분하며 거대한 네트워크가 만들어져야 하고 데이터들을 서로 나눌 수 있어야 한다. 그리하여 디지털화는 항상 협력을 의미할 수밖에 없다.

아울러 이 모든 계획들은 적합한 기술력 없이는 실현될 수 없다. 따라서 디지털 인프라를 만들어내는 것이 중요하며, 독일의 경우 이를 위해 “독일디지털도서관(Deutsche Digitale Bibliothek, 이하 DDB)”이 만들어졌다. 이 과제에 참여한 파트너들은 특정한 자료 및 내용에 대한 찾기가 보다 간단하게 되도록 하는 것, 실행된 프로젝트들이 지속적으로 이용가능하게 만드는 것 등을 중요한 과제로 삼고 있다.

2020년 현재 DDB는 가상 도서관으로서 총 2100여개의 도서관, 문서고, 박물관, 미디어 파크 및 연구소가 DDB 산하에 등록되었고, 3만 개에 달하는 문화 및 학문 관련 기구들이 연계되어서 이 공통의 플랫폼을 활용하고 있다. DDB를 유럽차원에서 보자면, 전유럽연합의 가상도서관인 유로피아나(Europeana)에 연결되어 있기도 하다.

DDB가 위로는 유로피아나에 연계되어 있지만, 그 산하에 매우 다양한 자료 포털과도 연계되어 있다. 대표적으로 신문포털(DDB-Zeitungsportal)을 들 수 있다. 독일처럼 저널리즘이 발전한 경우 특히 엄청난 자료들의 활용이 가능하다. 현재 1600년부터 1945년에 이르는 시기 동안 출간된 21,583 종에 달하는 다양한 신문들이 1654개 기관에 흩어져 있는 상황이다. 이러한 상황을 극복하기 위해 “DDB-Zeitungsportal”이 만들어졌다. 베를린 국립도서관(Staatsbibliothek zu Berlin), 드레스덴의 작센 연방주 도서관과 드레스덴 국가 및 대학 도서관,

프랑크푸르트의 정보인프라를 위한 라이프니츠 연구소와 칼스루에의 FIZ Karlsruhe 등 프로젝트 파트너들이 이에 참여하고 있고, 독일연구재단(DFG)이 이를 후원하고 있다. 이들은 향후 이 신문 자료 전체를 스캔해서 DDB (Deutsche Digitale Bibliothek)를 통해 이용가능하게 하는 것을 목표로 하고 있다. 디지털화된 자료가 디지털화된 형태로도 각각의 소장 기관에 남아있을 지라도 메타데이터의 가공은 DDB를 통해서 이루어지도록 계획되어 있다. 이 경우 Zeitungsportal에 필요한 기술 개발을 개별 기관이 아니라 계속해서 DDB의 인프라를 통해서 해결할 수 있는 것은 큰 장점으로 꼽히고 있다.

■ 지역 내에서의 연계

2020년 현재 독일역사학회 산하조직인 디지털인문학 연구그룹에서 언급하고 있는 독일 내 디지털 인문학 연구소군은 14개에 달한다. 가령 뮌헨의 경우 뮌헨 시내의 바이에른 아카데미, 뮌헨 대학 등의 연구소들이 “디지털 인문학 뮌헨”으로 묶인 결과 1개의 연구그룹을 구성하여 단일한 하나의 홈페이지를 운영하고 있는 상황이다. 따라서 디지털 인문학 연구의 지역 거점이 14개인 것으로 판단할 수 있는 셈이다. 지방분권이 강한 독일의 경우, 이 지역 거점을 발판으로 지역에 따라 특화된 방식의 디지털 인문학 연구가 이루어지고 있다는 점이 특징이다.

Berlin	Weizenbaum Institut für die vernetzte Gesellschaft
Berlin	Berlin-Brandenburgische Akademie der Wissenschaften (BBAW): Interdisziplinärer Forschungsverbund Digital Humanities in Berlin (if DH b)
Erlangen	Friedrich-Alexander-Universität Erlangen-Nürnberg: Interdisziplinäres Zentrum [IZdigital] – Digitale Geistes- und Sozialwissenschaften an der FAU
Frankfurt am Main	LOEWE-Schwerpunkt „Digital Humanities Hessen“
Giessen	Arbeitsstelle Digital Humanities am Zentrum für Medien und Interaktivität

Goettingen	Göttingen Centre for Digital Humanities (GCDH)
Koeln	Institut für Dokumentologie und Editorik (IDE)
Koeln	Cologne Center for eHumanities (CCeH)
Koeln	Historisch-Kulturwissenschaftliche Informationsverarbeitung
Mainz	Mainzer Zentrum für Digitalität in den Geistes- und Kulturwissenschaften (Mainzed)
Muenchen	Digital Humanities in München (dhmuc)
Passau	Passau Centre for eHumanities
Tuebingen	eScience Center Tübingen
Wuerzburg	Julius-Maximilians-Universität Würzburg

■ 대규모 연구인프라: CLARIN-D + DARIAH-DE = CLARIAH-DE

CLARIN-D는 언어가 연구에서 중심적인 역할을 하는 인문사회과학 연구자들을 위한 연구 네트워크를 구축하는 것을 목표로 하는 대규모 연구 인프라이고, DARIAH-DE는 교육, 연구, 연구데이터 및 기술 구성 요소 등 네 가지 축으로 구성된 연구인프라를 통해서 디지털적인 방법으로 행하는 인문 및 문화 과학 연구를 지원해왔다. 최근 두 네트워크의 인프라를 CLARIAH-DE에 통합하는 작업이 이루어지고 있는데, 이것은 독일과 유럽의 인문학, 문화 및 사회 과학에 대한 디지털적인 공동 연구를 위한 공통 공간을 마련하고자 하는 시도로 이해될 수 있다.

■ 교수채용

독일의 디지털 인문학 분야에는 2020년 1월 현재 정년, 비정년 트랙을 포함하는 79개의 교수 자리가 마련된 상태이다. 디지털 인문학 전반적인 분야는 그 가운데 29.7%, 언어학 및 문학 분야의 경우 40.6%, 그리고 다른 분야 및 대상

을 아우르는 것이 29.7%로 분석되고 있다. 디지털인문학이 대체로 디지털문학과 동일시되고 있는 것을 고려하자면, 실제로는 문학 분야가 디지털인문학 분야에서 주도적이라 볼 수 있을 것이다.

이 채용의 구성을 표로 만들면 다음과 같다.

	언어학	고고학	역사학	문헌학 및 기타
베를린자유대	컴퓨터문헌학/고대 세계언어의 데이터 과학	고고정보학		디지털인문학
베를린공대				Digital Provenance
뉘른베르크			고고학과 오리엔탈리 스틱을 위한 디지털인문학	컴퓨터문헌학
프랑크푸르트	(1) 텍스트 기술/컴퓨터 언어학 (2) 응용컴퓨터 언어학			
빌레펠트	(1) 컴퓨터 언어학 (2) 디지털 언어학		중세사/디지털인문학	
다름슈타트	(1) 언어기술학 (2) 디지털 문헌연구 (3) 독일언어학: 코퍼스와 컴퓨터 언어학 (4) 디지털문헌학-근대독일문학			(1) 컴퓨터문헌학/인문학 컴퓨팅 (2) 인문데이터 과학과 방법론
밤베르크		고고지리학의 정보처리	(1) 디지털 역사학 (2) 문화유산 보존에서 디지털 기술/디지털기념 기술	

	언어학	고고학	역사학	문헌학 및 기타
라이프치히	텍스트마이닝			(1) 컴퓨터인문학 (2) 컴퓨터 문헌학 (3) 컴퓨터이셔널 인문학
트리어	컴퓨터언어학			디지털인문학
괴팅엔	디지털인문학과 독일문학	고전 고고학과 그 디지털 방법론		(1) 디지털 인문학 (2) 디지털도상학
파사우				디지털 인문학
슈투트가르트	(1) 컴퓨터언어학 (2) 컴퓨터언어학			디지털인문학
퀘른		고고정보학		(1) 디지털인문학/ 역사문화적 정보처리 (2) 디지털인문학/ 언어정보처리 (3) 디지털 인문학/ 언어정보처리
퀘른공대				디지털자료에 대한 오픈 역세스와 처리
파더본	컴퓨터언어학 기초		디지털문화학	음악학/ 디지털음악편집/ 디지털인문학
마인츠				
포츠담	컴퓨터언어학 기초			
아헨				컴퓨터기반 인문 사회과학 방법론과 이론
할레				eHumanities
부퍼탈			중동지역에 중점을 둔 디지털 민 비주얼역사학	
기센			디지털역사학	

	언어학	고고학	역사학	문헌학 및 기타
에어랑엔- 뉘른베르크	코퍼스 와 컴퓨터언어학		문화사에 강조점을 둔 디지털인문학	
에나				(1) 도상학에 근거한 디지털인문학 (2) 기계학습과 언어기술에 초점을 둔 디지털인문학
마인츠전문 기술학교				디지털인문학
베를린공과 대학				
헴니츠공대				디지털인문학
로스톡				디지털인문학
윈스터			역사학과 문화학에서 디지털인문학	
헴니츠	연문학 및 디지털 언어학			
윈헨			디지털문화사/ 디지털 도상문화	
울덴베르크				학문과정과 디지털 미디어
힐데스하임				디지털인문학
베를린훅볼트			디지털역사학	

2008년에 최초로 디지털 인문학 교수채용이 이루어지기 시작한 이래 현재까지 30여개 대학에서 디지털인문학 관련 교수자리가 만들어진 상태이다. 특히 초기에는 주로 언어학 분야를 중심으로 자리가 만들어졌지만, 2018년부터 디지털 역사학 분야의 교수자리들이 만들어지고 있는 상황이다. 언어학, 고고학,

역사학 등으로 분야를 특정하지 않고 “디지털인문학”이라는 큰 범주로 교수자리가 만들어진 경우도 많다. 2019년에 신설된 경우들을 보면, 실제로 채용이 이루어지지 않은 경우가 많아서 그 분야 전문가를 찾기가 어려운 현실을 반영한다고 볼 수 있을 듯하다.

이처럼 디지털인문학으로 특화되지는 않았다고 하더라도 “디지털 인문학분야를 특별히 고려하는 역사학” 혹은 “디지털 인문학 분야에의 지식이 선호된다” 등의 추가조건이 붙은 경우까지도 고려한다면, 디지털 인문학 분야를 연구하는 독일 대학 교수들의 숫자는 훨씬 많다고 볼 수 있다. 훔볼트 대학, 빌레펠트 대학, 뮌헨 대학 등 역사학이 강세인 대학들에서 디지털 역사학 분야 교수를 채용하고 있는 경향 역시도 눈여겨볼 만한 대목이다.

그렇다면 디지털 인문학 분야에서 실제 교육은 어떤 방식으로 이루어지고 있는가? 디지털 인문학 분야 교육은 주로 석사과정에 개설되어 있고, 학부과정부터 디지털인문학이 개설되는 경우는 적다. 드물게 학부과정이 있는 라이프치히 대학의 경우 2016년부터 학부 과정이 개설되기 시작했다. 학과의 홈페이지에 따르면 “전통적인 문헌학과 디지털적인 방법의 교차점에 서서 고전학 전반에 대한 이해를 증진시키는 것을 목표”로 한다. 디지털 문헌학이 강세인 프로그램으로서 고전 아랍어, 그리스어, 라틴어 등을 포함하는 언어 교육을 포함한다는 점에서 특징적이다. 학부과정 프로그램 구성을 보면 다음과 같다.

프로그램	학점
디지털 인문학	25LP
인문학	60LP
컴퓨터공학	65LP+10 LP 전문화
학부 논문	10LP
핵심 자질 모듈	10LP

즉, 인문학 분야와 컴퓨터 공학 복수전공의 형태라고도 볼 수 있겠지만, 디지털 인문학 관련 과목들을 여럿 수강하게 함으로써 인문학과 컴퓨터 공학 사이의 연결을 용이하게 하는 것이 특징이다.

학과에서 내세우고 있는 직업전망으로는 교육분야, 미디어분야, 그리고 문화유산 처리와 관련되는 분야에서 자연언어처리, 정보처리, 소프트웨어 개발 분야의 업무 등이다. 결국 전통적으로 인문학 전공자가 취업하는 분야와 다르지 않으나, 결과적으로는 디지털과 관련되는 업무능력을 증진시킨 것으로 판단된다.

2.4. 역사학자들의 대응 양상

2.4.1. 미국역사학회의 대응

2.4.1.1. 초기 발전 과정

미국의 디지털 인문학 발전과정 초기 단계에서 역사학계의 집단적인 대응은 없었다. 개별 역사가들 중에서 관심이 있는 소수의 역사학자들이 다른 분야의 전문가들과 함께 협업을 통해 프로젝트를 진행하였다. 이 때 프로젝트들은 아카이브 구성이나 웹사이트 제작, 교육 콘텐츠 제작 등에 집중되어 있었고 전통적인 역사학 연구와는 직접적으로 관계가 없었기 때문에 사실상 역사학계 밖에서 진행되었다고 볼 수 있다. 다만 2001년부터 미국사학회가 “웹사이트 리뷰(Website Review)”를 통해 이러한 흐름들을 소개하는 작업은 하고 있었다는 사실을 보면 역사학계에서 디지털 역사학에 기본적인 관심은 가지고 있었음은 알 수 있다.

2.4.1.2. 2010년대 이후

이러한 양상은 2010년대 들어와서 변화하게 된다. 학계 밖에서 활동해오던 디지털 역사학자들이 전통 역사학자들과의 교류와 소통을 강조하기 시작했던 것이다. 소위 디지털 역사가들이 “디지털화 프로젝트”가 중심이었던 기존의 흐름에서 더 나아가 “학문으로서의 디지털 역사”를 강조하기 시작했다. 따라서 디지털 역사 결과물들을 적극적으로 전통 역사학 저널에서 소개할 것을 권장하고 2017년에 “디지털 역사로 주장하기”라는 워크숍을 통해 디지털 역사와 전통 역사의 접점을 찾으려는 시도도 진행했다.

이러한 흐름에 반응하여 전통 역사학계에서도 디지털 역사학에 대한 보다 적극적인 대응을 하기 시작했다. 이를 위해 2014년에 미국역사학회는 미국 최대의 역사학 연례학술대회에서 “디지털 역사 시작하기(Getting Started in Digital History)라는 워크숍을 개최하기 시작했고 미국역사학회는 홈페이지 내에서 다양한 디지털 역사 프로젝트들을 소개하는 페이지를 추가했다. 또한 디지털 역사 연구를 심사할 수 있는 기준도 마련하여 젊은 디지털 역사가들이 자신들의 활동을 인정받을 수 있는 길을 열어주기 시작했다.

2.4.2. 영국역사학회의 대응

2.4.2.1. 영국의 왕립 역사학회(Royal Historical Society)의 대중 역사학 시상

영국 왕립 역사학회는 1868년에 설립되어 영국 역사학자들에게 각종 지원과 함께, 출판, 세미나, 컨퍼런스 등을 주최하고 홍보하고 있다. 빅데이터 및 디지털이라는 주제에 대하여 별도의 컨퍼런스나 워크숍을 개최하고 있는 것은 아니지만 전반적인 동향을 따라가고 있다고 할 수 있다.

특히 왕립 역사학회는 2015년과 2018년에 ‘대중 역사학’에 관련된 많은 분야에 대하여 상을 수여하였는데, 여기서 디지털은 영화, 다큐멘터리, 박물관 및 전시회, 라디오 및 팟캐스트 등 많은 분야들과 함께 대중 역사학 분야에서의 주요한 통로로 인식되고 있음을 알 수 있다.

2015년	웹&디지털 분야	영국도서관의 ‘과학의 목소리들(Voices of Science)’ 영국도서관은 1940년대 이후 다양한 분야(환경 과학, 컴퓨터, 해양학, 물리학 등)에 걸친 과학사의 중요 장면들을 과학자들을 인터뷰한 구술사 기법으로 구성했다. (다음의 홈페이지에서 확인 가능. https://www.bl.uk/voices-of-science)
2018년	온라인 분야	러니메드 트러스트(Runnymede Trust)의 ‘우리의 이주 이야기(Our Migration Story)’ 이 웹사이트는 학교에서 교육을 일차 목적으로 러니메드 트러스트와 케임브리지 대학, 그리고 대학 연구자들의 협업으로 만들어졌다. 브리튼 섬의 역사 시기 전체에 대해, 중요한

		<p>역사적 장면들과 이주를 접하게 구성했다. 세부적으로는 각 시기와 테마들에 대해 이용가능한 온라인 리소스를 연결했고, 이 리소스들은 이주사에 대한 대표적인 디지털 프로젝트 결과물들을 반영하고 있다는 점에서 주목할 만하다. (https://www.ourmigrationstory.org.uk/)</p>
--	--	---

2.4.2.2. 컨퍼런스

디지털 인문학 중에서도 디지털 역사학을 주제로 열린 컨퍼런스는 2015년에 케임브리지 대학에서 열린 <빅데이터를 인간적으로 만들기: 디지털 시대의 역사학 (Making 'Big Data' Human: Doing History in a Digital Age)>을 들 수 있다. 하루 동안의 이 컨퍼런스에서는 다음의 몇 가지 예시와 같이 빅데이터 웹 아카이브와 연구 성과, 디지털 방법론이 소개, 논의되었다.

- 역사학 연구를 위한 빅데이터: 한사드에서 영국 웹 아카이브까지 (Jane Winters, Institute of Historical Research)
- 역사 연구에서 빅데이터의 한계와 그것을 어떻게 극복할 것인가 (Pim Huijnen, Utrecht University)
- 의학사 케이스북 연구 프로젝트 소개 (Lauren Kassell and Michael Hawkins, University of Cambridge)
- 디지털 인문학과 가족 인물연구: 아일랜드 이민자 가족 연구 (John Herson, Liverpool John Moores University)
- 텍스트 분석과 GIS를 이용한 19세기 신문 상의 공중 보건 표현 (Catherine Porter, Lancaster University)

디지털 인문학 전반을 주제로 한 컨퍼런스로는 매 2년마다 셰필드 대학 (University of Sheffield)이 주최하는 <디지털 인문학 컨퍼런스>를 들 수 있다. 2012년에 시작되어 올해 9월에 열릴 5번째의 컨퍼런스를 준비하고 있다. 역사학, 언어학, 고고학 등 인문학 내 다양한 분과에서 디지털 방법론을 이용한 프로젝트 연구 성과를 공유할 뿐만 아니라 기술적 제도적 측면에 대해서도 논의

한다. 그 동안의 프로그램에 대해서는 다음 홈페이지에서 확인 가능하다.
(<https://www.dhi.ac.uk/dhc2020/>)

2.4.3. 독일 역사학회

1895년에 만들어진 유서깊은 독일역사가협회(Verband der Historiker und Historikerinnen Deutschlands)는 디지털 인문학에 대한 관심이 증가하게 됨에 따라 역사학회 내에 디지털인문학 연구 그룹(Arbeitsgruppe Digitale Geschichtswissenschaft)을 만들었다. 이들은 2년마다 열리는 역사학대회에서 독립된 세션으로 발표를 할 뿐만 아니라, 독자적으로 매 2년마다 “디지털 역사학” 대회를 개최하고 있다.

2020년에 열리게 될 53회 독일 역사가 대회에도 디지털 역사학에 대한 섹션들이 있다. 2020년 역사학 대회의 모토가 “해석투쟁”이기 때문에, 디지털 역사학 섹션의 연장선 상에서, 다양한 디지털 도구들과 방법론, 그리고 그 활용에 대한 조망을 제시하는 것을 목표로 하고 있다.

그런가 하면 2020년 3월에는 “디지털 역사학 대회”역시도 개최될 예정이다. 2020년 학회에서는 세 가지 주제영역에 대해 다음과 같은 토론을 제안하고 있으며, 이는 향후 디지털 역사학이 진행되어 가는 과정에서 항구적으로 논의되어야 할 주제일 법하다.

먼저 연구과제들과 관련하여 디지털적인 방법의 잠재력이 측정되고 비판적으로 조명되어야 한다. 그 경우 디지털적인 방법이 역사학 발전에 어떤 기여를 할 수 있는지에 초점이 놓여져야 한다고 본다.

두 번째로는 역사학적인 연구, 그 주제, 실제 및 소통구조 혹은 출판 구조가 디지털 시대에서 어떻게 변화되는지, 그리고 디지털과 고전적인 해석학적인 방법론의 관계를 어떻게 새롭게 정립할지에 대한 논의도 필요하다.

마지막으로 디지털적인 변화가 학문후속세대 및 교육, 그리고 연구를 촉진하는데 있어 어떤 영향을 미치는지에 대해서도 논의되어야 할 것이다.

2.5. 디지털 인문학 연구의 미래

대다수의 디지털 인문학자들은 여전히 유보적인 태도를 취하고 있다. 디지털 인문학 연구가 투입되는 노동량에 비해 결과로서 나오는 서사는 단순하고, 그러한 정도의 결론은 기존의 작업방식인 역사가의 직관으로도 얼마든지 도달가능하다고 보는 입장이 대표적이다. 이에 대해 디지털 인문학자들은 디지털 인문학에 대한 근본적인 오해에서 비롯된 것으로 파악하고 있다. 독일의 대표적인 디지털 인문학자인 뤼디거 홀스(Ruediger Hohls)는 일반 역사가들이 디지털 인문학에 대해 기존의 문제제기를 위한 쉽게 투입될 수 있는 도구로 생각할 뿐, 방법론적이고 해석학적인 함의를 가지지는 않은 것, 즉 독자적인 하나의 분과 학문으로 이해하지 않고 있다는 점을 지적하고 있다. 그러나 그 자신 역시도 디지털 인문학이 기존의 역사학에서 제기되지 않는 어떤 문제를 다루고 있는지에 대해서 답하는 선으로까지 나아가고 있지는 않아서 아쉬움을 남긴다.⁹⁶⁾

현재까지 서구 선진국들에서 이루어진 디지털 인문학 연구 성과들을 살펴보면 감춰진 패턴들을 드러내주고, 고전적인 텍스트들이 새로운 방식으로 해석될 수 있도록 하며, 그 결과로 문서 사료 만으로는 파악할 수 없었던 측면을 새롭게 드러내주는 부분이 분명 존재한다는 점을 부인할 수는 없어보인다. 즉 디지털 인문학이 새로운 문제제기를 하고 있기보다는 새로운 증거들을 만들어 냄으로써 전통적인 내러티브에 대해 다시 생각하기 위해 디지털 인문학을 활용하고자 한다는 것을 보여주고 있는 듯하다.

뿐만 아니라 디지털 인문학의 성과에 대해서는 다양한 판단이 존재하더라도, 디지털 인문학이 학자들이 서로 어떻게 상호작용하며, 어떻게 사료에 접근하는지, 어떻게 지식 생산물을 만들어내고 어떻게 가르칠지 등 학문 생산 구조의 근본을 변화시키게 되리라는 점에 대해 부정하는 사람은 별로 없다. 결국 현재까지 디지털 인문학의 성과에 대해서는 회의할망정, 그것이 가진 잠재력, 즉 미래에 대해서는 의구심을 갖지 않는다고 볼 수 있는 것이다.

어떻든, 새로운 연구대상 및 현상이 확인되고, 새로운 학문분야가 등장하며,

96) <https://guides.clio-online.de/guides/arbeitsformen-und-techniken/digital-humanities/2018>

이에 대한 구조화가 이루어지고, 새로운 방법론과 탐구방식이 개발되는 수순을 밟아온 것이 새로운 학문 분야와 관련한 지금까지의 역사적 경험이다.⁹⁷⁾ 그리고 디지털인문학의 경우 이와 같은 수순을 이미 밟아가고 있는 상태이다. 영국과 독일의 경우를 보았을 뿐이지만, 두 나라 모두 60~80에 달하는 독자적인 교수직이 마련되고, 다양한 방식의 대규모 재정지원, 영국도서관 랩(British Library Labs), 독일디지털도서관(Deutsche Digitale Bibliothek) 등 대규모 인프라 구축을 위시하여 디지털 인문학 분야에 대해 대대적인 지원을 아끼지 않고 있다는 점 역시도 분명한 상황이다.

97) Valuation Studies? Our collective two cents, Valuation Studies(1), 2013: 11-30

3. 디지털 시대의 사회과학

■ 연구분야의 명칭: 디지털 사회과학

디지털-빅데이터 관련 연구를 하는 사회과학의 분야를 지칭하는데에, 디지털 사회과학(digital social sciences), 전자사회과학(e-social science), 컴퓨터학적 사회과학 (computational social science), 인터넷학 (Internet studies), 디지털 문화연구 (digital cultural studies), 사이버문화연구(cyber cultural studies) 등 다양한 용어가 사용되어 왔다. 그러나 최근 들어 “디지털 사회과학 (digital social science)”이 디지털 데이터를 통해 사회학적 연구 주제를 다루는 분야를 전반적으로 지칭하는 용어로 자리잡아가고 있는 추세이다.

■ 사회학에서 디지털 데이터 연구의 계보

“디지털 사회학”이라는 명칭이 최근에 정착되었다는 것이 사회학이 디지털 데이터를 소홀히 다루어왔다는 의미는 아니다. 기실, 사회학에서 디지털 데이터를 다루는 연구는 1990년대로 소급될 수 있다. 초기(1990년대)에는 미디어 사회학자와 문화연구자들이 주도한 사이버컬처 연구가 주를 이루었다. 이 시기의 주된 연구 대상으로는 인터넷의 대중화 이후 새로운 문화현상으로 주목받은 온라인 공동체, 다중 정체성, 하이퍼리얼리티 등을 꼽을 수 있다.

그런가 하면, 2000년대에 이르러서는 ICT가 공간적으로 구현되었고, 그 결과 일상의 대부분의 활동 - 쇼핑, 교육, 교통, 통신, 거버넌스 등 - 이 디지털 데이터에 의해 매개되면서 스마트 시티 (smart cities)에 대한 연구가 새롭게 부각되기에 이르렀다. 2010년대부터는 기존의 통계학과 디지털 연구를 결합한 빅데이터 연구가 크게 주목받게 되었다.

■ 디지털 사회학의 새로운 경향성

디지털 사회학의 연구는 매우 다원화되어있지만, 다음 세 가지 흐름은 기존의 사회학 연구에서 볼 수 없던 흐름이다.

먼저, 통계학자와 기업이 함께 작업하는 경향을 볼 수 있다. 예를 들어 대표적인 디지털사회학의 연구위원회인 “The Social Science Research Council (SSRC) - Digital Social Science Working Group” (<https://www.ssrc.org/programs/component/digital-culture/digital-social-science/>)은 마이크로소프트(Microsoft) 와 일리노이 대학(University of Illinois at Urbana-Champaign)의 도서관정보학(Library and Information Science) 교수 빅토리아 스토티든(Victoria Stodden)이 주도하고 있다.

두 번째 흐름은 담론과 텍스트 중심의 전통적인 경향을 벗어나 시각정보를 중시하는 경향이다. “컴퓨터학적 사회과학 (Computational Social Science)”이 주도하는 이 분야의 선구자는 2007년 설립된 Cultural Analytics Lab의 디렉터인 레프 마노비치(Lev Manovich)이다. 그의 연구실은 전세계적으로 발생하는 다양한 문화현상(셀피의 유행, 청년세대가 즐기는 일본 만화의 이미지 분석, 다양한 문화행사의 빅데이터화 등)을 이미지나 위치데이터의 시각화를 통해서 보여주고 있다.

세 번째로 주목할 점은 글로벌 연구협력이 강화되고 있다는 점이다. 빅데이터 등 글로벌 수준에서의 데이터 수집과 그 영향력을 고려하여 국가 간 연구협력이 다양해지고 있다. 영국 “경제사회연구위원회(Economic and Social Research Council, ESRC)”는 12개의 핵심 연구 노드(node)를 설정하여 주로 시뮬레이션, 데이터의 질적/양적 분석, 데이터 관리, 기술의 사회에 대한 영향력 등에 대한 국제적인 연구들을 수행하고 있다. 가장 최근에 착수된 대규모 국제협력 프로젝트로는 2020년 1월에 시작된 영국과 일본(The Japanese Science and Technology Agency, JST)의 협력 프로젝트를 꼽을 수 있다. AI가 사회 및 경제 분야에 미치는 전반적인 영향력(예를 들어, AI가 의료서비스 및 법적 결정에 미치는 영향력 등)을 파악하고자 하는 목표를 가진 이 프로젝트에 향후 3년 동안 18억을 투입할 예정인 것으로 알려지고 있다.

■ 주요 연구기관 사례

현재 디지털 사회학 분야를 주도하고 있는 두 연구소는 각각 하버드 대학과

옥스퍼드 대학에 있다. 각 연구소에서 진행한 대표적인 프로젝트를 표로 만들면 다음과 같다.

연구소명	대표적인 프로젝트
Berkman Klein Center for Internet & Society at Harvard University (https://cyber.harvard.edu)	학습을 통해 실천을 모색한다는 믿음 하에 학문적 성과와 실제 온라인 상에서 다양한 커뮤니티 활동에 도움이 되는 실효성 있는 결과물을 프로그래밍하는 것을 목표로 하고 있음. 주요 프로젝트 사례 - 하버드 로스쿨의 수업을 듣는 학생들에게 개별 수업의 강의계획안을 오픈하여 판례들을 디지털화하여 제공함으로써 비싼 판례집을 교재로 구입하지 않고도 공유할 수 있도록 플랫폼을 제공함
Internet Institute at University of Oxford (http://www.oii.ox.ac.uk)	디지털 경제, 정보지리학과 불평등, 거버넌스, 디지털라이프와 웰빙, 정보학을 위한 윤리학과 철학, 디지털 지식과 문화, 정보 거버넌스와 보안 등의 연구분야를 다루고 있음. 주요 프로젝트 사례 - <Social Media Narratives of Sustainable Food Consumption> 2019년 1월 - 2020년 1월의 일 년 간 소셜미디어의 텍스트 마이닝을 통해 지속가능한 음식 소비(예를 들어, 육류 소비의 감소, 대량생산 식품에서 대안적 단백질 식량 소비 등)를 하는 개인들의 소비행태를 분류하고 분석하여 세인즈베리(Sainsbury)와 같은 기업에게 대안적 음식 소비 행태에 대한 내러티브를 제공하여 기업의 활동에 변화를 주고자 함)

■ 디지털 사회학 연구가 제기하는 문제들

1. 사회학자들의 통찰력이 왜소화되고 있는 상황

현대 디지털 사회학은 정치학, 커뮤니케이션, 서지학, 정보학, 인류학, 심리학, 경제학, 문화연구, 컴퓨터학 분야 등과 학제간 연계를 추구하고 있으나, 실상은 디지털 데이터를 수집하고 분석할 수 있는 통계학, 서지학 분야가 방법론을 주도하고 있는 상황이다. 그 결과, 디지털 데이터를 수집 및 분석 할 수 있는 직접적인 툴을 학습하지 않은 사회학자들의 통찰력이 주목받기 어려운 구조적 문제점이 점점 심화되고 있다. 디지털 사회

학 내에서 1990년대까지 새로운 문화현상에 대한 통찰력을 제공해왔던 사회학자들의 입지가 점점 줄어들고 있는 형국인 것이다.

따라서 새로운 연구 대상이자 방법론으로서의 디지털 데이터가 사회 전반에 미칠 영향력에 대한 거시적, 통시적 시각을 담은 연구들은 점점 줄어들어 드는 반면, 연구방법론의 단순한 소개와 사례연구에 집중한 연구들이 점차 주목받는 불합리한 상황이 일상화되고 있다. 이러한 문제점을 극복할 방안을 모색하는 일이 시급한 상황이다.

2. 디지털 데이터 성격논쟁: 자료인가, 방법인가, 플랫폼인가

“Trace data”는 개인이 컴퓨터 시스템과 상호작용을 하고 남게 된 자료를 수동적으로 모은 것이다. 전화사용 내역에서부터 지금의 페이스북과 트위터 데이터까지를 망라한다. “Big data”는 “Trace data”의 일부라고 볼 수 있다. 서베이나 인터뷰와 같은 자가 보고식 (self-reporting) 데이터 형성보다 예러가 ‘적고’ 샘플의 수가 ‘방대’하기 때문에 유용한 연구 데이터로 여겨진다. 보통 이러한 데이터는 다음의 네 가지 방식으로 수집된다.

- (1) “scaping”: 웹 페이지와 데이터베이스에 저장된 데이터를 모으는 것
- (2) “API (application programing interface)”: 프로그램의 일종으로서, 사용자와 시스템이 상호작용하도록 만든 소프트웨어를 통해 데이터 모이기
- (3) 트래이스 데이터를 모으기 위해 구현된 프로그램과 소프트웨어를 통해 데이터를 직접 모으기
- (4) 디지털 플랫폼을 제작한 주체를 통해 직접 모으기

전통적인 사회연구 방식으로 인간 행위를 연구하는 것은, 인류학적인 방법과 비슷하다, 즉 연구자가 연구하고자 하는 사회현상에 직접 참여 혹은 관찰하면서 얻은 자료를 분석한다. 그러나, ethnographic practice (인터넷상의 디지털

현상을 주로 연구하는 것을 지칭)는 대개 이미 일어난 현상을 리뷰하는 형식이 되기 때문에, content analysis와 archival 분석과 비슷하다. 이러한 점 때문에, “trace ethnography”라고도 불리기도 한다 (Geiger and Ribes, 2011). (여기까지의 내용은, Hampton, K. N. (2017). Studying the digital: Directions and challenges for digital methods. Annual Review of Sociology, 43, 167-188. 의 내용 중 발췌)

과연 이러한 트레이스 데이터 혹은 빅 데이터는 넓게는 사회과학연구 좁게는 사회학 연구에 획기적 변화를 가져다 줄 것인가? 사회학내에서, 디지털을 바라보는 시각은 크게 세 가지로 나누어진다. “The digital”을 연구 대상으로 바라보는 것, 연구의 도구(tool)와 방법 (method)”, 그리고 학문의 수요자 및 대중과 소통하는 플랫폼 (platform)으로 바라보는 것이 그것이다 (Noortje Marres, Digital Sociology, The Reinvention of Social Research, 2017).

컴퓨터 과학자인 펜트랜드(Alex Pentland)에 의하면, 디지털 데이터는 인간의 상호 행위가 어떻게 네트워크로 연결되어 있는지 보여줄 수 있기 때문에, 디지털 데이터 안에 담겨있는 네트워크를 검토하는 것으로 사회의 컴플렉시티 (complexities of society)를 밝힐 수 있다. 펜트랜드가 의미하는 바는, 다시 말하면, 디지털 데이터를 연구하는 것이 곧 사회를 연구하는 것이 된다는 것이고, 데이터에서 유추된 패턴이 곧 인간 행위의 패턴이라는 것이다. 이러한 접근이 궁극적으로 시사하는 질문은, “그렇다면 디지털 사회학 (혹은 사회과학)은 디지털 사이언스(digital science)와 어떤 차이가 있는가?”이다.

펜트랜드의 제안, 즉 “데이터 자료 안의 네트워크 분석이 곧 사회 복잡성 분석이 된다”는 것은 사회연구의 대상과 방법에 대한 획기적 제안으로 보일 수 있다. 하지만, 이 제안은 고전적 사회학자들이 던진 질문들, 가령 막스 베버가 고찰한, ‘사회적 행위’는 무엇인가? ‘사회를 구성하는 것은 무엇인가?’, ‘인간 행위의 목적은 무엇인가? 인간과 사회에 대한 연구의 본질에 관한 질문과 맞닿아 있다. 즉, 인간의 행위와 사회적 행위는 늘 개인의 행위 목적, 즉 행위가 의미하는 바로서 타인의 이해(interpretation)를 전제하기에, ‘사회성(sociality)’에 기반을 둔다. ‘데이터가 곧 사회다’라는 명제가 타당하다면 ‘데이터의 어떤

면이 사회적인가?’를 밝혀야만 하는 것이다.

디지털 데이터는 디지털 플랫폼을 통해 매개된다. 즉 디지털 플랫폼을 통해 제공된 세상, 즉, 시스템에 인간이 상호 작용하던지, 혹은 인간과 프로그램이 상호작용하는 곳이, the digital이고, 그 행위가 남기는 기록(trace)도 the digital 이기 때문이다. 디지털 데이터는 플랫폼이라는 기반(infrastructure), 시스템과 프로그램, 그리고 인간의 행위가 함께 상호작용하여 만들어진 것이다. 예를 들어 특정 사용자들이 특히 어떤 앱을 많이 사용한다면, 이 현상은 어떻게 해석될 수 있을 것인가? 그 앱을 사용하는 사용자를 연구하는데 유의미한 것인가? 사용자가 이용하는 플랫폼의 자동 권장 시스템이라든가, 랭킹 시스템을 연구하는데 더 유의미한 현상인가? 혹은 그러한 시스템을 고안한 주체를 연구하는데 더 유의미한 현상인가?

보다 더 근본적인 질문으로 치환하자면, ‘디지털 데이터’는 자연스럽게 주어진, 인간행위들의 기록물과 같은 것으로 간주될 수 있는 것인가? 사회학자들이 “the digital”에 주목하고 있는 것은 다음과 같은 의문들이다. 첫째, 우리가 집합적으로 디지털 데이터의 무엇을 ‘사회적’으로 정의하는가라는 것, 그리고 그렇게 우리가 이해한 ‘사회적’이라는 개념이 우리의 ‘사회성’을 재정의하게 될 것인가라는 것이다. 둘째, 디지털 데이터를 생성하는 조건, 플랫폼, 시스템, 프로그램 등이 우리가 어떤 사회적 행위를 해야 할 것인가를 미리 큐레이트(curate)하고 있다는 것이다. 디지털 데이터를 통해 인간 행위를 연구하고자 한다면, 디지털 데이터란 기술과 사회성/인간상호행위(sociality), 그리고 지식이 상호작용하여 이루어진 것이라는 것을 고려해야 한다.

결국, ‘The digital’은 ‘자료’이기도 하고, ‘방법’이기도 하고, 또한 ‘개입’, 즉 ‘플랫폼’이기도 하다. 다시 말해, ‘디지털’로 통칭되는 디지털 데이터 및 디지털 테크놀로지는 사회적 삶의 도구이자, 그 자체로 사회적 행위의 대상이 되고 있다. 따라서, ‘디지털적인 방식으로’ 사회를 이해하기 위해서는 우선 누가 어떻게 어디서 사회연구에 적합한 데이터를 가져와야 하는가를 질문해야 하고 무엇보다 디지털의 어떤 부분이 ‘사회적’인 것인지에 대한 근본적인 성찰이 절실한 것이다.

4. 결론: 종합 분석

인간 사회의 전 분야를 쓰나미처럼 휩쓸고 들어오는 디지털화의 시대는 개인의 생활 방식, 개인들끼리의 교류 방식, 더 나아가 사회가 작동하는 방식을 총체적이고 근본적으로 변화시키고 있는 상황이다. 우리는 현재 디지털을 사용하는 선을 넘어서서 디지털 가운데서 살아가고 있다. 인간의 삶 전반을 디지털이라는 매체가 조건 짓는 이러한 상황은 여러 가지 전례 없는 현상들을 낳고 있으며, 이러한 변화가 긍정적일지 부정적일지는 현재의 우리가 어떻게 대처하느냐에 달려있을 수밖에 없다. 평자에 따라서는 농업혁명, 산업혁명 이상의 역사적인 전환점이라고도 평가되고 있는 현재 시점은 곧 새로운 사회구조의 원리와 더불어 새로운 방식의 학문이 출현하는 시점인 듯하다.

이 글에서는 먼저 개인정보의 문제와 젠더의 문제에 집중하여 디지털-빅데이터 시대가 낳은 사회적 변화의 양상을 조망하고자 했다. 인간의 생활 전반이 디지털에서 흔적으로 남고 있는 현실은 그 흔적, 즉 “trace data”를 그 자체로 “상품”으로 간주하려는 경향과 결부되고 있다. 디지털 사회로의 진입이 이미 이루어졌고 후퇴가 불가능할 뿐만 아니라 디지털-빅데이터 사회가 갖는 장점이 매우 큰 것을 고려할 때 “데이터 공동체”, 혹은 “연대적인 데이터 저장” 등 현재 사회학에서 제시되고 있는 방법들에 대해 성찰함으로써 개인데이터가 “상품”으로 직진하지 않도록 하기 위한 구체적인 방법을 찾아내기 위해 노력할 필요가 절실해 보인다.

젠더 문제의 경우 디지털-빅데이터 시대가 사회 진보의 측면에서 가능성과 한계를 동시에 안고 있음을 선명하게 보여주고 있다. 한편으로 디지털 전환 자체가 젠더에 따라 다르게 경험되고 있고, 디지털 세계에서도 기존의 남성중심적인 구조가 재생산되는 경향이 분명한 것이 현실이다. 그러나 다른 한편으로 빅데이터는 2000년대 이후 정체상태에 머물러 있는 양성평등에 한걸음 더 가까이 갈 수 있는 가능성을 내포하고 있기도 하다. 젠더 격차의 문제를 빅데이터를 통해서 입체적으로 보여줄 수가 있고, 그 결과로 구체적이고 적극적인 개입이 가능하다는 것이 여러 가지 경로를 통해서 드러나고 있는 것이다. 가령 유엔이 클린턴 재단 등 거대 민간 재단들과 공동설립한 Data2X의 경우 기존의 데이

터들을 모으는 데서 더 나아가 이 데이터들을 해체함으로써 걸핍된 데이터, 즉 여성에 대한 데이터를 채우는 것을 목표로 삼고 있다. 이와 같이 빅데이터의 입체성은 인간의 사회가 젠더의 측면에서 얼마나 기울어진 운동장인지를 보다 더 명징하게 보여주고 있고, 디지털을 통해 가능해진 국제협력의 진전을 통해서 과거보다 더 적극적이고 구체적인 방식의 대응을 가져올 가능성을 내포하고 있다.

디지털화가 내포하고 있는 사회 전반의 변화는 분과학문 단위에서 접근할 수 있는 범위를 넘어서 있다. 이러한 문제 상황에 대해 인문학과 사회과학 분야가 각각 어떻게 대응하고 있는지를 II부와 III부에서 살펴보았다. II부에서는 먼저 디지털인문학이 출현함으로써 새롭게 제기되고 여러 쟁점들, 즉 지식 생산에 있어 학자와 대중의 경계 허물기의 가능성, 디지털 인문학이 기존 위계질서의 전복에 기여하는지 혹은 재생산에 기여하는지, 디지털 인문학이 한때 학문 전체를 지배할 듯 보였던 계량경제사(Cliometrics)와 같은 양적연구방법 같은 전철을 밟아가게 될 것인지, 디지털 인문학 연구에 대한 평가 기준 마련 등 디지털 인문학 분야 특수성을 어떻게 다룰지 등의 문제점에 대해 검토하였다.

다음으로 디지털 인문학 연구에서 앞서가고 있는 미국사, 영국사, 그리고 고전학에서 보여준 최근 연구 성과물에 대해 검토하였다. 디지털화의 토대가 되는 데이터가 매우 분명하고 동시에 대량인 고전학 분야의 경우 일반 역사학 분야보다 연구와 교육 양 측면에서 모두 디지털화에 앞서 나가고 있어서 기존의 모든 프로젝트들이 디지털 연구로 전환되는 상황이라고 평가될 정도이다. 자료의 디지털화는 연구방법의 디지털화로 이어져서 지리정보, 네트워크 분석 등 다양한 방식으로 고대 세계를 입체적으로 재현하는 연구가 다양하게 진행 중이다. 이 디지털 연구의 성과는 대중과 공유되기에 용이하기 때문에 디지털화의 진전은 곧 일반 사회와 학계의 소통이 증대되는 방식으로 나아가고 있다.

미국과 영국의 디지털인문학 연구 성과물들에 비추어볼 때 디지털 인문학 연구는 과거에 대한 시뮬레이션을 가능케 하는 입체성을 내포하고 있다. 그 결과 젠더와 빅데이터의 관련성에서 드러났던 것과 마찬가지로, 2차원에 그치는 기존의 문서자료를 통해서서는 드러나지 못하는 “걸핍”에 대한 문제제기를 가능케

할 수 있다는 것을 알 수 있었다. 디지털 인문학이 전통적인 역사학 방법을 대체하거나 능가할 혁신을 기대하는 비평가들에게는 실망스러운 결론일 수도 있겠으나, 학자의 임무 및 가능성이 결국 “가능한 최선의 방법”으로 역사적 실체에 접근하기 위해 노력하는 것이라면 충분한 존재의의를 가진다고 생각할 수 있을 것이다.

다음으로 영국과 독일을 중심으로 각국 정부 및 역사가 단체에서 디지털 인문학 연구에 대해 어떻게 대응하고 있는지를 살펴보았다. 영국과 독일 정부는 디지털 인문학을 대한 회의와 호기심이 공존하고 있는 학계와 달리 매우 과감한 투자를 하고 있음을 볼 수 있었다. 학문 생산을 뒷받침하는 조직과 펀딩을 통해 양국 정부 공히 디지털 인문학 분야를 육성하기 위해 매우 포괄적인 노력을 기울이고 있음을 알 수 있다. 60~80개에 달하는 교수직의 수가 이를 입증한다. 또한 양국의 역사학계 역시도 매우 적극적으로 대응하고 있다. 영국 왕립 역사학회의 경우 디지털 분야에 “대중역사학상”을 수여하고 있으며, 정기적으로 관련 컨퍼런스를 진행하고 있다. 독일역사학회의 경우 “디지털인문학연구 그룹”이 산하조직으로 자리잡아서 2년마다 디지털역사학 학회를 개최함으로써 연구성과들을 학계 및 대중과 공유하고 있다.

사회과학 분야의 논의를 분석한 III부에서는 사회과학에서의 디지털 흡수 노력이 훨씬 더 포괄적이고 근본적인 논의의 차원에 이르러 있음을 보여준다. 포괄적이라 함은 디지털 사회학이 학계와 국가라는 전통적인 틀을 넘어서, 한편으로 마이크로소프트 등 기업이 직접 참여하는 프로젝트가 진행되고 있는가 하면, 다른 한편으로 글로벌한 차원에서의 연구 프로젝트들이 다수 진행되고 있다는 점이다. 근본적이라 함은 디지털 데이터의 성격에 대한 논쟁을 통해서 잘 드러나고 있다. 디지털 플랫폼을 통해서 매개되는 디지털 데이터에 남는 기록들, 즉 “남는 기록들(trace data)”을 사회적인 행위의 기록으로 살펴볼 수 있을 것인가, 그렇다고 한다면 디지털 데이터의 어떤 측면을 “사회적인 것”으로 볼 수 있을 것인가, 즉 보다 근본적으로는 기술과 사회성, 지식의 상호 작용을 어떻게 평가할 수 있을 것인가에 대한 성찰이 시도되고 있다.

독일의 매체학자로 주목받고 있는 프리드리히 키틀러(Friedrich Kittler)는 인

간의 사유는 구체적으로는 “철자, 신호, 정보처리작업, 수학적 계산과 연산자 등 기술적 표준들이 만드는 종합적인 효과”에 불과하다고 말할 정도로 기술에 종속적인 존재로 인간을 그려냄으로 반휴머니즘, 혹은 포스트휴머니즘적인 사유의 전범을 보여주었다. 매우 극단적이기는 하지만 21세기의 호모 사피엔스들 다수가 “로그아웃”이 불가능한 방식으로 생활하고 있음을 고려할 때 그 사유의 조건에 대한 근본적인 성찰과 그에 대한 인적, 물질 투자가 매우 절실한 국면에 있다는 점은 부인할 수 없다.

5. 정책 제언

구체적인 정책적인 제언



- I. 디지털-빅데이터 시대에 젠더 불평등이 재생산되지 않을 수 있도록 하기 위한 의식적인 노력이 필요
 1. 빅 데이터 수집, 활용, 정책 수립 과정 전반에 있어 젠더 문제와 관련한 지속적인 모니터링이 이루어질 수 있도록 해야 한다. 빅데이터를 통해서 글로벌 젠더 문제의 심각성이 더욱 분명하게 드러낼 수 있으리라 기대되고 있다.
 - 1) 빅데이터 시대는 젠더 문제에 대해 새로운 가능성과 위험을 함께 제기하고 있다. 빅데이터로 수집할 수 없는 것, 디지털화될 수 없는 사회적 사실은 이제 존재하지 않는 것이나 마찬가지이다. 따라서 빅데이터를 통해 젠더 불평등의 모습을 '보이게 하고,' 젠더 불평등을 내포하고 있던 기존의 빅데이터를 해체하여 새롭게 재편하려는 노력이 필요하다. 이로부터 빅데이터는 단순히 데이터가 아니라 구체적인 후속 조치와 새로운 문제제기를

촉발하는 “actionable” 데이터가 될 수 있다. 예컨대 2009년 베트남 통계국은 세계보건기구 및 여러 단체와 협업을 통해서 전국적인 가정 폭력 실태 조사를 처음으로 실시했으며, 이 통계에 대한 충격으로부터 여러 가지 정책적, 인식론적 변화가 일어나고 있다. 빅데이터 분석을 통해 여성에게 가해졌으며 침묵되어온 폭력의 실체를 드러내는 것만으로도 문제 해결에 한걸음 다가서게 된 것이라 볼 수 있다.

2) 이는 UN이 클린턴 재단 등 거대 민간 재단들과 공동설립한 Data2X가 주창하는 “젠더 데이터 혁명(Gender Data Revolution)”의 안건과도 일맥상통한다. 이는 빅데이터에 숨은 젠더 불평등을 일깨워 새로운 관점에서 데이터를 수집하고 조직화하는 동시에, 빅데이터를 젠더 평등에 가까이 가기 위한 통로로 활용하는 것을 의미한다. 앞에서 살펴보았듯이 거대 도시의 이동성(mobility)에 대한 빅데이터를 젠더 측면에서 재구성해 보았더니 표면상으로는 알 수 없었던 여성의 교육 기회와 사회 활동에 대한 장애물들(성희롱이나 성폭력의 위협)이 드러났다. 또한 새로운 빅데이터 애플리케이션이 아프리카의 소외된 여성들의 손에 쥐어졌을 때 이들의 역량을 강화하기 위한 도구가 되는 모습도 보았다. 따라서 젠더 평등을 위해서는 단지 빅데이터를 모으는 것(aggregate)뿐만 아니라 해체하여(disaggregate) 그 속의 ‘결핍된 데이터(missing data)’를 찾아내고 채우는 한편, 빅데이터를 통해 정책을 젠더 통합적인 방향으로 전환시키고 사회에서 소외된 여성들의 역량을 키워내야 한다. 젠더 문제가 폭발적으로 제기되고 있지만 이에 대한 백래시가 만만치 않은 한국에서도 이와 같은 데이터 중심의 접근법을 활용해 볼 만 하다.

2. 현재 데이터 세계의 젠더 불평등을 생각할 때 빅데이터 관련 분야에서 여성의 비율을 높이는 것이 중요한 포석일 수밖에 없으며, 여학생 교육 단계에서부터 STEM 분야 유리천장을 없애는 문제에 이르기까지 포괄적인 정책 마련이 필요하다. 이와 관련하여 다음과 같은 모델들을 참고할 만하다.

- 1) 독일의 Go Mint(National Pact for Women in MINT Careers) 프로그램: 2008년 전문과학 분야의 여성 참여를 촉진하기 위해 250명 이상의 정치 경제 과학 미디어 종사자들과 협력하는 국민협약을 말함. 남성중심적인 과학 공동체에서 소외되고 배제당하는 여성 과학 인력들을 네트워킹하여 이들에게 힘을 실어줌. 동시에 과학뿐만 아니라 여성의 과학 참여를 저해하는 사회적 인습과 싸우기 위하여 정치, 사업, 과학, 미디어가 협업하여 과학과 관련된 직군에 대한 왜곡된 이미지를 재고함.
 - 2) 2009년 20개국의 교육부가 모인, 즉 교육자들의 유럽 공동체 “Scientix”: STEM 부문의 과학 분과들에 중점을 두고 유럽스쿨넷의 조정 아래 재정 지원을 받아 여성에 대한 과학 교육을 확충하고자 함. 이 프로젝트는 젠더에 대한 정형화된 스테레오타입과 편견이 어떻게 여성의 학습 기회를 제한하여 이후 여성 과학 인력 감소에까지 이어지는지, 과학과 디지털사이언스 학계와 사업계가 남성중심주의가 어떻게 민주적인 의사결정 과정을 침해하는지, 젠더와 섹슈얼리티에 대한 사회적 편견이 어떻게 과학 연구 자체를 왜곡하는지 잘 보여줌. 이를 통해 진정한 “혁신(innovation)”은 젠더 평등과 함께 가는 것임을 주창.
- II. 디지털 데이터 생성 단계에서부터 인문사회학적인 개입이 이루어질 수 있는 방안 마련 필요
- 2000년대 초반 영화나 음악 등과 함께 인문사회학이 산업의 대상인 ‘문화 콘텐츠’로 협소하게 자리매김됨으로써, 실제 빅데이터나 인공지능 관련 기술들이 사회변동에 미칠 주요 논의에 대한 심도 깊은 논의를 진행시키는 방향으로 참여하지 못했다. 디지털 인문학 연구가 디지털에 압도되어 종속되는 방식으로 이루어지지 않기 위해서는 지속적이며 안정적인 디지털 인문사회학 연구를 수행할 수 있는 환경의 제공이 시급히 요구되는 상황이다.
1. ‘The digital’은 ‘자료’이기도 하고, ‘방법’이기도 하고, 또한 ‘개입’, 즉 ‘플랫폼’이기도 하다. 다시 말해, ‘디지털’로 통칭되는 디지털 데이터 및 디지털

털 테크놀로지는 사회적 삶의 도구이자, 그 자체로 사회적 행위의 대상이 되고 있다. 따라서, ‘디지털적인 방식으로’ 사회를 이해하기 위해서는 우선 누가 어떻게 어디서 사회연구에 적합한 데이터를 가져와야 하는가를 질문해야 하고 무엇보다 디지털의 어떤 부분이 ‘사회적’인 것인지에 대한 근본적인 성찰이 절실하다. 또한 데이터의 생산, 관리, 공유의 과정과 시스템에 대한 공공의 역할과 개인의 역할, 제도적 조치에 대한 고찰이 필요하다. 또한 빅데이터를 가공하여 새로운 지식을 생성할 경우, 이 새로운 지식의 과학성을 누가, 어떻게 점검할 것인가의 문제도 고찰해야 한다. 현재 진행되는 방식은 동료평가(peer review)의 형식으로 자기검열이 이루어지고 있다. 데이터를 다루는 방식과 테크놀로지를 알지 못하는 많은 인문사회과학자들이 빅데이터로 만들어지는 새로운 지식의 자기검열 과정에 어떻게 참여할 것인가의 문제도 고찰해야 한다.

2. 이는 인문학의 경우에도 마찬가지여서, “살아남은 모든 것”을 가지고 연구하던 시대로부터 “모든 것이 살아남는” 시대로의 전환이 이루어지고 있는 상황이다. 대규모로 쉽 없이 생성되는 데이터를 보존하는 데 따르는 제약 때문에 어떤 데이터는 사라지게 된다는 점을 고려하더라도, 디지털 시대의 데이터 규모는 이전 아날로그 시대와는 비할 바 없다. 어떤 사료를 어떤 방식으로 선택할 것인지에 대한 선택이 전적으로 당대인에 의해 이루어지던 데서 현재 역사가들의 영역이 훨씬 더 광범위해질 수밖에 없는 상황에 이르게 된 것이다. 또한 어떤 데이터를 남기고 어떤 것을 사라지게 할 것인가를 결정하는 것이 미래 학자들의 영역에 포함되는 상황 역시 예측 가능하다.
3. 압도적인 규모로, 그리고 무작위로 이루어지고 있는 디지털-빅데이터 시대가 인류 공동체 전체에 유익한 방향으로 나아갈 수 있도록 하기 위해서는 인문사회학자들이 함께 지혜를 모아 데이터 생성 단계에서부터 적극적으로 개입해야 할 필요성이 절실하다. 특히 디지털로 재편된 연구 환경에서

인문사회학자가 연구에서의 주도권을 갖기 위해서는 웹 환경의 메커니즘과 컴퓨터 기술에 대한 최소한의 이해를 갖추는 것을 목표로 해야 한다. 이미 웹 데이터베이스 없이 인문사회학 연구는 불가능한 상황이지만, 우리나라는 물론이고 서구 국가들에서도 구글 등의 민간 업체가 웹 데이터베이스 구축과 운용에까지의 주도권을 갖고 있으며 이에 대한 학계의 문제의식은 겨우 몇 년 전부터야 공유되고 있는 상황이다. 국가 차원에서의 적극적인 지원이 있다 하더라도 그 지원을 효율적으로 사용하기 위해서는 인문사회학자들이 데이터의 발굴, 구축, 해체, 재구축을 주도하고 기술적 과정에 개입할 수 있는 능력을 갖추어야 한다. 이를 통해서만 디지털 기술을 단순한 편의의 도구로 활용하는 것을 넘어 융합적 사고를 통해 디지털-빅데이터에 대한 인식론적, 존재론적 사고의 전환을 기대할 수 있을 것이다.

III. 인문, 사회과학 분야 디지털 인문학 연구 촉진을 위한 방안

1. 디지털 인문사회과학 육성: 디지털 기술과 인문학적인 문제제기, 양자 모두에 능한 인적 자원을 교육하는 것이 필요하나, 최종단계인 교수채용 비용을 갑자기 높이는 방식으로 대응하는 것은 현재 이 분야를 평가할 기준이 제대로 마련되어 있지 않고 평가할 수 있는 인력풀도 매우 협소한 상태에서 다소 위험한 선택일 수 있다. 학계 전체가 중지를 모아서 디지털 인문사회과학을 폭넓게 정의하고 이를 커리큘럼 내에 뿌리내리게 하기 위한 객관적이고 엄정한 기준을 마련할 필요성이 시급하고도 절실하다.
2. 펜과 책상만이 인문학 지식 생산의 필수조건이던 것과는 달리, 디지털 인문사회과학 연구는 근본적으로 협업을 통해서 이루어질 수밖에 없다. 이러한 협업은 학제간 장벽뿐만 아니라 국가간 장벽, 혹은 전문가와 비전문가 간의 장벽을 넘어설 것을 요구하고 있기도 하다. 이러한 협업이 가능하기 위한 기반은 개별 대학, 연구소 차원에서 마련될 수 없는 것이니만큼, 정부 차원의 적극적인 대응이 필요할 수밖에 없다. 미국, 영국, 독일 정부에서 시도하고 있는 구체적인 협업 장려 방식은 다음과 같다.

1) 일반사회와의 소통을 강화할 수 있는 방안 마련

- 대표적으로 미국사의 경우 “The Salem Witch Trials: Documentary Archive and Transcription Project”, “The Geography of Slavery in Virginia” 등 역사학 연구 과정 자체를 대중에게 공개하는 프로젝트들이 진행 중이다. 이는 최종 지식 생산물인 책, 특히 교과서를 통해서 고체화된 역사지식을 전달받아온 대중들에게 역사학 지식 생산 과정 전반에 대한 관심과 이해를 일깨우고 촉진하는데 기여할 수 있을 것이다.

2) 국제적 연구 협력 촉진 방안

- 영국 “경제사회연구위원회(Economic and Social Research Council, ESRC)”는 12개의 핵심 연구 노드(node)를 설정하여 주로 시뮬레이션, 데이터의 질적/양적 분석, 데이터 관리, 기술의 사회에 대한 영향력 등에 대한 국제적인 연구들을 수행하고 있다.
- NEH/DFG 기금: 독일의 연구재단과 미국의 인문학재단이 함께 NEH/DFG 펀딩을 마련하여 독일과 미국 학자들 간 디지털 인문학 교류를 촉진시키고 있다.

3) 디지털 웹 아카이브 구축

- 독일의 경우 독일디지털도서관(Deutsche Digitale Bibliothek)이, 유럽의 경우 유러피아나(Europeana)라는 디지털 도서관이 만들어졌다. 2020년 현재 DDB는 가상 도서관으로서 총 2,100여개의 도서관, 문서고, 박물관, 미디어 파크 및 연구소가 DDB 산하에 등록되었고, 3만 개에 달하는 문화 및 학문 관련 기구들이 연계되어 있고, 그 산하에 신문포털 등 다양한 자료포털과도 연계되어 있다. 이처럼 거대한 공통의 플랫폼을 활용할 경우, 필요한 기술 개발을 개별 기관이 아니라 계속해서 DDB의 인프라를 통해서 해결할 수 있는 것은 큰 장점으로 꼽히고 있다.

4) 초중고 커리큘럼으로 제도화

- 이미 놀이와 학습의 도구로서 디지털 미디어를 삶의 필수 조건으로 경험하고 있는 현재의 초등학생 및 중·고등학교 학생들에게도 디지털 인문학을 일찍 접할 수 있는 교육 시스템이 마련될 필요가 있다.

Part II

윈터스쿨

1. 해외 프로그램 사례

디지털사회과학이건 디지털인문학이건 결국 디지털 테크놀로지가 전제될 수밖에 없기 때문에 인문사회학자를 위한 디지털 테크놀로지 교육프로그램을 만드는 것이 매우 절실한 과제일 수밖에 없다. 결국 디지털 인문학은 인문학자가 디지털 도구 사용에 능해졌을 때 가능한 분야일 수밖에 없는 것이다.

그리하여 디지털 인문학 관련 학회의 경우, 발표와 토론이라는 전통적인 방식 이외에 컨퍼런스와 별도로 워크숍을 개최하고 있어 눈길을 끈다. 독일의 역사학계를 예로 들어 그러한 문제적 상황에 대한 대응 방식을 가늠해 보도록 하자. 2년마다 열리는 독일 디지털역사학회의 2018년 워크숍의 구조는 다음과 같았다.⁹⁸⁾

	내용
1 워크숍	역사학에서 디지털적인 교육방법과 협업: 소셜 미디어와 더불어 Digital Public History, 교육상 위키피디어 활용, 협업을 위한 툴들 등이 중요하다.
2 워크숍	마녀사냥 연구에 대한 세 가지 접근법: 먼저 마녀사냥 관련 사료에서 포착된 지표와 동기들을 연대기 순으로 분석하고, 다음으로 마녀사냥의 중심적인 테제들을 살펴본 후, 이 주제들을 디지털적인 방식의 작업과 결부시켜본다. 핵심적인 것은 <u>데이터 모델링의 문제</u> , 즉 사료로부터 “표”로의 이전이다. 이를 위해 세 가지 전적으로 다른 방법들을 활용하여 사료를 디지털적으로 분석하고 그 결과들을 서로 비교하게 된다. 질적인 내용 분석을 위한 MAXQDA, 네트워크 분석을 위한 GEPHI, 그리고 양적 분석을 위한 SPSS등이 그것이다.
3 워크숍	디지털 편집: 역사 사료의 주석달기: Oxygen XML Editor, Atom, Git를 활용한다.

98) <https://digigw.hypotheses.org/1660>

	내용
4 워크숍	교육에서 텍스트 마이닝과 “멀리서 읽기”: 동독 언론 자료 포털 사이트에서 컴퓨터의 도움을 받는 텍스트 분석하기: DiaCollo라는 툴을 활용하여 텍스트마이닝을 위한 다양한 활용을 시도해보고, 이를 교육에 어떻게 활용할지에 대해 파악하기로 한다.

즉, 디지털 인문학의 연구 성과를 발표하고 있을 뿐만 아니라, 그러한 연구에서 활용되는 디지털 툴에 대해서도 공유하고 있는 것이다.

2020년 3월에 열릴 “디지털 역사학 대회”의 구조 역시도 마찬가지로 패턴을 보여주고 있다. 이 학회는 학회 전에 개최되는 학회 전 워크숍, 바캠프(BarCamp), 그리고 본 학회 등 세 부분으로 나뉜다.

컨퍼런스 전에 개최되는 두 개의 워크숍을 보자.

	내용
1	웹 스크래핑과 자연언어 처리 파이선 Syntax와 Wortschaftz에 대한 기초개념을 전달하고, 중앙 모듈, 도입, 작업환경 등에 익숙해지도록 하는 것
2	de.hypotheses로 학문 블로그 만들기 블로깅의 개념적이고 학문적인 측면을 일반적으로 테마화하고 인문, 사회과학 분야에서 몇 가지 좋은 예들을 소개하면서, 참여자 전체가 de.hypotheses.org 라는 블로그 형태를 통해서 각자 블로그를 연습.

이처럼 Hands-on Workshop을 정식 컨퍼런스에 추가하여 인문학자들이 기술적 장벽을 넘어서도록 하는 것은 디지털 인문학 컨퍼런스의 한 전형으로 자리잡은 듯하다.

그런가하면 각국에서는 섬머스쿨의 형태로 디지털 인문학 활용 지식을 확산시키기 위해 노력하고 있다. 독일의 라이프치히 대학에서 2019년에 행한 썸머 프로그램은 다음과 같다. 디지털 인문학을 구성하는 다양한 방법들, 도구들에 대한 지식들을 제공하고 있고, 아울러서 디지털 인문학에서 제기할 수 있는 문제제기의 방식에 대해서도 설명하고 있다.

1일차

XML-TEI document encoding, structuring, rendering and transformation	Manuscripts in the Digital Age: XML-Based Catalogues and Editions	Digital Annotation and Analysis of Literary Texts with CATMA 6.0	All About Data - Exploratory Data Modelling and Practical Database Access
Stylometry	Humanities Data and Mapping Environments	Images of Image Machines. Theory and Practice of Interpretable Machine Learning for the Digital Humanities	Compilation, Annotation and Analysis of Written Text Corpora. Introduction to Methods and Tools
An Introduction to Neural Networks for Natural Language Processing - Applications and Implementation			

2일차

XML-TEI document encoding, structuring, rendering and transformation	Manuscripts in the Digital Age: XML-Based Catalogues and Editions	Digital Annotation and Analysis of Literary Texts with CATMA 6.0	All About Data - Exploratory Data Modelling and Practical Database Access
Stylometry	Humanities Data and Mapping Environments	Images of Image Machines. Theory and Practice of Interpretable Machine Learning for the Digital Humanities	Compilation, Annotation and Analysis of Written Text Corpora. Introduction to Methods and Tools
Searching Linguistic Patterns in Text Corpora for Digital Humanities Research			

이처럼 전체적인 내용에 대한 전반적인 조망을 시도하는 교육 프로그램이 제공되는 한편에, 특정한 톨에 대해 집중적으로 교육하는 방식으로 프로그램이 진행되기도 한다. 뮌헨 대학에서 2017년에 실시한 썸머 프로그램의 경우 “디지털 편집”을 중심으로 이루어졌다.

	월	화	수	목	금
I	개막	인문학 실제에서 데이터베이스	인문학 실제에서 데이터베이스	인문학 실제에서 데이터베이스	인문학에 실제에서 데이터베이스
II	학문적 팀워크로서 디지털 편집	XML과 함께 하는 디지털 편집 테크닉	XML과 함께 하는 디지털 편집 실제	XML/XSLT와 함께 하는 디지털 편집의 테크닉	XML/XSLT와 함께 하는 디지털 편집의 테크닉
III	학문적 팀워크로서 디지털 편집	학문적 팀워크로서 디지털 편집 (1. 컴퓨터 링퀴스틱스에서 디지털 편집의 활용 2. 편집을 위한 그래픽 데이터)	학문적 팀워크의 실제로서 디지털 편집	학문적 팀워크의 실제로서 디지털 편집	디지털 편집 프로젝트 공유
IV	HTML과 CSS				

2019년 고려대에서 이루어진 썸머 스쿨의 경우 라이프치히 대학과 마찬가지로 디지털 인문학 방법론에 대한 전체적인 개관을 시도하고 있음을 볼 수 있다.

	월	화	수	목	금
오전	디지털 인문학의 이해	인문 지식 콘텐츠 제작 로드맵	개방형 지식 데이터의 연결: 시맨틱 웹과 LOD	맞춤형 심화실습	디지털인문학 교육 프로그램 설계
오후	디지털 환경에서의 협업: 위키 소프트웨어의 이해+위키 문법의 사용법	디지털 세계에서 소통하는 지식의 설계: 온톨로지의 이해+온톨로지 설계 방법	인문 지식 데이터 편찬 실습: 인문 지식 데이터 탐색+네트워크 그래프 제작 방법	인문지식 콘텐츠 제작 실습	인문지식 콘텐츠 제작 결과 발표

연세대학교 디지털사회과학센터에서도 디지털 사회과학의 의의나 방법론, 교육 프로그램 등을 제공하고 있다. 2017년부터 데이터 분석에 관한 워크샵이 정기적으로 열리고 있으며, Python과 같은 데이터 분석을 위한 기초 언어 습득 및 활용을 배우는 방식으로 진행되고 있다. 예를 들어 Python으로 웹 데이터를 수집하는 방법을 기초적으로 배우고, 텍스트 전처리, 네트워크 분석, 감성분석, 토픽 모델링(LDA)등의 활용법을 익힌다. 다만, 텍스트 분석과정은 이미 Python에 대한 이해가 충분하고, 사전 지식 및 경험을 가진 경우에만 참여할 수 있다.

한국에서도 인문사회학자들을 대상으로 하여 디지털 테크놀로지를 교육하는 효율적인 방식이 무엇인지 모색해보고자 2월 6~8일 사흘간 서울대에서 윈터스쿨을 개최하였다. 고려대, 연세대, 라이프치히 대학, 뮌헨 대학 등의 사례를 바탕으로 하여 디지털 인문학 연구 방법에 대한 전반적인 개관을 시도할지, 혹은 특정한 틀에 대해 집중적으로 학습하여 활용가능한 수준으로 끌어올려야 할 지에 대해 지속적으로 논의를 하였다. 이 논의의 결과로 전반적인 디지털 인문학의 경향에 대해 조망을 하되, 디지털 분석(자연어 분석, 사회네트워크분석, 딥러닝 분석)에 대해 좀 더 집중적으로 학습하기로 결정하였다. 그 중에서도 인문학에서 사용 사례가 빈번하고 유용한 자연어 처리 분석 부분의 비중이 상대적으로 높았다.

이 프로그램에는 매일 10~11명이 참여하였다. 아래에서는 그 교육 내용이 무엇이었는지를 정리하고, 다음으로 프로그램에 대한 평가표를 통해서 향후 있을 디지털 테크놀로지 교육에서 참고할 모형을 만들어보고자 했다.

2. 윈터스쿨 개요 및 교육내용

- 일시: 2020. 02. 06 ~ 02. 08
- 장소: 서울대학교 14-423
- 강연자: 김바로 (중앙대학교 HK연구교수)
- 수강인원: 10~11명, 서양사 연구자 다수와 사회과학자 및 동아시아학자도 있었음.
- 교육내용: 사흘에 걸쳐 데이터 구축과 분석까지의 과정에 대하여 워크숍을 진행했고, 대략의 교육 내용은 다음 표와 같다.

날짜	교육 내용
1일차(2월 6일)	<p><데이터 구축> 알고리즘의 개념 데이터의 개념 - 인간가독형 데이터를 기계가독형 데이터로 전환 데이터의 종류 - 계층형 데이터, 관계형 데이터 데이터의 수집 - 한국의 공공 데이터 검토, 기타 데이터의 정제 - 정규표현식 사용법 데이터 구축 방식 - markup / relational / semantic model</p>
2일차(2월 7일)	<p><자연어 처리 I > 인문 데이터 자연어 분석 원시 말뭉치(코퍼스)를 데이터화 하여 분석하기 분석 방식들 소개 - 형태소 분석, word2vec 등 형태소 분석 실습</p> <ul style="list-style-type: none"> • 자연어 처리 방식 중 형태소 분석을 실습. • 구글 드라이브에서 구글 colab을 이용하여 한국어 자연어 분석 실습 • 역대 대통령 취임사를 데이터로 만들고 분석함. • 다양한 형태소 분석기와 각각의 특징 확인, 형태소 분석의 결과 해석 • 형태소 분석의 한계
3일차(2월 8일)	<p><자연어 처리 II, 사회네트워크 분석, 딥러닝> 인문 데이터 자연어 분석의 또 다른 방식인 word2vec 실습</p> <ul style="list-style-type: none"> • word2vec은 서로 연관을 가지는 단어들을 숫자로 표현 • 적절한 결과를 도출하기 위해 분석 조건값을 조정하고 해석을 하는 것의 중요성 <p>사회네트워크 분석의 유형과 표현 방법 딥러닝의 개념과 방법 시각화 맛보기</p>

3. 결론: 윈터스쿨에 대한 평가와 종합

윈터 스쿨에 대한 설문조사 중 객관식 문항을 표로 정리한 결과는 아래와 같다.
 (아래 표 참고)

- 디지털 인문학에서 활용되는 기법에 대한 사전 지식 여부에 대해서 알고 있다고 답한 응답자는 전체의 50%였다. 나머지 50%는 전혀 모르거나 들어본 적이 있다고 답하여 참여자의 절반은 디지털 인문학의 경험이 별로 없었음을 알 수 있다.
- 이처럼 디지털 인문학에 대하여 친숙하지 않은 참가자들도 많았지만 결과적으

로 모든 참가자들은 추후에 비슷한 프로그램이 열린다면 참여하겠다고 답하여, 디지털 인문학에 대한 높은 관심을 드러냈다. 참여자들은 디지털 인문학의 중요성에 대한 인식을 갖고 교육에 참여하였다고 볼 수 있을 것이다.

- 동시에 이번 프로그램이 향후 연구에 도움이 될 것이라고 생각하느냐 하는 질문에 대해서 60%가 긍정적으로 답했지만 절반에 육박하는 40%는 ‘보통’이라고 답하였다. 즉 위의 문항 결과와 대비해 보았을 때, 사흘간의 짧은 윈터스쿨을 통해서 본인의 연구에 디지털 방법론을 직접 적용할 수 있는 수준에까지 도달하지는 못했다는 정도로 이해하는 것이 타당할 듯하다.

	디지털 인문학에 대한 사전 지식	이번 프로그램이 향후 연구에 도움이 될 것이라고 보는지	향후 비슷한 프로그램에 참여할 의사가 있는지
참여자 1	3(알고 있다)	3(보통)	4(그렇다)
참여자 2	3	5(매우 그렇다)	5(매우 그렇다)
참여자 3	1(전혀 모른다)	5	4
참여자 4	2(들어본 적은 있다)	3	4
참여자 5	3	4(그렇다)	5
참여자 6	2	4	4
참여자 7	3	4	4
참여자 8	3	3	4
참여자 9	2	5	4
참여자 10	2	3	4
참여자 11	1	3	4
종합	2.27 들어본 적은 있다와 알고 있다 사이	3.81 보통이다와 그렇다 사이	4.18 그렇다와 매우 그렇다 사이

설문 결과에 대한 위의 추론은 아래 주관식 설문을 분석함으로써 다시 한 번 확인되고 있다. 특정한 디지털인문학 연구 방법에 집중하기보다 전체적으로 디지털인문학 전반에 대해 조망을 할 수 있었던 점에 대해 대부분의 참석자들이

만족을 표했다(아래 첨부자료 참고). 그러나 주로 서양언어로 연구하는 참석자들은 서양언어로 교육이 이루어졌기를 희망했으며, 본인들이 실제 연구하는 자료들로 실습하지 못한 부분에 대해 아쉬움을 피력했다. 즉, 사흘간의 집중 코스로는 연구에 직접 적용할 수 있는 지식수준에 도달할 수 없다는 아쉬움을 피력한 경우가 많았던 것이다.

특징	긍정적인 측면	개선될 점
디지털 인문학 분야 전반을 다양하게 배우도록 한 구성 관련	<ul style="list-style-type: none"> • 디지털 인문학 분야의 연구 전반의 틀을 파악할 수 있어서 좋았다. • 인문학 전공자가 빅데이터 시대에 어떤 기여를 할 수 있을지 판단할 수 있었음 	<ul style="list-style-type: none"> • 전반의 틀을 파악하는 과정이 힘겨웠었고, 각각의 주제에 대해서라면 시간이 짧아서 아쉬웠음. • 전체 개관을 짧게 한 후 특정 주제에 대해 집중적으로 배우는 편이 바람직할 수도. • 실제 연구와 결부되지 않는다면 연구의 틀을 파악한 결과가 오래 기억될 것 같지 않다.
실습이 병행된 점	<ul style="list-style-type: none"> • 강연자가 미리 만들어둔 틀을 통한 실습을 통해서 디지털 인문학이 실제로 어떻게 진행되는지 파악할 수 있었다 	<ul style="list-style-type: none"> • 강연자 이외에 실습 조교가 필요한 듯 • 수강생들의 실제 연구와 연관되는 주제를 미리 준비 한 이후 실습을 했다면 더 도움이 되었을 듯. • 실습 시간이 부족
분석언어가 한국어였던 점	-	<ul style="list-style-type: none"> • 서양사 전공자의 입장에서는 구체적인 작업을 위해 영어 자료로 작업하는 과정을 따로 배워야 할 것 같다.
수강생들의 사전 지식 부족	-	<ul style="list-style-type: none"> • 기초적인 개념들에 대한 정리가 필요 • 수강생들이 미리 관련 자료들을 읽어오도록 하는 것이 필요한 듯 • 수준에 맞는 분반이 필요→ 초보를 위한 특별반
강연자가 틀을 아는 선을 넘어 실제 그 분야 연구자였던 점	<ul style="list-style-type: none"> • 연구하는 과정에서 실제로 부딪히는 문제, 혹은 중요한 주제에 대한 연구자의 판단에 대해 함께 숙고해볼 기회가 있어서 좋았음 	-

특징	긍정적인 측면	개선될 점
소규모로 진행	<ul style="list-style-type: none"> • 중간 중간 자유롭게 질의응답이 이루어질 수 있는 구조 	-
단기간의 워크숍 형식	<ul style="list-style-type: none"> • 디지털 인문학과 관련되는 대강의 흐름을 파악할 수 있었음 	<ul style="list-style-type: none"> • 정기적인 프로그램의 형태로 교육될 수 있기를 희망함
한 명의 강연자로 된 구성	<ul style="list-style-type: none"> • 일관된 구성 	<ul style="list-style-type: none"> • 강연자의 전문분야가 아닐 경우 외부전문가를 활용하는 방법을 고려할 필요

이러한 경험을 토대로 향후 인문사회학자들을 위한 디지털 테크놀로지 교육 프로그램이 실시될 경우를 대비하여, 다음과 같은 제언을 할 수 있을 듯하다.

4. 디지털 인문학 교육의 정책적 제언

관련되는 정책상의 제언



1. 디지털 인문학 교육의 전체적인 방향성에 관하여

모든 디지털 프로젝트는 구조적으로 공통점이 있다. 기본적으로 모든 디지털 데이터는 파일이라는 형식 (files or assets) 으로 존재하고, 이 데이터는 데이터베이스(data base, structure)에 저장되어 있으며, 브라우저 (browser)를 통해 접근되고 시각화 되는 것이다. 따라서 디지털 인문사회 연구를 위해서는 이러한 디지털 자료(digital assets)에 구조 (structure)를 만드는 법, 만들어진 구조를 구성하는 법, 구성된 구조를 통해 데이터를 시각화(exhibit, visualization)하는 작업 등을 단계적으로 이해하고 학습하는 과정을 거치게 된다.

현재 한국에서 이루어지는 디지털 인문학관련 교육이 교육데이터 분석 도구의

학습과 응용에 중점을 두고 있으며, 데이터 자체에 대한 연구라던가, 데이터 선택 구축 등을 다루는 데이터 큐레이팅⁹⁹⁾에 관한 내용은 제외되어 있는 경우가 많다. 데이터 큐레이팅은 데이터 분석 뿐만 아니라, 데이터 가공 ‘프레임’과 데이터 공유 ‘시스템’도 포함한다. 인문학자들이 디지털 테크놀로지를 능숙하게 다루어야 하는 필요성은 결국 지금까지의 ‘연구자료’와는 질적, 양적으로 다른 새로운 자료가 등장했다고 여겨지기 때문에 대두되었다. 따라서 디지털 인문학을 위한 테크놀로지의 습득을 위해서는 ‘새로운 자료’의 성격은 무엇이며, 이 새로운 자료가 어떻게 새로운 지식의 생산에 기여할 것인지, 연구의 체계적 방법론은 어떻게 바뀌어야 하는지도 함께 연구되고 교육되어야 한다.

결국, 기술적인 내용 만이 아니라 관련되는 사회학적이고 철학적인 고찰이 동시에 병행됨으로써 근본적인 혁신을 이룰 수 있도록 해야 할 필요가 절실한 것이다.

II. 디지털 인문학 지식 확산을 위한 구체적인 방식

1. 디지털 인문학 관련 온라인 플랫폼 구축 필요

- 1) 해외에서는 유튜브나 전용플랫폼(數位人文創新課程典藏網, <http://tcdh.archive.nccu.edu.tw/>) 등을 통해서 디지털 인문학 관련 온라인 플랫폼 구축이 이루어진 경우도 있다.
- 2) 컴퓨터공학 중심의 알고리즘 교육이 아닌 인문학 중심의 실용적인 데이터 설계, 구축, 분석, 해석, 시각화 교육 플랫폼 마련이 필요해 보인다.

2. 도서관 중심의 디지털 인문학 센터 구축

- 1) 해외에서는 도서관이 문헌 중심에서 정보 중심으로 변화하면서, 국책연구기관의 지원 하에 디지털 인문학 센터를 설립하고, 디지털과 인문학의 융

99) 김연숙 (2018). 「데이터 큐레이션과 사회과학의 미래」, 『디지털 사회』, 8호 (온라인). 데이터를 특정한 주제와 목적에 따라 수집하고 구성하는 행위를 데이터 큐레이션이라고 한다.

합 연구 방법에 대한 다양한 교육 및 연구 지원이 실행 중이다.

2) 현재 한국에서도 관련 논의가 있으나, 가시적인 결과를 낳지는 못하고 있다. 도서관 관련 전공이 개설된 대학들에서 먼저 디지털 인문학을 커리큘럼의 일환으로 포함시키는 것이 필요해 보인다.

3. 학술대회에서의 디지털 인문학 톨 교육 필요

1) 해외 인문학 학술대회에서는 학술대회 사전프로그램으로 학술대회 전날 혹은 당일 오전에 디지털 인문학 톨에 대한 워크숍을 진행하고 있는 상황이다.

2) 한국에서도 학술대회 지원 조건에 디지털 인문학 관련 교육 항목을 추가하여, 인문학자가 새로운 디지털 기술에 접근할 수 있는 환경을 조성할 필요가 있다.

4. 디지털 인문학 교육을 위한 교수자 필요

1) 디지털 인문학 교육을 위해서는 인문학과 디지털 기술에 대해서 두루 이해를 갖춘 교수자가 필요하다.

2) 디지털 인문학 분야에 대해 평가할 기준도 마련되지 않았고, 그 전문성을 평가하기에 충분한 인력풀이 확보되어 있지 않은 것이 현실이다. 영국과 독일들의 경우를 보아도 다수의 교수채용이 이루어지고 있지만, 이 과정은 매우 신중하게 이루어질 필요가 있어 보인다. 디지털 인문학 분야 신진 연구자들에 대한 과대 평가도, 과소 평가도 부적절하며, 학계의 광범위한 논의에 근거한 신중한 접근이 절대적으로 필요하다. 디지털 인문학 향후 교육과 연구의 방향성을 결정하는 중요한 선택이 될 것이기 때문이다.

III. 디지털 인문학 교육 프로그램에 대한 구체적인 제안(예시)

1. 디지털 인문학 기본

디지털 인문학 기본은 데이터에 대한 필수 개념을 이해하고 실습을 하기 위해서 다음과 같은 예시를 제시할 수 있다. 핵심 교육 내용은 디지털 인문학에 대한 이해를 중심으로 하는 강의와 특정 주제에 대한 인문 데이터 구축을 중심으로 하는 실습이다.

주차	교육 내용
1	강의 소개 및 수업 내용과 과제 안내
2	인문데이터 개론
3	나의 디지털 연구실: 서버, 네트워크 이용 환경 조성 방법
4	XML(eXtensible Markup Language)의 이해
5	XML을 활용한 인문데이터의 제작 방법
6	위키 시스템의 이해와 실습
7	공간의 시각적 표현: Google Earth
8	관계의 시각적 표현: Network Visualization
9	인문학 대상에 대한 이해
10	인문지식 아카이브 제작 실습 (1) - 데이터 정의
11	인문지식 대상에 대한 현장 답사
12	인문지식 아카이브 제작 실습 (2) - 위키 시스템 활용
13	인문지식 아카이브 제작 실습 (3) - 지리정보시스템 활용
14	인문지식 아카이브 제작 실습 (4) - 네트워크 활용
15	기말 발표

2. 디지털 인문학 심화

디지털 인문학 심화는 현장에서 사용하는 데이터 관련 지식을 이해하고 실습을 하기 위해서 만들어졌다. 핵심 교육 내용은 XML과 RDB를 기반으로 하는 인문 데이터 구축과 운용 실습이다.

주차	교육 내용
1	강의 소개 및 수업 내용과 과제 안내
2	인문데이터 심화 개론
3	XML 이론 및 실습 (1)
4	XML 이론 및 실습 (2)
5	DTD 이론 및 실습 (1)
6	DTD 이론 및 실습 (2)

주차	교육 내용
7	(팀프로젝트)인문데이터 주제 탐색
8	중간고사
9	관계형 데이터베이스 개론
10	MSSQL 이론 및 실습 (1)
11	MSSQL 이론 및 실습 (2)
12	XML 데이터베이스 이론 및 실습 (1)
13	XML 데이터베이스 이론 및 실습 (2)
14	(팀프로젝트)XML 및 관계형데이터베이스 설계 (1)
15	(팀프로젝트)XML 및 관계형 데이터베이스 설계 (2)
16	기말고사

3. 디지털 인문학 분석

디지털 인문학 분석 분석은 다양한 디지털 분석 방법을 이해하고 특정 주제로 실습을 하기 위해서 만들어졌다. 핵심 교육 내용은 언어분석, 사회네트워크 분석, 공간분석, 딥러닝 분석을 기반으로 하는 인문 데이터 분석 실습이다.

주차	교육 내용
1	강의 소개 및 수업 내용과 과제 안내
2	인문데이터 분석 개론
3	인문데이터 분석과 활용 사례
4	인문데이터 분석 방법론 및 사례 분석 - 인문데이터 주제발굴
5	인문데이터 분석 방법론 및 사례 분석 - 인문지식 데이터 분석
6	인문데이터 분석 방법론 및 사례 분석 - 인문데이터 사회네트워크 분석
7	인문데이터 분석 방법론 및 사례 분석 - 인문데이터 공간 분석
8	중간고사
9	팀프로젝트)인문데이터 분석 주제발굴과 발표 1
10	팀프로젝트)인문데이터 분석 주제발굴과 발표 2
11	팀프로젝트)언어분석발표 1
12	팀프로젝트)언어분석발표 2
13	팀프로젝트)사회네트워크분석 발표
14	팀프로젝트)공간분석 발표
15	팀프로젝트)최종발표
16	기말고사

4. 디지털 인문학 시각화

디지털 인문학 시각화는 인문데이터, 인문데이터 분석 결과, 인문데이터 해석 결과에 대한 디지털 시각화 방법을 이해하고 실습하기 위해서 만들어졌다. 핵심 교육 내용은 디지털 시각화의 핵심 원리를 이해하고, 자신의 데이터, 인문 데이터 분석 결과, 인문 데이터 해석 결과를 토대로 디지털 시각화를 수행하는 실습이다.

주차	교육 내용
1	강의 소개 및 수업 내용과 과제 안내
2	인문데이터 시각화 개론
3	인문데이터 시각화 방법론 및 사례 분석 - 차트
4	인문데이터 시각화 방법론 및 사례 분석 - 테이블
5	인문데이터 시각화 방법론 및 사례 분석 - 그래프
6	인문데이터 시각화 방법론 및 사례 분석 - 맵
7	인문데이터 시각화 방법론 및 사례 분석 - 대시보드
8	중간고사
9	팀프로젝트)인문데이터 시각화 주제 발굴과 발표 1
10	팀프로젝트)인문데이터 시각화 주제 발굴과 발표 2
11	팀프로젝트)인문데이터 시각화를 위한 데이터 편찬 1
12	팀프로젝트) 인문데이터 시각화를 위한 데이터 편찬 2
13	팀프로젝트)인문데이터 시각화를 위한 데이터 분석
14	팀프로젝트)인문데이터 시각화 연습 발표
15	팀프로젝트)인문데이터 시각화 최종 발표
16	기말고사



디지털 인문학 윈터 스쿨 1일차

- 1일차: 데이터 구축
- 2일차: 자연어 처리
- 3일차: 네트워크 분석(자연어와 연동한 공기어 네트워크 분석) + 딥러닝

I. 알고리즘

알고리즘의 핵심: 변수/조건(if-else)/반복

알고리즘 교육 프로그램: code.org-> 초등학생용 교재

II. 데이터

1. 문자 데이터를 숫자로 변환 가능하도록 배정됨.
2. 한글의 경우: 과거와 달리 유니코드를 활용. 유니코드 내 CJK (China-Japan-Korea) 문자 Set을 차지. 그 외 동아시아 문자 Set도.
3. 인간 가독형 데이터 vs. 기계가독형 데이터
 - 기계가독형 데이터 만드는 방식: Semantic model/relational model/markup model
4. 기계가독형 데이터 모델의 변천
 - 1) 계층형 데이터 베이스 모델: 폴더들 간의 관계, 도서관 장서 정리 방식
어느 하나의 계층에 포함되지 못하는, 즉 중복 데이터 처리의 경우 문제가 발생
 - 2) 관계형 데이터 베이스 모델: 데이터 무결성 유지를 위해 분리가능한 모든 정보를 최대한 분리, 보존 => 상업적인 경우에 활용
 - 엑셀에서 활용하는 방식

- 이처럼 분리한 데이터를 연결해주는 것=SQL(Structured Query Language).
Select/from/where/order by/group by 등 간단한 용어들로 구성됨.

3) 비정형 텍스트 데이터 처리 방식=markup model ⇒ 학문의 경우에 많이 쓰임

- 컴퓨터가 알 수 있는 방식의 처리가 필요
- 인물의 경우: 시작점과 끝점에 대해 통일적인 표기를 하는 방식으로 인물임을 보여줌
- 동명이인을 구분하기 위해 개인 id를 부여. 즉 인물에 속성 정보를 부여

<인물 id=001> 총무공</>

- XML의 구성요소: Element, 시작 태그, 속성, 종료 태그 등
- XML 문서 구조: 항목, 항목명, 메타데이터, 본문 등

5. 코퍼스 분석

- 1) text/body/head/paragraph 등으로 구분지어줌. HTML이 표준 규정
- 2) tei(말뭉치 언어학에서 쓰이는 표준 규정) 활용. 그보다 하위 단위의 경우 언어에 따라 다름.
- 3) 말뭉치의 대표적인 활용: 사전 만들기. 형태소 분석
- 4) 말뭉치를 자동적으로 만들어내기: N-gram 방식. N-gram을 통한 유사 문서 탐색이 가능(야후와 달리 구글, 네이버 등은 이를 활용.)
 - 1-gram(행복은, 너의, 마음에, 있다), 2-gram(행복은 너의, 너의 마음에, 마음에 있다), 3-gram(행복은 너의 마음에, 너의 마음에 있다.)
 - N-gram 방식의 문제: 관련성을 밝힐 뿐, 관계 자체의 의미는 모름. 한국어처럼 조사가 많고 띄어쓰기가 통일적이지 않은 경우의 문제.
 - 토큰화(tokenization): 행복/은/너/의/마음/에/있다/.

6. 딥러닝 방식: Labeled Data. 이미지는 오히려 단순한 데이터(1~206까지의 숫자로 표기가 가능)

- 1) Data-Label 구조로 만들기.
- 2) 워드 임베딩(숫자로 바꾸어주는 방식): 가장 유명한 것이 워드투벡(word2vec: 워드를 벡터로) 방식(문맥을 고려해서, 즉 공기어를 고려해

- 서 숫자로 치환해주는 것)
- 단어 간의 의미를 파악. 한국과 우리나라의 경우. 코사인 유사도를 통해서 파악(공식이 존재). 이순신과 충무공의 거리가 가까움을 파악하게 됨. 의미의 연산까지도 가능해짐. 단어 단위에서 유사도를 알 수 있고, 문서 단위에서도 파악해 줌.
 - 단어 간의 의미 연산이 가능: 한국-서울+도쿄=일본
- 3) 전체적인 경향: 알고리즘/빅데이터/하드웨어가 증대 및 발전. 데이터 셋의 공유가 증가. 알고리즘 공유 사이트 증대(GitHub, TensorFlow, SQuAD 등이 예)
- 알고리즘 공유: 알고리즘을 논문으로 출판되는 것보다 공유 사이트를 통해서 입증하고 활용도를 높이도록 함.
- 4) 형태소 분석기가 새롭게 개발됨.
- 5) 딥 러닝의 경우 그 “과정”이 블랙박스. “해석 가능한 인공지능” 등에 대한 연구가 진행되고 있기도 하나, 현재까지는 블랙박스 상태.
- 6) Artificial Intelligence>Machine learning(programming)>deep learning
7. 딥러닝(=subsymbolic) 이외에 인공지능의 다른 방식: Symbolic (ontology=데이터 설계도)
- 1) expert system: if-else 방식
 - 2) 이를 구현하는 것이 시맨틱 웹
 - XML(비정형 텍스트에서 중요한 요소들을 마크 업. eXtensible Markup Language)의 한계: relation이 빠진 상태
 - 시맨틱 웹(ontology): 관계를 집어넣어주는 것. 레미제라블-저자-위고. 인간 인식과 최대한 비슷한 정보를 넣어주는 것
 - LOD(Linked open data): 서울시, 국립중앙도서관, 시설안전공단 등으로 분리되어 있는 정보들이 서로 연결될 수 있도록 함. 유로피어나(Europeana)가 대표적인 사례들

III. 데이터 수집

1. 공공데이터

1) 정부 3.0: 능동적 정보 공개

- 사례: 버스 노후화를 측정하기 위해 버스에 GPS를 붙인 끝에 버스 운행 정보로 변화하게 됨.

2) 데이터 3법: 개인 정보를 쓰면 안 되지만(상업적으로 의미있는 것이 개인 정보이기 때문에), 가명처리한 개인 신용 정보를 동의 없이 이용하거나 제공 ⇒ 정보를 최대한으로 공개하는 것이 최근의 추세.

3) 공공데이터포털: 한국(www.data.go.kr) XML, CSV, RDB), AI Hub

2. 수집 방식: 스크래핑(목적지가 있는 상태) & 크롤링(목적지 없이 떠도는 상태)

1) 웹스크래핑: 웹 사이트마다 방식이 다름

IV. 데이터 정제

1. 정제를 위한 툴: “정규 표현식”(공백 없애기/문장 단위).

1) 정규표현식의 기호들: 언어가 섞여있을 때 분리해내는데 유용

2) 컴퓨터 언어 처리는 기본적으로 “찾기 & 바꾸기”이기 때문에 정규표현식을 정확히 이해할 경우 형태소분석, 딥러닝 등을 쉽게 배울 수 있음.

3) Notepad ++로 실시

V. 데이터 구축방식

1. Markup 모델: XML 실습

1) HTML: 문서의 내용과 모양을 한꺼번에 취급

2) HTML과 달리 XML의 경우 DTD(문서의 구조), XSL(문서의 모양), XML(문서의 내용)이 구분됨

3) XML로는 연결고리가 파악이 안됨. 유의미한 것을 식별해 낼 뿐.

4) “,”, “|”의 차이. “.”의 경우 순서가 바뀌어서는 안 된다는 것.

5) 속성: 아이디값(필수값), 혹은 항목값

6) Markup 제작 프로그램: Oxygen XML editor

2. Relation Model: RDB(사전 등 정보의 집적을 위해)

- 1) 데이터베이스 설계의 3원칙: 제1정규화, 제2정규화, 제3정규화를 통해 데이터의 무결성에 이르게 함.
- 2) 1차 정규화: 테이블의 각 셀이 하나의 값을 갖도록 함.
- 3) 2차 정규화: 주 키(key)가 합성키인 경우
- 4) 3차 정규화: 주 키가 아닌 나머지 속성 사이의 관계의 문제. 2차와 달리 3차의 경우 선택적으로 진행할 수 있음.
- 5) 이를 위한 language: SQL
- 6) 추론이 불가능한 단순 데이터 관련성. 관련이 있다는 것을 보여줄 뿐, 어떤 관계인지 설명하지는 못함.

3. Ontology(지식 체계, 설계도)+시맨틱 웹 ⇒ 추론이 가능하다는 점에서 RDB보다 발전된 형태

- 1) 온톨로지 목록: BIBO, SKOS, Dublin-Core 등
- 2) 온톨로지의 기본 개념: 사건, 장소, 속성
- 3) Entity/individual/object(개체, 객체) ⇒ 속성(attribute/property) ⇒ 클래스(class) ⇒ 관계(클래스와 클래스 사이에 존재)
- 4) 기계가 읽을 수 있는 것(machine-readable, machine-understandable)
+ shared = 온톨로지의 가장 중요한 속성
- 5) Object properties의 사례들
 - inverse/functional(동일 판단)/inverse functional/transitive(파생관계)/symmetric(대칭)/asymmetric(비대칭 관계)/reflective(재귀적)/irreflexive(비재귀)
- 6) Protege 프로그램으로 실습



2일차 교육: 자연어처리

I. 자연어 처리 방식

1. 형태소분석(전통적인 방식) + 워드임베딩(딥러닝 기술이 포함된 내용. 업계에서는 딥러닝으로 간주하지 않음. 문자로 된 워드를 숫자로 변환한다는 의미)

II. 자연어 분석 이론(I): 형태소 분석

1. 형태소분석의 원리

- 1) 정규표현식에서 “찾아서 바꾸기” 원리와 같음.
- 2) 조사의 “은”과 명사 금“은”이 같지 않다는 것. 이를 구분해내는 것이 형태소 분석의 핵심
- 3) 말뭉치(코퍼스): 구문 분석 말뭉치
- 4) 형태소 분석의 원리: 사전처럼 단어 구분 + 문법 규칙 찾아냄.
- 5) 형태소 분석의 다양성
 - 사전의 다양성과 유사: 예컨대 사전에 따라 보통명사, 혹은 명사, 조사 혹은 격조사 구분을 하는 등 다양.
 - 문법 규칙도 다양하게 인지됨: 모든 예외사항을 다 문법규칙에 포함시킬지 혹은 예외로 간주할지에 따라 다름
 - 형태소 분석기가 현대 언어에 특화되는 방식으로 이루어짐.
- 6) 형태소 분석 자체는 파이선이나 r을 통해서 하고, 결과치 분석은 엑셀(대부분의 통계 분석이 가능)을 통해서
- 7) 결과 해석은 인문학의 영역 ⇒ 예시: 겹(의도), 마침표 빈도에 대한 해석
- 8) 언어분석의 방식: 토픽모델링(γ) + 감정분석(Δ)
- 9) 토픽모델링(γ): 텍스트 기준
 - TF-IDF-문서에서 키워드가 무엇인지 찾아주는 방식. 조사 등 무의미하게 많이 나오는 표현들을 빼주는 것.

- 명사 중심
- 전체 문장에서 다른 조합과 자주 등장할 때 보다 중요. 실습.
- 10) 감정분석(L): 모든 단어에 긍정과 부정 판별을 해준 후 식별
- 11) 워드 임베딩: 텍스트 전체를 숫자로 바꾸어서 활용
 - word2vec
 - word2tensor
- 2. 형태소 분석을 위한 파이선 설치
 - 1) 파이선을 컴퓨터에 설치하기의 어려움 ⇒ colab(구글이 사용을 간단하게 하기 위해 만들)
 - 파이선을 편리하게 사용하기 위한 보조툴: 주피터 노트북(파이선의 보조툴)
 - 2) colab의 경우 30분마다 끊기는 상황(클라우드에서 활용), 1시간 사용할 경우에도 끊김.
- 3. 형태소 분석
 - 1) 데이터 세팅 작업에 인문학자의 식견이 반영됨
 - id/타입(분류)/text
 - 2) pip install(설치하라는 명령어) -v python-mecab-ko
 - import(실행하라는 명령어) mecab
 - mecab = mecab.MeCab()
 - mecab.morphs("영등포구청역에 있는 맛집 좀 알려주세요.")--> 문장을 넣을 경우 형태소로 자동 분리
 - mecab.nouns(우리나라에는 무릎 치료를 잘하는 정형외과가 있는가): 명사만 분리해냄
 - mecab.pos(자연주의 쇼핑몰은 어떤 곳인가?): 품사 태깅 방식
 - 3) 형태소분석기의 사례: Twitter Korean text(띄어쓰기가 잘 맞지 않는 텍스트들의 경우)가 대표적인 사례
 - 4) 형태소 분석을 하지 않으면 변이가 너무 많아서 처리할 수가 없음.
- 4. 실행방식

- 1) 프로그램 설치와 실행
- 2) 데이터 불러오기
 - rawdata(저장한다는 파이선 용어) = pd(데이터표).read_csv(쉼표로 분리된 상태)(‘testko.csv’, header=0, encoding=utf-8(유니코드가 아닌 utf-8. 언어들에 섞여있을 경우), dtype=str, engine=python))
 - vs. tsv(탭으로 분리) ⇒ 자연어 텍스트에 나오는 “,”와 구분하기 위해 tsv를 더 자주쓰게 됨.
- 3) 데이터 형태소 분석하기

III. 자연언어 처리 방식(II): 워드 임베딩의 원리

1. 워드투백의 핵심: 분산 의미(Distributional Semantics)
 - 1) 의미연관을 생각
 - 2) 문자에 숫자(좌표)를 입력.
2. 워드투백의 개념(ㄱ): 차원을 표현
 - 1) 차원이 높아질수록 더 세분해서 볼 수 있음
 - 2) 그러나 차원이 높아지게 되면, 즉 숫자가 많아지게 되면 숫자연산이 많아져서 컴퓨터 성능에 제약을 받게 됨: 따라서 일반적으로는 100차원이 최적화된 차원. 인문학에서는 더 높아도 가능
3. 워드투백의 개념(ㄴ): window의 기본. 앞뒤로 연관되는 단어 찾아 넣기
 - 1) 한국어는 10을 기본값으로 하는 경우가 많으나 문헌마다 다름.
4. 워드투백의 개념(ㄷ): CBOW vs. Skip-Gram
 - 1) CBOW; 최상의 ~을 얻으리라. 추론하도록 하는 것.
 - 2) Skip-Gram: _ 깨달음_. 답을 알려주고 학습을 시키는 것.
 - 3) 역사학의 경우 단어가 충분히 많기 때문에 의미관계망을 파악함. 문헌에서의 의미를 파악하는 것으로 충분.
- 4) 상업적인 경우: Bert
 - 형태소 단위가 아니라 캐릭터 단위, 음소 단위까지 나아감

IV. 실습

1. Ubuntu: OS 프로그램
2. Mode name: 사양
3. 런타임 유형변경: GPU(딥러닝할 때). TPU의 경우 접속이 끊김.
4. KoNLP(Korean natural language process) : 파이선 한국어 NLP
5. mecab: 형태소 분석기
6. pandas=pd, numpy=np
7. # 녹색. 녹색 부분은 실행이 되지 않고, 사람이 읽는 부분. 녹색 부분은 컴퓨터가 읽지 못함.
8. 워드 클라우드: jasondavies.com/wordcloud



3일차 교육

I. 워드투백

1. 워드투백의 원리
 - 1) 그 원리를 알 수 있게 시각화해주는 프로그램
 - wevi: word embedding visual inspector (ronxin.github.io/wevi/): 2차원의 경우에서 어떻게 이루어지는지를 잘 보여줌. 차원이 높을수록 연관성을 가지는 다양한 요소들이 고려되는 것.
 - 좌표값을 어떻게 배정하는지를 시각화해서 보여줌. 반복할수록 가까우면 점점 더 가깝게, 관련성이 멀 경우 점점 더 멀게 나타나게 됨. 그에 근거하여 좌표값이 배정됨.
 - 2) 인문학자들이 도메인 전문가로서 세팅값을 바꿔가면서 그 결과에 대해 검증할 수 있음.
 - 3) 용어: size= 차원, window(=보는 범위), min_count(=최소출현. 공학에서는 100을 기준으로. 100번 이상 출현하지 않을 경우 제외한다는 의미. min_count=100. min_count=0(모든 단어를 다 넣어야 한다는 의미),

iter(=반복을 의미. iter=10인 경우가 많음). 세팅 값을 봐 가면서 조정해야 하는, 즉 매우 경험적인 방식이라고 볼 수 있음. sg=1(Sg.는 CBOW 일지, Skip-Gram 일지의 문제)

- 4) 워드투백: 의미연산을 가능하게 하다는 것이 더 큰 장점.
 - 가령 “한국-서울+동경=일본”이 의미연산의 예.
 - 워드투백의 핵심: 1. 자료에 대한 판단을 위해 인문학자의 개입이 필요, 2. 의미연산을 하게 됨.
 - * 토큰화: 띄어쓰기 단위로 분리. Erlebnis, Erlebnis“ 등을 같게 인식. 기본형으로 바꾸어주고, 문장부호 등을 제거해주는 등의 노이즈 제거 작업이 필요
 - * 메깅(형태소 분석기)을 까는 방법: pip install -v python-mecab-ko

II. 시각화의 이론

1. 예시

- 1) 형태소 중 상위 100개의 좌표를 보는 방식
 - 차원 축소(예컨대 TSNE, PCA(주성분 분석, 위키에 존재))의 방식을 통해 시각화를 할 수 있음.
2. 파이션 영어 기반. 한국어 폰트가 없어서 한국어는 깨짐. 폰트를 한국어로 세팅해 주어야.

III. 네트워크 분석의 이론

1. 분석의 예

- 1) Republic of Letter의 경우
- 2) 족보
- 3) 문학
 - 문학 작품에서 인물들 간의 관계
- 4) 단어들끼리의 관계
- 5) Emotion of London
 - 지리정보 분석. 개별적인 장소에서의 감정

2. 분석의 핵심

- 1) 노드(점/대상/사람) + 링크(선/연결/관계)
- 2) A와 B의 관계. 인물일수도, 물건일수도, 단어일수도.
 - 사람과 사람의 관계(중국 데이터. CBDB); 혈연, 서신 교환, 사제 관계 등

3. 핵심 개념

- 1) 마당발/허브, 연결자(Linker), 중심자(Center)
 - 마당발: 1차적인 관계가 많은 경우
 - 연결자: 그 존재가 없어졌을 때 연결이 끊기는 경우
 - 중심자: 영향력, 혹은 전파력. 보균자로 생각할 수 있음.

4. 네트워크 분석의 원리

- 1) 오일러의 과니히스베르크의 다리 건너기 문제. 점과 선으로 단순화시킴
- 2) 사회과학에서의 분석
 - 밀그램의 편지 전달 실험
 - UCINET 4(범용 네트워크 분석 프로그램)
 - 프리만의 중심성 측정 척도 개발
- 3) 네트워크 과학의 역사(지식 네트워크)
 - 가필드의 인용색인
 - 케슬러의 서지결합법
 - 스몰의 동시인용 분석
 - 켈론 등의 동시 단어 선택
- 4) 복잡계 네트워크
 - 에르되시와 레니의 무작위 네트워크(ER 알고리즘)
 - 와츠와 스트로가츠의 좁은 세상 네트워크

5. 네트워크 표현 방법

- 1) 에지 리스트
 - (Source(김바로), target(변윤경), weight(기본값이 1))
 - 방향성이 있을 경우(direct)/방향성이 없을 경우(undirect),

- 2) 인접 행렬
- 3) 그래프
6. 분석이론 저서
 - 1) “소셜 네트워크 분석”
 - 2) “네트워크 분석 방법론”
7. 저장하는 방식
 - 1) Rawdata를 CSV 파일 형태로 저장. 엑셀에서 저장
 - 2) 스스로 정제 해줌: 예컨대 쉼표 처리.
 - 3) 동명이인의 경우
 - 김바로 1, 2도 가능하나 많은 경우 ID 값 주기.

IV. 딥러닝

1. 인공지능: symbolic(인간이 지식을 전수해주어야. Ontology) + subsymbolic(지식 입력 없이 기계가 스스로 파악. 최소한의 지식을 입력)
2. 딥러닝의 특징
 - 1) $A=1, B=2, AB=C, C=?$ (전통적인 방식)
 - 2) 연산방식을 묻는 것=딥러닝: $A=1, B=2, A?B=C, C=3$
 - 딥러닝: 연산방식 자체가 인문학의 핵심일 수 있기 때문에, 딥러닝은 인문학적일 수 없음.
3. 딥러닝의 실제
 - 1) teachable Machine: <https://teachablemachine.withgoogle.com/>
 - 2) 답이 있는 문제에 대해 분류하는 데 있어 강점을 가짐.
4. 딥러닝의 기본 개념
 - 1) Batch Size: 최초의 크기를 어떻게 잡아줄 것인가.
 - 2) Learning rate(기울기): 기울기는 학습 결과로 할 수 있음. 기울기를 바꾸어 가는 각도
 - 3) Overfitting \Rightarrow 대부분의 경우 정확도가 0.8 정도에 그침
5. 딥러닝 방식

- 1) 단어 단위가 아니라 문장 단위로 파악.
 - = 예: 긍정 감정, 부정 감정으로 텍스트를 나누어 판단하는 경우
- 2) 작업 절차: 임베딩 ⇒ 딥러닝
 - 1층: 임베딩 레이어. 숫자를 받아들임. 100일 경우. 차원이 100차원인 것으로 생각할 수 있음.
 - 2층:
 - 결국 긍정인지, 부정인지 판별
 - 이 층들이 계속 늘어나는 것이 딥러닝
 - “Layers.dense”가 층
- 3) 용어
 - 모델을 만들어 구축하고(Compile)한 다음 돌림(epoch를 돌리는 것)
 - loss(오류 값이 없는지 검증. 숫자가 낮을수록 좋음)/accuracy(정확도. 1이 최고): 정확도가 너무 올라갈 경우(overfitting) 협소한 영역에 특화된 것.
 - 학습(Epoch를 한번 돌 때 마다 학습) 셋/검증 셋. val(validation)_locc/val_accuracy 0.3인 경우 학습 셋이 7, 검증셋이 3일 경우.
 - drop: 망각시키는 것.
 - 기울기를 다양하게 테스트해보면서 accuracy는 높고, loss는 낮은 조합을 찾아내는 것. 대체로 세팅 값을 건드리지 않아도 무방함.
- 4) 모델을 만드는 작업이 딥러닝에서 하는 일
- 5) fast.ai: 일반 사용자들이 간편하게 쓸 수 있는 AI 툴
- 6) NAS, AUTOML
 - 전이(transfer): 이미 학습된 것을 바탕으로 그 위에 추가학습 시키기.
 - * word2vec: <https://nlp.stanford.edu/projects/histwords/>
 - * bigkinds.or.kr
 - * 맵핑용 툴: Gephi, arcGIS, QGist(지리 정보 분석 무료 소프트웨어)

V. 종합 정리

1. 데이터 형식

- 1) 비정형데이터(XML), 정형데이터(RDB), RDF
- 2) 설계도를 전체로 보여주는 것: XML에서는 DTD, RDB에서는 스키마, RDF에서는 온톨로지

2-1. 데이터 분석: symbolic(방법론을 사람이 지정하도록 되어 있음)

- 1) 언어분석(자연어 분석, 형태소분석, 감정 분석), 사회 네트워크 분석, 지리 정보분석
- 2) 언어분석: 형태소 분석(사전+문법=해당 텍스트를 자동으로 분리), word2vec(딥러닝을 하기 위해 텍스트를 숫자로 바꾸는 방식. 반복학습. 원천적으로 답을 주지 않음)
- 3) 형태소 분석: 감정분석, 공기어 분석(함께 등장)
 - 형태소 분석(빈도, 상대빈도를 알 수 있음. 동음이의어 문맥 관계를 전혀 파악을 못함). 기본적으로 count에 의존
 - 문맥적인 상황을 파악하기 위해서는 word2vec이 유리
- 4) word2vec: 문자를 숫자로 전환. 차원 개념+윈도우(포커싱의 문제)+CBOW(빈칸 채우기)/Skip-gram(분산 의미. 단어와의 관계.)
 - 의미 연산을 가능하게 함.
 - 상대 관계성으로 뽑아내는 것: 클러스터링이 가능. 대통령-국민.
- 5) 토픽 모델링은 형태소 분석에 근거한 테크닉

2-2. 데이터 분석: subsymbolic(딥러닝. 필요한 것은 라벨 데이터. 즉 정답이 필요)

- 1) 연도측정, 장르 구분 등 정답이 있을 수 있는 내용에 대해서 가능
- 2) 분류가 가능한 데이터의 경우에 적합

참고문헌

■ 국내 문헌

김연숙 (2018). 「데이터 큐레이션과 사회과학의 미래」, 『디지털 사회』, 8호 (온라인)

■ 외국어 문헌

Arts & Humanities Research Council, Delivery Plan 2019, p. 2.

Arts & Humanities Research Council, Delivery Plan 2019, pp. 2-3.

Arts & Humanities Research Council, Delivery Plan 2019, pp. 3.

Arts & Humanities Research Council, Delivery Plan 2019, p. 12.

Arts & Humanities Research Council, “The Human World: The Arts and Humanities in Our Times,” AHRC Strategy 2013-2018, p. 15.

Chetty, Krish; Aneja, Urvashi; Mishra, Vidisha; Gcora, Nozibele; Josie, Jaya, “Bridging the digital divide in the G20: Skills for the new age,” Economics, Vol. 12, Iss. 2018-24, pp. 1-20.

Chris Alen Sula, S. E. Hackney, Phillip Cunningham, “A Survey of Digital Humanities Programs,” The Journal of Interactive Technology & Pedagogy 11 2017.

Gender-related Development Index (GDI) and Gender Empowerment Measure (GEM).

Ian Milligan and Robert Warren, “Chapter 6. Big Data and the Coming Historical Revolution: From Black Boxed to Models,” in Big Data in the Arts and Humanities: Theory and Practice, p.

Institute for Women’s Policy Research 2010.

- Lara Putnam, “The transnational and the text-searchable: Digitized sources and the shadows they cast”, *American Historical Review*, Vol. 121-2, 397.
- Paul Moore, “Chapter 5. From Big Data to Thick Data: Theory and Practice,” in *Big Data in the Arts and Humanities: Theory and Practice*, p. 56.
- Richard A. Hawkins, “Chapter 7. Use of Big Data in Historical Research,” in *Big Data in the Arts and Humanities: Theory and Practice*, eds. Giovanni Schiuma and Daniela Carlucci (CRC Press, 2018), pp. 79-80.
- Richard A. Hawkins, “Chapter 7. Use of Big Data in Historical Research,” in *Big Data in the Arts and Humanities: Theory and Practice*, eds. Giovanni Schiuma and Daniela Carlucci (CRC Press, 2018), pp. 81-82.
- Ronald Inglehart and Pippa Norris, *Rising Tide: Gender Equality and Cultural Change Around the World*, Cambridge University Press, 2003, pp. 8-18.
- Sharon Block, #DigEarlyAm: Reflections on Digital Humanities and Early American Studies, *William and Mary Quarterly*, 3d ser., 76, no. 4, October 2019 DOI: <https://doi.org/10.5309/willmaryquar.76.4.0611>.
- U.S. Bureau of Labor Statistics 2009.
- Valuation Studies? Our collective two cents, *Valuation Studies*(1), 2013: 11-30
- World Economic Forum, “The Global Gender Gap Report, 2018”에서 Figure 2: Global performance, 2018.

■ 웹사이트

- <http://digitalitaet-geisteswissenschaften.de/veranstaltungen/symposienreihe>
<http://dohistory.org/>
<http://dx.doi.org/10.5018/economics-ejournal.ja.2018-24>; Federica Minichiello, “Disciplines scientifiques et in?galit?s de genre : une pr?occupation mondiale,”
http://femicide-watch.org/search?f%5B0%5D=field_freehandtags%3A3545
<http://femicide-watch.org/sites/default/files/FemicideVolXI.pdf>

<http://gendergapgrader.com/>
<http://gendergapgrader.com/studies/airline-pilots/>
<http://gendergapgrader.com/studies/gender-gap-in-science/>
<http://hdr.undp.org/en/content/table-4-gender-development-index>
<http://maps.vista-info.net/gis/htm/IWMIBasinMaps/index.html>
<https://mellon.org/grants/>
<http://ninastem.aprende.sep.gob.mx/?>
<http://republicofletters.stanford.edu/>
<http://revolt.axismaps.com/>
<http://revolt.axismaps.com/project.html>
<http://salem.lib.virginia.edu/home.html>
<http://usg.maps.arcgis.com/apps/webappviewer/index.html?id=2c94e981a6c6485ab0e26bb680e0b063>
<http://usg.maps.arcgis.com/apps/webappviewer/index.html?id=eb6ca76e008543a89349ff2517db47e6>
<http://www.crit.rai.it/CritPortal/?notizia=1553&lang=en>
http://www.dfg.de/en/research_funding/programmes/infrastructure/lis/international/dfg_neh_humanities
http://www.dfg.de/foerderung/info_wissenschaft/2016/info_wissenschaft_16_36/index.html
<http://www.gender4stem-project.eu/?>
http://www.indusbasin.org/applications/indusbasingp/gender_map.php
http://www.mogef.go.kr/eng/lw/eng_lw_f002.do
<http://www.oecd.org/internet/bridging-the-digital-gender-divide.pdf>
<http://www.scientix.eu/?news>
http://www.slate.com/articles/life/the_history_of_american_slavery/2015/06/animated_interactive_of_the_history_of_the_atlantic_slave_trade.html
<http://www.thegovlab.org/project-gender-and-urban-mobility-in-chile.html>

<http://www2.vcdh.virginia.edu/gos/>
http://www3.weforum.org/docs/WEF_GGGR_2018.pdf
<https://www.acls.org/programs/digitalextension/>
<https://ahrc.ukri.org/research/fundedthemesandprogrammes/themes/digitaltransformations/bigdata/>
<https://api.parliament.uk/historic-hansard/index.html>
<https://archive.ph/20131202232836/http://www.cio.de/dynamicit/bestpractice/2281581/index.html>
<https://cursus.edu/articles/37738/les-femmes-et-le-disciplines-scientifiques-rapport-unesco>
<https://data2x.org/>
<https://data2x.org/resource-center/big-data-and-the-wellbeing-of-women-and-girls/>
https://data2x.org/wp-content/uploads/2019/09/StreetHarassmentWomensEdChoices_WBBrown.pdf
<https://data2x.org/wp-content/uploads/2019/10/MiningWebViolenceAgainstWomen.pdf>
<https://data2x.org/wp-content/uploads/2020/02/Impact-Case-Studies-Vietnam-4P.pdf>
<https://dataconomy.com/2016/03/big-data-reveals-big-gender-inequality/>
<https://digigw.hypotheses.org/1660>
<https://ecda.northeastern.edu/>
<https://ehistory.org/about>
https://en.wikipedia.org/wiki/Big_data#/media/File:Hilbert_InfoGrowth.png
<https://gender.the100questions.org/>
<https://guides.clio-online.de/guides/arbeitsformen-und-techniken/digital-humanities/2018>
<https://information.tv5monde.com/video/femicides-combien-de-femmes-tu>

ees-dans-le-monde
<https://journals.openedition.org/ries/6787>
<https://mila.ss.ucla.edu/>
<https://omeka.org/>
<https://scripto.org/>
<https://tellmaps.com/uis/gender#!/tellmap/79054752>
<https://tellmaps.com/uis/gender#!/topic/GENDER>
<https://thinkbig.enm.bris.ac.uk/projects/content-analysis-of-british-periodicals/>
<https://tropy.org/>
<https://www.balcanicaucaso.org/eng/Areas/Europe/Femicide-the-numbers-in-Europe-184329>
<https://www.bl.uk/projects/british-library-labs>
<https://www.bmbf.de/foerderungen/bekanntmachung-643.html>
<https://www.britishnewspaperarchive.co.uk>;
<https://casebooks.lib.cam.ac.uk>
<https://www.clintonfoundation.org/press-releases/clinton-foundation-and-gates-foundation-partner-measure-global-progress-women-and>
https://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/2018/170116_rede_strohschneider_neujahrsempfang_en.pdf
https://www.dfg.de/en/research_funding/principles_dfg_funding/digital_turn/conference/index.html
<https://www.equalmeasures2030.org/how-we-work/data-and-analysis/>
<https://www.flickr.com/commons>
<https://www.girleffect.org/what-we-do/mobile-platforms/tega/>
<https://www.historians.org/teaching-and-learning/digital-history-resources/resources-for-getting-started-in-digital-history>
<https://www.iwmi.cgiar.org/2018/11/indus-basin-gender-profile-mapper/>
<https://www.komm-mach-mint.de/Komm-mach-MINT/English-Information>
<https://www.lejdd.fr>

[https://www.moe.gov.my/?](https://www.moe.gov.my/)

<https://www.neh.gov/divisions/odh>

<https://www.pnas.org/content/early/2017/01/03/1606380114.full>

<https://www.slavevoyages.org/>

<https://www.thetimes.co.uk/archive/>

<https://www.undp.org/content/undp/en/home/sustainable-development-goals/goal-5-gender-equality.html>

<https://www.unglobalpulse.org/2018/03/new-gender-equality-and-big-data-report-takes-stock-of-innovation-efforts-to-close-gender-gaps/>

<https://www.unglobalpulse.org/sites/default/files/Gender-equality-and-big-data-en-20>

https://www.unglobalpulse.org/wp-content/uploads/old_site/Gender-equality-and-big-data-en-2018.pdf

<https://www.weforum.org/reports/gender-gap-2020-report-100-years-pay-equality>

Abstract

The era of digitalization, which sweeps across all fields of human society like tsunami, changes the way how individuals live and interact, and more interestingly to this project, how society functions and how knowledge will be produced. This situation, in which the entire human life is conditioned by the digital media, is creating a number of unprecedented phenomena. This project which deals with these comprehensive changes both in our ordinary life and academia consists of two parts.

Chapter I in the Part I focuses on the use of intimate personal data and gender relation in order to see the influence of social changes in the digital-big data era on our ordinary lives. The fact that human life as a whole leaves traces in digital world could possibly go with the commercialization of this “trace data”. Facing this problem, sociologists have emphasized on the importance of “data community” or “collegial data storage,” so that we could protect these intimate personal data from market-driven commercialization.

As for Gender, the digital-big-data era has both positive and negative aspects. The male-dominated structure continues to be reproduced well into the digital world. However, big data also implies the possibility of getting a step closer to gender equality in various ways. For example, Data2X, which the U.N. has co-founded with the Clinton Foundation and other private foundations, aims to find out data about women which have been deficiently collected and distributed. By aggregating and disaggregating data about women, we can enhance gender equality in a much efficient way.

Chapter 2 and 3 examines how Humanities and social sciences are responding to the changes in the digital-big-data era, respectively. First, chapter 2 deals with the new issues in the digital humanities, such as breaking down the boundaries between the professional scholars and the public in producing knowledge etc. Then, we examine the latest academic achievement of digital humanities by focusing on the American history, British history and Classical studies. Finally, we discuss the measures which were taken by the governments and professional scholars of Britain and Germany to promote digital humanities research.

Chapter 3 analyzes discussions on the digital-big-data era in the field of social sciences. Nowadays, the digital social science, a new branch of social sciences, goes well beyond the traditional academic and national boundaries and begins to ask questions how “social” the digital “trace” data can be.

Part II analyzes the three-day workshop program and the enquete. Based on this survey, we make a few suggestions such as making it a regular program rather than an one-time event, offering differentiated levels of learning programs and others.

Keyword : Digital Humanities, Digital Social Science, trace data