

디지털 인문학 분야의 국내외 연구 동향 분석*

An Analysis on Research Trends of Digital Humanities

정유경 (Yoo Kyung Jeong)**

초 록

본 연구의 목적은 디지털 인문학 분야의 국내외 주요 연구의 동향을 살펴보고 비교 분석하는 데 있다. 기존의 디지털 인문학 동향에 관한 연구들은 해외의 주요 연구사례와 국가정책, 발전 방향 등을 검토하는 연구들로, 디지털 인문학 분야의 세부 연구주제 및 하위 분야를 파악하는 것과는 거리가 있었다. 본 연구에서는 국내외 학술 문헌의 계량서지적 분석을 통해 디지털 인문학 분야의 지적 구조를 살펴보고자 한다. 이를 위해 Web of Science와 RISS로부터 2020년 3월까지 출판된 디지털 인문학 관련 국내외 학술 문헌을 수집하여 분석에 사용하였다. 이 자료들을 대상으로 구조적 토픽모델을 적용하여 세부 주제들을 살펴보았으며, 이 주제들의 시계열적 추이를 파악하였다. 분석결과, 국내외 공통적으로 인문학 기반의 융합 연구들이 수행되었으며, 주로 사료의 디지털화가 주된 연구의 관심사였다. 국내는 문화콘텐츠와 스토리텔링 관련된 연구주제가 두드러지게 나타났으며, 해외는 디지털화된 자료의 제공 측면에서 문헌정보학 관련 주제들이 주목을 받는 것으로 나타났다. 이를 통해 국내외 디지털 인문학 분야의 연구 공백을 파악하고 연구주제 확장이 가능한 영역들을 제안하였다.

ABSTRACT

The purpose of this study is investigate the research trend on digital humanities. Previous studies focused on analyzing representative cases and national policies, not overall research trends in the digital humanities field. To this problem, this study intends to identify the intellectual structure of the digital humanities by adopting bibliometric approach. In this study, 1,765 articles retrieved from Web of Science and 514 records from RISS were analyzed to investigate research trends. Structural topic models were applied to examine research topics and to grasp the time-series trend. The results show that humanities-based convergence studies and digitization were main research interests in both side. In Korea, research topics related to cultural contents and storytelling were prominent, while in terms of providing digitized data, library and information science field was one of the important research topic abroad.

키워드: 디지털 인문학, 연구 동향, 구조적 토픽모델, 디지털화, 문헌정보학
digital humanities, research trend, structural topic model, digitization, LIS

* 본 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017S1A6A3A01079581).

** 연세대학교 근대한국학연구소 HK연구교수(yk.jeong@yonsei.ac.kr)

■ 논문접수일자: 2020년 5월 26일 ■ 최초심사일자: 2020년 6월 5일 ■ 게재확정일자: 2020년 6월 22일
■ 정보관리학회지, 37(2), 311-331, 2020. <http://dx.doi.org/10.3743/KOSIM.2020.37.2.311>

※ Copyright © 2020 Korean Society for Information Management
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

정보기술의 급속한 발전과 더불어 인공지능 기술은 이제 사회 전반의 다양한 문제에 적용되고 이를 해결하는데 다양하게 활용되고 있다. 이러한 경향은 학문 분야에도 영향을 상당한 영향을 미쳤는데, 인문학 분야에서도 계량분석이나 기계학습 등 디지털정보기술과 데이터 분석기법을 적용한 연구 활동들이 이루어지고 있다. 이렇게 디지털정보기술의 도움을 받아 새로운 방식으로 수행되는 연구들이 하나의 경향을 이루게 되었으며, '디지털 인문학'이라는 인문학의 새로운 세부 학문영역으로 자리 잡게 되었다.

디지털 인문학은 정보기술과 인문학의 융합을 통해 이루어지는 다양한 연구, 교육활동을 지칭하는 것이다(김현, 2013). 디지털 인문학 분야의 가장 초기 단계의 연구 프로젝트는 로베르토 부사가 토마스 아퀴나스의 저작을 디지털화한 프로젝트로, 이처럼 초기 단계에 수행된 연구들은 자료를 디지털화하고 이를 공유하고 보존하기 위한 아카이브를 구축하는 방식으로 이루어져 왔다. 이후에는 축적된 자료들을 대상으로 "넓게 보기"의 방식으로 데이터를 조망하는 연구들이 점차 많이 수행되었다(김바로, 2019; 이재연, 2016; 허수, 2014).

이에, 본 연구는 정보기술 발달에 따른 인문학 분야의 연구 경향 변화와 디지털 인문학의 연구 동향을 파악하고자 '디지털 인문학'을 주제로 이루어진 학술연구들을 대상으로 계량서지학적 분석을 수행하였다. 이를 위해 국내외 학술데이터베이스로부터 디지털 인문학 관련 연구(국외 1,765건, 국내 514건)들의 서지정보

를 수집하여 국내외 디지털 인문학 분야의 연구 동향을 비교 분석하였다.

2. 선행 연구

디지털 인문학은 전통적인 인문학의 연구주체들을 다루면서도 정보기술의 도입을 통해 새로운 방식으로 인문학연구를 수행하는 것으로(김현, 2013), 타 학문 분야에 비해 비교적 최근에 인문학의 하위 분야로 논의되기 시작하였다. 디지털 인문학의 가장 초기 단계에서는 인문학 관련 자료들을 디지털화하여 보존하는 것을 주된 목적으로 삼았다. 디지털 인문학의 시초로 여겨지는 연구는 이탈리아의 로베르토 부사의 작업으로(1949~1974), 토마스 아퀴나스와 관련된 작가들의 라틴어 텍스트를 디지털 방식으로 편찬한 것이었다(Schreibman, Siemens, & Unsworth, 2008). 이후 텍스트의 전산화를 위한 마크업 언어의 개발을 통해 인문학 텍스트를 디지털화하는 경향이 이어졌다.

국내의 초기 디지털 인문학 프로젝트 또한 과거 문헌이나 기록물 대상의 거대한 디지털화 작업 프로젝트 위주로 이루어져 왔다. 1967년부터 시작된 에드워드 와그너와 송호준의 '문과방목' 데이터베이스화 작업이 그것이다. 이 프로젝트를 통해 약 14,000여 명의 문과 합격자와 친인척 데이터를 디지털화하여 분석하였다. 이후에는 『조선왕조실록 CD-ROM』의 편찬을 통해 본격적으로 방대한 양의 자료를 디지털화하고 탐색할 수 있도록 하는 아카이빙 작업들이 이어졌다. 또한 국내외 소재의 한국 고전적에 대한 목록을 제공하는 한국고전적종합목록

시스템¹⁾이나 한국학 역사자료를 전산화하여 데이터베이스로 구축하고 검색서비스를 제공하는 한국역사정보통합시스템²⁾처럼 국가주도의 대규모 프로젝트들도 수행되었다.

해외에서는 영국 셰필드 대학과 허트포드셔 대학에서 1690년부터 1800년 사이에 영국 런던 거주민의 삶에 관계된 고문서 240,000여 건을 대상으로 인명 정보 중 동일인을 추적해 사람들의 생애를 구성한 런던 프로젝트가 대표적인 디지털인문학 프로젝트로 꼽힌다.³⁾ 이 프로젝트는 고문서를 디지털화하여 아카이빙하는 방식에서 그치는 것이 아니라, 자료의 가공을 통해 새로운 지식을 발견하는 수단으로 활용될 수 있음을 보여주었다. 최근 수행되고 있는 대규모 디지털 인문학 프로젝트들은 디지털 인문학을 활용하여 인문학 연구 분야를 재정의하거나 새로운 학문 분야를 개척하고 디지털 플랫폼을 구축하는 등 디지털 인문학 연구의 패러다임이 자료의 디지털화에서 활용과 데이터에 기반한 지식 창출로 변해가고 있다.

특히, 이러한 경향은 연구 대상의 규모가 증가하면서 양적 분석방법을 적용하는 방식으로 변화되어왔다. 디지털화된 자료로부터 자동으로 색인어를 추출하거나, 개념 간의 관계망을 구축하여 네트워크나 지도를 통해 그 내용을 전달하는 방식 등의 시도들이 이루어져 왔으며, 이렇게 컴퓨터의 사용과 전산화적 측면을 강조한 ‘인문학 전산화(Humanities Computing)’ 또는 ‘전산 인문학(Computational Humanities)’이라는 용어가 사용되기도 하였다(김재연, 2017).

유사하게, 디지털화된 저작이나 저작에 대한 출판, 디지털 큐레이션과 보존을 통해 학문 분야에서 디지털 자원을 이용하는 ‘디지털 스칼라십’이 주목받기도 하였다(Rumsey, 2011). 연구자가 초고속 인터넷망과 첨단장비, 첨단정보를 활용해 연구를 진행하는 ‘e-사이언스’(조만형, 2007), 인문학지식의 정보기술을 의미하는 ‘인문정보학(cultural informatics)’ 또한 비슷한 맥락으로 볼 수 있다(김현, 2013). 이러한 용어들은 인문학보다는 연구환경이 디지털화되고 전산화된다는 측면을 강조한 것으로 연구환경의 변화를 지칭하는 용어들이라고 볼 수 있다. 반면, 디지털 인문학은 인문학에서 사용하던 연구 소재들은 여전히 유지하되, 이미 구축된 디지털 아카이브로부터 의미 있는 정보를 찾아내는 데이터마이닝이나, 시각화로 연구영역을 확장하는 방향으로 이 용어가 자리 잡게 되었다(김현, 임영상, 김바로, 2016).

데이터 분석에 기반한 양적 연구로는 ‘물결 21’ 코퍼스처럼 대규모로 구축된 말뭉치를 대상으로 단어의 빈도변화와 자료의 특성을 파악하는 연구가 수행되었다(최재웅, 이도길, 2014). 국외에서는 이미 이러한 텍스트의 정량적 분석이 수행되고 있었는데, 영문학에서 문학 텍스트를 대상으로 정량적 텍스트 읽기를 시도한 Moretti(2005)의 ‘멀리 읽기(distant reading)’가 그것이다. 이 연구들의 주요 특성은 주로 전체코퍼스의 특성이나 특정 대상 단어의 빈도에 기반한 분석을 수행했다. 이와는 달리 대상이 되는 주제어나 개체명의 연관어 또는 맥락을 파

1) <https://www.nl.go.kr/korcis/>

2) <http://www.koreanhistory.or.kr/>

3) <https://www.londonlives.org/>

악하고자 하는 시도들이 있는데, 특정한 범위 안에서 함께 사용된 단어를 파악하는 동시출현(co-occurrence) 기법이 그것이다. 20세기 전반의 잡지에 나타난 '제국'의 의미를 어휘망을 통해 살펴본 허수(2014)의 연구가 대표적이며, 이재연(2016)은 『개벽』에 나타난 주요 단어들의 공기어를 통시적으로 살펴보는 방식으로 동시출현 분석을 연구에 적용하였다.

단어의 동시출현 분석이 단어 수준에서 텍스트를 분석했다면, 단어나 문헌의 군집으로 연구 대상을 확장할 경우, 유사한 문헌끼리 군집화하는 문헌 클러스터링이나 문헌자동분류, 토픽모델링 등의 기법을 적용하여 말뭉치의 특성을 파악할 수 있다. 이진오와 김동진(2019)은 적벽가의 이본들을 계통 분류하기 위해 계층적 클러스터링 기법을 사용하여 완관계열과 창본 계열로 분류하였으며, 이와 유사한 연구로 심청전을 대상으로 한 계통분류를 수행하였다(권기성, 김동진, 2019). 최근에는 딥러닝 기법 중 하나인 Word2Vec을 활용하여 불경텍스트에서 주요 단어들의 유사어들을 살펴보고, 이를 활용하여 불교학 디지털 온톨로지 구축을 제안한 연구도 진행되었다(김바로, 2019).

전반적인 디지털 인문학의 연구 동향에 대한 논의는 주로 해외의 디지털 인문학 사례들을 살펴보는 것으로 시작되었다. 김바로(2014)는 한국의 디지털 인문학 발전을 위한 토대로 미국, 대만, 일본, 중국의 디지털 인문학 정책과 연구 현황을 살펴보았다. 미국과 대만의 경우 정부의 강력한 정책적 지원을 통해 디지털 인문학을 발전시켜 왔으며, 특히 대만은 이러한 정부의 지원에 힘입어 아시아 디지털 인문학 분야에서 선도적인 역할을 하고 있음을 언급하였다. 반면

일본과 중국의 경우, 미미한 정부지원과 함께 상대적으로 자료의 디지털화에만 치중하는 디지털 인문학 초기 발전 단계를 밟고 있다고 지적하면서, 한국의 디지털 인문학의 발전을 위해 고려할 수 있는 위한 다양한 해외 선진사례를 소개하였다. 김종대(2014)는 독일의 대학교육 사례들을 중심으로 디지털 인문학 교육에 대해 살펴보았다. 인문학의 위기와 미래사회에 대한 인문학의 자구책으로서 직업과 연계된 디지털 인문학 교육의 필요성에 대해 언급하였다. 손정훈과 김민규(2018)는 프랑스의 디지털 인문학 현황과 정책 방향을 소개하면서 한국 디지털 인문학 변화를 위해 연구자와 연구소를 묶어 인큐베이팅하고 행정적 지원을 하는 인프라 구축과 아카이빙 서비스제공, 성과 공유 인프라 구축, 국제적 인적 네트워크 구축 등을 언급하였다. 이지수(2017)는 국내의 일본학 관련 디지털 인문학의 주요 프로젝트를 검토하면서 일본학과 해양문명 관련 연구들이 문부과학성의 지원하에 발전하고 있음을 지적하였다. 이 연구들은 해외 연구의 학문적 경향을 검토하기 보다는 국가의 특성에 초점을 맞추어 프로젝트들을 소개하고 정책적 영향력 분석에 비중을 두는 경우가 많았다. 이러한 흐름은 2018년 발간된 『역사학보』 240호 특집에서 각국의 디지털 역사학에 대해 다루는 것으로 이어졌다. 독일, 프랑스, 영국, 미국, 대만의 다양한 사례를 통해 각국의 디지털 역사학 연구들이 경향을 살펴 보았다.

근래에 와서야 한국의 디지털 인문학 분야에 대한 계량 서지적 분석이 수행되었는데, Lee & Lee(2019)는 국내에 출판된 학술논문을 통해 연구 동향을 파악하고, 디지털 인문학에 대해

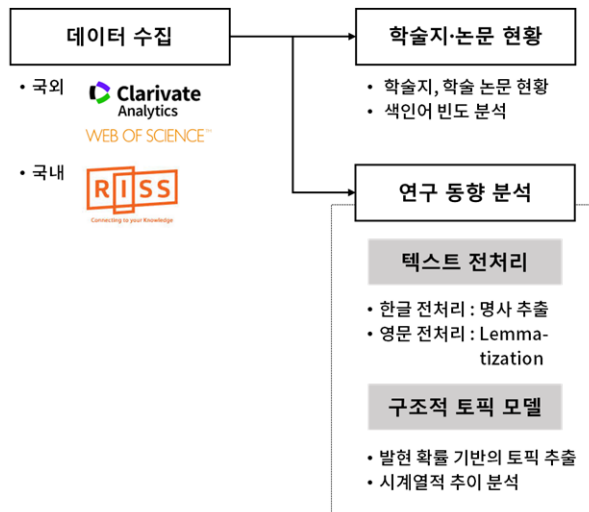
서 실제 연구자들이 가지고 있는 인식 조사를 수행하였다. 이 연구에서는 2017년까지 출판된 학술논문을 대상으로 하여 디지털 인문학 분야를 연구하는 저자의 수, 저자의 주요 연구 분야와 같은 연구자와 관련된 특성과 수록된 학술지들의 특성을 살펴보았다. 디지털 인문학과 관련된 논문을 출판한 연구자의 수가 점차 증가하는 추세였으며, 90% 이상의 연구자들이 인문·사회과학(Social Sciences and Humanities) 분야의 연구자들로 나타났다. 논문에 부여된 키워드를 인문·사회과학 분야와 과학기술 분야로 분류한 결과에서도 상당수의 용어가 인문학(General Humanities) 분야와 문학(Literature) 분야로 나타났다. 이러한 경향은 학술지의 학문 분야에서도 동일하게 인문·사회과학분야가 상당수를 차지하고 있었다. 이 외에도 이광석과 윤자형(2019)은 대학의 인문학 교육 측면에서 새로운 기술들을 어떻게 다루는지에 대해 논의하였다.

이처럼 디지털 인문학의 동향에 관한 논의들은 초기에 새로운 학문의 성격을 논하고 학문 영역을 구축하는 단계에서, 점차 세계 주요 국가들의 사례들을 중점적으로 살펴보는 방식으로 이루어져 왔으며, 최근에 와서야 학술논문들에 대한 서지적 분석이 이루어지기 시작하였다. 이에, 본 연구에서는 디지털 인문학과 관련된 학술지들과 학술논문들의 텍스트 분석을 통해 연구 동향을 살펴보고자 하였다.

3. 연구방법

3.1 연구의 개요

본 연구는 디지털 인문학 관련 학술 논문들의 지적 구조 및 연구의 동향을 파악하고자 계량정보학적 분석을 수행하기 위한 연구절차를 다음 <그림 1>과 같이 수립하였다.

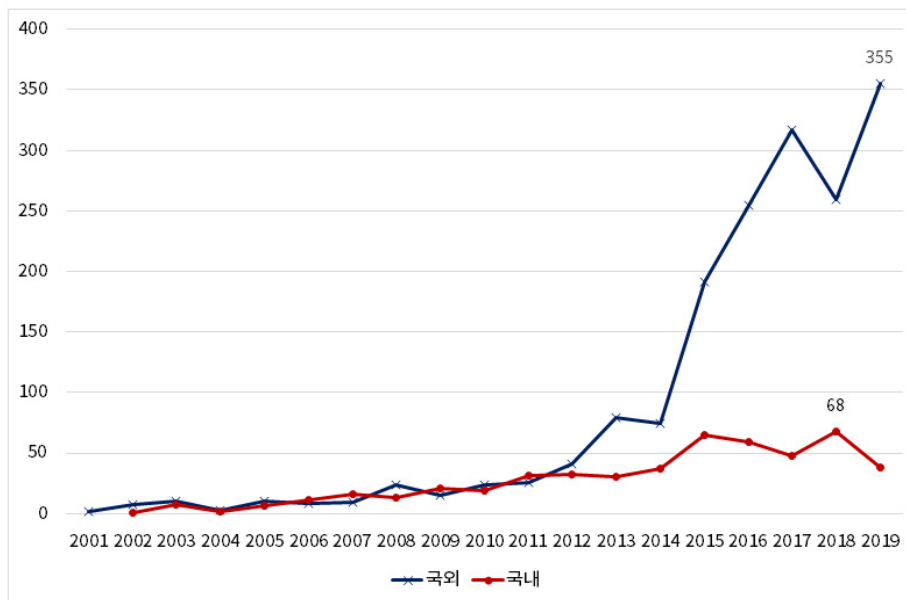


<그림 1> 연구의 개요

우선, 국내와 해외의 연구 동향을 비교 분석하기 위해 2020년 3월을 기준으로 국내의 주요 학술 데이터베이스로부터 학술지 논문의 메타데이터들과 초록을 수집하였다. 해외 동향으로는 Web of Science에 수록된 연구들의 서지정보를 수집하였으며, 국내 연구 동향을 위해서는 RISS 학술연구정보서비스로부터 데이터를 수집하였다. 국외에 비해 국내의 연구들이 상대적으로 적었기에, 가급적 다양한 국내 학술논문을 포함하여 비교하기 위해 DBPIA와 NDSL, KISS 등 여러 데이터베이스를 통합 검색할 수 있고 학술논문들의 서지정보와 초록을 모두 수집할 수 있는 RISS를 사용하였다. 각 데이터베이스를 대상으로 '디지털 인문학', 'digital humanities'라는 키워드를 사용하였으며, 그 결과로 해외에서 발행된 연구 논문 1,765건, 국내에서 수행된 연구 514건의 서지정보를 수집하였다. 다음 <그림

2>는 수집된 논문의 연도별 추이이다. 이 그래프는 디지털 인문학 연구의 증가 추이를 보여주기 위한 것으로 그 수가 작은 2001년 이전의 논문들과 2020년의 논문은 표에서 제외하였다. 디지털 인문학이라는 용어가 국내 연구에 등장한 2001년부터 2019년의 국내외 연구의 수를 표기하였다. 2001년 이전에는 해당 연도의 논문이 없거나 1~3건의 논문이 출판되었고, 2020년에는 국내 47건, 국외 9건이었다.

<그림 2>에서 확인할 수 있듯이 국내외 디지털 인문학 관련 연구는 2010년 이전까지는 미미하게 증가하는 추세였으나, 2015년 이후로는 국내외 연구 모두 상당한 증가를 보이는 것으로 나타났다. 그 시기는 국내에서 디지털 인문학의 이론적 체계화에 기여하는 저작들이나 연구 프로젝트들이 시작되는 시기와 맞물려 있다고 할 수 있는데, 2014년 한국연구재단의 “디지털



<그림 2> 연도별 디지털 인문학 관련 연구 문헌의 수(2001~2019년)

털 인문학 시각화 콘텐츠 개발” 사업이 시작되었으며, 2015년에는 『빅데이터 인문학, 진격의 서막: 800만권의 책에서 배울 수 있는 것들』이 번역되어 국내에 소개되었고, 2016년에는 디지털 인문학과 관련된 여러 전문가들이 모여 편저한 『디지털 인문학 입문』이 발간되어, ‘디지털 인문학’의 개념이 좀 더 보편화 되면서 디지털 아카이빙, 네트워크 분석, 텍스트마이닝을 적용한 연구들이 진행되었다(이재연, 2018). 국외에서도 이와 유사한 경향을 살펴볼 수 있는데, 2014~2016년을 전후로 ‘digital humanities’이라는 개념이 학술분야에 보편적으로 사용되었으며(이재연, 2018), 디지털인문학 관련 프로젝트와 디지털 아카이브 프로젝트들이 확대되었다.

다음으로는 수집된 연구들을 대상으로 디지털 인문학 분야의 대표적인 학술지들을 파악하고 수록된 논문들의 서지적인 특성을 살펴보았다. 주제적 측면에서의 연구 동향을 파악하기 위해서는 앞서 언급했듯 구조적 토픽모델을 사용하였고, 이를 바탕으로 시계열적 동향을 살펴보았다.

본 연구에서는 구조적 토픽모델을 적용하기 위해 학술논문들의 초록을 사용하였는데, 한글 초록과 영문 초록 각각을 대상으로 텍스트 전처리(text pre-processing)를 수행하여 토픽모델링의 입력 말뭉치로 사용하였다. 텍스트 전처리 과정은 말뭉치를 사용한 텍스트 분석 연구에서 결과의 신뢰도에 많은 영향을 미치는 중요한 부분으로, 일반적으로 형태소 분석과 어간 및 표제어 추출, 불용어 제거 등의 과정을 거친다. 한글 초록을 대상으로는 초록을 문장

으로 분리한 후 문장 단위의 형태소 분석을 통해 명사만을 추출하여, 초록 단위로 입력텍스트를 구성했다. 한글 형태소 분석을 위해서는 Python의 KoNLPy에 구현되어 있는 한나눔 형태소 분석기⁴⁾를 활용하였으며, 카이스트 태그셋의 체언(N) 중 수사(NN)를 제외한 명사들을 추출하여 사용하였다. 영문 초록의 경우, R의 textstem 라이브러리⁵⁾를 사용하여 표제어추출(lemmatization)을 수행하였고 이를 구조적 토픽모델의 입력텍스트로 사용하였다. 마지막으로 ‘디지털 인문학’, ‘digital humanities’, ‘digital humanity’와 같이 검색어로 사용되어 모든 문헌에 포함된 단어는 불용어로 간주하여 제외하였다.

3.2 구조적 토픽모델

토픽모델은 대규모 말뭉치로부터 잠재적 구조를 추출하는 비지도 학습 방법 중에 하나로, 단어를 벡터화하여 수치화시키는 단어 임베딩 기법과 맥을 같이한다. 토픽모델에서 가장 보편적으로 사용되는 잠재디리슈레탈당(LDA: Latent Dirichlet Allocation)은 랭킹이나 검색을 위한 잠재의미색인(LSI: Latent Semantic Indexing)에서 확률의 개념을 도입한 pLSI와 매우 유사하나, 확률분포가 디리슈레 분포를 따른다는 가정에서 차이가 있다(Ayadi, Maraoui, & Zrigui, 2015). LDA는 문서 안에 있는 단어들의 빈도에 기반하여 토픽을 추출하기 때문에, 학술 분야의 연구 동향을 파악할 때 연구자 주관에 영향을 받지 않고 텍스트의 통계적 분석

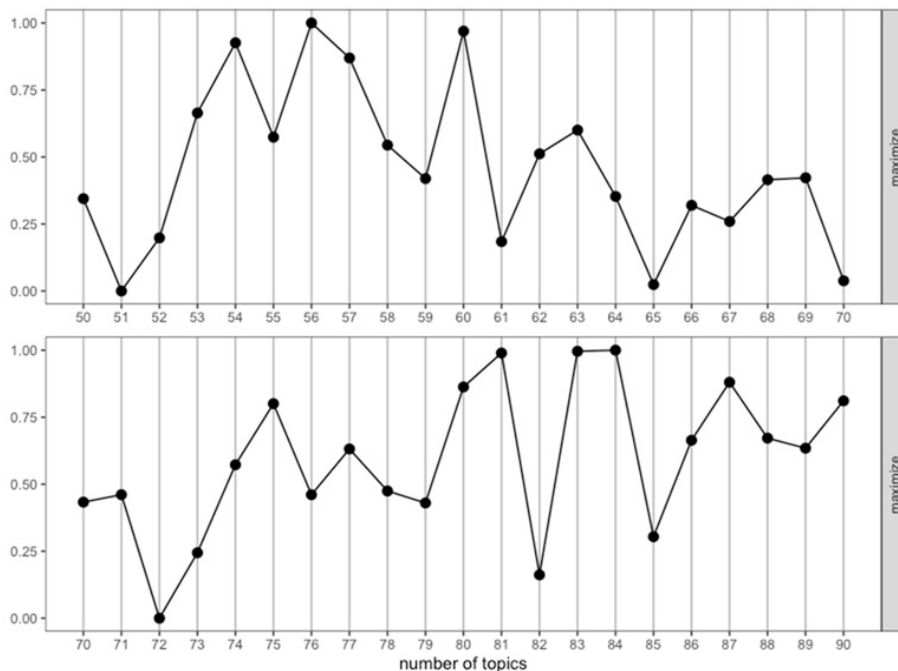
4) <https://konlpy.org/ko/latest/>

5) <https://github.com/trinker/textstem>

을 통해 연구주제를 파악할 수 있다는 장점이 있다(박종도, 2019). 그러나 추출된 토픽에 대한 적합성 측정의 문제와 사전에 토픽의 수를 임의로 선정해야 하는 문제, 토픽 간의 상관성이 존재할 가능성 등 여러 한계들이 제기되었다. 이러한 제한점을 해결하기 위해 LDA의 다양한 변형 모델들이 제시되었는데, 구조적 토픽모델은 문서에 존재하는 다양한 메타데이터를 활용하여 추출된 토픽모델의 적합성을 파악하고 각각의 토픽에 대한 발현확률의 통계검증을 수행하는 방식으로 문제점을 해결하고자 하였다(Roberts, Stewart, Tingley, Lucas, Leder-Luis, Gadarian, ... Rand, 2014). 본 연구에서는 디지털 인문학 분야의 세부 연구주제와 이에 따른 시계열적 동향을 살펴보기 위해 메타데이터로 학술논문의 출판연도를 사용하여 추이를 살펴보았

다. 구조적 토픽모델을 텍스트에 적용하기 위해서는 R의 stm 라이브러리(Roberts, Stewart, & Tingley, 2014)를 사용하였다.

앞서 언급하였듯, 토픽모델은 비지도 학습기반으로 토픽을 추출하기 때문에 모형의 적합성 평가가 어려운 면이 있다. 따라서 연구자가 자의적으로 토픽의 수를 사전에 지정하여 모형을 추정해야하기 때문에 선정된 토픽의 수가 적절한지, 추출된 토픽들이 원래의 말뭉치와 얼마나 일치하는지를 판단하는 것이 중요하다. 일반적으로는 최대우도추정과 잔차를 활용하여 토픽의 수를 선정하게 되는데, 본 연구에서는 Griffiths와 Steyvers(2004)가 제안한 로그우도 추정치를 사용하여 최적의 토픽 수를 산출하였다. 다음 <그림 3>은 토픽 수에 따른 추정치로 1에 가까울수록 적합성이 높다고 판단할 수 있다.



<그림 3> 토픽 수에 따른 로그우도 추정치(상: 국내, 하: 국외)

본 연구에서는 <그림 3>의 추정치를 기준으로 가장 값이 1에 가까운 국내 56개, 국외 84개를 토픽의 수로 선정하였다. 최종적으로는 구조적 토픽모델을 통해 추출한 토픽들 중 시계열 메타데이터 변수에 따른 발현확률(topical prevalence)이 통계적으로 유의한 토픽을 선별하여 국내는 16개, 국외는 25개의 토픽을 분석하여 제시하였다.

4. 연구 결과

4.1 학술지 및 학술논문 분석

본 절에서는 디지털 인문학 분야의 국내·외 주요 학술지들을 살펴보았다. 디지털 인문학 관련 연구들을 1건 이상 포함하고 있는 국내 학술지 및 학술대회 논문집은 총 262건이었으며, 해외의 학술지 및 학술대회 논문집들은 901건으로 나타났다. 디지털 인문학 분야의 국내외 주요 학술지와 논문의 수는 다음 <표 1>과 <표 2>에서 살펴볼 수 있다. <표 1>은 디지털

인문학 관련한 연구를 수록하고 있는 국내 학술지의 상위 10종으로 대부분 융복합 연구를 수행하는 기타인문학 분야로 분류되어있는 것을 알 수 있다. 디지털 인문학 관련 연구가 47건으로 가장 많이 실린 학술지는 『인문콘텐츠』지로 2014년 34호에서는 해외 디지털 인문학 동향에 관한 논문을 3편 수록하면서 본격적으로 디지털 인문학과 관련된 연구들을 다루기 시작하였다. 1970-1980년대에 창간된 전통적인 인문학 분야의 학술지들(『인문과학』, 『인문학연구』, 『열상고전연구』)이 있는 반면, 2000년대 이후에 생겨난 융복합 분야의 학술지들이 상위에 상당수 차지하고 있는 것을 알 수 있다. 분야로는 기타인문학, 기타국문학, 역사학 등 인문학 분야의 학술지들은 약 10% 내외로 디지털 인문학 분야의 논문들을 다루고 있었으나, 과학기술학이나 디자인, 예술 일반 분야의 학술지들은 수록논문의 수에 비해 매우 낮은 비율로 디지털 인문학 분야를 다루는 것으로 나타났다. 이러한 경향은 Lee & Lee(2019)의 연구에서 인문·사회학분야가 국내 디지털인문학 분야의 주류를 이루고 있는 것과 동일하게 나타났다.

<표 1> 디지털 인문학(DH) 관련 국내 학술지 상위 10종

순위	학술지명	창간연도	연구분야	논문건수	DH 논문건수
1	인문콘텐츠	2002	기타인문학	569	47
2	人文科學	1971	기타인문학	471	41
3	인문학연구	1974	기타인문학	741	25
4	디지털융복합연구	2003	과학기술학	5,446	24
5	열린정신 인문학연구	2000	기타인문학	353	23
6	디지털디자인학연구	2001	디자인	2,302	21
7	열상고전연구	1988	기타국문학	573	17
8	탈경계인문학	2008	기타인문학	200	17
9	기초조형학연구	2000	예술일반	4,079	14
10	개념과 소통	2008	역사학	162	12

〈표 2〉는 국외 학술지 상위 10종으로 Web of Science의 카테고리 정보와 Web of Science의 핵심 컬렉션에 수록된 논문 건수 및 디지털 인문학을 다루는 논문 건수이다. 상위 10종 중 'digital humanities'를 키워드로 가지고 있는 논문이 가장 많은 학술지는 *Digital Humanities Quarterly*로 107건의 논문을 수록하고 있었다. 국내 학술지와 가장 큰 차이는 주제 분야였는데, 국내의 학술지들이 주로 인문학과 융합 분야였다면, 국외 학술지의 경우 다학제적 성격의 'Humanities', 'Linguistics', 'History', 'Information

Science & Library Science'로 구성되어 있었다. 특히 국내 학술지에서는 없었던 'Information Science & Library Science' 분야의 학술지들이 상위에 있었다.

다음으로 디지털 인문학 관련 학술논문에 부여된 색인어를 통해 어떠한 세부 연구주제들을 다루고 있는지 살펴보았다. 각 논문에 부여된 색인어의 종류는 국내의 경우 3,843개, 국외의 경우에는 4,446개로 나타났다. 빈도 기준으로 가장 많이 나타났던 색인어는 다음 〈표 3〉과 같다.

〈표 2〉 디지털 인문학 관련 국외 학술지 상위 10종

순위	학술지명	창간연도	WoS 분야	논문건수	DH 논문건수
1	Digital Humanities Quarterly	2007	Humanities, Multidisciplinary	457	107
2	Digital Scholarship in the Humanities	1986	Humanities, Multidisciplinary; Linguistics	382	76
3	Literary and Linguistic Computing	1986	Linguistics: Language & Linguistics: Literature	342	56
4	College & Undergraduate Libraries	1994	Information Science & Library Science	161	31
5	Bibliothek Forschung und Praxis	1977	Information Science & Library Science	320	27
6	Historical Social Research	1979	History: History Of Social Sciences: Social Sciences, Interdisciplinary	1,067	25
7	International Journal of Humanities and Arts Computing	1994	Humanities, Multidisciplinary; Computer Science, Interdisciplinary Applications	82	25
8	Journal of Documentation	1945	Computer Science, Information Systems; Information Science & Library Science	1,839	22
9	Electronic Libraries	1983	Information Science & Library Science	1,696	16
10	PMLA-Publications of the Modern Language Association of America	1884	Literature	1,898	14

〈표 3〉 국내외 학술논문에 부여된 색인어

순위	국내 색인어	빈도	국외 색인어	빈도
1	인문학	40	Humanities	47
2	문화콘텐츠	39	Digital history	30
3	스토리텔링	25	Digital libraries	30
4	빅데이터	17	GIS	30
5	콘텐츠	12	Big Data	28
6	인공지능	10	Open Access	26
7	데이터베이스	9	Digitization	23
8	디지털시대	9	Academic Libraries	22
9	문화산업	9	Collaboration	22
10	융합	9	Data Visualization	20

우선, 국내외 연구 모두에서 가장 보편적으로 ‘인문학’, ‘Humanities’가 가장 많이 언급된 색인어로 나타났다. 이는 기존의 디지털 인문학 연구의 논의 중 다수가 인문학이라는 학문적 정체성 위에 다양한 기술이나 외부의 콘텐츠 등과 융합을 시도하는 것이라 볼 수 있다. ‘빅데이터’와 ‘Big Data’ 또한 국내외 연구에서 모두에서 주된 연구주제로 등장했다. 반면, 문화콘텐츠, 콘텐츠, 문화산업 등의 주제들은 국내의 연구에서 주로 다루어졌으나, 국외에서는 ‘Digital libraries’, ‘Academic Libraries’, ‘Open Access’와 같은 도서관 관련 주제들이 다루어졌다. 이는 앞서 국외 학술지에서 ‘Information Science & Library Science’ 관련 분야를 다루던 것과 같은 맥락이라고 볼 수 있다. 이러한 색인어의 차이를 살펴보았을 때, 국외에서는 인문학과 다학제적 융합 연구 이외에 문헌정보학과 관련된 연구내용 또한 비중있게 다루고 있음을 알 수 있다.

4.2 구조적 토픽모델 기반의 연구주제

국내의 디지털 인문학 분야의 주요 연구주제

와 동향을 파악하기 위해, 앞서 도출한 구조적 토픽모델링을 수행하여 토픽을 추출하였으며, 메타데이터에 따른 토픽 발현확률이 높은 국내는 16개, 국외는 25개의 토픽을 선별하였다.

〈표 4〉는 국내 학술논문을 대상으로 토픽모델링을 수행한 결과이다. STM 결과 추출된 단어들은, 상위 색인어 〈표 3〉에서 언급되었던 연구주제들과 유사한 결과를 보였으나, 좀 더 세분화된 연구주제들이 나타났다. 문화콘텐츠(T8), 온라인 콘텐츠(T14)와 같이 콘텐츠 관련 토픽과 내러티브적인 측면이나 스토리텔링을 다루는 스토리텔링(T7), 이야기/스토리(T16), 구비문학(T26)과 같은 주제들이 세분화되어 나타났다. 기술적 측면이나 방법론과 관련된 토픽으로는 방법론(T10), 도구/플랫폼(T29), 정보기술(T22), 텍스트(T49) 관련 토픽들이 등장하여 디지털 인문학의 세부 연구주제로 다루어지고 있었다. 상위 색인어에서는 나타나지 않았지만, 도서관(T32), 전자책(T42), 글쓰기(T3), 아카이브(T53)와 같이 도서관, 독서활동과 관련된 영역도 나타났다. Lee & Lee(2019)의 연구에서 중요 하위 분야로 나타났던 ‘교육’ 관련 주제는 토픽모델링 결과에서 중요하게 나

〈표 4〉 국내 학술논문의 STM 결과(16개 토픽)

순번	주제명	주제 단어
T1	빅데이터	빅데이터, 데이터, 디지털화, 인문학자, 컴퓨터, 정보학, 아이패드, 사물인터넷, HCI, 프로젝트
T3	글쓰기	글쓰기, 스토리, 디지털, 인문학, 테크놀로지, 수필문학, 영상매체, 가능성, 학술연구, 대중문학
T4	디자인	디자인, 디자이너, 창의성, 다이어그램, 프로세스, 전문인력, 건축가, 들뢰즈, 서비스, 커뮤니케이션
T7	스토리텔링	스토리텔링, 콘텐츠, 스토리, 미디어, 김시습, 가능성, 트랜스, 캐릭터, 국어국문학, 코러스
T8	문화콘텐츠	콘텐츠, 문화콘텐츠, 디지털콘텐츠, 불교문화, 활성화, 지역문화, 스마트, 지식정보, 문화원, 융합형
T10	방법론	방법론, 한국, 시각화, 프로젝트, 연구방법론, 연구자, 역사학, NEH, 영문학계, 조선시대
T14	온라인 콘텐츠	콘텐츠, 텍스트, 온라인, 오리지널, 플랫폼, 테마파크, 프로토타입, 증강현실, 문화콘텐츠, 입체적
T16	이야기/스토리	이야기, 정체성, 건국신화, 주인공, 애니메이션, 스토리, 고전문학, 국문학, 주몽신화, 환상성
T22	정보기술	전자문화, 상상력, 스마트폰, 시각화, 정보기술, 조선족, 효율적, 세부적, 세계관, 문화정보
T23	4차산업혁명	산업혁명, 디지털화, 인더스트리, 문화원, 노동, 산업, 문화자원, 페러다임, 시스템, 미래, 교육정책
T26	구비문학	구술성, 구비문학, 통시적, 재구성, 아날로그, 문화사, 인구학, 디지털, 핵심, 시사
T29	도구/플랫폼	저작도구, 플랫폼, 인문학, 고문헌, 디지털기술, 텍스트, 키워드, 분석, 정보기술, OCR
T32	도서관	도서관, 서비스, 공공도서관, 인프라, 디지털, 사서직, 지역문화, 의료서비스, 시스템, 지역학, 교육용
T42	전자책	전자책, 콘텐츠, 테크놀로지, 이해관계자, 출판사, 디지털화, 커뮤니케이션, 독서문화, 기성작가, 문학 개념
T49	텍스트	텍스트, 데이터, 시각화, 관계망, 네트워크, 창작자, 방법론, 키워드, 구조화, 빅데이터
T53	아카이브	아카이브, 데이터베이스, 조선시대, 데이터, 온톨로지, 디지털화, 시스템, 컴퓨터, 효율적, 표준화

타나지 않았는데, 세부 주제로 다뤄지기보다는 도서관의 교육용 인프라나 4차산업혁명 관련 교육정책 등의 주요 주제와 연관된 하위 주제로 나타났다.

국외에서 수행된 연구들의 경우, 국내의 연구주제들보다 좀 더 다양하고 세분화된 주제들로 토픽이 추출되었다. 'Text Annotation(T1)', 'Corpus(T60)', 'Text Analysis(T65)' 등 텍스트, 말뭉치 분석과 관련된 세부 주제가 나타났으며, 역사·문화 관련된 토픽은 'Cultural study(T14)', 'Historical Project(T21)', 'History and Culture(T49)'로 국내의 연구주제가 문화콘텐츠와 관련되어 있었다면, 국외에서는 역사학 관련 프로젝트들이 비중있게 다루어지는 것을 알 수 있었다. 데이터 자원에 관한 토픽(T3, T68)이나 시각화, 지도와 관련된 토픽인 'Image(T8)', 'City Map(T10)', 'Map(T30)' 또한 국내의

연구에 비해 많이 다루어졌으며, 특히 역사적 자료와 지도의 시각화가 주요 세부연구 주제로 나타났다. 정보기술과 관련된 토픽들인 'Technology(T37)', 'Technology Change(T55)', 'Information Science(T62)', 'Tool(T73)'들도 다양하게 나타났다.

반면, 국내의 연구에서는 중요하게 다루어지지 않았던 분야인 문헌정보학, 교육, 소셜미디어 등의 주제도 다루어졌다. 문헌정보학이나 기록·보존과 관련된 토픽은 국외의 연구들에서 훨씬 다양하게 다루어지고 있었다. 'Journal Publication(T18)', 'Library Service(T31)', 'Archive(T44)', 'Open Access(T51)', 'Print(T64)'와 같이 학술논문의 출판이나 학술지에 대한 이용, 도서관 봉사와 같은 토픽들이 그것이다. 'Education(T69)'의 경우에도 대학에서의 정보기술 교육에 관한 것으로, 국내에서는

〈표 5〉 국외 학술논문의 STM 결과(25개 토픽)

순번	주제명	주제 단어
T1	Text Annotation	text, annotation, read, content, analysis, early, technology, article, way, digitization
T3	Dataset	data, information, historical, semantic, system, dataset, big, spatial, knowledge, link
T7	Interdisciplinary Science	science, computer, interdisciplinary, new, philosophy, history, technology, problem, art, author
T8	Image	image, visual, work, object, collection, scholar, method, museum, archive, form
T10	City Map	city, computer, urban, map, study, make, archive, kittler, information, present, knowledge
T14	Cultural Study	cultural, publish, study, scholarly, edit, access, library, heritage, literature, scholarship
T18	Journal Publication	journal, science, publish, publication, citation, social, academic, index, article, scientific
T21	Historical Project	project, historical, resource, library, information, paper, purpose, provide, development, data
T30	Map	art, map, history, historical, method, project, spatial, question, geography, place
T31	Library Service	library, service, access, literature, university, community, provide, information, new, academic
T35	Manuscript	manuscript, project, make, text, essay, edition, material, discuss, article, transcribe
T37	Technology	information, development, technology, application, knowledge, new, content, activity, process, environment
T44	Archive	science, document, preservation, social, archive, research, sound, study, archival, library
T49	History and Culture	heritage, cultural, project, information, culture, medieval, development, system, platform, value
T51	Open Access	open, discourse, communication, science, scientific, practice, access, scholarly, social, publication
T55	Technology Change	technology, new, change, information, system, generation, social, critical, intelligence, year
T60	Corpus	analysis, corpus, read, apply, result, text, method, literary, term, distant
T62	Information Science	information, research, science, social, academic, data, organization, system, internet, practice
T64	Print	print, work, book, analysis, article, development, century, data, network, corpus
T65	Text	text, language, corpus, encode, tei, word, linguistic, resource, analysis, process
T68	Resource/ Data	source, data, historical, use, material, music, project, information, online, work
T69	Education	student, university, learn, teach, education, teacher, course, classroom, technology, skill
T73	Tool	technology, tool, human, education, age, care, machine, new, process, identity
T80	Social Media	twitter, read, medium, social, study, story, different, report, tweet, character
T84	Google Project	google, place, book, material, world, narrative, text, corpus, GAP, classical

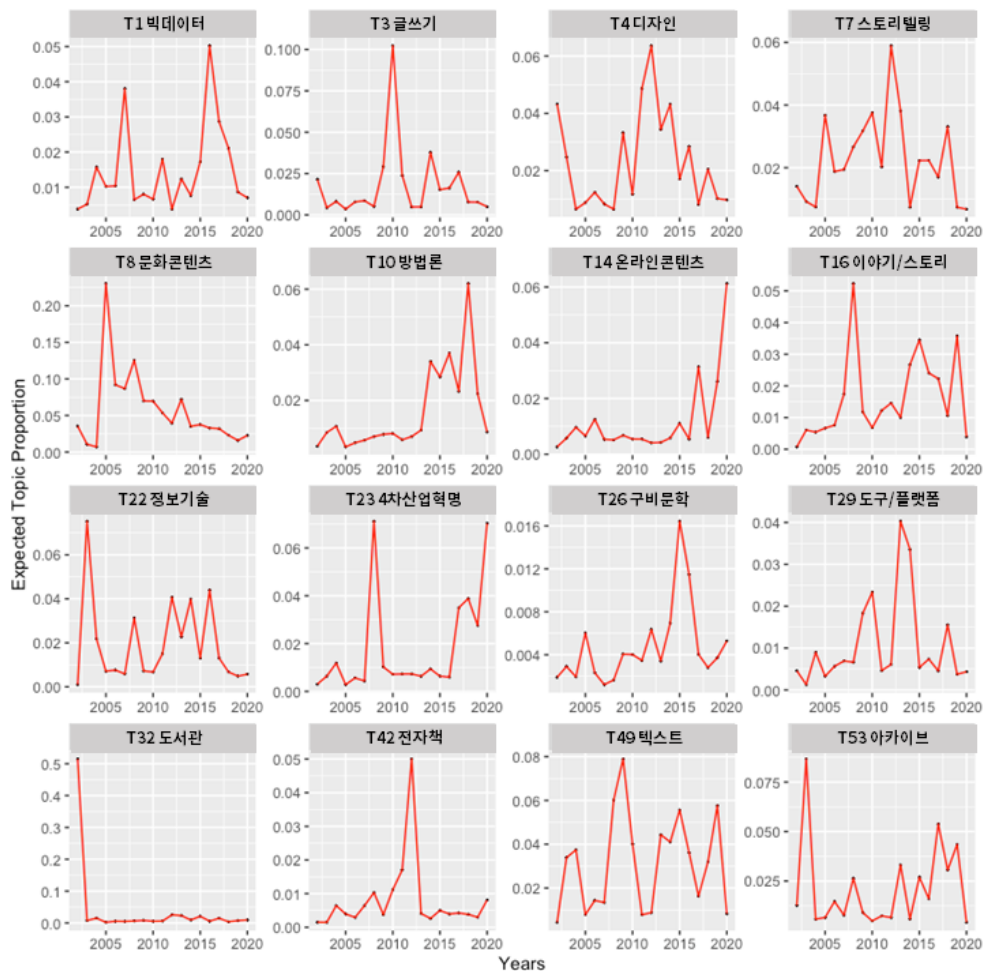
따로 다루어진 영역은 아니었다. 마지막으로 트위터와 연관된 'Social Media(T80)'와 'Google Books Library Project(T84)'와 같이 디지털 자원과 관련된 주제도 특징적으로 다루어졌다.

4.3 시계열 분석

본 절에서는 구조적 토픽모델을 통해 추출된 연구주제들이 시간이 지남에 따라 어떻게 변화

하는지에 관한 연구 동향을 살펴보았다. 토픽들이 시간에 따른 증가·감소 추세를 도출하기 위해 메타데이터 중 연도정보를 토픽에 대한 공변량으로 설정하여 토픽들의 추이를 추정하였다.

〈그림 4〉는 국내 연구논문들의 주요 토픽별 시계열 추이를 보여준다. 우선, 대부분의 토픽들은 시기별 토픽의 발현 비중이 0.1 이하로 매 시점에서 중요하게 언급되지 않는 것으로 나타



〈그림 4〉 국내 주요 토픽의 시계열적 발현 추이

났다. 그렇지만 세부적으로는 어떤 시기에 특정 토픽에 속한 단어들이 많이 나타났는지의 추이들은 파악할 수 있다. 예를 들어, 아카이브(T53) 토픽은 2003년에 상대적으로 많이 나타났으며, 그 이후로도 꾸준히 언급되었다. 실제로 조선 시대의 지도연구와 국학 자료, 고건축, 고문서 정리 등의 연구가 2003년에 이루어졌다.

특정한 연도에서 토픽이 많은 변화를 보이는 경우는 문화콘텐츠(T8)와 도서관(T32) 토픽이었다. 문화콘텐츠와 관련된 토픽은 2005년에 다루어질 확률이 높은 것으로 나타났으며 그 이후 점차 하강하는 추세를 보여주었다. 이와 관련하여 2005년에는 불교 문화, 삼국지 등 고전 문화콘텐츠와 관련된 연구들이 수행되었다. 도서관(T32)과 관련된 토픽은 2002년에 다루어질 발현확률이 높은 것으로 나타났는데, 디지털 시대로의 변화에서 문헌정보학과 정보전 문직이 어떠한 역할을 해야 하는지에 대해 다룬 연구들이 나타났다.

다음으로 국외에서 다루어진 연구들 또한 대부분 주요 토픽이 시기에 따라 발현할 확률이 높지 않았다. <그림 5>에서 볼 수 있듯이 토픽들의 기대되는 토픽의 발현 비중이 0.1 이하로 나타나는 것이 많았다.

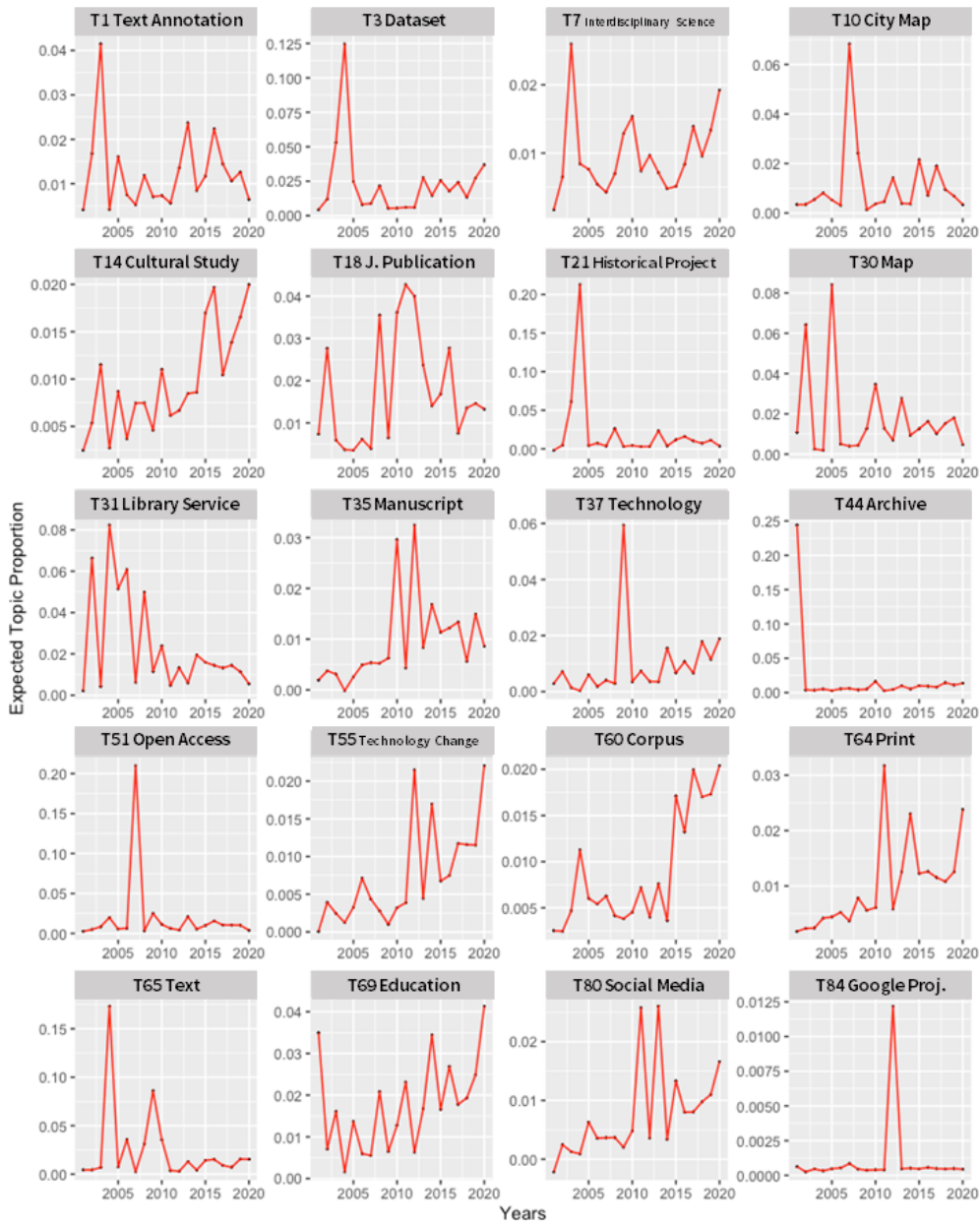
국외의 몇몇 토픽들은 해당 시기에 핵심적인 논문이나 사건에 영향을 받은 것으로 나타났다. 도시의 지도 시각화와 관련된 토픽(T10)은 2008년에 상대적으로 많이 다루어졌으며, 원고/글쓰기(T3) 토픽은 2010년대 중반, 기술 관련 토픽(T37)은 2009년, 아카이브(T44)는 2001년, 오픈 액세스(T51)는 2007년, 구글 프로젝트(T84)가 2012년 많이 나타났다. 문화 관련 토픽(T14)과 기술의 변화 관련 토픽(T55)은 점

차 증가하는 경향을 보였다. 아카이브의 경우 2001년 영국의 디지털 도서관 컬렉션과 같은 연구가 진행되었으며, 구글 북스 라이브러리 프로젝트와 관련된 경우, 해당 시기에 'Google's Digital Humanities Research Awards'와 'Google Books', 'Google Ancient Project(GAP)' 등의 이슈가 다루어졌다. 오픈 액세스 관련 토픽은 해당 시기에 'arXiv'와 디지털 자원의 보존과 접근의 측면에서 연구들이 이루어졌다.

역사 프로젝트(T21)와 관련된 주제는 2000년대 중반에 토픽의 발현확률이 높아지는데, 역사텍스트 자료들의 디지털화 작업과 구조적인 표현, 이에 관한 온톨로지 등의 연구들이 이 시기에 이루어졌다. 지도 및 시각화와 관련된 주제(T30)들 또한 2000년대 중반에 주로 이루어졌는데, 지도자료와 GIS 정보를 활용한 디지털화를 다룬 연구들이었다. 이처럼 2000년대 전반적으로는 역사자료, 지도자료, 문헌의 디지털화가 전반적으로 진행되었던 시기라고 볼 수 있다.

도서관 서비스(T31)와 관련된 토픽은 디지털 도서관과 더불어, 도서관 자원에 대한 검색과 자료접근 등이 연구주제로 나타났다. 도서관을 통해 자원에 대한 접근과 서비스를 제공하는 것과 관련된 연구주제가 주로 2000년대 중반에 이루어졌음을 알 수 있다. 텍스트 처리와 관련된 주제(T65)의 경우, 텍스트의 디지털화 과정에서 고려해야 할 요소 중 인코딩과 관련된 요소기술과 분석과 관련된 연구들이 해당 시기에 이루어졌다.

국내외 토픽 발현에 따른 연구경향의 차이를 살펴보면, 국내에서 주로 상승하는 연구주제로는 방법론과 도구/플랫폼에 대한 관심사가 커



〈그림 5〉 국내 주요 토픽의 시계열적 발현 추이

지는 것을 확인할 수 있었고, 문화콘텐츠와 관련된 주제들은 하강하는 추세를 보였다. 국외의 경우에는 기술변화나 소셜미디어, 교육, 말뭉치

관련 연구들이 전반적인 상승세를 보였으며, 시각화와 관련된 주제나 역사학 관련 프로젝트는 그 언급이 줄어드는 것을 확인할 수 있었다.

5. 결론 및 논의

본 연구에서는 2000년대 이후에 수행된 국내외 디지털 인문학 분야의 연구 동향을 비교하고 시계열적 추이를 살펴보았다. 기존의 디지털 인문학 동향과 관련된 연구들은 국가별 사례나 프로젝트 등을 소개하고 분석하여 앞으로 한국의 디지털 인문학이 나아갈 방향이나 전망 등을 다루는 질적 연구들이 다수를 차지하였다. 본 연구에서는 특정 국가나 사례에 연구의 범위를 한정하지 않고, 디지털 인문학이라는 큰 범주 안에 속하는 다양한 연구들을 수집하여 세부 연구주제들을 파악하고 그 추이를 살펴보려고 하였다. 이를 위해 2020년 3월까지 출판된 국내, 국외의 학술논문들의 서지정보와 초록을 수집하였으며, 이 데이터를 대상으로 초록에 구조적 토픽모델을 적용하여 분석을 수행하였다.

전반적으로 국내의 연구들은 역사학 분야의 사료 디지털화와 문화콘텐츠 측면에 디지털화가 주된 연구영역으로 나타났으며, 이에 따른 스토리텔링이나 국문학 분야의 문학이 주된 연구 관심사로 나타났다. 국외의 연구에서 역시 역사학, 문화콘텐츠가 주된 연구영역으로 나타났으며, 특징적으로는 지도와 같은 시각화자료에 대한 연구가 이루어졌다. 그러나 국외에서는 도서관의 서비스 제공이나 학술지의 출판, 오픈 액세스와 같은 문헌정보학 관련 분야가 두드러지게 나타났다. 이 외에도 소셜미디어에 대한 관심이나 구글 북스 라이브러리 프로젝트 등 특정 데이터 자원을 대상으로 한 연구주제들도 나타났다. 시계열적 분석에서는 각 주제들이 상승하거나 하강하는 주제들을 파악할 수 있었는데,

기술변화와 관련한 주제들은 국내외 모두에서 상승하는 경향을 보이는 것으로 나타났다.

이상의 연구 경향을 살펴보면 국내외에서 모두 디지털 인문학과 관련된 연구들은 아직까지 자료의 디지털화가 중요한 연구주제 중에 하나이며 특정한 연구방법론이나 이를 활용한 개별 연구들은 이제 막 시작하는 단계라는 것을 알 수 있다. 국외에서 도서관이 디지털 인문학 관련 자료의 제공창구와 정보기술 교육을 제공하는 기관으로서 언급되고 있으므로, 국내에서도 해당 주제들을 고려한 연구 범위의 확장이 가능할 것이다. 특히, 도서관 측면에서 디지털 인문학 성과물들을 이용자에게 서비스하는 방식의 연구들은 국내에서 거의 이루어지고 있지 않기 때문에 이 주제에 대한 논의가 필요하며, 기술 자원이나 방법론 등에 대한 확장 연구들도 수행될 수 있을 것이다. 또한, 국외에서는 텍스트에 대한 분석이 Corpus, Annotation, Manuscript 등 다양한 측면에서 이루어지고 있으므로, 국내에서도 텍스트에 세부적인 특성이나 분석방법을 다각화하는 연구들이 추가로 이루어질 수 있을 것이다.

디지털 인문학의 연구 대상과 관련해서는, 소셜미디어와 같은 데이터 자원이나, 구글 프로젝트, 기술변화와 관련된 주제들이 국외에서 중요한 주제로 발견된 것에 반해, 국내에서는 아직까지 역사, 문화콘텐츠에 대상이 한정되어 있음을 알 수 있었다. 따라서 연구 대상의 다각화를 통하여 국내 디지털 인문학 분야의 연구 외연 확장도 가능할 것이다.

마지막으로 본 연구의 제한점으로는 '인문정보학', '전산인문학', '인문정보학' 등 디지털 인문학을 지칭하는 다양한 키워드를 사용하지 않

고 데이터를 수집했다는 점을 들 수 있다. 데이터의 범위가 2000년대 초반부터로 한정되었고, Web of Science와 RISS라는 특정 학술 데이터베이스를 대상으로 자료를 수집하였기에 그 연구 대상이 한정적이었다. 이를 보완하기

위해 향후 연구로 디지털 인문학 분야를 지칭할 수 있는 다양한 표현들과 여러 학술데이터베이스를 망라적으로 포함하여 분석하고, 디지털 인문학 분야의 연구 동향을 보다 정확하게 파악할 필요가 있을 것이다.

참 고 문 헌

- 권기성, 김동건 (2019). 컴퓨터 문헌 분석 기법을 활용한 <심청전> 계통 분류 연구. *판소리연구*, 47, 167-205.
- 김바로 (2014). 해외 디털인문학 동향. *인문콘텐츠*, 33, 229-254.
- 김바로 (2019). 딥러닝으로 불경 읽기 - Word2Vec으로 CBETA 불경 데이터 읽기. *원불교사상과종교문화*, 80, 249-279.
- 김재연 (2017.9.5). 디지털 인문학, 인문학 변화를 이끌 것인가. *주간경향*, 1242.
<http://weekly.khan.co.kr/khnm.html?mode=view&artid=201708281902091&code=116>
- 김중대 (2014). 독일의 디지털인문학 동향. *철학과 문화*, 32, 55-76.
- 김현 (2013). 디지털 인문학 - 인문학과 문화콘텐츠의 상생 구도에 관한 구상. *인문콘텐츠*, 29, 9-26.
- 김현, 임영상, 김바로 (2016). *디지털 인문학 입문*. 서울: HUEBOOKs.
- 박중도 (2019). 토픽 모델링을 활용한 다문화 연구의 이슈 추적 연구. *한국문헌정보학회지*, 53(3), 273-289. <http://dx.doi.org/10.4275.KSLIS.2019.53.3.273>
- 손정훈, 김민규 (2018). 프랑스의 디지털인문학 동향과 정책. *프랑스문화예술연구*, 64, 125-151.
<https://doi.org/10.21651/cfaf.2018.64..125>
- 에레즈 에이든, 장바티스트 미셸 (2015). 빅데이터 인문학, 진격의 서막: 800만 권의책에서 배울 수 있는 것들. (김재중 옮김). *과주: 사계절*.
- 이광석, 윤자형 (2019). 국내 디지털인문학의 정착과 굴곡: 대학 교육과 미디어 테크놀로지의 불안정한 접속. *한국언론정보학보*, 95, 9-34. <http://dx.doi.org/10.46407/kjci.2019.06.95.9>
- 이재연 (2016). 키워드와 네트워크: 토픽 모델링으로 본 [개벽]의 주제 지도 분석. *상허학보*, 46, 277-334.
- 이재연 (2018). 디지털 시대의 인문학에서 디지털 인문학 시대로 한국문학에서 본 디지털 인문학 연구. *역사학보*, 240, 157-190. <https://doi.org/10.16912/tkhr.2018.12.240.157>
- 이지수 (2017). 디지털 인문학의 정의 및 연구 현황: 일본학 지식정보의 디지털화를 중심으로. *일본어문학*, 74, 41-62. <https://doi.org/10.18704/kjill.2017.09.74.41>

- 이진오, 김동건 (2019). 컴퓨터 문헌 분석 기법을 활용한 <적벽가> 이본의 계통 분류 연구. 한국콘텐츠학회논문지, 19(6), 1-9. <http://dx.doi.org/10.5392/JKCA.2019.19.06.001>
- 조만형 (2007). 국가 e-사이언스 전략이 필요하다. *The Science & Technology*, (12), 12-13.
- 최재웅, 이도길 (2014). 물결 21 코퍼스: 공개 웹 자원 및 활용 도구. *민족문화연구*, (64), 3-23. <http://dx.doi.org/10.17948/kcs.2014.64.3>
- 허수 (2014). 어휘 연결망을 통해 본 '제국'의 의미 - '제국주의'와 '제국'을 중심으로. *대동문화연구*, 87, 501-562. <https://doi.org/10.18219/ddmh.87.201409.501>
- Ayadi, R., Maraoui, M., & Zrigui, M. (2015, October). LDA and LSI as a dimensionality reduction method in arabic document classification. In *International conference on information and software technologies* (pp. 491-502). Springer, Cham.
- Gold, M. K., & Klein, L. F. (Eds.). (2016). *Debates in the digital humanities 2016*. U of Minnesota Press.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl 1), 5228-5235.
- Lee, J., & Lee, H. E. (2019). Digital humanities and new directions in South Korea. *Digital Scholarship in the Humanities*, 34(4), 772-790.
- Moretti, F. (2005). *Graphs, maps, trees: abstract models for a literary history*. Verso.
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2014). stm: R package for structural topic models. *Journal of Statistical Software*, 10(2), 1-40.
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open ended survey responses. *American Journal of Political Science*, 58(4), 1064-1082. <https://doi.org/10.18637/jss.v091.i02>
- Rumsey, A. (2011). *Scholarly Communication Institute 9 - New-Model Scholarly Communication: Road Map for Change*. University of Virginia Library <http://uvasci.org/institutes-2003-2011/SCI-9-Road-Map-for-Change.pdf>
- Schreibman, S., Siemens, R., & Unsworth, J. (Eds.). (2008). *A companion to digital humanities*. John Wiley & Sons.

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

Aiden, Erez, & Michel, Jean-Baptiste (2015). *Uncharted: Big data as a lens on human culture*.

- Riverhead Books, 2013.
- Choe, Jae-Woong, & Lee, Do-Gil (2014). Trends 21 corpus: Public web resources and search tools. *Korean Cultural Studies*, (64), 3-23. <http://dx.doi.org/10.17948/kcs.2014..64>.
- Hur Soo (2014). The meaning of 'jegook(帝國)' in corpus networks - centering on the analysis of 'imperialism' and 'empire'. *Journal of Eastern studies*, 87, 501-562. <https://doi.org/10.18219/ddmh.87.201409.501>
- Jo, Man-Hyeong (2007). The need of national 'e-Science' strategy. *The Science & Technology*, (12), 12-13.
- Kim, Ba-Ro (2014). Trend of digital humanities. *Korea Humanities Content Society*, 33, 229-254.
- Kim, Ba-Ro (2019). Reading buddhist scriptures with deep learning - Reading CBETA buddhist scriptures data with Word2Vec. *Won-Buddhist Thought & Religious Culture*, 80, 249-279.
- Kim, Hyeon (2013). Digital humanities: Cooperative scheme between humanities and cultural contents. *Korea Humanities Content Society*, 29, 9-26.
- Kim, Hyeon, Yim, Young-Sang, & Kim, Ba-ro (2016). *Digital humanities*. Seoul: HUEBOOKS.
- Kim, Jae Yon (2017, September, 5). Digital humanities, Will it lead to changes in humanities. *The Weekly Kyunghyang*, 1242. <http://weekly.khan.co.kr/khnm.html?mode=view&artid=201708281902091&code=116>
- Kim, Jong-dae (2014). Digital humanities in deutschland. *Philosophy & Culture*, 32, 55-76.
- Kwon, Ki-Sung, & Kim, Dong-Keon (2019). The classification of Sim Cheong-jeon through a computer analysis technique of bibliographies. *Journal of Pansori*, 47, 167-205.
- Lee, Jae-Yon (2016). Keywords AND networks - exploring the thematic maps of kaebŏk through topic modelling. *Sanghur Hakbo*, 46, 277-334.
- Lee, Jae-Yon (2018). From humanities in the digital age to the age of digital humanities: History of adopting DH in Korean literary studies. *The Korean Historical Review*, 240, 157-190. <https://doi.org/10.16912/tkhr.2018.12.240.157>
- Lee, Jin-O, & Kim, Dong-Keon (2019). A study on the classification of jeokbyeok-ga's version by the computer analysis technique of bibliographies. *Journal of the Korea Contents Association*, 19(6), 1-9. <http://dx.doi.org/10.5392/JKCA.2019.19.06.001>
- Lee, Ji-soo (2017). Definition and research status of digital humanities: Focusing on digitization of knowledge information in Japanese studies. *Korean Japanese of Japanese Language and Literature*, 74, 41-62. <https://doi.org/10.18704/kjill.2017.09.74.41>
- Park, Jong Do (2019). A study on issue tracking on multi-cultural studies using topic modeling. *Journal of the Korean Society for Library and Information Science*, 53(3), 273-289.

[http://dx.doi.org/10.4275.KSLIS.2019.53.3.273](http://dx.doi.org/10.4275/KSLIS.2019.53.3.273)

Son, Jeong Houn, & Kim, Min Kyu (2018). La tendance et la politique des humanités numériques en France. *Etudes de la Culture Française et des Arts en France*, 64, 125-151.

<https://doi.org/10.21651/cfaf.2018.64..125>

