

# 빅데이터 분석을 통해 본 한국 위키피디아의 지식형성 과정에 관한 연구\*

## A Study on the Knowledge Formation Process of Wikipedia in Korea through Big Data Analysis

이정연 (Jungyeoun Lee)\*\*

전수현 (Suhyeon Jeon)\*\*\*

### 초 록

본 연구는 대표적인 온라인 협업커뮤니티인 한국 위키피디아의 초기 2002년부터 2019년까지의 편집로그 빅데이터를 해체하여 공동협업과정을 시계열적으로 분석하였다. 공개된 오픈데이터의 표준화된 XML 문서편집 기록을 활용해 Python과 R을 이용하여 분석 요소를 추출하여 이를 활용하였다. 연구 분석 결과 한국 위키피디아 편집자의 참여 방법, 데이터 내용의 특징, 문서 생성의 추이 등을 설명할 수 있었다. 소수 편집자들의 적극적 활동과 대다수 편집자들의 느슨한 참여도 밝혀졌으며, 온라인에서도 나타나는 사회 문화적 특징이 한국 위키피디아에서도 나타났다. 집단지성을 지속화시키기 위해서는 새롭고 다양한 외부자원이 필수인데 신규 진입자들이 공동편집 커뮤니티에 정착하기 위한 다각적인 고려가 필요하며, 관리자 그룹의 고착화를 탈피하여 순환구조를 통한 개방성이 필요함을 제안하였다.

### ABSTRACT

This study analyzed the collaborative process in time series by dismantling the edit log big data of Wikipedia Korea, a representative online collaboration community, from early 2002 to 2019. Analysis elements were extracted from the document edit records, formatted in standardized XML, and analyzed using Python and R. The ways of editors' contribution, the characteristics of data contents, and the trend of document creation were explained by the analysis. An active contribution of a small set of editors and a loose participation of the majority were revealed. In addition, sociocultural characteristics that appear in online communities were also found in Wikipedia Korea. A new, diverse set of external resources is necessary to sustain the collective intelligence. An effort to settle new editors into the wikipedia community and an openness through circulation structure to avoid the exclusiveness of the management group are suggested.

키워드: 위키피디아, 위키백과, 빅데이터, 탈중심화, 온라인협업, 협업 플랫폼, 지식권력, 손수제작물  
Wikipedia, Wikimedia, CSCW, OPPC, Peer Production, Korea, Editors, UGC,  
User Generated Content

\* 이 논문은 2019년도 한국연구재단(NRF-2019R1I1A1A01059988)의 지원을 받아 연구되었음.

\*\* 이화여자대학교 이화사회과학원 연구교수(jyonlee@ewha.ac.kr) (제1저자, 교신저자)

\*\*\* 우아한형제들 데이터엔지니어(suhyeonok@gmail.com) (공동저자)

■ 논문접수일자: 2020년 5월 26일 ■ 최초심사일자: 2020년 6월 13일 ■ 게재확정일자: 2020년 6월 22일

■ 정보관리학회지, 37(2), 171-195, 2020. <http://dx.doi.org/10.3743/KOSIM.2020.37.2.171>

※ 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

온라인 네트워크 시대가 도래하여 지식정보 분야도 다양한 영역에서 혁신적인 변화가 출현하고 있는데 학술분야에서는 오픈 사이언스, 과학분야는 e-Science, 정부에서는 공공데이터 개방 등으로 확장되어 가고 있다. 이러한 변화는 단순히 온라인으로 연구와 업무를 수행하는 것을 포함하는 것 뿐 아니라 온라인 플랫폼을 기반으로 지식과 데이터의 공유와 개방을 시도하여 재생산되는 지식을 공동으로 이용하고 개방시키고자 하는 철학을 포함하고 있다.

정보기술의 혁신으로 인한 전통적인 지식 생태계의 변화는 외부에서 발전한 기술이 기존의 지식 생산의 방식을 어떻게 바꾸어가고 있는지에 대한 근원적인 질문이 필요한 시점이다. 지식의 생산과 배포는 시대의 중심에 따라 변화된다. 왕정 시대에는 왕의 권위에 의한 내용 집필과 관련 독자들이 있었다면, 근대 계몽주의 이후에는 합리적 이성과 전문성을 내세운 대학을 중심으로 한 지식인들의 권위가 있었다고 볼 수 있다. 시민사회의 등장과 온라인 시대는 그 권위가 대중에게로 옮겨지고 있으며, 그 단적인 예로 1768년부터 240여년의 백과사전의 명맥을 이어온 '브리태니커'가 인쇄본 출판을 중지하기에 이르렀으며 이를 대체하여 누구나 참여할 수 있는 온라인 협업 콘텐츠와 플랫폼이 그 자리를 잡아가고 있다.

본 연구는 정보 기술을 기반으로 확산하고 있으며, 집단지성 협업의 성공 모델로 평가되고 있는 온라인 지식백과인 위키피디아의 지식

생산과정을 살펴봄으로서 지식의 공동협업 생산, 개방, 공유 활동이 실제로 어떻게 이루어지고 있는지를 데이터를 통해 분석해 보고자 하였다. 동일한 위키 플랫폼을 사용하여 다양한 국가에서 콘텐츠를 생산하는 위키피디아는 언어 영역권 별로 문화적 특징과 온라인 지식 구조에 대한 연구가 활발히 진행 중이다. 반면 지금까지 우리나라에서 이루어진 위키피디아에 관한 연구는 주로 위키피디아의 지식 권력과 집단지성에 대한 담론적인 연구가 주를 이루고 있는데, 위키피디아 문서 내용의 편집기록 데이터를 기반 한 분석 연구는 미비한 상황이다.

따라서 본 연구는 온라인 플랫폼을 기반으로 하고 있는 정보기술과 자발적인 개별 편집자들이 에이전트(agent)로서 활동하고 공동의 상호작용을 통해서 지식을 생산하고 축적되는 위키피디아를 대상으로 공동협업 지식생산 과정을 객관적으로 분석해 보고자 한다. 이에 한국 위키피디아가 시작된 2002년부터 현재까지의 위키피디아 빅데이터 편집기록을 이용하여 지식 형성 과정을 시계열적인 분석을 통해 파악하고 집단지성의 협업 과정이 이루어지고 있는 현황을 이해해 보고자 한다.

## 2. 관련 연구

'위키피디아'<sup>1)</sup>는 모두가 참여하여 지식을 구성하는 온라인 백과사전이다. 최초의 위키시스템은 1995년 워드 커닝엄이 만든 위키위키 웹인데, 현재의 위키피디아는 2001년 1월에 지미

1) 한국위키미디어협회는 한국어 위키피디아를 '위키백과'로 명명하고 있으나 본 연구에서는 한국어로 기재되고 있는 한국 위키피디아로 명칭을 통일하고자 한다.

웨일즈와 래리 생어가 시작하였다. 미디어위키 데이터베이스를 기반으로 정보를 압축하여 저장하고 호출될 때마다 이를 브라우저로 송출하는 시스템으로 구성되어 있으며, 모든 편집기록이 저장되고 공개된다. 한국어 위키피디아는 2002년 10월 11일에 시작되어 오늘날까지 자발적으로 참여하는 편집자들의 공동작업에 의해 지속적으로 데이터 생산이 이루어지고 있다(진주완, 정철, 류철, 2018).

온라인협업 지식생산 플랫폼인 위키피디아에 대한 연구는 다양한 관점에서 수행되고 있다. 첫째, 컴퓨터 협업업무(CSCW: Computer-Supported Cooperative Work)라는 큰 틀에서 온라인 협업 과정에 미치는 영향을 다루고 있는 협업 플랫폼에 관한 연구이다(Balestra, Arazy, Cheshire, & Nov, 2016). 특히 온라인 공동 협업 커뮤니티(OPPC: Online Peer Production Communities)의 연구에서는 커뮤니티의 참여 형태와 조직 이탈의 원인을 밝히고 있다. 대표적으로 위키피디아, 리눅스, 리서치 게이트 등이 그 분야의 공동이익을 공유하는 개인들의 활동 커뮤니티로서 역할을 하고 있는데, 온라인 협업 플랫폼의 특징은 소규모 핵심 기고자에 의해 많은 양의 데이터가 작성되고 있음을 지적하고 있다(Jiang, Mirkovski, Wall, Wagner, & Lowry, 2018). 또한 전통적인 보상체계가 주어지지 않는 일에 대한 사회적 참여 동기에 관해서는 기술활용 사회참여(TMSP: Technology Mediated Social Participation)라는 주제로 연

구가 이루어지고 있다(Arazy, Lifshitz-Assaf, Nov, Daxenberger, Balesra, & Cheshire, 2017; Lieberman & Bar-Ilan, 2019).

위키피디아를 운영하고 있는 비영리단체인 위키미디어 재단은 협업 활동의 지원을 위하여 데이터의 공개와 연구를 진행하고 있다. 지식의 축적 과정을 통계수치로 매월 공개하고 있으며, 언어 커뮤니티별로 비교가 가능할 수 있도록 하고 있다(Wikipedia Statistics, 2018). 더불어 위키피디아 커뮤니티의 건강성 증진을 위해(자유로운 지식 접근과 기여의 방해요소와 다양성) 매년 연구를 진행하고 있다. 2018년 보고서에 따르면 참여그룹(편집자, 위키미디어 기관들, 프로그램 리더, 기술 공헌자)으로 나누어 조사한 결과 신규 커뮤니티(emerging community)의 젠더 다양성은 떨어지고 있음이 나타났으며, 평균 중위연령은 35-44세이고, 주로 글로벌 노스<sup>2)</sup> 지역의 편집자가 81%를 차지하며, 공헌자의 90%가 남성임을 밝혀내었다(Wikimedia Foundation, 2018). 위키피디아는 신규 진입 편집자(new editor)의 확보와 유지를 위하여 중간 규모의(월평균 300-3,000명) 편집자가 있는 커뮤니티를 선정한 결과 16위(773명)인 한국과 18위(615명)인 체코인 총 65명을 대상으로 현지조사(ethnography) 연구를 진행하였다. 신규 편집자의 배경, 관심사, 동기는 다양하였으며 위키피디아에 공헌을 할 수 있는 잠재성을 갖고 있지만 상호간의 의견이 달라 이해가 충돌<sup>3)</sup>할 수 있다고 지적하고 앞으로 이들을 이해하고

2) North-South divide라고 불리우며 사회경제적 그리고 정치적인 구분을 말한다. 일반적으로 글로벌 노스(북반구)는 북미 및 영국, EU, 러시아, 일본, 싱가포르, 한국, 호주 등을 말하며 Global South는 아프리카, 라틴 아메리카, 아시아 개발 국가, 중동 등을 지칭한다.

3) 위키피디아는 내용 작성에 의견이 다른 경우 각 문서와 편집자별 <토론> 공간에서 의견을 나누는 공간이 있다. 본 연구는 이 부분을 다루지 않았다.

지원하는 방안을 모색하는 것이 필요하다고 하였다. 또한 유관기관과의 협력이 필요하다고 보았는데, 한국의 경우 정부와 협력을 시도하고 있다고 하였다(Wikimedia Foundation, 2017).

둘째, 각 언어 영역권 위키피디아의 데이터를 기반으로 문화적 차이점에 관한 맥락 비교 연구 및 개별 언어국가에 대한 편집자들의 이용행태 연구가 있다. 전 세계적으로 309개의 언어커뮤니티가 있는데, 1위 문서 생산 커뮤니티는 영어권 위키피디아이고 2020년 5월 현재 6,076,841건이며, 한국어 커뮤니티는 494,315건으로 24위의 문서 수를 생산하고 있다(Wikipedia, 2020). Karkulahti와 Kangasharju(2012)는 아랍, 핀란드, 스웨덴, 한국을 대상으로 2009~2010년의 데이터를 분석하였다. 아랍권은 주로 목, 금, 토요일에 주로 편집이 이루어지고 있으며, 업무 시간대가 균등하게 나타난 것은 그 외의 국가와는 다른 결과를 보이고 있다. 아랍 언어 생활권 사람들의 주말이 금, 토요일이고, 타임존이 널리 퍼져있는 반면 한국인의 활동 시간대와 요일은 한국 타임존에 맞추어져 있음이 조사되었는데, 이는 대부분 지리적으로 한국에 있는 사람들이 편집하는 것<sup>4)</sup>임을 설명해주고 있다. 이 연구에서는 활동하는 편집자를 편집 수에 의해 카테고리 나누었는데 문서(articles)를 1,000번 이상 변경하는 그룹(heavy users), 51-1,000번 변경하는 그룹(Medium Users), 2-50번 변경하는 그룹(Light), 1번만 변경한 그룹(Random)으로 나누었다. Heavy users는 전체의 0.2%이며, 문서의 60%를 생산하며, Medium Users는 전체의 1% 편집자이며, 13%의 문서

를 생산한다. Light와 random 그룹은 전체의 99%를 차지하는 편집자이며, 29%의 문서를 생산한다. 다시 말하면 전체 2% 편집자들이 70% 이상의 문서를 생산하고 있음을 밝혀내었다. 최근에는 Polat와 Akbiyik(2019)가 영어, 독일어, 일본어, 한국어, 핀란드어 터키어 위키피디아 협업 혁신의 문화적 차이를 연구하였는데, 편집 활동자와 관리자의 수에 따라 사회적 구조계층의 차이점이 무엇인지를 분석하여 문화적으로 이해하고자 하였다. 문화적으로 협업과정이 다른 행동 구조가 나타남이 밝혀졌는데, 국가별 인구대비 위키피디아 활동 인구를 기술적인 통계적 비교에 의하면 인구 1백만명당 영어권, 독일권, 핀란드권, 일어권 그리고 한국어, 터키어 순으로 나타났다. 각 언어 영역별로 자발적인 기준으로 관리자를 선출하고 편집의 중립성을 위하여 가이드라인을 만들고 준수 여부 및 편집 권한의 정지 등의 권한을 갖게 된다. 관리자의 분포가 높을수록 덜 수직적인 구조가 되는데, 가장 수직적인 조직은 일본(0.4%), 터키(0.88%), 영어권(1.0%), 독일권(1.06%) 순서이며, 한국은 (1.4%), 핀란드(2.56%)로 나타났다. 이 연구는 동일 플랫폼을 활용하는 문화권별 차이를 이해하는데 도움을 주고 있으며, 온라인 편집자들의 행동과 협업 모델이 현실세계의 물리적 문화권의 영향을 받고 있음을 알 수 있었다.

셋째, 우리나라에서는 위키피디아 내용의 신뢰성에 관한 논의로 SCI 학술지를 대상으로 위키피디아 인용을 분석한 연구가 있었다(심원식, 변제연, 김민정, 2013). 지금은 이러한 논의

4) 타국에 거주하는 한국인들의 편집인은 희박하다고 보고 있다.

는 거의 찾아볼 수 없을 정도로 위키피디아 인 용이 매우 활발한 상황이다(이강봉, 2017). 황 주성과 최서영(2010)은 위키피디아와 유사한 네이버 지식인과의 비교연구를 수행하였는데, 네이버 지식인 기여자들은 객관적 지식 축적보 다는 필요에 의한 지식 제공 및 만족을 중시하 고, 개별적 집단지성인 반면 위키 기여자들은 지 식의 축적이나 발전 그 자체가 더 높았으며, 협 업적 집단지성임을 설명하였다. 한창진(2009) 은 네이버 지식인과 위키피디아 두 집단의 집 단지성 모델은 구조적으로 다른 방식으로 모델 화 되어 있기 때문에 생산 데이터가 축적이 되 면서 생산 방식 차이에 있어서 발생하는 이용 도가 확연히 차이가 있어 그 모델의 차이점이 나타날 것이라고 하였다. 최근 장순규, 강하영, 윤재영(2018)은 나무위키와 위키피디아의 비 교 연구에서 지식정보의 전달 방법이 다르고 그 역할이 다르다고 보았다. 위키백과는 신뢰 성과 질적인 지식정보를 뉴스나 논문, 미디어 등 외부 자료를 활용하여 명확하고 좁은 범위 의 정보를 전달하려는 반면에 나무위키는 흥미 성을 강조하여 긴 텍스트와 영상 및 이미지를 활용하고, 넓은 범위의 내용을 포괄하려고 한 다고 하였다. 이용자는 정보의 습득 목적과 신 뢰성 및 흥미성에 따라 선택이 달라지며, 위키 백과 보다는 관리자 개입이 덜한 나무위키의 내용과 이미지와 영상의 활용이 높고, 정보의 양도 많음을 밝혀내었다.

사회과학 분야의 연구에서는 위키피디아의 반

(反)전문가주의 특징과 집단지성 한계점 및 새 로운 지식권력의 문제를 제기하는 연구들이 있 다(손준우, 2012; 윤은호, 2012). 정승화(2016) 는 위키피디아 편집에 참여하는 참여자<sup>5)</sup>들의 인종과 성별 계급의 편중성을 비교하며 온라인 에서 출현하는 권력층인 관리자의 권한과 신규 진입자 정착의 어려움 등에 대하여 논의하였다. 또한 위키피디아는 기존 매체의 인용에 근거한 글쓰기를 허락하고 있는 정책이 기존의 제도적 구조를 충실하게 재생산한다는 한계점을 지적 하고 있다. 특히 위키피디아의 권위 구조<sup>6)</sup>와 집단 문화는 여성들을 배제하는 경향이 있고, 기존 지식체계에서 배제된 이들이 더욱 주변화 시킬 위험이 있으며, 대부분은 사소한 편집이 나 반달리즘에 방어하는 활동을 하고 있다고 하였다.

이와 같이 온라인 집단지성은 기존의 지식정 보에 비해 누구나 자발적으로 네트워크를 통해 언제 어디서든지 참여하여 만들 수 있고, 대규 모 협업을 통해 콘텐츠를 형성하고 있다. 국내 에서도 다양한 온라인 협업 플랫폼이 등장하고 있으며 편집 정책과 시스템 구조가 상이하 며 참여자들의 동기에 따라 내용과 목적이 다름을 알 수 있다. 또한 온라인 협업 조직도 새로운 형태의 조직 구조를 만들어가면서 플랫폼에 기 여하는 구성원들에 의해 그 특성이 형성되고 있다. 앞으로도 탈중심화 된 네트워크 공유 협 업 모델이 지속적으로 증가될 것으로 예상되는 이 시점에 그동안 이루어졌던 한국 위키피디아

5) 북미와 유럽을 중심으로 교육받은 20~30대의 젊은 남성이 다수의 편집 참여자라고 하였다.

6) 위키피디아는 일반편집자(editors)들이 있고 관리자(administrators)들이 약 1,600명 정도가 있다(전세계 위키피 디아). 이 관리자들 중 관리자 지명 및 해촉 권한을 갖게 되는 관료(bureaucrat)가 있는데 약 26명으로 구성되어 있다. 또한 조정위원회(arbitration committee)는 최종 의사결정기구 역할을 하고 있다(이항우, 2013).

에 대한 담론적인 연구를 확장하여 데이터 기반의 실증적 연구가 요구되는 상황이다.

### 3. 연구방법

#### 3.1 위키피디아 빅데이터 수집 방법

본 연구에서는 위키피디아에서 공개적으로 제공하고 있는 XML 형태의 대용량의 원본 데이터를 수집하여 가공하고 이를 활용하는 방안을 채택하였으며, 데이터를 다루기 위해 Python 언어를 바탕으로 한 전처리 기반 기술을 활용하였다. 박정원, 양수철, 문혜정(2017)의 연구에서도 위키피디아의 XML 데이터 원본을 Python과 R 언어로 전처리하여 필요한 요소들을 적절한 형태로 가공 활용하여 연구한 바 있다.

#### 3.2 분석 과정 및 방법

위키피디아는 여러 유형의 생산 데이터들을 공개 및 제공하고 있다. 본 연구에서는 이러한 공개 데이터 중 전체 한국어 문서의 생성 및 수정 등의 히스토리가 담긴 XML 형태의 데이터를 활용하였다. 해당 데이터는 한국어 위키피디아 문서가 최초로 생성된 시점인 2002년부터 최근까지의 모든 기록이 담겨있으며, 최신 일자의 데이터가 계속해서 업데이트되고 있다. 본 연구는 2002년부터 2019년 10월까지의 데이터를 사용하였다. 압축본의 용량은 12G이지만 원본의 용량은 300G가 넘는 방대한 사이즈의 데이터이다. 추가로 위키피디아의 넘겨주기

문서 목록을 바탕으로 일반문서를 구분하기 위해 'redirect.sql.gz' 파일을 활용하였다. 해당 파일은 문서 간 넘겨주기 구조를 담은 테이블 생성쿼리가 담겨 있는 SQL 데이터이다.

#### 3.3 분석 데이터 추출 요소

본 연구에서는 데이터를 분석에 활용하기 위한 정보 형태로 가공하는 작업을 우선적으로 수행하였다. 히스토리 데이터는 Python 언어를 사용하여 가공하였으며, '넘겨주기 문서' SQL 데이터는 파일에 포함된 테이블 생성 쿼리문을 활용하여 테이블을 생성한 후 이를 히스토리 데이터와 결합하여 넘겨주기 문서를 제거하기 위한 전처리 로직에 사용하였다.

히스토리 데이터는 <그림 1>의 좌측과 같은 XML 형태이며, 여러 요소가 다양한 태그로 구분되어 있다. 데이터에 담긴 정보는 <그림 1> 같은 구조이며, 이러한 각 요소에 해당하는 정보들을 추출하여 <표 1>의 데이터 테이블을 완성하였다.

전체 데이터를 한 번에 메모리에 적재시키기에 어려움이 있기에 한 라인씩 데이터를 불러오며, 문서명(Title), 문서 이름 공간(NameSpace), 문서 번호(Page-ID), 편집 번호(Revision-ID), 편집 일자 및 시간(Revision-Timestamp), 편집자-로그인(Revision-Contributor-Username / Revision-Contributor-ID), 편집자-비로그인(Revision-Contributor-IP) 항목을 추출하여 정제하였다. 또한 전체 문서 중, 일반 문서만을 분석 대상에 포함시키고자하여 위키피디아의 메타 페이지에서 공개하는 메타 정보를 활용하였다.



〈그림 1〉 위키피디아 XML데이터 원본 일부 및 히스토리 XML 데이터 자료 구조

〈표 1〉 분석 추출 요소

요소	설명
Title	문서 제목
Name Space	문서 이름 공간
Title ID	문서 번호
Revision ID	편집(작성/수정) 내역 번호
Revision Timestamp	편집(작성/수정) 연월일&시간
Revision-Contributor-Username	편집자 이름(로그인을 했을 경우)
Revision-로그인-Contributor-ID	편집자 ID(로그인을 했을 경우)
Revision-비로그인-Contributor-IP	IP(비로그인일 경우)
Text Size	편집 후 문서 전체 텍스트의 byte 수

'https://ko.wikipedia.org/wiki/위키백과:이름공간' 페이지를 통해 각 저장 공간 ID의 정의를 파악하여, 0번 저장 공간 번호를 가진 문서만 분석 대상에 포함시켰다. 추가로 문서가 시간이 지남에 따라 수정 내역의 트랜드가 어떻게 변화되는가를 파악하기 위해 수정 내역 별로 문서 전체 텍스트의 byte 수를 계산하여 파생변수로 생성하였다. 다만 히스토리 데이터에는 편집 완료 후 문서의 전체 텍스트가 남는 형태인데다가, 문서를 표현하는 모든 html 코드 등까지 포함되어 있어 이 점을 감안하고 데이터를 분석하였다.

이러한 과정을 거쳐 정제 및 파생변수 생성이 완료된 데이터는 R 언어를 통해 시각화 및 심화 분석을 수행할 수 있도록 데이터프레임 형태로 구성하였으며, 다음과 같은 형태이다. 〈그림 2〉를 살펴보면 '조선 세종'이라는 제목을 가진 위키피디아 문서는 위키피디아 내부 데이터베이스에서 2789번이라는 ID로 분류됨을 알 수 있다. 2004년에는 생성되었으며 그 이후 여러 사용자들에 의해서 편집되고 있는 기록이 담겨있다. 이 내용은 위키피디아 문서 메인의 우측 상단에 있는 역사보기 페이지를 통해서도 확인이 가능하다.



였는지, 새롭게 생성하였는지를 추적할 수 있다. 따라서 이 연구에서는 지난 18년간 한국 위키피디아 문서 약 47만 건, 편집자 약 68만여 명<sup>10)</sup>의 활동이 어떻게 이루어지고 있는지를 문서와 편집자<sup>11)</sup>(저자)를 중심으로 시계열적으로 분석해 보고자 하였다.

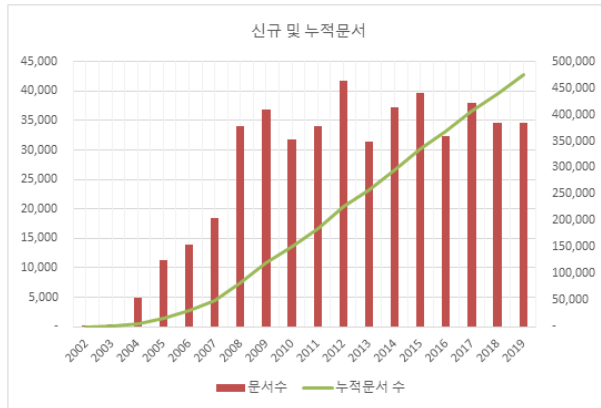
4.1.1 한국 위키피디아 문서 편집 현황 분석  
문서는 편집자들에 의해서 비동시적인 편집이 일어나면서 생성되고 있는데, 편집자가 관심 있는 주제에 관한 신규 문서 생성 및 다른 편집자가 작성해 놓은 기존 문서의 편집, 수정, 추가, 삭제, 재배열 등의 편집 과정을 통해 이루어진다. 한국 위키피디아 문서의 협업을 계량

적으로 분석하면 다음과 같다. 지난 18년간 축적된 문서는 총 474,947건이며, 연도별 신규 생성 문서 수는 다음의 <표 2>와 같다. 2002년 16건의 문서로 시작하여 2009년까지는 매년 지속적인 신규 문서 생산이 증가세를 보이다가 2009년부터는 등락을 보이면서도 안정적인 문서 생성이 이루어지고 있다(<그림 3> 참조). 평균적<sup>12)</sup>으로 매일 평균 78건, 매월 2,328건이, 매해 27,938건의 문서가 생성되고 있다.

문서의 연도별 추이를 살펴보면 2008년까지 공동편집자 수와 편집 횟수가 꾸준히 증가하고 있으며, 문서의 데이터의 양도 급격히 증가하고 있다. 그러나 2009년부터 편집 횟수와 공동편집자의 수가 서서히 감소하고 있음을 보여주

<표 2> 신규문서의 연도별 증가율

연도	신규 문서 수	전년대비 증가 수	증가율(%)	누적문서 수
2002	16			16
2003	454	438	▲2,738%	470
2004	4,823	4,369	▲962%	5,293
2005	11,238	6,415	▲133%	16,531
2006	13,989	2,751	▲24%	30,520
2007	18,354	4,365	▲31%	48,874
2008	33,952	15,598	▲85%	82,826
2009	36,769	2,817	▲8%	119,595
2010	31,745	-5,024	▼14%	151,340
2011	33,990	2,245	▲7%	185,330
2012	41,650	7,660	▲23%	226,980
2013	31,448	-10,202	▼24%	258,428
2014	37,187	5,739	▲18%	295,615
2015	39,697	2,510	▲7%	335,312
2016	32,417	-7,280	▼18%	367,729
2017	37,903	5,486	▲17%	405,632
2018	34,648	-3,255	▼9%	440,280
2019	34,667	19	▲0%	474,947
총 합계	474,947			



<그림 3> 연도별 신규생성 문서 추이

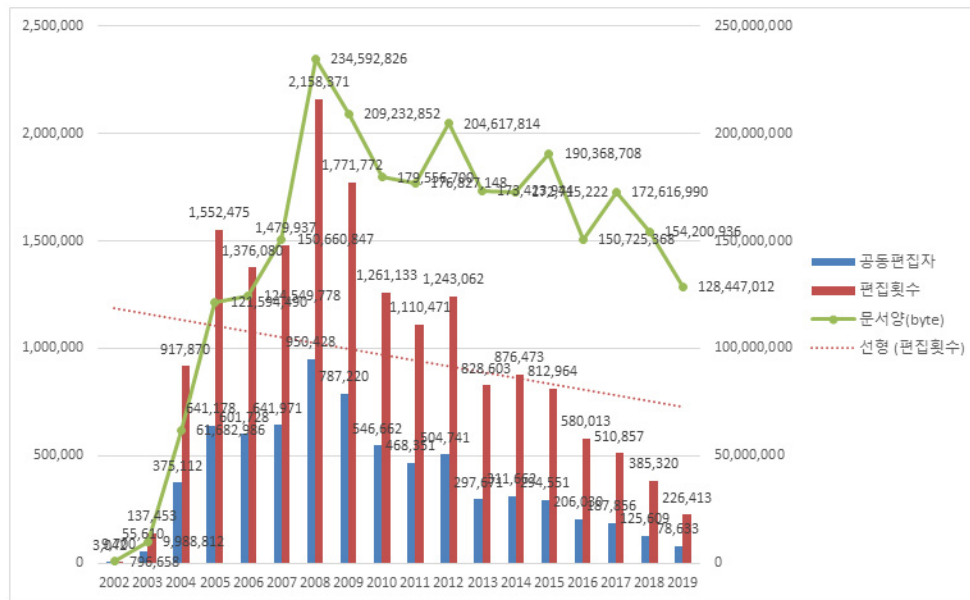
- 10) 통상 한국 위키피디아 통계에서 다루는 문서는 전체문서(토론, 넘겨주기 포함)이다. 본 연구는 편집과정을 분석하는데 목적을 두었기 때문에 넘겨주기를 받은 '기본 문서'에 작업을 한 로그인(등록자) 및 비로그인(미등록 IP로 접속) 편집자를 대상으로 분석하였다.
- 11) 위키피디아는 다수의 편집자들에 의해 공동으로 문서를 작성해 간다. 문서의 공동 저자들을 의미하는데, 위키피디아 공식 용어가 편집자(editors)이므로 본 연구도 '편집자'로 통일하고자 한다.
- 12) 연평균, 월평균, 일평균 계산은 2002년부터 2018년까지의 17년으로 처리하였다. 2019년은 10월까지의 데이터로서 2개월이 제외되기 때문이다.

고 있다. 반면 문서 텍스트의 양은 편집횟수와 공동편집자 수의 감소에 비해 보다 완만하게 적정선을 유지하고 있음을 보여주고 있다(〈그림 4〉 참조).

개별 문서에 참여한 편집자들의 수와 편집 횟수 그리고 데이터의 양을 문서가 만들어지는 연월일 기준으로 분석한 결과, 하루에 공동편집자와 편집 횟수 및 문서의 양은 다음의 〈표 3〉과

같이 일일 평균 1,144명이 2,817번의 편집을 하고 있으며, 총 427,549명 byte 문서데이터를 작업 하는 것으로 나타났다.

문서마다 편집에 참여하는 공동 편집인 수와 편집 횟수는 문서마다 상이하다. 평균적으로 문서 1개당 공동편집인 수와 편집 횟수를 측정해 보고자 하였다(〈표 4〉 참조). 평균 값은 공동편집자가 15명이 36회를 수행하여 문서 1개를 만



〈그림 4〉 연도별 문서의 공동편집자, 편집횟수, 문서양의 추이

〈표 3〉 일평균 문서 편집 현황

일 평균	공동편집자	편집 횟수	문서의 양(byte)
	1,144	2,817	427,549

〈표 4〉 평균 편집 횟수 및 편집자 수, 문서 양

문서 1개	공동편집자 수	편집 횟수	문서 양	총 문서
평균	15명	36회	5,509byte	
중위 값	7명	14회	2,580byte	

들어 간다. 반면에 중위 값은 공동편집자 7명이 편집 횟수를 14회 하는 것으로 나타난다. 이러한 차이가 나는 이유는 데이터에 극단 값이 있기 때문인데, 소수의 편집자가 상당히 많은 수의 편집을 하기 때문에 평균과 중위 값의 큰 차이가 나타나는 것이다. 문서의 양도 평균 값과 중위 값이 매우 다르게 나타나는데 문서에 따라 데이터 양의 차이가 많기 때문이다.

반면 1회 편집(한 명의 편집자)으로 작성된 문서의 건수는 21,737건(4.6%)에 달하는데 편집자 1인이 최초로 문서를 만들고 그 이후에 편집이 이어서 이루어지지 않은 경우를 말한다. 1~5회 편집으로 문서가 작성된 경우는 61,041건(62%)이다.

#### 4.1.2 한국 위키피디아 최다 참여 문서

문서에 참여하는 편집인 수와 편집 횟수가 많은 문서를 분석해 보았다. 다수의 편집자들이 여러 횟수를 거쳐 편집하고 있는 문서는 참여자들의 관심도가 높다는 것을 말한다. 18년간에 걸쳐 가장 많은 편집 횟수와 공동 편집자 수를 보이는 문서의 순위를 살펴보면 다음과 같다.

〈표 5〉와 같이 편집자 수의 참여가 높은 문

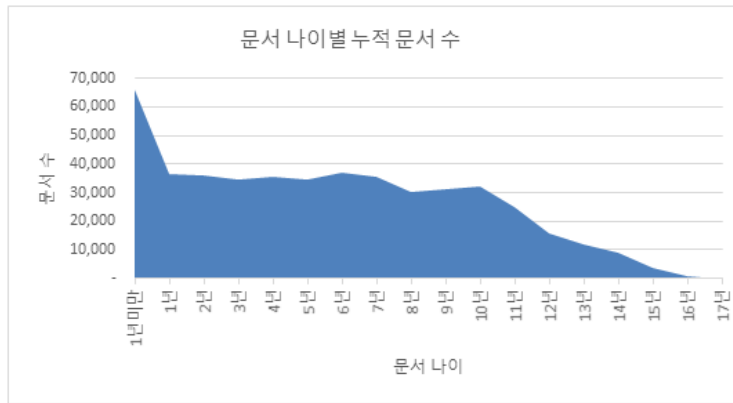
서는 주로 정치, 사회관련 주제가 주를 이루며, 초창기 연도부터 지금까지 꾸준히 업데이트가 이루어지고 있다. 반면에 편집 횟수가 높은 문서는 연예 및 스포츠 등 방송, 엔터테인먼트 분야의 주제가 대부분이며, 편집 시작 연도는 해당 주제(프로그램명)가 출현할 때 시작되어 최근까지 지속되고 있음을 알 수 있다. 주제 내용을 살펴보면, 학술적인 내용보다는 일반적이고 대중적인 주제 내용이 주를 이루고 있음을 알 수 있다.

#### 4.1.3 위키피디아 문서 나이

문서 47만 건을 개별 문서별로 편집이 시작된 연월일부터 최근까지 수정, 업데이트, 삭제된 편집연월일까지의 기간을 '문서 나이'라고 명명하여 분석해 본 결과 문서의 수명이 1년 미만인 경우가 전체의 22%이며, 1개월 미만도 66,060건으로 전체의 13.9%가 된다. 한편 10년 이상 지속적으로 편집이 이루어지는 문서의 비율도 20%가 된다. 전체 문서의 평균 나이인 편집 기간은 5.6년인 것으로 보아 편집자들이 관심 있는 주제에 대한 꾸준한 편집 수정을 하고 있음을 알 수 있다(〈그림 5〉 참조).

〈표 5〉 한국위키피디아 편집자 최다 참여 및 편집 횟수 문서 순위(2002-2019)

순위	TITLE	편집자 수	편집시작연도	최근 편집연도	순위	TITLE	편집횟수	편집시작연도	최근편집연도
1	대한민국	1,870	2003	2019	1	개그콘서트	15,049	2008	2019
2	개그콘서트	1,674	2008	2019	2	박정희	7,335	2004	2019
3	조선민주주의인민공화국	1,479	2003	2019	3	미스터리음악쇼 복면가왕	6,950	2015	2019
4	박정희	1,419	2004	2019	4	대한민국	6,934	2003	2019
5	소녀시대	1,411	2007	2019	5	대한민국 축구 국가대표팀	6,592	2005	2019
6	1박2일	1,352	2008	2019	6	FC 서울	6,571	2005	2019
7	뮤직뱅크	1,236	2008	2019	7	1박 2일	6,326	2008	2019
8	노무현	1,225	2003	2019	8	소녀시대	6,133	2007	2019
9	대한민국축구국가대표팀	1,184	2005	2019	9	뮤직뱅크	6,017	2008	2019
10	자유한국당	1,128	2004	2019	10	도전! 골든벨의 에피소드 목록	5,354	2013	2019



〈그림 5〉 문서 나이별 누적 문서 수

〈표 6〉 문서의 나이

문서나이	문서 수	백분율(%)
1년미만	66,060	13.91
1년	36,484	7.68
2년	36,131	7.61
3년	34,644	7.29
4년	35,385	7.45
5년	34,647	7.29
6년	36,828	7.75
7년	35,528	7.48
8년	30,072	6.33
9년	31,038	6.54
10년	32,092	6.76
11년	25,037	5.27
12년	15,933	3.35
13년	11,769	2.48
14년	8,762	1.84
15년	3,673	0.77
16년	829	0.17
17년	35	0.01
총합계	474,947	100

#### 4.2 편집자 분석을 통해 본 위키피디아의 특성

위키피디아는 두 가지 형태의 편집자를 허용한다. 본인의 아이디를 생성하고 로그인하여 편집하는 '로그인 편집자'와 로그인 하지 않고 편집하는 '비로그인 편집자'가 있다. 본 연구는 기본 문서 편집을 1회 이상 진행 한 로그인 편집자와 비로그인 편집자<sup>13)</sup> 모두를 분석하였다. 이 중에서 '봇'이라는 아이디의 편집 활동은 제외하였는데 그 이유는 분류 변경, 목록 갱신과 같은 간단한 작업을 자동 처리할 때 활동하는 것이기 때문이다. 따라서 본 연구에서는 '107개 봇'<sup>14)</sup>의 활동을 제외한 비로그인 편집자 568,941명, 로그인 편집자 115,023명을 포함한 전체 683,964명의 편집자들에 대하여 분석하였다.

##### 4.2.1 연도별 편집자의 편집활동 추이

위키피디아에 로그인을 하여 편집자로 활동한다는 의미는 본인의 아이덴티티를 가지고 문서편집 내역과 토론 등 활동을 지속적으로 하겠다는 것이며, 로그인 하지 않고 편집하는 활동은 본인의 활동을 기록으로 남기지 않고 편집에 참여하는 것을 의미한다. 편집자의 편집 문서 수와 편집 활동에 대한 분석을 하기 위하여 로그인 그룹 및 비로그인 그룹에 대한 편집 현황의 연도별 추이를 보면 다음과 같다(〈표 7〉 참조). 2002년에 13명의 편집자를 시작으로 위키피디아가 시작되었고 연도별로 꾸준히 증가세를 보이고 있는데 평균적으로 로그인 편집자에 비하여 비로그인 편집자의 수가 약 4배가 많다.

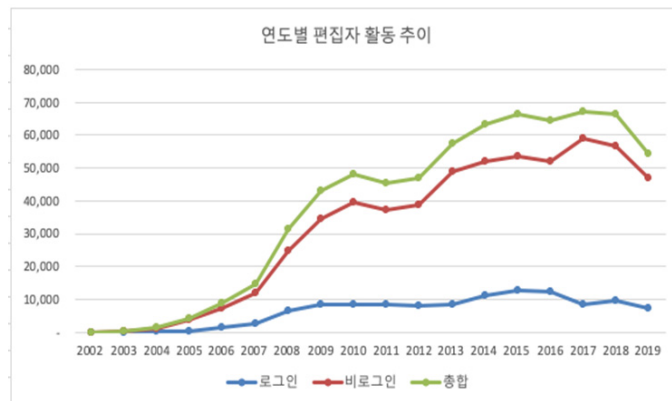
〈그림 6〉과 같이 연도별 편집자 활동 추이를 보면 초기에는 로그인과 비로그인 활동이 비슷

13) 비로그인 편집자는 IP 주소가 ID가 되어 로그가 기록되어 있으므로 이를 분석하였다.

14) 본 연구의 '봇'은 한국 위키피디아 봇 목록 121개 중 연구 대상 데이터와 매칭된 '봇'은 107개로 집계 되었다.

〈표 7〉 로그인 및 비로그인 그룹 편집 횟수

연도	로그인 편집자	편집문서 수	편집 횟수	연도	비로그인 편집자	편집문서 수	편집 횟수	연도	Total	편집문서 수	편집 횟수
	115,023	3,369,561	9,489,278		568,941	2,087,635	4,614,737		683,964	5,457,196	14,104,015
2002	2	1,803	2,850	2002	11	34	72	2002	13	1,837	2,922
2003	34	28,513	77,887	2003	263	948	1,376	2003	297	29,461	79,263
2004	191	45,448	124,911	2004	1,127	4,343	6,905	2004	1,318	49,791	131,816
2005	436	157,859	356,972	2005	3,763	14,352	25,696	2005	4,199	172,211	382,668
2006	1,370	179,238	427,109	2006	7,498	21,852	37,184	2006	8,868	201,090	464,293
2007	2,753	429,619	1,031,880	2007	11,899	40,309	69,428	2007	14,652	469,928	1,101,308
2008	6,425	462,236	1,178,859	2008	24,900	70,205	130,867	2008	31,325	532,441	1,309,726
2009	8,574	484,245	1,353,985	2009	34,541	135,980	284,922	2009	43,115	620,225	1,638,907
2010	8,426	367,492	1,000,986	2010	39,737	132,237	263,720	2010	48,163	499,729	1,264,706
2011	8,315	204,302	591,630	2011	37,098	135,492	292,670	2011	45,413	339,794	884,300
2012	8,112	235,894	716,680	2012	38,954	132,758	275,672	2012	47,066	368,652	992,352
2013	8,400	244,244	692,075	2013	49,092	158,787	381,418	2013	57,492	403,031	1,073,493
2014	11,324	116,659	621,703	2014	51,939	192,556	446,668	2014	63,263	309,215	1,068,371
2015	12,642	157,521	529,037	2015	53,639	247,026	594,338	2015	66,281	404,547	1,123,375
2016	12,520	101,075	339,414	2016	52,062	231,664	563,072	2016	64,582	332,739	902,486
2017	8,516	65,188	221,143	2017	58,846	230,453	512,185	2017	67,362	295,641	733,328
2018	9,717	55,945	141,862	2018	56,552	206,765	451,195	2018	66,269	262,710	593,057
2019	7,266	32,280	80,295	2019	47,020	131,874	277,349	2019	54,286	164,154	357,644

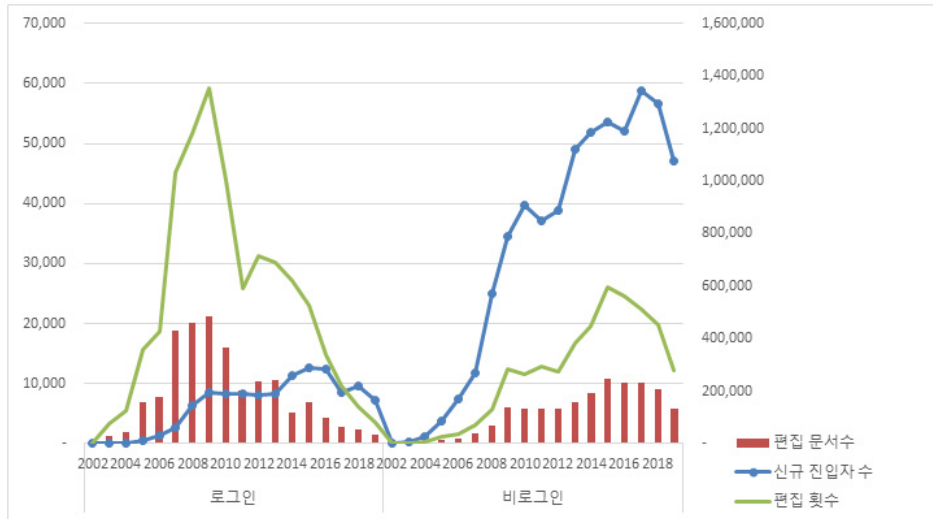


〈그림 6〉 연도별 편집자 활동 추이

하게 이루어지다가 비로그인 편집자들은 지속적으로 증가하는 반면에 로그인 편집인들은 2008년까지 활동이 증가하다가 등락을 보이면서 현재는 10년 전 활동량을 보이고 있다. 반면에 비로그인의 편집 활동은 최근까지 지속적인 성장세를 보이고 있다.<sup>15)</sup>

편집 문서의 수와 편집 횟수를 기준으로 〈그림 7〉과 같이 로그인 및 비로그인 편집인들의 활동을 분석해 보았다. 2009년까지 로그인 그룹의 편집 문서 수와 편집 횟수는 월등히 증가하였으나 2010년 이후로 편집 횟수는 급격히 감소되고 있으며 편집 문서 수도 감소되고 있

15) 2019년 성장곡선이 다소 떨어지게 보이는 것은 본 연구 대상의 데이터가 2019년 10월까지이므로 남은 두 달의 데이터가 반영되지 않았기 때문이다.



〈그림 7〉 로그인 및 비로그인 그룹별 편집 문서 수 및 편집 횟수

다. 반면 로그인 그룹에 비하여 비로그인 편집자의 신규 진입자 수는 최근까지 꾸준히 증가하고 있으며, 편집 횟수와 편집 문서는 2016년도까지 증가하다가 점차 안정 하향세를 접어들고 있는 상황을 볼 수 있다.

#### 4.2.2 편집자의 참여 기간

지난 18년 동안 편집자가 위키미디어에 참여한 기간을 살펴보기 위하여 편집자의 최초 편집일자에서 최종 편집일자까지의 편집 로그를 분석 해 본 결과 〈표 8〉과 같이 1개월 미만 참여하는 편집자가 84.97%가 되었고, 1년 미만이 전체의 93.42%에 이르며 평균 2.8개월 정도 참여하는 것으로 나타났다. 로그인 편집자들의 경우에는 평균 5.6개월 참여하여 비로그인 편집자들에 비해 머무는 시간이 높다. 하지만 대다수 편집인들은 1개월 미만 참여하여 편집하는 것으로 보아 편집자들이 지속해서 위키미디어에 흥미를 가지면서 편집하고 있지 않다고

사료된다. 반면 10년 이상 장기적으로 참여하는 편집자들도 0.1%를 차지하고 있다.

〈표 8〉 편집자의 참여 기간

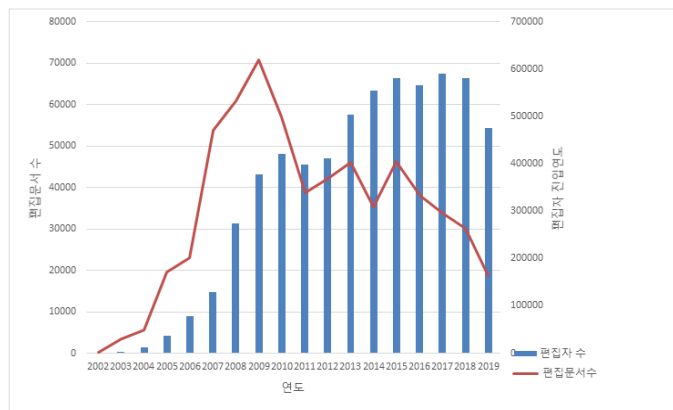
참여 기간	편집자수	백분율
1개월미만	581,167	84.97%
1개월~6개월	38,711	5.66%
6개월~1년	19,100	2.79%
1년 이상-2년 미만	18,646	2.73%
2년 이상-3년 미만	10,032	1.47%
3년 이상-4년 미만	5,960	0.87%
4년 이상-5년 미만	4,118	0.60%
5년 이상-10년 미만	5,538	0.81%
10년 이상-15년 미만	681	0.10%
15년 이상	11	0.002%
총	683,964	100.00%

#### 4.2.3 편집자의 참여 시기

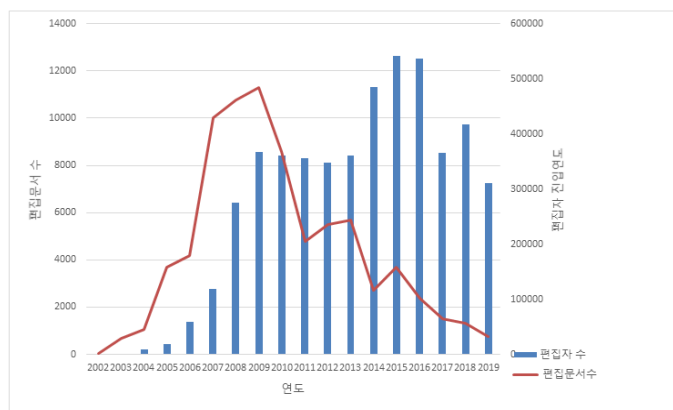
한국의 위키미디어는 2002년 10월에 문서가 생성되었는데, 편집자들이 위키미디어에 신규 참여한 시기를 추이해 보면 다음과 같다. 〈그림 8〉과 같이 로그인 비로그인을 합친 편집자들의

신규 참여는 2002년도에 13명(로그인 편집자 2인)으로 시작하여 2010년까지 꾸준히 증가세를 보이다가 안정선을 유지하였다. 2014년 이후에 다시 증가세를 보이고 있으며 평균 33,998명이 신규로 참여하는 것으로 나타났다. 그러나 신규 진입자가 일정 수준으로 매년 유지됨에도 불구하고 편집 문서의 수는 2010년 이후에 감소하는 것으로 나타난다. 특히 <그림 9>에 나타나는 바와 같이 로그인 편집자의 신규 진입연도에 따른 편집문서 수를 보면 최근 10년

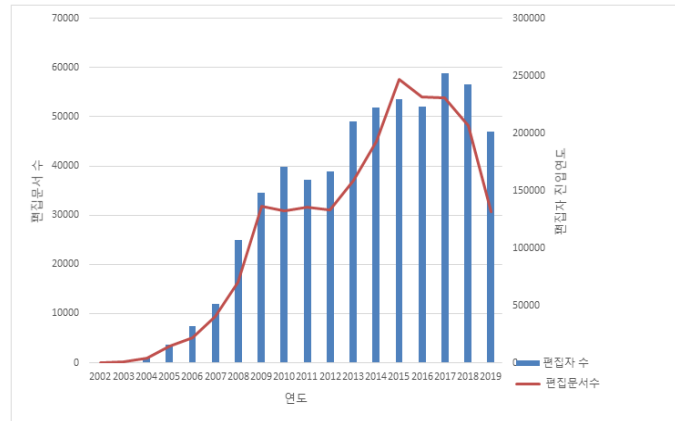
내에 연도별로 진입하는 신규 진입자들의 수와 관계 없이 편집문서 수는 급격히 감소하는 것으로 나타난다. 반면 비로그인 상태에서 편집하는 편집자들의 진입 연도와 편집 문서의 수는 비례하고 있다(<그림 10> 참조). 이와 같은 현상은 비로그인 편집자들의 진입이 비교적 안정되게 증가하고 있으며 또한 문서 편집도 꾸준히 이루어지고 있음을 알 수 있는데, 이들이 최근의 한국 위키피디아 문서 편집에 기여하는 것으로 분석된다.



<그림 8> 전체 편집자 진입연도 및 편집 문서 수



<그림 9> 로그인 편집자 진입연도 및 편집문서 수



〈그림 10〉 비로그인 편집자 진입연도와 편집문서 수

4.2.4 편집자의 문서 편집 현황 분석

위키피디아의 편집인들이 문서를 몇 회 편집을 하는지 분석하여 활동량을 추정하였다. 편집 횟수가 1회에 그친 경우가 전체 편집인들의 49%에 해당하였다. 또한 1~10회 내외로 편집한 편집인들이 전체의 90%를 차지하고 있었다(〈표 9〉 참조).

〈표 10〉은 편집 횟수 구간별 편집자 수를 분석을 보여주고 있는데, 100회 이상의 편집 횟수를 기록하는 편집자들은 1.3%에 해당한다.

〈표 10〉에 나타나듯이 편집 횟수가 10만 건 이상을 편집하는 소수의 편집인들이 있는데 이들은 모두 로그인 편집인들이며, 이들 8인이 1,495,118건의 편집 횟수와 51만 건<sup>16)</sup>의 문서

〈표 9〉 편집인의 편집 횟수(1-10회)

편집 횟수	편집인 빈도 수	백분율(%)	누적 백분율(%)
1	336,902	49%	49%
2	119,894	18%	67%
3	54,410	8%	75%
4	32,355	5%	80%
5	21,871	3%	83%
6	15,731	2%	85%
7	11,711	2%	87%
8	9,229	1%	88%
9	7,651	1%	89%
10	6,119	1%	90%
합계	615,873		90%

16) 8인이 대다수의 문서를 편집하고 있음을 보여주고 있다. 개인별 문서 편집은 데이터에서 중복처리 하여 분석하였지만 8인의 편집 문서는 중복될 수 있다.

〈표 10〉 편집 횟수 구간별 편집자 수

편집 횟수	편집자 수	백분율(%)	누적 백분율(%)
1-10회	615,873	90%	90%
11-100회	58,679	8.6%	98.6%
101-1,000회	7,994	1.2%	99.8%
1,001-10,000회	1,249	0.2%	99.9%
10,001-100,000회	161	0.02%	99.99%
100,000회 이상	8	0.001%	100%
총		683,964	

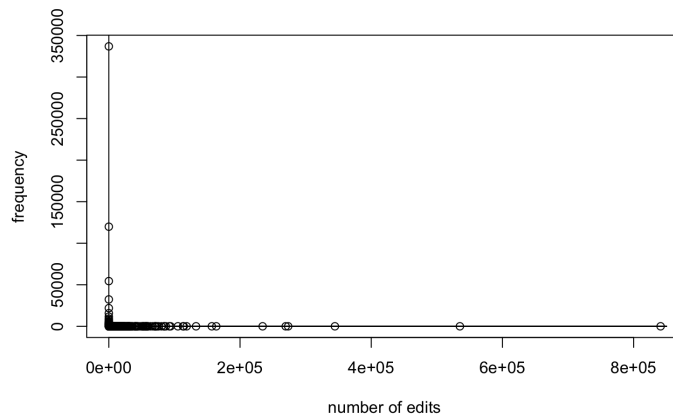
를 편집하고 있었다. 이중 최다 편집 횟수를 수행하는 1인은 34만의 편집 횟수와 16만 9천 건의 문서를 편집하고 있는데, 이는 전체 36%의 문서를 다루고 있음을 알 수 있다.

이와 같이 대부분의 편집자들은 매우 소극적인 편집을 하고 있음에 반하여 극소수의 편집자들이 왕성한 편집 횟수를 보이고 있음을 알 수 있다. 이러한 현상은 편집자의 편집 횟수의 평균 값에 영향을 미치고 있다. 전체 편집자의 평균 편집 횟수는 21회이지만, 중위 값은 '2'로 나타났다. 특히 로그인 그룹 편집자들의 평균 편집 값은 82, 비로그인 그룹의 평균 값은 8로 산출된 것은 소수의 대량 편집자들이 대부분

로그인 편집 그룹이기 때문이다.

편집인들의 편집 횟수를 다음의 〈그림 11〉과 같이 편집자들과 편집 횟수의 관계를 도표로 나타내면 정규분포가 아닌 다음과 같은 형태로 나타난다. 즉 다수의 편집자들은 적은 횟수의 편집을 하고 극소수의 편집자들이 대량의 편집을 하는 것으로 나타난다.

편집 문서 수를 살펴보면 1개의 문서만 편집한 편집자들은 전체의 66%인 484,095명이며, 문서 2개만 편집한 편집자들은 12%인 80,831명이었다. 즉 한 두 개의 문서만 편집한 편집자들이 전체의 78%에 해당함을 알 수 있다. 반면에 46명은 문서를 1만개 이상에서 16만개까지



〈그림 11〉 편집자들과 편집 횟수의 분포도

편집한 이력이 있는 것으로 나타났다. 이들 모두 로그인 편집자로 활동하고 있었다. 이와 같이 문서의 편집 횟수나 편집하는 문서의 수나 대부분 편집자들은 1~2회 정도 편집에 그치는데 반하여 소수의 대량 편집자들이 문서의 대부분을 다루고 편집하고 있음을 알 수 있었다.

지금까지 전체 편집자들의 문서 편집 현황 분석을 통해 한국 위키피디아의 전반적인 추이를 살펴보았다. 한편 현재 활동하고 있는 편집자(활동 편집자, active editor)를 파악하고자 하였다. 이를 위해서는 통상적으로 3개월 동안의 실제 편집 활동 통계를 기준으로 측정한다. 글로벌 위키피디아 범용 통계에는 평균 매월 5회 이상 편집하고 있는 편집자 수를 측정하는데, 한국 위키피디아 통계에는 월 평균 1회 이상 편집자로 구분한다. 본 연구에서는 최근 활동 내역인 2019년 8월, 9월 10월의 월 평균 1회 이상, 5회 이상, 100회 이상의 편집자를 추출하여 평균 활동 편집 수를 파악하였다. <표 11>의

통계에 나타나듯이 활동 편집자는 한 달에 5,476여명 정도가 되며, 자주 활동하는 편집인은 약 828명, 매우 열심히 활동하는 편집자는 평균 월 35명이라고 볼 수 있다. 최근 활동하는 활동 편집자들도 로그인 그룹 보다는 비로그인 그룹이 더 활발함을 알 수 있었다.

#### 4.2.5 위키피디아 편집 관리자 활동 현황

위키피디아 관리자(Wikipedia administrator)는 위키피디아 사용자 가운데 투표를 거쳐 관리자(sysops)로 지명되는데, 위키피디아 정책에 적합한 신뢰성과 정확성을 모니터링하고 관리하는 주요 업무를 수행한다. 따라서 관리자들은 다른 편집자들에 비해 기술적인 특권을 추가적으로 부여받고 있다. 부적절한 문서들을 삭제하고 문서들을 보호하며 위키피디아의 정책에 어긋나는 계정을 차단하는 역할을 한다.

한국 위키피디아 관리자<sup>7)</sup>는 총 22인으로 관

<표 11> 현재 활동 중인 편집자 (활동편집자) 통계

편집연월	1회 이상	5회 이상	100회 이상
2019년 10월	6,329명	907명	22명
로그인	767	226	5
비로그인	5,562	681	17
2019년 9월	5,072명	769명	47명
로그인	668	179	11
비로그인	4,404	590	36
2019년 8월	5,027명	809명	38명
로그인	709	202	9
비로그인	4318	607	29
<b>평균</b>	<b>5,476명</b>	<b>828명</b>	<b>35명</b>
평균 로그인	714명	202명	8명
평균 비로그인	4,761명	626명	27명

17) 위키피디아 관리자는 편집자(editors)들의 선거를 통해 선출된다. 관리자는 문서의 삭제나 보호, 잘못된 행동을 하는 사용자에 대한 제재 등 몇 가지 중요한 조치를 결정할 수 있다(진주완, 정철, 류철, 2018).

리자, 사무관, 인터페이스 관리자, 검사관, 기록 보호자의 역할로 활동하고 있으며, 한국 위키피디아의 정책 결정과 편집자의 제명이나 분쟁 조정, 수정 되돌리기 등의 편집에 있어서 중요한 역할을 수행하는 일을 하고, 일반 편집자들 중에 선출되어 활동한다.

본 연구에서 추출한 편집자들의 18년간 데이터와 한국위키피디아 관리자 목록<sup>18)</sup>을 대조하여 관리자별 편집 시작연도, 최근 편집연도, 편집 횟수, 편집 문서, 편집 기간 및 관리자 유형별로 <표 12>와 같이 정리하였다. 한국의 관리자 대부분이 위키피디아 초창기인 2003년부터 2010년 동안 편집 활동을 시작한 편집인들로 구성되어 있으며, 전원 모두 최근까지 편집 활

동을 지속적으로 하는 것으로 나타났다. 대부분이 10년 이상의 편집을 하였으며, 2016년에 편집을 시작하여 관리자가 된 회원은 1인 인 것으로 나타났다. 이들의 평균적인 편집 횟수는 35,518건 이며, 편집 문서 수는 16,731건 이고, 평균 편집 기간은 11년이다. 위에서 살펴본 바와 같이 가장 많은 편집 횟수로 문서를 편집하고 있는 1인이 관리자 중 한명으로 나타났다.

이와 같이 한국 위키피디아 관리자는 주로 초기 멤버들로 구성되어져 있음을 알 수 있다. 이러한 조직 구성은 기존 멤버들에 의한 의사 결정으로 편중될 가능성이 있다. 앞서 살펴 본 Polat와 Akbiyik(2019)의 연구에 설명된 바와 같이 편집 활동은 관리자의 수에 따라 사회적

<표 12> 한국 위키피디아 관리자 유형 및 세부 활동

관리자	시작연도	최근 편집연도	편집 횟수	편집문서	편집 기간	관리자 유형				
						관리자	사무관	인터페이스 관리자	검사관	기록보호자
1	2003	2019	13,133	2,987	16.2	관리자				
2	2003	2019	42,233	20,732	16.2	관리자	사무관			
3	2004	2019	1,837	1,110	15.4	관리자				
4	2004	2019	79,643	22,222	15.3	관리자				
5	2005	2019	91,497	52,808	14.5	관리자	사무관	인터페이스 관리자		
6	2005	2019	8,336	5,094	13.4	관리자	사무관			
7	2006	2019	5,968	3,496	13.2	관리자			검사관	기록보호자
8	2006	2019	17,059	9,062	13.1	관리자		인터페이스 관리자	검사관	
9	2006	2019	6,515	2,758	12.8	관리자				
10	2006	2019	30,606	10,624	10.5	관리자	사무관			
11	2007	2019	10,435	5,734	10.5	관리자				
12	2007	2019	6,318	3,133	12	관리자				
13	2007	2019	342,047	168,299	12	관리자		인터페이스 관리자		
14	2007	2019	14,731	3,336	11.9	관리자				
15	2008	2019	3,216	2,158	9.9	관리자		인터페이스 관리자		기록보호자
16	2009	2019	16,270	5,488	10.5	관리자				
17	2009	2019	7,105	2,972	10.3	관리자				
18	2009	2019	6,320	794	10.1	관리자			검사관	
19	2010	2019	5,625	1,530	9.8	관리자				
20	2010	2019	9,837	3,192	9.8	관리자	사무관			기록보호자
21	2010	2019	59,151	39,637	9.5	관리자				
22	2016	2019	3,509	925	3.1	관리자				

18) ko.wikipedia.org/wiki/위키백과:관리자\_목록

구조 계층이 나타난다. 한국은 관리자 비율이 높아 다른 나라에 비하여 비교적 수직적이지 않다고 밝혀졌지만, 본 연구결과 한국사회의 문화인 연공서열(가입 우선순위)의 위계질서가 한국 위키피디아 관리자 조직에서 나타남을 알 수 있다. 이와 같은 연구 결과는 선행연구에서 언급한 바와 같이 새로운 권위 유형의 재형성과 지식권력이 생겨나는 것을 증명해 주는 결과이며, 한국적인 오프라인 문화가 온라인 협업 문화에도 영향을 미치는 것으로 파악된다.

온라인 플랫폼도 물리적 세계와 마찬가지로 건강한 협업 토론과 공동 편집을 통해 지식 축적의 투명성을 유지하기 위해서는 건강한 조직 문화가 기반이 되어야 한다. 관리자 그룹의 장기화와 신규 진입자들이 관리자가 되기 어려운 환경은 기존 세력의 토착화에 따른 수직적인 문화를 가져올 수 있다. 또한 새로운 진입자들이 기존 플랫폼 문화에 적응하고 이를 수용하는데 어려울 수 있는 단점을 초래할 수 있다.

## 5. 결론

위키피디아는 누구에게나 열려있는 집단지성의 장이며 끊임없이 생산, 축적되고 있는 지식들이 무료로 공개되어 오늘날까지 이어지고 있다. 본 연구는 한국 위키피디아 온라인 협업 지식시스템의 시작 초기부터 현재까지 어떻게 진행되고 있는지를 살펴보기 위하여 지난 18년간의 편집 내역을 빅데이터를 분석해 보았다. 연구결과를 요약하면 다음과 같다.

첫째, 위키피디아는 모든 편집과정 데이터들이 표준화된 유형별 네임스페이스와 XML 형

태로 구축되어 있어, 문서와 편집 및 토론 등 집단지성이 이루어지는 과정이 객관적이고 투명하게 기록되고 있었다. 더불어 모든 데이터는 지식의 축적 결과 뿐 아니라 초기부터 현재까지의 협업 과정에 대한 공개로 출처의 투명성을 보여주고 있다. 이는 오픈 데이터의 특성인 개방성, 공개성, 투명성을 함의하고 있을 뿐 아니라 협업 플랫폼을 설계하고 데이터를 축적하며 서비스를 제공하는 일련의 과정들이 초기 시스템 설계에서부터 오픈 데이터에 관한 지속적인 정책과 유지 관리가 있음을 알 수 있다.

둘째, 한국 위키피디아는 지난 18년간 신규 문서의 생성이 꾸준히 증가하고 있으며, 월별로 평균 78건의 새로운 문서가 만들어지고 있다. 편집 횟수와 문서의 양도 2008년까지 꾸준히 증가하였으나 2009년부터는 편집 활동이 서서히 감소하고 있음을 보여주고 있다. 평균 문서의 나이는 5.6년으로 공동 편집인들의 꾸준한 참여 속에서 문서가 업데이트 되고 있음을 알 수 있다. 편집 횟수가 많은 문서의 주제는 연예, 오락 및 스포츠이며, 편집자 수가 가장 많이 모여서 작성한 문서는 주로 정치, 사회전반에 관한 주제로서, 학술적인 내용 보다는 일반적이고 대중적인 주제 내용이 주를 이루고 있다.

셋째, 비로그인 그룹의 편집은 로그인 편집에 비해 평균 4배가 많은 수로 활동하고 있다. 2009년 이전까지는 로그인 그룹의 편집 문서 수와 편집 횟수가 비로그인 그룹에 비해 월등히 높았으나, 2010년부터는 비로그인 활동이 서서히 증가하면서 안정세를 보이고 있다. 반면 로그인 그룹의 신규 진입이 편집 문서 수에 영향을 주지 않는 것으로 보인다. 오히려 비로

그인 그룹의 신규 진입과 '활동 편집자'가 문서 편집에 높게 기여하는 것으로 분석된다.

넷째, 대다수의 편집자들은 문서 편집을 1~2회 정도 하는 소극적인 참여를 하며 평균 참여율도 평균 5.6개월이다. 반면에 극소수의 매우 적극적인 편집자들이 10만 건 이상의 편집과 10년 이상의 참여율을 보이는 0.1%가 있음을 알 수 있었다. 이러한 현상은 최근 Yun, Lee, & Jeong(2019)의 연구에서 밝혀진 바와 같이 소수 편집자들의 영향력이 커지고 유지되는 패턴이 나타난다는 것을 실증적인 데이터로 보여 주고 있는 것이다.

다섯째, 한국 위키피디아 관리자는 본 연구 결과 총 22인으로 밝혀졌는데, 분쟁 조정, 문서 삭제, 문서 보호, 수정 되돌리기, 편집자의 제명 등의 조정 역할을 수행하고 있으며, 일반 편집자들 중에 선출되어 활동한다. 관리자들의 1인당 평균 편집 기간은 11년이었고 평균 편집 횟수도 일반 편집인들에 비해 월등히 높은 것으로 나타났다. 이들의 대다수가 2003년 이후 초창기 편집자들로 구성되어 있으며, 2011년 이후에 편집을 시작한 편집자 중에 관리자가 된 경우는 한 명이다. 온라인 협업 조직도 관리자에 의한 사회적 계층구조가 나타나고 있으며, 한국은 편집인 수에 비해 관리자의 수가 많지 않아 덜 수직적이라고 연구된 바 있으나(Polat & Akbiyik, 2019), 본 연구 결과 관리자 구성이 한국 사회의 문화인 연공서열(가입 우선순위)의 위계질서를 보여주고 있으며, 이는 기존 연구에서 밝혀진 바와 같이 새로운 권위유형의 재형성과 지식권력이 생겨나는 현상이라고 볼

수 있다.

전 세계적으로 위키피디아의 기술적 플랫폼은 동일하지만, 내용이 구축되는 과정에서 생겨나는 문화적 특징들이 각 언어 커뮤니티별로 오프라인과 동일하게 온라인에서도 적용되고 있음을 알 수 있었다. 기존의 백과사전은 편집자들이 해당 백과사전의 발간 목적, 편집 방향의 기준을 정하고 전문가 집단으로 편집으로 이루어졌다면, 위키피디아는 기본적인 다섯 원칙<sup>19)</sup>을 기반으로 각 언어 커뮤니티별로 규칙과 조정을 만들어간다.

한국 위키피디아의 경우, 신규 진입자가 꾸준히 이어지고 있지만 이들이 '활동하는 편집자'로 안착하는 과정이 용이해 보이지 않으며, 로그인 편집자들의 문서 편집 수가 확연히 줄어들고 있는 반면에 극소수의 편집인들의 활동은 매우 활발하다. 온라인 협업 커뮤니티가 시간이 축적될수록 소수의 편집자들이 다수의 편집을 하게 된다면, 지식의 주제와 내용은 편향된 관점을 가질 수 있는데 한국의 위키피디아도 이러한 현상을 보이고 있다. 이를 개선하기 위해서는 위키피디아 공동편집 참여의 다양성에 대한 다각적인 고려가 필요하다. 위키피디아에 '관심 있는 이들'이 '활동하는 편집자'로 전환되어 보다 다각적인 주제와 내용이 증진될 수 있도록 오랜 기간 활동해 온 편집자들이 열린 마음으로 함께 협업할 수 있는 방안을 마련하는 것이 필요하다. 특히 위키피디아는 다른 온라인 협업 플랫폼에 비하여 관리자들의 권한이 높은 편이므로 편집에 참여하여 방향성을 이끌어 가는 중요한 요소가 된다. 한국 위키피디아 역시

19) 위키피디아는 백과사전, 중립적 시각, 공유 콘텐츠, 다른 사용자의 존중, 엄격한 규칙이 없음이라는 다섯 원칙이 있다.

누구나 관리자가 될 수 있는 가능성이 열려있지만 실제로는 관리자 그룹이 고착화 되어 있는 상황이다. 앞으로는 보다 다양한 편집자들이 관리자로 진입할 수 있는 방안이 마련되어 관리자 그룹의 개방성을 도모하여야 한다. 기존 관리자 멤버들의 순환 구조가 이루어지지 않는다면 편향된 방향으로 갈 수 밖에 없으며, 이러한 현상은 공동 협업 플랫폼이 지속적으로 발전적으로 나아가는데 방해 요소가 된다.

온라인 협업 지식백과인 위키피디아는 지식의 협업, 공유, 개방을 목적으로 다수의 자발적인 편집자들이 모여 공동의 목적을 이루어가고 있으며 Sundin(2011)은 위키피디아가 오늘날 사회에 있어 필요한 정보활용능력 훈련의 장으로 가치 있다고 하였다. 네트워크 사회는 탈중심적 사회이며, 창의적인 다종의 소통과 협력을 통해 다양한 정보를 효과적으로 통합해 낼 때에만 집단지성이 유지될 수 있다. 특히 지식백과의 역할을 하는 위키피디아는 새롭고 다양한 외부자원의 끊임 없는 공급과 다수가 눈여

겨보면서 오류를 수정해가는 공동편집시스템을 지속적으로 유지하는 것이 매우 중요하다. 이러한 커뮤니티를 유지하기 위해서는 누구나 자유롭게 활동할 수 있는 기반을 마련하는 것이 필요하고 건강한 온라인 조직문화를 만들어 가는 것이 필요하다.

본 연구는 그동안 담론적으로 다루어왔던 한국의 위키피디아 연구를 확장하여 18년간의 위키피디아 편집 로그 빅데이터를 해체 분석하여 실증적인 데이터로 설명하였다. 한국 위키피디어가 보다 건강하고 지속적인 집단지성 협업공동체가 되는데 본 연구가 그 기초가 되기를 희망한다. 향후 온라인 협업 커뮤니티의 보다 깊은 이해를 위해서 위키피디아의 토론 영역을 포함한 공동 편집 과정에 대한 커뮤니케이션 분석과 편집 유형, 실제 편집자들의 참여 동기, 탈퇴자들에 대한 질적 연구 등의 심층적인 연구와 더불어 다양한 집단지성 협업 공동체에 대한 비교 연구가 지속적으로 보완되기를 희망한다.

## 참 고 문 헌

- 박정원, 양수철, 문혜정 (2017). 한국 위키피디아의 편집자 네트워크를 이용한 융합분야 탐색. 한국경영학회 통합학술발표논문집, 212-230.
- 손준우 (2012). 위키피디아 편집 논쟁을 통해 본 지식-주체의 형성: 한국어 위키피디아에 있어 저명성 논쟁의 레퍼토리 분석. 문화연구, 1(1), 187-243.
- 심원식, 변제연, 김민정 (2013). 학술논문에서의 위키피디아 인용에 관한 연구. 한국문헌정보학회지, 47(2), 247-264. <https://doi.org/10.4275/KSLIS.2013.47.2.247>
- 윤은호 (2012). 집단지성 및 백과사전의 한계 그리고 미래. 한국사전학회 학술대회 발표논문집, 81-90.
- 이강봉 (2017. 9. 27). 과학계 위키피디아 인용 급증. The Science Times.

- 이향우 (2013). *클릭의 사회학: 페이스북에서 위키피디아까지 디지털 민주주의 깊이 읽기*. 서울: 이매진.
- 장순규, 강하영, 윤재영 (2018). 지식공유를 위한 집단지성 온라인 서비스로서 위키서비스 고찰: 나무위키와 위키백과 비교를 중심으로. *한국디자인학회 학술발표대회 논문집*, 80-85.
- 정승화 (2016). 인터넷 백과사전 위키피디아와 지식 권력의 재구성. *인문과학*, 106, 87-119.
- 진주완, 정철, 류철 (2018). 위키백과, 우리모두의 백과사전. 서울: 사계절출판사.
- 한창진 (2009). 지식생산 방식에 따른 집단지성 구조 분석: 네이버 지식 IN과 위키피디아를 중심으로. *한국HCI학회 학술대회*, 1363-1373.
- 황주성, 최서형 (2010). 집단지성의 유형에 따른 참여자의 특성 분석: 위키백과와 지식IN에 대한 한미 비교 연구. *사이버커뮤니케이션 학보*, 27(4), 257-301.
- Arazy, O., Lifshitz-Assaf, H., Nov, O., Daxenberger, J., Balestra, M., & Cheshire, C. (2017). On the “How” and “Why” of emergent role behaviors in wikipedia. *CSCW 2017 conference: proceedings of the 2017 ACM conference on computer-supported cooperative work and social computing*, 2039-2051. <https://doi.org/10.1145/2998181.2998317>
- Balestra, M., Arazy, O., Cheshire, C., & Nov, O. (2016). Motivational determinants of participation trajectories in Wikipedia. In *proceedings of the 10th international conference on web and social media*. ICWSM 2016, 535-538.
- Jiang, L., Mirkovski, K., Wall, J.D., Wagner, C., & Lowry, P. B. (2018). Proposing the core contributor withdrawal theory (CCWT) to understand core contributor withdrawal from online peer-production communities. *Internet Research*, 28(4), 988-1028. <https://doi.org/10.1108/IntR-05-2017-0215>
- Karkulahti, O., & Kangasharju, J. (2012). Surveying wikipedia activity: Collaboration, commercialism, and culture. *The International Conference on Information Network 2012*, 384-389. <https://doi: 10.1109/ICOIN.2012.6164405>
- Lieberman, Y. S., & Bar-Ilan, J. (2019). Individual and collaborative information behaviour of Wikipedians in the context of their involvement with Hebrew Wikipedia *Information Research*, 24(4), paper 843. Retrieved from <http://www.informationr.net/ir/24-4/paper843.html>
- Polat, M., & Akbiyik, A. (2019). Analysis of cultural differences in collaborative innovation networks: Wikipedia. *Yönetim ve Ekonomi: Celal Bayar Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 26(2), 547-555. <https://doi.org/10.18657/yonveek.565210>
- Sundin, O. (2011). Janitors of knowledge: Constructing knowledge in the everyday life of Wikipedia editors. *Journal of Documentation*, 67, 840-862. <https://doi.org/10.1108/00220411111164709>
- Wikimedia Foundation (2017). *New editor experiences: Summary of findings from Korean and Czech Wikipedia*. Retrieved from

- [https://www.mediawiki.org/wiki/New\\_Editor\\_Experiences](https://www.mediawiki.org/wiki/New_Editor_Experiences)  
Wikimedia Foundation (2018). Community engagement insights 2018: Fostering learning to improve support to Wikimedia communities. Retrieved from [https://meta.wikimedia.org/wiki/Community\\_Insights/2018\\_Report](https://meta.wikimedia.org/wiki/Community_Insights/2018_Report)  
Wikipedia (2020). Wikipedia. Retrieved from <https://en.wikipedia.org/wiki/Wikipedia>  
Wikipedia statistics (2018). Wikipedia statistics: all languages, 2018. Retrieved from <https://stats.wikimedia.org/EN/TablesWikipediaZZ.htm>  
Yun, J. H., Lee, S. H., & Jeong, H. (2019). Early onset of structural inequality in the formation of collaborative knowledge in all Wikimedia projects. *Nature Human Behavior*, 3, 155-163. <https://doi.org/10.1038/s41562-018-0488-z>

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

- Han, Changjin (2009). Collective intelligence structure analysis: naver knowledge IN and wikipedia. HCI Society of Korea Conference Proceeding, 1363-1373.  
Hwang, Joo-Sung, & Choi, Seo-young. (2010). Analysis of participants' features in different collective intelligence models: Comparative analysis between Korea and U.S.A. *Journal of Cybercommunication Academic Society*, 27(4), 257-301.  
Jang, Soon-kyu, Kang, Ha-Young, & Yun, Jae-Young (2018). A study on the wiki-service of collective intelligence online service to share user knowledge: Focused on comparison Namuwiki and Wikipedia service. *KSDS Conference Proceeding*, 80-85.  
Jin, Ju-Wan, Chung, Chul, & Ryu, Chul (2018). *Wikipedia: The free encyclopedia*. Seoul: Sakaejeol Publisher. <https://doi.org/10.979.116094/2989>  
Joung, Seung-Hwa (2016). The web based encyclopedia wikipedia and the reconstruction of knowledge-power. *The Journal of the Humanities*, 106, 87-119.  
Lee, Hang-Woo (2013). *Sociology of Click: Deep reading of digital democracy from facebook to wikipedia*. Seoul: Imagine Publisher. <https://doi.org/10.978.8993985/924>  
Lee, Kang-bong (2017, September 27). Rapid citation of the scientific community on Wikipedia. *The Science Times*.  
Park, Jung-won, Yang, Soo-Chul, & Moon, Hye-Jung (2017). Discover integrated field by ko-wikipedia editor network. *Proceedings of the Korean Operations and Management Science Society Conference*, 212-230.

- Shim, Won-sik, Byun, Jae-yeon, & Kim, Min-jung (2013). Analysis of wikipedia citations in peer-reviewed journal article. *Journal of the Korean Library and Information Science Society*, 47(2), 247-264. <https://doi.org/10.4275/KSLIS.2013.47.2.247>
- Son, Joon-woo (2012). Formation of knowledge-subject through wikipedia editorial debate: 'Notability' debate in Korean wikipedia and available repertories. *Cultural Studies*, 1(1), 187-243.
- Yoon, Wn-Ho (2012). The collective intelligence and limitations of the encyclopedia, and the future. *Proceedings of the Korea lex Conference*, 81-90.

