

Big Data and the Prospects of Historical Research

- A study of research in modern and contemporary Korean history -

Minki Moon*

Introduction

At some point in Korean history, the term “Fourth Industrial Revolution” has become a word describing an age that will soon begin or has already begun. Klaus Schwab stated that “a ubiquitous and mobile internet, small powerful cheap sensors, artificial intelligence, and machine learning” are distinguishing characteristics of the Fourth Industrial Revolution,¹ and as such the key to this revolution is a historic development in information technology (IT) and digital revolution.

The term “Fourth Industrial Revolution” is not a newly coined term. It was used to describe the emergence of electronic engineering in 1955, the computer age in the 1970s, and the development of information and communication technology (ICT) in 1984. The same term was also used to discuss nanotechnology in the 1990s.² “The Fourth Industrial Revolu-

* Ph.D. candidate in Korean History at Korea University

1 Klaus Schwab, *Klaus Schwab-ŭi che 4-cha sanŏp hyŏngmyŏng* (Klaus Schwab’s Fourth Industrial Revolution), trans. Song Kyŏngjin, (Seoul: Saeroun Hyŏnjae, 2016), 25.

2 Pak Pyŏngwŏn, “Introduction: ingong chinŭng, robot, pik teit’ŏ-wa che 4-cha sanŏp hyŏngmyŏng,” (Introduction: AI, robotics, Big Data, and the Fourth Indus-

tion” was used early on in Korea as well. On October 4, 1983, the Federation of Korean Industries (FKI, Chŏn’guk Kyŏngjein Yŏnhaphoe) invited W. W. Rostow, an American economist known for his theory of modernization, to hold a lecture under the title “Korea and the Fourth Industrial Revolution: 1960-2000.” Rostow explained that the Fourth Industrial Revolution includes, “innovations in micro-electronics, communications, the offshoots of genetics, robotics, the laser, robots, and new synthetic materials.”³ This was from 36 years ago, but even then “technological innovation” was the main keyword used to describe the Fourth Industrial Revolution, as it is today.

Different people have different opinions regarding the appropriateness of the use of the term “Fourth Industrial Revolution” and its distinguishing characteristics. However, changes dubbed as “technological innovations” are occurring, and that is an undeniable fact. The Fourth Industrial Revolution has seen the emergence of numerous new technologies, which have and will continue to have an impact on various areas of society. In the FKI lecture mentioned earlier, Rostow accurately forecast that the “new technologies are likely to be ubiquitous, affecting virtually all sectors of the economy: the old basic industries; agriculture, forestry, and animal husbandry; and the service sectors from medical care to education.”⁴

The Fourth Industrial Revolution has a huge impact on the academia as well. The development of technologies that is driving the change to a digital environment is closely related to and has a huge impact on our

trial Revolution) *Future Horizon* 28 (February 2016): 4; Hong Sŏnguk, “Wae 4-cha sanŏp hyŏngmyŏngron-i munjein’ga?,” (What is the problem with the theory of the Fourth Industrial Revolution?) edited by Hong Sŏnguk, *4-cha sanŏp hyŏngmyŏng-iranŭn yuryŏng* (The phantom called the Fourth Industrial Revolution) (Seoul: Humanist, 2017), 32.

3 W. W. Rostow, “Han’guk-gwa che 4-cha sanŏp hyŏngmyŏng: 1960-2000,” (Korea and the Fourth Industrial Revolution: 1960-2000) *Chŏn’gyŏngnyŏn* (November 1983): 47.

4 W. W. Rostow, “Han’guk-gwa che 4-cha sanŏp hyŏngmyŏng: 1960-2000,” 50.

lives. Discussions on the development of technologies also emerged early on in the humanities, which led to the emergence of discourses on the digital humanities.⁵ The digital humanities explores how digital technology should be used in the disciplines of the humanities and how the humanities should change in the digital age. This is related to the question “How should academic disciplines, specifically historical studies, respond to and use the technological innovations of the Fourth Industrial Revolution?” This paper is an attempt to find an answer to this question.

One technology that never fails to appear in the discussion of the Fourth Industrial Revolution is big data, which forms the basis of other technologies. Science and technology are not the only fields that are built upon data. In fact, all academic disciplines are built on data. Historical studies, in particular, is based on data accumulated over a long period of time, known as historical sources. As a result, data and historical studies are inseparable.

And now is the age of big data. As a historical scholar, understanding and utilizing big data means looking for ways to organize and use a vast corpus of data, or digital historical sources. This paper aims to examine the significance of big data and the points of intersection between big data and historical studies. Furthermore, this study will explore the ways in

5 For the definition of the concept of digital humanities and relevant discourse in Korea, see: Ch'oe Hüisu, “Tijit'öl inmunhak-üi hyönhwang-gwa kwaje,” (The status and challenges of the digital humanities), *Sot'ong-gwa inmunhak* (Communication and Humanities) 13 (August 2011); Kim Hyöñ, “Tijit'öl inmunhak— inmunhak-gwa munhwa k'ont'ench'ü-üi sangsaeng kudo-e kwanhan kusang,” (Digital humanities—a cooperative scheme between the humanities and cultural contents), *Inmun k'ont'ench'ü* (Humanities contents) 29 (June 2013).

For the trend in digital humanities overseas, see: Kim Paro, “Haeoe tijit'öl inmunhak tonghyang,” (Trends in digital humanities), *Inmun k'ont'ench'ü* (Humanities contents) 33 (June 2014); Kim Tongyun, “P'ürangsu 'tijit'öl inmunhak'-üi inmunhakjök maengnak-gwa tonghyang,” (Digital humanities in France from the perspective of human sciences), *Inmun k'ont'ench'ü* (Humanities contents) 34 (September 2014).

which big data can be used by looking at the example of research on modern Korean history. The advent of new technologies naturally leads to changes in academic disciplines, and I would like to also look into the impact of technological innovations of the Fourth Industrial Revolution on the emergence of new historical writing and historical studies, as well as the role of historical researchers amid these changes.

The Significance of Big Data

1. Big data in the age of the Fourth Industrial Revolution

The term “big data” usually conjures up the image of a huge mass of data. But the vast volume of data is only half of what big data is. The definition of big data today is centered on the change in how we can make use of a myriad of data sources. Data is no longer regarded as static or stale but rather seen as a raw material used to create a new form of economic value.⁶ Bernard Marr asserted that big data has no particular value unless we use our insight to make them usable, and that is the key to using big data.⁷

There is no rigorous definition of big data, but experts in related fields generally support this sentiment. The International Data Corporation (IDC), a provider of market intelligence and IT technology services, defines big data as “a range of technologies used to maximize data processing capacity and costs following the change in the platform and hardware, including large-scale memory models and cloud computing.”

6 Viktor Mayer-Schönberger and Kenneth Cukier, *Pik teit'ō-ga mandūnūn sesang* (Big Data: A Revolution That Will Transform How We Live, Work, and Think), trans. Yi Chiyōn (Seoul: 21 Segi Puks, 2013), 17.

7 Bernard Marr, *Pik teit'ō: 4-cha sanōp hyōngmyōng-ūi ōnō* (Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results), trans. An Chunu and Ch'oe Chiūn (Seoul: Hakkojae, 2017), 7.

This means that big data does not simply refer to a huge volume of data but also has to meet other conditions, such as the high processing speed and the data management infrastructure. Similar sentiments are expressed in the following definition of big data: “A new generation of technologies and architectures designed to economically extract value from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and/or analysis.” Another definition is “technologies with broader capacities than traditional database management tools, which was limited to data collection, storage, management, and analysis, that are used to extract value and analyze results from a huge amount of structured and unstructured data sets.”⁸ In other words, big data cannot be defined without asking and answering the question about how to analyze and utilize data.

As can be seen in these various definitions, big data is a term that emerged from the dramatic change in the volume of data. Initially, the idea was that the volume of digital information had grown so large that the quantity to be examined no longer fit into the memory that computers use for data processing, and so engineers had to revamp the tools they use to analyze all of the data. This led to the emergence of new data processing technologies. The change of scale led to a change of state, and the quantitative change led to a qualitative one. As it became possible to handle data that was unable to be collected or analyzed using traditional formats and technologies, people can now work on a much larger scale to extract new insights or create new forms of value, which could not be done on a smaller scale.⁹

Although we usually use the general term “big data” as if referring to one specific technology, big data is actually a system of a complex set of technologies, including machine learning, natural language processing,

8 Chŏng Ujin, *Pik teit'ŏ-rŭl marhada* (About Big Data) (Seoul: Kŭlaudŭ Puks, 2013) 58-60.

9 Viktor Mayer-Schönberger and Kenneth Cukier, *Pik teit'ŏ-ga mandŭmŭn sesang*, 18-19.

various statistical techniques, and distributed parallel processing. Big data technology starts from having computers understand and digitize spoken and written languages that people use in their daily lives, instead of information made up of computer commands. The vast amount of data collected in this way are then programmed and analyzed to extract and deduce complex significance using various statistical techniques, machine learning, and other AI programs.¹⁰

Big data analysis is currently being applied in various areas. It has become customary for global enterprises to use the results of big data analysis for customer management and for manufacturing of new products and marketing. Big data is expanding its influence into the research and academic domains, and it is expected to continue diffusing into other parts of our society. Although most examples of the use of big data in academic disciplines come from the sciences, such as science and technology, engineering, and medicine, the necessity of big data in the humanities disciplines will become clear when we consider the possibilities and scalability of big data.

2. Big data and historical studies

If we were to classify historical studies, it is a discipline of the humanities. Although there are numerous research topics and methodologies in the field of history, the essence of this discipline is in analyzing and understanding humans. Most people probably think of “stories about the past” when they hear the words “historical studies.” So how can stories about the past come together with big data—one of the brand new technological innovations of our time?

Most of the examples on the use of big data mentioned in books or articles focus on businesses. As a result, people often look at big data from

10 Son Minsŏn and Mun Pyŏngsun, “Pik teit’ŏ sidae-ŭi Han’guk: Kalap’agosŭ-ga toejŭ anŭryŏmyŏn,” (Korea in the age of big data: how not to become another Galapagos) *Pik teit’ŏ-rŭl marhada*, 18-20.

the perspective of technology or business. However, a close look at what big data technologies are analyzing leads you to humans. This means that data collection and analysis undertaken by big data technologies are acts of collecting and analyzing human activity. Ultimately, the purpose of collecting and analyzing data is to understand and analyze humans. And the same purpose is at the heart of historical studies. Therefore the methodology for big data analysis is also a methodology for the humanities, and historical studies and big data come together with the same goal: to understand humans.

A unique characteristic of historical studies is that instead of analyzing human activities of the present day, it focuses on the analysis of past human activities. Significant information can be gained from analyzing fragmented and seemingly insignificant data from the past. One such example is the case of United States Navy officer Matthew Fontaine Maury. Maury used a vast amount of data that no one was interested in for a long period of time and found new navigable sea routes. Nicknamed the “Pathfinder of the Seas,” he was one of the people who recognized a core tenet of big data—that there is a special value in huge amounts of data that was lacking in small amounts.¹¹ In addition, the process of Maury examining fragmented and seemingly insignificant data and analyzing them to find a new meaning is very similar to the methodologies used in historical studies. On the surface, Maury seems to have analyzed natural forces, such as the wind and sea currents, but his achievements can be marked as that of the humanities in that he collected and analyzed data on past human activities of fighting nature and produced results.

As Maury’s case shows, finding useful information in a huge corpus of data is important for all disciplines. Named data mining, this is the process of discovering meaningful patterns and rules in large data sets using automatic or semi-automatic analytical tools. The term “mining” perfectly captures this process, as it likens extracting information from datasets to

11 Viktor Mayer-Schönberger and Kenneth Cukier, *Pik teit’ō-ga mandūnŭn sesang*, 139-145.

find hidden values to mining coal or crude oil.¹²

Similarly, all academic disciplines also use a similar process. Historical studies, in particular, is a discipline where data mining becomes a basic task. History scholars spend a lot of time and efforts in collecting the past data, and they produce results by extracting and analyzing meaningful information from them.

With the help of technological innovations of the Fourth Industrial Revolution, there have been huge developments in collecting and analyzing unstructured data. Unlike numeric data, unstructured data refers to information that does not have a predefined or structured model, such as images, videos, or documents. Traditional data, such as books, periodicals, documents, medical records, voice information, and image information, as well as data generated on the internet or via mobile devices, such as email, tweets, and blogs, are considered unstructured data. There are different types of unstructured data mining, such as text mining, which refers to the process of extracting useful information from huge volumes of textual data; web mining, the process of extracting meaningful information from web logs or search words generated in using the Internet; and opinion mining, the process of systematically analyzing subjective information, such as people's opinions, thoughts, attitudes, and emotions regarding certain issues, people, or events.¹³

Collecting and analyzing unstructured data is expected to become even more important in historical studies in the future. Even now, information relevant to the various incidents and facts that historical studies aim to analyze are continuously generated as unstructured data. The National Institute of Korean History (Kuksap'yŏnch'an Wiwŏnhoe) and other related organizations are currently collecting historical sources and digitizing them. Through their efforts, extensive data including scanned images of documents as well as voice recordings and videos are being collected. Historical sources are also being digitized as image files. Unstructured

12 Chŏng Yongch'an, *Pik teit'ŏ* (Big Data) (Seoul: Kŏmunikeishyŏn Puksŭ, 2013), 32.

13 Chŏng Yongch'an, *Pik teit'ŏ*, 42-49.

data collection is an important process that is currently ongoing and will continue in the future, along with technological innovations for convenience access and use of the collected unstructured data.

As examined so far, historical studies and big data share the purpose of analyzing human activities. And technologically, big data will be amazingly useful in researching history, both for structured and unstructured data. Although both tasks of collecting and analyzing data rely on individual researchers at this point, it is highly likely for the methods of data collection and analysis to change dramatically in the future.

Collecting Data and Writing History

1. The need to collect vast amounts of data

Historical studies is a discipline that involves restructuring, analyzing, and interpreting the past events using historical sources. At the same time, it also involves accumulating current sources and writing about them. Just as Chosŏn (1392-1910) dynasty court historians (*sagwan*, 史官) produced drafts and records of the events of their time, historians today have an obligation to collect sources and write the history of the present age. This is a role that is particularly required of scholars studying modern and contemporary Korean history. Of course, the sources they collect will then be historical sources used by future historians.

Therefore collecting and organizing big data as historical sources that show the way we are today and analyzing them will be one of the most important tasks in future historical studies. Primarily, historians must decide on the type of data and explore ways to collect, organize, and preserve them. This should be done in collaboration with archivists who are always thinking of ways to improve record and source management. In archival studies, there have been a number of studies on analyzing the advent of the Fourth Industrial Revolution and big data as well as their

impact and implications on record management and services.¹⁴ Writing contemporary history means analyzing the current Korean society, and a number of disciplines including historical studies are currently writing about the current times. Therefore it is important to bear in mind ways to carry out exchange with academic disciplines other than archival studies, such as social science, political science, and economics.

Humans create huge amounts of records. In the past, such records have been left by the intelligentsia and the rulers of society or by public institutions. In the modern age, however, the individual was discovered, and it became possible for numerous individuals to create their personal records. After learning how to write, people began to keep records of their daily lives. The publication of newspapers and periodicals also enabled the lives of numerous individuals to be recorded and preserved. And today, it has become possible for individuals to create personal records in real time on social media or online communities, as we can see from the numerous posts and images uploaded to these services. These are important historical sources for us to write a contemporary history and for future historians to write about the history of our times in later years. This is why Song Chuhyōng asserted that social media needs to go down in history as the topic and portrait of our society today.¹⁵

14 Yun Ch'ol, "Pik teit'ō-e kibanhān kirok chōngbo sōbisū-ūi panghyang," (Archival reference services based on big data) MA thesis, Hanshin University, 2013; Cho Pyōngch'ol and Yuk Hyōnsūng, "Pik teit'ō sidae munhwajōk kiōk pojōnsorosōūi yōngsang akaibū-ūi yōkhal," (The role of archive as a cultural storage in the age of big data) *Tijit'ol yungbokhap yōn'gu* (Journal of Digital Convergence) 12, no.2 (December 2014); An Daejin and Im Chinhūi, "Che 4-cha sanōp hyōngmyōng kisul-ūi kirok kwalli chōgyong pangan," (Application of Fourth Industrial Revolution technology to records management), *Kirokhak yōn'gu* (The Korean Journal of Archival Studies) 54 (October 2017); Chin Chuhyōng, "Kukka kirokwōn websait'ū-ūi pik teit'ō punsōk-gwa hwaryōng," (Big data analysis and use on the National Archives of Korea website) MA thesis, Myongji University, 2018.

15 Song Chuhyōng, "Kirok kwalli taesang-ūrosō SNS yōn'gu" (A study on records management by social media services), *Kirokhak yōn'gu* (The Korean Journal of Archival Studies) 39 (January 2014): 103.

It is also necessary for us to consider records kept by national and public institutions as big data and collect them. Numerous records in public institutions are already being generated as textual data and internet logs, which are being collected and preserved. But in addition to this, it is important to convert historical sources that have been recorded on paper into unstructured data (in an image format) and to collect and preserve them. The future development of big data technology will greatly increase the utility of image sources.

Publications, such as newspapers and periodicals, are already considered important historical sources, and there are online services that provide images of the original publications, which are often used in historical research. Historians studying modern or contemporary Korean history frequently use the Naver News Library service.¹⁶ Naver News Library provides images of original newspapers for *Kyunghyang Shinmun* (Kyōnghyang shinmun), *Dong-A Ilbo* (Tonga ilbo), *Maeil Business Newspaper* (Maeil kyōngje shinmun), *The Hankyoreh* (Hankyōre), published from 1920 to 1999. In addition, the contents of the newspapers are also available in the text format, allowing researchers to search for words within the text to locate the relevant article. *The Chosun Ilbo* (Chosōn ilbo) provides PDFs of its past issues on its website as well.¹⁷ Newspapers and periodicals should be handled as big data from now on, since most are available via print and also online. In fact, print media seems to be disappearing.

Most of the records that are generated today and will be used as sources in the future to write about our times are in the digital format. These are not simple data; rather they qualify as big data, which cannot be collected or analyzed using traditional technology.

All digital data, however, is at risk of being lost easily. It is particularly difficult to preserve the posts shared on social media or online communities. Once the person who created the record deletes it, that record is lost

16 Naver News Library (newslibrary.naver.com)

17 Chosun Ilbo Database Chosun (http://srchdb1.chosun.com/pdf/i_service/index.jsp)

forever. The records on social media and online communities seem highly preservable since they exist on web servers, but the problem is that these services are provided by companies. If companies cease to provide the services, all of the records created via these services will be lost.¹⁸ There is also a possibility of documents from national and public institutions and newspaper articles on the internet being deleted. Digital data is highly volatile, and therefore measures should be taken to collect and preserve them in order to prevent the loss of future historical sources.

So far I have only discussed ways to keep and preserve records that are being created today, with a focus on writing contemporary history. But it is also important to work on digitizing past records and sources. Once all the data are turned into big data and accumulated, it will create positive synergy in historical research, along with the development of AI, an important Fourth Industrial Revolution technology, which will greatly facilitate the management and analysis of unstructured digital data.

2. Methodologies for history writing using big data

Once progress has been made on the collection of big data of historical sources, the next step is data analysis and history writing. In this section, I will examine case studies of research on modern and contemporary Korean history and look for future possibilities. I would like to note that there are many different ways of history writing than the ones presented in this paper.

There is research that involves the analysis of posts on online communities and social media. Recognizing active participation of women and teenagers as one of the distinguishing characteristics of the 2008 candle-light rallies, Pak Ch'angsik analyzed the identity of women in their 20s and 30s who participate in political online communities as well as the characteristics of their communication.¹⁹ The widespread surge of online

18 Song Chuhyŏng, "Kirok kwalli taesang-ŭrosŏ SNS yŏn'gu," 122-123.

19 Pak Ch'angsik, "Chŏngch'ijok sot'ong-ŭi saeroun chŏnmang: 20-30 tae yŏsŏng-

communities and online social networks gave rise to a number of influential communities with strong political leanings, though not blatantly supporting certain politicians originally. Major examples of such communities are Ilgan Pesūt'ŭ Chōjangso (Ilbe) and Megalia. In his study on the culture of online communities, Pak Kabun also analyzed the posts on these sites to examine the participants' thoughts and culture.²⁰ There are a number of academic papers that examined and analyzed Ilbe, which had a particularly huge impact on Korean society.²¹ Other researchers also have analyzed posts on social media, such as Twitter and Facebook, mainly focusing on the political aspect, analyzing public opinion or expressions of political opinions.²²

dūrūi onlain chōngch'i k'ōmunit'i-rūl chungshimūro (New prospects of political communication: focusing on political online communities for women in their 20s and 30s), PhD dissertation, Kwangwoon University, 2010.

- 20 Pak Kabun, *Ilbe-ūi sasang* (The Ilbe Ideology) (Seoul: Owōr-ūi Pom, 2013); *Hyōmo-ūi mirōring: hyōmo-ūi sidae-wa Megalia sindūrom parobogi* (Mirroring hate: the age of hate and the Megalia syndrome) (Seoul: Pada Ch'ulp'ansa, 2016).
- 21 Han Yunhyōng, "Han'guk chwaup'a t'ujaeng-ūi hūrūm sogesō 'Ilbe'-rūl paraboda: Ilbe-nūn kijon-ūl chwaup'a-wa öttōke talmatko, tto tarūn'ga," (Looing at Ilbe in the flow of struggle between Korean leftists and rightists: how is Ilbe similar and different from the Korean leftists and rightists?) *Chinbo p'yōngnon* (The Radical Review) 57 (September 2013); Yun Pora, "Ilbe-wa yōsōng hyōmo: Ilbe-nūn ödiena itko ödiedo öpda," (Ilbe and misogyny: Ilbe is everywhere yet nowhere) *Chinbo p'yōngnon* (The Radical Review) 57 (September 2013); Ki Hakjun, "Int'ōnet kōmunit'i Ilbe Chōjangso-esō nat'ananūn hyōmo-wa yōlgwang-ūi kamjōngdonghak (Dynamics of cyber hate and effervescence: focusing on the Korean internet community Ilbe Chōjangso), MA thesis, Seoul National University, 2014.
- 22 Ryu Ch'ōlgyun and Yi Chuhi, "T'ūwit'ō-rūl t'onghan sōn'gō huboja-ūi sūt'orit'eling punsōk: 4:27 chaebogwōl sōn'gō kigan-ūi Ch'oe Munsun, Ōm Ki-yōng hubo-ūi t'ūwis-ūl chungshimūro," (A storytelling analysis of election candidates via Twitter: focusing on the Twitter accounts of Ch'oe Munsun and Ōm Ki-yōng up to the April 27 by-election), *Inmun k'ōnt'ench'ū* (Humanities Contents) 23 (December 2011); Chang Tōkjin, "T'ūwit'ō konggan-ūi Han'guk chōngch'i: chōngch'iin net'ūwōk'ū-wa yugwōnja net'ūwōk'ū," (Korean politics on Twitter: networks of politicians and voters) *Ōnon chōngbo yōn'gu* (Journal of Communi-

The analysis of online communities and social media is gaining traction since it has gradually become a methodology for examining people's awareness of reality or important social issues. Although the term "big data" is not directly mentioned in these research studies, they are classified as big data research since it involves far greater volumes of data than in the past.²³ The analysis of online communities first requires researchers to search for related posts, collect them and analyze them. Therefore in order to develop this as a methodology, it will be necessary to make improvements and changes are expected to occur in this area.

Continued development of big data technologies will help researchers expand their research interest into various areas instead of limiting it to a certain political event or aspect, since online communities and social media services are not just tools for gauging public opinion but a trove of data about our daily lives. Social media posts and blogs in particular are

cation Research) 48, no.2 (August 2011); Hong Chuhyŏn and Yi Ch'anghyŏn, "T'ŭwit'ŏ-esŏ hyŏngsŏngdoen chŏngch'ijŏk ŭigyŏn punsŏk-ŭl t'onghan punhwa-doen kongjung yŏn'gu: 10:26 Sŏul sijang chaebogwŏl sŏn'gŏ-rŭl chungshimŭro," (A study of the public typology based on the analysis of political opinions on Twitter: focusing on the October 26 by-election for the mayor of Seoul) *Han'guk ŏnnon chŏngbo hakbo* (Korean Journal of Communication and Information) 59 (August 2012); Pak Chiwŏn and Pak Hanu, "P'eisŭbuk p'aenp'eiji-ŭi tongshi taetgŭl teit'ŏ-rŭl iyonghan net'ŭwŏk'ŭ punsŏk: Taegu Kyŏngbuk yuryŏk huboja-rŭl chungshimŭro," (Social Network Analysis among Facebook Fanpage Co-commenters: Daegu-Gyeongbuk's Mayor-Governor Candidates) *Journal of the Korean Data Analysis Society* 16, no.6 (December 2014); Pak Chiyŏng and Pak Hanu, "Ŭimimang punsŏk-ŭl t'onghan P'eisŭbuk taejung yŏron-ŭi yŏkdongsŏng punsŏk: Sŏul kyoyukgam sŏn'gŏ-rŭl chungshimŭro," (An Exploratory Semantic Analysis in the Dynamics of Public Opinion on Facebook: A Case of the Superintendent of Seoul Office of Education in South Korea) *Journal of the Korean Data Analysis Society* 17, no.3 (June 2015).

- 23 Chang Tŏkjin analyzed a total of 77,425,090 tweets posted from 145 Korean politicians' accounts and 1,113,365 Korean accounts on Twitter, collected from August 1 to September 30, 2010. Chang Tŏkjin, "T'ŭwit'ŏ konggan-ŭi Han'guk chŏngch'i: chŏngch'iin net'ŭwŏk'ŭ-wa yugwŏnja net'ŭwŏk'ŭ," 81-82.

similar to daily journals or diaries kept by people in modern Korea. It is easy to find Korean historical research papers based on the analysis of personal diaries. Chŏng Pyŏnguk reconstructed the life and thoughts of Kang Sanggyu, a student in Kyŏngsŏng (present day Seoul) in colonial Korea, based on his personal diaries, and also identified the characteristics of the March 1st Movement in 1919 using personal diary entries from that time period.²⁴ Many researchers also used diary and journal data to analyze individuals' lives and their perceptions on various subjects, including tradition, modernity, colonialism, and nation.²⁵ There are also studies that analyzed the daily lives, rites, and public forums of farmers in rural villages, using *P'yŏngt'aek Taegok Ilgi* (Diary from Taegok, P'yŏngt'aek), written by a farmer in P'yŏngt'aek.²⁶

Another big transformation in historical studies is the change in the type of data that is being accumulated. Until now, most records that have

24 Chŏng Pyŏnguk, *Singminji puronyŏlchŏn: mich'in saenggak-i paetsogesŏ naonda* (A book of rebellious people in colonial Korea: insane ideas come from the pit of the stomach) (Seoul: Yŏksa Pip'yŏngsa, 2013); *1919 Samil undong-gwa ilgi charyo* (The March 1 movement in 1919 and journals), *Han'guk sahakbo* (The Journal for the Studies of Korean History) 73 (November 2018).

25 Chŏng Pyŏnguk and Itagaki Ryuta, edited, *Ilgi-rŭl t'onghae pon chŏnt'ong-gwa kŭndae, singminji-wa kukga* (The tradition and the modern, colony and nation seen through journals) (Seoul: Somyŏng Ch'ulp'an, 2013). This book contains a "list of sources and research on modern and contemporary journals.

26 Ahn Hye-Gyoung (An Hyegyŏng), "'P'yŏngt'aek ilgi'-rŭl t'onghae pon ilsaeng ūirye-wa sokshin," (The rites of passage and the folk beliefs in the 'Pyeongtaek Diary'), *Silch'ŏn minsokhak yon'gu* 18 (August 2011); Kim Young-Mi (Kim Yŏngmi), "'P'yŏngtaek taegok ilgi'-rŭl t'onghaesŏn pon 1960-70 nyŏndae ch'o nongch'on maŭl-ŭi kongnonjang, tonghoe-wa mashilbang," (Village office and *masilbang*, the public opinion sphere in Korean villages in early 1960s and 1970s) *Han'guksa yŏn'gu* (The Journal of Korean History) 161 (June 2013); Yang Yunju, "1960-1970 nyŏndae nongmin-ŭi ilsangsaenghwal-gwa chŏngch'esŏng-ŭi pyŏnhwa," (The daily lives of farmers in the 1960s and 1970s Korea and their identity change: a study of farmers' journals), MA thesis, Kookmin University, 2017.

been preserved were in the text form, such as diaries and collection of writing. Then images were added in records from online communities and social media, and now videos have become the main form of record keeping. With the increasing growth in the number of creators (dubbed one-person media in Korean), YouTube has seen an exponential increase in the amount of data uploaded to its servers. More and more people will begin to create records of their lives in the form of videos instead of texts and images, and those videos will be important historical sources that provides glimpses into the lives of people today. If it becomes possible to collect and use videos, a type of unstructured data, we will be able to see a dramatic development in historical research and other fields of study. Big data of our daily lives and culture will continue to grow with the development of the Internet of Things (IoT). The IoT records all of our activities that we are not even aware of and stores them as data. Access to this kind of data will enable historians to write about the basic necessities of our lives (where we live, what we eat, and what we wear), leisure activities, health conditions, and more.

History writing using newspapers and periodicals is a methodology that is widely used in different fields of research. Since these are valuable sources that show the general contemporary history, their importance will never fade in research. The difference will be in their usage, since articles in newspapers and periodicals will be stored as digital data. As mentioned earlier, the change in the media will necessitate the establishment of methods for collecting and organizing records and documents. Therefore big data in historical studies is expected to enable researchers to paint a more dynamic and multi-dimensional portrait of the times.

Currently, research on newspaper big data is in the initial stage. One of the studies on newspaper big data examined “Mulgyöl 21,” a large-scale newspaper corpus that the Center for Digital Humanities (Tijit’öl Inmunhak Sent’ö) at Korea University’s Research Institute of Korean Studies (Minjok Munhwa Yön’guwön) has compiled since 2008.²⁷ By

27 “Mulgyöl 21” received articles from four major Korean newspapers (Chosun Ilbo,

analyzing keywords and co-words extracted from newspaper big data, this study aimed to follow the change in the meaning of certain concepts over time and also studied the changes in keywords related to the Korean people's identity and ideological conflicts within the division system of the Korean Peninsula. Hur Soo (Hō Su) also traced the change in the meaning of the concept of *minjung* (often translated as “the masses”) based on *Dong-A Ilbo* articles from the 1980s.²⁸ This kind of corpus analysis certainly is one method of analyzing newspaper big data.

An examination of previous research and a look at the possibility of future development show that volumes of data have grown larger while the methodologies for history writing have remained the same. However, the vast amount of data will certainly show patterns and interpretations from different perspectives that could not be seen using past sources, and the far larger size of analytical samples than in the past will also increase objectivity. In the study of modern and contemporary Korean history, the use of big data is only in the initial stage. Over time, the development of big data technologies will lead to the emergence of new and innovative methodologies, and it will be up to historical researchers to explore and search for new options.

Dong-A Ilbo, Joongang Ilbo, and the Hankyoreh) since 2000. In the first half of every year, “Mulgyōl 21” also acquired articles from these newspapers from the previous year and has created a database of newspaper articles by polishing, processing, and storing them. Currently, “Mulgyōl 21” consists of about 2.57 million articles, with an annual average of 180,000 articles. The cumulative number of word segments in “Mulgyōl 21” amounts to about 600 million. Kim Ilhwan et al., *Kiwōdū, konggiō, kūrigo net'ūwōk'ū: shinmun pik teit'ō-ga poyōjunūn kōt* (Insights into keywords, co-words, and networks) (Seoul: Somyōng Ch'ulp'an, 2017), 57-58.

- 28 Hur Soo (Hō Su), “Net'ūwōk'ū punsōk-ūl t'onghae pon 1980 nyōndae ‘minjung’: *Tonga Ilbo* (*Dong-A Ilbo*)-ūi yongnye-rūl chungshimūro,” (Discussion of minjung by the citizens of modern Korean society in the 1980s: a network analysis of co-occurring words using the Dong-A ilbo corpus from 1975 to 1994) *Kaenyōm-gwa Sot'ong* (Concept and Communication) 18 (December 2016).

The Role and the Necessity of Research Historians

1. Areas that require value judgments

Collecting a vast amount of information that reflects the state of the times is an important task in historical studies. This is equivalent to the data collection phase in the big data domain. But ‘data collection’ in historical studies does not mean simply collecting all and any kind of data. The following issues need to be considered in collecting data from our times for future research.

First, there is the issue of filtering useful information from the vast amount of data known as big data. Numerous data is constantly generated, but many contain incorrect information. The surge of fake news, which has become a big social issue today, and manipulated images and videos shared by people are highly likely to become sources that result in writing an incorrect history. Crosschecking and critical examination of historical sources are the basic tenets of history writing, so people may not consider the flood of fake news as too much of a problem. However, if the volume of incorrect data overshadows that of correct data, assessing the information to identify the correct ones will become a much more complicated issue. Artificial intelligence may be able to help in this matter, but ultimately it is up to people to make the initial judgment on whether the given information has been manipulated or misrepresented. In terms of historical sources, critical examination is an area that is in dire need of researchers with expert training, and the importance of historical researchers will increase in this aspect.

Once information that has not been manipulated is collected, it needs to be organized, since not all unmanipulated data is meaningful or useful. It has become possible to record all information about human activities, but it is important to make value judgments on whether they are useful for history writing or whether they are meaningless garbage. Value judgments need to be made by humans—and by research historians specifically, for the purpose of history writing. Historians have been trained in

identifying useful information for explaining the spirit of the times from the vast amount of sources and numerous past incidents. This kind of expertise should be applied to collecting and organizing big data, as well as in prioritizing the sources in terms of their usefulness in describing the state of our times today.

Another important issue is that of privacy. Privacy has come to the fore with rapid digitization in many different parts of our lives, and it will become a big issue when personal records on social media and other digital services are collected and turned into big data. Companies that provide social media platforms and services have already faced problems of leaking personal information. In the past, personal information only consisted of the necessary information collected from a user when signing up for a service, but now it has come to mean all data produced by a user, including user activities and information about his private life, as well as posts uploaded by the user. All of this information will be included when collecting data for the purpose of historical research. Therefore once we identify what kind of data to collect, we need to check for possible problems and resolve them.

The results of a study on the consumer awareness of the use of personal information show that people disagree with the use of personal information without consent. Therefore it is essential to give people an option to decide on sharing their personal information, while those who wish to use people's personal information must make efforts to clearly explain the process and the scope in which personal information will be used.²⁹ Particularly, when collecting or using posts that have been uploaded to online communities or on social media, it is important to persuade people that the information they have created is not simply personal, as they are valuable "collective personal records of average people." Emails used to be

29 Kim Inhye and Yŏ Chŏngsŏng, "Pik teit'ŏ hwangyŏng-esŏui kaein chŏngbo hwaryong-e taehan sobija inshik," (Consumer perception on personal information and its usage in the big data environment) *Sobijahak yŏn'gu* (Journal of Consumer Studies) 28, no.6 (December 2017): 145.

considered personal records in the past, and therefore not managed, but with the increasing use of emails for business purposes, their importance has grown and are not kept as records. In archival studies, there have been discussions on the management and value of instant messages, voice recordings, text messages, and images shared via mobile phones by individuals or by small groups of people mainly for work purposes but often contain idle chatter.³⁰ It is important to make concerted efforts to ensure that personal records are not simply considered as personal information and become completely inaccessible.

The process of ruminating on the sensitive issue concerning privacy and devising a solution is also a task that research historians and other humans must undertake. Creating guidelines and thinking of ways to persuade people are not tasks in an area that computers and AI should handle. It is up to humans to decide on the standards of classifying useful information and to protect personal privacy.

One other thing that I would like to emphasize is the growing importance of interdisciplinary collaboration. It will be highly likely to discover different possibilities when research historians work with experts in other academic fields when collecting and filtering useful information. Research historians can work with archivists and law professionals concerning privacy issues in many different areas. More and more research methodologies are requiring interdisciplinary efforts instead of the efforts of individual research historians.

2. Research historians' insight and historical studies

Why do we need research historians in the process of analyzing big data, creating a narrative, and writing? The answer is in line with the reason I mentioned regarding the difficulty and the importance of making value judgments—history writing requires research historians' insight. This means that modern historical studies, which emphasizes objectivity and

30 Song Chuhyōng, "Kirok kwalli taesang-ūrosō SNS yōn'gu," 111.

positivity, will lose its significance in the age of the Fourth Revolution and advanced big data technology.

The development of big data technology cannot be discussed without the development of artificial intelligence. Currently, AI technology has advanced to the point of being able to write up economic reports and even a novel. With the right kind of data, articles about sports game results, stocks, and weather forecasts can be written up using AI much faster than by humans. Since these articles tend to have sentences that generally repeat the same patterns and do not require creativity, AI can quickly create articles given the proper data.³¹

AI can analyze a vast amount of data within a much shorter period of time compared to humans and will likely become better than humans in putting data together in a logical manner. This means AI will be better at identifying simple relationships between facts if accumulated big data is provided. If the goal of historical studies is to objectively demonstrate the events of the past, writing history will soon become a task undertaken by artificial intelligence based on big data.

Work that requires creativity, however, such as writing fiction, belongs in the human domain. No one would think that “creation” or “creativity” in historical studies means manipulating and fabricating facts. What I am asserting is that historical research requires human insight that cannot be provided by AI. Collecting and organizing big data are areas that require AI assistance, but conducting research and writing history based on the collected data must be undertaken by humans.

There has been an intense debate among research historians regarding criticisms against modern historical studies for having too much emphasis on objectivity and positivity, and also regarding new methodologies of

31 Yi Chōngyōp, “AI-rŭl t’onghan kŭlssŭgi-wa chakka-ŭi unmyōng: ‘konpuutaga shousetsu o kakuhi’-ŭl chungshimŭro” (A study on creative writing by artificial intelligence and the fate of the author: a case study of *The day a computer writes a novel*), *Hyōndae sosŏl yŏn’gu* (The Journal of Korean Fiction Research) 68 (December 2017)” 105.

history writing. Big data technology might provide a breakthrough in unexpected ways. Connecting the dots and writing history is an activity that is no longer limited to research experts. Today a lot of people are turning objectivity and proof into history.

Personally, I believe that research historians should approach history writing with *betroffenheit*, or *sangsim* (a feeling of shock, consternation, and concern), that Detlev Peukert described.³² Peukert explained that the “experience of Nazism” consists of two aspects, the first of which is the awareness and actions of the people who lived through the times, and the second is the discussion held by the future generation regarding the results of the past people’s actions and the evidence of their unique way of understanding what happened. Academically, these two different experiences must be handled separately in terms of analysis, though not completely severed. If the two are completely separated, history could stop individuals for experiencing *sangshim* and also obscure the fact that our question about Germany’s past is coming from our experiences today.³³ The gist of his argument is in line with that of Yi Yŏngnam’s, who discussed the “writing of a sympathetic history.”³⁴

Feelings of loss or sympathy are different words for insight that only human research historians can have. It may be *sagwan* (史觀), or the view of history, using a historical term. The reason I classified historical stud-

32 Translating Peukert’s “*betroffenheit*” as *sangshim* (a feeling of shock, consternation, and concern), Kim Hagi explains in the translator’s note: “German historians who wished to study and overcome the historical past of Nazism put forth *sangshim*—individuals agonizing over the Nazi past—as the most important mechanism. *Sangshim* prevents people from falling into oblivion and glamorizing the past, helps aggressors reconcile with the victims, and allows victims to forgive the aggressors.” Detlev Peukert, *Nach’i sidae-ŭi ilsangsa* (Volksgenossen und Gemeinschaftsfremde, National Comrades and Community Aliens), trans. Kim Hagi (Seoul: Kaemagowŏn, 2003), 14.

33 Detlev Peukert, *Nach’i sidae-ŭi ilsangsa*, 13-14.

34 Yi Yŏngnam, *P’uk’o-ege yŏksa-ŭi munbŏp-ŭl paeuda* (Learning the grammar of history from Foucault) (Seoul: P’urŭn Yŏksa, 2007), 10-103.

ies as a discipline in the humanities is because it is imperative for research historians to look at history from their own perspectives. We are gradually departing from an age when the expertise of research historians was needed in simple fact checking. In the age of the Fourth Industrial Revolution, ironically “human-ness” seems to be the quality that is required of humans.

Conclusion

Regardless of how we define the Fourth Industrial Revolution, change is already pervading our lives. Technological innovations are being developed across different sectors and making an impact. Academic disciplines are no exception to such an effect, and I decided to write this paper to think about the changes that such technological innovations may bring to the academia.

The most basic tasks in historical studies are to collect and organize historical sources. The advancement of big data technology will greatly transform these tasks, above all in collecting historical sources necessary for writing about our times. Collecting personal and public records, which was performed in previous historical research, will become more expansive in the future. Tasks that used to be undertaken by individual researchers will be performed by big data technology, which will easily collect a vast amount of data. The collected data will include the type of materials that did not exist in the past. Records from newspapers, periodicals, or public institutions as well as private records posted on online communities and social media will also be collected as historical sources. Not only numeric and textual data but also various types of unstructured data, such as images and videos, will be collected and used as a great historical source that enriches the historical writing and understanding of our times. Big data technology will also lead us to new methodologies for writing history, and this will require further research and development by research historians.

Big data technology is advancing alongside AI technology. The role of AI will grow bigger and more pronounced in big data collection and analysis. Therefore the more advanced big data technology becomes, the more people will question the role and the necessity of historical researchers. However, it is important to note that collecting and analyzing big data still requires value judgments. In a deluge of data, historians must distinguish which are true and which are false; they have to filter through worthless data and find useful information. Moreover, researchers must take great pains to think about and persuade people regarding the use of personal information in the fierce controversy over privacy. This is an area that requires expertise and human judgment.

These questions and problems all culminate with one issue: how should history writing change with the development of big data? There are a number of distinctive characteristics of historical studies and history writing, but most people still point to objectivity and positivity as the two important traits and values of historical studies and writing of history. This was the state of historical studies since Ranke. However, the development of big data and artificial intelligence makes us realize that such “objectivity and positivism” should not stop at simply verifying the facts, because the task of using the given information to connect the dots and link factual information will soon be undertaken by AI. Already, AI technology has advanced enough to write simple articles and economic reports. As a result, the development of big data technology emphasizes the need for researchers to provide historical insights, ironically. It is now up to research historians to explore and search for different ways to write history by reaching beyond objectivity and positivism. Of course, historians are not the only ones to undertake this task; researchers in all disciplines must work together to find solutions.

References

Books

1. Chǒng, Pyǒnguk. *Singminji puronyǒlchǒn: mich'in saenggak-i paetsogesǒ naonda* (A book of rebellious people in colonial Korea: insane ideas come from the pit of the stomach). Seoul: Yǒksa Pip'yǒngsa, 2013.
2. Chǒng, Pyǒnguk, and Itagaki, Ryuta, ed. *Ilgirŭl t'onghae pon chǒnt'ong-gwa kŭndae, singminji-wa kukga* (The tradition and the modern, colony and nation seen through journals). Seoul: Somyǒng Ch'ulp'an, 2013.
3. Chǒng, Ujin. *Pik teit'ǒ-rŭl marhada* (About Big Data). Seoul: Kŭlaudŭ Puks, 2013.
4. Chǒng, Yongch'an. *Pik teit'ǒ* (Big Data). Seoul: Kǒmunikeishyǒn Puksŭ, 2013.
5. Kim, Ilhwan, ed. *Kiwǒdŭ, konggiǒ, kŭrigo net'ŭwǒk'ŭ: shinmun pik teit'ǒ-ga poyǒjunŭn kǒt* (Insights into keywords, co-words, and networks). Seoul: Somyǒng Ch'ulp'an, 2017.
6. Marr, Bernard. *Pik teit'ǒ: 4-cha sanǒp hyǒngmyǒng-ŭi ǒnǒ* (Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results). Translated by An Chunu and Ch'oe Chiŭn. Seoul: Hakkojae, 2017.
7. Mayer-Schönberger, Viktor, and Kenneth Cukier. *Pik teit'ǒ-ga mandŭnŭn sesang* (Big Data: A Revolution That Will Transform How We Live, Work, and Think). Translated by Yi Chiyǒn. Seoul: 21 Segi Puks, 2013.
8. Pak, Kabun. *Ilbe-ŭi sasang* (The Ilbe Ideology). Seoul: Owǒr-ŭi Pom, 2013.
9. Pak, Kabun. *Hyǒmo-ŭi mirǒring: hyǒmo-ŭi sidae-wa Megalia sindŭrom parobogi* (Mirroring hate: the age of hate and the Megalia syndrome). Seoul: Pada Ch'ulp'ansa, 2016.
10. Peukert, Detlev. *Nach'i sidae-ŭi ilsangsa* (Volksgenossen und Ge-

meinschaftsfremde, National Comrades and Community Aliens). Translated by Kim Hagi. Seoul: Kaemagowŏn, 2003.

11. Schwab, Klaus. *Klaus Schwab-ŭi che 4-cha sanŏp hyŏngmyŏng* (Klaus Schwab's Fourth Industrial Revolution). Translated by Song Kyŏngjin. Seoul: Saeroun Hyŏnjae, 2016.
12. Yi, Yŏngnam. *P'uk'o-ege yŏksa-ŭi munbŏp-ŭl paeuda* (Learning the grammar of history from Foucault). Seoul: P'urŭn Yŏksa, 2007.

Articles & Thesis

13. Ahn, Hye-Gyoung. (An Hyegyŏng) “P'yŏngt'aek ilgi'-rŭl t'onghae pon ilsaeng ŭirye-wa sokshin.” (The rites of passage and the folk beliefs in the 'Pyeongtaek Diary') *Silch'ŏn minsokhak yon'gu* 18, (August 2011): 167-199.
14. An, Daejin, and Im Chinhŭi. “Che 4-cha sanŏp hyŏngmyŏng kisul-ŭi kirok kwalli chŏgyong pangan.” (Application of Fourth Industrial Revolution technology to records management) *Kirokhak yŏn'gu* (The Korean Journal of Archival Studies) 54, (October 2017): 211-248.
15. Chang, Tŏkjin. “T'ŭwit'ŏ konggan-ŭi Han'guk chŏngch'i: chŏngch'iin net'ŭwŏk'ŭ-wa yugwŏnja net'ŭwŏk'ŭ.” (Korean politics on Twitter: networks of politicians and voters) *Ŏnnon chŏngbo yŏn'gu* (Journal of Communication Research) 48, no.2 (August 2011): 80-107.
16. Chin, Chuyŏng. “Kukka kirokwŏn websait'ŭ-ŭi pik teit'ŏ punsŏk-gwa hwaryong.” (Big data analysis and use on the National Archives of Korea website) MA thesis, Myongji University, 2018.
17. Cho, Pyŏngch'ŏl, and Yuk Hyŏnsŭng. “Pik teit'ŏ sidae munhwajŏk kiŏk pojŏnsorosŏŭi yŏngsang akaibŭ-ŭi yŏkhal.” (The role of archive as a cultural storage in the age of big data) *Tijit'ŏl yungbokhap yŏn'gu* (Journal of Digital Convergence) 12, no.2 (December 2014): 1-10.
18. Chŏng, Pyŏnguk. “1919 Samil undong-gwa ilgi charyo.” (The

- March 1 movement in 1919 and journals) *Han'guk sahakbo* (The Journal for the Studies of Korean History) 73, (November 2018): 203-236.
19. Ch'oe, Hüisu. "Tijit'öl inmuñhak-üi hyöñhwang-gwa kwaje." (The status and challenges of the digital humanities) *Sot'ong-gwa inmuñhak* (Communication and Humanities) 13, (August 2011): 67-95.
 20. Han, Yunhyöng. "Han'guk chwaup'a t'ujaeng-üi hüürim sogesö 'Ilbe'-rül paraboda: Ilbe-nün kijon-ül chwaup'a-wa öttöke talmatko, tto tarün'ga." (Looing at Ilbe in the flow of struggle between Korean leftists and rightists: how is Ilbe similar and different from the Korean leftists and rightists?) *Chinbo p'yöngnon* (The Radical Review) 57, (September 2013): 14-32.
 21. Hong, Chuhyön, and Yi Ch'anghyön. "T'üwit'ö-esö hyöngsöngdoen chöngch'ijök üigyön punsök-ül t'onghan punhwadoen kongjung yön'gu: 10·26 Söul sijang chaebogwöl söñ'gö-rül chungshimüro." (A study of the public typology based on the analysis of political opinions on Twitter: focusing on the October 26 by-election for the mayor of Seoul) *Han'guk öñnon chöngbo hakbo* (Korean Journal of Communication and Information) 59, (August 2012): 138-161.
 22. Hur, Soo. (Hö Su) "Net'üwök'ü punsök-ül t'onghae pon 1980 nyöndaie 'minjung': *Tonga Ilbo* (*Dong-A Ilbo*)-üi yongnye-rül chungshimüro." (Discussion of minjung by the citizens of modern Korean society in the 1980s: a network analysis of co-occurring words using the Dong-A ilbo corpus from 1975 to 1994) *Kaenyöm-gwa Sot'ong* (Concept and Communication) 18, (December 2016): 53-95.
 23. Ki, Hakjun. "Int'önet kömunit'i Ilbe Chöjangso-esö nat'ananün hyömo-wa yölgwang-üi kamjöngdonghak." (Dynamics of cyber hate and effervescence: focusing on the Korean internet community Ilbe Chöjangso) MA thesis, Seoul National University, 2014.
 24. Kim, Hyön. "Tijit'öl inmuñhak—inmuñhak-gwa munhwa k'ont'ençh'ü-üi sangsaeng kudo-e kwanhan kusang." (Digital humanities—a cooperative scheme between the humanities and cultural

- contents) *Inmun k'ont'ench'ü* (Humanities contents) 29, (June 2013): 9-26.
25. Kim, Inhye, and Yō Chōngsōng. “Pik teit’ō hwangyōng-esōūi kaein chōngbo hwaryong-e taehan sobija inshik.” (Consumer perception on personal information and its usage in the big data environment) *Sobijahak yōn’gu* (Journal of Consumer Studies) 28, no.6 (December 2017): 129-148.
 26. Kim, Paro. “Haeoe tijit’ōl innumhak tonghyang.” (Trends in digital humanities) *Inmun k'ont'ench'ü* (Humanities contents) 33, (June 2014): 229-254.
 27. Kim, Tongyun. “P’ūrangsū ‘tijit’ōl innumhak’-ūi innumhakjōk maengnak-gwa tonghyang.” (Digital humanities in France from the perspective of human sciences) *Inmun k'ont'ench'ü* (Humanities contents) 34, (September 2014): 9-27.
 28. Kim, Young-Mi. (Kim Yōngmi) “‘P’yōngtaek taegok ilgi’-rül t’onghaesō pon 1960-70 nyōndae ch’o nongch’on maül-ūi kongnon-jang, tonghoe-wa mashilbang.” (Village office and *masilbang*, the public opinion sphere in Korean villages in early 1960s and 1970s) *Han’guksa yōn’gu* (The Journal of Korean History) 161, (June 2013): 383-416.
 29. Pak, Ch’angsik. “Chōngch’ijok sot’ong-ūi saeroun chōnmang: 20-30 tae yōsōngdūrūi onlain chōngch’i k’ōmunit’i-rül chungshimūro.” (New prospects of political communication: focusing on political online communities for women in their 20s and 30s) PhD diss., Kwangwoon University, 2010.
 30. Pak, Chiwōn, and Pak Hanu. “P’eisūbuk p’aenp’eiji-ūi tongshi taetgūl teit’ō-rül iyonghan net’ūwōk’ū punsōk: Taegu Kyōngbuk yuryōk huboja-rül chungshimūro.” (Social Network Analysis among Facebook Fanpage Co-commenters: Daegu-Gyeongbuk’s Mayor-Governor Candidates) *Journal of the Korean Data Analysis Society* 16, no.6 (December 2014): 3243-3253.
 31. Pak, Chiyōng, and Pak Hanu. “Ūimimang punsōk-ül t’onghan P’eisūbuk taejung yōron-ūi yōkdongsōng punsōk: Sōul kyoyukgam

- sön'gō-rül chungshimūro.” (An Exploratory Semantic Analysis in the Dynamics of Public Opinion on Facebook: A Case of the Superintendent of Seoul Office of Education in South Korea) *Journal of the Korean Data Analysis Society* 17, no.3 (June 2015): 1557-1571.
32. Pak, Pyōngwōn. “Introduction: ingong chinūng, robot, pik teit'ō-wa che 4-cha sanōp hyōngmyōng.” (Introduction: AI, robotics, Big Data, and the Fourth Industrial Revolution) *Future Horizon* 28, (February 2016): 4-5.
33. Ryu, Ch'ōlgyun, and Yi Chuhūi. “T'ūwit'ō-rül t'onghan sön'gō huboja-ūi sūt'orit'eling punsōk: 4·27 chaebogwōl sön'gō kigan-ūi Ch'oe Munsun, Ōm Kiyōng hubo-ūi t'ūwis-ūl chungshimūro.” (A storytelling analysis of election candidates via Twitter: focusing on the Twitter accounts of Ch'oe Munsun and Ōm Kiyōng up to the April 27 by-election) *Inmun k'ōnt'ench'ū* (Humanities Contents) 23, (December 2011): 237-261.
34. Song, Chuhyōng. “Kirok kwalli taesang-ūrosō SNS yōn'gu.” (A study on records management by social media services) *Kirokhak yōn'gu* (The Korean Journal of Archival Studies) 39, (January 2014): 101-138.
35. Yang, Yunju. “1960-1970 nyōndae nongmin-ūi ilsangsaenghwal-gwa chōngch'esōng-ūi pyōnhwa.” (The daily lives of farmers in the 1960s and 1970s Korea and their identity change: a study of farmers' journals) MA thesis, Kookmin University, 2017.
36. Yi, Chōngyōp. “AI-rül t'onghan kŭlssūgi-wa chakka-ūi unmyōng: 'konpuutaga shousetsu o kakuhi'-ūl chungshimūro.” (A study on creative writing by artificial intelligence and the fate of the author: a case study of *The day a computer writes a novel*) *Hyōndae sosōl yōn'gu* (The Journal of Korean Fiction Research) 68, (December 2017): 103-131.
37. Yun, Ch'ōl. “Pik teit'ō-e kibanhan kirok chōngbo sōbisū-ūi panghyang.” (Archival reference services based on big data) MA thesis, Hanshin University, 2013.
38. Yun, Pora. “Ilbe-wa yōsōng hyōmo: Ilbe-nūn ōdiena itko ōdiedo

õpda.” (Ilbe and misogyny: Ilbe is everywhere yet nowhere) *Chinbo p'yõngnon* (The Radical Review) 57, (September 2013): 33-56.

〈Abstract〉

Big Data and the Prospects of Historical Research **- A study of research in modern and contemporary Korean history -**

Minki Moon

The most basic tasks in historical studies are to collect and organize historical sources. The advancement of big data technology will greatly transform these tasks, above all in collecting historical sources necessary for writing about our times. The collected data will include the type of materials that did not exist in the past. And it will be collected and used as a great historical source that enriches the historical writing and understanding of our times. Big data technology will also lead us to new methodologies for writing history, and this will require further research and development by research historians.

Big data technology is advancing alongside AI technology. The role of AI will grow bigger and more pronounced in big data collection and analysis. However, it is important to note that collecting and analyzing big data still requires value judgments. In a deluge of data, historians must distinguish which are true and which are false; they have to filter through worthless data and find useful information. Moreover, researchers must take great pains to think about and persuade people regarding the use of personal information in the fierce controversy over privacy. This is an area that requires expertise and human judgment.

The development of big data and artificial intelligence makes us realize that such “objectivity and positivism” should not stop at simply verifying the facts. The development of big data technology emphasizes the need for researchers to provide historical insights, ironically. It is now up to research historians to explore and search for different ways to write history by reaching beyond objectivity and positivism. Of course, historians are not the only ones to undertake this task; re-

searchers in all disciplines must work together to find solutions.

Keyword: big data, artificial intelligence(AI), Fourth Industrial Revolution, historical research, historian

〈국문초록〉

빅데이터와 역사학 연구의 전망

문민기 (고려대학교 박사수료)

역사학 연구에서 가장 기본이 되는 것은 사료를 수집하고 분류하는 작업이다. 빅데이터 기술의 발달은 이 작업에 많은 변화를 가져오게 될 것이다. 무엇보다 당대사 서술에 필요한 사료들을 수집하는 데 큰 역할을 할 것으로 기대된다. 이때 수집되는 기록/데이터의 종류는 기존에 생각지 못했던 것들을 포함하게 되며, 이는 당대사의 서술을 더욱 풍부하게 만들어주는 훌륭한 사료가 될 것이다. 빅데이터 기술의 활용은 기존의 역사 쓰기 방식이 아닌 새로운 역사 쓰기 방법론을 가져다 줄 수도 있을 것인데, 이에 대해서는 앞으로 역사학 연구자들의 고민과 탐구가 더욱 필요하다.

빅데이터 기술은 인공지능 기술과 함께 발전해 나간다. 빅데이터의 수집과 분석에서도 인공지능의 역할은 점차 커지고 강조될 것이 틀림없다. 하지만 빅데이터 수집과 분석에 있어서 가치판단이 필요한 지점이 여전히 많다. 범람하는 데이터 속에서 그 진위 여부를 판단해야 하며, ‘쓰레기 데이터’ 속에서 유용한 정보를 찾아내야 한다. 또한 개인정보 보호와 관련된 첨예한 논쟁은 앞으로 관련 연구자들이 더 고민하고, 많은 이들을 설득하는 과정을 거쳐야 한다. 이는 전문성이 요구되는 지점이며, 아직 인간의 영역이라고 평가될 수 있는 부분이다.

빅데이터와 인공지능의 발달은 역사학의 ‘객관과 실증’이 단순한 사실 관계의 확인 수준에서 그쳐서는 안 된다는 것을 깨닫게 만든다. 그렇기에 빅데이터 기술의 발달은 역설적이게도 역사학 연구자가 역사적 통찰력을 발휘해야 할 필요성을 강조한다. 객관과 실증을 넘어서는 역사 쓰기를 위한 고민과 탐구 또한 역사학 연구자들의 몫이다. 물론 이것이 역사학 연구자만의 과제는 아니며, 다른 분과의 연구자들과 적극적으로 협업하면서 고민하고 해결해 나가야 할 과제이다.

주제어: 빅 데이터, 인공지능, 4차 산업혁명, 역사 연구, 역사가

