

# History and Trends of Data Education in Korea

## - KISTI Data Education Based on 2001-2019 Statistics

Jaehong Min<sup>1</sup>      Sunggeun Han<sup>2</sup>      Bu-young Ahn<sup>3\*</sup>

### ABSTRACT

Big data, artificial intelligence (AI), and machine learning are keywords that represent the Fourth industrial Revolution. In addition, as the development of science and technology, the Korean government, public institutions and industries want professionals who can collect, analyze, utilize and predict data. This means that data analysis and utilization education become more important. Education on data analysis and utilization is increasing with trends in other academy. However, it is true that not many academy run long-term and systematic education. Korea Institute of Science and Technology Information (KISTI) is a data ecosystem hub and one of its performance missions has been providing data utilization and analysis education to meet the needs of industries, institutions and governments since 1966.

In this study, KISTI's data education was analyzed using the number of curriculum trainees per year from 2001 to 2019. With this data, the change of interest in education in information and data field was analyzed by reflecting social and historical situations. And we identified the characteristics of KISTI and trainees. It means that the identity, characteristics, infrastructure, and resources of the institution have a greater impact on the trainees' interest of data-use education. In particular, KISTI, as a research institute, conducts research in various fields, including bio, weather, traffic, disaster and so on. And it has various research data in science and technology field.

The purpose of this study can provide direction for the establishment of new curriculum using data that can represent KISTI's strengths and identity. One of the conclusions of this paper would be KISTI's greatest advantages if it could be used in education to analyze and visualize many research data. Finally, through this study, it can expect that KISTI will be able to present a new direction for designing data curricula with quality education that can fulfill its role and responsibilities and highlight its strengths.

☞ Keyword : Data education, trainee, educational trends, Data usage and analysis, data

## 1. Introduction

As a public institution, Korea Institute of Science and Technology Information (KISTI) is making great efforts to apply and use up-to-date scientific and technological knowledge. To actively expand its infrastructures through HPC (high performance computing), data, and AI, KISTI is providing the right education to meet specific targets, such as specialized education using data and supercomputing for

researchers. For college students, KISTI has developed a platform called Education-research Integration through Simulation on the Net (EDISON), through which it provides science computing simulation and education. [1]

With the emergence of the Third Industrial Revolution and information age and the advancement of information and communications in the 1960s, KISTI imparted education with keywords, such as "informatization" and "information industry" according to the government policy, anticipating citizen participation in accessing, analyzing, and predicting any information that they want.

Since 2001, the statistical chart of KISTI's curriculums on data use shows examples, such as STN The Scientific & Technical Information Network, database storage repository for science and technology industries as well as patents Information Search, Basics of Information Search, and Patent Information Search. As the Fourth Industrial Revolution emerges, there has been emphasis on keywords, such as data analysis, AI based on big data, machine learning, and

<sup>1,2,3</sup> Dept. of science data school, Korea Institute of Science and Technology Information(KISTI), Daejeon, 34141, Korea

\* Corresponding author: Bu-young Ahn (ahnyoung@kisti.re.kr)

[Received 27 May 2020, Reviewed 15 June 2020(R2 31 August 2020), Accepted 9 September 2020]

☆ This work was supported by KISTI Program(K-20-L05-C01-S01) and the National Research Foundation of Korea(NRF) and the Center for Women In Science, Engineering and Technology (WISSET) Grant funded by the Ministry of Science and ICT(MSIT) under the Program for Returners into R&D in 2020 (WISSET-2020-168).

☆ A preliminary version of this paper was presented at ICONI 2019.

statistical forecasting. In particular, the World Economic Forum selected education as one of the fields that can bring about a rapid reform in the Fourth Industrial Revolution. And it implies that the need for education on data use is emphasized worldwide. Korea also requires all kinds of support from public institutions, such as establishing a futuristic education system and nurturing specialists in science and technology, aligned with the education and job policy. [2] In addition, there are changes in KISTI's data curriculums, such as newly opened courses, changed curriculum titles, or cancellations. This study analyzes the cumulative data on participants of each curriculum from 2001 to 2019, determines the interest in the major curriculums, and forecast the direction for data education in the future. Moreover, as a government-funded researcher, a new direction can be suggested in establishing a new curriculum that can fulfill the role and responsibility of KISTI.

## 2. Research methodology

For analysis, we summarized the current state of trainees in each curriculum educated by KISTI from 2001 to 2019. Participants of data education are mostly workers in the relevant industries and researchers, and each session of the program can accommodate up to 30 participants per lecture

hall. Table 1 summarizes the data use curriculums implemented for at least 10 years since 2001. These programs are offered twice a year.

From the Table 1, the shift of interest in education can be analyzed by reflecting social and historical situations. For this study, Google Trend was used to compare and analyze changes in Koreans' interest in the data education from 2006 to 2019. Google Trend was launched in 2006 and is a platform that analyzes the popularity of top search frequency in Google Search across the world. In addition, it provides suitable information according to the needs and interests of people in many fields such as IT, business, sport, education and so on. Above all, Google Trends has been accumulating a wide range of search terms for a long period from 2006 to now, and it is widely used in big data research because various search terms can be compared on this platform.[3, 4]

## 3. Results

By analyzing the records of data education for 20 years as shown in Table 1, we could summarize the main characteristics of data education in Korea as follows.

(Table 1) (KISTI data education curriculum statistics)

Curriculum	No. of participants	Year of opening	Operating period (years)	Note
Patent Information Search	990	'01-'18	18	Combined
STN Information Search	213	'01-'05	5	
Basics of Information Search	566	'03-'19	18	
Patent Map Expert	753	'02-'19	18	
Analysis of Industrial Market Research	940	'03-'19	17	
Technology Road Mapping	1,080	'03-'19	17	
R&D Planning, Management & Evaluation	954	'04-'19	16	
Future Technology Forecast	172	'09-'19	11	
Technology Valuation	177	'08-'19	10	
Python	119	'19	Newly opened	
Data Management Plan	19	'19	Newly opened	
Data Analysis with R	16	'19	Newly opened	
Deep Learning	7	'19	Newly opened	

### 3.1 Change in concept from information to data

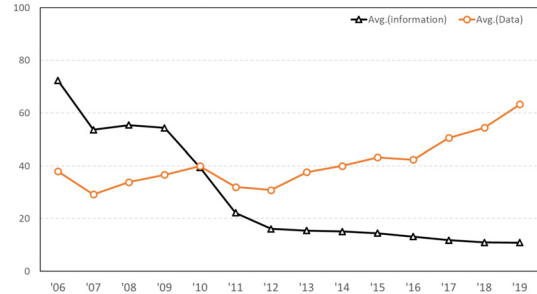
In the early 2000s, with the widespread use of the internet, the keywords of the information age were "informatization," "information search," or "information use". Currently, in the age of the Fourth Industrial Revolution, there are many keywords on big data and data. Information refers to defined and processed data, whereas data is raw and unprocessed. This indicates that the beginning of the Fourth Industrial Revolution naturally increased people's interests in original data as well as data processing, analysis, and visualization. In database, information and data are defined as follows.

**Information:** Output of processing and systematically organizing data so that it can be used in decision making.  
**Data:** Unprocessed fact or value merely observed, measured, or collected in the real world [5].

At this time there were changes in KISTI education. From 2001 to around 2010, most education programs were on "information" that was already defined, such as Basics of Information Search, Technology Roadmap, Practical Affairs of Technology Contracts, Patent Information Search, Analysis of Industrial Market Research, etc. Above all, the demand was highest for the curriculums in Table 1, as well as "STN Information Search" and "Basics of Information Search." However, as people adapted to the information age, "STN Information Search" class was closed due to low enrollment rates, and the curriculum for Basics of Information Search was changed to educating for "research data usage" on analyzing and using big data and even changing the title from information to data.

These changes also appear similarly in Korea society as well. Figure 1 is a graph using data from Google Trends, and it shows the changing trend of search keywords "information" and "data." The change is clear when the annual average number of searches for the two keywords are compared for 2006 and 2019. For 2010, search volume of information and data intersected. This shows that the conceptual change regarding information and data also appeared in the Korean society in general. It means that before 2010, around the time of the Fourth Industrial Revolution, people were interested in collecting

information defined by individual purpose. And also this is similar to the point of change for the Third and Fourth Industrial Revolution, as defined in the World Economic Forum.[6]



(Figure 1) (Annual average search volume of "information" and "data" in 2006-2019 by Google Trends)

### 3.2 Change in educational interest due to enhanced competencies of data users

In the early 2000s, there were many education programs on basic computer use and information search due to the supply of computers, ICT, and internet speed. Moreover, with the advancement of technology and the guidelines by the Ministry of Education in 2001, computer education has become mandatory from elementary school.

Accordingly, education on basic computer use has improved computer skills and information literacy of citizens for 10 years.[7] And also it can explain the constant decline of educational interests and trainees in Information search class for 10 years.

On the other hand, there were demands for education on producing and forecasting new outputs by analyzing data. Therefore 'Future Technology Forecast' and 'Technology Valuation' classes opened in KISTI curriculum. This can explain the growing interest in education on practical and operational information-use education of workers. And they have been employed based on the information utilization technology acquired through the school curriculum. Furthermore, data users are interested in using data to produce and predict new data or results. These changes will also affect the important survey for the establishment of

KISTI curriculum.[8]

### 3.3 Increasing interest in the general field of data

Data users' increased level of technology and knowledge also introduced many changes to their fields of interest in education. Recently, they are beginning to show an interest in the general lifecycle of data, such as data management, instead of just focusing on processing data.

In this era of big data, researchers produce various outputs by using public data. They are making plans on production, preservation, management, and utilization of a variety of data from the research stage, which is known as the data management plan (DMP). Sharing and utilization of data will contribute toward development into an innovative field of science. Recently, the Ministry of Science and ICT has begun discussing the full implementation of DMP. In light of this change, KISTI also opened a program on DMP, and many researchers are showing interest in the program. [9-11]

Data users' are also interested in programming languages. With an open data availability, those who analyze large datasets or are interested in AI can access the data easily. Thus, education on programming language, which was the domain for a few experts or academic majors in the past, is now provided to a range of participants including the general

public. Anyone can study programming systematically and professionally, to solve problems in a creative way. Recently, Python is being used as an essential tool in AI. This explains the high demand for Python, ranked No. 1, and R, ranked No. 5, in the results of analyzing the world rankings of programming languages in 2019. [12]

Based on this analysis, we predicted that there was a high demand for education in programming language and data analysis. For this reason, the basic education for Python, R, and deep learning was newly established in KISTI as shown Table 2.

We are also developing a curriculum that can analyze data using programming language such as 'data analysis using R', 'machine learning-based data analysis' and 'statistical data analysis'. Other examples include 'data analysis using Tableau'

### 3.4 Connection between the characteristics of the educational institution and trainees

KISTI, a government-funded research institute, is a public institution for the country and its people. Figure 3 is a graph showing five curriculums run by KISTI. It shows a steady demand for education for about 16 years, which indicates that the characteristics of the institution providing education affect the trainees' choice of curriculums. Regardless of the

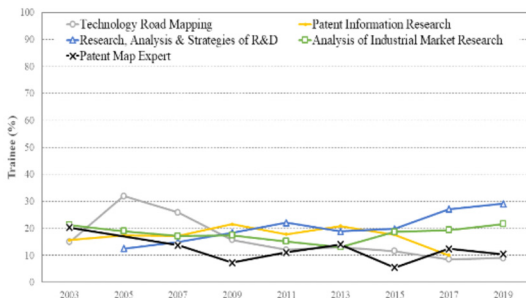
(Table 2) (List of data education newly opened in 2019)

Curriculum	Content
Python	- Learning how to analyze and visualize big data based on Python - Learning the basic skills of using big data
Data Management Plan	- Understanding domestic and international policies related to data sharing - Developing the ability to use the DMP method through cases of data management and sharing plans by country or institution
Data Analysis with R	- Understanding the basic concepts of analysis in mathematics/statistics using R - Developing the ability to analyze data through various cases studies and exercises, such as clustering, correlation analysis, regression analysis, time series analysis, etc.
Deep Learning	- Introducing deep learning techniques and understanding the principles - Conducting a deep learning project to resolve classification and forecast issues by using public data
Data Analysis with Excel	- Developing the ability to deal with data by using Excel - Learning basic statistical methods such as probability distribution, estimation, and testing, and performing analysis exercises

times, keywords such as R&D, market research, and patent have great significance to researchers and workers in the relevant industries.

These results have the following significance. It is necessary to activate education that can utilize the infrastructures (HPC, research data, AI, specialists), which are the strengths of KISTI. This indicates that the institution's identity, characteristics, infrastructures, and resources also have a great impact on the decision to take courses on data education.

KISTI conducts research in various fields such as bio, weather, traffic, and disaster and so on. And KISTI has lots of research data in these fields of science and technology. If these research data can be used for analysis and visualization education, it will be KISTI's greatest advantage. Therefore, data analysis classes using research data from the various field of science and technology will also be a good curriculum for KISTI's role and responsibility.



(Figure 2) (Demand for certain curriculums ('03~'19))

## 4. Conclusions

Analysis of KISTI's data education from 2001 to 2019 showed that there were constant changes in KISTI's data curriculums, driven by the industrial trends and keywords, including newly opened courses, changed curriculum titles, or cancellations. Contrarily, there were continuous demands for practical and operational curriculums since the participants were mostly researchers and workers in the relevant industries.

We determined the trends of data education in Korea based on the results of analysis. Furthermore, we will predict

the characteristics of future data education in Korea and set a new direction for curriculums.

- 1) It is necessary to open education programs on data use in general.

Users have gone past employing analyzed information. They are showing interest in the complete process of data use, such as data collection, analysis, use, and storage. To this end, there must be education on data use in general, a typical example being DMP education. DMP education was opened from 2019 to answer this need.

- 2) It is necessary to provide education and develop curriculum that reflects the knowledge level of users.

A reason that the number of trainees and interest in the basics of information decreased is because of the improved level of data/information-use. Therefore, it is necessary to open curriculum that can enhance expertise rather than a class on basic education on data utilization. The basic level courses can utilize the open platform (MOOC, Massive Open Online Courses) to strengthen basic learning through repetitive classes. It is a way to operate basic courses for data-use education that can be taken at any time without time and space limitations.

- 3) It is necessary to establish curriculums that are compatible with the identity of the institution.

There are many educational institutions offering courses on data. Therefore, each institution must have curriculums that convey its distinctiveness and identity. In particular, KISTI has its strength in using its own infrastructures (data, high performance computing, AI, etc.) in data education. For example, if data utilization education using highly demanding programming languages is provided, it would be good to develop an analytical curriculum using research data in the field of science and technology.

Educational courses on data analysis or use is increasing, but few institutions are offering long-term, systematic programs. KISTI is currently developing and operating

educational program that meets the trend. Since 1996, KISTI has been offering various data-use curriculum and is creating and running systematic and developmental curriculums that meet the needs of firms, institutions, and the government.

In addition, we are trying to respond sensitively and quickly to the changes in the world flow and the interest of trainees. Through this study, it can expect that KISTI will be able to design a data curriculum with high quality of education that can fulfill its role and responsibilities and highlight its strengths.

## References

- [ 1 ] N. R. On et al., "An Analysis of the Factors Affecting User Satisfaction in Computational Science and Engineering Platforms: A Case Study of EDISON", *Journal of Internet Computing and Services (JICS)*, Vol. 20, No.6, pp.85-93, 2019.  
<http://dx.doi.org/10.7472/jksii.2019.20.6.00>
- [ 2 ] J. Kim, "Impacts and Countermeasures of the Fourth Industrial Revolution on the Public Sector", *Industry focus*, Vol. 42, pp. 2-6, 2017.  
Retrieved from [https://www2.deloitte.com/content/dam/Deloitte/kr/Documents/public-sector/2017/kr\\_ps\\_issue-highlights\\_20170327.pdf](https://www2.deloitte.com/content/dam/Deloitte/kr/Documents/public-sector/2017/kr_ps_issue-highlights_20170327.pdf)
- [ 3 ] S. Jun et al., "The possibility of using search traffic information to explore consumer product attitudes and forecast consumer preference", *Technological Forecasting & Social Change*, Vol 86, pp.237-253, 2014. <https://doi.org/10.1016/j.techfore.2013.10.021>
- [ 4 ] S. Jun et al., "Ten years of research change using Google Trends: From the perspective of big data utilizations and applications", *Technological Forecasting & Social Change*, Vol 130, pp.69-87, 2018. <https://doi.org/10.1016/j.techfore.2017.11.009>
- [ 5 ] Kim, Y., "Introduction to database", 6thEd, p.20, Hanbit Academy, Korea, ISBN: 9788998756253, 2013.
- [ 6 ] M. Chung and J. Kim, "The Internet Information and technology Research directions based on the fourth industrial revolution", *KSII Transaction on internet and information systems*, Vol. 10, No.3, pp.1311-1319, 2016. <http://dx.doi.org/10.3837/tiis.2016.03.020>
- [ 7 ] J. Rheem, "Present State of Programming Language Education and Suggestions for Its Improvement", *Korean Institute for practical Engineering education*, Vol. 3,no. 1, pp. 56-61, 2011.  
<https://www.koreascience.or.kr/article/JAKO201123061362601.pdf>
- [ 8 ] Jung, S., "Data analyst curriculum development", pp.50-67, KISTI (Korea Institute of Science and Technology information) research Report, Korea, 2018. <https://doi.org/10.22810/2018KRR039>.
- [ 9 ] Lee, T. Choi H. and Jeon Y., "Informatics Education", 2nd Ed, p.69, Hanbit Academy, Korea, ISBN: 9791156642510, 2016
- [10] Regulations on the Management of State R&D Projects, Ministry of Science and ICT, Korea, 2019. Retrieved from <https://www.law.go.kr/lsInfoP.do?lsiSeq=215595&efYd=20200317#0000>
- [11] Do, J. et al, "Development of Data Science Curriculum based in Data Life cycle", p.22, KISTI (Korea Institute of Science and Technology information) research Report, Korea, 2019. <https://doi.org/10.22810/2019Kacademy002>
- [12] S. Cass, "The Top Programming Languages 2019- Python remains the big kahuna, but specialist languages hold their own", *IEEE Spectrum*, 2019. Retrieved from <https://spectrum.ieee.org/computing/software/the-top-programming-languages-2019>.

● 저 자 소 개 ●



**Jaehong Min**

2014 M.S. in Dept. of Chemical and Biomolecular engineering, Yonsei University, Seoul, Korea  
2017~ present, researcher at the Science data school, Korea Institute of Science and Technology Information (KISTI), Daejeon, 34141, Korea.  
Research Interests : Supercomputer, education, Computer science, Science and technology policy, data  
E-mail: jhmin@kisti.re.kr



**Sunggeun Han**

2000 M.S. in Dept. of Computer Engineering, Chonbuk National University, Jeonju, Korea  
2002~present, senior researcher at the Science data school, Korea Institute of Science and Technology Information (KISTI), Daejeon, 34141, Korea.  
Research Interests : Supercomputer, parallel process, AI, High performance computing, data  
E-mail: sgghan@kisti.re.kr



**Bu-young Ahn**

2008 Ph.D. in Library and Information Science Chungnam National University, Daejeon, Korea.  
1982~ present, senior researcher at the Science data school, Korea Institute of Science and Technology Information (KISTI), Daejeon, 34141, Korea.  
Research Interests : Supercomputer, Computational science and engineering, Education, data, High performance computing, Science and technology policy  
E-mail: ahnyoung@kisti.re.kr