

# 한국인 연구자의 논문 데이터 분석과 활용

## Analysis and Usage of Research Data of Korean Researchers

최원준, 김재훈, 김정환  
한국과학기술정보연구원 정보기반실

Wonjun Choi(cwj@kisti.re.kr), Jayhoon Kim(jay.kim@kisti.re.kr),  
Jeonghwan Kim(kimjh@kisti.re.kr)

### 요약

국내뿐만 아니라 전 세계적으로 과학기술정보 데이터는 꾸준히 증가하고 있다. 분야에 관계없이 하루에도 많은 양의 문헌이 만들어지고 배포되는 현 시점에서 연구자들이 이러한 문헌 정보를 효과적으로 사용하여 가치 있고 인간 생활에 유용한 연구 활동을 지원하기 위한 요구가 늘어나고 있다. 과거에는 연구자가 필요한 정보를 검색하는 수준에서 벗어나 이제는 데이터와 데이터를 연결시키고 활용 가능한 고부가가치 정보를 찾아내어 연구자에게 제시하는 서비스로 점점 변화하고 있다. 본 연구에서는 국내외의 과학기술정보 데이터 중에서 한국인 연구자가 작성한 문헌 데이터를 분석하여 활용 사례를 알아보고, 연구자들을 위한 서비스에 도움을 주는 정보를 제시하고자 한다.

■ 중심어 : | 과학기술정보 | 논문 데이터 | 학술논문 | 문헌 정보 | 정보 서비스 |

### Abstract

Science and technology information data is steadily increasing not only in Korea but also around the world. At present, dozens of documents are produced and distributed every day regardless of field, and researchers are increasingly demanding to use this information effectively to support research activities that are valuable and useful for human life. In the past, researchers are moving away from searching for necessary information, and now they are changing to services that connect data and data, find high value-added information that can be used, and present it to researchers. In this study, we analyze literature data created by Korean researchers from domestic and foreign scientific and technological information data, find use cases, and provide information to help researchers.

■ keyword : | Science Technology Information | Research Data | Academic Paper | Literature Information | Information Service |

## 1. 서론

국내뿐만 아니라 해외에서도 과학기술정보 데이터가 지속적으로 증가하고 있고 학술정보 유통을 위한 과학

기술정보 데이터 서비스 사업에 대한 지원이 확대되고 있다. 구글 학술 검색이나 Web of Science, Scopus 같은 서비스는 사용자들이 그 자체가 전자 정보원은 아니지만 하나의 데이터베이스로 인지하고 있다[1]. 특정 주

\* 본 연구는 한국과학기술연구원 연구과제로 수행되었습니다.

접수일자 : 2017년 10월 20일

수정일자 : 2017년 11월 13일

심사완료일 : 2017년 11월 13일

교신저자 : 김정환, e-mail : kimjh@kisti.re.kr

제 분야에서는 Web of Science보다 구글 학술 검색에서 더 많은 자료 양을 보여 주었으며, 많은 대학에서 링크리졸버에 구글 스칼라를 연결시켜 온라인 데이터베이스와의 연동 서비스를 지원하고 있다. 구글 학술 검색의 경우 학술논문과 관련된 정보를 제공하고 원문 전체를 검색할 수 있는 서비스를 지원하고 있다. 또한, 중복된 논문을 제거하는데 구글 학술 논문 검색이 매우 성공적임을 확인할 수 있다[2]. 하지만, 시소러스, 주제어 같은 통제어휘가 없는 것은 복잡한 검색이 요구되는 분야에서는 아쉬운 부분으로 지적된다. 많은 양의 데이터를 서비스하고는 있지만 구글이 지불수단까지 제공하고 도서관과 출판사 사이의 중간 매개자 역할을 하게 되면 거래의 단순화를 지원할지라도 독점화에 대한 우려가 존재한다.

현재는 데이터 서비스 패러다임이 서서히 변화하고 있다. 고부가가치 서비스를 창출하기 위해 데이터를 네트워크로 링크시키고 분석하여 가치 있는 서비스로의 영역으로 확대가 요구되고 있다. 2017년 9월 1일에 서비스를 오픈한 네이버 학술정보 서비스는 15억 건의 참고문헌을 101개의 정보원으로부터 자료를 입수하여 서비스를 하고 있다. 해당 분야 별, 연도 별 피인용 현황, 많이 참조한 문헌, 분야별 주요 연구 키워드, 등을 서비스함으로써 과거의 기본적인 지식 검색 서비스에서 벗어나 좀 더 차별화된 분석 서비스를 시작했다. 이와 같이 과학기술정보 데이터는 갈수록 증가함에 따라서 다양한 서비스 플랫폼이 나오는 시점에서 수요자 맞춤형 차별화된 서비스의 필요성이 생겨나기 시작했다. 분야별 분석 서비스가 그 중에 하나라고 볼 수 있으며, 데이터 자체의 신뢰성 또한 중요해 지고 있다. 데이터가 양적으로 많다 하더라도 업데이트와 신뢰성 검사가 지속적으로 수행되지 않으면 분석 서비스를 설계한다고 하더라도 제대로 된 결과 값을 얻을 수 없다. 본 논문에서는 신뢰성이 어느 정도 입증된 국내 연구자 데이터베이스를 토대로 국내 과학기술정보 데이터베이스에서 한국인 연구자의 성과물을 구분하고 새로운 차원의 논문 서비스를 설계하고 활용하기 위한 분석을 시도해 보고자 한다. 분야 별로 트렌드를 확인하고 미래를 예측하여 서비스하는 방식은 앞으로 필수적이 될 것이다. 한

국인 연구자 데이터베이스를 선정한 이유는 NTIS[3] 또는 연구재단 등에서 한국인 연구자 데이터를 가지고 있지만 한국인 연구자 전체를 대변하는 데이터베이스가 아니기 때문에 관련 자료와 비교 분석하는 것은 의미 있는 작업이 될 것이다. 또한, 한국인 연구자가 해외에 논문을 투고한 부분과 같이 구축하여 분석할 필요가 있다. NDSL[3]에서는 어느 정도 한국인 연구자가 해외에 투고한 논문을 확인할 수 있다. 따라서 NDSL[3] 데이터베이스에서 한국인 연구자 데이터를 구분하고 분석한 결과를 토대로 새로운 서비스를 위한 활용 방안을 알아보는 것은 의미가 있다. 본 논문에서는 국내 과학기술정보 데이터를 분석하여 활용 방안을 연구해 보고자 한다.

## II. 선행 연구

국내의 과학기술정보의 한 주축을 이루고 서비스하고 있는 NDSL[3] 데이터를 통해서 과거와 현재에 많은 연구가 이루어졌고 진행되고 있다. 그 예로 NDSL에 있는 과학기술정보 데이터를 가지고 연구자들의 정보요구를 파악하고 정보요구별 탐색행태를 분석한 논문[4], 연구자가 필요한 정보를 적시에, 적절한 경로를 통해서 공급하는 것이 곧 연구 활동의 경쟁력으로 평가하는 논문[5] 이 있는데 주로 수요기반의 서비스를 개발하기 위한 연구이다. 이밖에도 국내 대표 학술지를 중심으로 기록 관리학 연구자들의 학술 논문과 그 논문의 참고문헌을 조사, 분석하여 연구자들의 인용 행태와 특성 및 변화를 제시한 논문[6]이 있으며, 논문[7]의 경우에는 국내 문헌정보학 연구자들의 인용행태를 분석하여 어떤 학술지가 많이 인용이 되는지 에 대한 연구를 수행하였다. 논문[8]의 경우에는 학문 분야별 연구자들이 정보를 생산하기 위해 활용하는 정보원을 비교, 분석하여 분야별 선호도 차이를 비교하였다. 현재까지의 대부분의 국내 연구자 분석 논문이 주로 국내 대표 학술지를 참조했지만 한국인이 해외에 투고한 학술지와 한국인 논문을 인용한 해외 논문에 대한 학술지에 대한 정보가 아쉬운 부분으로 남아있다. 이 외에도 신진 연구자의

연구 환경이 향후의 연구 활동에 중요한 영향을 미치는지에 대한 연구[9], 학술 메타데이터의 저자 인용 분석과 연구 유사도 산출 및 연구자간의 관계분석 시스템 연구[10], 연구자들의 연구기획, 협업 연구자 선정 및 특정 분야의 전문가에 대한 정보 서비스에 관한 연구[11], 분야 별 연계도 분석 연구[12], 자기 인용 네트워크와 연구 이력을 분석한 연구[13], 정부의 연구비 지원이 대학 연구자의 논문 성과에 얼마나 영향을 미치는지에 대한 연구[14], 연구자의 정보이용행태에 분석을 활용한 효과적인 정보서비스 기초자료 연구[15], 참고문헌 분석을 통한 국방과학 연구자들의 인용행태 연구[16], 한국 문학 번역원과 한국, 미국, 일본의 국립도서관에서 한국인 저자 전거의 현황과 문제점에 관한 연구[17], 국내 행정학 분야의 연구자들 간에 연구수행 경향을 분석한 연구[18], 문헌정보학 분야 국제 학술지에 논문을 게재한 연구자의 소속기관 유형 및 전공분야, 기여도 분석에 관한 연구[19], 동일한 학문분야를 전공하는 연구자들의 연구행태를 분석하여 그 분야의 연구자들이 주로 이용하는 학술지들이 어떤 것들이 있는가를 조사한 연구[20] 등이 있다. 이와 같이 새로운 서비스와 동향을 파악하고 새로운 서비스를 설계하기 위해 기존의 과학기술 학술 정보 데이터를 분석했다는 점은 모두가 동일한 성격을 지니고 있다. 문헌정보학, 국방, 기록관리학 등의 다양한 학문 영역에서도 연구 행태를 확인하고자 하는 노력은 지속되고 있다는 것을 확인하였다. 본 논문에서는 국가과학기술정보 데이터베이스(NDSL[3])를 기반으로 식별된 한국인 연구자 데이터를 분석하여 고부가가치 정보를 제공하고 새로운 서비스 연구 개발을 위한 발판을 마련하고자 한다.

### III. 한국인 연구자 데이터 분석

한국인 논문 데이터 범위는 NDSL[3] 데이터에 있는 한국인 연구자 논문 식별 데이터를 참고하였다. 식별된 데이터의 정확도는 약 95.7% 정도의 신뢰 가능한 데이터를 선정하여 분석하였다. 식별 기준은 메일 주소, 공동 저자, 소속 명, 저널 명, 저자 키워드를 중심으로 식별되었다. 정확률 검증 방식은 사용자 참여 형 방식으

로 체크되었다. 식별된 데이터의 양은 약 1,118,545(9/19 기준)건으로 집계된 데이터를 분석하였다.

#### 1. 연도 별 한국인 연구자의 논문 인용 빈도 수

본 챕터에서는 연도 별로 한국인 연구자의 논문이 얼마나 인용이 되는지를 알아보기 위해 1998년도부터 2014년도까지 NDSL에서 수집한 데이터 중에 한국인 연구자로 식별이 된 데이터를 확인하였다. 주로 한국인 연구자가 국내 논문지에 출판한 논문은 언어가 한글로 작성이 되어 있기 때문에 국외에서 인용율이 높지 않아 한국인 연구자가 국외 논문지에 출판한 논문을 중심으로 통계 수치를 뽑아 보았다. 주로 2005년과 2010년도에 한국인 연구자의 논문이 인용이 많이 된 것으로 확인되었다[그림 1].

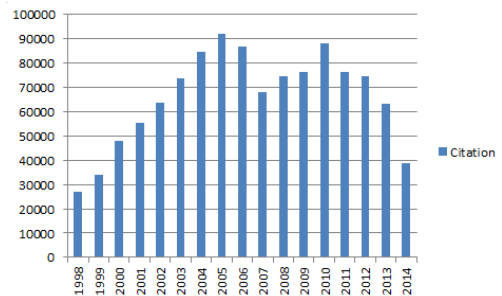


그림 1. 연도 별 한국인 논문의 인용 빈도 수

#### 2. 주제 별 한국인 연구자의 논문 수 비교

이번 챕터에서는 국내 논문 약 111만 건에 대해서 한국인 연구자의 논문 중 어느 분야가 많은 논문 수를 포함하는지 알아보기 위해 주제 분야가 정확한 정보 45만 건에 대한 논문 정보에서 논문 수가 7000건 이상인 분야를 산출했을 경우에 다음과 같았다[그림 2]. 대략 절반에 가까운 데이터를 가지고 분석했을지라도 어느 정도의 경향을 보여주기 때문에 데이터를 산출하였다. 주로 화학분야와 농업분야 그리고 컴퓨터 과학 등의 분야에 논문이 많이 출판된 것으로 확인되었다. 본 분석 자료는 어떤 분야에 논문 출판이 활발하게 진행되는지 확인이 가능하기 때문에 논문 출판이 덜 된 분야에 대한 지원과 논문 출판이 활발한 분야에 대한 예산지원 증가를 고려해 연구 성과율을 높일 수 있을 것으로 생각된다.



인기 있는 출판사를 알 수 있고 기관 간 과학기술정보를 공유할 때에 참고자료로 활용할 수 있을 것으로 여겨진다.

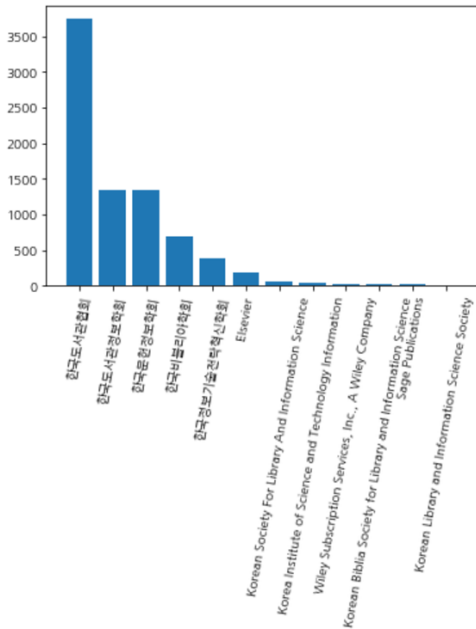


그림 5. 도서관과 정보학 분야 출판사의 논문 수 비교

### 6. 연도 별 한국인 연구자 논문 출판 수

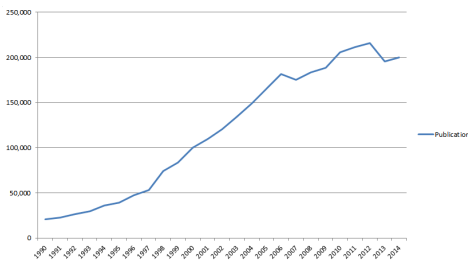


그림 6. 연도 별 한국인 연구자 논문 출판 수 비교

이번 챕터에서는 연도 별로 한국인 연구자의 논문이 얼마나 출판되었는지를 확인해 보고자 한다. 범위는 1990년도부터 2014년도까지이다. [그림 6]을 보면 꾸준히 논문 출판 수가 증가한 것을 확인할 수 있다. 지속적으로 한국인 연구자의 논문 출판 수가 증가하는 것으로 보아 앞으로는 한국인 연구자의 논문을 별도로 구축

하고 관리하여 국내 연구 성과물 평가지침으로 활용 가능할 것이다.

### 7. 특정 분야의 키워드 빈도 수

특정 분야의 키워드 빈도수를 연도 별로 확인할 수 있으면 어떤 연도에 어떤 키워드가 주로 이슈가 되는지 확인할 수 있으며, 트렌드에 맞게 관련 연구자에게 추천을 해주면 관련 연구자는 여러 논문을 읽지 않아도 현재의 트렌드에 맞추어 빠르게 연구 토픽을 설정하고 연구를 진행할 수 있기 때문에 연구 성과물의 생산성도 크게 증가하게 될 것이다. 본 논문에서는 의사소통, 원격 통신 분야를 선정하였으며, 2015년도에 해당 분야의 키워드가 어떤 것이 있는지 빈도수를 산출해 보았다. 4차 산업혁명과 관련이 있는 IoT 기술에서 빼놓을 수 없는 기술인 무선과 모바일 기술, 에너지 파워 기술 등이 주요 키워드로 나와서 관련 기술들이 점점 늘어날 전망으로 예상할 수 있다.

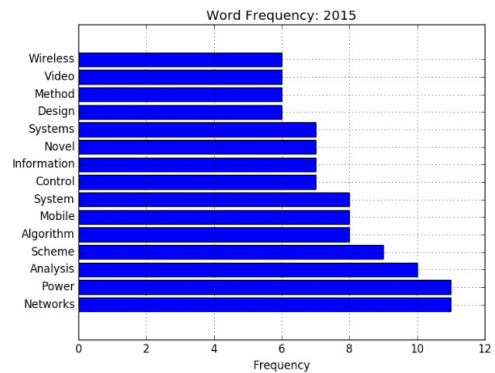


그림 7. 2015년도 의사소통, 원격 통신 분야 키워드 빈도 수

### 8. 기관 별 한국인 연구자의 빈도 수

한국인 연구자의 데이터 중에서 기관 값이 존재하는 데이터는 2017년 9월 7일 현재 대략 91만 건 정도 되며 이 데이터를 기반으로 총 기관 값을 산출해 보았을 때에 27만개의 기관이 산출되었다. 한국인 연구자가 1000명 이상인 기관을 그래프로 산출해 보았을 때 다음과 같았다. 본 분석 자료를 통하여 국내 연구자의 기관 분포율을 확인할 수 있다.

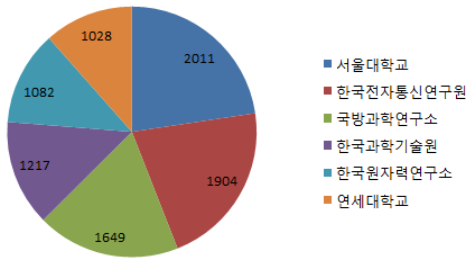


그림 8. 한국인 연구자가 1000명 이상인 기관(2017.09.17 기준)

### 9. 기관 별 한국인 연구자의 인용 건수

기관에 따른 한국인 연구자의 인용 건수를 산출하게 되면 기관 간 평가의 지표로도 활용이 가능하며, 연구자 수 대비 논문 인용 건수와의 상관관계 분석을 통해서 예산 계획을 고려할 때에 참고할 수 있는 자료가 될 수 있다. 본 논문에서는 연구자 수 대비 인용 수와 상관관계가 있는지를 조사해 보기 위해 수치를 산출하여 그래프로 나타내었다. [그림 9]는 인용 건수가 있는 한국인 연구자가 최소 15명 이상인 기관에 대해서 연구자 수에 대한 인용 지수를 산출해 보았을 때의 결과를 그래프로 나타낸 것이다. 그래프에서도 알 수 있듯이 연구자가 많다고 해서 인용 지수가 많이 나오지는 않았다. 이로써 연구자 수와 인용 지수와의 상관관계는 없는 것으로 확인되었다.

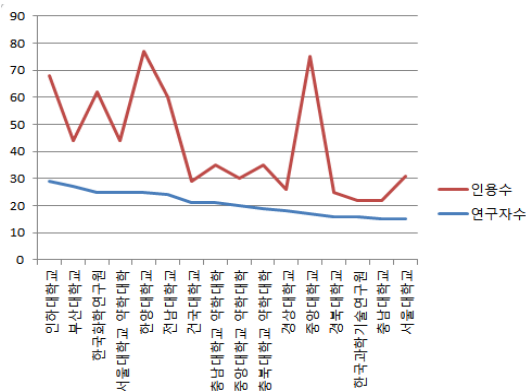


그림 9. 인용 건수가 있는 연구자의 기관 별 인용 수

### 10. 한국인 저자의 논문이 참고문헌으로 많이 인용하는 논문의 게재지

한국인 저자의 논문이 참고문헌으로 인용하는 논문의 게재지를 살펴보면 한국인 연구자가 참고문헌으로 인용하기 위해서 어떤 저널지를 주로 이용하는지 확인할 수 있고 수익 형 서비스로도 활용이 가능하다. 따라서, 본 논문에서는 한국인 저자의 논문이 참고문헌으로 인용하는 논문의 게재지를 확인하고 top 10 순위를 그래프로 나타내었다. [그림 10]의 그래프 수치는 논문 수를 나타낸 것이다. 주로 Science 지나 Nature 지 등의 저명한 저널지의 논문을 참고문헌으로 많이 인용하는 패턴을 보였다.

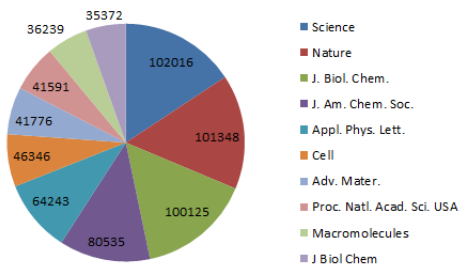


그림 10. 한국인 저자의 논문이 인용하는 논문의 게재지 Top 10

## IV. 결론

국가 과학기술정보 데이터가 대내외적으로 꾸준히 증가함에 따라 데이터 검색, 데이터와 데이터의 연결 서비스와 더불어 부가가치 정보를 발견하기 위한 시도가 요구되고 있다. 문헌 데이터의 연구 분석 자료를 토대로 국내 연구자의 연구 트렌드와 선호 키워드 등을 발견하고 관련 분야를 더욱 지원하고 발전시킬 수 있는 서비스를 도출하고자 본 논문에서는 기본이 되는 분석을 시도해 보았다. 본 연구의 장점은 새로운 정보에 대한 분석을 기존 데이터를 식별하여 가능하다는 점을 발견한 것이며 특정 분야의 심도 있는 분석은 아직 아쉬운 점으로 남는다. 이를 위해서 향후에는 보다 심도 있는 학술연구 분석을 진행할 예정이다.

## 참고 문헌

- [1] J. A. Yang and Y. J. Nam, "A Comparative Study on Metasearch System and Google Scholar in the Digital Library Environment," *Journal of the Korean Society for Information Management*, pp.45-54, 2010.
- [2] W. S. Shim, "Big Deal, Open Access, Google Scholar and the Subscription of Electronic Scholarly Contents at University Libraries," *Journal of the Korean Society for Information Management*, Vol.29, No.4, pp.143-163, 2012.
- [3] N. G. Kim, Y. J. Shin, G. C. Park, W. K. Joo, K. S. Choi, K. N. Choi, and J. S. Kim, "National R&D Paper Archiving Using NTIS & NDSL," *Korean Society For Internet Information*, pp.231-232, 2010.
- [4] S. H. Han and J. Y. Lee, "Searching Behavior of Scientists and Technologists Based on Their Information Needs," *Korea Society for Information Management*, pp.17-22, 2006.
- [5] J. Y. Lee, G. J. Bae, and J. S. Park, and Y. H. Moon, "Analysis of Information Usage Behavior of National Science and Technology Researchers," *Korea Technology Innovation Society*, pp.401-425, 2003.
- [6] Y. K. Chung, "A Reference Study on Archives and Records Management in the Journal of the Korean Society of Archives and Records Management during the Period of 2001-2010," *Journal of the Korean Society for Library and Information Science*, Vol.45, No.2, pp.367-391, 2011.
- [7] H. G. Choi, "A Study on the Research Patterns of Domestic Library & Information Researchers through LISA&ISA," *Journal of the Korean Society for Library and Information Science*, Vol.37, No.1, pp.61-80, 2003.
- [8] H. Y. Cho, "A Comparative Study on Researchers' Language Preference for Citing Documents in Different Subject Fieldsn," *Journal of the Korean Biblia Society for Library and Information Science*, Vol.21, No.1, pp.211-221, 2010.
- [9] C. H. Song, K. Y. Hwang, and Y. S. Choi, "A Study of Research Activities of Young Researchers in Science and Engineering in Korea," *Korea Technology Innovation Society*, pp.244-253, 2011.
- [10] H. S. Lee and H. J. Lee, "Development of the Journal Paper Metadata Analysis System for the Creation of the Research Similarity-Weighted Social Network and the Extraction of Researcher Relationship Information," *Journal of Korean Institute of Information Technology*, Vol.14, No.1, pp.143-163, 2016.
- [11] M. Y. Lee, P. Kim, and H. M. Jung, "Study on Researcher Researcher-centered Academic Information Analysis Services," *Korea Information Science Society*, pp.26-30, 2010.
- [12] J. Y. Lee, "Analyzing the Network of Academic Disciplines with Journal Contributions of Korean Researchers," *Journal of the Korean Society for Information Management*, Vol.25, No.4, pp.327-345, 2008.
- [13] J. Y. Lee, "Exploring a Researcher's Personal Research History through Self-Citation Network and Citation Identity," *Journal of the Korean Society for Information Management*, Vol.29, No.1, pp.157-174, 2012.
- [14] D. S. Han, D. H. Jang, and S. H. Han, and J. M. Yang, "An Empirical Study on the Impacts of Public Funding on the Research Performance of Academic Faculties," *Korean Public Administration Review*, Vol.42, No.4, pp.265-290, 2008.
- [15] L. J. Lee and S. J. Kim, "A Study on the

Literature Review of Information Use Behavior in Specialized Fields,” Journal of the Korean Bibliography Society for Library and Information Science, Vol.26, No.2, pp.129-153, 2015.

- [16] M. R. Kim and E. B. Lee, “A Study on the Citing Behavior of Military Scientists by Reference Analysis,” Journal of the Korean Society for Library and Information Science, Vol.47, No.2, pp.143-159, 2013.
- [17] S. I. Kim and Y. K. Chung, “A Comparative Study on Authority Records for Korean Writers Among Countries,” Journal of the Korean Society for Library and Information Science, Vol.49, No.1, pp.379-403, 2015.
- [18] S. J. Bae, “An analysis of research carried out by domestic administrative researchers,” Journal of the Korean Society for Library and Information Science, Vol.35, No.1, pp.123-141, 2001.
- [19] J. W. Lee and H. R. Bak, “Characteristics of Korean Researchers through Bibliometric Analysis of Papers Published in International LIS Journals,” Journal of Korean Library and Information Science Society, Vol.47, No.1, pp.217-242, 2016.
- [20] I. S. Cho and M. S. Han, “An Analysis on the Articles and Cited Journals Suggested by LIS Researchers of Korea,” Journal of the Korean Bibliography Society for Library and Information Science, Vol.18, No.2, pp.89-105, 2007.

저 자 소 개

최 원 준(Wonjun Choi)

정회원



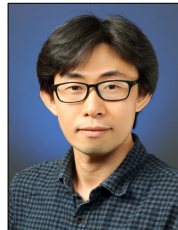
- 1999년 3월 ~ 2006년 2월 : 원광대학교 수리통계학(학사)
- 2013년 3월 ~ 2017년 2월 : 한국과학기술연합대학교 과학기술정보학(공학석사, 공학박사)
- 2017년 2월 ~ 현재 : 한국과학기술정보연구원 연구원

기술정보연구원 연구원

<관심분야> : 과학기술정보, 네트워크 분석, 데이터 분석

김 재 훈(Jayhoon Kim)

정회원



- 1999년 7월 ~ 2005년 12월 : 한국과학기술원 연구원
- 2006년 1월 ~ 2017년 현재 : 한국과학기술정보연구원 선임연구원
- 2008년 2월 : 성균관대학교 경영학과(경영학 석사)

학과(경영학 석사)

<관심분야> : 학술논문정보, 전자정보컨소시엄, 정보추구행태

김 정 환(Jeonghwan Kim)

정회원



- 2013년 : 충남대학교 문헌정보학(박사)
- 2014년 : 유럽핵입자물리연구소(CERN) 방문연구원
- 2016년 ~ 현재 : 영국물리학회 출판부(IOPP) 자문위원

▪ 2006년 ~ 현재 : 한국과학기술정보연구원 책임연구원, 실장

<관심분야> : 데이터베이스