

과학기술 통계·정보에 기반한 한국의 기초연구 효과성에 관한 연구

Study on Effectiveness of Korea's Basic Research based on S&T Statistics and Information

박귀순*, 석혜은**, 박진서***, 김해도*
한국연구재단*, 이화여자대학교**, 한국과학기술정보연구원***

Kwisun Park(kwisun_park@nrf.re.kr)*, Hyeeyun Seok(graceseok@gmail.com)**,
Jinseo Park(jayoujin@kisti.re.kr)***, Haedo Kim(hdkim@nrf.re.kr)*

요약

최근 과학기술정보 콘텐츠 분석 기반의 R&D 전략 수립이 중요시됨에 따라 국내·외 과학기술 통계·정보를 수집, 범주화하여 DB를 구축하고 이를 증거 기반의 기초연구정책을 수립하는데 활용하고자 하였다. 국내·외 49종의 과학기술 관련 보고서 등 자료를 수집하여 6,446개 지표 추출, 877개 통계·정보로 범주화, 이중 208개를 집중 관리 통계로 도출하였다. 이렇게 구축된 DB의 활용성 제시를 위해 논문정보 관련 심층분석을 통해 제한적으로나마 우리나라 기초연구의 효과성을 보여주하고자 하였다. 이를 위해, 고피인용 논문 창출과 R&D 투자 간의 통계기반 분석과 고피인용 논문의 국내·외 지식맵 비교·분석을 실시하였다. 그 결과, 기초연구비 증가가 우리나라의 보편적 연구활동도와 세계적 연구성과 창출에 긍정적 영향을 미친 것으로 나타났다. 또한 순수기초연구분야로 간주되어온 수학, 물리, 화학이 신성장동력이라 불리는 AI, 청정에너지 등 기술 발전의 기반이 되어왔음을 알 수 있었다.

■ 중심어 : | 과학기술 | 통계·정보 DB | 기초연구 | 기초연구 효과성 | 논문 지식맵 |

Abstract

As increasing the importance of the R&D strategy planning based on S&T evidence, the S&T DB was conducted to develop appropriate basic research support strategies by collecting 49 multiple-nations' statistics and information, extracting 6446 raw data, categorizing 877 indicators including 208 core indicators. An statistical and knowledge map analysis using the highly cited publication-related indicators was conducted to examine the expansion of DB utilization by demonstrating effectiveness of Korea's basic research. As a result, basic research investment have a strong influence on creating outstanding R&D outcomes and on producing a foundation of various S&T based-growth engines.

■ keyword : | Science & Technology | Statistics & Information DB | Basic Research | Effectiveness of Basic Research | Knowledge Map |

* 이 논문은 2016년도 정부(미래창조과학부)의 재원으로 수행된 연구임(2016R1A6A6913149,2016R1A6A6913152).

* 본 논문은 한국콘텐츠학회 2017 춘계 종합학술대회 우수논문입니다.

접수일자 : 2017년 09월 29일

심사완료일 : 2017년 10월 16일

수정일자 : 2017년 10월 16일

교신저자 : 박귀순, e-mail : kwisun_park@nrf.re.kr

I. 서론

최근 과학기술 정보 콘텐츠의 가치평가 모형 구축 등 과학기술 통계·정보와 관련된 콘텐츠의 중요성이 강조되고 이를 활용한 국가 과학기술 정책 수립과 진단의 필요성이 제기되고 있다[1-3]. 1967년 과학기술처 신설을 시작으로한 과학기술 50년을 맞이하여 그동안의 기초연구에 대한 지원에 대비한 성과 분석과 더불어 기초연구의 발전을 위한 증거기반의 중장기 비전·전략 수립이 요구되고 있다[2][4][5]. 이에 단편적인 연구개발 관련 투입 지표, 연구 결과물 관련 통계 제공에서 나아가 과학기술과 관련한 투입, 산출, 수준 진단 등을 위한 종합적이고 핵심적인 과학기술 통계·정보 제공의 필요성이 높아지고 있다. 또한 지속된 세계 저성장 기조에 따라 기초연구를 포함한 연구개발 투자의 효과성 증명에 대한 요구가 커지고 있어 증거에 기반한 합리적인 진단과 제공이 필요한 시점이다[6-8]. 기초연구에 대한 투자가 혁신과 미래를 이끌 중요한 기반적 요소임을 보여주기 위해서는 다양한 과학기술 통계·정보를 활용한 심층적인 효과성 분석이 지속적으로 필요할 것이다. Narin 외(1997)의 연구에 따르면 미국 상업특허의 73%가 학술연구 결과를 인용하였음을 알 수 있었으며 이는 기술혁신이 상당부분 기초 과학연구 결과에 기반하고 있음을 나타내고 있다[9]. 또한 미국 National Institutes of Health(NIH)는 1980~2007년 사이 NIH의 연구보조금을 지원받은 과제를 대상으로한 효과성 분석 연구를 통해 연구논문 30.8%가 산업특허(commercial patent)에 인용되었음을 밝혔다[8]. 직접적인 상업적 가치 창출보다는 시장의 성숙 등 지속적인 환경변화에 따라 예상치 않은 효과를 창출할 수 있는 기초연구의 장기적 효과성을 제시하고자 한 것이다[8]. 국외 사례와 같이 우리나라도 연구개발 투자 효과, 특별히 기초연구와 같이 직접적인 경제적 가치를 증명하기 어려운 R&D 투자에 대한 객관적인 효과성 분석이 필요할 것이다. 국내에서도 이미 공공 및 민간분야에서는 대규모 DB에 축적된 빅데이터로부터 정보를 분석하여 의사결정에 필요한 정보를 추출하여 그 정보를 활용하는 방법에 관심이 높아지고 있다[10]. 이러한 변화를 지원하기 위해 기반자료로 활용 가능한 국내·외의 다양한 과학기술 정보를

즉시 활용가능한 형태(예시, 디지털화된 통계표 형태)로 제공하는 것이 필요하나 이러한 형태의 정보를 통합적으로 제공해주는 기능은 미흡한 실정이다.

이에 본 연구에서는 국내·외 주요 과학기술 자료를 수집하여 통계지표를 도출하고 체계적인 범주화를 통해 광범위한 과학기술 통계·정보 DB를 구축 하였다. 이렇게 구축된 DB는 연구관리 전문기관, 부처 등 관계자에 제공하여 증거기반의 정책결정을 지원하고 통계·정보의 심층분석을 통한 이슈도출 등에 활용할 수 있을 것으로 기대된다. 이렇게 구축된 과학기술 통계·정보 DB의 구체적인 활용방안을 제시하기 위해 기초연구의 대표적 성과로 간주되는 학술논문 성과와 관련한 심층 분석 사례 연구를 수행하였으며 이를 통해 우리나라 기초연구 투자의 효과성을 제한적으로나마 보여주고자 하였다. 사례분석은 첫 번째 다양한 연구개발 투자와 연구성과 창출 간 연계성에 관한 통계적 측면의 분석, 두 번째 고피인용(피인용 상위 1%) 논문의 지식맵 트렌드와 국내·외 트렌드 비교·분석을 실시하였다.

II. 관련 연구

본 연구는 기초연구정책과 같은 국가 과학기술과 관련한 정책수립을 지원하기 위해 활용도가 높은 과학기술 통계·정보와 관련한 DB를 체계적이고 포괄적으로 구축하고자 하였다. Organization for Economic Co-operation & Development(OECD)는 Main Science & Technology Indicators(MSTI), 영국은 Research & Development Expenditure, 한국은 Korea R&D Statistics(KRS), 연구성과정보시스템(성과마루), National Science & Technology Information Service(NTIS), 를 통해 다양한 자원 투입 지표, 과제 지원 지표, 연구 결과물 관련 지표 등 기본 통계량 중심의 과학기술 통계 정보를 제공하고 있다[11-15]. OECD MSTI[11]는 연구비 투입, 연구 인력, 국가별 경제력 등 국제적 비교가 가능한 정보를 비교적 다양하게 제공하고 있으나 국가별 거시적인 통계치로서 상세 분석을 통한 정책수립에는 한계점을 가진다. 영국의 Research & Development Expenditure[12]의 경우는 연구개발 투자에 대한 비교

적 자세한 정보를 수집할 수 있으며 사용자 관점에서 이해하기 쉬운 형태로 관련 정보를 제공하고 있어 DB 구성 틀을 벤치마킹 할 수 있으나 영국에 국한된 정보라는 제한점을 가진다. KRS[13]와 성과마루[14]는 다차원 기초·원천연구 지원 현황(시계열, 사업별), 성과정보(정부연구개발사업 과제별 논문·특허 등 성과소개서), 연구자 네트워크, 연구보고서 등을 제공하고 있으나 한국연구재단을 통해 지원된 과제와 관련된 정보만을 제공하고 있다. NTIS[15]의 경우에는 우리나라 국가연구개발사업 지원 현황, 참여인력 등에 관한 가장 광범위한 정보를 수집·제공하고 있으며 국가연구개발사업조사분석, 국가연구시설장비, 과학기술통계와 관련한 온라인 통계분석 서비스를 제공하고 있다. 그러나 이러한 정보는 국내 현황과 관련된 정보라는 제한점을 가진다. 한국과학기술정보연구원의 NDSL (National Digital Science Library)[16]과 한국과학기술기획평가원의 발간자료 검색서비스[17]는 다양한 과학기술 관련 문헌정보(논문/특허/보고서 등)를 제공하고 있으나 즉시 활용 가능한 디지털화된 통계표 형태가 아닌 원문 제공 서비스 형태로써 사용과 상세정보 검색에 제한점을 가진다. 종합하여 보면, 국내·외 다양한 주체들이 과학기술과 관련한 통계와 정보를 제공하고 있으나 거시적인 정보 제공으로 상세 비교·분석에 제한적이거나, 자국의 정보만을 제공하거나 원문제공으로 상세정보 검색·활용의 편의성이 떨어지는 등의 제한점을 가진다. 더불어 이러한 정보가 분산적으로 수록되어 필요에 따라 각각의 서비스를 개별적으로 검색해야 하며 필요한 정보를 얻을 수 있는 서비스를 인지하지 못하는 경우 정보수집에 상당한 시간을 투자해야 하는 문제가 발생할 수 있다.

기초연구의 대표적인 성과물은 학술논문으로 미국 National Research Council에 따르면 기초연구 수행의 핵심주체인 대학의 지식은 8가지 경로를 통해 확산되는데 학술논문의 출판은 첫 번째 경로인 고등교육인력 양성 다음으로 제시되었다[18]. 이에, 동 연구에서는 과학기술 통계·정보 DB에 구축된 학술논문과 관련한 시계열 통계·정보의 다각적인 심층분석을 통해 DB의 효과적인 활용방안을 제시하고자 하였다. 이를 위해 통계 분석(statistical analysis)과 지식맵(knowledge map)

분석을 실시하였다. 연승민 외(2016)는 연구비 투입, 과제 특성(연구원수·연령, 수행기간 등) 등 연구성과에 미치는 영향요인에 관한 광범위한 선행 연구 결과를 제시하였다[19]. 동 선행연구에 따르면 연구비 투입(정부연구비, 민간연구비)과 연구성과(SCI 논문, 특허) 간 관계를 통계적 모형 검증(상관분석, 회귀분석, 분산분석 등)을 통해 정의하고자 하였으며, 분석결과 모두 정(+)의 상관관계를 보이는 것으로 나타났으며 연구비 증가가 반드시 연구성과 창출과 정비례하지 않을 수도 있음을 제시하였다[19-23]. 제시된 선행연구 분석 결과를 통해 연구비 투입과 성과창출 간의 관계를 정의하는데 통계적 분석이 효과적인 방법론 중의 하나임을 알 수 있었다. 다만, 선행연구들이 특정사업을 대상으로 분석하였고 연구비 중 정부연구비(민간 매칭 포함)만을 고려하였다는 점에서 기초연구비 투입의 효과성을 제시하는데는 다소 제한적일 수 있다. 학술논문 등 기술문헌을 활용한 기술 동향조사·분석이 활발해 지고 있으며 국가연구개발사업의 과제기획 시 논문특허의 사전조사를 실시하도록 하고 있다[23][24]. 최근에는 computational power 증가, 네트워크 분석력 강화, 다양한 분석툴 개발로 지식맵을 활용한 정보의 직관적 분석이 수행되고 있다[25-28]. 지식맵은 문서의 대표 단어를 기반으로 동시출현 관계·빈도에 기초하여 연구주제를 clustering 하고 시각화한 Bibliometric Map으로 특정분야의 연구 주제, 연구주제의 특징과 연계성 파악에 활용할 수 있다[26]. 학술논문 정보 기반의 지식맵을 활용하여 mobile-health, 리튬의 독성 연구, 지역별 핵심연구 트렌드 분석 등 다양한 연구결과가 발표되고 있다[28-30]. 동 연구에서는 학술논문 키워드의 빈도, 점유율의 증가율을 밀도맵과 네트워크맵으로 도출하여 다각적인 측면에서 분석하고자 하였다.

III. 과학기술 통계 · 정보 DB 구축

과학기술 통계·정보 DB 구축은 1단계 데이터 수집(연구재단 등 국내 기관 보유자료, 국외 자료), 2단계 지표 추출 및 데이터 범주화·코드화, 3단계 DB 구축·활용

으로 진행되었다. 1단계 데이터 수집을 위해 국내·외 49종의 과학기술 관련 보고서, DB, 링크 등의 연도별 자료를 on/offline으로 수집하였다. 수집된 49종의 자료는 31개의 국내 과학기술 관련 통계·정보와 18개의 국제 과학기술 관련 통계·정보로 구성되었다. 수집된 자료별로 제공기간은 다양하나 최대 1999년부터 2016년까지 정보를 취합하였다. 2단계로 수집된 49종의 자료로부터 6446개의 개별 지표를 추출하였고 수집된 자료의 구조와 특성 분석을 기반으로 계층화된 범주화 체계를 도출하였다. 1차 범주화를 통해 5개 그룹(① 국가, ② 대학, ③ 정부연구소, ④ 산업계, ⑤ 연구재단), 2차 범주화를 통해 4개 그룹(① 예산투입, ② 인력투입, ③ 프로세스, ④ 산출), 3차 범주화를 통해 2개 그룹(① 국내 통계·정보, ② 국제 통계·정보), 4차 범주화를 통해 산출에 한하여 5개 그룹(① 논문, ② 특허, ③ 기술료, ④ 사업화/창업, ⑤ 기타)으로 분류하였다. 4단계로 계층화된 체계를 기반으로 877개 통계·정보로 범주화하고 이중 208개를 집중 관리 통계로 도출하였다.

표 1. 주요 국내/국제 과학기술 통계·정보 수집 자료

주요 수집 자료명	
국내	투자중심 : 부처(과학기술정보통신부, 교육부) 사업지원실 적서, 국가연구개발사업 조사분석보고서, 국가연구개발사업 예산편성결과, 정부기초연구비 비중
	성과중심 : 국가연구개발사업 성과분석보고서, 피인용 상위 1% 논문실적 비교분석, SCI 분석연구
	인력중심 : 대학연구활동실태조사, 교육통계연보, 취업통계연보
	프로세스 : 산학협력실태조사, 과학기술혁신역량평가, 기술수준평가보고서
종합보고 : 과학기술연구개발활동조사보고서, 과학기술통계백서	
국제	성과중심 : WIPO 및 USPTO IP 보고서, ONS 통계
	프로세스 : IMD 국제경쟁력평가/Global Innovation Index, 과학기술혁신역량평가보고서, Bloomberg Innovation Index, Nature Index
종합보고 : OECD MSTI, NSF Science & Engineering Indicators	

표 2. 과학기술 통계·정보 범주화 및 코드화 예시

1st digit		2nd digit		3rd digit		4th digit	
코드	내용	코드	내용	코드	내용	코드	내용
1	국가	1	투입 (예산)	1	국내	1	-
				2	국제	2	-

2	투입 (인력)	1	국내	1	-
		2	국제	2	-
3	프로세스	1	국내	1	-
		2	국제	2	-
4	산출	1	국내	1	논문
				2	특허
				3	기술료
				4	사업화/창업
				5	기타
	2	국제	1	논문	
			2	특허	
			3	기술료	
			4	사업화/창업	
			5	기타	

표 3. 국내·외 과학기술 통계·정보 DB 구축 및 가공 단계

범주	(1단계) 자료수집/데이터 추출				
	투입 (예산)	투입 (인력)	프로세스	산출	합계
국가	860	381	57	1,271	2,569
대학	862	608	81	1,748	3,299
정부연구소	5	8	-	7	20
산업계	76	29	1	21	127
연구재단	179	9	-	243	431
합계	1,982	1,035	139	3290	6,446

범주	(2단계) 데이터범주화/가공				
	투입 (예산)	투입 (인력)	프로세스	산출	합계
국가	84	45	27	140	296
대학	50	33	10	395	488
정부연구소	2	5	0	0	7
산업계	3	6	0	0	9
연구재단	23	2	0	52	77
합계	162	91	37	587	877

범주	(3단계) 데이터범주화/가공(주요통계)				
	투입 (예산)	투입 (인력)	프로세스	산출	합계
국가	15	9	12	75	111
대학	8	11	-	21	40
정부연구소	2	5	-	-	7
산업계	3	6	-	-	9
연구재단	13	-	-	28	41
합계	41	31	12	124	208

IV. 과학기술 통계·정보 기반 기초연구 투자 효과성 분석

1. 통계분석 기반 기초연구 투자의 효과성 분석

우수한 학술논문 창출은 연구개발의 성과, 특별히 기초연구의 대표적인 성과로 간주되어 오고 있다. 기초연구 투자가 우리나라의 세계적인 연구성과 창출에 미치는 영향과 가장 주요한 요인을 판별하여 그 효과성을 보여주하고자 가설을 수립하고 통계적 검증을 실시하였다. 과학기술 통계·정보 DB로 구축된 최근 15년(2000년~2015년) 간 우리나라의 다양한 주체별 연구개발 투자와 고피인용 SCI 논문(1%, 0.1%, 0.01%) 창출 정도의 기술통계분석(descriptive statistics analysis), 피어슨 상관분석(pearson correlation analysis), 회귀분석(regression analysis)을 설계하였다. 한국연구재단은 2000년~2011년을 대상으로 고피인용 SCI 논문 분석을 최초로 추진하여 매년 관련 보고서를 발간하고 있어 분석기간을 최근 16년으로 선정하였다. 우리나라 총연구개발비, 정부연구개발비, 기초연구비는 OECD MSTI(2017.2.) 자료를 기반으로 하였다[11]. 고피인용 SCI 논문 자료는 분석기간에 해당하는 ‘주요국의 피인용 상위 1% 논문실적 비교분석 보고서’에 수록된 자료를 활용하였다[31].

- <가설1> 연구비 증가가 우리나라의 보편적 연구활동도(SCI 논문수) 증대와 세계적 연구성과 창출에 영향을 미쳤을 것이다.
- <가설2> 보편적 연구활동도(SCI 논문수) 증대가 세계적 연구성과 창출에 영향을 미쳤을 것이다.
- <가설3> 국가총연구개발비, 정부연구개발비, 기초연구비 중 SCI 피인용 상위 1% 논문 성과 창출에 기초연구비가 가장 큰 영향을 미쳤을 것이다.

수집된 자료는 SAS9.4를 이용하여 통계분석을 실시하였다[32]. 주요 지표변수들에 대한 기술통계는 평균, 표준편차로 제시하였고 주요 지표변수들 간의 관련성을 분석하기 위해 피어슨 상관분석을 실시하였다. 그 결과, 연구비 관련 변수(총연구개발비, 정부연구개발비, 기초연구비)들의 증가와 SCI 논문수 증가 및 고피인용

논문 창출(1%, 0.1%, 0.01%)과 상관이 높고(.76~.99) 통계적으로도 유의한 것으로 판별되었다($p < .05$). 우리나라의 경우 양적성과의 증가에 비해 질적성과가 저조하다는 비판을 검증하고자 양적성과(SCI 논문)와 질적성과(고피인용 논문) 간 연계성 분석을 실시하였다. 피어슨 상관분석 결과 SCI 논문 수 증대와 피인용 1%, 0.1%, 0.01% 논문 창출 간에 상관이 높고(.78~.98) 통계적으로도 유의한 것으로 판별되었다($p < .05$). 또한 SCI 논문 수의 논문 피인용 1%에의 효과를 단순회귀분석(simple linear regression)으로 검증해본 결과 통계적으로 유의한 것으로 나타났다($t = 21.6, p < .05$).

표 4. 기술통계분석결과(N=16)

	평균(M)	표준편차(S)
피인용 상위1% 점유율(%) (1)	2.18	0.82
피인용 상위0.1% 점유율(%) (2)	2.33	1.47
피인용 상위0.01% 점유율(%) (3)	3.21	2.85
피인용 상위1% 논문수 (4)	259.38	147.41
SCI 논문수 (5)	34416.00	14620.00
총연구개발비(백만 USD) (6)	43882.00	19355.00
정부연구개발비(백만 USD) (7)	13118.00	5589.00
기초연구비(백만 USD) (8)	7349.00	3854.00

표 5. 피어슨 상관분석결과(N=16, $p < .05$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	1							
(2)	.910	1						
(3)	.748	.828	1					
(4)	.991	.929	.740	1				
(5)	.983	.925	.788	.985	1			
(6)	.978	.927	.773	.988	.995	1		
(7)	.969	.896	.765	.970	.991	.990	1	
(8)	.985	.933	.786	.992	.995	.997	.990	1

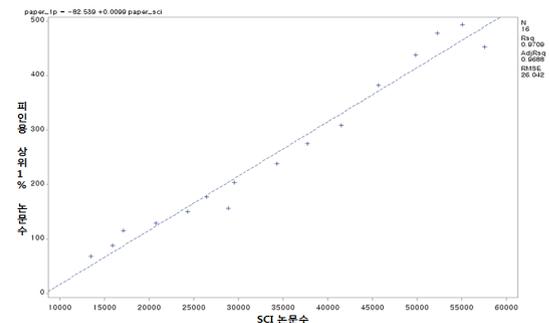


그림 1. SCI논문수와 피인용 상위1% 논문수 간 산점도와 회귀직선

국가총연구개발비, 정부연구개발비, 기초연구비 중 피인용 상위 1% 논문 성과 창출에 가장 영향을 미치는 지표변수가 무엇인지를 확인하기 위하여 중다회귀분석을 실시하였다. 변수 선택의 방법으로 단계적 변수선택(stepwise variable selection)방법을 사용하였고, 입력변수의 기준값과 제거변수의 기준값은 .15를 사용하였다. 그 결과 기초연구비와 정부연구개발비가 순차적으로 선택되었는데 이 경우 독립변수의 다공선성(multicollinearity)을 진단하는 통계량인 분산팽창지수(variance inflation factor; VIF)가 49로 산출(>10의 경우, 다공선성이 매우 높다고 판단하여 모형을 신뢰하기 어려움)되어 두 개의 지표변수가 모두 모형에 투입되는 것은 적절하지 않은 것으로 판단하였다[33]. 모형의 전반적인 적절성을 결정계수 R^2 , F 값과 유의확률로 확인한 후, 독립변수의 상대적 기여도를 나타내는 표준화계수(beta), t 값과 유의확률로 독립변수의 효과를 확인하였다. 그 결과 SCI 피인용 상위 1% 논문 성과 창출에 기초연구비가 가장 큰 영향을 미친 것으로 나타났다($t=52.37, p<.05$).

최근 연구개발비 절대투자금액과 투자비중 간의 중요도에 대한 이슈가 회자되고 있어 추가적으로 기초연구비 절대투자금액과 기초연구비 투자 비중(총연구개발비 중 비중) 중 피인용 상위 1% 논문 성과 창출에 더 큰 영향을 미치는 변수가 무엇인지 확인하고자 하였다. 이를 위하여 중다회귀분석을 실시하였고 기초연구비와 기초연구비 투자 비중을 독립변수로 피인용 상위 1% 논문수를 측정하는 모형에서 기초연구비가 더 영향을 주는 변수로 나타났으며($R^2=.94, t=23.47, p<.05$), 기초연구비 투자 비중이 동시에 포함된 모형($R^2=.98$)과도 큰 차이를 나타내지 않았다.

2. 지식맵 기반 기초연구 트렌드 분석

과학기술 통계·정보 DB로 구축된 정보를 활용하여 대표적 기초연구분야인 수학, 물리, 화학 분야의 고피인용된 학술논문 키워드 기반의 분석을 통해 시계열적 연구동향과 시사점을 도출하였다. 최근 10년(2006~2015년) 간 Scopus에 등재된 수학, 물리, 화학 분야의 과학기술논문 중 피인용 상위 1% 논문의 키워드 밀도와 중

가울을 분석하여 지식맵을 구축하고 지식맵에 내제된 연구 트렌드를 분석하였다.

연구 트렌드 분석을 위해 키워드 간의 관계를 중심으로 살펴보았다. 일반적으로 텍스트에서 단어들 간의 관계를 분석하는 방법은 단어들의 동시출현 관계와 빈도를 중심으로 한 연결망분석(Network Analysis), 단어들의 유사도에 기반한 잠재의미분석(Latent Semantic Analysis), 기계학습과 자연어처리 분야에서 확률 및 통계모형에 기반하여 문서와 단어 집합에서 토픽을 찾는 토픽모델링(Topic Modeling) 등 다양하다[34][35]. 학술논문의 경우, 저자 키워드와 색인어 등 구조화된 문서의 대표 단어들이 존재하고 직관적으로 이해할 수 있는 동시출현 관계와 빈도에 기반한 co-wording mapping을 통해 연구 트렌드를 도출하고 있다[36]. 본 연구에서는 네덜란드 Leiden 대학교에서 개발한 계량서지학적 네트워크 분석과 이를 시각화 해주는 소프트웨어인 VOSviewer를 활용하여 3대 기초연구분야 별 지식맵을 구성하는 연구주제가 무엇인지 확인(clustering)하고, 연구주제 간의 관계를 시각화(mapping) 하였다. VOSviewer에서 두 단어 간의 유사도는 두 단어가 한 논문에서 동시출현한 횟수에 따라 비례하고, 두 단어가 다른 단어들과 동시 출현한 빈도의 합에 반비례 한다[37]. 특정 단어의 밀도는 이웃하는 단어의 수와 이웃하는 아이템의 중요성을 반영한다. 따라서 Density view는 전체적인 맵의 구조를 개괄하는데 유용하며 맵에서 어느 영역이 가장 중요한 영역인지 판단할 수 있다. 맵에서 특정 지점의 색은 밀도로 표현(파랑-초록-노랑-주황-빨강)되는데 특정 지점 주변에 아이টে이 많고 weight가 크면 빨강색에 가깝게 표시, 반대의 경우 파랑색으로 표시된다. 빨강색은 관련성이 높은 단어(노드)가 밀집되어(dense) 분포하고 있다는 것을 의미한다.

3개 기초연구분야 지식맵 구축을 위해 저자키워드와 인덱스키워드를 활용하였고 중요 트렌드 강조를 위해 분야별 특성(논문수 등)을 고려한 논문 출현 빈도 기반의 임계값(cut-off)을 설정하였다. 연구분야별 키워드의 시계열 분석을 위해 전세계의 해당분야별 피인용 상위 1% 논문의 키워드를 3개 주기로 나누어 키워드 밀

도 맵을 도출하고 10년 전·후 트렌드를 비교하였다(주기1: 2006-2009, 주기2: 2009-2012, 주기3: 2012-2015). 또한 연구분야별 떠오르고 있거나 지속적으로 강한 세부분야의 과학과 전세계와 한국의 트렌드를 비교·분석하기 위해 10년간 키워드의 빈도 및 점유율의 증가율 맵을 전세계(한국 포함)와 한국 맵으로 나누어 분석하였다.

분석 결과, 수학분야의 경우 초기에는 전통적 수학 키워드 그룹이 강세였고, 이후 인공지능(Artificial Intelligence, AI)의 기반 기술인 다양한 알고리즘 기술(Evolutionary/Genetic Algorithm, Neural Network)이 부상하였고 바이오 관련하여 고유영역(Genetics, Genomics 등)을 구축하였다. AI는 알고리즘 기술과 바이오 관련 클러스터와 연계되어 있었으며 바이오 관련 영역의 키워드 점유율 증가율이 가장 큰 것으로 나타났다. 기후변화와 관련한 키워드(Climate Change, CO₂ 등)는 중요도는 상대적으로 낮으나 점유율의 증가율이 커지고 있다. 전세계 트렌드와 비교하여 한국은 전통적 수학 키워드 중심으로 세계적 수준의 연구가 진행 중인 것으로 분석되었다.

물리분야의 경우 초기에 전통적 물리 키워드(Optics, Ray 등)와 신소재(Graphene 등) 기술의 밀도가 높았고, 이후 이러한 키워드와 응용기술과 관련한 키워드 간의 연계성이 높아지고 이러한 그룹이 주요영역으로 진입하였다. 특별히 신소재기술의 경우 초기에 그래핀 중심에서 반도체, MoS2(그래핀 다음의 차세대 2차원 소재), 페로브스카이트 등 키워드가 다양화되고 고유의 영역을 형성하였다. 또한 10년간의 키워드 점유율 증가율 분석 결과 리튬전지, 태양전지 등과 같은 친환경에너지 관련 키워드가 주요영역으로 분석되었다. 한국은 그래핀, 페로브스카이트, 나노 중심의 신소재, 태양전지 등의 친환경에너지 기술이 주요영역으로 세계적 트렌드와 유사한 것으로 분석되었다. [그림 2]와 [그림 3]에 나타난 바와 같이 다양한 신소재 기술(빨강색으로 표기된 영역 중 MoS2, Graphene 등)과 친환경에너지 관련 기술(빨강색으로 표기된 영역 중 Solar cells, Lithium batteries 등)이 주요영역으로 부상하는 트렌드는 물리 분야 내 세부분야인 응집물리에서는 더욱 뚜렷하게 보

이고 있다. [그림 2]와 [그림 3]에 따르면, 친환경에너지 관련 기술이 중요 영역으로 자리잡은 점은 전세계와 한국이 유사하였으나 신소재 측면에서는 전세계는 MoS2 중심으로 한국은 Graphene 중심으로 연구되고 있는 것으로 나타났다.

화학분야의 경우 초기에는 화학구조, 합성, 전기화학 등 전통적 분야 관련 키워드가 주요영역을 이루고 있었고 이후 청정에너지, High-tech 장비, 신소재, 바이오 관련 기술 영역이 주요영역으로 진입 및 부상하였다. 청정에너지 기술로는 리튬전지, 태양전지 등이, High-tech 장비로는 SEM(Scanning Electron Microscopy), TEM(Transmission Electron Microscope), EIS(Electrochemical Impedance Spectroscopy) 등이, 신소재 기술로는 그래핀, 탄소복합소재 등을 중심으로 영역을 형성하였다. 바이오 관련 기술로는 Human 키워드를 중심으로 형광염료, 약물전달시스템 등이 주요영역을 형성하였다. 한국은 전세계 트렌드와 비교하여 주요 영역의 키워드의 다양성은 한정되었으나 전체적으로 유사한 것으로 분석되었다.

표 6. 분석대상 키워드 개요

분야	분석 논문수 (한국)	분석 키워드		
		키워드수	Cut-off 기준	커버리지
수학	6,527 (216)	378	20회 이상	62%
물리	14,877 (709)	356	40회 이상	91%
화학	7,483 (315)	395	60회 이상	83%

표 7. 수학분야 지식맵 분석

분야	주요영역*(주변 관심 키워드)	
10년 전·후 비교	전	Problem solving(optimal), Numerical method
	후	Algorithm(artificial intelligence, evolutionary/genetic algorithm), Genetics(genomics, human, climate change) Chaotic systems(Lyapunov functions, neural network) Differential equations(numerical methods)
세계 vs. 한국 비교	세계	Problem solving, Algorithm, Nonlinear equations, Chaotic systems(Lyapunov functions, NN) Genetics(genomics, human, AI, Climate change)
	한국	Lyapunov Krasovskii functions(time varying control systems), Fixed point(nonexpansive mapping), magnetohydrodynamics

* 지식맵에서 빨강으로 표기된 분야

표 8. 물리분야 지식맵 분석

분야		주요영역(주변 관심 키워드)
10년 전·후 비교	전	Optics, Graphene, Film/Carbon(carbon nanotube)/Oxide, Magnetic/Spin/Energy, Ray/Spectroscopy
	후	Optics(Fiber, Laser)/Plasmon, Graphene(semiconductor, MoS2), Nanoparticle, Magnetic
세계 vs. 한국 비교	세계	Perovskite(solar cells), Lithium ion battery(energy storage, supercapacitor), MoS2, Sensor
	한국	Graphene(perovskite), Nanoparticle(nanowire, carbon nanotube), Solar Cell, Sensor(biosensor)

표 9. 화학분야 지식맵 분석

분야		주요영역(주변 관심 키워드)
10년 전·후 비교	전	Chemical Structure(Molecular structure)/Synthesis, Nanoparticles(nanostructure materials), Electrochemistry
	후	Chemical Structure(Molecular structure), Nanoparticles(nanostructure materials, graphite), Graphene, Carbon(electrontransport), Scanning Electron Microscopy/Transmission Electron Microscope, Solar Cells, Electrical Performance/Lithium Battery, Human(Fluorescent dye, drug delivery systems)
세계 vs. 한국 비교	세계	Electrical Performance(Lithium), Transmission Electron Microscope/Electrochemical Impedance Spectroscopy, X-ray Diffraction, Graphene, Perovskite(Heterojunction), Carbon, Chemical Structure, Nanoparticles, Human(Fluorescent dye, drug delivery systems),
	한국	Lithium Ion Battery/Supercapacitor/Electrodes, Graphene/Perovskite, X-ray Diffraction, Solar Cells, Transmission Electron Microscope, Chemical Structure, Nanoparticles, Human(nonhuman, animals)

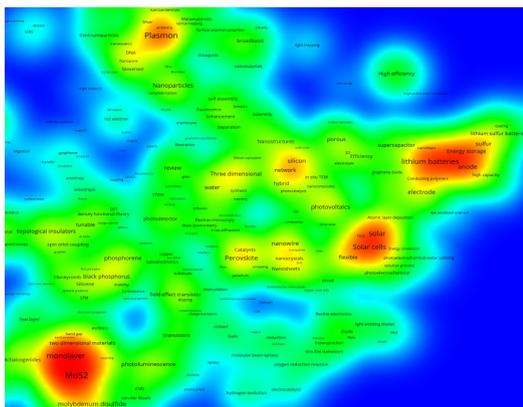


그림 2. 응집물리분야, 10년 간 점유율 증가율(전세계)

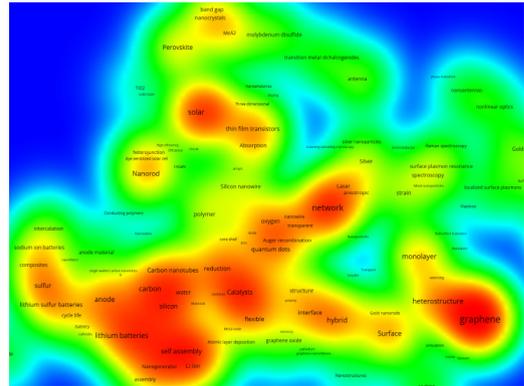


그림 3. 응집물리분야, 10년 간 점유율 증가율(한국)

V. 결론 및 시사점

본 연구에서는 증거기반의 국가 기초연구지원을 위한 정책을 수립하는데 활용하고자 핵심적이고 활용도가 높은 과학기술 통계·정보를 포괄적으로 수록한 DB를 구축하고자 하였다. 이를 위해 국내·외의 49종의 과학기술 관련 보고서 등 자료를 수집하여 6,446개 지표 추출, 877개 통계·정보로 범주화, 이 중 208개를 집중 관리 통계로 도출하였다. 다양한 스케일과 차원의 통계·정보를 통합적으로 수집하여 즉시 활용가능한 통계표 형태로 구축함으로써 정책결정자 등의 합리적 판단과 효율적인 재원 배분을 지원하는데 효과적으로 활용될 수 있을 것으로 기대된다. 이렇게 구축된 DB의 활용성 제시를 위해 논문정보 관련 심층분석을 통해 제한적으로나마 우리나라 기초연구의 효과성을 보여주고자 하였다. 학술논문은 기초연구의 대표적인 결과물로서 기초연구 수행의 핵심주체인 대학의 지식 확산의 대표적인 채널이기도 하다. 질적·양적 학술논문 창출 성과와 R&D 투자 간의 통계 기반 분석과 고피인용 논문의 지식맵 분석을 실시하였다. 첫 번째, 학술논문 성과와 연구비 투자 간 관계에 관한 통계 분석 결과, 기초연구비 증가가 우리나라의 보편적 연구활동도(SCI 논문수) 증대와 세계적 연구성과 창출(고피인용 논문)에 모두 영향을 미친 것으로 나타났다. 또한 양적성과(SCI 논문)와 질적성과(고피인용논문) 간 연계성 분석 결과, 상호

간에 정(+)의 상관관계인 것으로 판별 되었다. 국가총 연구개발비, 정부연구개발비, 기초연구비 중 기초연구비가 질적성과(고피인용 논문) 창출에 가장 큰 영향을 미치는 것으로 분석되었다. 이를 통해, 신정부 국정과제 등을 통해 발표된 연구자 주도 기초연구비 2배 확대 정책 추진의 필요성에 대한 한 가지 증거를 제시할 수 있을 것으로 기대된다. 두 번째, 지난 10년간의 피인용 상위 1% 논문 기반의 지식맵 트렌드 분석 결과, 순수기초 연구분야로 간주되어온 수학, 물리, 화학이 신성장동력이라 불리는 AI, 빅데이터 분석, 바이오 관련 기술, 청정에너지 등 기술 발전의 기반이 되어왔음을 알 수 있었다. 물리와 화학 분야의 경우 청정에너지 관련 기술과 신소재 기술이 부상하는 전세계 트렌드와 한국이 유사한 것으로 나타났다. 이는 순수기초 연구 외에 기술의 발전, 시장의 성숙 등에 환경변화에 따라 상용화가 가능한 응용기술로의 기술융합이 이루어진 결과로 유추된다. 수학의 경우, 전세계는 다양한 알고리즘 기술의 충분한 성숙과 4차 산업혁명 관련 기술의 발전으로 AI, 바이오 관련 분석기술 등의 부상이 뚜렷한 트렌드로 나타났다. 이에 반해, 한국은 전통적인 수학분야가 여전히 강세인 것으로 나타났는데 이는 지식맵 구축의 범위를 피인용 상위 1% 논문에 한정하여 규모는 작으나 점점 증가하고 있는 키워드를 보여주지 못한 것으로 추측된다. 다만 수학 관련 분야는 4차 산업혁명 관련 핵심기술의 견고한 기반을 제공하기 위해 중요한 영역으로 다양한 연구주체의 지원이 필요할 것으로 판단된다.

향후 동 연구를 통해 구축된 기초연구과제 정보, 기초연구과제 수행을 통해 창출된 성과 정보(논문, 특허 등)를 기반으로 구축한 지식맵의 3P분석(Project, Paper, Patent)을 통해 기초연구비 투자에 대한 효과성에 대한 심층분석이 가능할 것으로 기대된다. 지속적인 통계·정보 DB의 현행화와 함께 활용성, 개방성, 접근성 제고를 위한 방안을 모색하고자 한다.

* 본 연구내용은 저자의 개인적인 견해이며 정부, 한국연구재단 등 저자 소속기관의 공식적인 견해와는 다를 수 있습니다.

참 고 문 헌

- [1] 성태웅, 전승표, 변정은, 박현우, “과학기술정보 콘텐츠의 가치평가모형 프레임워크 연구,” 한국콘텐츠학회논문지, 제16권, 제11호, pp.421-433, 2016.
- [2] 장경수, 박귀순, 김해도, 천기우, 기초연구지원 정책 수립을 위한 과학적 통계·정보 분석에 관한 연구, 한국연구재단, 2017.
- [3] 김미정, 이흥권, 최태진, “기초연구의 예산배분체계 및 관리 패러다임 전환에 관한 小考,” 기술혁신학회지, 제14권, 제2호, pp.177-204, 2011.
- [4] 과학기술정보통신부, 과학기술정책연구원, 과학기술 50년사 1편/과학기술의 시대별 전개, 과학기술정보통신부, 2017.
- [5] 장경수, 박귀순, 김해도, 조석민, 황은희, 조영돈, 박숙미, 기초연구의 진단과 미래 과제, NRF ISSUE REPORT 2017-3호, 한국연구재단, 2017.
- [6] 박성욱, “국가 지식정보시스템 개발의 경제적 효과 분석,” 정보관리연구, 제39권, 제3호, pp.73-94, 2008.
- [7] 이우성, 박미형, 김보현, 기초·원천연구 투자의 성과 및 경제적 효과분석, 과학기술정책연구원, 2014.
- [8] E. Dolgin, “NIH grants yield windfall,” Nature, Vol.544, Issue7648, pp.14-15, 2017.
- [9] F. Narin, K. S. Hamilton, and D. Olivastro, “The increasing linkage between U.S. technology and public,” Research Policy, Vol.26, pp.317-330, 1997.
- [10] 이강용, 남궁현, 심재철, 조기성, 류원, “공공분야에서의 빅데이터 활용을 위한 지식자산(Knowledge Base) 구축,” 정보과학회지, 제30권, 제6호, pp.40-46, 2012.
- [11] http://stats.oecd.org/Index.aspx?DataSetCode=MSTL_PUB
- [12] <https://www.ons.gov.uk/economy/governmentpublicsectorandtaxes/researchanddevelopmentexpenditure>

- [13] <https://stats.nrf.re.kr/>
- [14] <http://rmd.nrf.re.kr/>
- [15] <http://www.ntis.go.kr/ThMain.do>
- [16] <http://www.ndsl.kr/>
- [17] <http://www.kistep.re.kr/c3/sub7.jsp>
- [18] National Research Council, *Managing University intellectual property in the public interest*, The National Academies Press, 2010.
- [19] 연승민, 김슬기, 김재수, 이병희, “정부연구비 지원 변동이 기초연구성과에 미치는 영향 분석: IT 분야를 중심으로,” *콘텐츠학회논문지*, 제16권, 제5호, pp.157-171, 2016.
- [20] 교육과학기술부, *2008 연구개발활동조사보고서*, 교육과학기술부, 2008.
- [21] 장금영, “연구개발투자의 성과에 영향을 미치는 요인에 관한 연구 : 정부의 산업기술개발사업을 중심으로,” *기술혁신연구*, 제18권, 제1호, pp.75-98, 2010.
- [22] 장덕희, 강길모, 한동성, 도수관, “지식재산(특허) 생산에 대한 정부연구비 지원의 영향 분석: 우리나라 이공계 전공분야 대학교수들의 연구활동 결과를 중심으로,” *정책분석평가학회보*, 제24권, 제2호, pp.1-26, 2014.
- [22] 최태진, *국가연구개발사업의 유형별 성과분석을 통한 전략적 연구관리체계 구축에 관한 연구*, 건국대학교, 박사학위논문, 2007.
- [23] 이호신, 이일형, 김경호, “기술문헌DB로부터 기술동향분석을 위한 기술정보 분석시스템 개발에 관한 연구,” *한국콘텐츠학회논문지*, 제2권, 제4호, pp.67-75, 2002.
- [24] 이우형, 김윤명, 손성혁, 윤문섭, “연구기획 사전 타당성 분석을 위한 과학기술 지식맵 개발,” *대한산업공학회/한국경영과학회 2004 춘계 학술대회 논문집*, 한국경영과학회, pp.15-18, 2004.
- [25] 이방래, 여운동, 이준영, 이창환, 권오진, 문영호, “계량정보분석시스템으로서의 KnowledgeMatrix 개발,” *한국콘텐츠학회논문지*, 제8권, 제1호, pp.68-74, 2008.
- [26] 이방래, 이준영, 김도현, 노경란, 양명석, 권오진, 최광남, 김한준, “과학기술 지식맵의 형태적 분류와 정보분석 관점의 지식맵 사례 도출,” *한국콘텐츠학회논문지*, 제13권, 제11호, pp.461-476, 2013.
- [27] 이준영, *VOSviewer를 활용한 과학기술 매핑*, 한국과학기술정보연구원 2017.
- [28] N. J. Eck and L. Waltman, “Software survey: VOSviewer, a computer program for bibliometric mapping, *Scientometrics*, Vol.84, pp.523-538, 2010.
- [28] W. M. Sweileh, S. W. Al-Jabi, A. S. AbuTaha, H. Z. Sa’ed, F. M. Anayah, and A. F. Sawalha, “Bibliometric analysis of worldwide scientific literature in mobile-health: 2006 - 2016,” *BMC Medical Informatics and Decision Making*, Vol.17, No.1, p.72, 2017.
- [29] S. E. H. Zyoud, W. S. Waring, W. S. Sweileh, and S. W. Al-Jabi, “Global research trends in lithium toxicity from 1913 to 2015: A bibliometric analysis,” *Basic & Clinical Pharmacology & Toxicology*, 2017.
- [30] Z. Hu, F. Guo, and H. Hou, “Mapping research spotlights for different regions in China,” *Scientometrics*, Vol.110, No.2, pp.779-790, 2017.
- [31] 전세림, 정은희, 조석민, 이광희, 장경수, *2005-2015 주요국의 피인용 상위 1% 논문실적 비교분석 보고서*, 한국연구재단, 2016.
- [32] SAS Institute (2015). *Base SAS 9.4 Procedures Guide*. SAS Institute.
- [33] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*, John Wiley & Sons, 2012.
- [34] L. Leydesdorff, “Science Visualization and Discursive Knowledge,” in B. Cronin, and R. Sugimoto, eds. *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, pp.167-185, The MIT Press, 2014.
- [35] L. Leydosdorff and A. Nerghe, “Co-word Maps and Topic Modeling: A Comparison Using

Small and Medium-Sized Corpora (N< 1,000),” Journal of the Association for Information Science and Technology, Vol.68, No.4, pp.1024-1035, 2017.

[36] M. Callon, J. Law, and A. Rip, eds., *Mapping the Dynamics of Science and Technology*, The Macmillan Press Ltd., 1986.

[37] N. J. van Eck and L. Waltman, “Software survey: VOSviewer, a computer program for bibliometric mapping,” *Scientometrics*, Vol.84, pp.523-538, 2010.

저 자 소 개

박 귀 순(Kwisun Park)

정회원



- 2001년 : 울산대학교 환경공학 (공학사)
- 2009년 : Old Dominion Univ., USA(토목공학 석사 및 박사)
- 2009년 ~ 현재 : 한국연구재단 정책연구혁신센터 선임연구원

<관심분야> : 과학기술정책 및 기후변화대응 전략 수립, 국가연구개발사업 지원 체계 구축 및 효과성 분석

석 혜 은(Hyeeun Seok)

정회원



- 1998년 : 이화여대 통계학(이학 석사)
- 2016년 : 이화여대 심리학(심리 측정 및 통계) 석박사통합 박사 수료
- 2016년 ~ 현재 : 이화여대/충남대

시간강사

<관심분야> : 회귀분석, 비선형모형, 통계전산, 데이터 모델링, 베이지안

박 진 서(Jinseo Park)

정회원



- 1995년 : KAIST 경영과학(공학사)
- 1998년 : 고려대학교 과학학 협동과정(과학관리학, 석사)
- 2008년 : 고려대학교 과학학 협동과정(과학관리학, 박사)

▪ 2002년 ~ 현재 : KISTI 과학계량연구실 선임연구원
<관심분야> : 과학기술 거버넌스, 지식네트워크의 진화, 과학기술과 사회(STS)

김 해 도(Haedo Kim)

정회원



- 1991년 2월 : 건국대학교 공업화학(학사)
- 2002년 8월 : 충남대학교 특허법무대학원(석사)
- 2006년 8월 : 충남대학교 특허협동과정(박사)

▪ 1994년 1월 ~ 현재 : 한국연구재단 책임연구원
<관심분야> : 과학기술정책, 특허법, 기술사업화 및 창업