

# 소셜 빅데이터 기반 국제연구협력 트렌드 분석\*

노영희, 남윤서†

건국대학교 문헌정보학과, 지식콘텐츠연구소

본 연구는 국제연구협력의 트렌드를 분석하기 위해 소셜빅데이터를 수집하여 현황을 파악하고 주요 토픽을 도출하고자 하였다. 이를 위해 TEXTOM을 활용하여 구글, 네이버, 다음 등의 주요 포털 사이트 내 2018년부터 2022년 현재까지의 국제연구협력 관련 소셜 빅데이터를 수집하였다. 수집된 데이터는 텍스트 마이닝 과정을 통해 정제된 단어를 도출하였고 이를 바탕으로 빈도분석 및 중심성 분석, N-gram 분석, 매트릭스 분석, 토픽 분석 등을 수행하였다. 연구결과, 국제연구협력이 연구소와 대학을 중심으로 활성화되어 있으며 COVID-19 사태에 따라 백신에 관한 연구협력이 집중적으로 이루어졌음을 확인하였다. 그 외 스포츠, 경제, 교육, 산림, 과학 등의 연구주제가 상위 키워드로 도출되었다. 현재 국제연구협력 내 주요 담론은 센터를 중심으로 과학 분야의 공동연구에 관한 담론, 대학 중심의 학술적 국제연구협력 활동, 국내 산업체 단위의 기술개발 협력 활동, 세계경제에 관한 국제적 학술교류에 관한 담론 등으로 나타났다. 주요 토픽으로는 범죄방지, 백신, 바이오, 세계경제 등의 연구주제에 대한 세부적인 국제연구협력과 정보와 정책지원 측면에서의 토픽이 도출되었다. 이러한 연구 결과는 국제연구협력 내 핵심 키워드를 살펴보고 현재 국제연구협력 내 형성된 담론과 주제를 살펴봄으로써 향후 국제연구협력의 주요 쟁점을 예측하는데 유용한 자료로 사용될 수 있을 것으로 보인다. 또한 협력주체와 국가간의 네트워크 분석을 통해 현재 활성화되어 있는 분야와 그렇지 못한 분야를 확인하여 보다 심층적인 국제연구협력 지원방안을 수립하기 위한 기초연구로서 활용될 수 있을 것으로 보인다.

**주요어:** 국제연구협력, 빅데이터 분석, 소셜빅데이터, 텍스트마이닝, LDA 토픽 모델링

\* 본 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2017S1A5B4055679).

† 교신저자(Corresponding Author) : 남윤서, 건국대학교 지식콘텐츠연구소 연구원, 충북 충주시 충원대로 268 교수연구동 209호, E-mail : hongbi0810@kku.ac.kr

노영희, 건국대학교 문헌정보학과 교수(제1저자)

■ 최초투고일 : 2022년 8월 16일 ■ 심사마감일 : 2022년 9월 28일 ■ 게재확정일 : 2022년 10월 19일

## 1. 서론

최근 급속한 글로벌 연구환경 변화에 대응하고 국가경쟁력을 선점을 위해 미국을 중심으로 EU 및 일본 등은 전략적이고 체계적인 R&D 국제협력력을 추진하고 있다. 이러한 추세에 따라 세계 각국은 사회, 과학, 경제, 정치 등의 다양한 분야에서 국제협력력을 함으로써 다양성 및 전문성을 강화하고자 노력하고 있다.

우리나라 정부는 2010년 OECD 개발원조위원회(DAC) 가입 이후 국제개발협력기본법에 따라 2차례의 국제개발협력 기본계획(1기: 2011-2015, 2기: 2016-2020)을 수립·추진하였다. 이를 통해 ODA 정책의 효과성 및 효율성 제고를 위해 지속적인 개선을 추진해왔다. 그러나 분절화, 개발효과성 미흡, 성과평가 체계의 미숙 등 다양한 문제점들이 노정되고 있다(김수현, 손욱, 2021). 더불어 과학기술 경쟁력을 강화하기 위해 전략적이고 체계적인 국제협력 및 공동연구의 추진에 대한 중요성이 강조되고 있지만 글로벌 R&D 비중은 여전히 낮은 실정이다. 한국과학기술정보연구원(2018)에 따르면, 최근 3년간 출연(연) 25개소에 투입된 국제협력비는 연평균 40.9%로 감소 추세이며 국제공동연구 과제수도 연평균 21.4% 감소하고 있다. 또한 전체 연구사업비로 보았을 때, 국제협력비 비중은 최근 3년간 8.04%에서 3.54%로 감소하였다.

이러한 국제연구협력력을 효과적으로 지원하기 위해서는 기존 국제연구협력력의 동향을 살펴 볼 필요가 있다. 따라서 본 연구에서는 소셜 빅데이터를 활용하여 국제연구협력력에 관한 데이터를 수집하고 분석함으로써 국제연구협력력의 동향을 파악하고자 하였다. 먼저, 수집된 비정형 소셜 데이터를 텍스트마이닝 기법을 통해 데이터를 정제하고 단어빈도 분석과 TF-IDF 분석을 수행하여 핵심

키워드와 주요 키워드를 도출하였다. 다음으로 감성사전을 기반으로 국제연구협력에 대한 소셜네트워크상의 인식을 파악하기 위해 감성어를 추출하여 분석하고, CONCOR 분석을 통해 국제연구협력 내 담론을 파악하고자 하였다. 이후 Word-level Semantic Clustering 분석과 LDA 토픽 모델링을 통해 신기술 분야 산학협력 네트워크 내에서 가장 이슈화되고 있는 주제를 파악하였다. 다음으로 개체명인식 분석을 통해 국가 및 협력기관별 현황을 파악하고 각 국가 및 기관별 국제연구협력 네트워크를 분석하였다.

본 연구는 국제연구협력 내 핵심 키워드를 살펴보고 현재 국제연구협력 내 형성된 담론과 주제를 살펴봄으로써 향후 국제연구협력의 주요 쟁점을 예측하는데 유용한 자료로 사용될 수 있을 것으로 보인다. 또한 협력주체와 국가간의 네트워크 분석을 통해 현재 활성화되어 있는 분야와 그렇지 못한 분야를 확인하여 보다 심층적인 국제연구협력 지원방안을 수립하기 위한 기초연구로서 활용될 수 있을 것으로 보인다.

## 2. 선행 연구

국제연구협력력의 동향을 파악하고 활성화 방안을 마련하기 위해 다양한 방식으로 국제연구협력력을 분석하고자 하는 연구가 수행되고 있다. 국제연구협력에 대한 동향을 분석한 연구 중 김수현과 손욱(2021)은 1997년부터 2018년까지의 『OECD 국제개발협력보고서』를 대상으로 텍스트 데이터를 구성하고 인공지능(AI)을 이용하여 지난 20여년간의 국제개발협력력의 키워드를 분석하였다. 또한 국제개발협력력의 목표로 선언된 SDG의 17개 목표를 대상으로 어느 목표가 더욱 중시되고 있는지 추세의 변화를 파악하고 이를 통해 목표의 가

중치를 시산하고자 하였다. 연구결과, 국제개발협력의 키워드는 지속적으로 변화하여 왔으며, 특히 2010년대에 들어 기후변화와 환경 보존적 개발원조가 부각되는 가운데 민간의 투자 촉진과 개발 투자 유인에 대한 논의가 증시되는 경향을 보이고 있음을 확인하였다. 텍스트마이닝 결과 기후변화 대응이 가장 중요한 것으로 나타났으며, 글로벌 파트너십, 빈곤종식, 건강한 삶 및 웰빙, 성평등 및 여성 자력화, 포용적 양질의 교육, 양질의 일자리와 경제성장이 그 다음으로 중요한 것으로 나타났다.

주한나 등(2020)은 ICT분야의 국제개발협력 연구동향을 분석하기 위해 2002년부터 2020년간의 학술논문을 체계적으로 분석하였다. KCI와 국회도서관 데이터베이스 내에 개발도상국의 경제사회적 발전이라는 맥락에서 작성된 논문들을 추출하고자 하였다. 연구 결과, ICT 관련 국내 연구는 2010년대 이후 증가 추이를 보였으나 여전히 양적, 질적 성장이 요구되며, 지역 및 주제의 다양화, 수용도 연구의 확대가 필요한 것으로 나타났다. 나아가 ICT ODA가 독자적인 학문 분야로 발전하기 위해 이론적 토대에 기반을 둔 연구가 강화되어야 하며, 전문 학술공동체를 비롯한 체계적인 연구 인프라가 구축되어야 함을 확인하였다.

한재광(2021)은 2001~2020년 20년간 한국에서 발표된 국제개발협력을 주제로 시민사회와 관련한 논문의 현황을 밝히고, 그 결과가 일정 시기에 따라 의미있게 변화했는지를 분석하였다. 이를 위해 2001년에서 2020년까지 한국연구재단 등재/등재후보 학술지에 게재된 63편의 논문을 대상으로 체계적문헌고찰 방식을 활용해 논문 수, 주제, 내용, 학문분류, 작성자, 기존 연구결과 활용도를 기준으로 분석했고, 효과적인 내용분석을 위해서는 단어빈도분석을 실시하였다. 한국 국제개발협력 시민사회 관련 연구는 논문 및 연구자 수가 작

다는 한계를 가지고 있으나, 20여 년 동안 논문 수는 점차 증대하고, 연구주제와 내용도 다양해지고 있다고 밝혔다.

노영희, 이광희, 장로사(2019)는 국제연구협력 관련 논문 총 101편을 대상으로 국가별 비교분석을 수행하여 우리나라의 국제연구협력 실태를 살펴보고, 향후 활성화 방안을 제안하였다. 연구 결과, 국제연구협력 주제 관련 논문을 출판한 국가는 총 20개국으로, 순위는 영국(27편, 26.7%), 네덜란드(20편, 19.8%), 미국(19편, 18.8%), 브라질과 스페인(5편, 5.0%), 러시아(4편, 4.0%) 등의 순으로 나타났다. 한국은 하위권(2편, 2.0%)으로서 아직까지 국제연구협력 수준이 저조한 것으로 밝혀졌다. 이에 향후 대학의 연구자들, 연구 분야별로 국제 연구협력 주제 관련 논문을 활성화시킬 수 있는 사업을 확대함으로써 한국의 국제적인 연구역량을 강화해야한다고 언급하였다.

한편 노영희, 박우정(2021)은 국제연구협력 관련 참고자료 정보원 데이터베이스를 구축하여 학회지, 사전편람, 법령자료 등 연구에 기반이 되는 정보원을 분석하여 국제연구협력 현황을 파악하고자 하였다. 학회지는 교육, 문학, 국제문제, 과학 등을 중심으로 발간되고 있으며, 2000년대 초반 이후부터 꾸준히 학회지가 창간되고 있는 것을 파악되었다. 법령자료의 경우 국내를 제외하고 미국, 일본 등의 현황이 다수 나타났으며, 규정, 국제, 법령등에 관한 주제어가 높게 나타났으며 발행기관은 UN, 과학기술정보통신부, 외교부 등으로 높게 나타났다. 연구 결과, 국제연구협력의 지원 사업과 연구 내용이 보편적으로 연관성을 보였으며 보고서 키워드 분석 결과에서도 유사성을 가지고 있음을 알 수 있다. 따라서 보다 다양한 주제의 국제연구협력이 수행되기 위해서는 이를 지원하는 사업 역시 다양화되어야 할 것이라 제안하였다.

국제연구협력을 주제로 논문, 보고서 등을 바탕으로 연구동향을 분석한 논문은 많지만, 소셜 빅데이터를 활용하여 축적된 데이터를 분석한 연구는 부족하였다. 따라서 본 연구에서는 국제연구협력을 주제로 작성된 모든 정보원의 텍스트 문서를 수집하여 총체적인 국제연구협력의 분석을 수행하고자 한다.

### 3. 연구설계 및 방법론

#### 1) 연구 설계 및 내용

본 연구에서는 국제연구협력 동향을 파악하기

위해 빅데이터 분석의 대표적인 기술인 텍스트 마이닝(Text-mining)을 통해 단어빈도 분석과 TF-IDF 분석을 실시하였다. 단어빈도 분석을 통해 전체 데이터 중 등장 빈도가 높은 단어를 파악하여 핵심 키워드를 선정하고, TF-IDF 분석을 통해 해당 데이터의 중요도에 따라 단어를 나열하였다. 이후 키워드 간 연관성을 파악하기 위해 N-gram 분석을 통해 도출된 키워드 간의 네트워크를 확인하였다.

다음으로 협력주체별 국제연구협력 동향을 분석하기 위해 개체명인식 분석을 통해 기관명을 도출하여 기관유형별 분석을 수행하고, 2-mode 매트릭스 분석을 통해 각 협력주체별 국제연구협력 네트워크를 도출하였다. 다음으로 국가별 국제

단계	연구내용	세부연구내용
1단계 선행연구	관련 문헌 조사	<ul style="list-style-type: none"> <li>국제연구협력 빅데이터 관련 문헌 분석</li> </ul>
2단계 데이터 수집	데이터 수집	<ul style="list-style-type: none"> <li>TEXTOM을 활용하여 국제연구협력과 신기술 분야 소셜 빅데이터 수집</li> </ul>
3단계 데이터 분석	텍스트마이닝을 통한 국제연구협력 키워드 분석	<ul style="list-style-type: none"> <li>단어빈도 분석을 통한 신기술 분야 국제연구협력 핵심 키워드 도출</li> <li>TF-IDF 분석을 통한 신기술 분야 국제연구협력 중요도 도출</li> <li>N-gram 분석을 통한 연쇄단어 네트워크 도출</li> </ul>
	협력주체별 국제연구협력 동향 분석	<ul style="list-style-type: none"> <li>개체명인식 분석을 통한 협력주체별 동향 분석</li> <li>2-mode 매트릭스 분석을 통한 협력주체별 네트워크 도출</li> </ul>
	국가별 국제연구협력 동향 분석	<ul style="list-style-type: none"> <li>개체명인식 분석을 통한 국가별 동향 분석</li> <li>2-mode 매트릭스 분석을 통한 협력주체별 네트워크 도출</li> </ul>
	감성 및 담론 분석	<ul style="list-style-type: none"> <li>감성 분석을 통한 신기술 분야 국제연구협력 내 감성인식 도출</li> <li>CONCOR 분석을 통한 신기술 분야 간 국제연구협력 담론 그룹화 도출</li> </ul>
	토픽 분석	<ul style="list-style-type: none"> <li>LDA 토픽 모델링을 통한 주요 토픽 도출</li> <li>Word-level Semantic Clustering 분석을 통한 연구주체별 토픽 도출</li> </ul>
3단계 결과 분석 및 결론	논의	<ul style="list-style-type: none"> <li>데이터 분석 결과를 바탕으로 유의미한 결과 도출 및 논의</li> </ul>
	결론 및 제언	<ul style="list-style-type: none"> <li>연구결과를 기반으로 향후 연구를 위한 개선방안 및 시사점 제언</li> </ul>

<그림 1> 연구내용 및 절차

연구협력 동향을 분석하기 위해 개체명인식 분석을 통해 국가명을 도출하여 상위 5개국을 선정하고, 1-mode 매트릭스 분석을 통해 각 협력 주체별 국제연구협력 네트워크를 도출하였다. 이후 감성사전을 기반으로 문서 내 인식 및 감성을 파악하는 감성분석을 수행하여 소셜 네트워크상의 인식과 평가를 확인하고, 상관관계 분석을 반복적으로 수행하여 적절한 수준의 유사성 집단을 찾아내는 CONCOR 분석을 통해 담론을 도출하였다. 이후 LDA 토픽 모델링을 통해 문서 데이터 내의 핵심 주제를 도출하여 전체 신기술 분야 국제연구협력 활동 중 주요 토픽을 분석하고, 문서 내 단어들의 공출현 관계를 기준으로 군집화하는 방법인 Word-level Semantic Clustering을 통해 각 신기술 분야 내 주요 토픽을 도출하였다.

## 2) 데이터 분석 방법 및 내용

데이터를 수집하는 채널로 네이버, 다음, 구글을 선정하였으며, 자료검색을 위한 검색어는 ‘국제연구협력’을 선정하였다. 자료 분석 기간은 현재를 기준으로 5년간의 동향을 총체적으로 살펴보기 위해 2018년 1월 1일부터 2022년 6월 30일까지의 데이터를 수집하였다.

본 연구에서는 국제연구협력 트렌드를 알아보기 위해 소셜 빅데이터를 수집하고 목적에 따라

텍스트마이닝 분석, 매트릭스 분석, 토픽 분석을 수행하였다. 우선 국제연구협력 활동 양상을 파악하기 위해 단어빈도 분석과 TF-IDF 분석을 수행하여 상위 핵심 키워드를 도출하였다. 단어빈도 분석은 전체 데이터 내 빈출 키워드를 도출하고자 수행하였으며 단순 빈출도를 통해 파악하기 어려운 중요 키워드를 도출하기 위해 문서 내의 단어 가중치를 계산하는 방식인 TF-IDF 분석을 수행하였다. N-gram 분석을 통해 키워드별 연쇄단어를 도출하고, 개체명인식 분석을 통해 협력주체 및 국가별 키워드를 도출하였다. 이후 국가 간 연관성을 파악하기 위해 키워드의 동시출현관계(keyword co-occurrence matrix)를 파악하여 키워드 간 구조적 배열을 수치로 산출한 1-mode 매트릭스 분석을 수행하여 각 노드 간의 연관관계를 네트워크로 시각화하였다. 이후 협력주체별 국제연구협력 활동을 알아보기 위해 2-mode 매트릭스를 활용하여 각각 행과 열을 구성하여 국제연구협력 활동 네트워크를 도출하였다. 감성 분석을 통해 문서 내 감성단어를 분석하고 CONCOR 분석으로 담론을 그룹화하여 도출하였다. 이후 LDA 토픽모델링과 Word-level Semantic Clustering 분석 기법을 현재 국제연구협력 내 주요 토픽과 각 분야별 토픽을 도출하여 분석을 수행하였다.

<표 1> 분석데이터 정보

구분	내용
수집도구	TEXTOM
수집범위	네이버(웹문서, 블로그, 뉴스, 카페, 지식인, 학술정보전체) 다음(웹문서, 블로그, 뉴스, 카페) 구글(웹문서, 뉴스, 페이스북)
수집기간	2018년 1월 1일 ~ 2022년 6월 30일
검색어	국제연구협력

&lt;표 2&gt; 분석 구조도

구분	분석방법	내용
텍스트 마이닝 분석	단어빈도 분석	국제연구협력 데이터 내 핵심 키워드 도출
	TF-IDF 분석	국제연구협력 데이터 내 중요 키워드 도출
	N-gram 분석	신기술 분야 국제연구협력 데이터 내 연쇄단어 네트워크 도출
	개체명인식 분석	기관명, 지역명 분석을 통한 협력주체 및 국가별 키워드 도출
매트릭스 분석	1-mode 매트릭스 분석	국가별 국제연구협력 네트워크 도출
	2-mode 매트릭스 분석	협력주체별 국제연구협력 네트워크 도출
감성 및 담론 분석	감성분석	국제연구협력 내 감성인식 도출
	CONCOR 분석	국제연구협력 담론 그룹화 도출
토픽 분석	LDA 토픽 모델링 분석	국제연구협력 활동 내 주요 토픽 도출
	Word-level Semantic Clustering	국제연구협력 활동 내 연구주체별 토픽 도출

## 4. 연구결과

### 1) 텍스트마이닝을 통한 국제연구협력 키워드 분석 결과

#### (1) 단어빈도 분석을 통한 국제연구협력 핵심 키워드 도출

국제연구협력의 동향을 파악하기 위해 단어빈도 분석을 수행하여 신기술 분야와 관련한 상위 핵심 키워드를 도출하였다. 선정된 상위 30개 핵심 키워드의 빈도 및 백분율을 보면 <표 3>과 같다. 텍스트마이닝을 통해 수집한 체언 중 관련, 중, 등, 것 등 특정한 의미를 내포하지 않는 의존명사를 제외하고, 검색어로서 필연적으로 상위를 차지할 수 있는 국제, 연구, 협력 키워드를 제외하여 키워드를 정제하였다. 분석 결과, 국제연구협력에 관한 소셜 빅데이터 내에서 가장 많이 언급된 단어는 개발, 기술, 한국, 사업, 대학 등으로 나타났다. 이 중 유의미한 지표는, 연구주체에 해당하는 키워드인 백신, 경제, 과학, 교육, 스포츠 등이 상위 30위 내에 위치하면서 해당 주제가 국제연구협력이 가장 활발히 일어나고 있는 연구분야임을 알 수 있다.

#### (2) TF-IDF 분석을 통한 국제연구협력 주요 키워드 도출

단어빈도 분석은 모든 데이터 내의 단어를 단순 분석하는 방법으로, 여러 문서 내에서 중요도를 판단하기에는 어렵다는 단점이 존재한다. 일반적으로 TF값이 높을 경우 문서에서 해당 단어가 중요하다고 인지할 수 있으나, 분석하고자 하는 모든 문서에서 자주 나타나는 문서 빈도가 높은 단어일 수도 있다. 이처럼 단순 빈도에 의존하여 단어의 중요도를 분석할 때 모든 문서에서 자주 출현하는 상투어 등 발생할 수 있는 문제점을 보완하기 위해 단어 가중치 기법이 제안되었는데(김진호 외, 2017), 본 연구에서는 단어 가중치 기법 중에서 가장 많이 알려진 TF-IDF 분석을 수행하였다. 그 결과, 국제연구협력에서 가장 중요한 단어로 개발, 기술, 백신, 연구소, 사업, 지원, 한국, 기관, 센터 등이 나타났다. 앞선 단어빈도분석 결과와 대조되는 점은, 빈도분석결과 14위에 위치하던 백신이 TF-IDF 분석에서는 3위에 위치하면서 COVID-19 사태에 따라 백신(3757.925)에 관한 연구 협력이 상대적으로 활발히 일어났음을 확인하였다. 이를 뒷받침하는 다른 근거로 단어빈도 분석 시 상위권에 들지 못한 코로나(2077.533)

<표 3> 단어빈도분석을 통한 국제연구협력 상위 30개 핵심 키워드

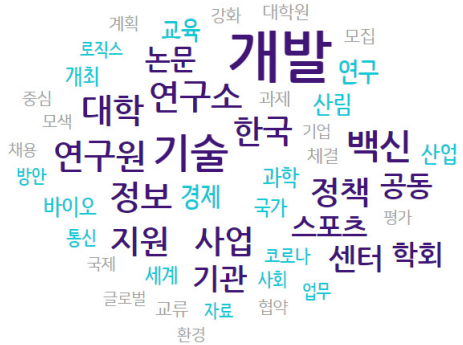
no.	키워드	N	%
1	개발	5556	2.033%
2	기술	2631	0.963%
3	한국	2171	0.794%
4	사업	1956	0.716%
5	대학	1936	0.708%
6	연구소	1813	0.663%
7	지원	1760	0.644%
8	기관	1666	0.610%
9	연구원	1514	0.554%
10	센터	1473	0.539%
11	공동	1431	0.524%
12	정보	1345	0.492%
13	정책	1320	0.483%
14	백신	1144	0.419%
15	경제	1136	0.416%
16	산업	1095	0.401%
17	과학	1091	0.399%
18	교육	1053	0.385%
19	개최	995	0.364%
20	국가	990	0.362%
21	논문	989	0.362%
22	세계	929	0.340%
23	학회	907	0.332%
24	사회	808	0.296%
25	방안	770	0.282%
26	교류	751	0.275%
27	업무	724	0.265%
28	스포츠	699	0.256%
29	자료	695	0.254%
30	강화	689	0.252%

가 TF-IDF 상위 30개 주요 키워드로서 도출된 것을 들 수 있다. 또한 연구소(3582.145), 대학(2101.915)이 상위권으로 도출되어 국제연구협력은 학계 및 연구계를 중심으로 이루어지고 있음을 시사한다. 이후 15위에 정보(3522.436)가 도출된 결과는 국제연구협력을 원하는 수요자들이 정보를 찾고, 이용하고 있음을 유추해볼 수 있다. 정

보 이전의 키워드들은 협력의 주체, 협력 방식, 연구 주제 등과 직접적인 관련이 있는 검색어이나 정보라는 키워드가 상위권에 도출된 것은 국제연구협력 정보에 대한 관심이 나타나는 상당히 주요한 지표로 보인다. 잇달아 사업(3379.693), 정책(2773.867), 지원(3379.693)이 주요 키워드로 도출되면서 국제연구협력의 지원을 위한 전반의 움직임

<표 4> TF-IDF분석을 통한 국제연구협력 상위 30개 주요 키워드

no.	키워드	TF-IDF	no.	키워드	TF-IDF
1	개발	5929.881	16	논문	2866.377
2	기술	4369.783	17	연구	2861.982
3	백신	3757.925	18	정책	2773.867
4	연구소	3582.145	19	경제	2752.778
5	대학	3539.345	20	교육	2659.661
6	사업	3379.693	21	산림	2559.487
7	지원	3363.216	23	산업	2457.023
8	한국	3359.9	24	과학	2445.357
9	기관	3148.576	25	개최	2341.574
10	센터	3124.048	26	국가	2239.923
11	연구원	3020.244	27	세계	2152.331
12	학회	2933.039	28	코로나	2077.533
13	공동	2929.593	29	사회	2035.569
14	스포츠	2885.966	30	자료	2034.387
15	정보	2882.269	30	업무	1956.658



이 일어나고 있으며 정책에 대한 필요성을 바탕으로 구체화되고 있음을 예측할 수 있다. 이 외로 스포츠(2885.966), 경제(2752.778), 교육(2659.661), 산림(2559.487), 산업(2457.023), 과학(2445.357) 등의 연구 분야를 유추할 수 있는 키워드가 도출되어 관련 분야의 국제협력연구가 진행되고 있음을 확인하였다.

**(3) N-gram 분석을 통한 키워드별 영향 관계 분석 결과**

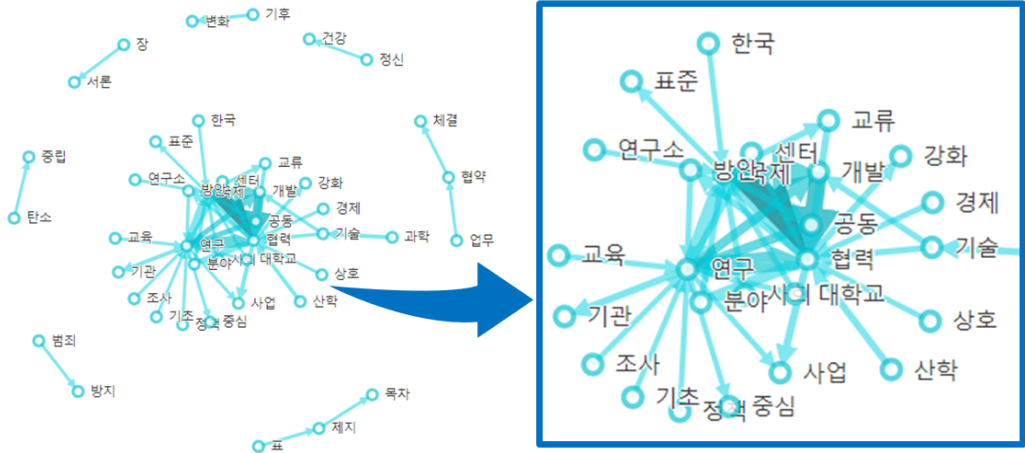
N-gram 분석은 문장 구조 분석을 통해 연속된 단어 간의 관계를 알고자 할 때 주로 활용되는데, 특정 명사가 전체 텍스트 안에서 어떤 명사와 함께 쓰였고 영향관계가 어떠한지 알아보하고자 할 때 사용되는 분석 기법이다. 본 연구에서는 신기술 분야 산학협력에 관한 전체 소셜 빅데이터를 대상으로 N-gram 분석을 수행하여 키워드 간 연관성을 파악하고자 하였다. <표 5>의 분석 결과를 보면, 국제, 연구, 협력, 개발, 공동, 기관, 사

업 등의 단어가 강한 연쇄빈도를 가지고 상호 연관 언급되는 것으로 나타난다. 이후 백신은 연구소와 높은 횟수로 연쇄적으로 도출되었으며, 다시 국제는 백신과 연결되는 결과를 나타내었다. 이는 국제백신연구소(IVI)가 국내 정부부처, 유관기관 등과의 협약 및 교류를 맺은 지난 2년의 결과가 반영된 것으로 유추할 수 있다. 20, 21위에 위치한 바이오, 로직스, 유의 키워드는 지난 20년 국제백신연구소와 유바이로직스가 코로나19 백신 임상개발 협력을 위해 양해각서(MOU)를 체결한 사건이 반영되었음을 나타낸다. 17, 18위에서는 정보, 통신, 정책, 연구원이 이어서 높은 연쇄성을 띄고 있는데, 국내외 ICT 정책연구 협력 네트워크를 구축하고자 다양한 연구를 수행하는 정보통신정책연구원(KISDI)에 대한 결과인 것으로 예상된다. 다음의 결과로 국제연구협력은 산업체와 학계가 함께 하는 산학협력의 형태로도 이뤄지고 있으며, 국제연구협력 분야의 협력정보에 대한 수요와 공급이 일어나고 있음을 유추할 수 있었다.



<표 5> N-gram 분석을 통한 상위 30개 연쇄 단어

no.	단어1	단어2	N	no.	단어1	단어2	N
1	국제	협력	3832	16	기술	협력	469
2	개발	협력	2793	17	정보	통신	452
3	국제	개발	2420	18	정책	연구원	450
4	연구	협력	1544	19	바이오	로직스	444
5	협력	연구	1472	20	유	바이오	444
6	국제	연구	1243	21	산학	협력	438
7	연구	개발	795	22	협력	국제	437
8	공동	연구	664	23	협력	정보	431
9	연구	기관	590	24	연구	국제	411
10	협력	사업	547	25	협력	방안	390
11	백신	연구소	540	26	협력	학회	363
12	국제	백신	537	27	범죄	방지	351
13	연구	논문	532	28	분야	국제	350
14	과학	기술	523	29	협약	체결	345
15	국제	공동	480	30	기술	개발	336



2) 협력주체별 국제연구협력 동향 분석 결과

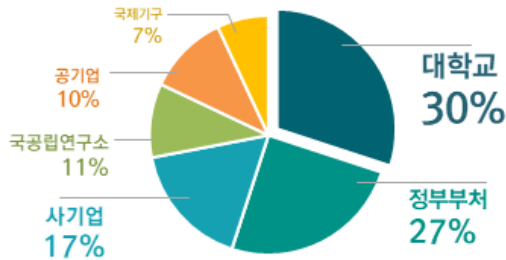
(1) 개체명인식 분석을 통한 국제연구협력 동향 분석

국제연구협력의 협력주체별 활동 양상을 파악하기 위해 개체명인식 분석을 수행하여 유관한 상위 핵심 키워드를 도출하였다. 선정된 핵심키워드별 분석 결과를 나타내면 <표 3>과 같다. 도출

한 기관을 유형별로 분류해보았을 때, 대학교, 정부부처, 사기업·공기업, 국공립연구소, 국제기구로 분류해볼 수 있다. 분석 결과, 국제연구협력을 주도적으로 수행하는 주체는 대학 기관이며 이후로 정부부처, 사기업, 국공립연구소, 공기업, 국제기구 순으로 나타났다. 이는 대학 내 국제협력을 위한 부설기관이 설립된 경우가 많아, 함께 언급된 빈도가 많은 까닭으로 보여진다. 이후 정부부

<표 6> 개체명인식 기관명 분석 결과 및 비율

no.	키워드	N	no.	키워드	N
1	정부	365	16	한국은행	25
2	서울대학교	114	17	국립보건연구원	25
3	유네스코	94	18	교육부	25
4	고려대학교	94	19	한국연구재단	24
5	유엔	84	20	성균관대학교	24
6	경희대학교	82	21	삼성전자	23
7	현대제철	75	22	제주대학교	22
8	금오공대	72	23	해군사관학교	21
9	연세대학교	72	24	부경대학교	21
10	삼성	64	25	법무부	21
11	인천대학교	51	26	통일부	21
12	교통안전공단	34	27	한국기계연구원	20
13	한국선급	28	28	외국어대학교	20
14	대우	27	29	해양수산부	20
15	한화	26	30	외교부	20



치는 국제연구협력을 수행하는 주체이자 지원하는 협력기관으로서 언급되는 경우가 많은 것을 고려하여 많은 비율을 차지하는 원인을 유추할 수 있다.

**(2) 매트릭스 분석을 통한 협력주체별 국제연구협력 네트워크 분석**

국제연구협력의 협력주체별 활동 양상을 파악하기 위해 개체명인식 분석을 수행하여 유관한 상위 핵심 키워드를 도출하였다. 분석 결과, 대학이 가장 활발한 국제연구협력 활동을 수행하고 있으며 특히 사업, 협약, 논문의 키워드에서 가장

높은 연관성을 보이고 있어 유관 사업 수행, 타 기관과의 협약 체결, 논문을 바탕으로 학술 연구를 주로 수행하고 있음을 알 수 있다. 정부부처의 경우 전체적인 네트워크에서 상대적으로 연결성이 미비하나, 이는 직접적인 연구를 수행하는 것이 아닌 관련 연구 및 사업을 진행하여 지원하는 기관인 까닭으로 보인다. 기업의 경우 기술, 개발과 가장 높은 연관성을 보이고 있으며, 국공립연구소가 가장 다양한 유형의 협력을 고루 진행하고 있는 것으로 나타났다. 이러한 협력주체는 대체적으로 연구, 기술, 개발, 학술교류(세미나, 학술대회)순으로 협력을 수행하는 것으로 나타났다.

<표 7> 매트릭스 중심성분석을 통한 협력주체별 네트워크

구분	대학교	정부부처	기업	국공립연구소
연구	2284	600	829	2264
개발	721	232	297	866
기술	321	171	396	681
사업	537	134	225	294
협약	237	11	44	194
세미나	173	9	2	171
논문	105	17	19	30
인력	104	23	32	97
학술대회	38	2	4	84

대학의 국제연구협력 네트워크	정부부처의 국제연구협력 네트워크
기업의 국제연구협력 네트워크	국공립연구소의 국제연구협력 네트워크

### 3) 국가별 국제연구협력 동향 분석 결과

#### (1) 개체명인식 분석을 통한 국제연구협력 동향 분석

국제연구협력의 협력주체별 활동 양상을 파악하기 위해 개체명인식 분석을 수행하여 유관한 상위 핵심 키워드를 도출하였다. 한국을 제외하고 높은 빈도를 보인 국가는 미국, 중국, 북한, 유럽연합, 일본으로 국가 간의 국제연구협력의 비율을 확인할 수 있다.

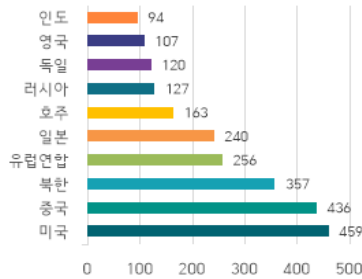
#### (2) 매트릭스 분석을 통한 상위 5개국 국제연구협력 네트워크 분석

상위 5개국의 매트릭스 분석 결과, 한국, 미국, 중

국, 일본, 북한, 유럽 순으로 높은 연결중심성을 보였다. 이 중 북한의 경우, 수집된 원문데이터 검토를 통해 상호연구협력 및 교류보다 대북협력의 차원에서 지원 및 개발협력에 대한 연구가 국내에서 중점적으로 진행되고 있음을 확인하였으며 이러한 결과는 네트워크 시각화의 일방향 화살표 도식을 통해 드러난다. 위세중심성은 해당 국가와 연결된 다른 국가들의 중심성까지 고려하여 도출된 지수로 동일하게 한국, 미국, 중국, 일본, 북한, 유럽 순으로 나타나, 타국과의 협력관계가 얼마나 중요한지 확인할 수 있다. 매개중심성은 다른 노드와의 연결을 위해 의존성이 높은 노드를 의미하는데, 본 국제협력 관계에서는 각국 간에 선행관계 또는 의존관계가 존재하지 않으므로 동일한 수치가 도출되었다.

<표 8> 개체명인식 국가명 분석 결과 및 비율

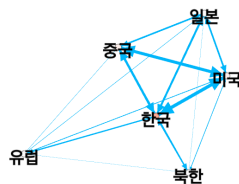
no.	키워드	N	no.	키워드	N
1	한국	1987	16	아프리카	74
2	미국	459	17	동아시아	66
3	중국	436	18	태평양	65
4	북한	357	19	몽골	59
5	유럽연합	256	20	캐나다	55
6	일본	240	21	북극	54
7	호주	163	22	우크라이나	48
8	러시아	127	23	라오스	45
9	독일	120	24	동남아	41
10	영국	107	25	스위스	40
11	인도	94	26	인도네시아	38
12	동북아	89	27	미얀마	37
13	스웨덴	88	28	싱가포르	33
14	프랑스	80	29	캄보디아	33
15	베트남	80	30	남극	30



<표 9> 매트릭스 중심성 분석을 통한 상위 5개국 국제연구협력 네트워크 분석

단어	연결중심성	매개중심성	위세중심성
한국	60.909	0.064	0.538
미국	54.727	0.064	0.534
중국	48	0.064	0.481
북한	17.636	0.064	0.2
유럽(연합)	14.091	0.064	0.158
일본	32.636	0.064	0.359

상위 5개국의 국제연구협력 네트워크



#### 4) 국제연구협력 감성 및 담론 분석 결과

##### (1) 감성 사전을 기반으로 한 국제연구협력의 감성 분석

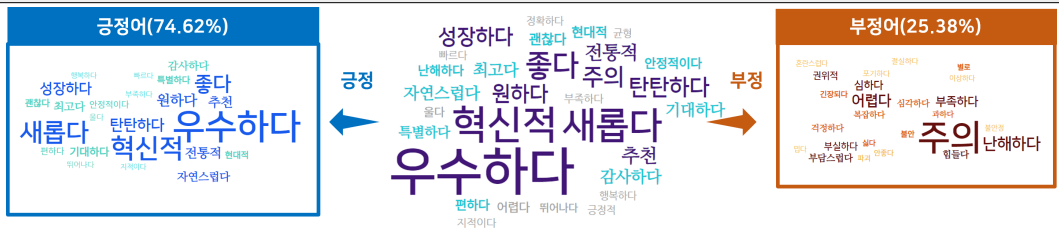
국제연구협력에 대한 소셜네트워크상의 인식을 파악하기 위해 감성분석을 수행하였다. 감성사전을 기반으로 감성에 관한 키워드를 수집하여 이를 수치화하였다. 그 결과, 긍정적인 의견이 74.62%를 차지하며 우수하다, 성장하다, 좋다, 성장하다, 탄탄하다, 안정적이다 등의 평가와 혁신적, 원하

다, 새롭다, 기대하다, 특별하다 등의 미래지향적인 평가가 우세하게 나타났다. 반면, 부정적인 평가는 25.38%로 난해하다, 어렵다, 부족하다, 울다, 힘들다, 절실하다 등의 평가가 주를 이뤘다. 추출한 전체 감성어를 흥미, 호감, 통증, 슬픔, 분노, 두려움, 놀람, 기쁨, 거부감의 9가지 감정으로 분류하였을 때, 긍정적인 측면으로는 호감, 흥미에 대한 감정이 우세하여 국제연구협력에 대한 전반적 평가는 호감형 평가이며 미래의 발전가능성에 대한 기대를 내포하는 평가가 주를 이루고 있다.

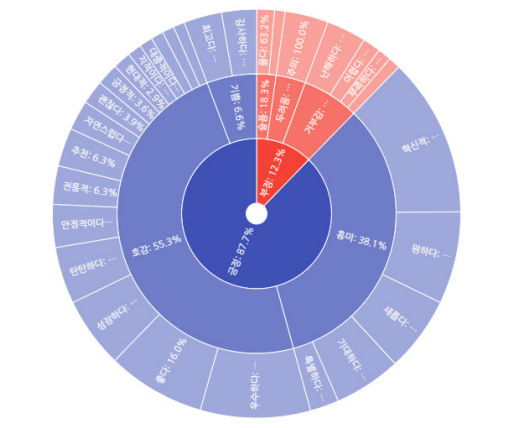
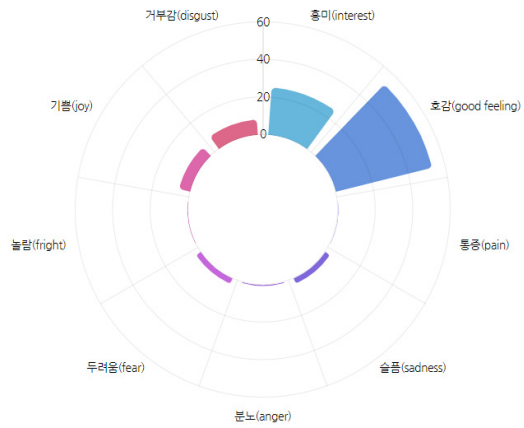
<표 10> 감성분석 결과

구분	빈도(건)	감성강도비율(%)	빈도비율(%)
긍정	1915/2589	74.62/100.0	73.97/100.0
부정	674/2589	25.38/100.0	26.03/100.0

감성단어 워드클라우드



단어 감성분석	감성단어 분석
---------	---------



\* 감성단어를 9가지 감정으로 표현한 결과 흥미와 호감이 우세함

\* 긍정 단어의 전체 비율을 나타내는 파이차트 분석 결과

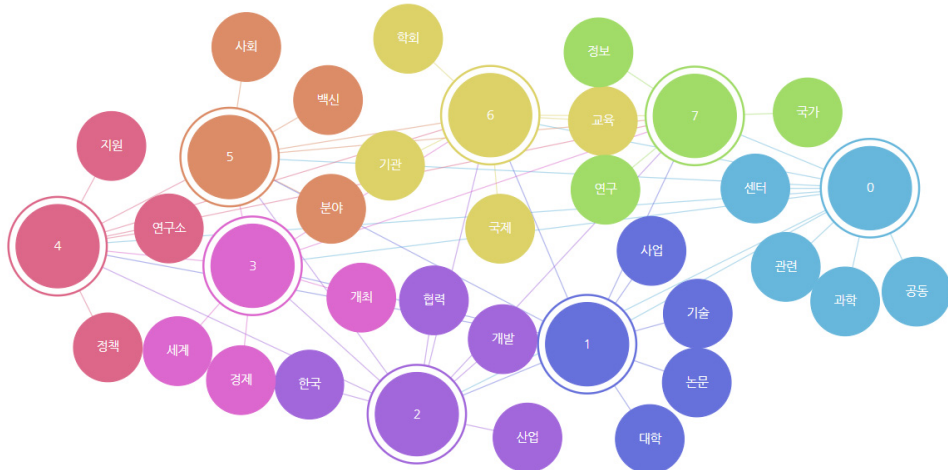
**(2) CONCOR 분석을 통한 국제연구협력 담론 분석**

국제연구협력에 대한 담론을 파악하기 위해 공출현 단어 간의 상관관계를 이용하여 단어 간의 관계 패턴에 따라 군집화하는 CONCOR 분석을 수행하였다. 군집화된 그룹을 나타낸 결과는 다음과 같다. G0의 경우, 공동, 과학, 관련, 센터의 키워드가 공출현하는 단어로 군집되었다. 이는 센터를 중심으로 과학 분야의 공동 연구 혹은 개발이 일어나고 있음을 시사한다. G1의 경우, 기술, 논문, 대학, 사업 등의 단어가 군집되었는데 이는 대학이 중심이 되어 논문, 사업, 기술 등에 관한 국제연구협력에 대한 활동이 주로 이루어지고 있음을 유추할 수 있다. G2의 경우 개발, 산업, 한국, 협력이 군집되어, 한국의 산업체 단위에서 개발 협력을 수행하고 있음을 알 수 있다. G3의 경우 개최, 경제, 세계의 키워드가 군집되었는데, 세계

경제에 관하여 세미나 또는 학술대회 등의 행사를 개최함으로써 직접적인 소통이 활성화되고 있음을 알 수 있다. G4의 경우, 연구소, 정책, 지원이 군집되었는데 이는 연구소 단위에서 정책 관련 연구가 이뤄지며 이를 지원하기 위한 노력이 이루어지고 있음을 유추할 수 있다. G5의 경우 백신, 분야, 사회가 군집되었는데 이는 국제사회가 백신을 급하기 위해 국제연구협력을 수행한 관련 정보가 군집된 것으로 보인다. G6의 경우 교육, 국제, 기관, 학회가 군집되었는데 이는 경제협력개발기구(OECD)가 지난 20년 국제학업성취도 평가 연구를 발표함에 따라 관련 학회 및 기관이 교육에 관하여 활발한 논의가 일었던 것으로 유추된다. 마지막으로 G7의 경우 국가, 연구, 정보가 군집되었는데, 이를 통해 국제연구협력에 관한 연구정보를 제공하기 위한 정부 차원의 지원이 이뤄지고 있음을 예측할 수 있다.

<표 11> CONCOR 분석 결과

군집	단어	군집	단어
G0	공동, 과학, 관련, 센터	G4	연구소, 정책, 지원
G1	기술, 논문, 대학, 사업	G5	백신, 분야, 사회
G2	개발, 산업, 한국, 협력	G6	교육, 국제, 기관, 학회
G3	개최, 경제, 세계	G7	국가, 연구, 정보



## 5) 국제연구협력 주요 토픽 분석 결과

### (1) LDA 토픽 모델링 분석을 통한 국제연구협력 토픽 분석

LDA 토픽 모델링은 대량의 문서군으로부터 주제를 찾아내기 위해 유사한 의미를 가진 단어들을 집단화 하는 알고리즘이다. 본 연구에서는 LDA 토픽 모델링을 통해 신기술 분야 산학협력 네트워크 내에서 가장 이슈화되고 있는 주제를 파악하여 동향을 분석하고자 한다.

토픽 모델링에서 원의 크기는 해당 토픽의 빈도수를 나타내므로 가장 핵심 토픽으로 해석할 수 있다. 토픽에서 가장 많은 비중을 차지하고 있는 토픽 1번은 국내에서 이루어지는 국제연구협력에 대한 포괄적인 이슈를 다루고 있다. 기술, 대학, 공동, 사업, 센터, 연구소, 기업, 기관, 지원, 논문, 정보 등의 주제를 기준으로 국내 국제연구협력 전반에 대해 논의하고 있는 것을 확인할 수 있었다. 더불어 과제, 성과, 변화, 교수 등의 키워드가 함께 도출되면서 국제연구협력의 추진은 교수의 성과향상에 직접적인 영향을 주고 있음이 나타났다. 그 다음으로 많은 비중을 차지하는 토픽 3번은, 토픽 1번과 중첩되어 있음과 동시에 그 토픽 내에 속해있는 형태를 띠고 있다. 토픽 3번은 국내 국제연구협력 중에서도 범죄방지, 백신, 바이오 등의 세부 분야에 따른 구체적인 국제연구협력 이슈가 다루어지고 있음을 시사한다. 또한 정책, 지원 등의 키워드가 함께 도출되면서 해당 분야의 국제연구협력은 법적제 기반과 지원하에 활발히 이루어지고 있음을 확인할 수 있다. 3번째로 큰 비중을 차지하는 토픽 4번의 경우, 국제연구협력 내 교육과 글로벌 시장 내 세계 경제에 관한 토픽으로 이루어져있다. 세계, 글로벌, 경제, 사회, 방안, 국가 등의 키워드가 주를 이루고 있으며 이러한 분야에 대한 연구가 데이터를 기반으로 이

루어지고 있음을 나타내었다. 다음으로 2번 토픽의 경우, 국제연구협력 내 업무협약, 양해각서 등의 협약 체결을 통해 이뤄진 국제연구협력 성과에 대한 키워드로 이루어져 있다. 이 중 백신, 금융, 경제, 반도체 분야의 상호 협약이 가장 많이 이루어졌음을 나타내고 있으며, 이를 통해 관련 기술을 개발하거나, 연계 사업을 추진하거나, 관련 시스템을 구축하고자 하는 노력이 수행되었음을 확인할 수 있다. 다음으로 2번 토픽과 동일한 비중을 차지하는 6번 토픽의 경우, 중국, 미국과의 국제협력에 대한 이슈를 담고 있다. 사회경제적 흐름에 따른 사회 이슈에 영향을 받은 국제협력에 대한 데이터가 주로 나타났다. 가장 중심으로부터 떨어져 있으며 작은 비중을 차지하는 토픽 5번의 경우, 핵심 토픽으로 보기는 어려우나 이후 주요 토픽으로 떠오를 수 있는 잠재력을 가지고 있다. 토픽 5번에서는 국제연구협력에 관한 정책에 대한 주제로 군집되어 있으며 정보, 공고, 취업, 인재, 지원 등의 키워드가 함께 도출되면서 국제연구협력의 법적, 사회적 기반 마련을 위한 전반의 담론이 형성되어있음을 확인하였다.

### (2) Word-level Semantic Clustering 분석을 통한 국제연구협력 토픽 분석

현재 신기술 분야 산학협력 내 주요 토픽을 분석하기 위해 문서 내 단어들의 공출현 관계를 기준으로 벡터화하여 인접 단어를 같은 군집으로 묶어 토픽을 도출하는 Word-level Semantic Clustering 분석을 수행하였다.

가장 많은 언급이 있었던 분야는 스포츠주제 분야로, 경영학, 에이전트, 마케팅, 축구, 프로 등의 연관 단어가 군집되어있다. 이때 경영학 분야의 군집단어가 스포츠 군집 단어가 유사한 것을 보아 스포츠와 경영학 분야의 연관성이 파악되는 것으로 나타난다. 두 번째로 언급이 많았던 분야

<표 12> LDA 토픽 모델링 결과



<표 13> Word-level Semantic Clustering 분석 결과

단어	군집단어	N
에너지	현대제철, 안동일, 중립, 철강, 탄소	475
코로나	퇴치, 수단, 신종, 감염증, 재가동	432
우주	정거장, 탐사, 달, 아르테미스, 인공위성	264
농업	농촌, 식량, 기술원, 기획회의, 미작	242
스포츠	경영학, 에이전트, 마케팅, 축구, 프로	665
환경	대기, 플라스틱, 오염, 보전, 수질	593
바이오	의약품, 사이언스, 헬스, 로직스, 판매	242
경영학	스포츠, 마케팅, 에이전트, 축구, 호주	141
공학	생명, 나노, 공학부, 가천, 서원일	142
관광	레저, 동북아, 한중일, 문화, 백제	130
반도체	배터리, 전기차, 시갈, 공장, 동맹	129



는 환경 분야로, 대기, 플라스틱, 오염, 보전, 수질 등의 연관단어가 도출되었으며 에너지 분야가 다음으로 언급이 많은 것을 보아 현재 환경, 에너지 분야에 대한 국제연구협력이 중점적으로 수행되는 것으로 예측할 수 있다. 관광 분야의 경우 동북아, 한중일을 중심으로 국제연구협력이 활발히 일어나고 있음을 확인하였다.

## 5. 논의

글로벌 시대가 도래함에 따라 경제 활동과 사회 문화적 교류가 더 이상 한 국가 내에 국한되지 않고 전 세계를 배경으로 이루어지고 있으며, 이러한 시대적 흐름에 맞추어 국제연구협력에 대한 중요성과 필요도가 더욱 높아지고 있다(노영희, 박우정, 2022). 국제연구협력은 연구 성과 향상을 위한 전략 중 하나로 떠오르고 있어 다양한 분야의 국제연구협력이 수행됨에 따라 효과적인 연구 협력의 수행과 지원에 대한 방안 마련이 필요해진 상황이다(노영희, 노지윤, 2021).

본 연구 결과를 바탕으로 국제연구협력 동향에 관하여 다음의 세가지 부분의 논의점을 도출하자면 다음과 같다. 첫째, 수집된 데이터의 연간 추이를 통해 향후 국제연구협력이 계속해서 증가할 것으로 보이며, 그 속도가 가속화되고 있다. 2018년도와 2019년도는 7천건 대가 수집되었으나, 2020년 8천건 이상이 수집되었으며 2021년 6천건으로 감소하였다가 다시 2022년 8천건 이상이 수집되었다. 21년도 수집량이 눈에 띄게 저조한 원인은 COVID-19 팬데믹에 대한 여파로 국가 간의 교류가 감소하고 세계경제가 침체된 영향으로 보인다. 그러나 주목할 만한 점은, 연구의 수행 시기를 고려하였을 때 2022년도의 경우 상반기 데이터만 수집하였음에도 불구하고 상당한 수집량을 보이

면서 관련 소셜데이터가 폭발적으로 급증하였으며 따라서 향후 국제연구협력이 계속해서 빠르게 증가할 것으로 예측된다.

둘째, 국제연구협력의 주요 연구주제는 국제사회의 이슈에 많은 영향을 받고 있으며, 여타 비주류 연구 분야의 국제연구협력 활성화를 위해 다양한 지원이 필요할 것으로 보인다. 분석을 통해 도출된 연구주제는 스포츠, 경제, 교육, 산림, 과학 등으로 나타났다. 이 중 스포츠의 경우 가장 높은 빈도 수를 보였는데, 이는 국가의 국제적 위상을 높이하고자 진행하는 올림픽 등의 국제스포츠대회가 계속해서 발전하고 있으며 이러한 움직임은 단순히 외교적 수단을 넘어 글로벌 사회에 공헌하고 경제 개발에 기여하는 ‘국제스포츠교류협력’으로 그 개념이 확장되고 있음을 나타내는 지표로 보인다(조현주, 2020). 이 외에 세계적인 경제침체 상황에 따라 경제 분야의 국제연구협력이 활발히 진행되고 있으며, 환경오염이 국제적인 문제로 대두됨에 따라 환경 분야와 관련된 에너지, 산림, 원자력, 미세먼지 등의 토픽이 함께 도출되었다. 따라서 국제연구협력은 국제사회에 당면한 이슈에 따라 밀접한 영향을 받고 있으며, 이 외의 비주류 연구분야가 국제연구협력을 수행하기 위해서는 이들을 지원하기 위한 다양한 사업들이 수행되어야 할 것으로 보인다.

셋째, 국제연구협력을 지원하는 기반에 대한 담론이 형성되고 있는 것으로 보이며 국제연구협력 활성화를 위한 인프라 조성이 향후 국제연구협력 내 주요 토픽이 될 것으로 보인다. TF-IDF 분석 결과 정보, 정책, 지원 등의 키워드가 도출되었고, 이어 담론 분석에서 동일 맥락상에서 국가, 연구, 정보 키워드가 군집되면서 국가적 차원의 정책 지원과 연구정보 지원 등의 인프라에 관한 담론이 형성된 것을 확인하였다. 또한 LDA 토픽 모델링에서 가장 작은 비중을 차지한 5번 토픽에서

는 정책, 정보, 지원, 인재 등의 내용을 담은 데이터가 도출된 것을 보아, 아직 국제연구협력의 인프라에 관한 담론이 지배적이지는 않으나 관련 주제로 담론이 이뤄지고 있는 단계이며 국제연구협력 활성화를 위한 인프라 조성이 향후 주요 담론이 될 가능성이 높은 것으로 예측된다.

## 6. 결론 및 제언

본 연구는 국제연구협력에 관한 소셜빅데이터를 수집하여 글로벌 R&D의 동향과 특성을 파악하고자 하였다. 주요 연구결과를 요약하자면 다음과 같다. 첫째, 단어빈도분석과 TF-IDF 분석을 통해 국제연구협력 내 상위 키워드를 분석한 결과, 개발, 기술 등의 단어가 가장 많이 등장하였으며 연구소, 대학이 상위권에 등장하면서 국제연구협력이 학계와 연구계를 중심으로 활성화되어 있음을 확인하였다. 다음으로 백신 키워드가 TF-IDF 3위에 위치하면서 COVID-19 사태에 따라 백신에 관한 국제연구협력이 집중적으로 이루어졌음을 알 수 있었으며, 이 밖에 정보, 사업, 정책, 지원 등의 키워드를 통해 국제연구협력정보에 대한 관심과, 사업 및 정책적 지원에 대한 데이터가 많이 형성되어있음을 유추하였다. 연구주제에 해당하는 키워드는 스포츠, 경제, 교육, 산림, 과학 등이 상위 키워드로 관찰되었다.

둘째, 국제연구협력의 감성분석 결과, 긍정적인 의견이 74.62%를 차지하며 우수하다, 성장하다, 좋다, 탄탄하다 등의 평가가 도출되었다. 25.38%의 부정적인 의견으로는 난해하다, 어렵다, 부족하다 등의 평가가 주를 이루고 있음을 확인하였다. 다음으로 CONCOR 분석을 통해 담론을 분석한 결과 크게 센터를 중심으로 과학 분야의 공동연구에 관한 담론, 대학 중심의 학술적 국제연

구협력 활동, 국내 산업체 단위의 기술개발 협력 활동, 세계경제에 관한 국제적 학술교류에 관한 담론이 주요 담론으로 도출되었다. 그 밖에 연구소 중심의 정책 관련 연구활동, 국제사회의 백신에 관한 연구협력, 교육 분야의 국제연구협력, 국제연구협력정보 지원을 위한 국가적 차원의 지원에 대한 담론이 형성되어 있음을 확인하였다.

셋째, 국제연구협력 주요 토픽 분석을 수행한 결과, 국제연구협력 내 가장 큰 토픽은 국내 대학, 센터, 기업, 연구소에서 이루어지는 공동연구 활동에 대한 전반에 관한 사항이 나타났으며 해당 토픽 내에 범죄방지, 백신, 바이오 등의 연구주제에 대한 세부적인 활동이 일어나고 있는 양상을 확인하였다. 다음으로는 세계경제, 국제사회에 관한 연구협력과 교육에 관한 토픽과 백신, 금융, 경제, 반도체 분야의 업무협약을 통한 교류가 활성화되어있음을 알 수 있는 토픽이 도출되었다. 이러한 업무협약을 통해 유관 기술의 개발, 시스템 구축 등의 활동이 이루어졌음을 확인하였다. 그 외로 중국과 미국의 국제정세와 밀접한 이슈가 언급된 토픽, 국제연구협력의 정책과 채용, 공고, 취업, 인재 등의 이슈에 관한 토픽이 도출되었다. World-level Sementic Clusturing 분석을 통해 스포츠 분야가 경영학, 마케팅과 함께 언급되며 융합이 이루어지고 있음을 예측할 수 있었으며, 환경, 에너지 분야의 연구가 활발히 이루어지고 있음을 확인하였다.

넷째, 개체명인식 분석과 네트워크 분석을 통해 협력주체별 동향을 분석한 결과, 대학, 정부부처, 사기업, 국공립연구소, 공기업, 국제기구 순으로 기관 유형의 비중에 대한 현황을 확인했으며 네트워크 분석을 통해 대학이 가장 활발한 국제연구협력 활동을 수행하고 있으며 특히 사업, 협약, 논문의 키워드에서 가장 높은 연관성을 보이고 있어 유관 사업 수행, 타기관과의 협약 체결, 논문

을 바탕으로 학술 연구를 주로 수행하고 있음을 도출하였다.

다섯째, 개체명인식 분석과 네트워크 분석을 통해 국가별 동향을 분석한 결과, 미국, 중국, 북한, 유럽연합, 일본 순으로 언급빈도가 많은 것을 확인하였고 이를 바탕으로 상위 5개국 네트워크 분석을 수행한 결과 미국, 중국, 일본, 북한, 유럽연합 순으로 높은 연결성을 보이고 있다.

본 연구는 국제연구협력의 동향을 특성을 파악하기 위해 소셜빅데이터를 활용하여 분석하였다는

점에 의의가 있다. 그럼에도 불구하고 연구의 진행 상 다소 한계점을 내포하고 있다. 이러한 한계점을 개선하기 위한 향후 연구를 다음과 같이 제안한다. 본 연구에서 수집된 데이터는 국내의 소셜빅데이터만을 수집하였기 때문에 실제 국제연구협력이 일어난 모든 관계를 반영하지 못하였다. 따라서 향후 연구에서는 국제연구협력 관계를 나타내는 직접적 지표를 대상으로 분석하여 실질적인 동향을 파악할 수 있도록 하는 것이 필요하다.

### 참 고 문 헌

- 김수현 · 손욱 (2021). 국제개발협력의 키워드 변화 텍스트마이닝 분석: 『OECD 국제개발협력보고서』 (1997-2018). <국제개발협력연구>, 13(1), 19-39.
- 노영희 · 광우정 (2021). 국제연구협력 참고자료 정보원 분석을 통한 국제연구협력 특징 및 동향 분석에 관한 연구 <한국자치행정학보>, 35(4), 93-117.
- 노영희 · 노지윤 (2022). 국제연구협력 동향 및 발전 방향에 관한 연구: 국제연구협력 연구보고서 분석을 중심으로 <한국콘텐츠학회 논문지>, 22(3), 476-487.
- 노영희 · 이광희 · 장로사 (2019). 국가별 국제연구협력 주제 연구동향 비교분석. <한국콘텐츠학회 논문지>, 19(3), 288-301.
- 조현주 (2020). 국제스포츠교류협력 전담 체계 구축 방향 고찰. <스포츠엔터테인먼트와 법>, 23(3), 159-176.
- 주한나 · 정효림 · 이희진 (2020). 한국의 ICT 국제개발협력 연구 동향 분석: 2002-2020년 학술 논문의 체계적 문헌 분석. <국제개발협력연구>, 12(3), 33-55.
- 한국과학기술정보연구원 (2018). <국제협력 기반구축 및 협력활성화 연구>. 세종: 과학기술정보통신부.
- 한재광 (2021). 한국에서의 국제개발협력 시민사회 연구동향 분석: 2001~2020년 학술논문 분석을 중심으로 <국제개발협력연구>, 13(1), 59-79.

## An Analysis of Trends in International Research Cooperation Based on Big Data

Younghee Noh, Yoonseo Nam

Konkuk University, Research Institute for Knowledge Content Development & Technology

This study was conducted to analyze the trends in international research cooperation, Social big data were collected to identify the current status and derive the major topics in this field. To this end, TEXTOM, a big data collection and analysis software system, was used to collect social big data related to international research cooperation from 2018 to 2022 from major portal sites such as Google, Naver, and Daum. The collected data were subjected to frequency, centrality, N-gram, matrix, and topic analyses after refinement through a text mining process. The study results confirmed that international research cooperation was actively centered on research institutes and universities, and research cooperation on vaccines was being intensively conducted due to the coronavirus disease 2019 (COVID-19) pandemic. In addition, research topics such as sports, economy, education, forest, and science were derived as the top keywords. Currently, the main discourse in international research cooperation was found to be on joint research in the field of science, academic international research cooperation activities centered on universities, and technology development cooperation activities at the domestic industry level. The main topics derived were, detailed international research cooperation and information and policy support on research topics such as crime prevention, vaccines, bio, and the world economy. These research results are expected to provide useful data for predicting major issues relating to international research cooperation in the future by examining the main keywords concerning the same and the discourse and topics of interest within current areas of international research cooperation. In addition, it can be used as a basic study to establish in-depth support measures for international research cooperation by identifying currently active and non-active fields through network analysis between cooperative entities and countries.

*Keywords: International Research Cooperation, Big-Data analysis, Social Bigdata, Textmining, LDA Topic Modeling*