

유망연구주제의 조기 탐색 기법 연구*

- 사회과학분야를 대상으로 -

A Study on Method for Early Detection of Emerging Research Topic in Social Sciences

구 본 진 (BonJin Koo)**

장 덕 현 (DurkHyun Chang)***

목 차

- | | |
|-----------------|----------|
| 1. 연구의 배경 및 필요성 | 4. 분석 결과 |
| 2. 선행연구 | 5. 결 론 |
| 3. 연구의 내용 및 방법 | |

초 록

본 연구는 사회과학 분야를 대상으로 향후 유망연구주제로 성장할 가능성이 있는 연구주제를 조기에 예측하기 위해 altmetrics의 활용 가능성을 모색하고자 하였다. 이를 위해, 연구주제의 도출을 위한 토픽모델링 분석, 연구주제의 계량서지 지표와 altmetrics 지표의 공통적 특징을 추출하기 위한 주성분분석, 계량서지 지표와 altmetrics 지표의 시차 간 관련성과 영향 관계를 확인하기 위한 교차상관 분석 및 회귀분석, 유망연구주제 도출을 위한 상관분석을 수행하였다. 분석결과, altmetrics 지표 중 capture 수는 논문 수, 저자 수, 인용 수의 선행지표로 활용 가능하며, 그 경우, 유망연구주제를 2년에서 3년 정도 빠르게 파악하는 것이 가능하므로 유망연구주제의 조기 탐지를 위해 해당 지표를 사용할 수 있을 것으로 판단된다. 본 연구는 기존 연구에서 주로 활용되어 온 인용 및 키워드 기반의 지표들에서 발생하는 시간적 한계를 극복하고 즉시성을 담보할 수 있는 지표로서 altmetrics의 활용 가능성을 확인하였다는 점과 주성분분석을 활용하여 미래 유망연구주제를 도출함으로써 유망연구주제의 예측을 위한 방법론적 확장 가능성을 제공하고 향후 연구주제 탐지 모델 설계에 활용할 수 있는 기반을 마련하였다는 점에서 의의가 있다.

ABSTRACT

This study aimed to investigate promising research areas in the field of social science by exploring the possibility of employing altmetrics to serve the purpose. To facilitate the research, five analytical methods were conducted: 1) topic modeling to derive research topics, 2) PCA to extract common features of bibliometric indicators and altmetrics indicators, 3) cross-correlation analysis and 4) regression analysis to identify time differences between bibliometric and altmetrics indicators, and define their relation, and 5) correlation analysis to finalize emerging research topics. The findings of the study revealed that the captures, among the altmetrics indicators, can be used as a leading indicator of the number of articles, authors, usages, and it can contribute to identifying emerging research topics two to three years earlier. This study has significance as it identified the applicability of altmetrics in overcoming the limitation of time lag that occurs in citation and keyword-based indicators used in previous studies, and proved it to be an indicator to ensure immediacy in the detection of emerging research topics. It also demonstrates the methodological extensibility for predicting emerging research topics by using PCA and provides the foundation for designing detection models for emerging research topics in the future.

키워드: 유망 주제, 연구동향, 토픽모델링, 주성분분석, 시계열 교차상관분석

Emerging Topics, Research Trends, Altmetrics, Topic modeling, Principal Component Analysis(PCA), Time-series Cross-Correlation Analysis

* 이 논문은 제1저자의 박사학위논문을 축약한 것임.

** 부산대학교 문헌정보학과 강사(bjkoo@pusan.ac.kr / ISNI 0000 0005 0269 148X) (제1저자)

*** 부산대학교 문헌정보학과 교수(dchang@pusan.ac.kr / ISNI 0000 0004 6477 1402) (교신저자)

논문접수일자: 2025년 4월 24일 최초심사일자: 2025년 5월 8일 게재확정일자: 2025년 5월 10일

한국문헌정보학회지, 59(2): 43-72, 2025. <http://dx.doi.org/10.4275/KSLIS.2025.59.2.043>

© Copyright © 2025 Korean Society for Library and Information Science

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 연구의 배경 및 필요성

학문 분야에서 발전 가능성이 있는 연구주제를 탐지하고 발굴하는 것은 잠재력있는 연구의 선별, 유망 연구주제에 대한 경쟁력 선점, 우수 연구의 조기 진행, 연구 상황에 대한 객관적 파악 및 미래 성과 예측을 통한 향후 발전 방향 결정 등에 매우 중요하다. 특히, 정보기술의 발달과 지식의 양적 증가, 학문의 학제적 성향 강화 등으로 인해 새로운 연구 분야가 빠르게 출현하고 다양한 연구 분야에서 통합과 세분화가 빈번하게 이루어지는 등 연구 분야들이 복잡하게 연결되어 있다. 이로 인해 현재 부상(浮上)하고 있는 새로운 연구주제의 발견이나 연구 분야 간 연관성을 파악하는 데 어려움이 있으며, 따라서 새로운 연구주제의 동태(動態)와 발전을 파악하는 것이 더욱 중요해졌다.

학문 분야에서 예측이라는 것은 새롭게 등장하는 연구주제나 기존의 연구주제 중 새롭게 부상하는 연구주제를 발굴하고 탐지하는 것으로 이해할 수 있으며, 따라서 발전 잠재력이 있는 연구주제를 최대한 조기에 식별하는 것이 유용하다. 이에 따라 해외의 경우, 다양한 방법론을 적용하여 연구주제를 탐지하고자 하는 노력이 이루어져 왔으며, 유망연구주제의 탐지에 관한 연구가 꾸준히 수행되고 있다. 그러나 현재 새로운 연구주제의 출현을 예측하기 위해 사용되는 지표들은 주로 연구동향 분석을 위해 사용하던 인용 및 키워드 기반의 지표들로, 대부분 해당 분야의 연구 성과가 어느 정도 축적된 이후 식별되기 때문에 연구주제가 새롭게 출현하거나 성장하는 단계에서 새로운 연구주제를 감지하는 것은 불가능하다. 또한 최근 들

어 예측을 위해 머신러닝 기법이 주로 활용되고 있지만 유망연구주제의 예측에 사용되는 변수들의 영향력보다는 예측력 자체에 초점을 맞추는 경우가 많으며(Xu et al., 2019; Liang et al., 2021), 분석에 사용하는 변수 역시 이전과 큰 차이가 없다.

따라서 유망연구주제를 조기에 탐지하기 위해서는 즉시성을 담보할 수 있는 지표를 함께 검토하는 것이 필요하다고 할 수 있으며, 이에 본 연구에서는 altmetrics 지표의 활용 가능성을 모색하고자 한다. altmetrics는 뉴스 기사, 블로그, 각종 SNS 등에서 언급 또는 게재된 수, 서지관리도구 등에서 구독된 데이터 등 다양한 측정 데이터의 조합을 통해 연구 성과의 영향력을 측정하는데, 학술적 영향력뿐만 아니라 사회적 영향력 평가 도구로서 연구의 영향력 평가에 새로운 관점을 제공하고 있어 그 유용성을 인정받고 있다. 특히, altmetrics는 소셜 미디어 등 웹에서의 반응을 지표로 하고 있기 때문에 연구의 영향력을 실시간으로 측정할 수 있으며, 연구 트렌드를 파악하는 데 도움이 될 수 있다.

이에 따라 본 연구에서는 연구주제를 조기에 예측하기 위한 지표로써 altmetrics의 활용 가능성을 검토하고자 한다. 구체적으로 기존의 연구주제 탐지에 주로 활용되던 계량서지 지표들과 altmetrics 지표 간에 시차와 방향성을 분석하여 유망연구주제의 조기 탐지에 효과적인 altmetrics 지표를 확인하고, 계량서지 지표와 altmetrics 지표를 사용하여 도출한 유망연구주제를 비교함으로써 altmetrics의 유용성을 검증하고자 한다.

2. 선행연구

유망연구주제를 파악하기 위해 수행된 초기 연구들은 대부분 유망연구주제의 탐지 그 자체 보다는 주로 국가, 기관, 연구자, 연구 영역의 생산성과 영향력 측정이나 도서관의 연구지원 서비스 제공을 위한 목적으로 수행되었다. 따라서 대부분의 연구가 연구동향 분석의 일부로 유망 영역을 제시하고 있으며, 새로운 연구 영역의 출현이나 성장을 파악하여 미래를 예측함으로써 발전 가능성이 있는 연구 영역을 제안하고 있다. 이들 연구는 주로 논문 수나 인용수를 분석하여 특정 연구 분야의 현황과 미래 트렌드를 예측한 Ball과 Tunger(2006)의 연구와 디스플레이 분야 논문 및 특허데이터와 이들의 인용 데이터를 분석한 안세정 외(2012)의 연구가 대표적이다.

이후 인용의 패턴을 분석하여 새로운 영역의 출현이나 성장을 식별하는 연구들이 이루어졌다. 대표적으로 Small(2006), Takeda와 Kajikawa(2009), Iwami et al.(2014)의 연구가 있으며, 이들은 인용 네트워크 분석을 통해 연구영역의 구조를 파악하고 변화를 추적함으로써 유망 연구영역을 제안하였다.

이처럼 인용 수나 인용 패턴 분석의 경우, 해당 연구 영역에 대한 관심의 정도를 파악할 수 있으며, 나아가 성장 가능성을 확인할 수 있는 지표로 활용된다. 하지만 논문 출판 시점과 인용에 이르기까지 걸리는 시간차로 인해 유망 주제의 탐지에 있어 인용 지표 사용의 적절성에 대한 의문이 제기되면서 유망 주제를 탐지하기 위한 지표가 더욱 다양한 분석 단위를 대상으로 확장되었다. 그 중 많이 활용된 방법은

키워드 분석으로, Lee(2008)는 정보보안 분야를 대상으로, An과 Wu(2011)는 줄기세포 분야를 대상으로 동시단어분석을 수행하여 해당 분야의 최신 연구 동향과 유망 주제를 추적하였다. Ohniwa et al.(2010)은 생명과학 분야를 대상으로 출현이 증가한 키워드를 식별하고, 이들을 대상으로 동시인용 분석을 수행하여 유망 주제를 식별하였다. 이후 키워드를 활용한 분석은 토픽모델링 등 텍스트마이닝 분석에 꾸준히 활용되었고, 대표적으로 김현정 외(2015), Weismayer와 Pezenka(2017), Alattar와 Shaalan(2021)의 연구가 있다.

최근에는 유망연구주제의 탐지를 위해 2개 이상의 지표 및 방법론을 결합한 형태의 연구들이 수행되고 있으며, 특히 인용 분석과 키워드 분석을 결합한 연구가 많다. 대표적으로 Boyack과 Klavans(2010)는 하이브리드 접근법으로 동시인용분석, 서지결합분석, 직접인용분석과 인용 문헌 기반의 서지결합분석의 네 가지의 주요 인용 기반의 접근법을 활용하여 분석을 수행하였다. Guo et al.(2011)은 키워드, 저자, 참고문헌의 세 가지 지표를 사용하여 유망 연구 영역의 구조적 특징을 분석하였으며, 새로운 저자의 유입, 학제적 참고문헌의 인용 증가, 단어 사용의 급격한 증가 순으로 시간적 상관관계가 있는 것으로 나타났다. Glänzel과 Thijs (2012)는 신흥 유망연구주제를 탐지하고 레이블링하기 위한 핵심문헌의 사용에 대한 연구에서 서지결합분석과 텍스트마이닝을 결합하여 하이브리드 클러스터 분석을 수행하였다. 최근의 연구로 Yu et al.(2022)은 stress granules 관련 연구를 대상으로 동시인용분석, 키워드동시출현분석, 버스트분석을 통해 향후 유망 연구 영역을

예측하였다.

선행연구 분석을 통해 유망연구주제의 탐지와 관련된 대부분의 연구가 과학기술 분야를 대상으로 하여 수행되었다는 것과 유망연구주제의 탐지를 위해 사용한 지표들이 주로 인용 및 키워드 기반의 지표를 활용하고 있음을 알 수 있다. 특히, 대부분의 연구들에서 인용 및 키워드 기반의 지표와 관련 방법론을 활용하고 있는데, 이들 지표는 기존의 연구 동향 분석 및 영향력 분석에 주로 사용하던 지표 및 방법론들로 연구주제의 탐지에도 여전히 이들이 많이 활용되고 있음을 알 수 있다. 인용 및 키워드 기반의 지표들은 연구주제에 대한 관심도나 연구주제의 성장 가능성 및 역동성, 연구주제의 영향력을 보여준다는 점에서 유망연구주제 탐지를 위해서도 활용 가능하다. 그러나 인용 기반의 지표들은 인용 데이터가 축적되는 데 걸리는 시간이 있기 때문에 급부상하는 연구주제나 새로운 연구주제를 빠르게 탐지하는 것이 어려울 뿐 아니라 학문 분야마다의 시간차로 인해 상이한 학문 분야에 적용하여 비교하기에 적절하지 않다.

또한 키워드 기반의 지표 및 방법론 역시 비슷한 문제를 가지고 있는데, 연구주제를 대표할 수 있는 키워드는 해당 연구 커뮤니티와 연구자들에게 관련된 키워드들이 이미 인지되고 해당 키워드가 널리 사용된 이후에나 연구주제를 대표할 수 있는 키워드로서 식별되기 때문에 인용과 마찬가지로 급부상하는 연구주제나 새로운 연구주제를 빠르게 탐지하는 것이 쉽지 않다. 따라서 이들 지표의 한계를 개선할 수 있는 새로운 지표의 적용 가능성을 검토할 필요가 있으며, 특히 연구주제의 급부상성 및 신규성을 좀 더 빨리

포착하기 위해 즉시성을 담보할 수 있는 지표를 검토해야 할 필요가 있다고 볼 수 있다.

이상의 선행연구 검토 결과에 따라 본 연구에서는 그동안 이루어지지 않았던 사회과학 분야를 대상으로 하여 유망연구주제를 탐지하고자 한다. 아울러 유망연구주제의 탐지 시점을 좀 더 앞당기는 것이 가능한지 검토하기 위하여 즉시성을 담보할 수 있는 지표로서 altmetrics 지표의 적용 가능성을 확인하고자 하며, 이러한 측면에서 본 연구는 이전의 연구들과 차별성을 가진다.

3. 연구의 내용 및 방법

본 연구는 유망연구주제의 조기 탐색을 위한 altmetrics 지표의 활용 가능성을 검토하기 위해 토픽모델링, 주성분분석, 교차상관 분석, 회귀분석, 상관분석의 분석 방법을 활용하고자 한다. 대상은 사회과학 분야이며, 연구주제 도출을 위한 시간적 범위는 2013년부터 2022년까지이다. 도출된 연구주제들의 계량서지 지표와 altmetrics 지표는 논문 수 최고점 시점을 기준으로 이전 30년 치를 수집하며, 분석에는 논문 수 최고점 시점을 기준으로 5년 전 시점부터 30년 전 시점까지 총 25년 치의 데이터를 활용한다.

이를 위해 먼저 연구주제를 선정해야 하는데 연구주제의 선정을 위해서는 토픽모델링 분석을 수행한다. 토픽모델링 분석을 통해 도출된 토픽들의 등장 비중의 추이 및 논문 수와 키워드의 신규 등장 여부를 근거로 연구주제를 선행 유망연구주제와 미래 유망연구주제 후보군으로 나누어 도출한다. 분석은 1기(2013년~2017년)와 2기(2018년~2022년), 두 개의

시기로 나누어 실시하며, 1기의 토픽 중에서 선행 유망연구주제를, 2기의 토픽 중에서 미래 유망연구주제 후보군을 선정하였다. 본 연구에서 사용하는 선행 유망연구주제와 미래 유망연구주제 후보군의 정의와 도출 기준을 구체적으로 설명하면 <표 1>과 같다.

먼저 선행 유망연구주제는 어느 정도 성숙 단계에 이르렀다고 할 수 있는 연구주제이다. 선행 유망연구주제의 도출은 토픽의 연도별 등장 비중 추이와 함께 연구의 성과를 측정하는 대표적인 양적 방법으로써 논문 수를 기준으로 활용하였으며, 등장 비중이 감소 추세인 토픽 중에서 논문 수가 더 이상 증가하지 않는 토픽을 선행 유망연구주제로 도출하였다.

다음으로 미래 유망연구주제 후보군은 향후 유망연구주제가 될 가능성이 있는 신규 연구주제로 정의한다. 후보군의 도출은 신규 키워드의 등장 여부를 기준으로 하였는데, 이는 Guo et al.(2011)의 연구 결과를 근거로 유망연구주제의 경우 해당 주제를 대표하는 키워드의 사용이 증가한다는 점을 고려하여, 이전에는 등장하지 않았던 새로운 키워드가 포함되어 있는 경우에는 상대적으로 새로운 연구주제일 것이라고 판단하였기 때문이다. 이에 근거하여 2기 토픽의 핵심 키워드에 2기의 상위 100위 키워드에 최초 등장한 키워드가 3개 이상 포함되어 있는 토픽

을 미래 유망연구주제 후보군으로 도출하였다.

이어서 선행 유망연구주제와 미래 유망연구주제 후보군의 계량서지 데이터와 altmetrics 데이터를 수집하였다. 계량서지 데이터는 Web of Science(WoS) 데이터베이스를 통해 수집하였으며, altmetrics 데이터는 Scopus에서 제공하는 API를 활용하여 수집하였다.

다음은 선행 유망연구주제의 공통 추세를 도출하는 단계로 주성분분석을 활용하여 계량서지 지표와 altmetrics 지표를 단일 차원으로 축소, 이들 지표의 공통적 추세를 추출하고, 주성분분석 결과로 도출된 계량서지 지표와 altmetrics 지표의 공통적 추세 간에 유의한 시차가 존재하는지 확인하고 검증하기 위한 교차상관 분석과 회귀분석을 수행하였다.

마지막은 유망연구주제 도출 및 altmetrics 지표의 사용가능성을 확인하는 단계이다. 먼저 선행 유망연구주제의 공통적 추세와 미래 유망연구주제 후보군과의 상관관계 분석을 통해 선행 유망연구주제와 유사한 패턴을 보이는 연구주제를 파악한다. 이는 미래 유망연구주제 후보군 중에서 선행 유망연구주제의 공통적 특징과 유사한 패턴으로 성장하는 모습을 보이는 주제가 향후 유망연구주제가 될 가능성이 크다는 점을 전제로 한다. 이어서 선행지표로 사용 가능한 altmetrics 지표를 활용하여 상관관계

<표 1> 선행 유망연구주제와 미래 유망연구주제 후보군의 정의와 도출 기준

구분	선행 유망연구주제	미래 유망연구주제 후보군
정의	연구주제의 라이프 사이클상 성숙 단계에 이르렀다고 할 수 있는 연구주제	향후 유망연구주제가 될 가능성이 있는 신규 연구주제
도출 기준	2013년~2017년에 등장하는 토픽 중 등장 비중이 감소 추세인 동시에 더 이상 논문 수가 증가하지 않는 토픽	2018년~2022년에 등장하는 토픽 중 해당 시기의 상위 100위 키워드에 최초 등장한 키워드가 토픽의 핵심 키워드에 3개 이상 포함된 토픽

분석을 수행하며, 계량서지 지표와 altmetrics 지표를 사용하여 도출한 유망연구주제를 비교함으로써 altmetrics의 유용성을 검증하였다.

본 연구에서 활용한 분석 지표와 분석 도구는 다음과 같다. 먼저 계량서지 지표로 논문 수, 저자 수, 인용 수, 피인용 수, 신규 키워드 수를, altmetrics 지표로 PlumX Metrics의 Citations, Usage, Captures, Mentions, Social media를 활용한다. 연구 주제 도출을 위한 토픽모델링은 R 소프트웨어를, 주성분분석과 교차상관분석, 회귀분석, 상관분석에는 stata 18.0과 SPSS 27.0을 사용하였다. 또한 본 연구에서 연구주제의 도출을 위한 토픽모델링 분석 대상 저널의 선정은 Web of Science의 SSCI를 기준으로 하였으며, 주제 분야는 2023년 『국가과학기술표준분류체계』의 사회과학 분야를 기준으로 하였다. 이는 SSCI에 색인되어 있는 저널의 WoS 주제 카테고리의 경우, 인문학 분야 또는 과학기술 분야와의 교차 영역이라고 볼 수 있는 주제가 다수 존재하였고, 해당 주제에 포함되어 있는 논문들이 대부분 SCIE와 AHCI에 함께 색

인되어 있을 뿐 아니라 내용상 사회과학 분야와 거리가 있는 경우가 상당수 발견되었기 때문이다. 따라서 본 연구에서는 1차로 WoS의 SSCI에 색인되어 있는 저널에 배정된 WoS 주제 카테고리를 확인한 후 2차로 이를 『국가과학기술표준분류체계』의 사회과학 분류의 중분류와 매핑하는 과정을 거쳤다.

2023년 기준, SSCI에 색인되어 있는 저널에 배정된 WoS 주제 카테고리는 총 57개였으며, 이들 주제를 『국가과학기술표준분류체계』와 매핑하였다. 그 결과, 총 23개의 『국가과학기술표준분류체계』 주제 카테고리 중 18개의 주제에 45개의 SSCI 주제가 포함되었으며, 어디에도 포함되지 않는 12개의 SSCI 주제 카테고리는 분석에서 제외하였다.

이상의 과정을 거쳐 선정된 총 18개 주제에 해당하는 45개의 SSCI 주제 카테고리를 대상으로 하여 2013년부터 2022년까지 10년간 매년 IF(Impact Factor) 최상위 저널을 분석 대상 저널로 선정하였으며, 최종 분석대상 저널과 논문 수는 <표 2>와 같다.

<표 2> 18개 주제별 저널 수 및 논문 수

순	주제	저널 수	논문 수		순	주제	저널 수	논문 수	
			2013~2017	2018~2022				2013~2017	2018~2022
1	정치외교학	11	419	283	10	행정학/정책학	6	235	338
2	경제학	2	162	233	11	지리학	3	184	73
3	경영학	2	197	119	12	국제/지역개발	22	2,829	5,819
4	회계학	3	342	1,658	13	관광학	3	463	958
5	사회학	23	2,089	3,855	14	신문방송학	7	347	122
6	사회복지학	6	734	766	15	심리과학	33	5,158	5,914
7	인류학	5	556	1,441	16	생활과학	3	246	511
8	교육학	7	697	390	17	문헌정보학	4	325	776
9	법학	5	196	101	18	여성학	6	212	417
계							151	15,391	23,774

4. 분석 결과

4.1 주요 연구주제 및 핵심 키워드 도출

앞서 선정한 18개 주제의 151종 저널에 게재된 39,165건의 논문을 대상으로 토픽모델링을 수행하여 시기별로 주요 연구주제를 식별하고, 이를 근거로 선행 유망연구주제와 미래 유망연구주제 후보군 식별 및 영역별 핵심 키워드를 도출하였다. 큰 범주에서는 비슷한 주제라 할 지라도 학문 분야마다 중심 주제가 상이하거나 또는 적용하는 방법론에 차이가 있을 수 있으며, 각 학문적 특성을 반영한 키워드들이 사용될 수 있기 때문에 18개의 주제 분야별로 각각 토픽모델링을 수행하였고 데이터 수집은 WoS 데이터베이스를 대상으로 하였다. 토픽모델링 분석은 저자 키워드를 활용하였으며, 분석 대상 논문 중 저자 키워드가 제시되어 있지 않은 논문의 경우에는 논문명과 초록을 활용하였다. 또한 1기(2013년~2017년)와 2기(2018년~2022년), 두 개의 시기로 나누어 시기별 주요 연구주제와 각 주제를 대표하는 핵심 키워드를 도출하고 이를 토대로 선행 유망연구주제와 미래 유망연구주제 후보군을 선정하였다.

저자 키워드, 논문명, 초록에 대한 정제 작업은 먼저 공백 처리, 대소문자 통일, 문장부호 및 특수문자 제거의 전처리 과정을 거쳐 불용어 삭제, 어근 동일화(stemming)의 순서로 진행하였다. 데이터 전처리 및 분석은 R 소프트웨어를 통해 이루어졌으며, 자연어 처리를 위한 'NLP' 패키지, 텍스트마이닝을 위한 'tm' 패키지, 토픽 모델링 분석을 위한 'topicmodels', 'ldatuning' 패키지를 사용하였다. 토픽 수의 결정은 'ldatuning'

패키지의 FindTopicsNumber() 함수를 활용하였다. Griffiths와 Steyvers(2004), Cao et al. (2009), Arun et al.(2010), Deveaud et al. (2014)이 제안한 LDA 모형의 모형 적합도 지수를 확인하여 모형의 성능이 크게 향상되지 않는다고 판단되는 토픽의 수를 해당 주제의 최적 토픽 수로 하였으며, 18개 주제 분야별로 각각 수행하여 각 주제 분야별 최적의 토픽 수를 정하였다. 토픽의 주제명은 토픽별로 도출된 상위 빈도 키워드를 기반으로 하였고, 필요한 경우 토픽별로 등장 확률이 높은 논문을 추가로 확인하여 토픽의 주제를 해석하고 이름을 붙였다.

분석 결과, 18개 주제 분야에서 적게는 8개에서 많게는 25개의 토픽이 도출되었으며, 각 주제 및 시기별 토픽의 수는 다음 <표 3>과 같다.

4.2 선행 유망연구주제 및 미래 유망연구주제 후보군 도출

4.2.1 선행 유망연구주제

앞서 수행한 토픽모델링 분석을 토대로 선행 유망연구주제를 선정하였다. 선행 유망연구주제의 선정 과정은 다음과 같다. 먼저 선행 유망연구주제 후보군으로 각 주제 영역별로 1기에 등장하는 연구주제의 연도별 토픽 등장 비중이 감소하거나 뚜렷한 증가세가 없이 비슷한 비중을 유지하고 있는 토픽을 추출하였다. 다음으로 이들 토픽의 논문 수를 파악하여 현재까지 지속적으로 논문이 증가하고 있는 토픽을 제외하였는데, 이는 해당 주제 분야 내에서 차지하는 비중이 감소하는 토픽이라 하더라도 논문 수가 지속적으로 증가하고 있는 경우라면 여전

〈표 3〉 18개 주제 분야의 시기별 토픽 수

순	주제	토픽 수		순	주제	토픽 수	
		1기	2기			1기	2기
1	정치외교학	11	12	10	행정학/정책학	13	10
2	경제학	12	11	11	지리학	9	10
3	경영학	13	10	12	국제/지역개발	12	14
4	회계학	13	10	13	관광학	10	9
5	사회학	14	15	14	신문방송학	12	12
6	사회복지학	13	11	15	심리과학	21	25
7	인류학	12	11	16	생활과학	12	13
8	교육학	12	13	17	문헌정보학	8	9
9	법학	13	11	18	여성학	15	11

히 성장하고 있는 연구 영역일 가능성을 완전히 배제할 수 없기 때문이다.

연도별 비중 변화 분석을 통해 1차로 총 104개(전체의 46.2%)의 토픽이 선행 유망연구주제 후보군으로 도출되었다. 이 104개의 토픽을 대상으로 연도별 논문 수를 파악하여 논문 수가 계속 증가하고 있는 토픽을 제외하였으며, 기준 연도는 2020년으로 하였다. 2020년 이후로도 논문 수가 계속 증가하는 토픽을 제외한 결과 최종적으로 45개(전체 토픽의 20.0%)의 토픽이 도출되었으나, 이 중 논문 수가 최고점에 도달한 연도를 기준으로 과거 25년 치의 논문 수가 1,000건이 채 되지 않는 3개의 토픽(인류학의 'land property', 행정학/정책학의 'public accountability', 여성학의 'homeless women')을 제외하였으며, 결과적으로 총 42개(전체 토픽의 18.7%)의 토픽을 선행 유망연구주제로 최종 선정하였다. 이 42개의 선행 유망연구주제와 영역별로 논문 수가 최고점에 도달한 연도는 다음 〈표 4〉와 같다.

4.2.2 미래 유망연구주제 후보군

다음으로 선행 유망연구주제와의 상관관계

분석을 통해 미래 유망연구주제 도출에 활용할 대상으로서의 미래 유망연구주제 후보군을 선정하였다. 선정은 2기의 토픽을 대상으로 하였으며, 영역별로 토픽 내에 등장하는 핵심 키워드의 등장 시기를 기준으로 하였다. 이는 각 영역의 시기별 상위 100개의 키워드를 기준으로 1기에는 등장하지 않으면서 2기에 새롭게 등장한 키워드가 각 토픽의 핵심 키워드에 많이 포함되어 있는 경우 상대적으로 영향력이 크지 않은 새로운 연구들이 많이 포함되어 있을 가능성이 크다고 판단하였기 때문이다. 또한 앞서 유망연구주제의 탐지와 관련된 선행연구 중 키워드 분석을 활용한 연구들에서 확인할 수 있는 바와 같이 유망연구주제의 경우 해당 주제를 대표하는 키워드의 사용이 증가한다는 점을 고려하였다. 따라서 2기 토픽의 핵심 키워드 5개 가운데 상위 100개 키워드에 처음으로 등장한 키워드가 3개 이상 출현하는 토픽을 미래 유망연구주제 후보군으로 선정하였다. 이렇게 도출한 미래 유망연구주제 후보군은 다음 〈표 5〉와 같으며, 총 41개의 토픽이 도출되었다.

〈표 4〉 선행 유망연구주제

번호	주제	토픽명	논문 수 최고점 연도	번호	주제	토픽명	논문 수 최고점 연도
1	정치	international dispute	2020	22	행정학/	public service motivation	2019
2	외교학	identity politics	2020	23	정책학	local government	2020
3	경제학	cultural economy	2020	24	지리학	economic geography	2019
4	회계학	bank debt	2020	25		climate change geography	2019
5		price mode	2020	26		emotional geography	2019
6	사회학	gender differences	2020	27		urban geography	2019
7		sex offender	2020	28	국제/ 지역개발	crash injury severity	2020
8		children's education	2020	29		pedestrian safety	2020
9	사회 복지학	autism	2020	30		safety performance	2020
10		language disabled children	2020	31		Africa land politics	2020
11		cerebral palsy	2020	32		house politics	2020
12		intellectual disability	2020	33	관광학	visitor satisfaction	2020
13		problem behavior syndrome	2020	34		cultural tourism	2020
14	인류학	African-American discrimination	2020	35		travel socialization	2020
15		acculturation	2020	36		seasonality	2020
16		land grabbing	2019	37		tourism model	2020
17		rural development	2019	38	신문 방송학	information process model	2020
18	교육학	intellectual disability	2020	39		media opinion	2020
19		autism	2020	40		media avoidance	2020
20	행정학/ 정책학	performance goal	2020	41		politics agenda	2020
21		job satisfaction	2020	42		online communication	2020

〈표 5〉 미래 유망연구주제 후보군

번호	주제	토픽명	핵심 키워드
1	정치 외교학	Indo-Pacific strategy and China	<u>china</u> , <u>indopacific</u> , <u>power</u> , <u>strategy</u> , <u>region</u>
2		Indo-Pacific and Asean	<u>indopacific</u> , <u>asean</u> , <u>state</u> , <u>foreign</u> , <u>region</u>
3		internal norm	<u>intern</u> , <u>norm</u> , <u>state</u> , <u>law</u> , <u>human</u>
4	회계학	cryptocurrency	<u>cryptocurrency</u> , <u>bitcoin</u> , <u>trade</u> , <u>carbon</u> , <u>data</u>
5		stock market	<u>market</u> , <u>stock</u> , <u>social</u> , <u>policy</u> , <u>uncertainty</u>
6		volatility spillover	<u>covid</u> , <u>volatil</u> , <u>pandemic</u> , <u>spillover</u> , <u>network</u>
7	사회학	transnational sustainability	<u>sustainability</u> , <u>transnational</u> , <u>migrant</u> , <u>citizenship</u> , <u>stakeholder</u>
8		circular economy	<u>social</u> , <u>theory</u> , <u>economy</u> , <u>circular</u> , <u>identity</u>
9		sustainable environment	<u>sustainable</u> , <u>environment</u> , <u>green</u> , <u>development</u> , <u>technology</u>
10	사회 복지학	intimate partner violence	<u>violence</u> , <u>partner</u> , <u>work</u> , <u>intimate</u> , <u>return</u>
11		mental health	<u>health</u> , <u>mental</u> , <u>trauma</u> , <u>care</u> , <u>disorder</u>
12	인류학	rural politics	<u>polity</u> , <u>rural</u> , <u>population</u> , <u>youth</u> , <u>transition</u>
13	법학	faithful execution	<u>law</u> , <u>faith</u> , <u>execution</u> , <u>article</u> , <u>violence</u>
14		intersectionality theory	<u>theory</u> , <u>intersection</u> , <u>board</u> , <u>stagger</u> , <u>dominant</u>
15		holistic defense	<u>expert</u> , <u>defense</u> , <u>proficiency</u> , <u>holistic</u> , <u>criminal</u>
16		bankruptcy	<u>risk</u> , <u>bankruptcy</u> , <u>digital</u> , <u>wallet</u> , <u>consume</u>

번호	주제	토픽명	핵심 키워드
17	행정학/ 정책학	environmental policy	policy, carbon, china, environment, economy
18		covid-19 policy	policy, covid, social, capacity, integration
19		EU governance	european, union, interest, govern, trilogu
20		student subsidy	college, student, increase, program, benefit
21		climate change policy	climate, change, policy, mitigation, adaptation
22		immigrant children education	immigrant, increase, applicant, children, education
23		Paris agreement	agreement, climate, school, paris, student
24	지리학	health geography	information, hiv, male, economy, market
25		more-than-human geography	asia, southeast, urban, rural, morethanhuman
26	국제/ 지역개발	entrepreneurship education	performance, digital, entrepreneurship, education, transformation
27		sustainable development	sustainable, develop, environment, green, goal
28	관광학	emotional tourism	tourism, emotion, wellbeing, change, learn
29		tourism share economy	tourism, airbnb, share, hospitality, economy
30	신문 방송학	visual pattern	pattern, evaluation, differ, brand, visual
31		product image	product, advertise, image, agency, benefit
32		advertise algorithm	advertise, algorithm, approach, user, patern
33		gender in advertising	advertise, effect, gender, role, consumption
34	생활과학	IPV risk factor	minor, risk, factor, childhood, ipv
35	여성학	gender inequality	gender, inequality, change, organization, equality
36		gender stereotype	sexism, gender, stereotype, prejudice, identity
37		gender pay gap	gender, gap, discrimination, pay, social
38		covid and gender intersectionality	intersection, gender, covid, race, career
39		sexual objectification	sexual, harassment, objectification, assault, gender
40		neoliberal feminism	feminism, neoliberalism, gender, affect, feminist
41		self-objectification	social, body, media, image, selfobjectification

※ 2기에 처음 등장한 키워드는 밑줄로 표시함

4.3 선행 유망연구주제의 주요 특성 및 공통 추세 분석

4.3.1 선행 유망연구주제의 주요 특성

이상의 과정을 거쳐 최종 수집된 42개 선행 유망연구주제의 특징을 나타내는 지표들의 시계열적 패턴을 분석하였다. 선행 유망연구주제의 특징을 나타내는 지표로 계량서지 지표는 논문 수, 저자 수, 인용 수, 피인용 수, 신규 키워드 수를, altmetrics 지표는 PlumX Analytics의 PlumX Metrics(citations, usage, captures, mentions, social media) 지표들을 활용하였다. 연구주제별로 논문 수가 최고점에 달한 연도를

기준 시점으로 하여 과거 30년간의 데이터를 확인하여 각 지표들의 시계열 패턴을 확인하였다.

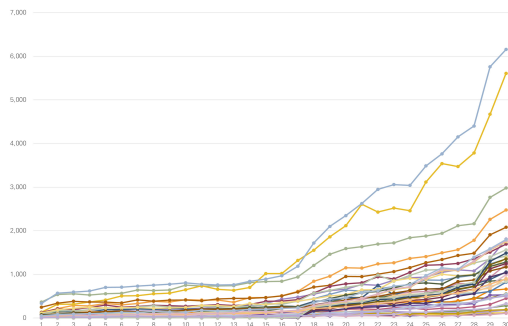
(1) 계량서지(bibliometrics) 지표

42개 선행 유망연구주제의 계량서지 지표값은 다음 <표 6>과 같다.

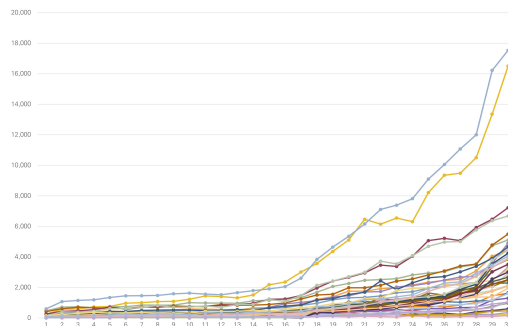
과거 30년간의 총 논문 수는 495,652건, 연평균 논문 수는 11,801건이었다. 논문 수 최고점 연도를 기준으로 과거 30년간 생산된 논문 수의 시계열적 패턴은 <그림 1>과 같으며, 논문 수 최고점 시점을 기준으로 약 13~14년 전 시점 이후부터 논문 수가 급격하게 증가하는 것으로 볼 수 있다.

〈표 6〉 선행 유망연구주제(42개)의 계량서지 지표값

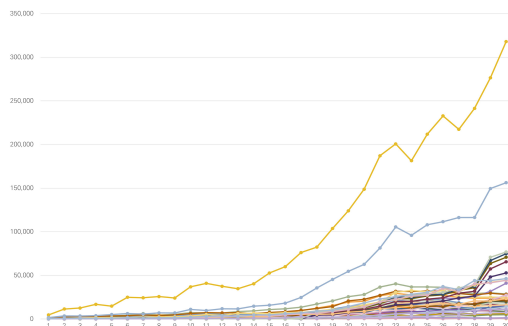
구분	논문 수	저자 수	이용 수	피인용 수	신규 키워드 수
총계	459,652	1,151,139	13,008,537	15,600,591	1,835,302
연평균	11,801	27,408	309,727	371,443	43,698
논문 1편당	-	2.3	26.2	31.5	3.7



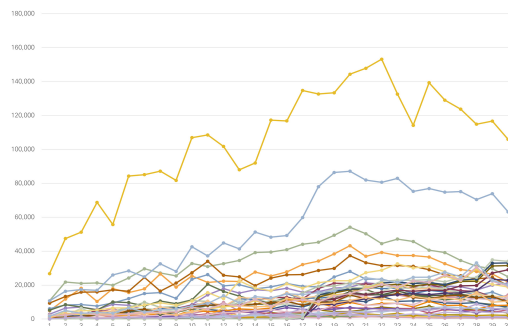
〈그림 1〉 선행 유망연구주제의 30년간 논문 수



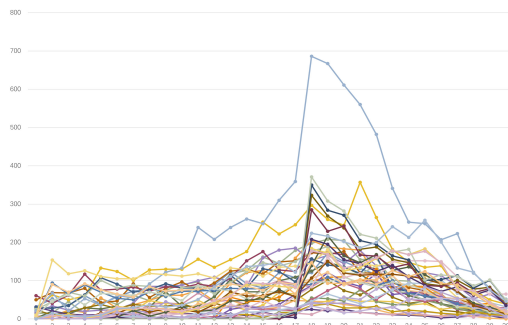
〈그림 2〉 선행 유망연구주제의 30년간 저자 수



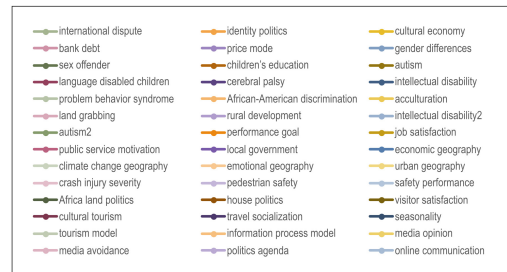
〈그림 3〉 선행 유망연구주제의 30년간 이용 수



〈그림 4〉 선행 유망연구주제의 30년간 피인용 수



〈그림 5〉 선행 유망연구주제의 30년간 신규 키워드 수



과거 30년간의 총 저자 수는 1,151,139명, 연평균 저자 수는 27,408명이었으며, 논문 1편당 저자 수는 평균 2.3명이었다. 42개 연구주제별 저자 수의 시계열적 패턴은 다음 <그림 2>와 같다. 저자 수 역시 논문 수와 비슷한 추세의 증가세를 보이고 있으며, 논문 수와 마찬가지로 논문 수 최고점 시점 기준 약 13~14년 이전 시점부터 증가하기 시작하는 것을 볼 수 있다. 이는 연구 성과의 대표적인 예로서 논문의 수가 증가하는 것과 동시에 해당 주제의 연구에 참여하는 연구자의 수가 함께 증가하기 때문인 것으로 판단할 수 있다.

이어서 과거 30년간의 연도별 이용 수를 파악하였다. 총 이용 수는 13,008,537건이었으며, 연평균 309,727건이 이용되는 것으로 나타났고, 논문 1편당 평균 이용 수는 26.2건이었다. 이용 수의 시계열 패턴은 <그림 3>과 같다. 이용 수 역시 논문 수나 저자 수와 증가 추세가 크게 다르지 않은 모습을 보이고 있어 논문 수의 증가와 함께 거의 동시에 이용이 증가하는 것으로 볼 수 있다. 구체적인 시점을 보면, 논문 수 최고점 시점을 기준으로 약 12~13년 전 시점 이후부터 급격하게 증가하는 것으로 판단할 수 있어, 시기적으로 보았을 때, 논문 수와 저자 수가 증가하고 난 1~2년 이후에 이용 수가 증가하는 것으로 판단할 수 있다. 다시 말해 유망연구주제의 경우에는 논문이 출판되고 난 이후 수개월에서 1~2년 이내에 이용되는 경향이 있다고 볼 수 있다.

다음으로 과거 30년간 피인용 수를 분석한 결과, 총 피인용 수는 15,600,591건이었으며, 연평균 371,443건의 인용을 받는 것으로 나타났고, 논문 1편당 평균 피인용 수는 31.5건이었다.

피인용 수의 경우, <그림 4>와 같이 지속적, 선형적으로 증가하다가 평균적으로 논문 수가 최고점에 이르는 시기보다 8~10년 정도 이전 시점에 가장 높은 피인용 수를 나타내고 있으며, 이후로 서서히 감소하는 모습을 볼 수 있다. 출판된 연구의 피인용이 특정 시점에 가장 많다는 것은 이 시기에 생산된 연구가 해당 연구주제를 대표하는 핵심 연구들이라고 해석할 수 있다. 즉, 유망연구주제의 경우, 논문 수 최고점 시점을 기준으로 8~10년 이전 시기에 해당 연구주제를 대표하는 핵심 연구들이 많이 생산되는 것으로 볼 수 있으며, 따라서 해당 시기가 연구의 영향력이 가장 큰 시기라고 판단할 수 있다.

마지막으로 과거 30년간 매년 신규로 등장하는 키워드의 수를 파악하였다. 키워드는 저자 키워드만을 대상으로 하였다. 제시된 총 키워드의 수는 1,835,302개였으며, 논문 1편당 평균 3.7개의 키워드가 제시되어 있었다. 이 중 중복을 제외한 총 키워드 수는 727,611개였으며, 이들을 대상으로 영역별로 빈도수 상위 키워드의 최초 출현 시점을 분석하였다. 해당 영역에서 많이 제시된 해당 영역을 대표하는 핵심 키워드의 최초 출현 시점과 그 증감의 경향을 통해 선행 유망연구주제의 시계열적 추세를 파악하고자 하였다. 다만 영역별로 제시된 키워드의 수가 상이하므로 등장 빈도 상위 10% 키워드만을 분석에 활용하였다. 신규 키워드 등장 시점의 시계열 패턴을 분석한 결과(<그림 5> 참조), 영역별로 많이 등장하는 핵심 키워드들이 최초로 등장하는 시점은 논문 수 최고점 시점을 기준으로 11~13년 이전 시점인 것으로 나타났다. 즉, 유망연구주제의 경우, 논문 수 최고

점 시점 이전 11~13년에 해당 주제를 대표하는 용어들이 해당 연구 커뮤니티에 인지되고, 사용되기 시작하는 것으로 판단할 수 있다.

(2) altmetrics 지표

다음으로 본 연구에 활용할 altmetrics 지표로 PlumX Metrics의 5개 지표(citations, usage, captures, mentions, social media)를 수집하였다. 앞서 설명한 바와 같이 altmetrics 지표는 Scopus에서 제공하는 API를 활용하는 파이썬 라이브러리인 PlumXMetrics()를 활용하여 수집하였으며, 선행 유망연구주제의 서지정보에 포함되어 있는 DOI를 사용하였다. 그러나 서지정보에 DOI가 누락되어 있거나 DOI에 오류가 있어 PlumX Metrics 지표를 확인할 수 없는 경우가 있었다(〈표 7〉 참조). 그러나 누락 건수와 오류 건수가 총 44,552건으로 전체의 8.99%에 불과하여 해당 경우를 제외하고 altmetrics 지표를 수집하여 분석에 활용하였다.

42개 선행 유망연구주제의 altmetrics 지표값은 다음 〈표 8〉과 같다.

과거 30년간의 총 citation 수는 17,578,659

건이었으며, 연평균 citation 수는 418,540건으로 나타났고, 논문 1편당 평균 citation 수는 38.8건이었다. 논문 수 최고점 연도를 기준으로 과거 30년간 생산된 논문 수의 시계열적 패턴은 다음 〈그림 6〉과 같다. citation 수의 경우, 평균적으로 논문 수가 최고점에 이르는 시기보다 8~10년 정도 이전 시점에 가장 높으며, 이후 서서히 감소하는 모습을 볼 수 있다. citation 수는 앞서 살펴본 계량서지 지표의 피인용 수와 유사한 패턴을 보이고 있는데, 이는 PlumX가 제공하는 citation 지표의 경우, 정책문서 인용이나 특허 인용을 포함하고는 있으나 핵심이 되는 주요 정보원이 인용 색인 및 원문 데이터 베이스들로 정보원이 중복되기 때문인 것으로 판단된다.

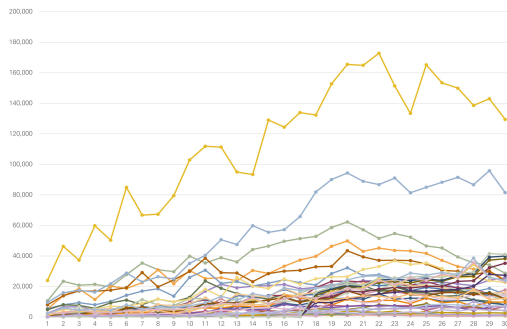
다음으로 과거 30년간의 연도별 usage 수를 파악한 결과, 총 usage 수는 514,444,703회였으며, 연평균 usage 수는 12,248,683회로 나타났고, 논문 1편당 평균 usage 수는 1,136.8회였다. usage 수의 시계열 분석 결과(〈그림 7〉 참조), 대체로 논문 수가 최고점에 이르는 시기를 기준으로 약 13년 전 이전 시점에 급격히 이용이 증

〈표 7〉 분석 대상 전체 논문 대비 DOI 누락 및 오류가 발생한 논문 비율

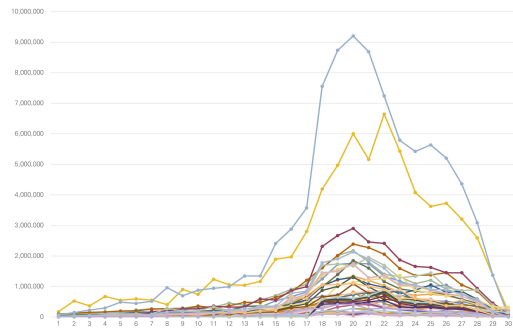
구분	총 논문 수	DOI 누락 건수	DOI 오류 건수	분석에 활용한 DOI 건수
계(건)	495,652	43,125	1,427	451,100
비율(%)	100	8.70	0.32	91.01

〈표 8〉 선행 유망연구주제(42개)의 altmetrics 지표값

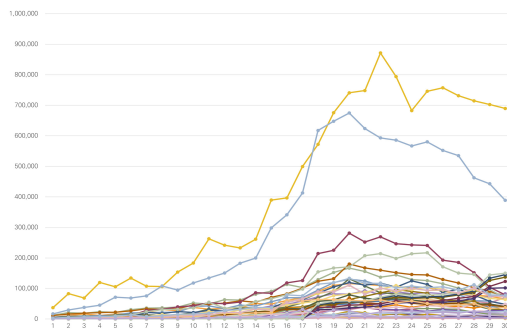
구분	citations	usage	captures	mentions	social media
총계	17,578,659	514,444,703	64,394,470	192,379	38,713,919
연평균	418,540	12,248,683	1,533,202	4,581	921,760
논문 1편당	38.8	1,136.8	142.3	0.4	85.6



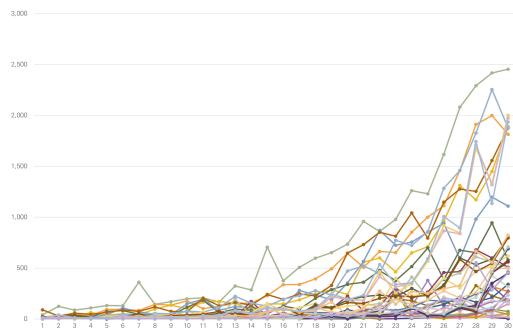
〈그림 6〉 선행 유망연구주제의 30년간 citation 수



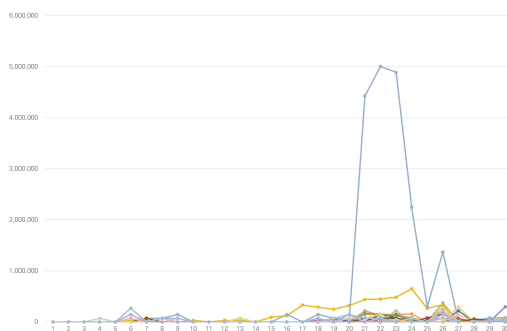
〈그림 7〉 선행 유망연구주제의 30년간 usage 수



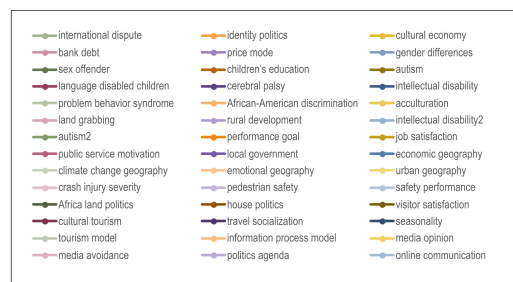
〈그림 8〉 선행 유망연구주제의 연도별 capture 수



〈그림 9〉 선행 유망연구주제의 연도별 mention 수



〈그림 10〉 선행 유망연구주제의 연도별 social media 수



가하였으며, 이용이 급격히 증가한 이후 3~5년 내에 최고점에 이르고 이후 급격히 감소하는 것으로 판단된다. 특이한 점은 altmetrics 지표의

usage 수는 계량서지 지표의 이용 수와는 다르게 이용이 정점을 찍은 이후 급격한 하락세로 돌아서는 추세를 보인다는 것이다. 이는 altmetrics

지표의 정보원이 학술 데이터베이스뿐만 아니라 프리프린트(preprint) 서버나 리포지토리를 다수 포함하고 있다는 데에서 그 이유를 찾을 수 있을 것이다. 물론 학문 분야마다 프리프린트 서버나 리포지토리 등록 및 이용 비중에 편차가 있을 수 있으므로 해석에 주의할 필요는 있다. 그러나 일반적으로 많은 연구자들이 연구 성과가 출판되기 이전에는 다양한 프리프린트 서버나 리포지토리에 연구 성과를 등록, 공개하면서 이를 통한 이용이 급증하다가 저널에 논문이 게재되거나 공개되고 난 이후에는 학술 데이터베이스 등을 통한 이용이 증가하면서 altmetrics 지표의 usage 수는 감소하는 것으로 판단할 수 있다.

이어서 과거 30년간의 연도별 capture 수를 파악하였다. 총 capture 수는 64,394,470회였으며, 연평균 capture 수는 1,533,202회, 논문 1편당 평균 capture 수는 142.3회였다. capture 수는 대체로 논문 수가 최고점에 이르는 시기보다 약 10~11년 전 이전 시점에 최고점에 달하는 것으로 판단할 수 있으며, 이후 감소하는 모습을 보인다(〈그림 8〉 참조).

다음으로 과거 30년간의 mention 수는 총 192,379회였으며, 연평균 mention 수는 4,581회, 논문 1편당 평균 mention 수는 0.4회였다. mention 수의 경우, 연구 분야 간 편차가 존재하였다. 다른 영역에 비해 상대적으로 평균 mention 수나 논문 1편당 mention 수가 많은 영역은 정치외교학, 사회학, 신문방송학 분야였는데, 이는 사회과학 분야 중에서도 현실과 밀접한 관련이 있거나 현실을 반영하는 연구 문제들을 많이 다루는 분야들로 다른 영역에 비해 뉴스 및 언론 보도나 블로그 등에 상대적으

로 많이 언급되기 때문인 것으로 볼 수 있다. 시계열 패턴 분석 결과(〈그림 9〉 참조), mention 수의 경우, 논문 수가 최고점에 이르는 시기보다 약 13~14년 이전 시점부터 증가하기 시작하여 논문 수 최고점 시점까지도 계속 증가하는 모습을 보이는 것이 특징이다. 이는 사회과학 분야의 특성상, 선행 유망연구주제들이 연구주제가 성장하는 단계에서는 각 학문 영역 내부에서만 논의되어 오다 이후 사회적으로 관심을 받고 사회가 주목하는 이슈나 쟁점이 되면서 언론 보도나 블로그 게시물 등으로 이어지기 때문인 것으로 판단된다.

마지막으로 과거 30년간의 연도별 social media 수는 총 38,713,919회였으며, 연평균 social media 수는 921,760회, 논문 1편당 평균 social media 수는 85.6회였다. social media 수의 시계열적 패턴을 보면(〈그림 10〉 참조), 대체로 논문 수가 최고점에 이르는 시기보다 약 6~7년 정도 이전 시점에 가장 많은 것으로 나타났다. 그러나 다른 PlumX Metrics 지표들과는 달리 social media 수의 경우 연구주제별 편차가 매우 큰 것이 특징인데, 논문 1편당 social media 수가 가장 많았던 영역은 교육학으로 평균 383.1회의 intellectual disability와 평균 249.0회의 autism이었고, 가장 적었던 영역은 인류학 주제의 African-American discrimination, acculturation, land grabbing, rural development로, 모두 평균 social media 수가 10회가 채 되지 않는 모습을 보였다. 이는 많은 연구자가 altmetrics의 단점으로 지적하고 있는 바와 같이 연구의 영향력이나 연구 자체가 가지고 있는 가치와는 별개로 포착이 어려운 altmetrics 지표, 특히, social media에서 언급이 잘 되지

않는 일부 연구주제가 존재한다는 것을 확인할 수 있다.

4.3.2 선행 유망연구주제의 공통 추세

다음으로 앞서 살펴본 42개 연구주제의 계량서지 지표와 altmetrics 지표의 공통적 추세를 생성하기 위해 주성분분석을 실시하였다. 선행 유망연구주제의 특징 지표들은 영역별 논문 수가 최고점에 달한 연도의 30년 전 시점부터 5년 전 시점까지 25년 치의 데이터를 활용하였다. 이는 본 연구에서 최종적으로 미래 유망연구주제로 도출되는 주제들은 선행 유망연구주제의 공통적 추세와 유사한 패턴으로 성장하고 있는 주제로, 현재 시점까지의 추세가 선행 유망연구주제의 추세와 유사하다면 적어도 향후 5년까지는 성장하는 연구주제가 될 가능성이 크다는 것을 의미하기 때문이다. 그러나 현재 시점을 분석에 포함하게 되는 경우, 이미 성숙 단계에 이른 주제들이 미래 유망연구주제로 도출될 가능성을 배제할 수 없고, 이는 유망연구주제의 예측이라는 측면에 적합하지 않으므로 논문 수 최고점 연도의 5년 전 시점을 기준으로 과거 25년 치의 데이터를 활용하였다.

주성분분석에 앞서 42개의 선행 유망연구주

제 분야들의 각 지표값에는 편차가 있으므로 이들 데이터를 정규화하는 과정을 거쳤으며, 주성분분석을 실시한 이후 첫 번째 주성분을 기준으로 하여 각 지표를 대표하는 단일 차원의 새로운 지표를 생성하고, 그 추세를 그래프로 나타내었다.

(1) 계량서지 지표의 공통적 추세

공통적 추세 분석을 위해 활용한 25년치의 계량서지 지표(논문 수, 저자 수, 인용 수, 피인용 수, 신규 키워드 수)에 대한 기술통계분석 결과는 다음 <표 9>와 같다.

다음으로 계량서지 지표를 대상으로 주성분분석을 실시하였으며, 지표별로 주성분 1의 고유값(eigenvalue), 차이(difference), 분산비율(proportion)을 <표 10>에 나타내었다. 논문 수의 경우, 첫 번째 주성분이 전체 데이터 분산의 약 93.4%를 설명한다고 할 수 있으며, 논문 수, 저자 수, 인용 수의 경우 설명력이 90%를 상회하는 수준으로 주성분의 개수를 하나만 활용해도 논문 수, 저자 수, 인용 수 각각의 전체 데이터를 충분히 설명할 수 있다고 볼 수 있다. 반면 키워드 수의 경우, 주성분 1의 설명력이 약 69.1%로 다른 지표에 비해 약간 낮은 모습을 보이고 있는 것이 특징이다.

<표 9> 계량서지 지표들의 기술통계분석 결과

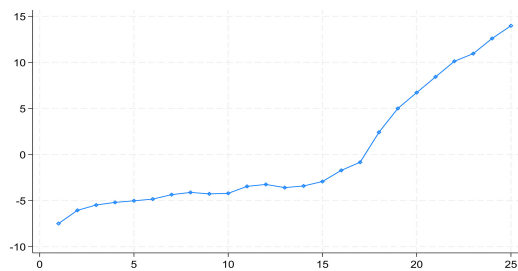
구분	논문 수	저자 수	인용 수	피인용 수	신규 키워드 수
전체	291,070	620,678	7,167,761	11,976,455	71,017
연평균	11,643	24,827	28,670	479,058	2,841
표준편차	8,193.09	20,438.12	307,058.02	231,613.89	1,817.12
최댓값	3,488	9,103	211,831	153,166	686
최솟값	0	0	0	0	0
0인 데이터(%)	35(3.33)	35(3.33)	35(3.33)	35(3.33)	135(12.86)

〈표 10〉 계량서지 지표들의 주성분분석 결과(주성분 1)

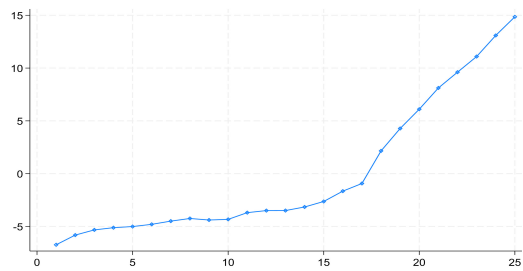
지표	eigenvalue	difference	proportion
논문 수	43.8773	42.4260	0.9336
저자 수	43.7570	42.0237	0.9310
이용 수	45.1664	44.4966	0.9610
피인용 수	35.7821	32.1433	0.7613
키워드 수	32.4527	27.3694	0.6905

이상의 과정을 거쳐 생성된 논문 수, 저자 수, 이용 수, 피인용 수, 신규 키워드 수의 5개 지표

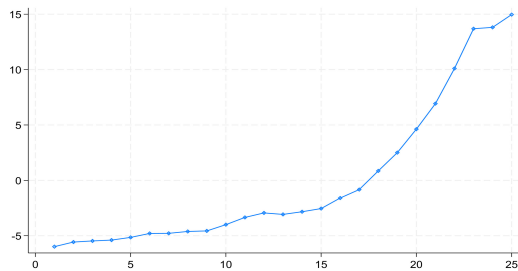
의 공통적 추세를 그래프로 나타내면 다음 〈그림 11〉~〈그림 15〉와 같다.



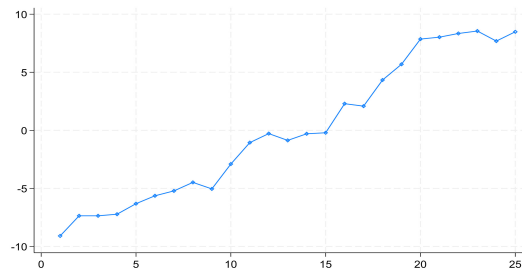
〈그림 11〉 논문 수의 공통적 추세



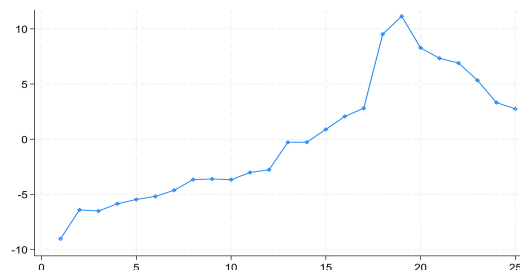
〈그림 12〉 저자 수의 공통적 추세



〈그림 13〉 이용 수의 공통적 추세



〈그림 14〉 피인용 수의 공통적 추세



〈그림 15〉 신규 키워드 수의 공통적 추세

분석 결과, 논문 수와 저자 수는 논문 수 최고점을 기준으로 14~15년 전 시점 이전까지는 완만한 성장세를 보이다가 그 이후에 급증하는 추세를 보이고 있다. 이는 해당 연구주제가 성장하면서 대표적인 연구 성과인 논문의 수가 증가함과 동시에 연구에 참여하는 연구자의 수도 함께 늘어나는 것이라고 볼 수 있다. 다음으로 인용 수는 논문 수 최고점을 기준으로 13~14년 전 시점에 급격히 증가하여 7년 전 시점까지도 계속 증가하고 있다. 피인용 수는 논문 수 최고점을 기준으로 지속적, 선형적으로 계속 증가하다가 10년 전 시점을 지나 그 이후부터는 큰 증가세 없이 비슷한 수준을 유지하는 것으로 볼 수 있으며, 연구주제의 핵심 키워드의 수는 논문 수 최고점 시점에서 약 12~13년 전에 급증하는 것으로 볼 수 있다.

이를 종합해 보면, 유망연구주제는 논문 수가 최고점에 이르는 시기보다 약 13~14년 전 시점부터 논문 수와 저자 수가 증가하기 시작하여, 논문 수와 저자 수 증가 이후 약 1~3년 후에 인용 수가 급증하고, 이후 3~5년 후에 신규 키워드의 수가 급증하는 모습을 보이고 있다. 따라서 논문 수가 최고점에 이르는 시점을 기준으로 최대 13~14년 전 시점에야 유망연구주제의 예측이 가능할 것으로 판단할 수 있

며, 기존의 계량서지 지표를 활용하는 경우, 인용 분석과 키워드 분석을 많이 활용하므로 일반적으로 유망연구주제의 예측이 가능한 시점은 논문 수 최고점을 기준으로 최대 10~13년 이전 시점으로 볼 수 있다.

(2) altmetrics 지표의 공통적 추세

본 연구에서 활용한 altmetrics 지표로서 25년 치의 PlumX Metrics(citations, usage, captures, mentions, social media) 지표에 대한 기술통계분석 결과는 <표 11>과 같다.

altmetrics 지표를 대상으로 주성분분석을 실시하였으며, 지표별로 주성분 1의 고유값(eigenvalue), 차이(difference), 분산비율(proportion)을 <표 12>에 나타내었다. usage 수와 capture 수의 경우, 첫 번째 주성분이 전체 데이터 분산의 약 92.5%, 88.8%를 설명한다고 볼 수 있으며, 따라서 이 두 지표의 경우 주성분의 개수를 하나만 활용해도 각 지표를 충분히 설명할 수 있다고 할 수 있다. 반면 social media 수의 경우, 주성분 1의 설명력이 약 30.6%로 다른 지표에 비해 확연히 낮은 모습을 보이고 있다. 이는 주성분분석이 여러 변수 간의 상관관계를 토대로 상관관계가 높은 변수들을 하나로 묶는데, social media 지표의 경우, 지표값이 0인 경우가 전체

<표 11> altmetrics 지표들의 기술통계분석 결과

구분	citations	usage	captures	mentions	social media
전체	13,291,125	434,251,523	47,021,361	85,504	30,553,033
연평균	531,645	17,370,061	1,880,854	3,420	1,222,121
표준편차	276,446.57	17,676,909.74	1,582,862.99	3,384.56	2,205,430.07
최댓값	172,772	9,202,530	871,217	1,261	5,005,467
최솟값	0	0	0	0	0
0인 데이터(%)	42(4.00)	43(4.10)	43(4.10)	114(10.86)	288(27.43)

〈표 12〉 altmetrics 지표들의 주성분분석 결과(주성분 1)

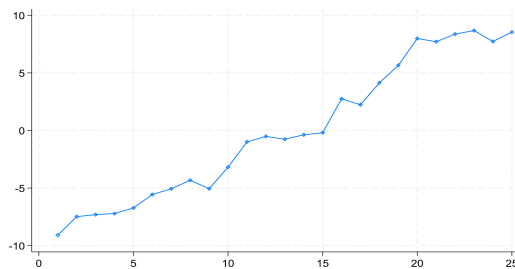
지표	eigenvalue	difference	proportion
citations	36.0425	32.2330	0.7669
usage	43.4523	42.1810	0.9245
captures	41.7242	39.8385	0.8877
mentions	31.7545	27.3965	0.6756
social media	14.3984	5.3816	0.3063

의 27.43%에 달하고, 지표값을 측정하는 정보원의 특성상 주제나 시기별로 지표값의 편차가 큰 데서 그 원인을 찾을 수 있을 것이다. 그러나 본 연구에서의 주성분분석의 목적은 차원 축소를 통하여 하나의 공통적 추세를 생성하는 것이므로 상대적인 중요도가 제일 큰 주성분 1을 그대로 활용하였다.

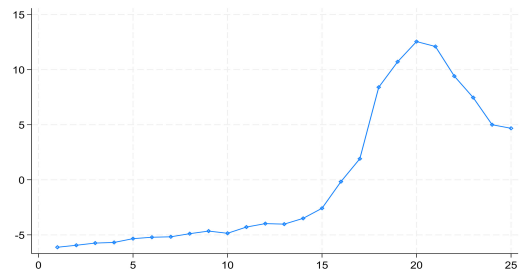
이상의 과정을 거쳐 생성된 citations, usage, captures, mentions, social media의 5개 지표

의 공통적 추세를 그래프로 나타내면 다음 〈그림 16〉~〈그림 20〉과 같다.

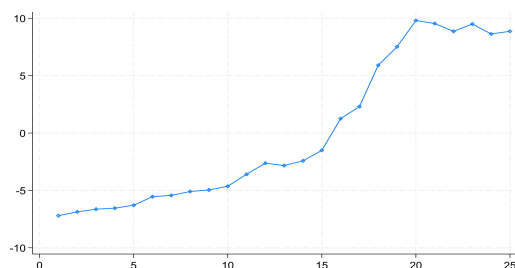
citation 수는 논문 수 최고점을 기준으로 10년 전 이전 시점까지는 선형적으로 성장하는 모습을 보이며, 그 이후에는 비슷한 수준을 유지하는 것으로 나타났다. 다음으로 usage 수는 논문 수 최고점을 기준으로 15년 전 시점에 급격히 증가하기 시작하여 10년 전 시점에 최고점에 도달하며, 이후 급격히 감소하는 것으로



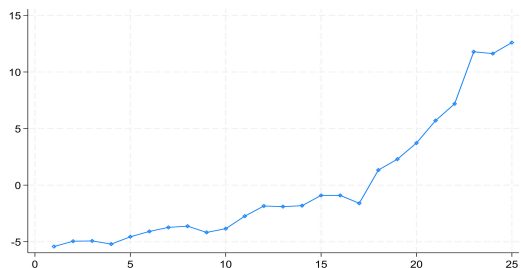
〈그림 16〉 citation 수의 공통적 추세



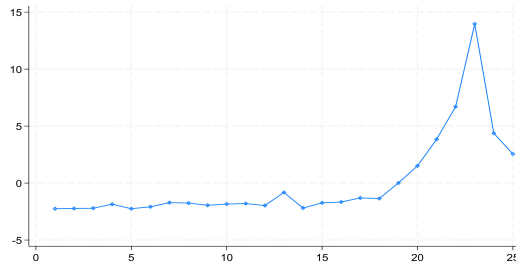
〈그림 17〉 usage 수의 공통적 추세



〈그림 18〉 capture 수의 공통적 추세



〈그림 19〉 mention 수의 공통적 추세



〈그림 20〉 social media 수의 공통적 추세

나타났다. capture 수는 최고점에 이르기까지는 usage 수와 유사한 성장세를 보이지만 최고점에 도달한 이후 비슷한 수준을 유지하며 서서히 감소하는 것으로 보인다. 다음으로 mention 수는 논문 수 최고점 시점으로부터 약 13년 전 시점에 급증하여 7년 전 시점에 최고점에 이르며, social media 수는 논문 수 최고점 시점보다 약 11~12년 전 시점까지는 거의 변화가 없다가 이후 증가하기 시작하여 7년 전 시점에 최고점에 이르고 있다. mention 수와 social media 수 역시 최고점에 이르는 시점은 유사하나 해당 시점 이후 mention 수의 경우 비슷한 수준을 유지하거나 상승하는 데 반해 social media 수의 경우 급감하는 모습을 보이고 있는 것이 차이점이다.

4.4 유망연구주제 탐지를 위한 선행지표로서 altmetrics 지표의 유용성 검증

4.4.1 교차상관 분석

다음으로 선행 유망연구주제의 계량서지 지표와 altmetrics 지표 간에 시차 간 관련성이 있는지 확인하기 위하여 교차상관 분석을 통한 시계열 상관관계 분석을 수행하였다. 시계열 데이터의 상관관계 분석을 위해서는 비정상 시계열을 정상 시계열로 변환해야 한다. 따라서 주성분분석을 통해 생성된 계량서지 지표와 altmetrics 지표의 공통 추세의 비정상성을 판단하기 위해 자기상관분석을 수행하였다. 분석 결과에 따라 전체 데이터를 차분 변환하여 정상 시계열로 변환한 후 교차상관 분석을 실시하였으며, 분석 결과는 〈표 13〉과 같다. 양의 시차가 발생하는 경우 계량서지 지표가 altmetrics 지표보다 선

〈표 13〉 계량서지 지표와 altmetrics 지표의 교차상관 분석 결과(시차)

구분		altmetrics 지표				
		citations	usage	captures	mentions	social media
계량 서지 지표	논문 수	-	+1	-2	+4	+6
	저자 수	-	+1	-2	0	+6
	이용 수	-	-5	-3	0	0
	피인용 수	0	+1	0	-	-
	신규 키워드 수	-	0	0	+5	+4

행하여 나타난다는 의미이며, 음의 시차가 발생하는 경우 계량서지 지표가 altmetrics 지표보다 후행하여 나타난다는 의미로, 음의 시차가 발생하는 경우 계량서지 지표의 선행지표로서 altmetrics를 활용할 수 있다고 해석할 수 있다.

교차상관 분석 결과, 논문 수, 저자 수, 이용 수, 신규 키워드 수는 usage, captures, mentions, social media와 유의한 교차상관이 존재하였으며, 피인용 수는 citations, usage, captures와 유의한 교차상관이 존재하는 것으로 나타났다.

논문 수, 저자 수, 피인용 수는 usage 수보다 1년 선행하는 것으로 나타났는데, 이는 이 세 개의 계량서지 지표값의 변화가 usage 수의 변화보다 먼저 발생하는 것으로 해석할 수 있다. 또한 신규 키워드 수는 usage 수와 시차수가 0으로, 이는 시차에 상관없이 신규 키워드 수의 변화와 usage 수의 변화가 동시에 발생하는 것이라고 할 수 있다. 그러나 이용 수의 변화는 usage 수보다 5년 후행하는 것으로 나타났는데, 이는 usage 수의 변화가 먼저 발생하고 그 5년 후에 이용 수의 변화가 나타나는 것이라고 할 수 있다.

그러나 capture 수의 경우, 논문 수, 저자 수보다 2년을, 이용 수보다는 3년을 후행하는 것으로 나타났으며, 이는 capture 수의 변화가 먼저 발생하고, 2년 후에 논문 수와 저자 수의 변화가, 3년 후에 이용 수의 변화가 나타나는 것이라고 볼 수 있다. 다만 피인용 수와 신규 키워드 수는 capture 수와 시차수가 0으로 피인용

수와 신규 키워드 수는 capture 수와 동시에 변화하는 것으로 해석할 수 있다.

다음으로 mention 수는 논문 수보다 4년을, 신규 키워드 수보다는 5년을 후행하는 것으로 나타났으며, 저자 수와 이용 수는 mention 수와 시차수가 0으로 동행 관계에 있는 것으로 나타났다. social media 수의 경우, 논문 수, 저자 수보다는 6년을, 신규 키워드 수보다는 4년을 후행하는 것으로 나타났으며, 이용 수와는 동행관계에 있는 것으로 나타났다. 특히 altmetrics 지표 중 citations는 피인용 수 외에는 유의한 교차상관관계가 나타나지 않았으며, 피인용 수와도 선행 관계 없이 동행하는 것으로 나타났다.

4.4.2 회귀분석

이상의 교차상관 분석 결과에 따르면 계량서지 지표에 선행하여 영향을 미치는 altmetrics 지표는 논문 수와 capture 수, 저자 수와 capture 수, 이용 수와 usage 수, 이용 수와 capture 수였으며, 논문 수와 저자 수보다 capture 수가 2년을, 이용 수보다 usage 수가 5년, capture 수가 3년을 선행하는 것으로 나타났다. 이에 따라 선행성이 존재하는 altmetrics 지표들을 대상으로 각 지표별로 해당하는 시차만큼 시기를 조정하여, 회귀모형으로 선행지표인 altmetrics 지표와 계량서지 지표의 영향 관계를 추정하였다.

첫째, 논문 수와 capture 수의 회귀분석 결과는 <표 14>와 같다.

<표 14> capture 수(시차 - 2)가 논문 수에 미치는 영향

변수	비표준화 계수		표준화계수	t(p)	F(p)	R ²
	B	SE	β			
(상수)	1.225	.212		5.776***	787.673***	.975
captures	1.014	.036	.988	28.066***		

*** p<0.001

분석 결과, F 값의 통계량은 787.673이고 유의확률이 0.000으로 이 회귀모형은 유의미한 것으로 나타났다. 회귀모형의 결정계수(R^2)는 0.975로, 97.5%의 설명력을 나타내고 있다고 할 수 있으며, 추정된 모형의 진단을 위한 잔차 분석 결과에 의하면 오차항에 대한 Durbin-Watson 통계량은 0.948이었다. 정확한 분석을 위한 잔차의 정규성 검정 결과, Kolmogorov-Smirnov와 Shapiro-Wilk 통계량의 유의확률이 각각 0.200, 0.784로 0.05보다 커 정규성을 만족하는 것으로 나타났다. 따라서 추정된 회귀모형은 적합하다고 할 수 있다.

둘째, 저자 수와 capture 수의 회귀분석 결과는 다음 <표 15>와 같다.

분석 결과, F 값의 통계량은 1157.466이고 유의확률이 0.000으로 이 회귀모형은 유의미한 것으로 나타났으며, 회귀모형의 결정계수(R^2)는 0.984로, 98.4%의 설명력을 나타내고 있다고 할 수 있다. 추정된 모형의 진단을 위한 잔차 분석 결과에 의하면 오차항에 대한 Durbin-Watson 통계량은 1.157로 자기상관이 존재하지

않는 것으로 나타났다. 정확한 분석을 위한 잔차의 정규성 검정 결과, Kolmogorov-Smirnov와 Shapiro-Wilk 통계량의 유의확률이 각각 .200, .926으로 0.05보다 커 정규성을 만족하는 것으로 나타났다. 따라서 추정된 회귀모형은 적합하다고 할 수 있다.

셋째, 이용 수와 usage 수의 회귀분석 결과는 다음 <표 16>과 같다.

분석 결과, F 값의 통계량은 187.810이고 유의확률이 0.000으로 이 회귀모형은 유의미한 것으로 나타났으며, 회귀모형의 결정계수(R^2)는 0.917로, 91.7%의 설명력을 나타내고 있다고 할 수 있다. 그러나 추정된 모형의 진단을 위한 잔차 분석 결과에 의하면 오차항에 대한 Durbin-Watson 통계량은 0.303으로 자기상관이 존재하는 것으로 나타났고, 잔차의 정규성 검정 결과 역시, Kolmogorov-Smirnov와 Shapiro-Wilk 통계량의 유의확률이 각각 .014, .003으로 0.05보다 작아 정규성을 만족하지 않는 것으로 나타났다. 따라서 회귀모형의 기본 가정을 충족하지 못하므로 추정된 회귀모형은 적합하다

<표 15> capture 수(시차 - 2)가 저자 수에 미치는 영향

변수	비표준화 계수		표준화 계수	t(p)	F(p)	R^2
	B	SE	β			
(상수)	.890	.159		5.602***	1157.466***	.984
captures	.950	.028	.992	34.022***		

*** $p < 0.001$

<표 16> usage 수(시차 - 5)가 이용 수에 미치는 영향

변수	비표준화 계수		표준화 계수	t(p)	F(p)	R^2
	B	SE	β			
(상수)	3.268	.488		6.692***	187.810***	.917
usage	1.101	.080	.958	13.704***		

*** $p < 0.001$

〈표 17〉 capture 수(시차 - 3)가 이용 수에 미치는 영향

변수	비표준화 계수		표준화 계수	t(p)	F(p)	R^2
	B	SE	β			
(상수)	2.106	.165		12.731***		
captures	1.125	.029	.994	38.609***	1490.628***	.988

*** p<0.001

고 할 수 없다.

넷째, 이용 수와 capture 수의 회귀분석 결과는 다음 〈표 17〉과 같다.

분석 결과, F 값의 통계량은 1490.628이고 유의확률이 0.000으로 이 회귀모형은 유의미한 것으로 나타났으며, 회귀모형의 결정계수(R^2)는 0.988로, 98.8%의 설명력을 나타내고 있다고 할 수 있다. 추정된 모형의 진단을 위한 잔차 분석 결과에 의하면 오차항에 대한 Durbin-Watson 통계량은 1.067이었다. 정확한 분석을 위한 잔차의 정규성 검정 결과, Kolmogorov-Smirnov와 Shapiro-Wilk 통계량의 유의확률이 각각 .171, .089로 0.05보다 커 정규성을 만족하는 것으로 나타났다. 따라서 추정된 회귀모형은 적합하다고 할 수 있다.

4.5 미래 유망연구주제 도출

앞의 회귀분석 결과에 의하면, 계량서지 지표에 선행하는 altmetrics 지표 중 유의한 지표는 capture 수인 것으로 나타났으며, 계량서지 지표 중 논문 수, 저자 수에 대하여는 각각 2년을, 이용 수에 대하여는 3년을 선행하여 영향을 미치는 것으로 나타났다. 따라서 본 장에서는 계량서지 지표를 대상으로 한 상관분석 결과와 논문 수, 저자 수, 이용 수 지표 대신 시차를 적용한 altmetrics 지표를 활용한 상관분석 결과

를 비교하고자 한다.

4.5.1 계량서지 지표와 미래 유망연구주제 후보군과의 상관관계 분석

우선 계량서지 지표를 사용하여 미래 유망연구주제를 도출하기 위한 상관관계 분석을 실시하였다. 이는 미래 유망연구주제 후보군 중 선행 유망연구주제의 공통적 추세와 유사한 패턴을 보이는 영역이 향후 유망한 연구주제가 될 가능성이 클 것이라는 것을 전제로 한다. 따라서 미래 유망연구주제 후보군 중에서 선행 유망연구주제와 유사한 성장 패턴을 보이는 영역을 도출하고자 하였으며, 앞서 추출한 선행 유망연구주제의 공통적 특징과 미래 유망연구주제 후보군과의 유사성을 파악하기 위하여 상관관계 분석을 실시하였다. 앞서 계량서지 지표 분석을 통해 도출한 선행 유망연구주제의 5가지 특징 모두에서 유의수준 $p<0.01$ 에서 양(+)의 상관성을 보이는 영역을 미래 유망연구주제로 선정하였다. 선행 유망연구주제의 5가지 지표별 공통적 추세와 미래 유망연구주제 후보군과의 상관관계 분석 결과는 다음 〈표 18〉과 같다.

분석 결과, 총 14개의 토픽이 선행 유망연구주제와 5가지 지표 모두에서 통계적으로 유의한 양의 상관관계가 나타났다. 이에 따라 선행 유망연구주제의 공통적 추세를 따라 성장 가능

〈표 18〉 계량서지 지표와 미래 유망연구주제 후보군과의 상관관계 분석 결과

번호	논문 수	저자 수	이용 수	피인용 수	신규 키워드 수
1	.939**	.961**	.687**	.296	.321
2	.915**	.951**	.457*	.013	.156
3	.924**	.950**	.416*	-.210	.029
4	.821**	.840**	.942**	.790**	.603**
5	.953**	.958**	.975**	.838**	.687**
6	.798**	.817**	.854**	.641**	.695**
7	.967**	.960**	.914**	.761**	.644**
8	.975**	.967**	.919**	.250	.021
9	.944**	.929**	.960**	.849**	.674**
10	.919**	.958**	.632**	.298	-.683**
11	.971**	.971**	.691**	.492*	-.788**
12	.456*	.111	-.118	-.708**	-.491*
13	.941**	.939**	.419*	.032	.722**
14	.850**	.888**	.454*	-.391	.641**
15	.792**	.853**	.317	-.463*	.719**
16	.930**	.953**	.586**	-.049	.727**
17	.793**	.837**	.735**	.381	.553**
18	.887**	.891**	.858**	.402*	.670**
19	.767**	.815**	.675**	.225	.549**
20	.794**	.813**	.713**	.301	.447*
21	.865**	.897**	.769**	.517**	.632**
22	.883**	.883**	.643**	.666**	.652**
23	.838**	.887**	.732**	.499*	.613**
24	.792**	.839**	.306	.158	-.085
25	.854**	.851**	.330	-.031	-.374
26	.976**	.982**	.903**	.643**	.826**
27	.971**	.978**	.947**	.684**	.514**
28	.946**	.962**	.910**	.803**	.741**
29	.921**	.934**	.896**	.820**	.788**
30	.950**	.956**	.856**	.422*	.524**
31	.965**	.977**	.877**	.556**	.615**
32	.973**	.978**	.914**	.681**	.726**
33	.967**	.970**	.873**	.639**	.702**
34	.913**	.916**	.743**	-.030	.367
35	.833**	.898**	.416*	-.109	.236
36	.933**	.925**	.643**	-.099	.264
37	.884**	.915**	.599**	-.018	.249
38	.890**	.902**	.370	-.080	.149
39	.943**	.948**	.600**	.001	.033
40	.925**	.942**	.607**	.079	.164
41	.901**	.920**	.629**	.141	.293

** p<0.01, * p<0.05

성이 있는, 즉, 향후 유망 주제로 부상할 가능성이 큰 영역은 cryptocurrency, stock market, volatility spillover(이상 회계학), transnational sustainability, sustainable environment(이상 사회학), climate change policy, immigrant children education(이상 행정학/정책학), entrepreneurship education, sustainable development(이상 국제/지역개발), emotional tourism, tourism share economy(이상 관광학), product image, advertise algorithm, gender in advertising(이상 신문방송학)인 것으로 판단

할 수 있다.

4.5.2 선행지표를 활용한 상관관계 분석

다음으로 계량서지 지표 중 논문 수, 저자 수, 인용 수에 선행하는 altmetrics 지표를 활용하여 상관관계 분석을 실시하였다. 앞서 교차상관 분석에서 확인한 바와 같이 capture 수의 경우 논문 수와 저자 수에 비해 2년 선행하고, 인용 수에 비해 3년을 선행하므로, 해당 시차를 적용하여 상관관계 분석을 실시하였으며, 그 결과는 다음 <표 19>와 같다.

<표 19> 선행 지표(altmetrics)를 활용한 상관관계 분석 결과

번호	(-2)captures	(-2)captures	(-3)captures	피인용 수	신규 키워드 수
1	.947**	.951**	.714**	.296	.321
2	.930**	.948**	.446*	.013	.156
3	.949**	.960**	.405	-.210	.029
4	.763**	.755**	.900**	.790**	.603**
5	.926**	.909**	.957**	.838**	.687**
6	.722**	.712**	.804**	.641**	.695**
7	.944**	.904**	.932**	.761**	.644**
8	.961**	.919**	.930**	.250	.021
9	.904**	.857**	.945**	.849**	.674**
10	.927**	.948**	.646**	.298	-.683**
11	.973**	.971**	.710**	.492*	-.788**
12	.455*	.077	-.176	-.708**	-.491*
13	.949**	.943**	.419	.032	.722**
14	.842**	.876**	.442*	-.391	.641**
15	.812**	.873**	.317	-.463*	.719**
16	.945**	.960**	.601**	-.049	.727**
17	.836**	.873**	.758**	.381	.553**
18	.878**	.859**	.850**	.402*	.670**
19	.820**	.868**	.697**	.225	.549**
20	.847**	.869**	.740**	.301	.447*
21	.886**	.898**	.784**	.517**	.632**
22	.919**	.919**	.678**	.666**	.652**
23	.876**	.914**	.758**	.499*	.613**
24	.833**	.883**	.293	.158	-.085
25	.896**	.895**	.305	-.031	-.374

번호	(-2)captures	(-2)captures	(-3)captures	피인용 수	신규 키워드 수
26	.969**	.950**	.925**	.643**	.826**
27	.962**	.941**	.967**	.684**	.514**
28	.953**	.948**	.921**	.803**	.741**
29	.941**	.939**	.919**	.820**	.788**
30	.961**	.965**	.895**	.422*	.524**
31	.980**	.980**	.916**	.556**	.615**
32	.977**	.969**	.945**	.681**	.726**
33	.972**	.957**	.908**	.639**	.702**
34	.912**	.914**	.762**	-.030	.367
35	.847**	.884**	.390	-.109	.236
36	.915**	.903**	.647**	-.099	.264
37	.895**	.919**	.616**	-.018	.249
38	.865**	.845**	.339	-.080	.149
39	.938**	.938**	.616**	.001	.033
40	.919**	.913**	.616**	.079	.164
41	.890**	.883**	.629**	.141	.293

** $p < 0.01$, * $p < 0.05$

분석 결과, 선행지표를 활용한 상관관계 분석 결과에서도 계량서지 지표를 활용하여 상관관계 분석을 수행한 결과와 동일한 결과가 도출되었으며, 특히 시차를 적용한 capture 수를 적용하여 도출된 지표들의 상관계수를 확인하였을 때 계량서지 지표를 활용한 분석 결과와 대동소이한 모습을 볼 수 있어, 향후 유망연구주제의 탐지에 있어 선행성을 높이는 데 해당 지표를 활용하는 것이 효과적이라고 할 수 있다.

5. 결 론

학문 분야에서 발전 가능성이 있는 연구주제를 탐지하고 발굴하는 것은 잠재력 있는 연구의 선별, 발전 가능성 있는 연구주제에 대한 경쟁력 확보, 학술적으로 우수한 연구의 조기 진행 등 연구자 개인 연구의 방향 설정에 유용하

다. 아울러 국내외 연구 상황에 대한 객관적 파악 및 미래의 성과 예측을 통한 방향 결정 등에 활용될 수 있어 기관 및 국가적 차원에서도 유망연구주제의 탐지와 발굴은 중요하다. 특히, 정보기술의 발달과 학술정보지식의 증가, 학문 분야의 학제적 성향의 강화 등으로 인해 새로운 연구주제의 동태(動態)와 발전을 파악하는 것이 매우 중요해진 만큼 현재의 현황 파악 중심의 연구 동향 분석에서 나아가 새로이 부상하는 연구주제 발견의 필요성이 더욱 크다고 하겠다.

본 연구는 이러한 문제의식 아래 사회과학 분야를 대상으로 향후 유망연구주제로 성장할 가능성이 있는 연구주제를 식별하고자 하였으며, 아울러 유망연구주제의 조기 탐지를 위한 altmetrics의 활용 가능성을 가늠해 보았다. 이를 위해 2013년부터 2022년까지의 사회과학 분야 저널 151종에 게재된 논문 39,165건을 대상

으로 토픽 모델링을 수행하여 사회과학 분야의 주요 연구주제를 파악하고, 선행 유망연구주제와 주제별 핵심 키워드 및 미래 유망연구주제 후보군 선정을 위한 신규 키워드를 도출하였다. 이어서 선행 유망연구주제의 공통적 추세를 추출하기 위한 주성분분석과 선행지표로서 altmetrics 지표의 유용성을 검증하기 위한 교차상관 분석 및 회귀분석을 수행하였으며, 미래 유망연구주제를 선정하기 위한 상관관계 분석을 수행하였다.

분석 결과, altmetrics 지표 중 capture 수가 논문 수, 저자 수에 대하여는 2년을, 인용 수에 대하여는 3년을 선행하는 것으로 나타났으며, 도출된 유망연구주제들은 계량서지 지표를 활용한 경우와 선행지표를 적용한 경우가 동일하게 나타났다. 따라서 altmetrics 지표 중 capture 수는 논문 수, 저자 수, 인용 수의 선행지표로서 활용가능할 것으로 판단되며, 이 경우, 유망연구주제를 2~3년 정도 빠르게 파악하는 것이 가능하다. 따라서 유망연구주제의 조기 탐지에 있어 선행성을 높이기 위해서는 해당 지표의 사용을 고려할 수 있을 것이다.

이 연구는 사회과학 분야 전체를 대상으로 하였으므로 거시적 측면에서 사회과학 분야의 연구 영역을 파악할 수 있었던 반면, 각 분야를 자세히 들여다보기에는 어려움이 있었다. 특히 이는 토픽모델링 분석을 통해 연구주제를 도출하는 과정에서도 한계로 작용하였는데, 일부 주제의 경우 토픽의 핵심 키워드들이 매우 일반적으로 사용되는 단어들이거나 거시적 관점의 키워드가 도출되어 토픽의 특징과 주제를 파악하기 위해 원문을 참고하여 토픽이 의미하는 바를 분석해야 하는 경우가 다수 발생하였다. 뿐만 아니라 본 연구에서는 토픽모델링 결

과로 도출된 핵심 키워드들을 각 연구주제 해당 문헌의 검색을 위한 검색어로 활용하였는데, 해당 키워드가 일반적인 단어이거나 거시적 관점으로 사용되는 키워드인 경우, 각 토픽의 특징을 잘 나타내는 키워드를 활용했을 때보다 검색 결과의 정확도가 떨어질 수 있어 검색 결과의 적합성 측면에서 학문 분야마다 차이가 발생할 수 있다는 점 역시 한계이다.

또한 본 연구에서 미래 유망연구주제 후보군으로 선정한 토픽들은 현재 시점에서 신규 키워드들이 포함되어 있는 토픽들로, 이들 토픽이 다른 토픽으로 변화하거나 완전히 다른 새로운 토픽을 형성할 가능성을 배제할 수 없으며, 코로나19와 같이 급격한 사회 변화로 인하여 추세적 특징 없이 특정 연구가 단기간에 폭발적으로 증가하는 경우에는 본 연구에서 제안한 방법을 적용하기 어려우므로 학문 분야 및 연구주제의 역동성과 변화에 대한 완전한 대처는 어렵다는 것 역시 본 연구의 한계라 할 수 있다.

아울러 본 연구에서 선행지표로 활용했던 altmetrics 지표의 경우, altmetrics의 짧은 연혁 뿐만 아니라 altmetrics의 주요 정보원으로 활용되는 웹 상의 각종 매체 및 소셜 미디어의 조작 가능성이나 학문 분야별 및 연구자별 이용 경향의 차이 등으로 인해 데이터에 편향이 존재할 수 있으므로 altmetrics에 대한 관심의 정도에 따라 특정 학문 분야나 연구주제의 경우 측정 값이 왜곡될 가능성도 존재한다. 이는 본 연구에서 활용한 altmetrics 지표 중 social media 지표에서도 확인할 수 있었던 바, 따라서 유망연구주제의 탐지를 위해 다양한 측면을 포괄할 수 있는 지표를 고려하는 동시에 기존의

계량서지 지표를 보완하는 측면에서 altmetrics 지표의 활용을 고려하여야 하며, 신뢰성과 유효성의 측면에서 altmetrics에 대한 지속적인 검증이 이루어져야 할 필요가 있을 것이다.

그러나 본 연구는 기존의 연구 동향 분석에서 활용되던 인용 및 키워드 기반의 계량서지 지표를 활용하여 새로운 연구주제의 출현을 예측함으로써 발생하는 시간적 한계를 극복하고 즉시성을 담보할 수 있는 지표로서 altmetrics의 유용성을 확인하였으며, 이를 통해 유망연구주제의 조기 탐지를 위한 altmetrics 지표의

잠재적 활용 가능성을 확인하였다는 데 의의가 있다. 또한 본 연구에서는 미래 유망연구주제 도출하기 위해 주성분분석을 활용하였는데, 최근 머신러닝을 활용한 다양한 예측모델을 구축할 때 차원 축소 기법으로 많이 사용하는 주성분분석을 활용하여 사회과학 분야의 미래 유망연구주제를 발굴하였다는 점에서 연구 동향 파악 및 유망연구주제 예측을 위한 방법론적 확장 가능성과 향후 보다 정교한 연구주제 탐지 모델의 설계에 활용될 가능성을 제공한다는 점에서도 의의를 찾을 수 있을 것이다.

참 고 문 헌

- 김현정, 조남옥, 신경식 (2015). 항공산업 미래유망분야 선정을 위한 텍스트 마이닝 기반의 트렌드 분석. *지능정보연구*, 21(1), 65-82. <https://doi.org/10.13088/jiis.2015.21.1.65>
- 안세정, 심위, 이준영, 권오진, 노경란 (2012). 과학계량학 기법을 이용한디스플레이 연구영역의 트렌드 탐지. *한국전자통신학회논문지*, 7(6), 1343-1351. <https://doi.org/10.13067/JKIECS.2012.7.6.1343>
- Alattar, F. & Shaalan, K. (2021). Emerging research topic detection using filtered-LDA. *AI*, 2(4), 578-599. <https://doi.org/10.3390/ai2040035>
- An, X. Y. & Wu, Q. Q. (2011). Co-word analysis of the trends in stem cells field based on subject heading weighting. *Scientometrics*, 88(1), 133-144. <https://doi.org/10.1007/s11192-011-0374-1>
- Arun, R., Suresh, V., Veni Madhavan, C. E., & Narasimha Murthy, M. N. (2010). On finding the natural number of topics with latent dirichlet allocation: Some observations. *Advances in Knowledge Discovery and Data Mining*, 6118, 391-402. https://doi.org/10.1007/978-3-642-13657-3_43
- Ball, R. & Tunger, D. (2006). Bibliometric analysis-a new business area for information professionals in libraries? support for scientific research by perception and trend analysis. *Scientometrics*, 66(3), 561-577. <https://doi.org/10.1007/s11192-006-0041-0>

- Boyack, K. W. & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately?. *Journal of the American Society for information Science and Technology*, 61(12), 2389-2404.
<https://doi.org/10.1002/asi.21419>
- Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72(7-9), 1775-1781.
<https://doi.org/10.1016/j.neucom.2008.06.011>
- Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document Numérique*, 17(1), 61-84.
<https://doi.org/10.3166/dn.17.1.61-84>
- Glänzel, W. & Thijs, B. (2012). Using 'core documents' for detecting and labelling new emerging topics. *Scientometrics*, 91(2), 399-416. <https://doi.org/10.1007/s11192-011-0591-7>
- Griffiths, T. L. & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl_1), 5228-5235. <https://doi.org/10.1073/pnas.0307752101>
- Guo, H., Weingart, S., & Börner, K. (2011). Mixed-indicators model for identifying emerging research areas. *Scientometrics*, 89(1), 421-435. <https://doi.org/10.1007/s11192-011-0433-7>
- Iwami, S., Mori, J., Sakata, I., & Kajikawa, Y. (2014). Detection method of emerging leading papers using time transition. *Scientometrics*, 101, 1515-1533.
<https://doi.org/10.1007/s11192-014-1380-x>
- Lee, W. (2008). How to identify emerging research fields using scientometrics: an example in the field of Information Security. *Scientometrics*, 76(3), 503-525.
<https://doi.org/10.1007/s11192-007-1898-2>
- Liang, Z., Mao, J., Lu, K., Ba, Z., & Li, G. (2021). Combining deep neural network and bibliometric indicator for emerging research topic prediction. *Information Processing & Management*, 58(5), 102611. <https://doi.org/10.1016/j.ipm.2021.102611>
- Ohniwa, R., Hibino, A., & Takeyasu, K. (2010). Trends in research foci in life science fields over the last 30 years monitored by emerging topics. *Scientometrics*, 85(1), 111-127.
<https://doi.org/10.1007/s11192-010-0252-2>
- Small, H. (2006). Tracking and predicting growth areas in science. *Scientometrics*, 68(3), 595-610.
<https://doi.org/10.1007/s11192-006-0132-y>
- Takeda, Y. & Kajikawa, Y. (2009). Optics: a bibliometric approach to detect emerging research domains and intellectual bases. *Scientometrics*, 78(3), 543-558.
<https://doi.org/10.1007/s11192-007-2012-5>

- Weismayer, C. & Pezenka, I. (2017). Identifying emerging research fields: a longitudinal latent semantic keyword analysis. *Scientometrics*, 113(3), 1757-1785.
<https://doi.org/10.1007/s11192-017-2555-z>
- Xu, S., Hao, L., An, X., Yang, G., & Wang, F. (2019). Emerging research topics detection with multiple machine learning models. *Journal of Informetrics*, 13(4), 100983.
<https://doi.org/10.1016/j.joi.2019.100983>
- Yu, H., Chen, Q., & Pan, Y. (2022). A bibliometric and emerging trend analysis on stress granules from 2011 to 2020: a systematic review and bibliometrics analysis. *Medicine*, 101(29), e29200.
<https://doi.org/10.1097/MD.00000000000029200>

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- Ahn, Sejung, Shim, We, Lee, June Young, Kwon, Oh-jin, & Noh, Kyungran (2012). Trends detection of display research areas by bibliometric analysis. *The Journal of the Korea Institute of Electronic Communication Sciences*, 7(6), 1343-1351.
<https://doi.org/10.13067/JKIECS.2012.7.6.1343>
- Kim, Hyun-jung, Jo, Nam-ok, & Shin, Kyung-shik (2015). Text mining-based emerging trend analysis for the aviation industry. *Journal of Intelligence and Information Systems*, 21(1), 65-82. <https://doi.org/10.13088/jiis.2015.21.1.65>