

문헌정보학분야 해외 연구 동향 및 유망 주제 분석 연구*

Research on Overseas Trends and Emerging Topics in Field of Library and Information Science

구 본진 (Bon Jin Koo)**
장덕현 (Durk Hyun Chang)***

목 차

- | | |
|---------------------------|----------|
| 1. 연구의 배경 및 필요성 | 4. 분석 결과 |
| 2. 유망 연구 주제의 개념 및 선행연구 고찰 | 5. 결론 |
| 3. 연구의 방법 | |

초 록

이 연구는 문헌정보학 분야의 연구 동향 분석을 통해 문헌정보학의 핵심 연구 영역을 파악하고 향후 유망 연구 주제로 부상할 가능성이 있는 주제를 식별하고자 하였다. 이를 위해 문헌정보학 분야의 국외 학술지 5종을 대상으로 지난 30년간 (1993~2022)의 학술논문 11,252건에서 40,897개의 저자 키워드를 수집하였으며, 저자 키워드를 활용한 키워드 분석을 통해 문헌정보학 분야의 핵심 연구 영역을 파악하였다. 이어서 논문수, 저자수, 공저논문 비율, 피인용 수를 활용하여 주성분분석과 상관관계분석을 통해 문헌정보학 분야의 미래 유망 연구 주제를 도출하였다. 분석 결과, 향후 문헌정보학 분야의 유망 연구 주제는 '머신러닝/알고리즘'과 '연구 영향력'이었으며, 이외에도 소셜미디어와 빅데이터분석, 자연어 처리, 연구 트렌드 분석, 연구성과 평가 등이 향후 주요한 연구주제로 성장할 가능성이 있는 것으로 나타났다.

ABSTRACT

This study aimed to investigate key research areas in the field of Library and Information Science (LIS) by analyzing trends and identifying emerging topics. To facilitate the research, a collection of 40,897 author keywords from 11,252 papers published in the past 30 years (1993-2022) in five journals was gathered. In addition, keyword analysis, as well as Principal Component Analysis (PCA) and correlation analysis were conducted, utilizing variables such as the number of articles, number of authors, ratio of co-authored papers, and cited counts. The findings of the study suggest that two topics are likely to develop as promising research areas in LIS in the future: machine learning/algorithm and research impact. Furthermore, it is anticipated that future research will focus on topics such as social media and big data, natural language processing, research trends, and research assessment, as they are expected to emerge as prominent areas of study.

키워드: 문헌정보학, 연구동향, 유망 주제, 키워드분석, 주성분분석

Library and Information Science, Research Trends, Emerging Topics, Keyword Analysis, Principal Component Analysis(PCA)

* 이 논문은 2020년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임
(NRF-2020S1A5B5A17088401).

** 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: 2023년 7월 18일 최초심사일자: 2023년 7월 31일 게재확정일자: 2023년 8월 12일
한국문헌정보학회지, 57(3): 71-96, 2023. <http://dx.doi.org/10.4275/KSLIS.2023.57.3.071>

※ Copyright © 2023 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

특정 분야의 연구 동향을 분석하는 것은 그 동안의 연구의 흐름을 파악하고, 연구 주제의 시대적 변천 과정 및 핵심 영역을 확인할 수 있어 해당 학문 분야의 변화와 발전을 위한 기반이 될 수 있으며, 아울러 부상하고 있는 새로운 연구 주제의 발견과 예측을 가능하게 한다는 점에서 중요하다. 과거에는 연구 동향 분석이 주로 학문 영역 내 혹은 학문 영역 간 관계와 구조를 이해하기 위한 목적 또는 해당 기관이나 연구자의 연구 실적을 분석하여 생산성과 영향력을 측정하기 위한 목적으로 수행되었으나, 최근 들어 연구 동향 분석을 통해 새롭게 등장하는 연구 영역을 식별하고 선점하거나 학문 분야의 향후 발전 방향을 예측하고자 하는 노력들이 이루어지고 있다.

특히, 정보기술의 발달과 과학기술지식의 양적 증가, 학문의 학제적 성향 강화 등으로 인해 새로운 연구 분야가 빠르게 출현하고, 다양한 연구 분야에서 분야 간 통합이나 세분화가 이루어지는 등 많은 연구 분야들이 복잡하게 연결되어 있다. 이로 인해 새로운 연구 주제의 발견이나 연구 분야 간 연관성을 파악하는데 어려움을 겪게 되면서 새로운 연구 주제의 동태(動態)와 발전을 포착하는 것이 연구 동향 분석의 또 다른 주요한 목적이 되었다.

또한 빅데이터 시대에 들어서면서 공공 및 민간 영역에서 생산되는 방대한 데이터를 통해 현재의 주요 이슈를 탐지하고 모니터링하며, 미래를 예측하기 위한 연구가 활발히 진행되고 있다. 특히, 코로나19의 확산으로 인해 질병의

확산 양상을 예측하고, 감염 양상과 변화를 추측하는 등 바이러스 확산 추이를 초기에 탐지하고 진단하려는 시도가 확산되면서 ‘예측’의 중요성이 강조되었으며, ‘예측’을 위한 연구가 다시 확산되는 움직임도 보이고 있다.

학문 분야에서 ‘예측’이라는 것은 새롭게 등장하는 연구 주제나 기존의 연구 주제들 중 새롭게 부상하는 연구 주제를 발굴하고 탐지하는 것으로 이해할 수 있다. 이는 연구 동향 분석이 기존에 이루어져왔던 해당 학문 분야 내 연구 주제의 현황 파악 중심에서 시간이 지남에 따라 이들 연구 주제가 어떻게 변화하고 발전해 가는지 분석하고 아울러 특정 연구 영역이 향후 어떻게 발전할 것인가를 파악함으로써 미래의 연구 경향을 예측하려는 방향으로 나아가야 함을 의미한다. 즉, 학문 분야에서 발전 가능성 있는 연구 주제를 탐지하고 발굴하는 것은 학술적으로 우수한 연구를 조기에 진행할 수 있으며, 발전 잠재력이 있는 연구를 선별하고, 미래의 성과를 예측하는 등 학문 분야의 발전에 방향성을 제시할 수 있다. 따라서 현황 파악 중심의 연구 동향 분석에서 나아가 새로이 부상하는 연구 주제 발견의 필요성이 있다고 볼 수 있다.

이에 따라 본 연구는 문헌정보학 분야를 대상으로 연구의 동향을 파악하여, 문헌정보학 연구가 어떻게 변화, 발전해 왔는지를 밝히고 문헌정보학 분야의 새로운 연구 주제를 도출함으로써 향후 문헌정보학 연구가 지향해야 할 방향과 시사점을 제안하고자 하였다.

2. 유망 연구 주제의 개념 및 선행연구 고찰

2.1 유망 연구 주제의 개념

유망 연구 주제는 대부분 특정 학문 분야의 연구 동향 분석 결과의 일부로 제시되어 있어 실질적으로 유망 연구 주제를 정의하고 있는 연구는 그리 많지 않다. 그러나 일부 연구에서 제시하고 있는 정의를 살펴보면 다음과 같다. 유망 연구 주제는 주로 'Emerging trend', 'Emerging topic', 'Emerging research topic', '유망연구영역' 등 연구자마다 각기 다른 용어를 사용하고 있는데, Kontostathis et al.(2004)은 유망 추세(emerging trend)를 시간이 경과함에 따라 관심과 유용성이 높아지는 연구 주제로 정의하였으며, 안세정 외(2012)는 유망 연구영역을 시간이 경과됨에 따라 관심과 유용성이 높아지고 주목받는 영역이라고 정의하였다. 또한 Wang(2018)의 연구에서는 유망 연구 주제(Emerging research topic)를 어느 정도의 지속성과 상당한 과학적 영향력의 특징을 가지고 있으며, 급진적인 신규성과 상대적으로 빠른 성장을 보이는 연구 주제로 정의하였다.

이상의 정의를 토대로 유망 연구 주제는 이전에는 존재하지 않았던, 새롭게 출현한 연구 주제, 그리고 기존에 이미 존재하고 있으나 주목받지 못하다가 최근에 새롭게 주목을 받고 있는 연구 주제를 모두 포함하는 개념으로 정의할 수 있다.

2.2 선행연구

그동안 다양한 학문 분야를 대상으로 해당

분야의 연구 동향을 분석하고 유망 영역을 파악하려는 시도는 꾸준히 이루어져 왔다. 그 중 연구 동향 분석과 함께 유망 연구 영역을 도출한 연구들을 살펴보면 다음과 같다.

Mane & Börner(2004)는 연구 영역의 구조와 진화 분석에 버스트 텀지 알고리즘을 적용하였으며, PNAS에 발표된 전체 논문을 대상으로 하여 주요 주제를 도출하고, 연구 주제의 진화와 패러다임의 변화를 확인하였다. 또한 Ball & Tunger(2006)는 논문 수와 인용 수를 분석하여 특정 연구 분야의 현황과 미래 트렌드를 예측하였으며, 특정 주제 영역의 추세를 분석하기 위해 계량서지학적 지표를 활용할 수 있음을 주장하였다.

정의섭 외(2008)는 건설교통 기술 관련 분야를 대상으로 과학계량학 분석을 수행하였으며, 구체적으로 고في인용 논문을 대상으로 국가별 동향 분석을, 상관관계 분석을 통해 분야 간 용·복합 관계를 확인하였고, 서지결합분석을 활용한 클러스터링 분석을 통해 유망연구영역군을 도출하였으며, Takeda & Kajikawa(2009)는 광학 분야, 특히 OLED와 관련된 학술논문을 대상으로 인용 네트워크의 구조를 분석하고, 이를 위상적 클러스터링을 군집화하여 연구 영역의 구조와 유망 연구 영역을 탐지하였다.

이재윤 외(2011)는 LED 분야를 대상으로 논문 수와 피인용 빈도를 분석하여 전략 다이어그램을 구성, 유망 영역을 제시하였으며, 연구 성장 추세를 반영할 수 있는 표준 성과 지수(NPI)와 표준 성장 지수(NGI)를 제안하였고, 안세정 외(2012)는 디스플레이와 관련된 영역을 대상으로 논문 및 특허 데이터와 인용 데이터를 분석하여 해당 영역의 발전과정과 향후

발전 가능성이 있는 영역을 제안하였다.

Hu et al.(2013)은 중국 문헌정보학의 지식 구조를 분석하기 위해 동시단어분석을 수행하였으며, 이를 2차원의 전략 다이어그램과 네트워크 지도로 표현하여 중국의 문헌정보학 지식 구조의 전체적인 특징을 파악하고, 주요 핵심 영역과 잠재력이 있는 유망 주제를 도출하였다. 또한 고재창, 조근태, 조윤호(2013)는 기술경영 분야를 대상으로 키워드 네트워크 분석을 수행하여 컴포넌트의 크기를 통해 새로운 연구 주제의 판단이 가능하다고 하였으며, 기존의 키워드의 연결정도와 신규 키워드와의 연결 정도 분석을 통해 새로운 연구주제의 생성을 확인할 수 있다고 하였다.

김현정, 조남옥, 신경식(2015)은 항공산업 관련 논문을 대상으로 토피모델링을 적용하여 항공 분야의 연구 동향을 파악하고 미래 유망 분야를 전망하였고, Weismayer & Pezenka(2017)는 마케팅 및 관광 분야의 새로운 연구 영역을 찾기 위해 관련 논문의 저자 키워드에 대한 텍스트마이닝 분석을 실시하였으며, 마케팅 및 관광 분야의 관련 연구 분야를 식별하고 향후 추세를 예측하였다. 또한 Chen & Pan(2022)은 stress granules 관련 연구를 대상으로 동시 인용분석, 키워드동시출현분석, 버스트분석을 통해 향후 유망 연구 영역을 예측하였다.

이처럼 학문 분야의 동향을 분석하고, 유망 연구 주제를 도출하기 위해 수행된 연구들을 보면 주로 논문 수와 피인용 수 분석, 인용 패턴 분석, 키워드 분석 등을 활용하고 있음을 알 수 있다. 특히 인용 패턴이나 키워드 분석의 경우 이를 시각화할 수 있는 네트워크 분석, 클러스터링, 텍스트마이닝 기법을 상호보완적으로 활용하는 경우가 많으며, 최근 들어서는 2개 이상의 지표와 방법을 결합하여 복합적으로 활용하고 있는 것을 확인할 수 있는데, 이는 유망 연구 주제를 도출하는 데 있어 방법론적 측면에서 단일의 분석 지표나 방법의 단점을 보완하기 위한 노력이라 볼 수 있다. 따라서 본 연구에서는 유망 연구 주제 분석을 위해 고려할 수 있는 또 다른 방법으로 주성분분석을 활용하고자 하였다.

용하는 경우가 많으며, 최근 들어서는 2개 이상의 지표와 방법을 결합하여 복합적으로 활용하고 있는 것을 확인할 수 있는데, 이는 유망 연구 주제를 도출하는 데 있어 방법론적 측면에서 단일의 분석 지표나 방법의 단점을 보완하기 위한 노력이라 볼 수 있다. 따라서 본 연구에서는 유망 연구 주제 분석을 위해 고려할 수 있는 또 다른 방법으로 주성분분석을 활용하고자 하였다.

3. 연구의 방법

3.1 연구 내용 및 방법

본 연구는 문헌정보학 분야의 핵심 연구 영역을 확인하고 향후 문헌정보학 분야의 유망 연구 주제는 무엇인지 식별하기 위한 연구로, 1993년부터 2022년까지 30년간 문헌정보학 내에 어떤 연구 주제들이 있어 왔으며, 시기별로 어떻게 변화해 왔는지, 그리고 그러한 패턴 속에서 향후 문헌정보학 분야의 유망 연구 주제는 무엇인지 파악하기 위한 시사점을 얻고자 하는데 목적이 있다.

이를 위해 문헌정보학 분야의 연구 논문을 수집하여 주요 현황을 파악한 후, 키워드 분석을 통해 문헌정보학 분야의 연구 영역 도출 및 이를 통한 연구 동향을 파악하고, 기존의 연구 주제들에서 확인할 수 있는 주요 특징을 근거로 미래 유망 연구 주제를 파악하고자 하였다. 구체적으로 연구 영역 도출 및 동향 분석을 위해 동시출현단어 분석을 수행하며, 이를 통해 문헌정보학 분야의 선형 유망 연구 주제들을

선정하였다. 다음으로 주성분 분석과 상관관계 분석을 수행하여 문헌정보학 분야의 미래 유망 연구 주제를 파악하고자 하였다. 앞서 선정한 선행 유망 연구 주제들의 특성을 주성분 분석을 통해 단일의 차원으로 축소하고 이를 미래 유망 연구 주제 그룹에 적용하여 향후 유망 연구로 부상할 가능성이 있는 연구 주제를 도출하고자 하였다.

동시출현단어분석은 특정 분야의 텍스트 데이터에서 주요 키워드를 추출하고 키워드 간의 동시 출현 빈도를 근거로 이들 간 연결관계를 살펴보는 분석 기법이다(이수상, 2014). 국내 외의 다양한 연구들에서 연구 동향을 분석하기 위한 대표적인 방법으로 많이 활용되고 있으며, 어떤 연구 주제가 어떤 클러스터를 형성하고 있는지, 해당 클러스터에서 가장 영향력 있는 주제는 무엇인지 등을 파악하는데 효과적이다. 따라서 시기별 연구 분야의 성장 패턴과 특정 키워드가 어떤 클러스터에 포함되어 있는지 파악하고, 이를 바탕으로 향후 유망 연구 영역을 제안하기 위하여 동시출현단어 분석을 수행하였다.

주성분분석(principal component analysis, PCA)은 대표적인 차원 축소 기법 중 하나로 원 데이터의 분포를 최대한 보존하면서 고차원 공간의 데이터를 저차원 공간으로 변환한다(Jolliffe, 2002). 즉, 주성분분석은 기준의 변수를 조합하여 새로운 변수를 만들어내는 기법이다. 김선호, 이문환, 배덕호(2016)는 3개의 자료를 활용하여 가뭄 해석을 위한 자연가뭄지수(NDI)를 산정하기 위해 주성분분석을 활용하였으며, 민인식(2018)은 주성분분석을 통해 KOSIS(국가통계포털)에서 제공하는 100대 통계지표 중

56개 지표를 이용하여 우리나라를 대표하는 하나의 지표를 생성하는 예를 제시한 바 있다. 이외에도 김성호, 김지표(2020)는 다양한 특허지표를 활용하여 주성분분석을 통해 특허 경쟁력 지수를 산출하고, 기업 간 비교에 적용한 바 있다. 국외의 경우, Radovanović, Filipović, & Golušin (2018)은 유럽 28개국의 다양한 에너지 안보 및 경제 지표 9개를 주성분 분석을 통해 단일 차원으로 축소, 에너지 안보의 지리경제적 지표를 생성하고 해당 지표값을 종속변수로 활용한 바 있으며, Dai, Xiong, & Zhou(2021)는 24개 국가의 경제정책불활성 지수를 주성분 분석을 통해 하나의 단일 차원으로 축소하고 글로벌 금융시장의 변동성과의 상관관계를 분석하였다. 따라서 이들 연구를 근거로 본 연구에서는 선행 유망 연구 주제의 추세를 생성하기 위해 주성분분석을 활용하였다.

본 연구에서 주성분분석에 사용할 지표는 논문 수, 저자 수, 공저논문 비율, 피인용 수의 네 가지이다. 선행연구 분석에서 알 수 있는 것처럼 연구 동향 분석 또는 유망 연구 주제의 탐지에 활용하는 지표들은 논문 수, 인용 수, 인용 패턴 분석, 키워드 분석이 대표적이라고 할 수 있다. 그러나 인용 패턴이나 키워드의 경우 대부분 네트워크 등의 시각화 기법이 함께 사용되고, 키워드 수의 경우 저자마다 제출하는 키워드의 수에 편차가 있으므로 단일 지표로 활용하기에는 적절하지 않다고 판단하였다. 따라서 본 연구에서는 연구 동향 분석 및 유망 연구 주제 탐지에 활용되는 대표적 단일 지표로서 논문 수, 인용 수와 함께 저자 수와 공저논문 비율을 활용하고자 한다.

논문 수의 경우, 학술 연구의 활동이 얼마나

활발하게 이루어지고 있는가를 나타내는 대표적인 지표라고 할 수 있으며, 인용 수는 해당 문헌의 우수성과 영향력을 나타내므로 해당 문헌이 다른 연구 주제나 연구 분야에 미칠 수 있는 영향력으로 이해할 수 있다. 다음으로 저자 수의 경우, 연구 주제가 성장하는 단계에서는 새로운 저자의 유입 속도가 빠르게 나타나는 측면(Guo, Weingart, & Börner, 2011)을 고려하여 선정하였고, 공저논문의 비율은 문헌정보학 분야에서 증가하는 공동 연구의 영향력 측면(양기덕, 김선욱, 이혜경, 2021)을 고려하여 선정하였다.

본 연구의 전반적 프레임워크는 다음 <그림 1>과 같으며, 동시출현단어분석 및 시각화는 VOSviewer 1.6.19를, 주성분분석 및 상관관계 분석은 Stata 18.0을 활용하였다.

3.2 데이터 수집

분석 대상 학술지의 선정은 JCR의 IF를 기준으로 하였으며, 선정 학술지에 게재된 논문의 키워드는 Web of Science(WoS) 데이터베이스를 통해 수집하였다. 2021년 기준 JCR의 Information Science & Library Science 범주에 속한 문헌정보학 분야 저널은 총 84종이다. 그러나 JCR의 경우, 복수의 범주를 배정받을 수 있으므로 문헌정보학 분야와의 관련성을 고려하기 위하여 Walters와 Wilder(2016)의 연구에서 선정한 문헌정보학 분야 핵심 학술지 31종을 대상으로 IF 상위 5종의 학술지를 선정하였다.

Walters와 Wilder(2016)의 연구에서 도출한 31종의 학술지 중 26종이 JCR의 Information Science & Library Science 범주에 포함되어 있었으며, 이 중 본 연구의 시간적 범위인 1993년부



<그림 1> 연구 프레임워크

터 2022년까지 WoS에 색인되어 있는 저널 중 IF 상위 5위 학술지를 선정하였다. 선정된 저널의 저널명, IF, JCR의 Information Science & Library Science 카테고리 내 순위는 다음 <표 1>과 같다.

다음으로 WoS 데이터베이스를 통해 이상의 과정을 거쳐 선정된 5종의 학술지에 1993년부터 2022년까지 30년 동안 게재된 논문을 검색, 수집하였다. WoS에 직접 학술지명을 입력하여 검색을 수행하였으며, 문헌의 유형을 논문만으로 한정하여 수집하였다. 그 결과, 수집된 총 논문의 수는 11,252건이었으며, 이들 논문으로부터 40,897개의 저자 키워드를 수집하여 분석에 활용하였다(<표 2> 참조).

다음으로 키워드 네트워크 분석을 위해 추출된 키워드에 대한 정제 작업을 진행하였다. 이를 위해 먼저 모든 키워드의 띠어쓰기 교정, 조사 및 특수문자를 삭제하였으며, 이후 유사어와 동의어를 추출하여 키워드를 통일하는 작업

을 진행하였다. 이 과정에서 제시된 키워드는 저자가 제시한 키워드의 원형을 기준으로 하였으며, 데이터 정제 작업은 Knowledge Matrix Plus를 활용하였다.

4. 분석 결과

4.1 연도별 일반현황

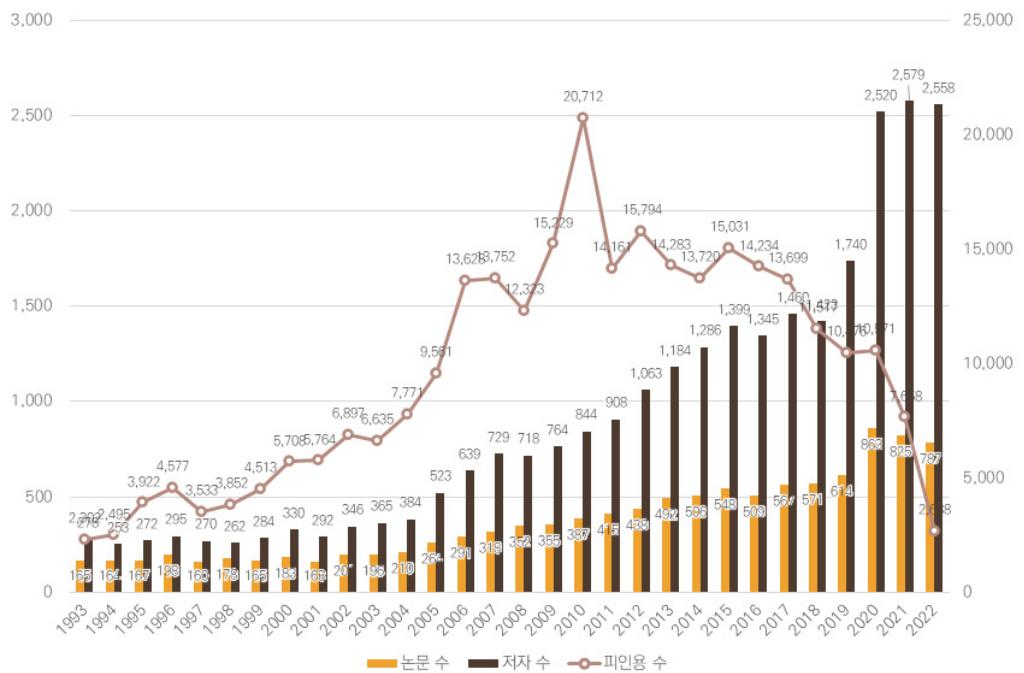
분석 대상 5종의 저널에 게재된 논문 수는 1993년 대비 2022년에 376.97% 증가하였고, 저자 수는 812.37% 증가하여 높은 성장세를 보이고 있으며, 이 중 공저논문의 비율은 1993년 50.91%에서 2022년 90.85%에 달한다. 또한 해당 저널에 게재된 논문은 총 287,006회 인용되었으며, 논문당 평균 인용 횟수는 25.5회였다(<그림 2>, <그림 3> 참조).

<표 1> 분석 대상 학술지 목록

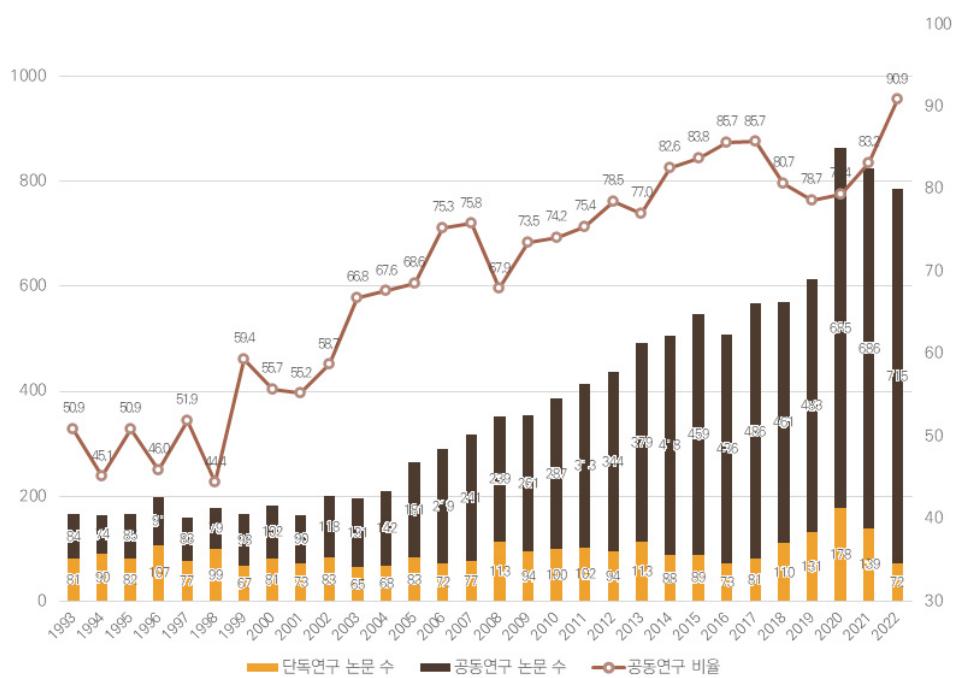
순	저널명	2021 IF	JCR 카테고리 내 순위
1	Government Information Quarterly	8,490	8/84
2	Information Processing & Management	7,466	12/84
3	Scientometrics	3,801	29/84
4	Library & Information Science Research	3,209	36/84
5	Journal of Information Science	2,462	43/84

<표 2> 학술지 별 논문 수 및 키워드 수

순	저널명	논문 수	키워드 수
1	Government Information Quarterly	1,275	4,914
2	Information Processing & Management	2,582	10,381
3	Scientometrics	5,245	20,084
4	Library & Information Science Research	752	496
5	Journal of Information Science	1,398	5,022
계		11,252	40,897



〈그림 2〉 연도별 논문 수, 저자 수, 피인용 수 추이



〈그림 3〉 연도별 공동논문 비율

4.2 문헌정보학 분야 연구 동향

4.2.1 연도별 출현 빈도 상위 키워드 비교
 1993년부터 2022년까지를 10년씩 3개의 시기로 구분하여 시기별로 학술지 논문의 저자 키워드를 빈도수 상위 20개씩 비교한 결과는 다음 <표 3>과 같다. 다만 1993년부터 2002년까지의 경우, 해당 시기에 총 1,744건의 논문이 게재되었으나 저자 키워드가 제시된 논문이 119건, 제시된 키워드는 438개에 불과하였으며, 이 중 415개의 키워드가 빈도수 2회 이하로 등장한 키워드들이었다. 따라서 해당 시기의 경우, 상위 5위까지의 키워드만을 제시하였다.

4.2.2 키워드 네트워크 분석

다음으로 문헌정보학 분야의 주요 연구 영역을 파악하기 위해 VOSviewer를 활용하여 키워드 네트워크 분석을 실시하였다. VOSviewer의 단어 맵핑 옵션은 attraction parameter는 2, repulsion parameter는 0으로 설정하였고, 클러스터링 옵션은 resolution 1.0, 최소 클러스터 크기는 1로 설정하여 클러스터링 하였다. 키워드 동시 출현빈도의 계산은 fractional counting을 적용하였다. 연구 영역의 시기별 변화와 차이를 파악하기 위해 10년 단위로 구분하여 분석하였으며, 클러스터별 대표 키워드를 빈도순으로 상위 5개 까지 제시하였다. 이후 각 클러스터 별로 대표 키워드들이 출현한 주요 논문을 직접 확인하는 과정을 거쳐 도출된 키워드의 맥락을 파악하고, 이를 바탕으로 클러스터에 대한 해석을 덧붙였다.

(1) 1기(1993년~2002년)

이 시기의 총 저자 키워드 수는 438개이고,

등장 빈도 5회 이상인 키워드는 7개였다. 이들 키워드를 대상으로 한 키워드 네트워크는 다음 <그림 4>와 같으며, 링크 수는 13개, total link strength는 15.50이었다. 총 3개의 클러스터가 나타났으나 이 중 단 하나의 키워드만을 포함하고 있는 클러스터 3은 분석에서 제외하였다.

클러스터 1은 정보 검색(information retrieval) 분야로, 정보 검색 시스템에서 정확도 향상을 위한 문서 클러스터링(clustering) 방법과 이를 위한 효과성 평가(evaluation) 연구가 주를 이루고 있으며, 유용성 평가를 위해 주로 이용자 연구(user study)를 수행하고 있는 것을 볼 수 있다.

클러스터 2는 정보 행태(information behavior) 분야로, 웹(web) 사용자의 검색 행태에 초점을 맞추어 이용자 행태를 이해하기 위한 연구를 주로 수행하는 것으로 볼 수 있다.

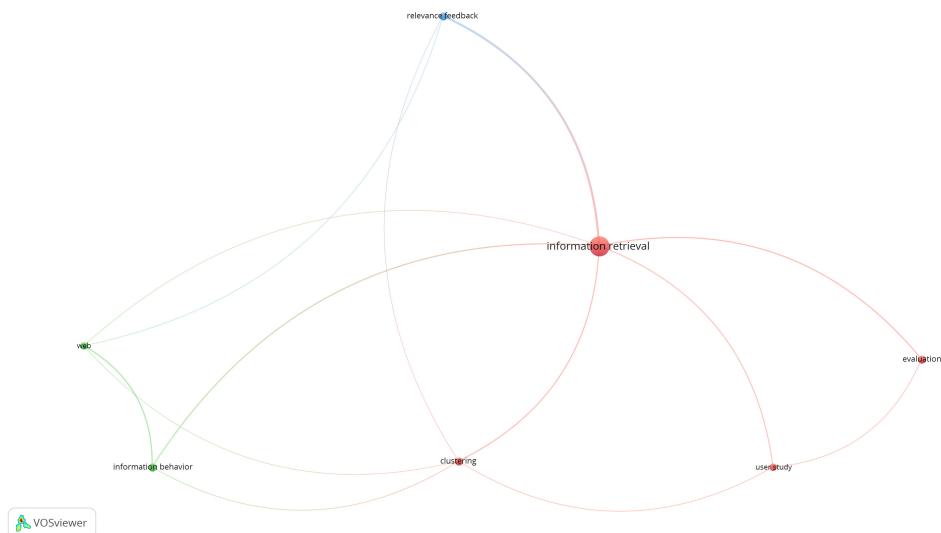
다만 주의할 것은 앞서 언급한 대로 이 시기에 저자 키워드가 제시되어 있는 논문이 전체의 9.8%로 극히 적어 해당 시기에 도출된 클러스터를 주요 연구 영역이라고 단정하기에는 무리가 있다는 점이다. 그러나 해당 시기에 진행된 연구 주제의 일부를 확인할 수 있다는 점에서 분석의 의의가 있다고 볼 수 있다.

(2) 2기(2003년~2012년)

이 시기의 총 저자 키워드 수는 5,905개이고, 등장 빈도 5회 이상인 키워드는 218개였다. 이들 키워드를 대상으로 한 키워드 네트워크는 다음 <그림 5>와 같으며, 링크 수는 1,321개, total link strength는 1,045.50이었고 10개의 클러스터가 나타났다.

〈표 3〉 시기별 상위 20위 키워드 비교

1기(1993~2002)			2기(2003~2012)			3기(2013~2022)		
순	키워드	출현 빈도	순	키워드	출현 빈도	순	키워드	출현 빈도
1	information retrieval	37	1	bibliometric analysis	140	1	bibliometric analysis	562
2	clustering	6	2	information retrieval	135	2	citation analysis	228
2	evaluation	6	3	e-government	111	3	citation	184
2	information behavior	6	4	citation analysis	70	4	scientometric	180
2	relevance feedback	6	5	citation	55	5	social media	174
3	user study	5	5	h-index	55	6	altmetrics	140
3	web	5	6	impact factor	48	6	e-government	140
4	automatic indexing	4	7	evaluation	44	7	h-index	131
4	data mining	4	7	information behavior	44	8	machine learning	121
4	expert system	4	8	digital library	28	9	social network analysis	116
5	access method	3	8	knowledge management	28	10	web of science	114
5	citation analysis	3	8	ontology	28	11	twitter	110
5	digital library	3	9	scientometric	27	12	deep learning	109
5	experiment	3	9	web	27	13	information retrieval	108
5	hypertext	3	10	search engine	26	14	topic modeling	99
5	indexing	3	11	social network analysis	25	15	impact factor	95
5	information retrieval system	3	12	co-authorship	24	15	natural language processing	95
5	interface design	3	12	collaboration	24	16	sentiment analysis	92
5	machine learning	3	13	china	23	17	research evaluation	91
5	neural network	3	13	internet	23	18	text mining	89
5	search engine	3	13	ranking	23	19	collaboration	82
5	sgml	3	13	web search	23	20	clustering	78
5	trec	3	14	natural language processing	22			
			15	machine learning	21			
			15	query expansion	21			
			16	data mining	20			
			16	university	20			
			17	clustering	19			
			17	research evaluation	19			
			17	text categorization	19			
			17	webometrics	19			
			18	social media	18			
			18	text mining	18			
			19	knowledge sharing	17			
			19	metadata	17			
			19	patent	17			
			19	publication	17			
			20	accessibility	16			
			20	network analysis	16			
			20	question answering	16			
			20	xml	16			



〈그림 4〉 키워드 네트워크(1993~2002)

〈표 4〉 클러스터 별 주요 키워드(1993~2002)

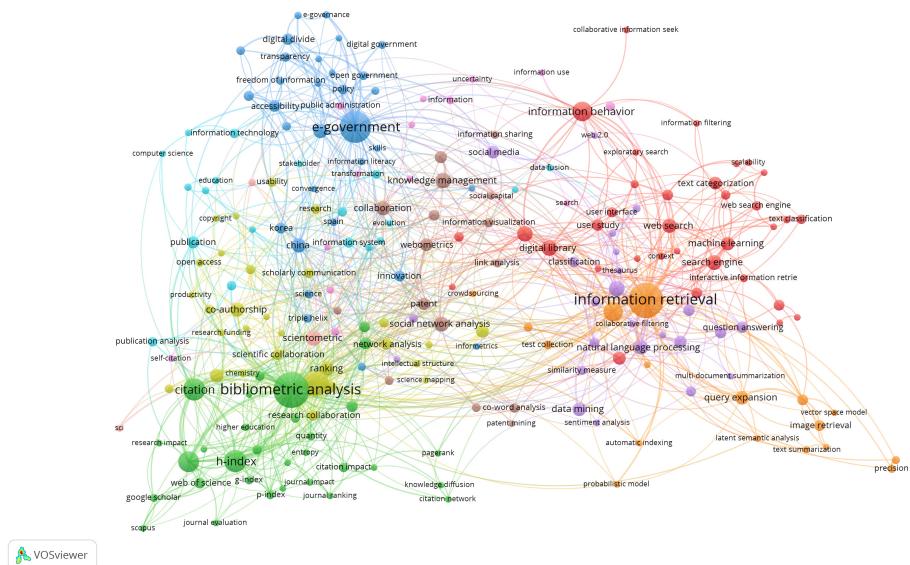
클러스터 No.	주요 키워드
클러스터 1	information retrieval, clustering, evaluation, user study
클러스터 2	information behavior, web
클러스터 3	relevance feedback

클러스터 1은 정보 행태(information behavior) 분야로, 앞선 시기에 이어서 웹 검색(web, web search)과 관련된 이용자의 정보 검색 행태와 함께 정보 검색 플랫폼으로 디지털 도서관(digital library)과 구글과 같은 웹 검색엔진(search engine)과의 비교 연구가 주로 이루어진 것을 볼 수 있다.

클러스터 2는 계량서지분석(bibliometric analysis)에 대한 연구로, WoS 데이터베이스를 대상으로 저널의 질(quality)을 평가하는 내용의 연구가 주를 이루고 있으며, 주로 h-index, impact factor와 같은 인용(citation) 기반의 지표를 활용하고 있다.

클러스터 3은 전자정부(e-government) 관련 연구로, 중국(china)의 행정 개혁(innovation)을 위한 전자정부의 구축과 이에 대한 이용자의 불평등한 접근(accessibility) 및 정보 격차(digital divide)에 대한 연구가 이루어지고 있다.

클러스터 4는 인용 분석(citation analysis) 분야로, 대학(university)의 연구 성과 평가(research evaluation) 및 순위화(ranking)를 위한 연구가 주로 이루어지며, 대표적인 연구 방법으로 공동작성(co-authorship) 네트워크 분석 및 클러스터링(clustering) 분석을 활용하는 연구가 주를 이루고 있다.



〈그림 5〉 키워드 네트워크(2003~2012)

〈표 5〉 클러스터 별 주요 키워드(2003~2012)

클러스터 No.	주요 키워드
클러스터 1	information behavior, digital library, web, search engine, web search
클러스터 2	bibliometric analysis, citation, h-index, impact factor, nanotechnology, quality, web of science
클러스터 3	e-government, china, internet, accessibility, digital divide, innovation
클러스터 4	citation analysis, co-authorship, ranking, university, clustering, research evaluation
클러스터 5	ontology, natural language processing, data mining, social media, metadata
클러스터 6	publication, information system, network, information technology, publication analysis, research productivity, case study
클러스터 7	information retrieval, evaluation, query expansion, relevance feedback, image retrieval
클러스터 8	knowledge management, social network analysis, collaboration, webometrics, knowledge sharing, patent
클러스터 9	relevance, assessment, information, public administration, simulation
클러스터 10	scientometric, author, sci

클러스터 5는 온톨로지(ontology), 메타데이터(metadata) 구축과 관련된 연구로, 소셜 미디어(social media)와 같은 대규모 비정형 다차원 데이터를 처리하기 위한 데이터 마이닝(data mining), 자연어 처리(natural language processing) 등의 연구가 주로 이루어지고 있다.

클러스터 6은 학술 출판물(publication), 출판 분석(publication analysis)을 통한 연구 생산성(research productivity) 분석으로, 주로 정보 기술(information technology), 정보 시스템(information system) 분야를 중심으로 특정 국가나 저널을 비교하는 사례연구(case study)

형태로 이루어졌으며, 네트워크(network) 분석 방법을 많이 활용하고 있다.

클러스터 7은 정보 검색(information retrieval) 분야로, 검색 결과의 평가(evaluation)와 함께 이미지 검색(image retrieval) 효율성을 향상시키기 위해 적합성 피드백(relevance feedback)을 사용한 질의 확장(query expansion)에 대한 연구가 주를 이루고 있다.

클러스터 8은 지식 경영(knowledge management) 관련 연구들이며, 특허(patent)를 활용한 지식의 공유(knowledge sharing)와 협업(collaboration) 패턴을 webometrics 및 소셜네트워크분석(social network analysis) 방법을 활용하여 분석한 연구들이다.

클러스터 9는 공공 행정(public administration) 시스템의 구축 및 서비스에 있어 정보(information) 적합성(relevance)의 평가(assessment)와 관련된 연구가 주를 이루고 있으며, 컴퓨터 기반의 모의 실험(simulation) 방법을 주로 활용하고 있다.

클러스터 10은 scientometric 분석 관련 연구들로, WoS의 SCI를 대상으로 문헌을 수집하고, 저자(author) 관련 분석을 수행한 연구들이 대부분을 차지하고 있다.

(3) 3기(2013년~2022년)

이 시기의 총 저자 키워드 수는 15,524개이고, 등장 빈도 5회 이상인 키워드는 729개였다. 이를 키워드를 대상으로 한 키워드 네트워크는 다음과 〈그림 6〉과 같으며, 링크 수는 8,320개, total link strength는 5,376.00이었고 14개의 클러스터가 나타났다.

클러스터 1은 소셜미디어(social media) 관

련 연구들로, 전자정부(e-government)에서의 시민 참여를 위한 twitter와 같은 소셜 미디어 활용 및 전자정부 서비스 제공을 위한 도구로서 소셜미디어의 유용성 및 잠재력에 대한 연구들이 포함되어 있으며, 특히, 코로나(covid-19) 유행 정부에 대한 신뢰에 소셜미디어가 미치는 영향, 인공지능(ai)과 관련한 소셜미디어 빅데이터 분석 연구 등이 주를 이루고 있다.

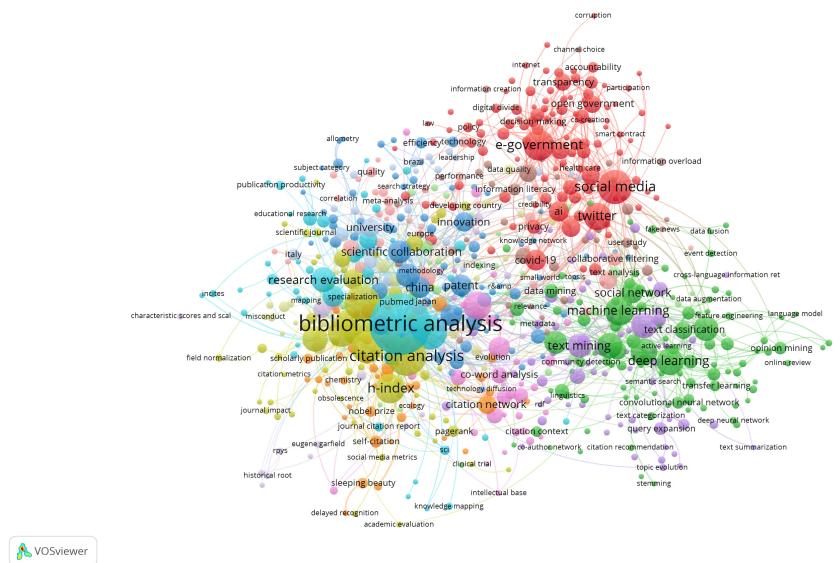
클러스터 2는 머신러닝(machine learning), 딥러닝(deep learning) 기술을 적용한 연구들로, 모델 구축을 위해 토픽 모델링(topic modeling) 및 감성분석(sentiment analysis)을 기반으로 하거나 클러스터링(clustering) 알고리즘을 적용한 연구가 주를 이루고 있다.

클러스터 3은 협업(collaboration, scientific collaboration) 패턴에 대한 연구가 주를 이루고 있으며, 주로 특허(patent) 데이터를 활용하여 학제적(interdisciplinary) 성격의 연구 분야의 협업의 효과에 대해 규명한 연구들이 있다.

클러스터 4는 인용 분석(citation analysis) 연구로, WoS 데이터베이스를 활용하여 알트 메트릭스(altmetrics) 지표와 기존의 인용 기반 지표(citation, h-index)를 비교하는 연구가 주요 연구들이라고 할 수 있다.

클러스터 5는 자연어 처리(natural language processing), 텍스트마이닝(text mining), LDA (latent dirichlet allocation) 등 온톨로지(ontology)를 구성하기 위한 새로운 방법들을 위키피디아(wikipedia)의 기사를 대상으로 제안하는 연구들로 구성되어 있다.

클러스터 6은 계량서지분석(bibliometric analysis) 관련 연구들로, scientometric, 공저자 분석(co-authorship)을 이용하여 연구 성과를 평



〈그림 6〉 키워드 네트워크(2013~2022)

〈표 6〉 클러스터 별 주요 키워드(2013~2022)

클러스터 No.	주요 키워드
클러스터 1	social media, e-government, twitter, covid-19, ai
클러스터 2	machine learning, deep learning, topic modeling, sentiment analysis, clustering
클러스터 3	collaboration, scientific collaboration, interdisciplinary, patent, china
클러스터 4	citation analysis, citation, altmetrics, h-index, web of science
클러스터 5	natural language processing, text mining, latent dirichlet allocation, ontology, wikipedia
클러스터 6	bibliometric analysis, scientometric, research evaluation, co-authorship, research performance
클러스터 7	citation network, content analysis, science, citation count, computer science
클러스터 8	information behavior, evaluation, information literacy, misinformation, personalization
클러스터 9	social network analysis, network analysis, co-word analysis, research trend, co-citation
클러스터 10	recommender system, ranking, impact, indicator, google scholar
클러스터 11	productivity, network, text analysis, emotion analysis, semantic analysis
클러스터 12	university, higher education, research assessment, university ranking, bibliometric indicator
클러스터 13	journal ranking, india, scientific impact, economics, pubmed
클러스터 14	citation impact, literature review, normalization, learning, characteristic scores and scales

가(research evaluation, research performance) 하는 연구가 주를 이루고 있다.

클러스터 7은 인용 네트워크(citation network) 관련 연구들로, 과학적 영향력을 더욱 정교하게 평가할 수 있도록 기존의 인용 수(citation

count) 기반의 분석에서 인용의 내용(content analysis)을 고려하는 연구들이 이루어지고 있는 것을 확인할 수 있다.

클러스터 8은 정보 행태(information behavior) 분야로, 잘못된 정보(misinformation)를 평가

(evaluation)하기 위한 정보 리터러시(information literacy) 관련 연구와 개인화(personalization)된 정보 요구에 대한 내용이 주를 이루고 있다.

클러스터 9는 소셜 네트워크 분석(social network analysis, network analysis)으로, 키워드동시출현분석(co-word analysis) 및 동시인용(co-citation) 분석을 활용하여 연구 트렌드(research trend)를 분석하는 연구들이 대부분이다.

클러스터 10은 추천 시스템(recommender system) 관련 연구로, 검색 결과의 순위(ranking)와 관련하여 저널의 영향력(impact)을 식별하기 위한 지표(indicator)들의 값이 데이터베이스별로 차이가 발생하는지에 대해 규명한 연구들이 주를 이루며, 주로 구글 스칼라(google scholar)와 비교하는 연구들이 대다수 포함되어 있다.

클러스터 11은 저자 및 저널 등 학문적 생산성(productivity)을 측정하고 분석하기 위한 연구들로, 주로 서지 데이터베이스(bibliographic database)를 대상으로 수집된 데이터를 활용하여 네트워크(network) 분석, 텍스트분석(text analysis), 감정 분석(emotion analysis) 등을 수행한 연구가 있다.

클러스터 12는 대학 등 고등교육기관(university, higher education)의 순위(university ranking)를 측정하는데 있어 필수적이라고 할 수 있는 연구 성과 평가(research assessment)를 위한 계량서지적 지표(bibliometric indicator)와 관련된 연구들로 구성되어 있다.

클러스터 13은 저널의 순위(journal ranking)와 과학적 영향력(scientific impact)에 대한 연구들이 포함되어 있으며, 이 시기의 경우 주로 경제학(economics) 분야의 저널을 대상으

로 순위를 측정한 연구와 pubmed를 활용한 과학 문헌의 영향력 측정 연구들이 다수 포함되어 있다.

클러스터 14는 인용 영향력(citation impact)과 관련된 연구들로, 인용 영향력을 측정하는 지표들에 대한 문헌 연구(literature review), 인용 영향력을 측정하기 위해 인용 횟수를 정규화(normalization)하여 분석한 연구들과 인용 영향력을 확인하기 위한 새로운 척도로서 CSS (characteristic scores and scales)를 적용한 연구들이 주를 이루고 있다.

4.3 유망 연구 주제 도출

본 연구에서는 향후 문헌정보학 분야에서 유망할 것으로 예상되는 연구 주제들을 도출하고자 한다. 이를 위해 문헌정보학의 하위 분야에서 어느 정도 성숙기에 접어들었다고 판단할 수 있는 키워드를 선정 유망 연구 주제로, 향후 유망 주제로 부상할 가능성이 있는 키워드를 미래 유망 연구 주제로 명명하였다. 다음으로 선정 유망 연구 주제 선정 및 공통적 특징 추출, 미래 유망 연구 주제 후보군 도출, 선정 유망 연구 주제의 공통적 특징과 미래 유망 연구 주제 후보군과의 상관관계 분석, 최종 유망 연구 주제 선정의 절차로 문헌정보학 분야에서의 유망 연구 주제를 도출하였다.

4.3.1 선정 유망 연구 주제 선정 및 공통적 특징 추출

문헌정보학 분야의 선정 유망 연구 주제를 선정하기 위한 과정은 다음과 같다. 선정 유망 연구 주제의 선정은 이상의 키워드 분석을 토대로

2기에는 등장하지만 3기에는 등장하지 않는 키워드가 다수 포함되어 있는 2기의 클러스터 1과 클러스터 7을 대상으로 하였다. 2기와 3기에 모두 높은 빈도로 등장하는 키워드의 경우 현재까지도 지속적으로 활발하게 연구가 이루어지고 있는 주제라고 볼 수 있으므로 여전히 성장 중인 연구 주제일 가능성은 배제할 수 없으며, 그러한 경우에 선행 유망 연구 주제로 보기에는 적절하지 않다고 판단하였기 때문이다.

먼저 2개 클러스터에 등장하는 상위 키워드를 각각 5개씩 1차로 추출하였다. 다음으로 해당 클러스터 내에서 2기에는 등장하지만 3기에는 등장하지 않는 키워드를 2차로 추출하였다. 2차로 추출된 키워드는 20개였으나 이 20개의 키워드 중 2기에는 평균 인용 빈도가 상위권이

면서, 현재 시점을 기준으로 해당 키워드를 포함하는 논문의 수가 크게 증가하지 않거나 또는 감소하는 모습을 보이는 키워드만 추출하였다. 이렇게 추출한 키워드는 총 10개였으며, 1차 추출 시 포함된 키워드가 2개 포함되어, 중복 포함된 경우를 고려하면 총 18개의 키워드가 도출되었다. 마지막으로 분석에 사용할 최종 키워드는 2022년 기준 직전 30년간 해당 키워드를 포함하고 있는 논문의 수를 기준으로 300건 이상의 논문이 출판된 키워드로 하였는데, 이는 연평균 논문 수가 10편이 되지 않을 정도로 적은 경우 시계열적 추세를 생성하기에는 미흡하다고 판단하였기 때문이다.

결과적으로 10개의 키워드를 최종 선정하였으며, 분석에 사용할 키워드는 다음 <표 7>과 같다.

<표 7> 분석 대상 키워드

구분	순	키워드	논문 수	최종 선정 여부
클러스터 1	1	information behavior	3,017	○
	2	digital library	2,454	○
	3	web	3,851	○
	4	search engine	522	○
	5	web search	850	○
클러스터 7	1	information retrieval	1,077	○
	2	evaluation	2,700	○
	3	query expansion	18	×
	4	relevance feedback	24	×
	5	image retrieval	92	×
2013년 이후 미등장 키워드 (평균인용빈도 순)	1	web search engine	304	○
	2	transaction log analysis	55	×
	3	web	3,851	○
	4	collaborative information seeking	115	×
	5	language modeling	242	×
	6	automatic indexing	83	×
	7	search engine	522	○
	8	domain analysis	653	○
	9	user interface	541	○
	10	test collection	281	×

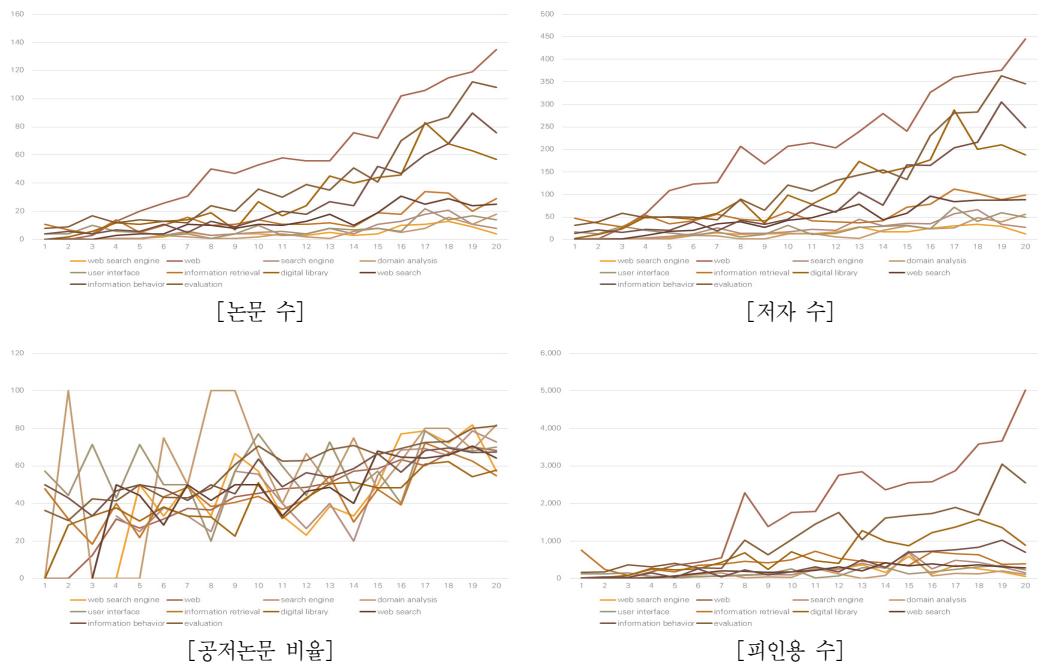
다음으로 최종 10개의 키워드를 대상으로 논문 수, 저자 수, 공저논문 비율, 피인용 수의 공통적 추세를 생성하기 위해 주성분분석을 실시하였다. 주성분분석을 통한 공통 추세의 생성에는 2012년을 기준으로 과거 20년의 데이터를 활용하였으며, 최종 수집된 10개 선행 유망 연구 주제의 논문 수, 저자 수, 공저논문 비율, 피인용 수를 대상으로 2012년을 기준으로 한 지난 20년 간의 시계열적 패턴은 다음 <그림 7>과 같다.

다음으로 10개 선행 유망 연구 주제의 공통적 추세를 생성하기 위해 논문 수, 저자 수, 공저논문 비율, 피인용 수의 4개 지표에 대하여 주성분분석을 실시하였다. <표 8> 및 <그림 7>에서 보는 바와 같이 이 10개의 키워드는 각 지표마다 편차가 있으므로 이를 데이터를 정규화하는 과정을 거쳤으며, 이후 제1주성분(PC1)을 기준으로 각 지표를 대표하는 단일 차원의 새로운 지표를 생성하였다. 주성분분석을 통해 2012년을 기준으로 과거 20년 기간 동안의 선행 유망 연구 주제의 지표별 공통적 추이를 도출하였으며, 그 결과는 다음 <그림 8>과 같다.

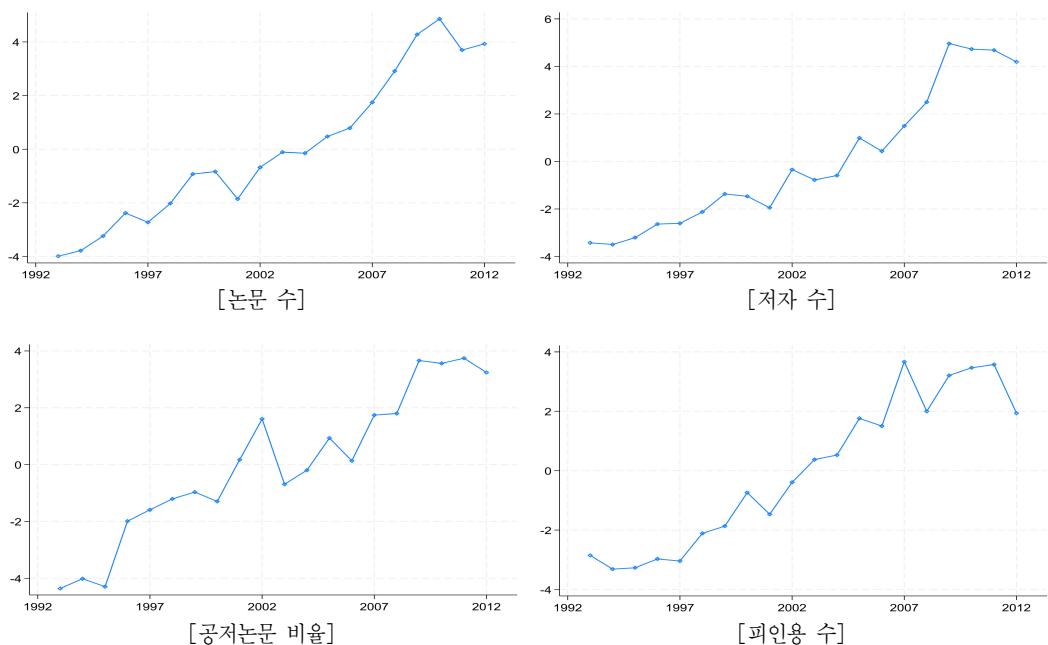
이를 통해 알 수 있는 선행 유망 연구 주제의 특징을 종합하면 다음과 같다. 첫째, 논문 수는 지속적, 선형적으로 증가하는 추세를 보이고 있다. 논문 수의 경우, 2012년을 기준으로 약 5~6년 전부터 급증하며, 2년 전 시점에 가장 많은 논문 수를 보이고 있다. 둘째, 저자 수 역시 논문 수와 마찬가지로 지속적, 선형적으로 증가하는 추세를 보이고 있으며, 2001년까지는 완만한 성장세를 보이다가 그 이후부터 급격히 증가하기 시작하여, 2012년을 기준으로 3년 전 시점에 저자 수가 가장 많은 모습을 보이고 있다. 셋째, 공저논문의 비율은 다른 지표들과 비교할 때 상대적으로 앞선 시기부터 증가세를 보이고 있으며, 일부 시기를 제외하면 그 증가 폭도 큰 편이다. 이는 공동연구가 증가하고 있는 현재의 추세와도 관련이 있지만 연구 분야가 성장하면서 해당 분야에 관심을 갖는 연구자가 증가하고, 더불어 공동연구가 활발하게 이루어지는 데에서 기인했을 것으로 판단할 수 있다. 넷째, 피인용 수는 2012년 기준 약 5년 전 시점에 가장 높은 모습을 보이고 있다. 특히, 공

<표 8> 분석 대상 키워드의 전반적 특징

순	키워드	논문 수	저자 수	공저논문 비율	피인용 수
1	information behavior	3,017	7,509	76.5	21,162
2	digital library	2,454	5,151	60.5	21,162
3	web	3,851	8,385	63.6	58,147
4	search engine	522	1,152	64.4	5,611
5	web search	850	1,945	66.6	7,241
6	information retrieval	1,077	2,314	60.9	11,281
7	evaluation	2,700	6,891	74.9	39,312
8	web search engine	304	672	67.4	3,730
9	domain analysis	653	1,668	77.2	5,773
10	user interface	541	1,296	72.1	5,868
합계		15,969	36,983	684.1	179,287



〈그림 7〉 선행 유망 연구 주제(10개)의 시계열 패턴(20년)



〈그림 8〉 선행 유망 연구 주제(10개)의 공통적 추세

저논문의 비율은 논문 수, 저자 수와는 달리 상당히 시간상으로 앞서 큰 폭으로 증가하는 경향을 보이는데, 다른 지표들과 함께 사용한다면 유망 연구 주제를 파악하는 유의미한 지표로 활용할 수 있을 것으로 판단하여 이 네 가지 지표를 모두 상관분석에 활용하였다.

4.3.2 미래 유망 연구 주제 탐색 대상 키워드 도출

새롭게 등장한 키워드를 많이 포함하고 있는 클러스터에는 논문 수, 저자 수, 인용 수 등이 상대적으로 많지 않은 새로운 연구들이 포함되어 있을 가능성이 높다. 따라서 각 클러스터 별로 3기에 처음으로 등장한 키워드를 출현빈도 순으로 상위 5개를 선정하여 총 70개의 키워드를 분석에 활용하였다(〈표 9〉 참조).

4.3.3 선행 유망 연구 주제의 공통적 특징과 미래 유망 연구 주제 후보군과의 상관 관계 분석

다음으로 미래 유망 연구 주제를 선정하기 위해 선행 유망 연구 주제와 유사성을 보이는 키워드를 도출하고자 하였다. 본 연구는 선행 유망 연구 주제의 논문 수, 저자 수, 공저논문 비율, 피인용 수의 추세와 유사한 모습을 보이는 주제가 유망한 연구 주제가 될 가능성이 클 것이라는 것을 전제로 하고 있다. 이를 위해 앞서 수행한 주성분 분석을 통해 생성된 선행 유망 연구 주제의 공통적 추세와 미래 유망 연구 주제 후보군 간에 유사성을 확인하기 위한 피어슨 상관계수 추정을 실시하였다. 예측의 불확실성이라는 특성을 고려하여 논문 수, 저자 수, 공저논문 비율, 피인용 수 모두에서 유의한

〈표 9〉 클러스터별 신규 키워드

클러스터 1	클러스터 2	클러스터 3	클러스터 4	클러스터 5
twitter	deep learning	collaboration network	altmetrics	lda
covid-19	topic modeling	emerging technology	jurnal	wikipedia
ai	neural network	efficiency	bibliometric study	semantic similarity
big data	diversity	knowledge flow	mendeley	linked data
blockchain	link prediction	scientific output	citation distribution	information quality
클러스터 6	클러스터 7	클러스터 8	클러스터 9	클러스터 10
social science	citation count	misinformation	research trend	recommender system
research output	scientific publication	information need	co-citation	impact
scientific productivity	sleeping beauty	academic career	bibliographic coupling	science policy
research area	nobel prize	information overload	vosviewer	technology
journal citation report	scholarly publication	network dynamics	keyword analysis	data analysis
클러스터 11	클러스터 12	클러스터 13	클러스터 14	
text analysis	university ranking	scientific impact	literature review	
emotion analysis	gender difference	economics	learning	
bibliographic database	internationalization	pubmed	css	
semantic analysis	comparative study	retraction	climate change	
recommendation	framework	science of science	creativity	

상관성을 보이는 키워드를 선정하는 것이 가장 적절하다고 판단하였고, 결과적으로 상관관계 분석을 통해 논문 수, 저자 수, 공저논문 비율, 피인용 수의 4개 지표 전체에서 선행 유망 연구 주제와 $p < 0.01$ 수준에서 양(+)의 상관성을 보이는 키워드를 도출하였다.

선행 유망 연구 주제의 공통적 특징과 미래 유망 연구 주제 후보군과의 상관관계 분석 결과는 다음 〈표 10〉과 같으며, 선행 유망 연구 주제의 공통적 패턴을 따라서 유망 연구 주제로 부상할 가능성이 높은 키워드는 twitter, ai, big data, deep learning, topic modeling, neural network, link prediction, emerging technology,

efficiency, lda, semantic similarity, information quality, social science, research area, scientific publication, scholarly publication, research trend, vosviewer, keyword analysis, impact, science policy, technology, data analysis, university ranking, gender difference, framework, scientific impact, learning의 총 29개 키워드였다.

4.3.4 최종 유망 연구 주제 선정

이상의 상관관계 분석 결과를 토대로 향후 유망 연구 주제로 부상할 가능성이 높다고 예상할 수 있는 키워드를 가장 많이 포함하고 있는 클러스터를 최종 유망 연구영역으로 선정하

〈표 10〉 상관관계 분석 결과

키워드	논문 수	저자 수	공저논문 비율	피인용 수
twitter	.889**	.908**	.812**	.673**
covid-19	.548*	.634**	.585**	.476*
ai	.700**	.776**	.596**	.653**
big data	.864**	.902**	.855**	.750**
blockchain	.769**	.824**	.792**	.509*
deep learning	.682**	.753**	.578**	.585**
topic modeling	.836**	.889**	.635**	.624**
neural network	.727**	.805**	.624**	.678**
diversity	.871**	.915**	.192	.077
link prediction	.774**	.803**	.746**	.586**
collaboration network	.927**	.948**	.520*	.635**
emerging technology	.905**	.934**	.597**	.746**
efficiency	.912**	.950**	.636**	.657**
knowledge flow	.866**	.912**	-.147	.105
scientific output	.914**	.936**	.172	.285
altmetrics	.871**	.914**	.522	.759**
journal	.862**	.925**	.946**	.507*
bibliometric study	.878**	.914**	.235	.754**
mendeley	.854**	.894**	.095	.748**
citation distribution	.803**	.815**	.273	-.106
latent dirichlet allocation(lda)	.741**	.827**	.665**	.686**
wikipedia	.900**	.902**	.680**	.340
semantic similarity	.839**	.884**	.673**	.607**
linked data	.896**	.928**	.473*	.685**

키워드	논문 수	저자 수	공저논문 비율	피인용 수
information quality	.918**	.960**	.692**	.694**
social science	.916**	.955**	.794**	.781**
research output	.908**	.938**	.636**	.314
scientific productivity	.901**	.921**	.399	.382
research area	.918**	.954**	.777**	.789**
journal citation report	.854**	.896**	.267	-.084
citation count	.797**	.882**	.333	.175
scientific publication	.882**	.931**	.591**	.707**
sleeping beauty	.680**	.635**	.600**	.137
nobel prize	.667**	.668**	.401	.323
scholarly publication	.882**	.927**	.748**	.661**
misinformation	.618**	.706**	.751**	.542*
information need	.882**	.908**	.850**	-.586**
academic career	.888**	.936**	.298	.352
information overload	.884**	.886**	.627**	.546*
network dynamics	.895**	.912**	.602**	.463*
research trend	.888**	.932**	.600**	.785**
co-citation	.839**	.849**	.488*	.606**
bibliographic coupling	.637**	.708**	.482*	.616**
vosviewer	.650**	.699**	.757**	.643**
keyword analysis	.837**	.878**	.653**	.793**
recommender system	.876**	.924**	.579**	.535*
impact	.907**	.950**	.807**	.741**
science policy	.916**	.947**	.815**	.781**
technology	.928**	.966**	.869**	.733**
data analysis	.904**	.943**	.857**	.774**
text analysis	.883**	.890**	.313	.668**
emotion analysis	.855**	.892**	.598**	.619**
bibliographic database	.838**	.909**	.605**	.304
semantic analysis	.917**	.937**	.200	.710**
recommendation	.872**	.933**	.835**	.416
university ranking	.880**	.915**	.784**	.612**
gender difference	.889**	.938**	.616**	.613**
internationalization	.930**	.953**	.171	.163
comparative study	.897**	.943**	.523*	.685**
framework	.909**	.951**	.797**	.676**
scientific impact	.909**	.946**	.647**	.710**
economics	.787**	.859**	.720**	.412
pubmed	.762**	.747**	.495*	-.135
retraction	.674**	.723**	.634**	.356
science of science	.602**	.706**	.710**	.121
literature review	.920**	.946**	.734**	.477*
learning	.904**	.948**	.901**	.598**
characteristic scores and scales	.675**	.640**	.630**	.496*
climate change	.804**	.789**	.542*	.347
creativity	.841**	.855**	.434	.572**

** p<0.01, * p<0.05

〈표 11〉 문헌정보학 분야 미래 유망 연구 주제

영역	대표 주제	대표 신규 키워드
클러스터2	머신러닝과 알고리즘	deep learning, topic modeling, neural network, link prediction
클러스터10	연구 영향력	impact, science policy, technology, data analysis

였다. 앞서 도출된 29개의 키워드 중 클러스터 2와 클러스터 10에서 가장 많은 4개의 키워드가 포함되어 있었으며, 3개의 키워드가 포함된 클러스터는 클러스터 1, 클러스터 5, 클러스터 9, 클러스터 12였다. 가장 많이 포함하고 있는 2개의 클러스터를 문헌정보학 분야의 미래 유망 연구 주제로 최종 선정하였고, 해당 클러스터에 포함된 신규 키워드를 구체적으로 살펴보면 다음 〈표 11〉과 같다.

클러스터 2에서 유망 연구 주제로 부상할 가능성이 큰 신규 키워드는 deep learning, topic modeling, neural network, link prediction이었다. 즉, 데이터 처리와 분석을 위한 머신러닝 및 딥러닝 기법의 적용과 이를 구현하기 위한 알고리즘 관련 연구가 여전히 주요한 연구 주제가 될 것으로 예상된다.

클러스터 10에서 유망 연구 주제로 부상할 가능성이 큰 신규 키워드는 impact, science policy, technology, data analysis이었다. 이는 향후 연구의 영향력과 과학 기술 정책과의 관련성에 대한 연구들이 주요 연구 주제가 될 가능성이 있으며, 데이터 분석을 기반으로 하는 연구들이 중심이 될 것으로 예상된다.

이 외에도 3개 이상의 키워드가 포함되어 있는 클러스터를 살펴보면 해당하는 연구 주제는 소셜미디어와 빅데이터 분석(클러스터 1), 자연어 처리(클러스터 5), 연구 트렌드 분석(클러스터 9), 연구성과 평가(클러스터 12)이며,

이들 주제와 관련된 연구 역시 향후 주요한 연구 주제라고 할 수 있을 것이다.

그러나 이 외에 다른 클러스터의 경우, 향후 유망 연구영역으로서의 영향력은 상대적으로 적다고 볼 수 있으며, 시간이 지날수록 점차 비중이 낮은 연구주제가 될 가능성이 있다.

5. 결 론

이 연구에서는 문헌정보학 분야의 연구 동향 분석을 통해 시기별 핵심 연구 영역을 파악하고, 향후 유망 연구 주제로 부상할 가능성이 있는 주제를 도출하였다. 이를 위해 1993년부터 2022년까지 문헌정보학 분야 IF 상위 학술지 5종의 논문 11,252건에서 수집한 저자 키워드를 대상으로 동시출현단어분석을 수행하였으며, 미래 유망 연구 주제를 도출하기 위해 주성분분석 및 상관관계분석을 활용하였다. 그 결과로 선정된 문헌정보학 분야의 미래 유망 연구 주제는 ‘머신러닝과 알고리즘’, ‘연구 영향력’이었다. 물론 본 연구의 결과로 도출된 주요 키워드들이 현재 시점에 속해 있는 클러스터가 시간이 지나면서 다른 클러스터로 분화하거나 진화하여 새로운 클러스터를 형성할 가능성을 배제할 수는 없다. 그러나 본 연구는 새로운 키워드를 많이 포함하고 있는 클러스터의 경우, 향후 유망 연구 주제로 부상할 가능성이 있는 새로운 연구들이 포함되어 있을 것

이라는 점을 전제로 하고, 해당 클러스터 아래에 포함되어 있는 신규 키워드들을 유망 연구 주제 탐색 대상 그룹으로 설정한 후 과거에 유망했던 연구 주제들과 유사한 특징을 보이는 주제들을 유망 연구 주제로 선정하였다.

이 연구에서 아쉬운 점은 문헌정보학 분야 전체를 연구 대상으로 함으로써 문헌정보학의 하위 분야를 보다 세밀하게 들여다보는데 한계가 있다는 것과 본 연구의 결과로 도출된 유망 연구 주제에 대한 정성적 검증이 부재하다는 것이다. 또한 본 연구에서 분석대상으로 선정한 저널이 5종에 불과하고, Scopus 데이터베이스와 비교할 때 WoS 데이터베이스 내 데이터의 누락으로 인하여 예측 범위가 제한적이므로 연구 결과의 일반화에는 무리가 있으며, 방법론적 측면에 있어 본 연구에서 미래 유망 연구 주제를 선정하기 위해 활용한 상관관계분석이 두 변수 간의 선형적 관계를 파악하는 방법이므로 키워드가 출현하는 시점의 차이를 고려하지 못했다는 한계도 존재한다. 따라서 연구결과의 적용 가능성은 높이기 위해서는 문헌정보학의 하위 분야를 대상으로 연구를 진행할 필요가 있으며, 아울러 데이터 누락 등 WoS 데이터베이스의 한계를 극복하기 위해 데이터 수집 범위 및 대상 저널의 범위를 확대한 추가 연구가 필요하다.

그러나 이 연구는 거시적인 측면에서 문헌정보학 분야의 연구 동향을 분석하고, 시기별로 연구 주제가 어떻게 변화하는지 그 흐름과 방향을 파악하였다는 점에서 의의가 있다. 또한 최근 들어 머신러닝 기반의 예측 모형 구축 시 모델의 정확도를 높이기 위한 차원 축소 기법으로 주성분분석을 많이 활용하는 점을 고려할 때, 미래 유망 연구 주제의 발굴을 위해 효율성과 객관성을 담보할 수 있는 방법으로서 주성분분석을 적용해봄으로써 문헌정보학 분야에서의 예측 연구의 확장 가능성을 제공했다는 데에 학문적 의의가 있다고 하겠다.

따라서 후속 연구에는 문헌정보학의 하위 분야 및 타 분야를 대상으로 해당 방법론을 검증할 필요가 있으며, 정성적 검증이 함께 이루어져야 할 필요가 있다. 또한 시계열상관분석을 통해 출현 시점에 대하여 키워드 간의 시간 차를 고려한다면 더욱 세밀한 예측이 가능할 것이다. 아울러 이 연구에서는 선행 유망 연구 주제의 특성을 분석하기 위해 연구의 영향력을 대표할 수 있는 단일 지표로서 논문 수, 저자 수, 공저논문 비율, 피인용 수의 4개 지표를 기준으로 하였으나, 유망 연구 주제의 특징을 보다 잘 나타낼 수 있는 추가 지표를 규명하기 위한 연구도 필요할 것으로 보인다.

참 고 문 헌

- 고재창, 조근태, 조윤호 (2013). 키워드 네트워크 분석을 통해 살펴본 기술경영의 최근 연구동향. *지능정보연구*, 19(2), 101-123. <http://dx.doi.org/10.13088/jiis.2013.19.2.101>
- 김선호, 배덕호, 이문환 (2016). 주성분분석을 활용한 자연가뭄지수 산정 및 평가. *한국수자원학회논문집*, 49(6), 565-577. <https://doi.org/10.3741/JKWRA.2016.49.6.565>

- 김성호, 김지표 (2020). 주성분분석을 활용한 우리나라 주요 기업의 특허 경쟁력 분석. 산업혁신연구, 36(4), 213-241. <https://doi.org/10.22793/indinn.2020.36.4.009>
- 김현정, 조남옥, 신경식 (2015). 항공산업 미래유망분야 선정을 위한 텍스트 마이닝 기반의 트렌드 분석. 지능정보연구, 21(1), 65-82. <http://dx.doi.org/10.13088/jiis.2015.21.1.65>
- 민인식 (2018). Stata를 이용한 KOSIS 100대 지표 웹 크롤링. The Korean Journal of Stata, 5(2), 29-38.
- 안세정, 심위, 이준영, 권오진, 노경란 (2012). 과학계량학 기법을 이용한 디스플레이 연구영역의 트렌드 탐지. 한국전자통신학회논문지, 7(6), 1343-1351. <https://doi.org/10.13067/JKIECS.2012.7.6.1343>
- 양기덕, 김선옥, 이해경 (2021). 국제 및 국내 문헌정보학 분야의 연구성과 비교 분석. 한국문헌정보학회지, 55(1), 365-392. <https://doi.org/10.4275/KSLIS.2021.55.1.365>
- 이수상 (2014). 언어 네트워크 분석 방법을 활용한 학술논문의 내용분석. 정보관리학회지, 31(4), 49-68. <https://doi.org/10.3743/KOSIM.2014.31.4.049>
- 이재윤, 김관준, 강대신, 김희정, 유소영, 이우형 (2011). 계량서지적 기법을 활용한 LED 핵심 주제영역의 연구 동향 분석. Journal of Information Science Theory and Practice, 42(3), 1-26. <https://doi.org/10.1633/JIM.2011.42.3.001>
- 정의섭, 여운동, 조대연, 서일원 (2008). 과학계량학적 정보분석을 통한 건설교통분야의 유망연구영역 도출. 한국콘텐츠학회논문지, 8(2), 231-238. <https://doi.org/10.5392/JKCA.2008.8.2.231>
- An, L., Yu, C., & Li, G. (2014). Visual topical analysis of chinese and american library and information science research institutions. Journal of Informetrics, 8(1), 217-233.
- 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, 561-577. <http://dx.doi.org/10.1007/s11192-006-0041-0>
- Cano, V. (1999). Bibliometric overview of library and information science research in Spain. Journal of the American Society for Information Science, 50(8), 675-680.
- Dai, P. F., Xiong, X., & Zhou, W. X. (2021). A global economic policy uncertainty index from principal component analysis. Finance Research Letters, 40, 101686.
- Guo, H., Weingart, S., & Börner, K. (2011). Mixed-indicators model for identifying emerging research areas. Scientometrics, 89(1), 421-435.
- Hu, C.P., Hu, J.M., Deng, S.L., & Liu, Y. (2013). A co-word analysis of library and information science in China. Scientometrics 97, 369-382. <https://doi.org/10.1007/s11192-013-1076-7>
- Jolliffe, I. T. (2002). Principal Component Analysis (2nd ed.). New York: Springer.
- Kontostathis, A., Galitsky, L. M., Pottenger, W. M., Roy, S., & Phelps, D. J. (2004). A survey

- of emerging trend detection in textual data mining. In Berry, W. Michael eds. Survey of Text Mining. New York: Springer, 185-224.
- Mane, K. K. & Börner, K. (2004). Mapping topics and topic bursts in PNAS. PNAS, 101(1), 5287-5290. <https://doi.org/10.1073/pnas.0307626100>
- Radovanović, M., Filipović, S., & Golušin, V. (2018). Geo-economic approach to energy security measurement-principal component analysis. Renewable and Sustainable Energy Reviews, 82, 1691-1700.
- Takeda, Y. & Kajikawa, Y. (2009). Optics: a bibliometric approach to detect emerging research domains and intellectual bases. Scientometrics, 78(3), 543-558.
- Uzun, A. (2002). Library and information science research in developing countries and Eastern European countries: a brief bibliometric perspective. The International Information and Library Review, 34(1), 21-33.
- Walters, W. H. & Wilder, E. I. (2016). Disciplinary, national, and departmental contributions to the literature of library and information science, 2007-2012. Journal of the Association for Information Science and Technology, 67(6), 1487-1506.
- Wang, Q. (2018). A bibliometric model for identifying emerging research topics. Journal of the association for information science and technology, 69(2), 290-304.
- Weismayer, C. & Pezenka, I. (2017). Identifying emerging research fields: a longitudinal latent semantic keyword analysis. Scientometrics, 113(3), 1757-1785.
- 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, July 22, 2022. <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>
- Jung, Eui-Seob, Yeo, woon dong, Cho, Dae-yeon, & Suh, Il-Won (2008). Emerging research field selection of construction & transportation sectors using scientometrics. The Journal of the Korea Contents Association, 8(2), 231-238.

- Kho, Jaechang, Cho, Keuntae, & Cho, Yoon Ho. (2013). A study on recent research trend in management of technology using keywords network analysis. *Journal of Intelligence and Information Systems*, 19(2), 101-123. <http://dx.doi.org/10.13088/jiis.2013.19.2.101>
- 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. <http://dx.doi.org/10.13088/jiis.2015.21.1.65>
- Kim, Seon-Ho, Bae, Deg-hyo, & Lee, Moon Hwan (2016). Estimation and assessment of natural drought index using principal component analysis. *Journal of Korea Water Resources Association*, 49(6), 565-577. <https://doi.org/10.3741/JKWRA.2016.49.6.565>
- Kim, Sung-Ho & Kim, Jipyo (2020). Analysis of the patent competitiveness of major Korean companies using PCA. *Journal of Industrial Innovation*, 36(4), 213-241. <https://doi.org/10.22793/indinn.2020.36.4.009>
- Lee, Jae Yun, Kim, Pan Jun, Kang, DaeShin, Kim Hee jung, Yu, So-Young, & Lee, Woo-Hyoung (2011). A bibliometric analysis on LED research. *Journal of Information Science Theory and Practice*, 42(3), 1-26. <https://doi.org/10.1633/JIM.2011.42.3.001>
- Lee, Soosang (2014). A content analysis of journal articles using the language network analysis methods. *Journal of the Korean Society for Information Management*, 31(4), 49-68. <https://doi.org/10.3743/KOSIM.2014.31.4.049>
- Min, Insik (2018). Web Crawling of KOSIS 100 indicators using Stata. *The Korean Journal of Stata*, 5(2), 29-38.
- Yang, Kiduk, Kim, SeonWook, & Lee, HeyKyung (2021). Comparison of research performance between domestic and international library and information science scholars. *Journal of the Korean Society for Library and Information Science*, 55(1), 365-392. <https://doi.org/10.4275/KSLIS.2021.55.1.365>