

대학생 대상 정보 리터러시 연구의 키워드 네트워크 분석 및 토픽 모델링

Keyword Network Analysis and Topic Modeling in an Information Literacy Study of Undergraduate Students

이다현 (Da-Hyeon Lee)*

신동희 (Donghee Shin)**

초 록

정보 리터러시는 정보화 사회를 살아가는 모든 사람들에게 필요한 역량이지만, 대학생은 학업 수행과 직업 준비의 과정에서 정보 리터러시를 갖춰야 할 필요성이 특별히 높은 시기이다. 본 연구에서는 KCI에 등재된 대학생 대상 연구를 대상으로 진행한 정보 리터러시 관련 연구의 영문 초록을 대상으로 빈도 분석, 네트워크 분석, 토픽 모델링을 수행하여 대학생 대상 정보 리터러시 연구의 동향을 파악하고자 하였다. 빈도분석과 키워드 네트워크를 분석하고 그 결과를 비교하여 주요 키워드와 후속 연구 주제를 도출하였으며, 토픽모델링 결과 8개의 하위 토픽을 도출하여 주 연구 분야를 관찰하였다. 대학생 대상 정보 리터러시는 주로 교육적 목적을 중심으로 연구가 진행되었으며, 간호 정보와 분석 모델 개발이 주요 하위 토픽으로 드러났다.

ABSTRACT

Information literacy is a necessary competency for all people living in the information society, but undergraduate students are especially in need of information literacy in the process of academic performance and career preparation. In this study, we conducted frequency analysis, network analysis, and topic modeling on the English abstracts of information literacy-related research on undergraduate students listed in KCI to identify trends in information literacy research on undergraduate students. The main keywords and subsequent research topics were derived by analyzing the frequency analysis and keyword network and comparing the results, and eight subtopics were derived from the topic modeling to observe the main research areas. Information literacy for college students was mainly studied for educational purposes, and nursing information and analysis model development were the main subtopics.

키워드: 정보 리터러시, 대학생, 텍스트 마이닝, 빈도 분석, 토픽 모델링, 네트워크 분석
information literacy, undergraduate student, text mining, frequency analysis, topic modeling, network analysis

* 숙명여자대학교 문헌정보학과 석사과정(2331672@sookmyung.ac.kr) (제1저자)

** 숙명여자대학교 문헌정보학과 조교수(shindh@sookmyung.ac.kr) (교신저자)

■ 논문접수일자: 2024년 8월 15일 ■ 최초심사일자: 2024년 8월 26일 ■ 게재확정일자: 2024년 9월 13일

■ 정보관리학회지, 41(3), 249-268, 2024. <http://dx.doi.org/10.3743/KOSIM.2024.41.3.249>

※ Copyright © 2024 Korean Society for Information Management

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. 서론

1.1 연구의 배경 및 목적

정보통신 기술과 모바일 기기의 발전으로 인해 누구나 손쉽게 정보를 생산하고 공유할 수 있는 환경이 조성되었다(신승용, 2021). 이로 인해 정보 생산량은 급격하게 증가하고 있으며, 이는 기술과 사회적 혁신의 기반이 되고 있다. 그러나 정보화에 긍정적인 면만 있는 것은 아니다. 최근에는 생성형 AI의 등장 및 확산으로 인해 정보화의 부작용에 대한 우려가 높아지고 있다. 정보화의 문제점의 예시로는 학습 대상의 데이터 오염으로 인한 잘못된 정보, 이용자의 선호에 따라 특정 주제만을 선별적으로 노출하여 발생하는 정보의 편향, 생성형 AI가 부정확한 정보를 생성하는 데이터 환각인 할루시네이션(Hallucination) 등이 있다(한운수, 류현숙, 2023). 이러한 오염된 정보의 위험에서 올바르게 정보를 습득하고 활용하기 위해 습득한 정보의 사실 여부와 유용성을 판단할 수 있는 능력인 정보 리터러시(Information Literacy)(이수상, 2007)가 어느 때보다도 중요한 능력으로 부상하고 있다.

영미권 국가에서는 정보 리터러시의 중요성을 인식하여, 일찍이 관련 연구와 보고서를 통해 정보 리터러시의 기준과 모델을 개발하였다(이혜영, 남태우, 2008). 미국 대학 및 연구 도서관 협회(ACRL, Association of College & Research Libraries)에서 제시한 고등교육을 위한 정보 활용능력 표준(Information Literacy Competency Standards for Higher Education)은 정보 리터러시를 “평생 학습의 기반”이며,

“모든 학문 분야, 교육 환경, 교육 수준에서 필요한 능력”으로 정의한(ACRL, 2000, 2) 대표적인 지침이며, 호주와 뉴질랜드에서도 정보 활용능력연구소(ANZIIL, Australasian and New Zealand Institute for Information Literacy)에서 Australian and New Zealand Information Literacy Framework(Doskatsch, 2002)를 발간하여 학부 과정을 통한 평생 학습자 개발을 목적으로 정보 리터러시에 대한 프레임워크를 생성하였다. 이들 선행 연구에서는 정보 리터러시를 정보화 사회에서 모두에게 필요한 역할로 보고 있으나, 향후 사회인으로서 다양한 정보 문제를 겪을 준비단계에 있는 대학생들에 특히 주목하고 있다(김태훈, 2010). 대학생은 교과학습의 일환으로 리포트 형식의 과제 연구를 수행하고, 차후 사회의 전문 인력으로 성장하는 준비 단계이며, 이 시기에 형성된 문제 해결 능력과 비판적 사고 능력은 향후 평생 학습을 좌우한다(이수상, 2007). 또한 대학생은 높은 수준의 일반지식을 갖추어 정보 활용능력을 향상하고, 이를 토대로 정보와 정보기술의 복잡성을 극복하여 자기 주도적 학습의 토대를 갖추는 필요가 있다(이혜영, 남태우, 2008). 이러한 특성을 기반으로 대학생의 정보 리터러시에 대한 연구가 국내외에서 다수 진행되었다. 대학생들의 정보 활용능력 연구는 학생들의 정보 관리, 검색 및 통신 기술을 향상시키는 밑바탕이 되므로 디지털 시대에 잘 적응할 수 있도록 하는 데 매우 중요하다(Liming et al., 2024).

이미 리터러시 관련 연구가 다수 진행되었으나, 리터러시라는 단어의 의미적 모호함으로 인해 유사한 개념이 혼재되어 사용되는 경우가 많으므로, ‘정보 리터러시’를 명확하게 한정하

여 연구를 진행하는 것이 필요하다. 또한 대학교육과 관련한 리터러시 연구는 선행 연구나 설문조사를 활용한 평가 모델 개발에 초점이 맞추어져 있으며 문헌정보학, 교육학, 간호학 등 각 학문 분야에서 독자적으로 연구가 이루어지고 있다. 본 연구에서는 다양한 학문 분야에서 이루어지는 대학생 대상 정보 리터러시 연구를 통합적으로 살펴보기 위하여 텍스트 마이닝 분석 기법을 사용하였다. 텍스트 마이닝은 고유의 목적으로 작성된 텍스트 데이터를 분석하므로 응답자와 연구자의 편견이 배제되기 때문에(김기옥, 2020) 여러 학문 분야를 포괄하기에 적합하다. 이를 위해 ‘대학생’과 ‘정보 리터러시’ 두 가지 키워드를 동시에 포함한 논문이 최초로 등장하는 1999년부터 2024년 7월 사이의 학술논문을 수집하여 텍스트 마이닝을 수행하였다. 해당 논문에는 일본어(1건), 중국어(1건)를 포함하여 국문 초록이 존재하지 않는 논문도 다수 존재하므로, 영문 초록을 대상으로 텍스트 마이닝을 진행하였다. 분석 도구로는 R version 4.3.1을 사용하였다.

2. 선행 연구

연구를 진행하기에 앞서 대학생 대상 정보 리터러시 관련 선행 연구와 리터러시 관련 논문에서 텍스트 분석 기법을 사용한 연구를 검토하여 관련 논의가 발전된 과정을 살펴보았다.

2.1 정보 리터러시 관련 선행 연구

리터러시(Literacy)는 1880년대부터 사용된

단어로, 초기에는 ‘문맹’의 반대 의미로 ‘읽고 쓸 수 있는 능력’을 의미했다(박주현, 2018; Oxford English Dictionary, 2018). 이 단어는 시간이 지남에 따라 다양한 문맥에서 의미가 파생되었고, 현재에는 특정 주제, 혹은 미디어를 ‘읽을’ 수 있는 능력, 혹은 ‘특정 분야에 대한 지식이나 경쟁력’으로 의미가 확장되었다(ACRL, 2016).

최근에는 정보 기술의 발전으로 인하여 인터넷을 비롯한 디지털 환경에서의 리터러시 개념이 점차 중요해지고 있으며, 이러한 경향에 따라 독서, 정보, 컴퓨터, IT, ICT 등 특정 주제와 리터러시가 결합된 복합 리터러시가 사용되고 있으나, 교육 현장에서는 이러한 개념을 혼재하여 사용하고 있다(박주현, 2018).

그 중 정보 리터러시의 개념은 Zurkowski의 1974년 보고서 ‘The Information Service Environment, Relationships and Priorities’에서 최초로 사용되었다(Zurkowski, 1974). 이 보고서는 정보 문해(Information Literate)를 특정 주제에 대한 지식을 탐색할 수 있는 능력으로 정의하며, 탐색 과정에 필요한 도구, 기법, 기관이 프레임워크를 형성한다고 설명하였다.

오의경, 장혜란(2005)은 선행 연구를 분석하여 대학생의 리터러시 측정 도구를 개발하였으며, 각 요인과 요인 사이의 관계와 영향력을 통계적으로 분석하였다. 그 결과 성별, 전공 등 개인적 배경 요인 외에 도서관 이용교육 노출도와 교수진의 조언, 정보원 이용 경험 등을 리터러시 영향 요인으로 도출하였다.

이수상(2007)은 대학생의 과제 해결 과정 능력의 요건을 도출하여 정보 리터러시 모형을 제시하였으며, 정보 리터러시의 개념을 “정보의 요구를 식별하고, 요구된 정보를 찾아내고, 찾아

진 정보를 비판적으로 평가하고, 새로운 지식의 생산에 적용하며 관리하는 등의 정보 활용 전반에서 요구되는 일정 수준의 능력들의 집합”으로 정의했다.

김태훈 (2010)은 문헌 연구로 정보 활용능력의 주요 요소를 도출하여 ‘정보요구’, ‘정보자원’, ‘정보검색’, ‘정보분석’, ‘정보관리 및 소통’, ‘정보윤리’의 6개 영역으로 구분하였다. 각 영역을 반영한 설문지를 활용하여 신학 대학교 재학생을 대상으로 정보 활용능력을 평가하였으며, 이 결과를 바탕으로 정보 활용능력 증진 방안을 제시하였다.

조민정, 구미옥 (2021)은 간호학 전공 대학생을 대상으로 간호 정보 활용 역량 교육 프로그램을 개발하였으며, 대조 실험을 통하여 그 효과를 검증하였다. 해당 프로그램은 간호 정보 활용 역량, 문제해결능력, 자기주도적 학습능력, 근거기반실무역량 향상에 효과적이었다.

Liu et al.(2024)은 중국 광저우 지역의 대학생들의 정보 활용능력 수준을 측정하는 설문조사 연구를 수행하였다. 약 1,700명의 설문을 분석하여 정보 활용능력에 영향을 미치는 요인으로 학년, 출신 고교, 관련 교과 과정 수강 여부, 전자 기기 사용 여부 등을 도출하였다. 또한 정보 검색, 정보 관리, 커뮤니케이션 기술 측면에서 개선의 필요성을 제시하며 관련 커리큘럼의 필요성을 강조하였다.

2.2 텍스트 마이닝 적용 선행 연구

텍스트 마이닝은 “정형 또는 비정형 텍스트 데이터에서 자연어처리과정과 데이터마이닝 기법을 통하여 방대한 텍스트에서 의미 있는 정

보를 추출해내고, 가공하는 것을 목적으로 하는 기술”(배규용 외, 2013, 1429)이다. 텍스트 마이닝 기법을 활용한 리터러시 연구에는 다음과 같은 논문이 있다.

장수현, 남영준(2022)는 WoS와 KCI에 수록된 리터러시 관련 논문을 대상으로 키워드 분석과 토픽 모델링을 수행하여 국내외 리터러시 연구의 주요 키워드를 비교하였다. 특히 전체 리터러시 관련 연구와 문헌정보학 분야의 리터러시 연구의 연구 동향, 연구량 변화 시기, 핵심 키워드를 비교 분석하여 ‘정보 활용 교육’, ‘학교 도서관’, ‘사서교사’ 등의 키워드를 문헌정보학 분야의 주요 리터러시 관련 키워드로 도출했다.

이창봉 외(2021)는 ‘리터러시’ 키워드의 연구 동향과 학술 데이터별 양상을 나타내기 위해 토픽 모델링과 네트워크 분석을 활용하였다. 2017년 이후 리터러시 관련 연구가 급증한 것을 확인할 수 있었으며, ‘디자인 이미지와 디지털 미디어’, ‘문화예술, 교양정보’, ‘교육 정책, 교육 내용’, ‘디지털 사회 변화’라는 네 개의 하위 주제를 도출하였다.

윤지혜, 정유경(2022)은 문헌정보학 분야 4개 학술지에서 정보 활용 교육과 관련된 논문을 선정하여 초록 텍스트를 대상으로 LDA 기법을 사용한 토픽모델링을 수행하였다. 하위 토픽으로 ‘미디어 리터러시’, ‘정보 활용 교육의 교수학습방법’, ‘정보원 이용 및 판단’을 도출하였으며, 그 결과를 활용하여 정보 활용 교육의 내용적, 방법론적 제안을 제시하였다.

임정훈 (2022)은 빈도분석, 키워드 네트워크 분석, 토픽모델링을 활용하여 문헌정보학 분야에서 정보 활용 교육 연구의 동향을 분석하였다. 분석 결과 ‘정보’, ‘도서관’, ‘교육’, ‘교사’ 등

의 키워드를 중심으로 네트워크가 형성되었으며, 토픽모델링의 결과로는 ‘교육과정’, ‘정보 활용 교육 효과’ 등 15개의 토픽을 도출하였다. 도출 결과를 바탕으로 정보 활용 교육 대상의 확대, 교육과정 도입에 대한 논의의 필요성, 미디어 및 디지털 리터러시 연구의 필요성을 제안하였다.

기존의 대학생 대상 정보 리터러시 관련 연구는 주로 선행 연구를 정성적으로 검토하여 리터러시의 개념이나 평가 요소를 도출하는 방식으로 진행되었다. 최근에는 간호, 교육 등 연구 대상의 전공 계열에 따른 정보 리터러시 연구가 수행되고 있으나, 이는 각 학문 분야에서 독자적으로 진행되고 있다. 또한 2022년 이후 전 학문 분야의 리터러시 연구의 동향을 살피는 논문이 발표되기도 했으나, 그 중에서도 대학생의 정보 리터러시에 주목한 학제적 연구는 미비하다. 이번 연구에서는 전 학문 분야에서 대학생이라는 대상에 한정된 정보 리터러시를 특정하여 기존의 연구 동향과 구별되는 특징을 발견하고자 한다.

3. 연구 방법

3.1 연구 대상 선정

KCI에 수록된 논문 중 ‘대학생’과 ‘정보 리터

러시’ 키워드를 모두 포함하고 있는 논문을 선정했다. 수집 과정에서 관련 주제를 다루고 있는 논문이 누락되는 것을 방지하기 위하여 동의어, 유의어를 검색어에 포함하였다. 동의어 관계는 이창봉 외의 2021년 연구와 네이버 사전을 참조하였다(이창봉 외, 2021).

정보 리터러시의 동의어로는 ‘information literacy’, ‘정보문해력’, ‘정보 문해’, ‘정보 활용능력’, ‘정보 소양’, ‘정보소양’을 선정하였으며, 대학생의 동의, 유사어로는 ‘고등교육’, ‘고등 교육’, ‘대학’, ‘higher education’, ‘college’, ‘university’, ‘undergraduate’를 선정했다(〈표 1〉 참조).

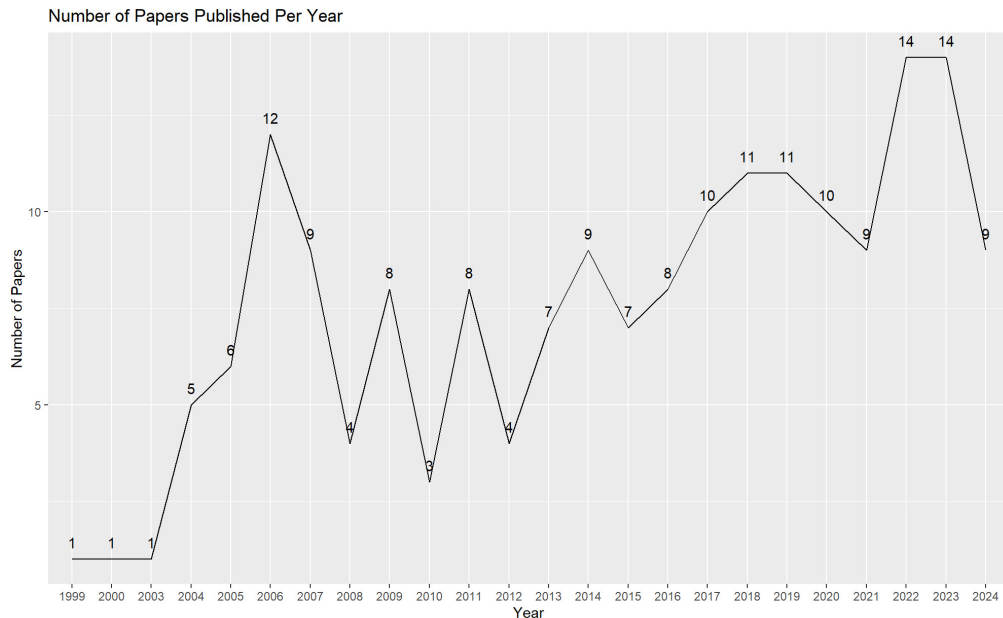
KCI 검색 결과 총 271건이 반환되었으며, 해당 논문, 제목, 키워드, 초록을 검토하여 다음 사항에 해당하는 경우 분석 대상에서 배제하였다.

먼저 논문 제목, 혹은 초록에서 연구 대상이 명확하게 ‘고등학생’, ‘고령층’ 등 명백하게 대학생이 아닌 것을 확인할 수 있는 경우이다. 단 성인 일반을 연구 대상으로 하는 경우 학력 수준 혹은 재학여부를 변인으로 포함하고 있는 경우는 분석 대상에 포함하였다.

마찬가지로 ‘미디어 리터러시’, ‘디지털 리터러시’ 등 다른 종류의 리터러시를 주제로 하며, 해당 개념과 정보 리터러시를 명백하게 구별하고 있는 논문의 경우 역시 분석 대상에서 제외하였다. 단 건강정보 리터러시 등 특정 분야의 정보 리터러시를 주제로 하고 있는 경우에는 분석 대상에 포함하였다. 최종적으로 총 271건

〈표 1〉키워드 동의어 및 유의어 관계

NO	대표어	동의어 및 유의어
1	정보 리터러시	information literacy, 정보문해력, 정보 문해, 정보 활용능력, 정보 활용능력, 정보 소양, 정보소양
2	대학생	고등교육, 고등 교육, 대학, higher education, college, university, undergraduate



〈그림 1〉 대학생 대상 정보 활용능력 연구 논문의 연도별 발행 현황

의 논문 중 89건을 제외한 181건을 분석 대상으로 확정하여 해당 논문의 영문 초록을 크롤러(Crawler)를 활용하여 수집하였고, OCR을 제공하지 않는 2건의 경우 직접 입력하여 수집하였다(〈그림 1〉 참조).

연도별로 대상 논문의 분포를 살펴보면, 연구가 진행되는 초창기인 1999년부터 2003까지의 대학생을 대상으로 하는 정보 리터러시 연구는 1년에 1건 이하로 부진하였으나, 2004년을 기점으로 지속적으로 관련 연구가 발표되었다. 특히 IFLA에서 평생교육을 위한 정보 활용능력 가이드라인(Guidelines on Information Literacy for Lifelong learning)을 발표한 2006년에는 관련 연구가 12건으로 급증하였다. 이 가이드라인은 교육 프로그램에 종사하는 정보 전문가들이 모든 도서관 환경에서 정보 리터러시 프로그램을 수행할 수 있도록 정보 리터러

시의 개념을 명확히 하고, 국제적인 기준을 제시하였으며 프로그램에의 적용 방안을 제시했다(Lau, 2006). 정보 리터러시에 대한 국제적 논의가 활발해짐에 따라 국내에서도 해당 분야에 대한 관심이 높아진 것으로 보인다. 이후 2008년부터 상승과 하락을 거듭하며 2015년 이후부터는 매년 7건 이상의 논문이 생산되고 있다. 이는 장수현, 남영준(2022)의 논문에서 KCI의 리터러시 주제 논문의 연도별 발행 추이와 유사하며, 이창봉 외(2021)가 제시한 연구의 도입기(~2008), 증가기(2009~2016), 급증기(2017~)와 대체로 유사하게 나타났다.

3.2 분석 방법

3.2.1 전처리

전처리 과정에는 영문 초록의 모든 텍스트를

활용하였다. 영어 키워드의 경우 한글 키워드와 다르게 동일한 단어가 문맥에 따라 다양한 품사로 활용될 수 있고, 정규화 과정을 거쳐 어원에 따라 의미를 통합하게 되므로 명사 외에도 모든 품사를 분석에 활용하였다.

동일한 의미의 단어가 표현에 따라 별개의 키워드로 분산되는 것을 방지하기 위해 R 4.3.1 버전을 사용하여 텍스트 전처리를 진행하였다. 먼저 'bix6', 'AI', 'IT' 키워드는 그 자체로 의미 있는 키워드라고 판단하여 각각 'bigsix', 'artificialintelligence', 'informationtechnology'로 정규화 과정에서 제외되지 않도록 하나의 키워드로 인식하였다. 다음으로 stringr 패키지의 tolower 함수를 사용하여 모든 문자를 소문자로 변환하였고, str_replace_all 함수로 특수 문자와 숫자를 제거했다. 'Nursing information literacy', 'Health information literacy', 'ILI', (information literacy instruction), 'problem solving' 등 단어 구가 특정 의미를 갖는 경우 해당 단어들이 토큰으로 나누어져 의미가 달라지는 경우를 방지하기 위해 하나의 키워드로 인식하기 위해 총 55개의 어휘를 수정하였다. 또한 단어들의 표제어를 추출하는 lemmatize_words 함수로 동일한 어원의 키워드들을 하나의 키워드로 집중시켰으며, stop_words 데이터 셋을 활용하여 분석에 활용하기 어려운 전치사, 대명사 등 728개 키워드를 제거하였다. 또한 연구 대상 논문을 수집 과정에서 검색어로 사용된 'information literacy', 'student', 'university', 'college', 'undergraduate' 키워드는 대다수의 초록에 포함되어 있으므로 유의미한 분석을 도출해 내기 어려워 분석 과정에서 제외하였다.

3.2.2 단어 빈도 분석(Term Frequency Analysis)

단어빈도분석은 키워드가 문서에서 출현한 빈도를 바탕으로의 문서에서 키워드의 분포를 파악하는 기초적 분석 방법이다(윤지혜, 정유경, 2022). 본 연구에서는 영문 초록에서 키워드의 등장 횟수를 누적하여 각 키워드가 전체 문서에서 등장하는 빈도를 나타내는 CF(collection frequency, 장서 빈도)를 적용하였다. CF는 전체 데이터에서 해당 용어의 중요성을 평가하는데 유용하므로(Zafer & Hakan, 2012) 상위 100개 키워드와 각 키워드의 누적 출현 횟수를 분석하여 주요 개념을 파악하였다.

3.2.3 네트워크 분석(Network Analysis)

네트워크 분석은 노드와 에지라는 개념을 사용하여 분석 대상의 관계를 시각적으로 나타낸다(이수상, 2014). 본 연구에서는 동시 출현 관계를 기반으로 전처리 과정을 거친 키워드를 노드로 설정하고, 키워드가 각 논문의 초록에서 동시에 등장하는 횟수를 에지로 설정하여 키워드 네트워크를 분석하여 각 키워드의 의미적 연결 관계와 각 키워드의 중요성을 파악하였다. 네트워크 그래프는 노드의 크기와 위치, 에지의 거리와 굵기 등을 활용하여 노드의 연관 관계를 시각적으로 표현하므로(이수상, 2018) 연구에서 중심이 되는 키워드를 명시적으로 파악할 수 있으며, 의미적 연관성이 높은 키워드 쌍을 도출해낼 수 있다. 본 연구에서는 노드 간의 물리적 거리와 이론적 최단 거리의 유사성이 높은 Kamada-Kawai 레이아웃을 사용했다. Kamada-Kawai 레이아웃은 대칭성이 높고 에지의 교차를 최소화하므로 시각적 관찰이

용이하다(Hasal et al., 2017)는 장점이 있다.

또한 각 노드의 연결 중심성(degree centrality), 근접 중심성(closeness centrality), 매개 중심성(betweenness centrality)을 계산하여 네트워크 내에서 각 노드의 역할을 분석할 수 있다. 연결 중심성은 각 노드와 직접적으로 연결된 이웃 노드의 개수, 즉 각 노드의 에지 수를 기준으로 모든 종류의 네트워크에서 계산된다. 근접 중심성과 매개 중심성은 네트워크 내에서의 최단 거리에 기반하여 계산된다(Freeman, 1978). 근접 중심성은 한 노드에서 다른 모든 노드로 도달하는 경로를 고려하며, 매개 중심성은 해당 특정 노드가 다른 노드의 최단 거리에서 등장하는 비율을 기반으로 결정된다(Koschutski & Schreiber, 2008). 중심성 분석을 활용하면 시각적으로 관찰하여 도출한 네트워크에서 키워드의 역할을 객관적 수치로 검증할 수 있다.

3.2.4 토픽 모델링(Topic Modeling)

토픽 모델링은 키워드가 각 토픽에 존재할 확률을 계산하고, 각 문헌에서 모든 키워드의 분포를 고려하여 문헌의 구조를 예측하는 문헌 분석 모델이다(박자현, 송민, 2013). 토픽 모델링을 활용하면 문서 집합 내의 내재적 의미를 탐색할 수 있다(윤지혜, 정유경, 2022). 토픽 모델링 기법의 예시로는 LDA(Latent Dirichlet Allocation, 잠재 디리클레 할당), STM(structural topic modeling, 구조적 토픽 모델링), 토픽의 관계 포착에 로지스틱 정규분포(logistic normal distribution)를 적용한 CTM(Correlated, Topic Model, 상관토픽모델링) 등이 있다(김기욱, 2020).

이번 연구에서는 가장 대표적인 토픽 모델링 알고리즘인 LDA 기법을 사용하였다. LDA는

Blei(2003)에 의해 제시된 확률 기반 토픽 모델링 기법으로, 문서 내의 텍스트 데이터를 분석하여 문서 내의 토픽과 구성 비율을 분석한다(Blei, 2003). LDA는 특정 주제에서 키워드가 사용될 확률을 기반으로 잠재적으로 유의미한 주제를 추론하기에 유용하다(임정훈, 2022).

토픽 생성 과정에서 토픽의 주제가 자동으로 도출되는 것은 아니므로, 토픽 별 주요 키워드의 연계성을 고려해서 연구자가 직접 주제명을 부여했다(유은지 외, 2012). 각 토픽의 주요 키워드는 dplyr 라이브러리의 arrange(-beta) 함수를 사용하여 도출한 베타(β) 값을 기준으로 판단했다. 베타 값은 해당 키워드가 토픽에서 등장할 확률을 0과 1의 범위로 나타내는 수치이며, 베타 값이 1에 가까울수록 토픽에서 의미적 중요성이 크다.

토픽 생성 이후 논문 초록을 이루는 모든 키워드의 베타 값을 합산하여 각 논문과 주제적으로 가장 밀접한 토픽을 연결하였으며, 전체 논문에서 각 토픽의 비율을 분석하였다.

4. 연구 결과

본 연구에서는 텍스트 마이닝 기법 중 단어 빈도 분석, 중심성 네트워크 분석, LDA 기반 토픽 모델링을 수행하였다. 먼저 빈도 분석으로 등장 횟수가 높은 키워드를 파악하였고, 이후 동시 출현 관계에 근거하여 키워드 네트워크에서 키워드의 중심성과 상관관계를 관찰하였다. 마지막으로 토픽 모델링으로 하위 주제를 도출하고, 각 토픽 별 핵심 키워드와 각 토픽에 해당하는 논문의 수를 파악하였다.

전처리 과정에서 토큰의 분산을 방지하기 위해 수정한 <표 2>의 어휘 중 변환한 키워드가 의미를 파악하기 어려운 경우, 시각화 과정에서 원 의미를 고려하여 대표어로 다시 변환하여 표시하였다.

4.1 키워드 빈도 분석

키워드 빈도분석을 수행하여 전체 초록 텍

스트에서 등장한 회수에 따라 정렬하여 상위 100개의 키워드를 <표 2>로 나타내었다. 전체 초록 텍스트에서 가장 많이 등장한 키워드는 'education'로, 총 312회 등장했다. 'information' 299회, 'competency' 222회, 'library' 149회, 'learn' 146회의 출현 횟수가 다음으로 많았다.

빈도 행렬을 분석한 결과 다음과 같은 결론을 도출했다. 첫째, 'education' 312회, 'teach' 96회, 'curriculum' 84회, 'academic' 82회, 'class'

<표 2> 빈도 분석 결과

키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도
education	312	problem solving	73	instruction	44	influence	34
information	299	communication	72	media	44	information literacy	34
competency	229	service	69	online	44	instruction	34
library	149	improve	68	collect	43	measure	34
learn	146	survey	66	include	43	school	34
ability	142	content	65	librarian	43	experience	33
nurse	134	level	65	questionnaire	42	require	33
literacy	119	knowledge	64	examine	41	characteristic	32
skill	118	conduct	60	korea	41	learner	32
base	116	relate	59	nursing information literacy	41	satisfaction	32
program	115	major	57	critical	40	support	31
analyze	101	class	55	item	40	utilization	31
analysis	97	subject	55	perception	40	internet	30
teach	96	process	53	relationship	40	material	30
effect	94	development	52	variable	40	plan	30
model	94	suggest	52	standard	39	read	30
develop	89	provide	50	change	38	source	30
curriculum	84	difference	49	core	38	awareness	29
factor	83	significant	47	job	38	direct	29
teacher	83	test	47	positive	37	effective	29
academic	82	user	47	computer	36	evaluate	29
digital	82	follow	46	identify	36	strategy	29
design	79	system	46	korean	36	technology	29
write	75	apply	44	understand	35	affect	28
data	74	educational	44	future	34	art	28
						concept	28

55회 등 대학 현장에서의 교육 및 학업과 관련된 키워드가 다수 높은 빈도로 나타났다. 이러한 키워드는 분석 대상 논문을 수집하는 과정에서 사용된 키워드와 밀접한 주제적 연관성이 있으며, 대학 교육 과정에서의 리터러시 교수, 학습 관련 연구가 다수 존재하는 것이 원인으로 예상된다.

둘째, 학생의 역량과 관련된 'competency' 222회, 'ability' 142회, 'skill' 118회, 'improve' 68회 등의 키워드도 다수 등장했다. 대학생의 학업 수행 능력의 일환으로 정보 리터러시를 파악하거나, 혹은 리터러시와 학업 수행 능력의 연관성을 연구한 논문이 다수 존재하였으며, 해당 논문에서 이러한 키워드를 반복적으로 사용한 것이 원인으로 파악된다.

셋째, 간호 및 보건 의료 관련 키워드가 일부 나타났다. 'nurse' 134회, 'nursing information literacy' 41회 외에도 'health' 16회, 'electronic medical record' 15회 등의 키워드로 간호 및 보건의료 분야의 정보 리터러시 연구가 유의미한 빈도로 이루어지고 있음을 알 수 있다.

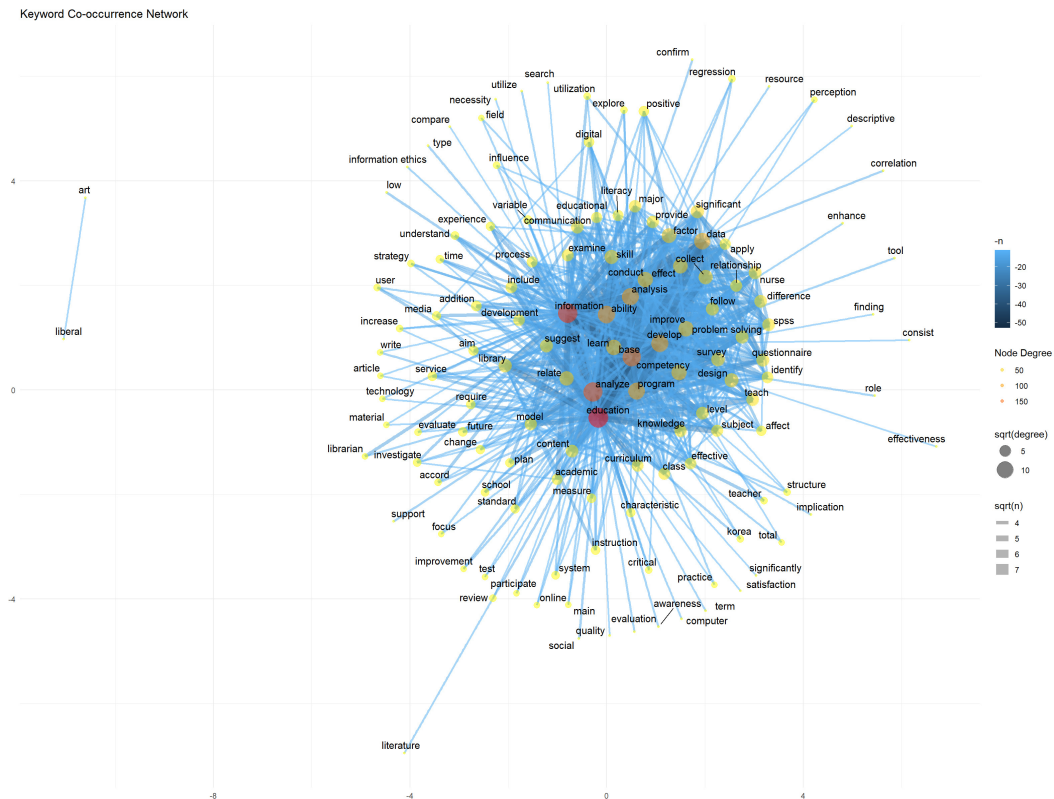
마지막으로, 'digital' 82회 외에 'data' 74회, 'online' 44회, 'computer' 36회, 'internet' 30회 등 온라인 및 디지털 환경과 관련된 키워드가 다수 등장했다. 온라인에서 오프라인으로 변화하는 교육 환경에서 대학생의 정보 리터러시 측정, 강화, 교수 방안 연구가 활발하게 진행되었으며, COVID-19로 인해 온라인 형태의 비대면 수업이 확대됨에 따라 학습 환경에 따른 정보 리터러시 능력의 변화에 대한 관심이 반영되었음을 유추하였다.

4.2 네트워크 분석

키워드를 노드로, 키워드 간의 동시 출현 관계를 에지로 설정하여 네트워크 그래프를 작성하여 〈그림 2〉로 나타내었다. 노드는 키워드의 단순 빈도에 비례하여 크기와 색상을 조절하였으며, 에지는 동시 출현 회수에 따라 색상과 굵기를 나타내었다. 노드와 에지가 지나치게 중첩되지 않도록 동시 출현 회수가 10회를 초과하는 키워드로 제한하여 총 132개 키워드의 801쌍의 동시출현관계로 네트워크를 구성했다. 이후 각 노드의 연결 중심성, 근접 중심성, 매개 중심성 수치를 산출한 뒤 소수점 3자리까지 반올림하였다.

총 132개의 노드에서 연결 중심성, 근접 중심성, 매개 중심성이 가장 높은 키워드는 빈도 분석에서 출현 빈도가 가장 높았던 'education' (연결 1.46, 근접 0.78 매개 0.26)으로 나타났다. 'information'은 네트워크 분석에서 연결 중심성 1.36, 근접 중심성 0.76, 매개 중심성 0.22로 중심성 값의 합이 두 번째로 높게 나타나 네트워크 전반에서 주요한 키워드임을 알 수 있다. 'analyze'(1.25, 0.73, 0.17), 'base'(1.1, 0.69, 0.05), 'base'(1.1, 0.69, 0.1) 등이 네트워크 그래프에서 주요 키워드로 파악되었다.

연결 중심성, 근접 중심성, 매개 중심성 사이에는 높은 상관관계가 나타났다. 각 키워드의 중심성 값의 순위를 비교한 결과, 세 중심성의 순위는 대체로 유사하게 나타났다. 그러나 'art'와 'liberal' 키워드의 경우 근접 중심성이 1.0로 높게 나타났으나 연결 중심성은 0.015, 매개 중심성은 0으로 낮게 나타났다. 이 키워드 쌍은 다른 키워드와 연결 관계를 갖지 않는 별도의



〈그림 2〉 동시 출현 관계 기반 키워드 중심성 네트워크

고립된 하위 커뮤니티를 형성하고 있다. ‘art’와 ‘liberal’ 키워드는 네트워크 전체에서는 중요한 역할을 수행하고 있지 않아 대학생 대상 정보 리터러시라는 연구 주제와 직접적 관련성이 적으나, 특정한 문맥에서 의미적 연결성을 강하게 갖는 것으로 보인다. 두 키워드가 동시에 등장하는 논문을 살펴본 결과, 두 키워드는 별개의 키워드가 동시에 출현하는 것이 아니라 ‘교양 교육’을 의미하는 ‘liberal art’의 형태로 사용하고 있음을 파악할 수 있었다.

또한 그래프를 구성하는 132개의 키워드 중에서 약 60%에 해당하는 78개의 키워드의 매개 중심성 수치가 0으로, 예지가 하나인 것으로 나

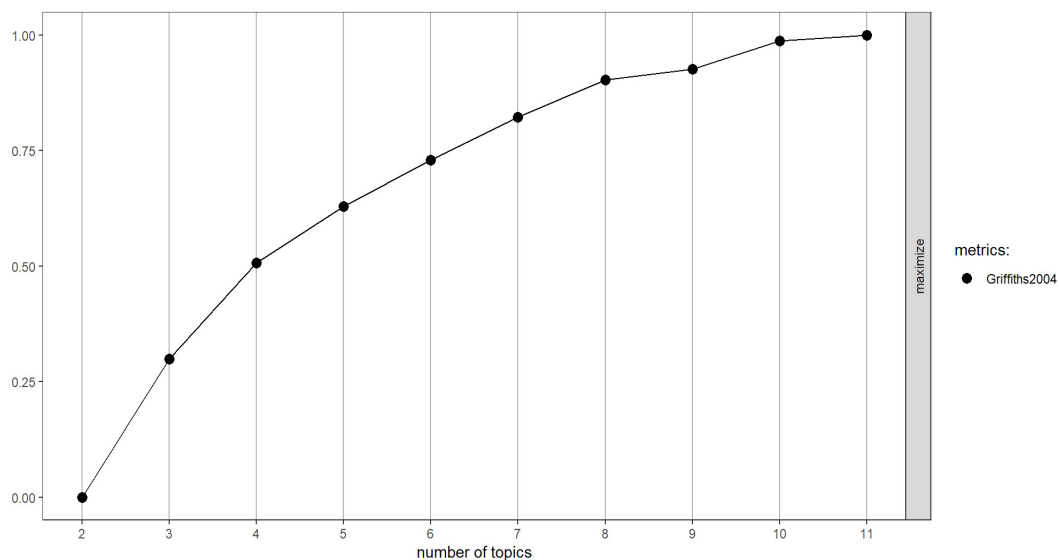
타났다. 이는 네트워크 그래프가 나머지 40%에 해당하는 소수의 키워드에 중심성이 집중된 집중적 구조를 띄고 있음을 의미한다. 또한 매개성이 낮은 키워드는 연결 중심성과 근접 중심성도 낮게 나타나, 현재까지 해당 분야의 연구 빈도가 낮았음을 알 수 있다. 따라서 이러한 키워드를 살펴보면 새로운 연구 분야를 탐색하는 데 유용할 것이다. 매개 중심성이 0으로 나타난 키워드는 ‘write’, ‘online’, ‘computer’, ‘information ethics’, ‘social’, ‘technology’, ‘literature’, ‘descriptive’, ‘perception’ 등이 있다. 이 키워드 중 ‘online’, ‘computer’, ‘technology’ 등 온라인 환경, 혹은 기술 관련 정보 리터러시와 관련된 단어들은

연구 대상 논문 수집 과정에서 ‘디지털 리터러시’, ‘ICT 리터러시’ 등을 다룬 논문을 인위적으로 제거하였기 때문에 이번 연구 대상에 포함되지 않아 중요성이 드러나지 않은 것으로 예상된다. 이외의 키워드를 살펴보면 ‘social’과 ‘write’에서 인문 사회학 분야의 정보 리터러시 관련 연구가 상대적으로 미비한 것으로 보이며, ‘literature’, ‘descriptive’ 등의 키워드를 보아 문헌 리뷰나 기술적 연구 등의 방법론은 상대적으로 적게 사용되는 것으로 보인다. 또한 대학생의 글쓰기와 정보 활용능력의 연관성, 정보 윤리와 리터러시의 연관성 또한 독립적으로 연구가 진행되고 있음을 파악할 수 있었다. 따라서 이러한 키워드를 종합적으로 고려하여, ‘인문학 글쓰기에 있어 정보 리터러시의 역할’, ‘온라인 정보 활용에서의 정보 윤리와 리터러시’ 등의 연구 가능한 주제를 도출할 수 있을 것이다.

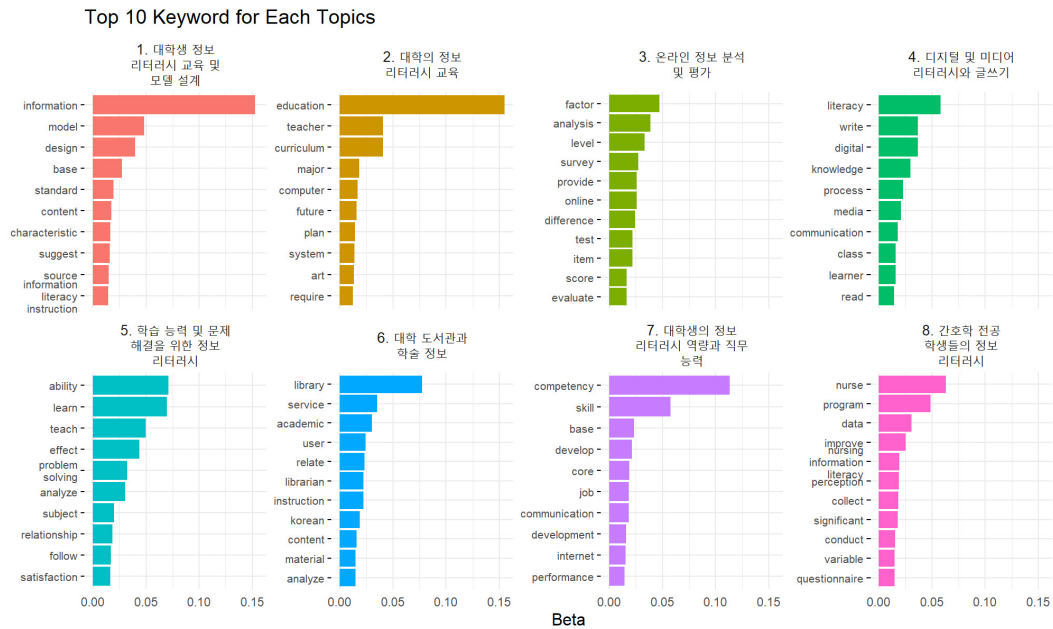
4.3 토픽 모델링

〈그림 3〉은 Griffiths와 Steyvers가 2004년 논문에서 최초로 활용한 토픽 모델링 기법인 Greffiths2004(Griffiths & Steyvers, 2004)를 활용하여 적합도를 평가하여 적정 토픽의 수를 산출한 과정을 나타낸 그래프이다. 이 모델은 단어의 동시 출현 관계를 식별하여 대규모 데이터를 분류하고 요약하는 데 강점이 있다(Lowrie et al., 2024). R의 topicmodels 패키지를 사용하여 2개에서 10개의 토픽을 생성하여 x축은 생성한 토픽의 개수, y축은 Greffiths2004 값을 0과 1 사이의 값으로 정규화한 수치인 그래프를 도출했다. y축의 점수 증가가 둔화되는 시점인 8개가 가장 적합하다고 판단했으며, 이 때 Greffiths2004 복잡도는 -78024.73이며, 정규화 수치는 0.914이다.

〈그림 4〉는 각 토픽에서의 베타 값 상위 10



〈그림 3〉 Greffiths2004 활용 토픽 수 산출



〈그림 4〉 토픽 별 주요 키워드

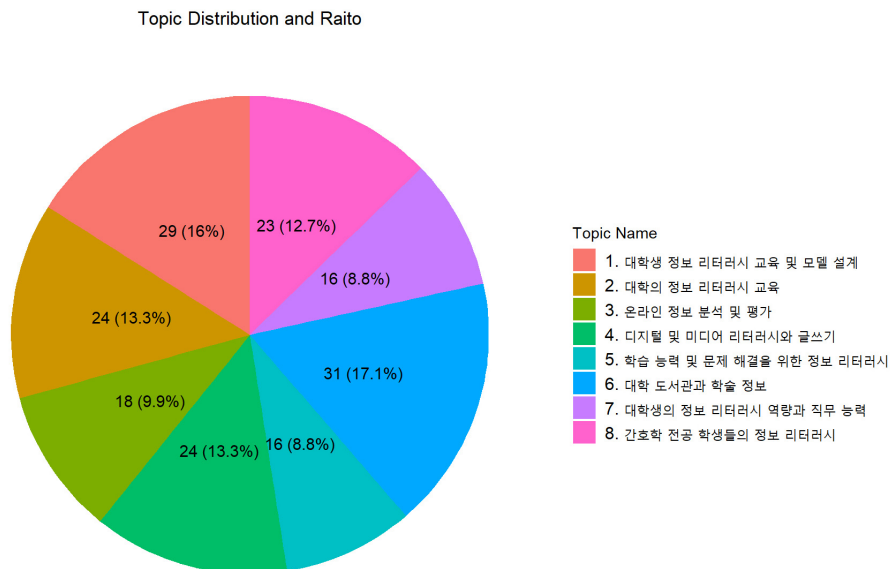
개 키워드와 각 토픽의 제목을 나타낸 그래프이며, 〈그림 5〉는 각 토픽에 해당하는 논문의 개수와 전체 논문 수에서 각 토픽이 해당하는 비율을 파이 그래프로 나타낸 결과이다.

첫 번째 토픽은 대학에서 정보 리터러시를 측정하거나, 관련 인식을 연구하거나, 평가 기준을 제시하여 그 효과를 분석하는 논문이 주로 속해 있다. 이 토픽의 주요 키워드는 ‘model’, ‘design’, ‘standard’, ‘suggest’ 등이 있으며, 총 29개의 논문이 이에 해당하여 전체의 약 16%를 차지한다. 이 토픽의 이름은 ‘대학생 정보 리터러시 교육 및 모델 설계’로 설정하였다.

두 번째 토픽은 ‘대학의 정보 리터러시 교육’이며, 약 13.3%에 해당하는 24개의 논문이 이 토픽에 해당된다. 주요 키워드는 ‘education’, ‘curriculum’, ‘major’, ‘teacher’ 등 대학의 교육 과정과 단어들 이 주로 나타났다. 이 토픽에 해

당하는 논문들은 대학 수업에서의 정보 리터러시 교육의 실용성, 효율적인 정보 리터러시 교육 운영 방안 등을 다루며, 특히 교육학 전공 학생을 대상으로 진행된 연구가 다수 포함되었다. 해당 토픽에서 나타난 ‘education’, ‘teacher’ 등의 키워드는 예비 교육자인 교육대학원 학생과 주제적 연관성이 나타난다.

세 번째 토픽은 ‘online’, ‘factor’, ‘analysis’ 온라인과 웹 기반의 정보를 수용, 활용, 판단하는 능력과 연관성이 있으며, 토픽의 이름은 ‘온라인 정보 분석 및 평가’이다. 이 토픽에 해당하는 논문은 전체의 약 9.9%며, 총 18건이다. 이 토픽은 상대적으로 다른 토픽에 비해 키워드들의 베타 값이 고르게 분포한다. 이 토픽에 해당하는 논문들은 소셜 네트워크, 온라인 튜토리얼, e-health 등 다양한 주제가 광범위하게 분포되어 있다.



〈그림 5〉 토픽 별 원문 분포 비율

네 번째 토픽은 ‘디지털 및 미디어 리터러시와 글쓰기’로, 정보 시각화, 개발, 정보 접근 역량 등과 관련되어있다. 이 토픽에 해당하는 논문의 수는 24건이었으며, ‘digital’, ‘media’, ‘communication’ 등 디지털 콘텐츠와 관련된 키워드와 ‘write’, ‘read’ 등 전통적 문해력 개념과 연관된 키워드가 동시에 나타난다. 이 토픽 또한 키워드의 베타 값이 비교적 고르게 분포하며, 미디어를 활용한 글쓰기 수업, 리터러시를 통한 문화 학습, 디지털 기술과 글쓰기의 협업 수업 등 복합적이고 학제적인 연구가 주로 분포한다.

‘학습 능력 및 문제 해결을 위한 정보 리터러시’는 다섯 번째 토픽으로, 가장 적은 16건의 논문이 이에 해당하여 전체의 8.8%를 차지했다. 이 토픽의 주요 키워드는 ‘ability’, ‘effect’, ‘problem solving’, ‘relationship’ 등 정보 리터러시와 대학생의 학업 성취도, 학업 능력, 교수

효과의 상관관계를 연구하였다. 문제 해결 능력, 자기주도적 학습 능력 등 학업 분야의 역량 및 성취도와 정보 리터러시의 연관성이 해당 토픽의 주요 관심사이다.

‘대학 도서관과 학술 정보’는 ‘library’, ‘service’, ‘academic’, ‘user’ 등을 주요 키워드로 하는 여섯 번째 토픽이다. 이 토픽에는 대학 도서관, 이용자 교육, 혹은 학술 정보 서비스, 학술 정보 검색과 관련된 31건의 논문이 분포한다. 이 토픽에 해당하는 논문들은 수업과 연계하거나, 혹은 대학 도서관이 독자적으로 정보 리터러시 프로그램을 기획하고 운영하는 방안을 제시하거나, 진행한 프로그램을 평가하는 내용을 다루며, 사서교사와 관련된 일부 논문 또한 여섯 번째 토픽에 포함되었다.

일곱 번째 토픽은 ‘대학생의 정보 리터러시 역량과 직무 능력’이며, ‘competency’, ‘skill’, ‘develop’, ‘job’ 등 정보 리터러시를 대학생의

핵심 역량 중 하나로 판단하여 실무 역량, 직업 기초 역량과의 상관관계를 연구하였다. 이 토픽에 해당하는 논문들은 외국인, 비서학전공생, 생산 관리 분야 등 특정 전공 분야에서 정보 리터러시와 해당 직업 성취도, 혹은 구직 행태의 상관관계를 연구하는 논문이 다수 분포했다. 이 토픽에 해당하는 연구는 총 16건이며, 전체의 8.8%이다.

여덟 번째 토픽은 전체 12.7%인 23개의 논문이 해당되는 '간호 및 건강 정보 리터러시'이다. 'nurse', 'nursing information literacy' 등 간호, 보건 관련된 키워드가 나타났다. 다양한 분야의 정보 리터러시 중에서도 간호 분야와 관련하여 활발하게 연구가 진행되고 있으며, 간호 정보의 경우 직업 환경에서 적응 능력과 밀접하게 연관된다는 점이 연구가 활발하게 진행되는 원인으로 유추할 수 있다. 간호 및 보건 분야 관련 연구는 다른 전공과 구별되어 독자적 토픽으로 나타날 정도로 다수의 연구가 활발히 진행되고 있으며, 다른 주제와 별개로 독립적 연구가 이루어지고 있음을 알 수 있다.

5. 논의 및 결론

본 연구는 다양한 학문 분야에 걸친 대학생 을 대상으로 하는 정보 리터러시 관련 연구를 텍스트 마이닝 기법 중 빈도분석, 네트워크 분석, 토픽 모델링을 사용하여 분석하였다. 빈출 키워드와 키워드 네트워크를 통해 연구 동향을 살펴보고, 하위 토픽을 통해 주요 연구 주제를 살펴본 결과 다음과 같은 결론을 도출했다.

첫째, 전체 텍스트에서 가장 등장 빈도가

높은 키워드는 'education'이며, 'information', 'competency', 'library'가 그 뒤를 이었다. 빈출 상위 키워드에서는 학업 관련 키워드의 출현 빈도가 높게 나타났다. 또한 학생의 역량 관련 키워드, 디지털 및 온라인 정보 관련 키워드, 간호 및 보건 의료 관련 키워드도 다수 관찰되었다.

둘째, 동시 출현 관계에 근거하여 키워드 네트워크 그래프를 생성하였으며, 각 노드의 연결 중심성, 근접 중심성, 매개 중심성을 계산하였다. 분석 결과 'education'과 'information'의 중심성이 가장 높아 네트워크에서 다른 키워드와 연관성이 높은 중심적 역할에 위치함을 파악하였으며, 키워드 네트워크는 소수의 노드에 연결성이 집중된 집중적 구조를 보이고 있다. 또한 중심성 수치가 낮은 키워드에 주목하여 'art'와 'liberal'이 고립된 하위 커뮤니티를 형성한 것을 확인하였으며, 상대적으로 연구가 적게 이루어진 키워드를 도출하여 향후 가능한 연구 주제를 제시하였다.

빈도 분석 결과와 네트워크 분석 결과를 비교하여 다음과 같은 사실을 발견하였다. 키워드의 등장 빈도와 중심성 값은 대체로 양의 상관관계가 나타났다. 그러나 'competency'와 'learn'은 빈출 순위가 각각 3위, 5위로 높으나 중심성 수치가 낮게 나타났다. 이외에도 'library', 'nurse', 'literacy' 키워드 또한 빈출 순위에 비해 중심성 수치가 낮게 나타났다. 이러한 키워드는 특정 하위 주제를 다루는 맥락의 연구에서 자주 등장하지만 키워드 네트워크에서의 역할은 미약하다. 즉, 이러한 키워드 관련 주제는 독립적으로 연구가 진행되어 왔음을 시사한다. 반면 'analyze', 'analysis', 'develop', 'program', 'data' 등의 중심성이 높으나 빈도가 낮게 나타났다. 이러한

키워드는 다양한 맥락에서 사용되기 때문에 다른 키워드들과의 연결성이 높게 나타나 연구의 맥락이나 흐름을 파악하기에 유용하다. 이러한 키워드에는 주로 연구 방법론이나 검증 방법에 관련된 키워드가 주로 분포한다. 해당 키워드를 살펴보면 많은 설문 조사를 통계적으로 분석하는 연구가 다수 진행되었음을 짐작할 수 있다. 그러므로 다른 집단을 대상으로 하는 연구에서 각 설문 도구의 변화를 살펴보는 연구도 유의미한 후속 연구가 될 수 있다.

마지막으로 토픽모델링으로 8개의 토픽을 도출하고 연구 대상 논문들의 분포 비율을 관찰하였다. 6번 토픽 '대학 도서관과 학술 정보'에 가장 많은 31개의 논문이 분포하였으며, 가장 적은 연구가 진행된 토픽은 16개의 논문이 분포하여 8.8%를 차지하는 '학습 능력 및 문제 해결을 위한 정보 리터러시'와 '대학생의 정보 리터러시 역량과 직무 능력'이다. 상대적으로 해당하는 논문의 수가 적은 대학생의 능력 및 역량 관련 연구에서 향후 후속 연구가 활발히 이루어질 수 있을 것이며, 정보 리터러시와 학

업 연관성, 직무 능력의 상관관계를 복합적으로 분석하는 연구도 기대할 수 있을 것이다.

본 연구는 특정 분야에 한정하지 않고 대학생의 정보 리터러시 관련 연구를 학제적으로 검토하였다는 점에서 그 의의가 있다. 문헌정보학 외에도 간호학, 교육학 등의 분야에서 리터러시 관련 연구가 활발하게 이루어지고 있으나, 그 외 전공에서는 정보 리터러시가 미비하다는 점에 주목하여 다른 전공 분야와 관련하여 후속 연구 주제를 제시하였다.

그러나 디지털, 미디어 특정 분야의 리터러시를 정보 리터러시와 구분하는 경우 해당 논문을 분석 대상에서 의도적으로 제거했기 때문에 전반적인 리터러시 연구 동향이 명확하게 드러나지 않았다. 또한 시계열 연구를 진행하지 않아 연구 동향의 변화를 명확하게 살피지 못했다는 점에서 한계가 있다. 그러므로 향후 거시적인 시점에서 대학생 대상 리터러시 연구를 살펴 전공별 차이, 혹은 주요 키워드와 토픽의 변화를 밝히는 유의미한 후속 연구가 진행되기를 기대한다.

참 고 문 헌

- 김기옥 (2020). 텍스트마이닝을 활용한 소비자학 연구 동향 분석. 소비자학연구, 31(5), 19-27.
<http://doi.org/10.35736/JCS.31.5.2>
- 김태훈 (2010). 대학생의 정보 활용능력 평가에 관한 연구: 신학대학교 재학생을 중심으로. 석사학위논문, 충남대학교.
- 박자현, 송민 (2013). 토픽모델링을 활용한 국내 문헌정보학 연구동향 분석. 정보관리학회지, 30(1), 7-32. <http://doi.org/10.3743/KOSIM.2013.30.1.007>
- 박주현 (2018). 독서 · 정보 · ICT · 디지털 리터러시의 개념화 모델 개발 연구. 한국도서관 · 정보학회지, 49(2), 267-300. <http://doi.org/10.16981/kliss.49.2.201806.267>

- 배규용, 박주현, 김정선, 이영섭 (2013). 텍스트 마이닝 기법을 활용한 기후변화관련 식품분야 논문초록 분석. *한국데이터정보과학회지*, 24(6), 1429-1437. <http://doi.org/10.7465/JKDI.2013.24.6.1429>
- 신승용 (2021). 지식정보화사회에서 SNS의 사회적 기능 향상에 따른 부작용 해결방안을 위한 멀티미디어-리터러시 교육 법제화의 당위성에 관한 연구(Ajzen의 계획된 행위 이론을 기반으로). *문화기술의 융합*, 7(3), 89-94. <https://doi.org/10.17703/JCCT.2021.7.3.89>
- 오의경, 장혜란 (2005). 대학생의 정보리터러시에 영향을 미치는 요인 분석. *한국문헌정보학회지*, 39(4), 207-231. <http://doi.org/10.4275/KSLIS.2005.39.4.207>
- 유은지, 김정철, 이춘열, 김남규 (2012). 시맨틱 텍스트 마이닝을 위한 온톨로지 활용 방안. *정보시스템연구*, 21(3), 137-161. <http://doi.org/10.5859/KAIS.2012.21.3.137>
- 윤지혜, 정유경 (2022). 토픽모델링을 활용한 정보활용교육 연구주제 분석 및 교육내용 제안. *정보관리학회지*, 39(4), 1-21. <http://doi.org/10.3743/KOSIM.2022.39.4.001>
- 이수상 (2007). 우리나라 대학생의 정보 리터러시 수준에 대한 실태조사. *한국문헌정보학회지*, 41(1), 85-103. <https://doi.org/10.4275/KSLIS.2007.41.1.085>
- 이수상 (2014). 네트워크 분석 방법론. 서울: 논형.
- 이수상 (2018). 네트워크 분석방법의 활용과 한계. 서울: 청람.
- 이창봉, 윤영, 한승규 (2021). 토픽모델링과 네트워크 분석을 활용한 리터러시 연구의 동향. *리터러시연구*, 12(6), 121-163. <https://doi.org/10.37736/KJLR.2021.12.12.6.05>
- 이혜영, 남태우 (2008). 대학생의 정보 활용능력 척도 개발 연구. *한국문헌정보학회지*, 42(4), 503-532. <https://doi.org/10.4275/KSLIS.2005.39.4.207>
- 임정훈 (2022). 키워드 네트워크 분석과 토픽모델링을 활용한 정보활용교육 연구 동향 분석. *정보관리학회지*, 39(4), 23-48. <http://doi.org/10.3743/KOSIM.2022.39.4.023>
- 장수현, 남영준 (2022). 문헌정보학 분야의 리터러시 연구 동향 분석. *정보관리학회지*, 39(3), 263-292. <http://doi.org/10.3743/KOSIM.2022.39.3.263>
- 조민정, 구미옥 (2021). 간호대학생을 위한 간호정보활용역량 교육프로그램 개발 및 효과 검증. *한국간호교육학회지*, 27(2), 210-222. <http://doi.org/10.5977/jkasne.2021.27.2.210>
- 한윤수, 류현숙 (2023). 공공분야 챗GPT 활용에 따른 학습 데이터 오염과 “할루시네이션(환각)” 개선 방안(Issue Paper 124). *한국행정연구원*, 1-10.
- Association of College and Research Libraries (2000). *Information Literacy Competency Standards for Higher Education*. Chicago: American Library Association.
- Association of College and Research Libraries (2016). *Framework for Information Literacy for Higher Education*. Chicago: American Library Association. <http://www.ala.org/acrl/standards/ilframework>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine*

- Learning Research, 3, 993-1022.
- Doskatsch, I. (2002). Australian and New Zealand Institute for information literacy (ANZIIL): new body to champion information literacy. Australian Academic & Research Libraries, 33(2), 113-116. <https://doi.org/10.1080/00048623.2002.10755187>
- Freeman, L. C. (1978). Centrality in social networks conceptual clarification. Social Networks, 1(3), 215-239. [http://doi.org/10.1016/0378-8733\(78\)90021-7](http://doi.org/10.1016/0378-8733(78)90021-7)
- Griffiths, T. L. & Steyvers, M. (2004). Finding scientific topics. Proceedings of the National Academy of Sciences, 101, 5228-5235. <https://doi.org/10.1073/pnas.0307752101>
- Hasal, M., Nowakova, J., & Platos, J. (2017). Three-dimensional graph drawing by Kamada-Kawai method with Barzilai-Borwein method. In 2017 IEEE Symposium Series on Computational Intelligence (SSCI), 1-7. <http://doi.org/10.1109/SSCI.2017.8285432>
- Koschützki, D. & Schreiber, F. (2008). Centrality analysis methods for biological networks and their application to gene regulatory networks. Gene Regulation and Systems Biology, 2008(2). <https://doi.org/10.4137/grsb.s702>
- Lau, J. (2006). Guidelines on Information Literacy for Lifelong Learning. International Federation of Library Associations and Institutions.
- Liu, L., Wang, S., Chen, J., Li, H., Gui, J., & Zhang, R. (2024). Information literacy of higher vocational college students in digital age. Journal of Librarianship and Information Science. <http://doi.org/10.1177/09610006231224445>
- Lowri, W., Eirini, A., Laura, A., & Pete, B. (2024). Topic modelling: going beyond token outputs. Big Data and Cognitive Computing. <http://doi.org/10.3390/bdcc8050044>
- Oxford English Dictionary (2023). Literacy. In Oxford English Dictionary. <https://doi.org/10.1093/OED/6268494087>
- Zafer, E. & Hakan, A. (2012). Nonlinear transformation of term frequencies for term weighting in text categorization. Engineering Applications of Artificial Intelligence, 25(7), 1505-1514. <https://doi.org/10.1016/j.engappai.2012.06.013>
- Zurkowski, P. G. (1974). The Information Service Environment: Relations and Priorities. National Commission on Libraries and Information Science. Related Paper 5.

• 국문 참고문헌에 대한 영문 표기

(English translation of references written in Korean)

Bae, Kyu Yong, Park, Ju-Hyun, Kim, Jeong Seon, & Lee, Yung-Seop (2013). Analysis of the

- abstracts of research articles in food related to climate change using a text-mining algorithm. Journal of the Korean Data & Information Science Society, 24(6), 1429-1437.
<http://doi.org/10.7465/JKDI.2013.24.6.1429>
- Han, Yoon Su & Lyu, Hyeon-Suk (2023). Contamination of Learning Data and Improvement Measures for “Hallucination” in the Use of ChatGPT in the Public Sector (Issue Paper, 124). Korea Institute of Public Administration, 1-10.
- Jang, Su Hyun & Nam, Young Joon (2022). A study on the research trends on literacy in library and information science. Korean Society for Information Management, 39(3), 263-292.
- Jo, Mean Jung & Gu, Mee Ock (2021). Development and effects of a nursing information literacy competency education program for undergraduate nursing students. The Journal of Korean Academic Society of Nursing Education, 27(2), 210-222.
- Kim, Kee Ok (2020). Analysis of research trends in consumer science through text mining. Journal of Consumer Studies, 31(5), 19-27. <http://doi.org/10.35736/JCS.31.5.2>
- Kim, Tae Hoon (2010). A Study on the Evaluation of Information Literacy of University Students: Focusing on the Undergraduate Students of C Theological University. Master’s thesis, Chungnam National University.
- Lee, Changbong, Yun, Young, & Han, Seungkyu (2021). Trends in literacy research using topic modeling and network analysis. The Korean Journal of Literacy Research, 12(6), 121-163.
<https://doi.org/10.37736/KJLR.2021.12.12.6.05>
- Lee, Soo-Sang (2007). A survey research on information literacy level in Korean university students. Journal of the Korean Society for Library and Information Science, 41(1), 85-103.
<https://doi.org/10.4275/KSLIS.2007.41.1.085>
- Lee, Soo-Sang (2014). Network Analysis Methods. Seoul: NonHyung.
- Lee, Soo-Sang (2018). Network Analysis Methods Applications and Limitations. Seoul: Cheungram.
- Lim, Jeonghoon (2022). Analysis of research trends in information literacy education using keyword network analysis and topic modeling. Korean Society for Information Management, 39(4), 23-48. <http://doi.org/10.3743/KOSIM.2022.39.4.023>
- Oh, Eui-Kyung & Chang, Hye-Rhan (2006). Information literacy: Identification of factors affecting undergraduate students. Journal of the Korean Society for Library and Information Science, 39(4), 207-231.
- Park, Ja Hyun & Song, Min (2013). A study on the research trends in library & information science in Korea using topic modeling. Journal of the Korean Society for Information Management, 30(1), 7-32. <http://doi.org/10.3743/KOSIM.2013.30.1.007>

- Park, Juhyeon (2018). A study on the development of conceptualization model for reading, information, ICT, and digital Literacy. *Journal of Korean Library and Information Science Society*, 49(2), 267-300. <http://doi.org/10.16981/kliss.49.2.201806.267>
- Rhee, Hey Young & Nam, TaeWoo (2008). The study on undergraduates' information literacy scale development. *Journal of the Korean Society for Library and Information Science*, 42(4), 503-532. <https://doi.org/10.4275/KSLIS.2005.39.4.207>
- Shin, Seungyong (2021). Study on justification of the legislation of multimedia-literacy education to solve side effects of improving social functions of SNS in the knowledge information society (based on Ajzen's theory of planned behavior). *The Journal of the Convergence on Culture Technology*, 7(3), 89-94. <https://doi.org/10.17703/JCCT.2021.7.3.89>
- Yu, Eunji, Kim, Jung Chul, Lee, Coon Yeui, & Kim, Namgyu (2012). Using ontologies for semantictext mining. *The Journal of Information Systems*, 21(3), 137-161. <http://doi.org/10.5859/KAIS.2012.21.3.137>
- Yun, Jihye & Jeong, Yoo Kyung (2022). A study of information literacy curriculum using topic modeling. *Journal of the Korean Society for Information Management*, 39(4), 1-21. <http://doi.org/10.3743/KOSIM.2022.39.4.001>