OPEN ACCESS

텍스트 마이닝을 활용한 대학도서관과 공공도서관의 장서개발 정책 키워드 분석

A Keyword Analysis of Collection Development Policies of University and Public Libraries Using Text Mining

이 다 현 (Da-Hyeon Lee)*
신 동 희 (Dong-Hee Shin)**

목 차

1. 서 론
2. 선행 연구
3. 연구방법
4. 연구 결과
5. 결 론

초 록

장서개발 정책은 도서관 서비스의 근간이며, 도서관의 성격과 정체성이 드러나는 성문화된 문서이다. 본 연구에서는 국립중앙도서관에서 원문 추출이 가능한 장서개발 정책 관련 원문 11개를 대상으로 빈도분석, 토픽 모델링, 네트워크 분석을 수행하였다. 장서개발 정책과 관련된 주요 키워드를 추출하고, 해당 키워드 간의 관계를 분석하였다. 그 후 대학도서관과 공공도서관의 관종별 장서개발 정책의 특징을 도출하기 위해 파이 계수 분석을 수행하였다. 분석 결과 '자료', '도서관', '장서개발', '이용자', '수집' 등의 키워드가 빈도 분석과 네트워크 중심성에서 주요 키워드로 나타났으며, 파이 계수 분석 결과 대학도서관은 '대학(0.73)', '구축', '학생', '대상', '비용' 등의 키워드가 나타나므로, 이용자의 학술적 요구와 디지털 자원에 대한 논의가, 공공도서관은 '성인', '설문조사', '특징', '종교' 등 다양한 이용자계층의 정보 요구와 관련된 키워드가 주요 쟁점으로 드러났다.

ABSTRACT

For this article, we conducted frequency analysis, topic modeling, and network analysis on eleven texts related to collection development policy found in the National Library of Korea. We deduced the main keywords related to collection development policies and analyzed the relationship between them. We subsequently conducted a pie coefficient analysis to identify the characteristics of collection development policies of university libraries and public libraries by category. The results showed that keywords such as "material," "library," "collection development," "user," and "collection" were the main keywords in frequency analysis and network centrality. Meanwhile, the pie coefficient analysis revealed that keywords such as "university," "construction," "student," "target," and "cost" were prevalent in university libraries, indicating that the academic needs of users and the discussion of digital resources were primary issues, while keywords related to the information needs of various user groups－including "adults," "survey," "feature," and "religion" －appeared in public libraries.

키워드: 장서개발정책, 텍스트 마이닝, 빈도 분석, 토픽 모델링, 네트워크 분석
Collection Development Policy, Text Mining, Frequency Analysis, Topic Modeling, Network Analysis

---

* 숙명여자대학교 문헌정보학과 석사과정(2331672@sookmyung.ac.kr / ISNI 0000 0005 1377 6147) (제1저자)
** 숙명여자대학교 문헌정보학과 조교수(shindh@sookmyung.ac.kr / ISNI 0000 0004 9239 1001) (교신저자)
논문접수일자: 2024년 1월 22일　최초심사일자: 2024년 2월 2일　게재확정일자: 2024년 2월 14일
한국문헌정보학회지, 58(1): 285-302, 2024. http://dx.doi.org/10.4275/KSLIS.2024.58.1.285

※ Copyright © 2024 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. 서 론

장서개발은 도서관 장서를 입수 및 구축하기 위한 활동을 의미하며, 이용자 요구분석, 장서개발지침 수립, 장서선정, 입수, 장서관리 및 폐기, 장서평가 등의 과정으로 구성되어있다(문헌정보학용어사전, 2020). 즉, 장서개발정책은 자료의 선택에서 제적까지를 계획, 실천하는 도서관의 핵심 활동이라고 할 수 있다, 이를 성문화한 공식 문서를 장서개발 정책은 이를 위한 정책 문서이다(이진경, 이지연, 2011). 따라서 장서개발 정책은 도서관 서비스의 전 과정과 연관되어 도서관의 자원을 효율적으로 활용하기 위해 소장할 장서의 종류와 수량을 결정하는 기준이 된다.

현재까지 국내에서 수행된 장서개발 정책과 관련된 연구는 '공공 도서관 장서관리매뉴얼(국립중앙도서관, 2010)', '대학도서관의 참고 장서개발 정책 기본요소 개발에 관한 연구(최재황, 2013)', '전문도서관의 전자자료 장서개발 정책에 관한 연구(박희숙, 2004)'와 같이 개별 기관의 장서개발정책을 제시하는 방향으로 주로 진행되어왔으며, 실무 현장에서의 자료를 분석한 연구는 미비하다. 그러므로 본 연구에서는 교육자료, 매뉴얼 등 실무 현장에서 사용되는 문서를 대상으로 연구를 진행하였다.

또한 단순히 장서개발정책과 관련된 키워드만을 선정하는 것에서 더 나아가 관종별 장서개발정책의 주요 키워드를 살피고자 한다. 이를 위해 장서개발정책 분야 연구 동향을 연구한 신유미(2019)의 논문에 근거하여 가장 많은 연구가 수행되고 있는 관종인 대학도서관과 공공도서관의 장서개발 정책을 연구 대상으로 선정하였다.

대학도서관과 공공도서관은 주 이용자와 소장 자원, 운영 목적에 차이가 있다. 대학도서관은 학술적 목적에서 연구에 중점을 주는 교수, 학생 등 연구자가 주된 이용 대상이며 이들의 연구를 보조하여 학문적 발전을 추구하는 것이 대학 도서관의 주요 목적이다. 반면 공공도서관은 지역 내의 다양한 연령층과 직업군의 이용자가 문화, 교육, 정보 등 다양한 목적으로 이용한다. 그러므로 공공도서관은 지역 주민의 다양한 요구를 수용할 필요가 있다(안문수, 2011). 이러한 주 이용 대상과 운영목적의 차이가 각 도서관의 특성을 규정하며, 이러한 특성이 각 도서관의 장서개발 정책에 반영되어 있을 것으로 예상한다.

이 연구의 목적은 텍스트 마이닝 기법을 활용하여 대학도서관과 공공도서관의 장서개발 정책 관련 데이터를 분석하는 것이다. 텍스트 마이닝은 다량의 데이터를 빠르게 처리하고 핵심 키워드와 키워드 간 패턴을 식별하는 데 적합하다(Arshad et al., 2020). 이러한 특성을 바탕으로 장서개발 정책 분야 실무 자료의 텍스트 데이터를 분석하여 키워드의 빈도분석을 진행하고, 토픽모델링을 활용하여 장서개발 정책이 활용되고 있는 세부 분야를 살펴보고자 한다. 마지막으로 각 키워드 간 네트워크 분석을 진행하여 장서개발 정책과 관련된 키워드간의 연관성을 살피고, 파이 계수를 활용하여 대학도서관, 공공도서관과 관련된 장서개발 정책 관련 키워드를 분석하여 관련 쟁점의 유사점과 차이점을 명확히 하고자 한다. 이러한 목적을 달성하기 위해, 본 연구는 다음과 같은 연구 질문을 설정하였다.

- 질문 1) 장서개발 정책의 주요 주제 및 토픽은 무엇인가?
- 질문 2) 대학도서관과 공공도서관의 장서개발 정책과 관련된 주요 키워드의 유사점과 차이점은 무엇인가?
- 질문 3) 키워드 간의 관계와 네트워크는 어떠한 양상을 나타내는가?

본 연구는 대학도서관과 공공도서관의 장서개발 정책에 대한 이해를 높이고, 분석 결과를 토대로 기존의 장서개발 정책을 검토하거나 향후 장서개발 정책 수립 및 개선 과정에 활용할 수 있는 중요한 통찰력을 제공할 것으로 기대된다. 또한, 이 연구는 텍스트 마이닝 기법을 적용한 도서관 및 정보학 분야 연구에 이바지할 것이다.

2. 선행 연구

본 연구를 위한 선행연구로, 문헌정보학과 장서개발 관련 분야에서 텍스트 분석을 활용한 연구를 검토하였다.

장서개발이란 "도서관 장서를 개발하는 것과 관련된 제반 활동"(문헌정보학 용어사전, 2000)을 의미한다. 유의어로는 '장서구축', '장서구성'이 있다. '장서관리'라는 용어와 혼재되기도 하나 장서개발은 주로 입수 및 구축에, 장서관리는 운영적 측면에 중점을 둔다. 『장서관리론』(윤희윤, 2020)은 장서관리의 기초이론과 정보매체의 역사, 도서관 자료의 종류와 특성을 논하고, 자료의 개발 방안, 장서보존의 이해와 전략, 장서평가, 폐기 등 장서관리의 기본적 개념에서 실무의 내용까지 포괄적으로 기술되어 있다.

2.1 장서개발 정책 관련 선행연구

먼저 각 도서관의 장서개발 정책을 분석하여 중심 키워드를 연구한 논문은 다음과 같다.

윤희윤, 김종애, 오선경(2020)은 선진국 대도시 공공도서관과 국내의 지역대표도서관 장서개발 정책을 분석하였다. 이 연구에서는 '자료수집력 강화, 장서개발 충실화, 장서구성의 편향성 최소화, 장서보존 최적화' 등 지역 대표 도서관의 장서개발 정책에 필요한 핵심 원칙을 세우는 데 기여하였다. 이 원칙은 '서문, 장서개발의 주체와 방법, 장서개발 기본원칙, 수집방법별 개발지침, 자료유형(매체)별 개발지침, 주제별 개발지침, 취약계층별 개발지침, 연계・협력지침, 보존관리 및 공동보존서고 지침, 정책문서 관리주체 및 개정주기'의 10개 영역의 구성체계와 이에 근거한 3개의 세부지침인 '아날로그 장서개발, 디지털 장서개발, 장서 보존관리'로 구성되었다.

이진경, 이지연(2011)은 서울 내의 공공도서관의 이용자 만족도와 장서 이용 행태를 분석하여 공통적인 장서개발 정책 요소를 도출하였으며, 이를 바탕으로 공공도서관 장서개발 정책의 방향성을 제시하였다. 장서개발 정책에 포함되는 주요 키워드로는 '선정', '수집', '구성', '개발', '관리'가 있다.

한국사립대학교 도서관 협의회(2014)는 사립대학도서관 협회 회원 기관의 웹사이트의 자료선정 기준을 분석하여 미국 도서관 협회와 류인석의 이론을 적용하였다. 그 결과 장서개발 정책이 '자료의 선정과 관한 사항, 장서개발위원회,

기증 및 교환, 기타사항, 상호협력, 예산, 정책의 개정'을 공통으로 포함하고 있음을 확인했다.

2.2 텍스트 마이닝 적용 선행연구

텍스트 마이닝이란 "자연어 텍스트 기반 문헌 클러스터링, 문헌분류, 온톨로지 학습을 통하여 하나의 패턴이나 경향을 파악하는 기법"(문헌정보학 용어사전, 2023)을 의미한다. 문헌정보학 분야에서 텍스트 마이닝 기법을 활용한 연구는 다음과 같다.

박자현, 송민(2013)은 LDA(잠재 디리클레 할당)기반 토픽모델링 방법을 활용하여 국내 문헌정보학 연구동향을 분석하였다. 국내 주요 문헌정보학 분야 학술지에 수록된 논문을 대상으로 토픽 모델링을 수행하여 결과를 문헌정보학 주제분류표와 비교하고, 동시 출현 단어의 빈도를 기준으로 학술지별 주제 키워드를 연결하였다.

최형욱, 최예진, 남소연(2018)은 KCI에 등재된 인용지수가 높은 문헌정보학 분야 3개 학술지의 저자 키워드를 선정하여 동시 출현 관계를 시계열 분석하였다. 저자 키워드의 네트워크 분석 이후 삼각매개중심성(rTBC)과 상대적 최근접이웃중심성(rNNC) 분석을 수행하여 키워드 및 세부 연구 주제를 파악하였다.

신유미, 박옥남(2019)은 2003년부터 2017년까지의 장서개발관리 분야 논문의 저자 키워드를 분석하였다. 시계열 분석, 동시 출현 단어 네트워크 분석, 연결중심성 분석 및 매개중심성 분석을 수행하여 중심 연구 키워드로 '오픈 액세스', '기관 레포지터리', '장서개발', '학술지'를 선정하였다.

Illescas, L.과 Sucozhanay D.(2018)는 에콰도르 Cuenca 도서관의 과학 분야 장서 개발에 텍스트 마이닝 방법론을 적용하였다. 2012년부터 2017년까지 Cuenca 도서관의 과학 분야 영어 논문의 출판과 인용 현황을 데이터 웨어하우징, 데이터 마이닝, 서지학적 연구기법을 결합한 비블리오마이닝 방법론을 적용했다. 전처리한 텍스트를 대상으로 빈도분석과 네트워크 분석을 진행한 결과, 건강, 공학, 심리학 분야의 연구가 가장 활발히 진행되고 있음을 알 수 있다.

3. 연구방법

3.1 연구 대상 선정

본 연구는 대학도서관과 공공도서관의 장서개발 정책의 주요 키워드를 분석하기 위하여 관련 주제의 연구 보고서를 대상으로 텍스트 마이닝을 수행하였다. 먼저 연구 대상으로 국립중앙도서관에서 '장서개발'을 검색어로 입력한 경우 반환되는 소장 자료 중 원문에서 텍스트를 추출할 수 있는 소장 자료 11건을 대상으로 하였다. 선별 데이터 전체를 대상으로 장서개발 정책에 관련된 키워드를 추출하기 위해 빈도분석을 수행하여 결과를 막대그래프와 워드 클라우드로 나타내었다. 이후 토픽 모델링을 진행하여 각 원문을 가장 주제와 밀접한 토픽에 분류하였다. 이후 선정된 키워드간의 관계를 네트워크 분석으로 시각화하였으며, 파이계수를 활용하여 '대학도서관', '공공도서관'과 관련된 키워드를 선정하였다.

연구 대상 중 일부 자료는 하나의 원문 내에

서 각 목차마다 개별적인 내용을 다루고 있다. 이 경우 해당 원문 전체를 하나의 토픽으로 분류하기에는 적합하지 않으므로, 목차별로 내용이 구별되는 1번『도서관장서개발과정』, 4번『장서개발(B)』, 5번『장서개발(A)과정』은 각 목차별로 원문을 분리하여 각각 토픽모델링을 진행하였다(〈표 1〉 참조).

3.2 분석 방법

3.2.1 텍스트 전처리

텍스트 전처리는 텍스트 원문에서 불필요한 요소를 제거하여 유의미한 데이터만을 정제하는 과정이다.

국립중앙도서관의 원문 데이터를 텍스트로 변환하는 과정에서 발생하는 인식 오류를 불용어로 처리하였다. 변환된 원문은 특수문자 제거, 형태소 분석, 표현형 통일 과정의 순으로 전처리를 시행했다. 한글과 로마자 외의 문자를 제거하고 Simplepos09 함수를 사용하여 명사, 동사, 형용사, 외국어만을 추출했다. 이 중 동사와 형용사는 활용형에 따라 다른 키워드로 분산되는 것을 방지하기 위해 기본형인 '-다' 형태로 표현형을 통합하였다.

또한 동일한 의미를 갖는 다양한 표현을 하나의 키워드로 통합하였다. 예를 들어 '장서 개발'이라는 유의미한 구의 띄어쓰기를 제거하여 '장서개발'과 하나의 키워드로 처리했다. 또한 문헌정보학 용어사전의 참조 관계를 바탕으로 동의어로 사용되는 '장서구성', '장서형성', '장서구축' 등도 '장서개발'로 통일하여 동일 의미를 가진 단어들을 하나의 키워드로 변환하였다.

〈표 1〉 국문제목 분석 대상 OCR 데이터 원문

NO	오프라인제어번호	서명	저자명	페이지
1	KMO200327650	도서관장서개발과정	국립중앙도서관	127 p.
2	KMO200848559	공공도서관의 장서개발 현 주소와 발전방향 모색	경기도	61 p.
3	KMO201033394	공공도서관 장서관리매뉴얼	국립중앙도서관 도서관연구소	248 p.
4	KMO201337278	장서개발(B)	국립중앙도서관 사서교육문화과	240 p.
5	KMO201411357	장서개발(A)과정	국립중앙도서관	200 p.
6	KMO201439077	경기도 대표도서관 장서개발정책 및 운영방안 연구 최종보고서	연세대학교 산학협력단	xviii, 259 p.
7	KMO201471990	도서관 장서개발 연구	충남대학교	ix, 23, 177 p.
8	KMO201755724	대학도서관 정책 및 발전계획 수립을 위한 컨설팅 과정	한국교육학술정보원	150 p.
9	KMO201759826	대학도서관 장서개발 정책 및 운영 전문가 과정	한국교육학술정보원	150 p.
10	KMO201766087	국립세종도서관 장서개발정책 수립 연구	대구대학교 산학협력단	xxv, 201 p.
11	KMO202268039	부산지역 도서관 활성화 계획 연구: 부산도서관 장서개발지침(분담수서 등) 수립 관련: 최종보고서	부산광역시 부산도서관	vi, 177 p.

3.2.2 단어 빈도 분석

빈도 분석은 키워드가 텍스트 내에서 얼마나 자주 사용되었는지를 측정하는 분석 방법이다. 전처리 과정을 거친 키워드가 각 원문에서 등장하는 횟수를 산출하고, 동일 키워드의 등장 회수를 누적하여 모든 원문에서 가장 등장 횟수가 많은 키워드의 목록을 구성했다. 빈도 분석의 결과는 막대그래프와 워드 클라우드로 시각화한다.

3.2.3 토픽 모델링

토픽 모델링은 문헌에서 숨겨져 있는 주제들을 찾아내기 위해 개발된 통계 추론 모델로(박자현, 송민, 2013) 단어와 문서의 관계를 활용하여 주제를 추론하고 문서를 토픽별로 분류하는 데 활용된다. 하이퍼 파라미터 튜닝으로 유사한 키워드를 군집하고 원문을 해석하기에 가장 적합한 토픽의 수를 결정했다. 이에 따라 토픽 모델링 기법 중 가장 대표적인 LDA(잠재 디리클레 할당) 알고리즘을사용하여 유관 키워드를 군집하여 적절한 토픽의 주제를 설정하였다. 토픽과 각 토픽별 주요 키워드는 그래프로 시각화했다. 도출된 토픽과 각 문서의 키워드 등장 빈도에 기반하여 각 원문과 가장 주제적 연관성이 높은 토픽을 매칭했다.

3.2.4 네트워크 분석

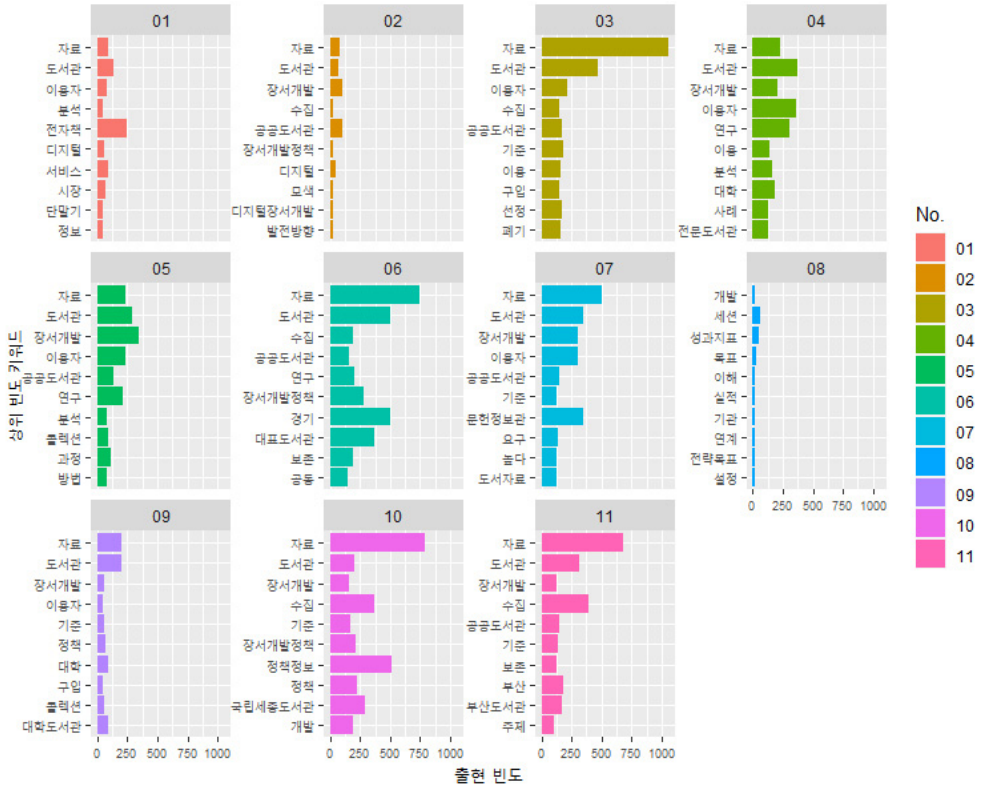
네트워크 분석이란 텍스트 내에서 중요하게 사용된 중심 키워드를 파악하고 키워드간의 연결성을 파악하는 데 유용한 분석 기법이다(Li Jiaen, 최영준, 2023). 네트워크 관계는 동시출현 관계를 기반으로 형성된다. 동시출현 빈도는 원문 내에서 두 개의 키워드가 함께 나타나는 회수를 의미하며, 동시 출현 빈도가 높은 키워드 간에는 의미적으로 연관이 있다고 판단된다. 또한 한 키워드가 동시출현관계에 있는 키워드의 수가 많을수록 연결 중심성이 높다고 표현하며, 이는 해당 분야에서 핵심적 키워드임을 의미한다(Freeman, 1979). 네트워크 분석의 결과는 네트워크 그래프의 형태로 시각화하며, 연결 중심성과 커뮤니티 관계를 파악할 수 있다.

4. 연구 결과

텍스트 마이닝 기법 중 빈도분석, 토픽 모델링, 네트워크 분석을 활용하여 연구를 진행했다. 첫째, 각 원문과 전체 원문을 대상으로 키워드의 빈도를 분석한 후 다수 키워드에서 빈출된 키워드와 누적 빈도수가 높은 키워드를 비교했다. 둘째, 토픽 모델링을 수행하여 토픽별 키워드에 근거하여 분석 대상 원문을 각 토픽에 주제적으로 연결 지었다. 마지막으로 네트워크 그래프와 파이 계수를 활용하여 도출된 키워드간의 관계를 살피고, 파이 계수로 대학도서관의 장서개발 정책, 공공도서관의 장서개발 정책과 특징적으로 관련이 높은 키워드를 도출하였다.

4.1 키워드 빈도분석

분석 대상 원문 11개의 텍스트를 분석하여 각 원문별로 빈출 키워드를 10개씩 추출하였다. 〈그림 1〉은 각 원문별로 도출된 키워드와 등장 횟수를 막대그래프로 나타낸 결과이다.



〈그림 1〉 원문 별 키워드 빈도 그래프

8번 원문『대학도서관 정책 및 발전계획 수립을 위한 컨설팅 과정』을 제외한 모든 원문에서 '자료'와 '도서관' 키워드의 등장 빈도가 첫 번째, 두 번째로 높았다. 또한 '장서개발'은 7개의 원문에서, '이용자', '공공도서관'은 6개의 원문에서, '기준', '수집'은 5개 원문에서 상위 빈출 10개 키워드에 포함되었다. 5개 이상의 원문에서 상위 빈출 키워드에 포함된 단어들을 제외하고 각 원문에서 나타나는 빈출 키워드를 살펴보면 다음과 같다.

1번 원문의 특징적인 빈출 키워드는 '분석', '전자책', '디지털'이 있으며, 2번 원문은 '디지털', '모색', '디지털장서개발', '발전방향' 등의 키워드가 관찰되었다.

3번 원문의 경우 '구입', '선정', '폐기' 등의 키워드가 나타나고 4번 원문의 주요 키워드는 '연구', '이용', '분석', '대학', '전문도서관'이 있으며, 5번 원문에서는 '콜렉션', '분석', '과정', '방법' 등의 키워드의 빈도가 높게 나타났다.

6번 원문에서는 '연구', '경기', '대표도서관', '보존', '공동'이라는 키워드가 등장한다. 7번 원문에서는 '문헌정보관', '요구', '높다' 키워드가 대표적이며, 8번 원문의 경우 '개발', '세션', '성과지표', '전략목표' 등 타 원문에서 주로 나타나지 않는 키워드들이 주로 관찰되었다.

9번 원문의 주요 키워드는 '대학', '대학도서

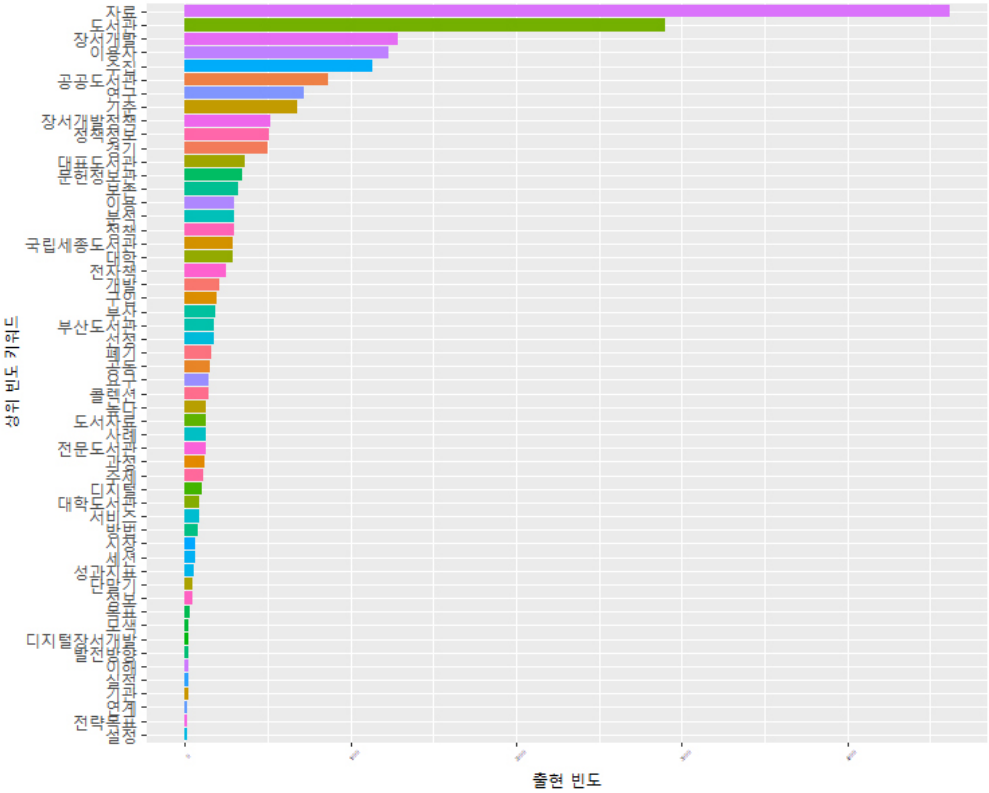
관', '콜렉션', '정책'이 있으며, 10번, 11번 원문에서는 각각 '국립세종도서관'과 '정책정보', '정책', 그리고 '보존', '부산', '부산도서관' 키워드가 두드러진다.

전체 텍스트에서 가장 많이 등장하는 키워드는 4,618번 등장한 '자료'였으며 그 다음으로 등장 빈도가 높은 키워드는 '도서관', '장서개발', '이용자', '수집', '공공도서관', '연구', '기준', '장서개발정책', '정책정보' 등이 있다. 5개 이상의 원문에서 공통적으로 나타나는 키워드인 '자료', '도서관', '장서개발', '이용자', '공공도서관', '기관', '수집'의 상위 7개 키워드는 등장 빈도순 상위 10개 키워드 내에 포함되었다.

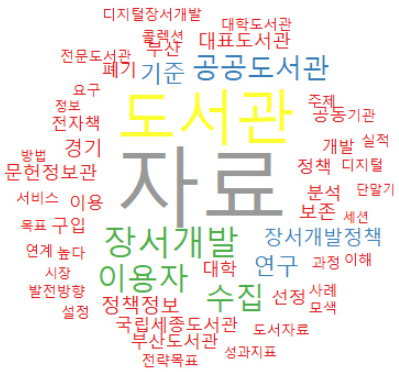
〈그림 2〉는 전체 원문의 빈출 키워드와 누적 출현 회수를 그래프로 나타낸 결과이며, 〈그림 3〉은 그래프를 워드 클라우드 기법을 활용하여 나타낸 결과이다. 워드 클라우드는 그래프 내에서 키워드의 크기와 색을 다르게 제시하여 키워드의 출현 빈도를 직관적으로 나타낸다(나지영, 2022).

4.2 토픽 모델링

분석 대상 원문에 잠재되어있는 주요 연구 주제를 찾기 위해 토픽모델링을 수행하였다. 각 토픽은 문서에서 추출된 키워드의 패턴을 통해 추론한다.



〈그림 2〉 누적 키워드 빈도 그래프



〈그림 3〉 누적 키워드 워드 클라우드

빈도분석에서 관측 빈도가 가장 높은 7개 단어 중 '공공도서관' 키워드는 관종에 따른 장서개발정책 차이를 살피고자 하는 이번 연구 주제에 필수적인 키워드이며, '장서개발' 키워드를 제외하는 경우 토픽의 주제가 명확하게 드러나지 않아 토픽모델링 키워드에 포함하였다. 그 외의 5개 키워드는 토픽의 특성을 명확하게 드러내지 못하므로 모델링 과정에서 제외하였다.

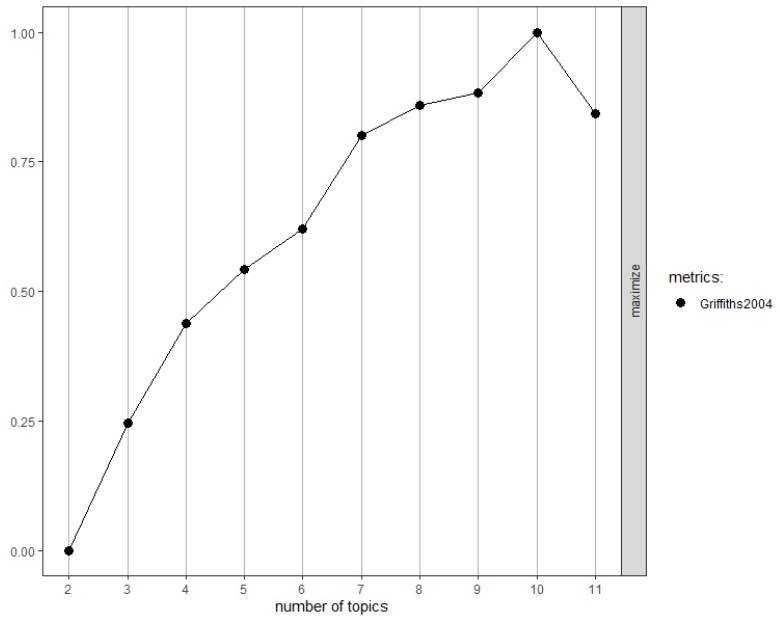
먼저 Gibbs 방식으로 하이퍼 파라미터 튜닝을 수행하여 2-11개의 토픽을 생성하였다. 그래프 상에서 성능 지표가 크게 향성하지 않으면서 주제를 가장 잘 나타낼 수 있도록 키워드 군집이 이루어진 토픽의 수를 7개로 설정하였다.

이후 각 토픽에서 할당 확률이 가장 높은 상위 10개의 키워드를 순서대로 추출하여 각 단어 간의 연계성을 고려하여 토픽의 이름을 선정하였다. 〈그림 4〉는 토픽과 토픽별 주요 빈도 키워드를 도출하여 빈도에 따라 그래프로 정렬한 결과이다. 〈표 2〉는 토픽 모델링의 결과로 도출된 7개의 토픽과 각 토픽 별 상위 할당 확률 단어, 각 토픽과 주제적 연관이 가장 높은 원문의 번호를 표로 나타내었다.

각 토픽의 이름과 주요 키워드, 선정된 원문을 살펴보면 다음과 같다. 1번 토픽은 '대학도서관', '대학', '교수' 등 대학도서관 관련 키워드와 '콜렉션', '전자저널', '단행본' 등의 자료 관련 키워드가 주로 나타났으므로 토픽 이름을 "대학 도서관의 자료 및 서비스 관리"로 설정하였다. 1번 원문 『도서관장서개발과정』의 일부 목차와 4번 원문 『장서개발(B)』의 일부, 8번 원문 『대학도서관 정책 및 발전계획 수립을 위한 컨설팅 과정』, 9번 원문 『대학도서관 장서개발 정책 및 운영 전문가 과정』이 이에 해당한다.

2번 토픽의 이름은 "공공도서관의 자료 선정과 관리 기준"으로 설정하였다. '공공도서관' 키워드 외에 '선정', '기증', '구입', '폐기' 등 장서관리 전반에 관련된 키워드들이 주로 나타나며, 3번 원문 『공공도서관 장서관리매뉴얼』과 4번 원문의 일부 목차, 5번 원문 『장서개발(A)』의 일부 목차가 이에 해당한다.

'문헌정보관', '공공도서관', '장서개발', '예산', '요구', '분석', '지역' 등의 키워드가 주로 나타난 3번 토픽의 이름은 "공공도서관 장서개발의 지



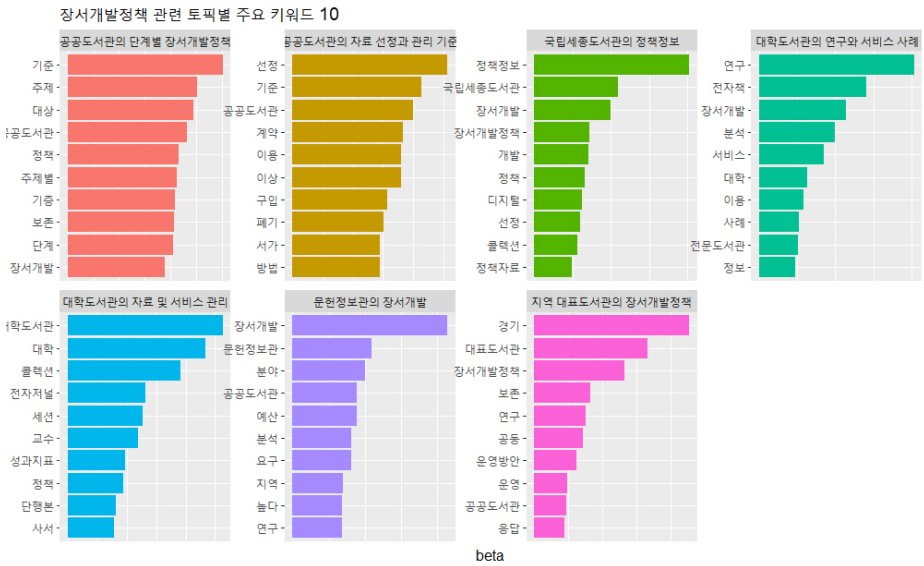
〈그림 4〉 토픽 수 산정

〈표 2〉 토픽 이름 및 주요 키워드

Topic No.	Topic Name	Term
Topic 1	대학도서관의 자료 및 서비스 관리	대학도서관, 대학, 콜렉션, 전자저널, 교수, 세션, 성과지표, 정책, 단행본, 사서
Topic 2	공공도서관의 자료 선정과 관리 기준	선정, 기준, 공공도서관, 계약, 이용, 이상, 구입, 폐기, 서가, 방법
Topic 3	공공도서관 장서개발의 지역적 요구	장서개발, 문헌정보관, 분야, 공공도서관, 예산, 분석, 요구, 지역, 높다, 연구
Topic 4	정책정보 특화 도서관으로서의 국립세종도서관	정책정보, 국립세종도서관, 개발, 정책, 장서개발, 장서개발정책, 디지털, 선정, 콜렉션, 정책자료
Topic 5	지역 대표도서관 장서개발정책	경기, 대표도서관, 장서개발정책, 보존, 연구, 공동, 운영방안, 운영, 응답, 공공도서관
Topic 6	공공도서관의 단계별 장서개발정책	기준, 주제, 대상, 공공도서관, 정책, 주제별, 기증, 보존, 단계, 장서개발
Topic 7	대학도서관의 연구와 서비스 사례	연구, 전자책, 장서개발, 분석, 서비스, 대학, 이용, 사례, 전문도서관, 정보

역적 요구"로 설정하였다. 5번 원문의 일부와 7번 원문『도서관 장서개발 연구』가 이 토픽과 연관성이 높다.

4번 토픽은 "정책정보 특화 도서관으로서의 국립세종도서관"으로 정했다. '국립세종도서관', '정책정보', '정책', '정책자료' 등의 키워드가 주로 나타나며 4번과 5번 원문의 일부와 10번 원문『국립세종도서관 장서개발정책 수립 연구』



〈그림 5〉 토픽 별 주요 키워드

를 해당 토픽에 할당하였다. 그러나 4-5 원문과 5-9 원문은 국립세종도서관과 주제적으로 관련성이 없으므로, 원문의 내용에 따라 적절한 토픽에 분류되지 않았다.

5번 토픽은 '공공도서관', '대표도서관' 등 도서관 관종과 관련된 키워드와 '보존', '연구', '운영방안' 등 장서개발정책 일반과 관련된 키워드가 나타나 "지역 대표 도서관의 장서개발정책"으로 명명했다. 6번 원문 『경기도 대표도서관 장서개발정책 및 운영방안 연구 최종보고서』가 이 토픽에 해당한다.

11번 원문 『부산지역 도서관 활성화 계획 연구: 부산도서관 장서개발지침(분담수서 등) 수립 관련: 최종보고서』는 토픽 6 "공공도서관의 단계별 장서개발정책"과 주제적 연관성이 높게 나타났다. 이 토픽에는 '주제', '주제별', '대상', '기증', '보존' 등의 키워드가 포함된다.

마지막 토픽인 7번에서는 '대학', '전문도서관', '연구', '서비스', '정보' 등이 나타난다. "대학도서관의 연구와 서비스 사례"에서는 1번, 4번, 5번 원문의 일부와 2번 『공공도서관의 장서개발 현 주소와 발전방향 모색』이 설정되었으므로, 이 토픽도 원문의 내용과 적절한 연관성을 갖추었다고 보기 어렵다.

4.3 네트워크 분석

네트워크 분석의 결과로 나타나는 네트워크 그래프는 키워드간의 상호관계가 시각적으로 드러나므로 추출한 키워드 간의 네트워크 관계를 살펴보고 원문의 주제를 나타내는 키워드를 파악하기에 적합한 연구 방법이다. 전처리 과정을 거쳐 추출한 키워드와 다른 키워드간의 연결성을 분석하여 네트워크 구조상 중심에 위치하는 키워드를 도출하였다.

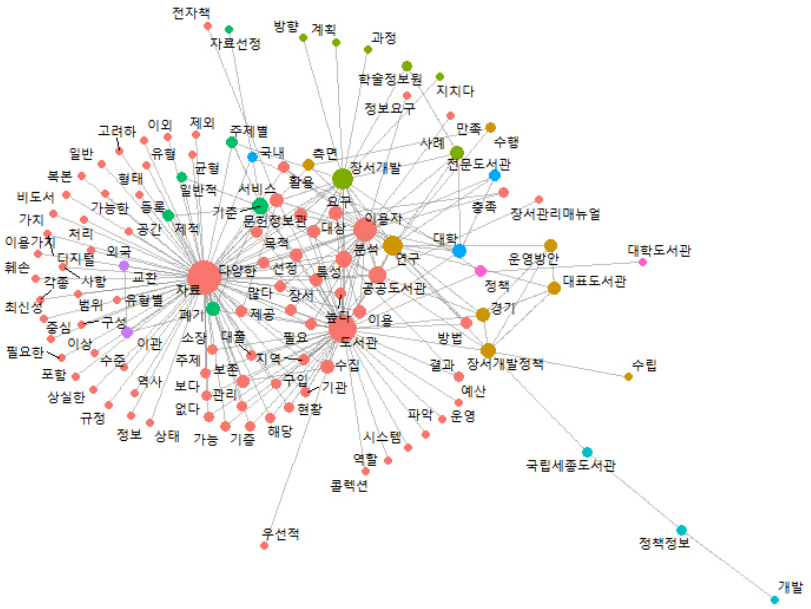
4.3.1 키워드 네트워크

빈도분석에서 추출한 키워드의 동시 출현 관계를 분석하여 동시 출현 빈도가 45회를 초과하는 단어쌍만을 네트워크 형태로 표현하였다. 동시 출현 빈도에 따라 단어를 연결하는 엣지의 색상과 명암을 다르게 설정하였다. 원문에서 함께 등장하는 빈도가 가장 높은 단어쌍은 총 413회 함께 등장한 '도서관과 자료'이다. 그 뒤로 '수집과 자료'(404회), '도서관과 장서개발'(278회), '도서관과 이용자'(263회), '이용과 자료'(237회) 순으로 동시출현 빈도가 높게 나타났다. 동시 출현 빈도가 높은 단어쌍의 키워드는 빈도분석에서 등장 빈도가 높게 나타난 키워드와 일치하고, 네트워크 그래프 상에서 다수의 키워드와 동시출현 관계에 있으므로 빈도분석에서 주요 키워드로 설정된 키워드가 키워드 네트워크에서 연결 중심성이 높은 키워드로 나타남을 알 수 있다.

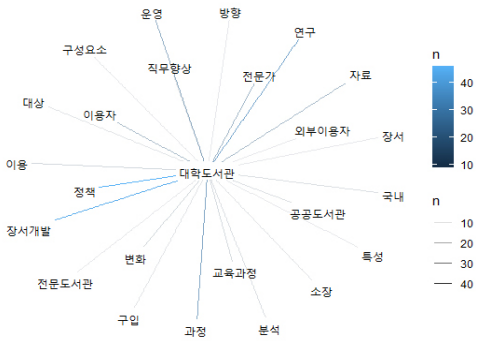
〈그림 6〉은 연결 중심성과 커뮤니티 변수를 추가하여 키워드 네트워크에서의 연결성과 키워드간의 주제 관계를 더욱 명확히 나타낸 그래프이다. 커뮤니티를 색상으로 표시하여 관련 키워드의 유사 관계와 네트워크 관계를 시각적으로 관찰할 수 있다. 네트워크 그래프에서 나타난 커뮤니티의 개수가 7개이므로, 토픽 모델링에서 설정한 토픽의 개수가 타당함을 확인할 수 있다.

4.3.2 주요 키워드 별 네트워크

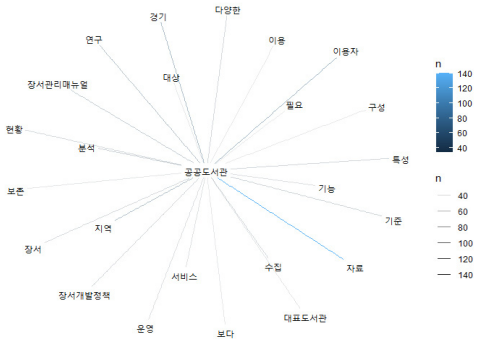
대학도서관, 공공도서관과 관련성 있는 키워드를 별도로 추출하여 관종별 장서개발정책 관련 키워드를 도출하고자 한다. 동시 출현 빈도 외에 파이 계수를 추출하여 주요 키워드와 특징적으로 사용된 단어를 분석하였다.



〈그림 6〉 커뮤니티 중심성 키워드 네트워크



〈그림 7〉 대학도서관 키워드 네트워크

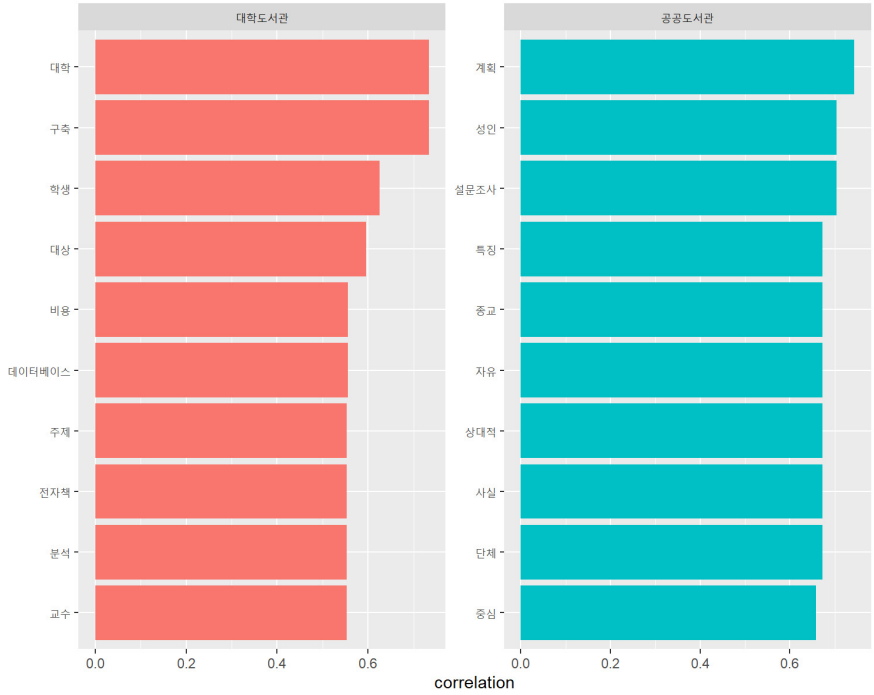


〈그림 8〉 공공도서관 키워드 네트워크

대학도서관과 가장 동시 출현 빈도가 높은 키워드는 '정책'으로 총 46회 함께 등장하였다. 그 외에 동시 출현 빈도가 높은 키워드는 '장서개발', '연구', '과정', '운영' 등 있다.

반면 공공도서관과 동시 출현 빈도가 가장 높은 키워드는 '도서관'(총 152회)으로 나타났다. 그 다음으로 동시 출현 빈도가 높은 키워드로 '자료', '장서개발', '이용자', '지역' 등이 있다.

'장서개발', '연구' 등의 키워드는 대학도서관과 공공도서관에서 모두 연관성이 높게 나타나



〈그림 9〉 대학도서관, 공공도서관 파이 계수

므로 관종별 장서개발 정책의 차이를 비교하기 위한 키워드로 적합하지 않다. 이러한 일반적인 키워드를 제거하고 특징적인 키워드만 도출하기 위해 파이 계수를 활용하였다. 파이계수에서 +1에 가까울수록 키워드 사이의 관련성이 높으며 -1에 가까울수록 관련성이 적음을 의미한다.

파이계수 분석 결과 대학도서관 관련 키워드 중에서 파이 계수가 가장 높은 키워드는 '대학(0.73)' 이며, '구축', '학생', '대상', '비용', '데이터베이스', '주제', '전자책', '분석', '교수' 키워드가 등장하였다. 해당 키워드를 종합하면 대학 도서관의 주 이용자 계층의 요구사항 분석과 데이터베이스와 전자책 등 디지털 자원에 대한 장서개발정책에 대한 논의가 중점적으로 이루어지고 있음을 알 수 있다.

반면 공공도서관의 장서개발정책에 대한 키워드 중 파이계수가 가장 높게 나타난 것은 '계획(0.74)'이었다. '성인', '설문조사', '특징', '종교', '자유', '상대적', '사실', '단체', '중심'이라는 키워드를 통해 다양한 이용자 계층의 요구사항을 파악하고 적용하기 위해 설문조사기법을 주로 활용하고 있으며, '성인', '단체', '자유', '상대적' 등의 키워드를 통해 공공도서관을 이용하는 다양한 이용자 계층을 고려하고 있다.

5. 결 론

도서관의 서비스는 자료의 수집, 관리, 보존하여 이용자에게 제공하는 것을 기반으로 점차 확장되어왔다. 그러므로 도서관의 장서는 도서관 서비스의 질을 결정하는 핵심 요소이다. 장서개발 정책은 도서관 자료를 장서를 체계적이고 일관적으로 구성하기 위한 성문화된 기준이므로, 도서관 서비스의 질을 향상시키기 위해서는 자관의 이용자의 특성에 적합한 장서개발 정책이 필수적이라고 할 수 있다. 이번 연구에서는 각 도서관이 장서개발정책을 수립하는 경우에 고려해야 하는 주요 키워드를 살펴보기 위하여 텍스트 마이닝 기법을 활용하여 장서개발 정책과 관련된 실무 자료를 분석하였다.

특히 도서관 관종에 따라 장서개발정책 수립 시 고려할 키워드를 유형화하기 위하여 대학도서관과 공공도서관이라는 키워드를 중심으로 관련 키워드 네트워크를 도출하였다. 본 연구는 대학도서관과 공공도서관 간의 관종 차이가 장서개발 정책 관련 실무 자료에서 미치는 차이를 관찰한다는 점에서 그 의의가 있다. 연구결과를 요약하면 다음과 같다.

첫째, 전체 원문에서 누적 출현 빈도가 가장 높은 키워드는 '자료'이며, 다음으로 출현 빈도가 높은 키워드는 '도서관', '장서개발', '이용자', '수집', '공공도서관', '연구' 등이 있다. 전체 원문에서 출현하는 빈도와 각 키워드가 빈출되는 원문의 개수 사이에는 높은 상관관계가 있었다.

둘째, 하이퍼 파라미터 정규화 결과 7개의 토픽과 이에 해당하는 키워드가 산출되었다. 각 토픽은 '대학도서관의 자료 및 서비스 관리', '공공도서관의 자료 선정과 관리 기준', '공공도서관 장서개발의 지역적 요구', '정책정보 특화 도서관으로서의 국립세종도서관', '지역 대표도서관 장서개발정책', '공공도서관의 단계별 장서개발정책', '대학도서관의 연구와 서비스 사례'로 명명했으며, 토픽의 키워드와의 주제적 연관성

에 따라 분석 대상 원문을 연결하였다.

셋째, 키워드 간의 동시출현관계를 바탕으로 네트워크 모델링을 수행하였다. 동시출현 관계가 가장 높은 단어쌍은 '도서관'과 '자료'이며, 빈도분석에서 출현 빈도가 높은 키워드들이 동시출현관계 네트워크에서 연결 중심성 및 커뮤니티 중심성이 높은 키워드로 나타났다. 대학도서관, 공공도서관과 관련된 키워드의 네트워크 관계와 파이계수를 살펴보았을 때 대학도서관은 '정책', '장서개발', '연구' 등의 키워드와 함께 등장하는 빈도가 높았으며, 파이 계수가 높은 키워드는 '대학', '구축', '학생', '데이터베이스' 등이 있었다. 반면 공공도서관 키워드와 동시 출현 관계가 높은 키워드는 '자료', '장서개발', '이용자'가 있었으며, 파이 계수를 중점으로 살펴보면 '계획', '성인', '설문조사' 키워드가 관찰된다.

이번 연구의 결과는 다음과 같이 활용할 수 있을 것으로 기대한다. 먼저 각 기관의 장서개발정책 재정 및 개정에 활용할 수 있을 것이다. 빈도분석 및 네트워크 분석의 결과에서 '수집', '전자책', '폐기', '디지털' 등의 키워드가 반복적으로 나타났다. 이는 디지털 접근성이 높아짐에 따라 도서관 서비스가 전통적인 인쇄자료에서 ebook과 전자저널 등 디지털자료로 확장되는 것과 관련이 있다. 그러므로 장서개발정책을 재・개정하는 경우 디지털 형태의 자료의 수집 및 관리와 관련된 논의를 충분히 진행하여야 할 것이다. 인쇄 자료와 구별되는 디지털 형태의 자료의 특성을 충분히 파악하고, 인쇄자료와 디지털 자료의 수서 비중, 전자 자료에의 접근 제한 등을 장서개발 정책에 명시해야 할 것이다.

네트워크 분석 결과, 대학도서관은 이용자와 연구 관련 키워드가 주로 드러났다. 이는 주 이용자들의 연구를 학술적 측면에서 보조한다는 대학 도서관의 운영 목적과 연관이 깊다. 그러므로 장서개발정책 수립 시 대중서보다는 학술도서에 중점을 두어 수집 분야, 희망도서 등을 고려해야 할 것이다. 반면 공공도서관은 지역사회의 특성, 특히 인구통계학적 분포와 지리학적 위치와 관련된 장서개발 정책 키워드가 주로 나타난다. 그러므로 각 도서관이 위치한 지역의 인구통계학적, 지리적 특성을 분석하여 이용자의 요구사항을 구체적으로 파악할 필요가 있을 것이다.

이러한 점에 입각해 각 기관의 운영 목적과 주 이용자층의 요구사항에 부합하도록 장서개발정책을 설립한다면, 통합적이고 장기적인 관점에서 도서관의 장서를 수집, 관리, 폐기 및 보존할 수 있으므로 장서의 질이 향상될 수 있고, 더 나아가 이용자의 장서 이용 만족도를 극대화할 수 있을 것이다.

그러나 이 연구의 대상에서 각 도서관의 장서개발 정책 원문 및 학술 및 학위 논문이 제외되어 있어 장서개발 정책 분야의 학술적 연구 주제에 대한 분석이 이루어지지 않았다. 또한 시계열 분석을 수행하지 못했기 장서개발 정책 분야의 키워드 동향의 변화를 반영하지 못하는 한계가 있다. 향후 연구에서는 다양한 텍스트 마이닝 분석 방법을 활용하여 장서개발 정책에 대한 연구를 보다 심도 있게 진행할 수 있을 것으로 기대한다.

참 고 문 헌

박자현, 송민 (2013). 토픽모델링을 활용한 국내 문헌정보학 연구동향 분석. 정보관리학회지, 30(1), 7-32. http://doi.org/10.3743/KOSIM.2013.30.1.007

박희숙 (2004). 전문도서관의 전자자료 장서개발정책에 관한 연구. Journal of Information Science Theory and Practice, 35(1), 29-49.

신유미 (2019). 장서개발관리 분야 최근 연구동향 분석. 석사학위논문, 상명대학교 대학원 문헌정보학과.

신유미, 박옥남 (2019). 장서개발관리 분야 최근 연구동향 분석에 대한 연구. 정보관리학회지, 36(2), 105-131. http://doi.org/10.3743/KOSIM.2019.36.2.105

안문수 (2011). 대학도서관과 공공도서관 협력 네트워크. 디지틀도서관, 64, 22-37.

윤희윤 (2020). 장서관리론(완전개정 제4판). 서울: 한국도서관협회.

윤희윤, 김종애, 오선경 (2020). 대도시 공공도서관 장서개발정책 분석과 제언. 정보관리학회지, 37(3), 51-75. http://doi.org/10.3743/KOSIM.2020.37.3.051

이진경, 이지연 (2011). 이용자 장서이용행태 분석을 통한 공공도서관의 장서개발에 관한 연구. 한국비블리아학회지, 22(3), 151-171.

정보봉사분과위원회 (2014). 대학도서관 자료선정규정에 관한 사례 연구. 사대도협회지, 15(-), 1-14.

최재황 (2013). 대학도서관의 참고 장서개발정책 기본요소 개발에 관한 연구. 한국도서관・정보학회지, 44(1), 409-432. http://doi.org/10.16981/kliss.44.1.201303.409

최형욱, 최예진, 남소연 (2018). 문헌정보학 분야의 지적구조 및 연구 동향 변화에 대한 시계열 분석: 2003년부터 2017년까지. 정보관리학회지, 35(2), 89-114. http://doi.org/10.3743/KOSIM.2018.35.2.089

Li Jiaen, 최영준 (2023). 텍스트 마이닝을 통한 ESG 연구동향 분석. 기업경영연구, 30(2), 133-166.

Listopia (2000. 6. 16.). 문헌정보학용어사전. 출처: https://www.listopia.net/glossary/49364?_ART001591058

Arshad, M., Khan, A., Ahmed, P., & Shah, N. A. (2020). Next generation data analytics: text mining in library practice and research. Library Philosophy and Practice, 1-12.

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. Social Networks, 1(3), 215-239. https://doi.org/10.1016/0378-8733(78)90021-7

Illescas, L., Sucozhanay, D., & Siguenza-Guzman, L. (2018). Integrating textmining and citation analysis in the decision-making process for library collections. INTED2018 Proceedings, 7450-7456. https://doi.org/10.21125/inted.2018.1754

• 국문 참고자료의 영어 표기
(English translation / romanization of references originally written in Korean)

An, Mun Su (2011). A study of cooperation network between college library and public library. Korea Digital Library Forum, 64, 22-37.

Choi, Hyung Wook, Choi, Ye-Jin, & Nam, So-Young (2018). Time series analysis of intellectual structure and research trend changes in the field of library and information science: 2003 to 2017. Journal of Korea Society for Information Management, 35(2), 89-114. https://doi.org/10.3743/KOSIM.2018.35.2.08

Choi, Jae Hwang (2013). A study on the primary parts of reference collection development policy in academic libraries. Journal of Korean Library and Information Science Society, 44(1), 409-432. http://dx.doi.org/10.16981/kliss.44.1.201303.409

Information Services Division (2014). A case study of a university library collection policy. Korean Association of Private University Libraries, 15, 1-14.

Lee, Jin-Kyung & Lee, Jee-Yeon (2011). A study of collection development by analyzing users' collection-use behaviors in public libraries. Journal of Korean Biblia Society for Library and Information Science, 22(3), 151-171.

Li, Jiaen & Choi, Young Jun (2023). Analysis of major research trends in ESG using topic modeling. Korean Corporation Management Review, 30(2), 133-166.

Listopia (2000, June 16). Library and Information Science Glossary. Available: https://www.listopia.net/glossary/50081?

Park, Hee Sook (2004). A study on the collection development policy of electronic materials of special libraries. Journal of Information Science Theory and Practice, 35(1), 29-49.

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

Shin, You Mi & Park, Ok Nam (2019). An analytical study on research trends of collection development and management. Journal of the Korean Society for Information Management, 36(2), 105-131. http://doi.org/10.3743/KOSIM.2019.36.2.105

Shin, You Mi (2019). An Analytical Study on Research Trends of Collection Development and Management. Master's thesis, Sangmyung University, Korea.

Yoon, Hee Yoon (2020). Collection Management (4th ed.). Seoul: Library Association.

Yoon, Hee Yoon, Kim, Jong Ae, & Oh, Seon Kyung (2020). Analysis and implication of the

collection development policy of public libraries in major cities. Journal of the Korean Society for Information Management, 37(3), 51-75.
http://dx.doi.org/10.3743/KOSIM.2020.37.3.051