Journal of Korean Library and Information Science Society
Vol. 53, No. 2

OPEN ACCESS

다이나믹토픽모델링을 활용한 문헌정보학 분야의 토픽 변화 분석*

Analysis of Research Topic Trend in Library and Information Science Using Dynamic Topic Modeling

김 선 욱(SeonWook Kim)**
양 기 덕(Kiduk Yang)***
이 혜 경(HyeKyung Lee)****

〈 목 차 〉

Ⅰ. 서 론
Ⅱ. 관련 연구
Ⅲ. 연구방법
Ⅳ. 분석결과
Ⅴ. 결 론

요 약 : 본 연구는 2001년부터 2020년까지 문헌정보학 SSCI 85종 학술지에 게재된 55,442편의 학술논문의 논문제목과 초록을 기반으로 다이나믹토픽모델링을 수행하여, 문헌정보학 분야의 연도별 흐름에 따른 연구 주제 추이를 분석하였다. 그 결과, 10개의 토픽에서 도서관경영(장서개발 및 관리, 도서관평가, 도서관 지식경영, 기획 및 활성화), 정보학(계량정보학, 정보이용행태・이용자연구, 의료정보, 정보시스템), 도서관 서비스(도서관교육・정보리터러시), 도서관체계(도서관 시책 및 정책)에 따른 4개의 대분류를 파악하였다. 연도별 흐름에 따라 정보학 영역의 경우, 계량정보학 연구 주제가 학술지단위에서 논문단위로 변화되고 있었으며, 최근 도서관경영 영역의 경우, 이용자의 의견과 감정에 관련한 연구가 최근 등장하였다. 도서관서비스 연구영역은 20년간 안정적인 연구 주제로 그 양상이 보다 심화되고 견고해졌음을 확인할 수 있었다. 그리고 최근에는 모바일과 소셜미디어와 관련한 연구가 진행 중인 것으로 나타났다. 한편, 정보학영역 하위주제로 의료정보와 관련한 연구가 비중 있게 등장해, 문헌정보학의 간학문적인 특징이 잘 나타난 결과라 판단하였다.

주제어 : 계량서지학, 계량정보학, 문헌정보학, DTM, 다이나믹토픽모델링, 동적토픽모델링

ABSTRACT : This study applied dynamic topic modeling on titles and abstracts of 55,442 academic papers in 85 SSCI journals from 2001 to 2020 in order to analyze research topic trend in library and information science. The analysis revealed four major themes of library management, informatics, library service, and library system in 10 major topics. The results also showed subtopics in information science and library management topic areas to change over time, while library service remained stable over 20 years to establish itself as a robust topic. In addition, medical information emerged as a significant sub-topic of informatics, thus exemplifying the interdisciplinary characteristics of library and information science field.

KEYWORDS : Bibliometrics, Library and Information Scicence, DTM, Dynamic Topic Modeling

* 이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2019S1A5C2A02081987).
** 경북대학교 일반대학원 문헌정보학과 박사과정 수료
(seonwook.kim@knu.ac.kr / ISNI 0000 0004 9360 1074) (제1저자)
*** 경북대학교 사회과학대학 문헌정보학과 교수(kiyang@knu.ac.kr / ISNI 0000 0004 7878 6533) (교신저자)
**** 경북대학교 일반대학원 문헌정보학과 박사과정 수료(keilee@knu.ac.kr / ISNI 0000 0004 9344 4179) (공동저자)

• 논문접수: 2022년 5월 25일 • 최초심사: 2022년 5월 29일 • 게재확정: 2022년 6월 14일
• 한국도서관・정보학회지, 53(2), 265-284, 2022. http://dx.doi.org/10.16981/kliss.53.2.202206.265

※ Copyright © 2022 Korean Library and Information Science Society
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.

www.kci.go.kr

Ⅰ. 서 론

토픽 모델은 전처리 된 단어가 존재하는 다량의 문서 군에서 확률적 분포 통계를 활용하여 주제(topic)를 발견하는 일종의 텍스트 마이닝 기법이다. 토픽 모델의 가장 전통적 모델이라 할 수 있는 잠재의미 분석, LSI(Latent Semantic Indexing) 혹은 LSA(Latent Semantic Analysis)(Landauer, Foltz, & Laham, 1998)는 단어의 빈도만으로 나타낼 수 없는 잠재적(Latent) 중요도와 유사성을 파악할 수 있게 한 최초의 발상으로, 단순히 문서 내 단어와 단어 간의 평면적 행렬을 벗어나 벡터라는 공간에서의 단어의 위치를 계산하여 유사성과 확률을 나타내었다. 그런데도 본 기법은 단어의 중의적 의미를 파악하는 데에 무리가 있다는 한계점이 드러나, 최근 일반적으로 가장 많이 사용되고 있는 잠재 디리클레 할당 LDA(Latent Dirichlet Allocation) (Blei, Ng, & Jordan, 2003)이 등장하였다. 이러한 토픽 모델링의 가장 큰 특징은 하나의 문서에서 하나의 토픽을 발견하는 일반적 방법과 달리, 하나의 문서에서 여러 개의 토픽의 존재를 파악하는 것으로 융합 성격을 내포하는 문서들을 파악할 때 매우 유용하기에(박주섭, 홍순구, 김종원, 2017), 연구자들은 토픽모델링을 활용하여 각 연구 동향이나 사회 영향 등의 복합 현상을 파악하고자 하였으며, 디지털 환경의 변화와 머신러닝의 급격한 발전은 LDA 이외에도 다양한 토픽모델의 파생 군들을 탄생시켰다. 그러나 여전히 가장 대표적인 토픽모델로써 LDA가 활용되고 있는데, LDA의 경우, 정적인 시간에서의 문서를 파악해 주제어를 발굴하므로, 즉, 연도를 고려하지 않기 때문에 어떠한 이슈나 주제 분야의 전반적인 통찰을 파악하는 것에는 도움이 되나, 시류를 읽어내기에는 무리가 있다(우창우, 이종연, 2020).

한편, 토픽모델링 기법을 활용하여 문헌정보학 연구 동향을 분석한 연구는 2013년 박자현과 송민의 연구 이후 극히 드물며, 특히 연도를 활용하여 시류를 파악하는 토픽모델링 기법을 활용한 사례는 찾아볼 수 없었다.

따라서 본 연구는 2001년부터 2020년까지 약 20년간 Web of Science 문헌정보학 분야의 SSCI 등재 학술지 85종에 게재한 학술논문의 초록을 기반으로 문헌정보학계의 연구 주제와 시간에 따른 그 추이를 살펴보고자 하였다. 본 연구는 LDA를 기반으로 하여 각 문서의 연도를 고려한 파생 토픽모델기법 중 하나인 다이나믹토픽모델링(DTM)을 사용해 문헌정보학의 주제 분야의 변화를 고찰해 20년간의 문헌정보학 주제변화를 시계열로 확인하고자 하였으며, 그 결과로 20년간 수행된 문헌정보학 분야의 연구 주제와 최근의 추이를 파악하여, 연구자들이 향후 전개할 연구 양상과 문헌정보학의 정체성에 부합하는 다양한 연구 주제의 발견을 도모하고자 한다. 더불어 새로운 토픽모델링 기법의 수행이 여타 연구의 기초자료의 역할을 수행할 수 있을 것으로 기대한다.

www.kci.go.kr

Ⅱ. 관련 연구

본 연구는 다이나믹토픽모델링기법을 활용하여 문헌정보학분야의 연구 주제를 파악하고, 그 흐름을 살펴보고자 하였다. 우선 전반적인 문헌정보학 분야의 연구 동향이나 추이, 현황 등을 고찰하는 연구는 최근까지도 지속해서 진행되고 있었다(박자현, 송민, 2013; 박준형, 오효정, 2017; 손정표, 2003; 송정숙, 2010; 양기덕, 김선욱, 이혜경, 2021; 오세훈, 2005; 이명희, 2005; 이재윤, 2015; 진설아, 송민, 2016; 최형욱, 최예진, 남소연, 2018).

이 중, 토픽모델링을 활용한 연구는 박자현과 송민(2013), 진설아와 송민(2016) 등으로, 박자현과 송민(2013)은 국내 문헌정보학 주요학술지 4종에 수록된 학술논문의 초록을 대상으로 LDA 토픽모델링을 수행하여 도출된 결과를 문헌정보학 주제분류표와 비교 분석하고 연도별 추이를 도출하여 연구 동향을 파악하였다. 그 결과 정보학 영역이 1990년대 중 후반 활발히 연구되다가 2000년대 중반부터 감소세였던 반면에 도서관 서비스 및 평가 영역은 2000년대에 활발히 연구가 진행되었고, 서지학 분야의 하강세를 밝혔다.

진설아와 송민(2016)은 토픽모델링을 기반으로 JCR의 사회과학 분야 중 정보학 분야의 상위 20종의 학술지를 대상으로 학제성을 측정하였는데, 토픽모델링이 학제성을 파악하는 데에 유용성과 텍스트 기반의 다양성과 응집성 등을 파악하고, 학술지별 네트워크를 분석해 정보학 분야를 해석하였다.

한편, 타 학문에서 최근 토픽모델링을 활용하여 연구 동향을 분석한 연구로는 학술지『리터러시연구』의 연구 동향을 파악하기 위하여 토픽모델링 분석을 수행한 이지영(2020)의 연구, ICT 분야의 국가연구개발사업의 주요토픽과 동향을 파악하기 위하여 토픽모델링을 수행한 우창우, 이종연(2020)의 연구, 토픽모델링을 통하여 국내 패션디자인 연구 동향을 분석한 장남경, 김민정(2017)의 연구 등이 있었으며, 대체로 토픽모델링 기법을 진행한 연구의 경우, LDA를 활용하는 것으로 나타났다.

이에 본 연구에서 수행하고자 하는 다이나믹토픽모델링기법을 활용하여 연구 동향을 분석한 연구를 살펴본 바, 우창우, 이종연(2020)은 2015년부터 2020년까지 약 6년간의 SW 및 AI 연구과제 데이터를 기반으로 다이나믹토픽모델링을 수행하여 D(Data), N(Network), A(A.I)분야의 연구 동향을 파악하였다. 그 결과, 2015년부터 약 2년간은 데이터와 네트워크 연구가 활발히 수행되었으며, 이후부터 인공지능 연구가 성장한 것을 확인할 수 있었다. 또한, 다이나믹 토픽모델링의 결과는 LDA 결과보다 연구 동향과 연구 주제를 파악하는 데 더 유용함을 피력하였다.

김동훈, 오찬희, 주영준(2021)은 네트워크 분석기법과 더불어 다이나믹토픽모델링 기법을 활용하여 2013년부터 2020년간 국내에서 발표된 블록체인 관련 연구를 대상으로 그 동향을 분석하였다. 특히, 다이나믹토픽모델링에서는 초록 데이터를 활용해 시계열 주제 분석을 시행하였으며, 각 주제에서 연도별 증가한 키워드(알고리즘, 네트워크, 플랫폼, 인공지능 등)와 하락한 키워드(유통, 금융)

www.kci.go.kr

등을 통해 주제의 연도별 흐름을 파악하였다.

이처럼 선행연구들을 고찰하면서 아직까지 다이나믹토픽모델링을 사용하여 문헌정보학 분야의 연구 주제 등을 파악한 연구는 살펴볼 수 없었다. 주로 우리 학계에서 사용된 토픽모델링기법은 위에서 언급한 바와 같이 Blei, Ng, Jordan(2003)가 고안한 잠재디클리레할당(Latent Dirichlet Allocation, LDA)으로, 본 기법은 최근에도 연구자들이 학문이나 문헌의 주제를 분석하고자 할 때 가장 많이 사용하는 대표적인 기법이다.

LDA는 문서 안의 모든 단어는 하나의 뭉치(bag of words)에 들어가고, 그 단어 간의 순서는 관계없다는 가정에서 시작해, 현재 나타나는 단어들을 통해서 한 문서에서 어떠한 주제가 존재할지를 추정하는 확률 모델로, 대상 문서가 지니는 토픽분포와 토픽이 지니는 단어분포를 추정한다. 이에 연구자는 문서의 토픽분포를 조율할 수 있는 알파(α)와 토픽 내에서 단어의 분포를 조율할 수 있는 베타(β)를 설정할 수 있다(Blei, Ng, & Jordan, 2003). 이렇게 구현된 LDA 모델은 최적의 토픽 개수(k)를 산정하여 설정한 토픽 개수 k에 대해 토픽을 보여준다. LDA 이전에도 확률잠재의미할당(PLSA: Probabilistic Latent Semantic Allocation)이라는 토픽모델링기법이 사용되었으나, 말뭉치의 크기가 커질수록 그 학습을 과도하게 학습시켜 그 결과의 오류를 야기시킬 수 있다는 단점으로 나타났고, 이를 보완하고 개선한 LDA기법은 안정적인 결과를 도출하고, 확장하기 수월한 모델로 자리 잡아 현재까지도 대표적인 토픽모델링기법으로 많은 연구자들이 활용하고 있다.

한편, 본 연구에서 사용하는 다이나믹토픽모델링(Blei & Lafferty, 2006)은 LDA를 고안한 Blei와 Lafferty가 2006년 고안한 기법으로 LDA 기반에서 시간적 변수를 도입한 것이다. 이에 LDA에서 도출한 토픽이 매우 정적인 상태에서의 토픽 결과라면, 다이나믹토픽모델링은 토픽의 변화 양상이나, 토픽에 해당하는 단어의 추이 등을 시계열로 파악할 수 있어 시간에 따른 주제의 진화(evolution)를 분석할 수 있다.

그러므로 본 연구는 다이나믹토픽모델링 기법을 사용하여 문헌정보학 분야의 주제를 파악함과 더불어 연도별 토픽 내 단어의 상승과 하강 등의 주제 변화 추이를 파악하여 약 20년간의 문헌정보학분야의 연구 주제를 고찰해보고자 한다.

Ⅲ. 연구방법

1. 데이터 수집

본 연구는 다이나믹토픽모델링을 활용하여 2001년부터 2021년까지 약 20년간의 문헌정보학의 연구 주제의 분포와 변화양상을 파악하고자, Web of Science에 등재된 문헌정보학분야의 SSCI

www.kci.go.kr

학술지 86종에 게재된 학술논문 188,674편의 서지정보를 수집하였다. 그러나 학술지 "Library Journal"의 경우, 20년간 게재된 논문이 총 97,726건이었다. 이는 본 학술지를 제외한 85종에 게재된 학술논문 편수인 90,948건보다 6,678건이나 많은 것이다. 본 연구에서는 "Library Journal"이 이상치로 판단하고, 이를 제외한 총 85종의 90,948편의 학술논문의 서지정보를 사용하였다(〈표 1〉 참조).

이렇게 수집한 논문 90,948편의 서지정보에서 본 연구가 사용한 데이터는 논문의 제목과 출판년도, 초록이었으며, 언어가 'English'가 아닌 데이터, 출판년이 누락된(null) 데이터, 초록이 포함한 문자열의 단어 수가 25개 미만인 데이터는 제외하였다. 출판사 Elsevier는 학술논문 초록작성에 있어 최대 작성 길이는 한정하나, 최소 작성에 대하여서는 언급하지 않고 있지만, Deveci(2019)는 한 초록이 최소 하나의 문장을 포함하고 있다고 가정하고, 25개 미만의 단어를 지닌 초록은 분석에서 제외한 바, 본 연구에서도 이와 같은 처리과정을 거쳤으며, 최종적으로 토픽모델링 분석에 사용한 학술논문 데이터는 55,442편이다.

〈표 1〉 수집 학술지 목록

순번	학술지명	논문 수
1	African Journal of Library Archives and Information Science	227
2	Aslib Journal of Information Management	320
3	Canadian Journal of Information and Library Science-Revue Canadienne Des Sciences De L Information et De Bibliotheconomie	727
4	College & Research Libraries	1,522
5	Data Base for Advances in Information Systems	266
6	Data Technologies and Applications	112
7	Econtent	2,744
8	Electronic Library	1,833
9	Ethics and Information Technology	380
10	European Journal of Information Systems	892
11	Government Information Quarterly	1,318
12	Health Information and Libraries Journal	774
13	Informacao & Sociedade-Estudos	615
14	Informacios Tarsadalom	449
15	Information & Culture	207
16	Information & Management	1,353
17	Information and Organization	183
18	Information Development	694
19	Information Processing & Management	1,819
20	Information Research-An International Electronic Journal	1,854
21	Information Society	832
22	Information Systems Journal	596
23	Information Systems Research	870

www.kci.go.kr

순번	학술지명	논문 수
24	Information Technology & Management	311
25	Information Technology & People	506
26	Information Technology and Libraries	617
27	Information Technology for Development	396
28	International Journal of Computer-Supported Collaborative Learning	344
29	International Journal of Geographical Information Science	1,953
30	International Journal of Information Management	1,730
31	Investigacion Bibliotecologica	503
32	Journal of Academic Librarianship	2,301
33	Journal of Computer-Mediated Communication	622
34	Journal of Documentation	1,470
35	Journal of Enterprise Information Management	371
36	Journal of Global Information Management	346
37	Journal of Global Information Technology Management	264
38	Journal of Health Communication	1,993
39	Journal of Information Science	1,155
40	Journal of Information Technology	571
41	Journal of Informetrics	1,063
42	Journal of Knowledge Management	955
43	Journal of Librarianship and Information Science	962
44	Journal of Management Information Systems	916
45	Journal of Organizational and End User Computing	220
46	Journal of Scholarly Publishing	588
47	Journal of Strategic Information Systems	458
48	Journal of the American Medical Informatics Association	3,559
49	Journal of the Association for Information Science and Technology	1,340
50	Journal of the Association for Information Systems	571
51	Journal of the Australian Library and Information Association	340
52	Journal of the Medical Library Association	1,776
53	Knowledge Management Research & Practice	610
54	Knowledge Organization	793
55	Law Library Journal	1,390
56	Learned Publishing	1,114
57	Library & Information Science Research	838
58	Library and Information Science	139
59	Library Collections Acquisitions & Technical Services	596
60	Library Hi Tech	885
61	Library Quarterly	973
62	Library Resources & Technical Services	766
63	Library Trends	901
64	Libri-International Journal of Libraries and Information Studies	237

www.kci.go.kr

순번	학술지명	논문 수
65	Malaysian Journal of Library & Information Science	282
66	Mis Quarterly	938
67	Mis Quarterly Executive	261
68	Online Information Review	1,809
69	Portal-Libraries and the Academy	1,084
70	Profesional De La Informacion	1,479
71	Qualitative Health Research	2,816
72	Reference & User Services Quarterly	2,861
73	Reference Services Review	261
74	Research Evaluation	627
75	Restaurator-International Journal for the Preservation of Library and Archival Material	370
76	Revista Espanola De Documentacion Cientifica	458
77	Scientist	8,428
78	Scientometrics	4,838
79	Serials Review	821
80	Social Science Computer Review	1,008
81	Social Science Information Sur Les Sciences Sociales	611
82	Telecommunications Policy	1,431
83	Telematics and Informatics	931
84	Transinformacao	327
85	Zeitschrift Fur Bibliothekswesen Und Bibliographie	1,277
총 계		90,948

2. 데이터 전처리

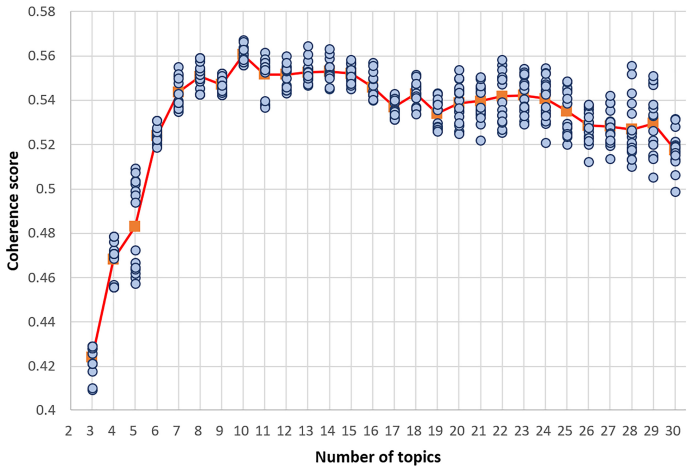
본 연구는 보다 정확한 분석을 위해서 텍스트 전처리를 다음과 같이 수행하였다.

첫째, 괄호 특수문자 '( ), [ ], { }' 앞뒤의 공백문자를 추가하고, 둘째, 연속한 빈칸을 하나의 빈칸으로 대체하였다. 셋째 하이픈(-)을 공백문자로 대체하였다. 넷째, 자동전처리모델에서 '%'를 명사로 인식하는 경우가 있으므로 삭제하였다. 다섯째, '(C) 2020 Elsevier Ltd. All rights reserved 및 Published by Elsevier Ltd'와 같이 출판사 Elsevier가 초록 내 임의로 삽입한 저작권 문구가 분석 대상의 데이터의 약 17%에 존재하여 해당 문구를 삭제하였다. 여섯째, 위와 같이 수동으로 처리한 데이터를 spaCy 모델을 활용하여 자동 전처리를 진행하였다. spaCy(https://www.spacy.io)는 파이썬 기반의 자연어처리 오프소스 라이브러리로, 66개 이상 언어에 대해 지원하고 있으며, 본 연구는 이 중 'en_core_web_lg' 언어 모델을 사용하였다. 한편, 자동 전처리를 진행할 때는 대문자를 소문자로 변환하고, 표제어추출(lemmatization)에서 명사의 기본형을 추출하고, 최대 3어절(trigram)까지 고려하였다.

www.kci.go.kr

3. 분석방법

본 연구는 Web of Science 중 문헌정보학분야의 학술지 85종에 게재된 학술논문들의 제목과 초록 데이터를 활용하여 다이나믹토픽모델링을 진행하였다. 토픽모델링의 보다 정확한 결과를 획득하기 위하여, 위에서 설명한 바와 같이 수집데이터에 대하여 수동 및 자동으로 데이터 전처리를 수행하였다. 더불어 본 연구는 최적의 토픽 개수를 선정하기 위하여 〈그림 1〉과 같이 3부터 30까지의 범위의 토픽 개수에서 데이터 일관성(coherence score)을 분석하였다.[1] 그 결과, 토픽 개수 k값이 10일 때 가장 최적의 결과를 나타낼 수 있는 것으로 판단하였다.



〈그림 1〉 토픽개수 k를 도출하기 위한 Coherence 분포

Ⅳ. 분석결과

〈표 2〉는 토픽을 10구분으로 구분하여 다이나믹토픽모델링기법을 수행한 결과로써 5개년 별로 문헌정보학 분야의 토픽을 도출한 것이다. 각 토픽의 주제는 오세훈(2005)의 문헌정보학 주제분류 분석 도구에서의 대주제영역 및 하위주제를 활용하여 부여하였다.

총 10개 토픽 중 정보학(토픽 1, 7, 8, 10), 도서관 경영(토픽 2, 3, 4, 5), 도서관 서비스(토픽 6),

1) 본 연구는 토픽 개수를 선정하기 위하여 perplexity score도 연산을 진행하였으나, 토픽개수가 증가할수록 분산하는 경향이 확인되어, perplexity score는 본 연구가 수행하는 토픽모델링의 하이퍼파라미터 성능을 판별하는데에는 무리가 있다고 판정하였으므로, Coherence score를 활용하였다.

www.kci.go.kr

도서관 체제(토픽 9) 등 4개의 대 주제영역이 등장하였으며, 도서관경영과 정보학으로 총 토픽 4개를 차지하면서 현재 문헌정보학 분야의 다양한 연구 주제를 내포하고 있음을 추측할 수 있었다. 이외 도서관 서비스와 도서관 체제에 대한 주제영역이 존재하는 것으로 나타났다.

도서관경영은 장서개발과 관리(토픽 2), 도서관 평가(토픽 3), 도서관 지식경영이나, 기획 및 활성화(토픽 4, 5) 등을 하위주제로 연구가 진행되었으며, 정보학은 각각 계량정보학(토픽 1), 디지털도서관 중 정보이용행태 등의 이용자 연구(토픽 7), 정보시스템(토픽 10) 등이 나타났다. 더불어 오세훈(2005)의 문헌정보 주제분류표에는 등장하지 않으나, 본 연구의 분석결과 의료정보(토픽 8)가 등장하였고, 의료정보의 경우 건강 및 의료적 수치를 취합하여 정량적 분석을 진행하기에, 이를 정보학의 세부분야로 판단하고, 정보학 대주제영역의 하위주제로 편성하였다. 대주제영역 도서관 서비스(토픽 6)는 특히 이용자교육과 정보리터러시 분야가 혼재되어 있었으며, 도서관 체제(토픽 9)는 도서관의 시책이나 정책 등의 하위주제로 연구 논의가 다수 진행된 것을 확인할 수 있었다.

각 대주제영역의 주제 토픽을 세부적으로 살펴보면, 도서관경영의 하위주제인 장서 관리의 경우, 'print'나 'collection', 'book' 등의 키워드 확인으로 인쇄 매체 관련 장서 및 자원과 관련한 연구는 여전히 상위 연구 주제로서 자리 잡고 있었으며, 2010년 이후 'web'이나 'repository', 'website' 등의 키워드가 상위로 등장하면서 도서관에서의 온라인 장서 관리 혹은 저장 및 보존에 따른 연구가 수행되기 시작하였음을 추측할 수 있었다. 그리고 토픽 3의 주제인 도서관 평가에서 2015년 이후부터 'review', 'emotion' 등의 키워드가 상위에 등장하는 것으로 나타나, 도서관 평가에서 이용자들의 의견이나 만족도, 감정 등을 활용한 연구가 나타나는 것을 추측할 수 있었다.

한편 정보학 관련 토픽 1 계량정보학의 연구 주제는 꾸준히 학술지(journal)와 인용(citation) 중심으로 진행되고 있었으며, 2010년부터 키워드 'article'이 상승하여 학술지 단위의 연구에서 논문 단위의 연구가 등장하고 있음을 추측할 수 있는데, 2015년 키워드 'collaboration'의 등장과 이후 'author', 'researcher' 등의 키워드 상승은 논문 편당의 저자나 공동연구 등과 같은 계량정보학적 연구가 특히 연구자들에게서 진행되고 있음을 파악할 수 있었다. 토픽 7의 정보이용행태의 경우 꾸준히 이용자나 대중들의 정보 행위와 관련된 연구가 주로 진행된 것과 더불어 2020년 키워드 'privacy'의 등장은 이용행위에 대한 개인정보 및 이용자 개인의 정보행태 향유와 관련한 연구 수행이 상승한 것으로 추측하며, 2010년 이후 키워드 'intention' 등장을 통하여 최근 이용자들의 정보 검색 혹은 정보 이용행위의 동기나 의도, 요구 등을 주제로 한 연구가 활발히 수행 중에 있음을 파악할 수 있었다. 토픽 10 정보시스템의 경우 2000년대 초반 정보시스템 자체의 설계나 구축을 위한 검색이나 데이터베이스에 관한 연구가 수행되었다면, 현재는 각 문서의 자질이나 검색을 위한 방법 및 알고리즘 등 보다 구체적인 검색 구축에 관한 연구가 수행되고 있었다.

도서관 서비스의 토픽6 도서관교육과 정보리터러시는 특정한 세부주제의 집중 없이 포괄적으로 현재까지 연구가 수행되고 있었으며, 도서관 체제 토픽 9의 경우 도서관 정부 시책에 대한 프로젝트를

www.kci.go.kr

수행한 결과로서의 연구에 대한 주제로 2020년의 'trajectory'의 경우, 실제 초록에서 GPS나 지역의 물리적 거리나 도서관 시책에 필요한 거리상 검토 등에 사용된 키워드로 나타나 도서관 관련 정책에서 지역적 유대나 행정 구역 등에 대하여 자세히 검토하고 있음을 확인할 수 있었다.

〈표 2〉 5개년 별 토픽

토픽 번호	대분류	하위주제	연도				
			2001	2005	2010	2015	2020
1	정보학	계량정보학	journal	journal	journal	journal	journal
			science	citation	citation	publication	citation
			publication	publication	publication	citation	publication
			citation	indicator	indicator	article	article
			indicator	article	article	country	field
			field	science	index	field	indicator
			country	author	field	indicator	researcher
			article	discipline	science	science	author
			author	scientist	author	author	science
			scientist	field	country	collaboration	impact
2	도서관경영	장서개발 및 관리	library	library	library	library	library
			service	service	service	service	service
			access	access	access	access	collection
			collection	collection	resource	resource	resource
			resource	resource	collection	collection	access
			reference	material	web	repository	website
			librarian	print	repository	web	book
			user	record	content	book	repository
			material	originality_value	originality_value	website	user
			print	book	user	record	material
3	도서관경영	도서관평가	quality	consumer	quality	consumer	consumer
			consumer	quality	product	quality	product
			product	product	trust	product	platform
			trust	trust	consumer	trust	customer
			team	web_site	team	customer	review
			effect	website	website	review	quality
			instrument	effect	customer	effect	app
			customer	customer	effect	rating	effect
			music	price	market	website	emotion
			web_site	risk	price	emotion	commerce
4	도서관경영	도서관 지식경영	knowledge	theory	theory	practice	practice
			theory	concept	concept	theory	work
			process	knowledge	process	concept	community
			concept	approach	development	work	concept
			work	practice	practice	knowledge	theory
			approach	process	approach	framework	process
			context	work	framework	context	design
			practice	article	system	process	system
			article	development	work	approach	approach
			action	context	context	design	way

www.kci.go.kr

토픽 번호	대분류	하위주제	연도				
			2001	2005	2010	2015	2020
5	도서관경영	기획 및 활성화	technology	system	knowledge	knowledge	innovation
			business	business	organization	business	knowledge
			organization	technology	business	innovation	firm
			firm	government	firm	technology	business
			management	management	government	government	government
			system	firm	innovation	organization	performance
			industry	organization	capability	firm	industry
			government	industry	technology	management	technology
			company	implementation	industry	system	capability
			innovation	process	performance	industry	relationship
6	도서관 서비스	도서관교육 정보리터러시	student	student	student	student	student
			program	librarian	learning	librarian	librarian
			learning	learning	program	learning	university
			course	program	education	program	education
			skill	skill	skill	education	skill
			education	course	literacy	literacy	program
			training	training	librarian	skill	literacy
			literacy	education	training	school	job
			librarian	literacy	course	course	learning
			faculty	faculty	school	university	course
7	정보학	정보이용행태 이용자연구	internet	internet	internet	user	medium
			group	user	user	medium	user
			communication	use	behavior	behavior	behavior
			use	medium	community	internet	use
			medium	behavior	use	use	privacy
			computer	group	medium	intention	factor
			behavior	computer	factor	factor	intention
			community	communication	group	motivation	finding
			user	people	intention	community	survey
			people	community	behaviour	behaviour	internet
8	정보학	의료정보	health	health	health	health	health
			patient	patient	patient	patient	patient
			woman	woman	woman	intervention	woman
			care	care	intervention	care	care
			intervention	health_care	care	woman	intervention
			health_care	child	hospital	child	disease
			physician	intervention	child	participant	hospital
			child	author	disease	cancer	child
			family	physician	cancer	health_literacy	healthcare
			disease	cancer	participant	risk	health_care
9	도서관체제	도서관 시책 및 정책	model	model	network	network	network
			network	network	model	model	city
			map	map	scale	location	location
			market	policy	competition	map	pattern
			simulation	cost	broadband	city	path
			telecommunication	error	simulation	distance	region
			competition	simulation	location	simulation	model
			object	telecommunication	cost	method	simulation
			agent	market	area	pattern	trajectory
			policy	competition	error	structure	distance

www.kci.go.kr

토픽 번호	대분류	하위주제	연도				
			2001	2005	2010	2015	2020
10	정보학	정보시스템	document	search	search	search	method
			system	document	document	method	feature
			search	retrieval	method	document	search
			retrieval	query	system	text	model
			method	web	classification	system	text
			user	method	text	classification	task
			text	language	retrieval	feature	classification
			term	term	query	task	approach
			set	text	term	query	algorithm
			database	user	task	retrieval	dataset

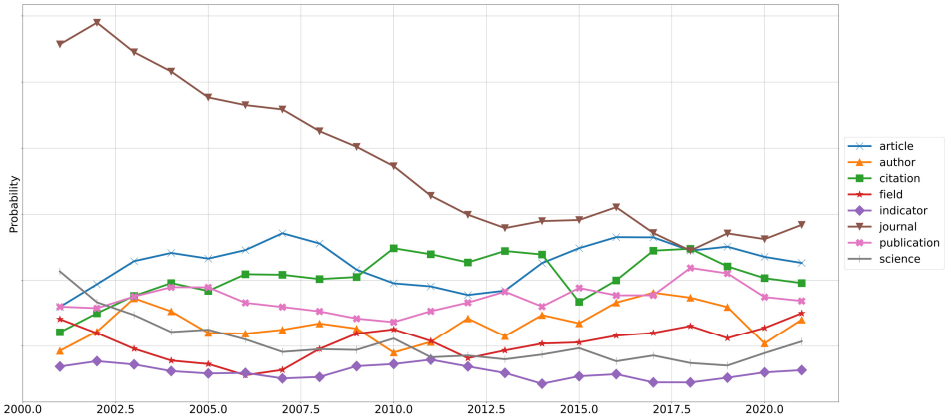
〈그림 2〉에서 〈그림 5〉는 정보학 분야의 토픽들의 키워드의 시계열 추이를 나타낸 것이다. 〈그림 2〉는 계량정보학 분야로 위에서 설명했던 바와 같이 'journal' 단위의 연구가 여전히 주요 내용이나, 2002년 이후 급격히 하강추세를 나타내고 있으며, 상대적으로 'article' 단위의 연구가 상승세임을 확인 할 수 있다. 또한 계량정보학적 분석에 대한 연구로 키워드 'field'가 미세하게나마 상승세로 다양한 학문분야별 계량정보학적 분석이 행해지고 있음을 추측할 수 있다.

〈그림 4〉는 토픽 8의 의료정보와 관련한 연구 주제로 이 분야의 연구 주제는 일반 건강적 내용을 포함하고 있었으나, 2020년 이후 'health' 키워드의 급격한 상승세는 Covid 19 시대와 맞물린 사회 현상을 반영한 것으로 추측할 수 있다.

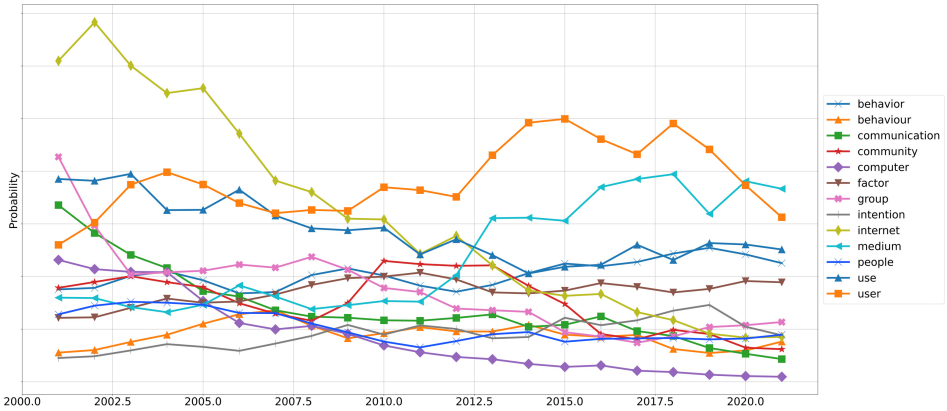
한편, 〈그림 5〉는 토픽 10의 정보시스템 연구 주제의 키워드 추이로써, 20년간 상위 키워드들이 조화롭게 연구의 흐름을 이끌어, 편향된 주제로 정보시스템 관련 연구를 진행하지 않고, 최근까지 안정적인 연구 경향을 나타낸 것으로 보였다. 그러나 상대적으로 키워드 'system'의 하락세는 정보시스템 연구 중 직접 시스템을 수립하고 구축하였던 2000년대 초반 연구를 기점으로 다소 감소하고 있음을 추측할 수 있었으며, 〈표 2〉에서 확인한 것처럼 시스템 설계 방법과 각 검색의 자질, 검색방식 등에 관한 연구가 구체적으로 진행됨을 〈그림 5〉를 통하여 보다 선명히 확인할 수 있었다.

〈그림 3〉은 토픽 7의 정보이용행태와 이용자 연구에 관련한 키워드 추이로, 본 그림에서 괄목할 만 한 점은 상위권이었던 키워드 'Internet'의 하락세이다. 1990년대 인터넷이 등장하고 성장하면서 본 학계도 인터넷 기반의 정보검색 이용행태 관련 연구가 2000년대 초반까지 이어져 왔었으나, 2000년대로 접어들면서 대중들의 온라인 정보향유가 일상화되고, 매개체가 단순히 컴퓨터와 인터넷을 넘어서 모바일 디바이스로 이동하는 등의 시류에 편승하여 컴퓨터와 인터넷을 주제로 한 연구가 타 연구 주제에 반해 급감하게 되고, 대중(medium)들의 정보 이용행태 자체와 의도를 분석하고 고찰하는 연구가 상대적으로 성장하고 있음을 확인할 수 있었다.

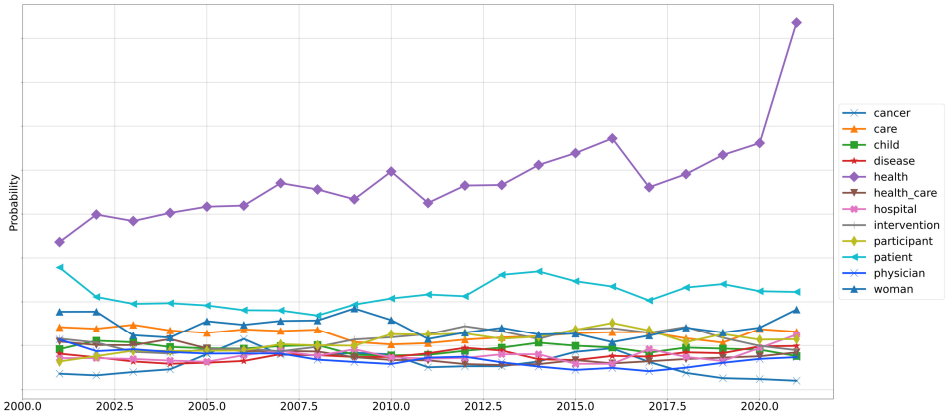
www.kci.go.kr



〈그림 2〉 토픽 1(계량정보학)의 연도별 키워드 추이

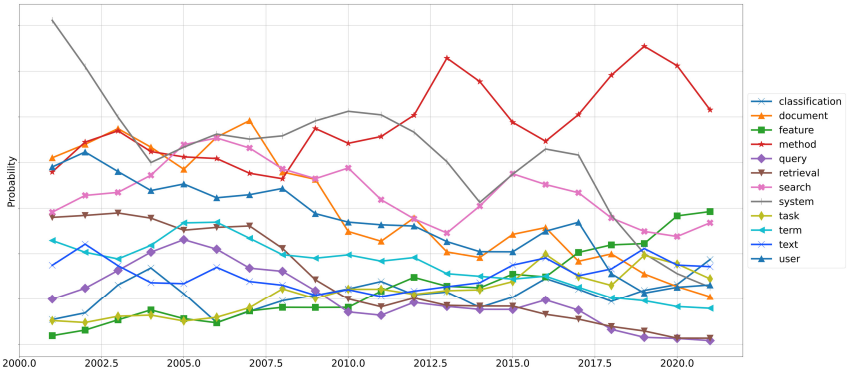


〈그림 3〉 토픽 7(정보이용행태 이용자연구)의 연도별 키워드 추이



〈그림 4〉 토픽 8(의료정보)의 연도별 키워드 추이

www.kci.go.kr



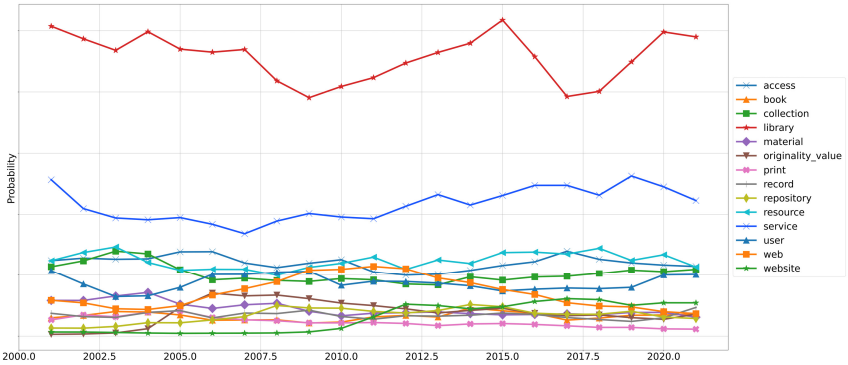
〈그림 5〉 토픽 10(정보시스템)의 연도별 키워드 추이

〈그림 6〉에서 〈그림 9〉는 도서관 경영과 관련한 토픽들의 키워드 시계열 추이이다.

〈그림 6〉은 토픽 2 도서관 장서 관리에서 나타난 키워드 추이로, 도서관과 서비스 키워드가 연도 흐름과 관계없이 주요 연구 분야로 자리 잡고 있었으며, 〈표 2〉에서 나타난 바와 같이, 'web'과 'website', 'repository'의 키워드가 2010년 이후 미세하게 상승세를 나타내며 관련 웹 자원이나 장서와 저장 등에 관한 연구가 시작되고 있었음을 파악할 수 있었으나, 여전히 주요 분야는 단행본 장서 연구임을 〈그림 3〉을 통해서 확인할 수 있었다.

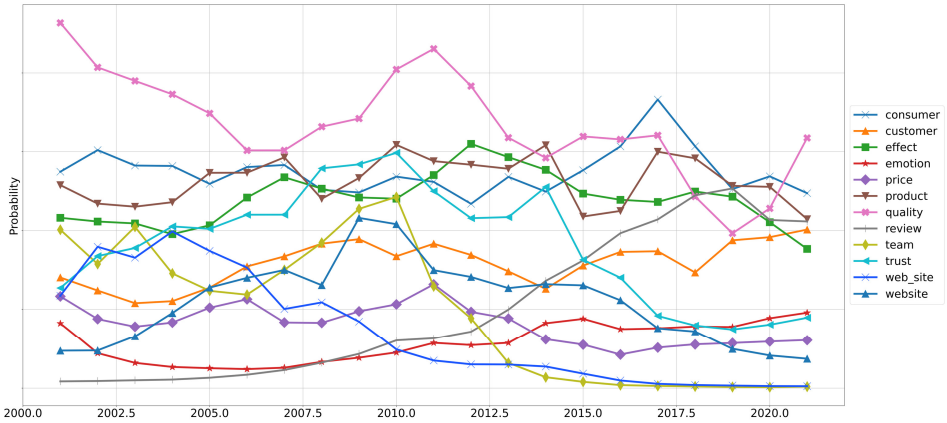
〈그림 7〉은 도서관 평가 관련 주제로 상위 키워드들이 해마다 변동 폭이 달라 전반적으로 주제 내에서 다양한 연구가 수행되어왔음을 추측할 수 있었다. 그런데도 특히 'review' 키워드의 지속적인 상승세는 위의 연구 분야에서 이용자 또는 소비자들의 만족이나 의견을 반영한 연구의 비중이 높아지고 있음을 시사한다.

〈그림 8〉과 〈그림 9〉는 각각 토픽 4와 5에 해당하는 키워드의 흐름으로, 도서관을 실제 경영하고, 도서관에서 제공하는 지식을 운영하는 등의 키워드들이 급격한 변화를 겪으면서, 시대의 흐름에 따른 도서관경영의 변화에 편승한 연구가 행해져 왔음을 파악할 수 있었다.

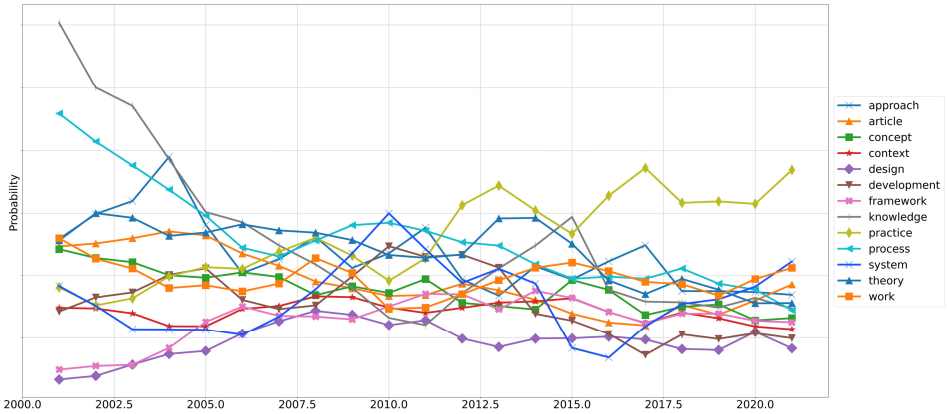


〈그림 6〉 토픽 2(장서개발 및 관리)의 연도별 키워드 추이

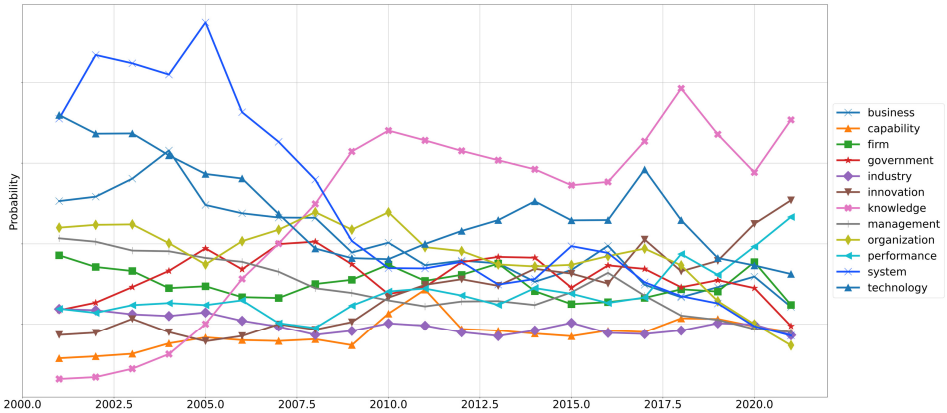
www.kci.go.kr



〈그림 7〉 토픽 3(도서관평가)의 연도별 키워드 추이



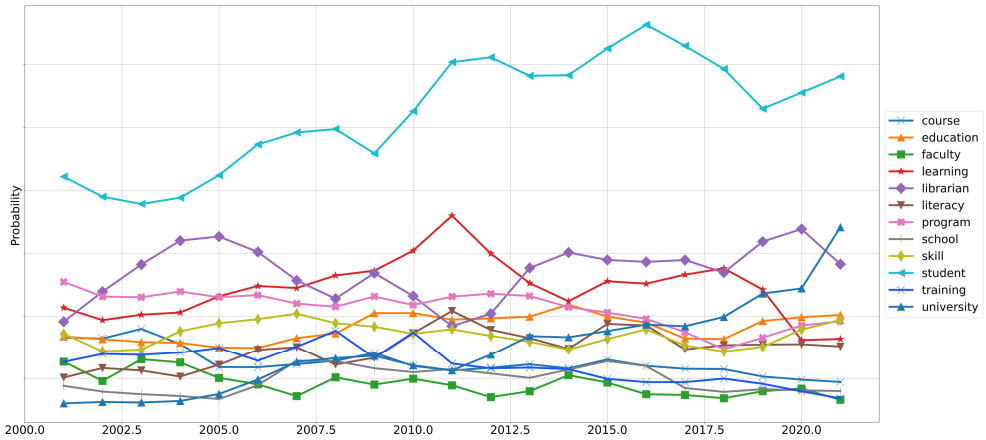
〈그림 8〉 토픽 4(도서관 지식경영)의 연도별 키워드 추이



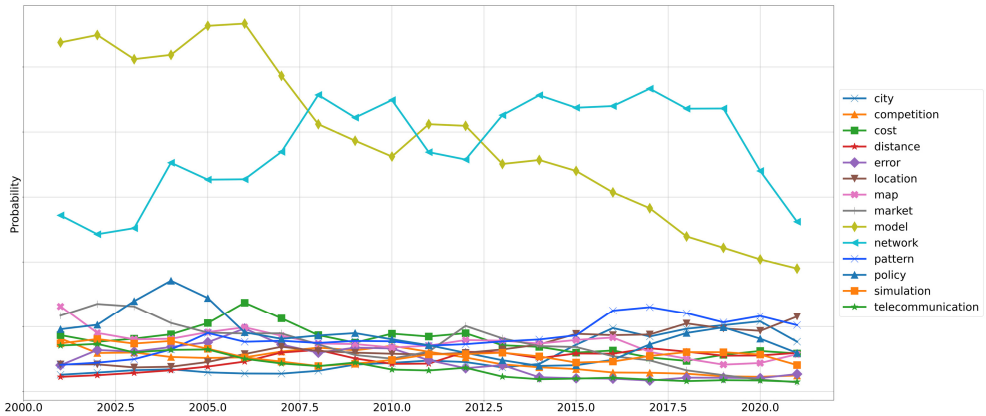
〈그림 9〉 토픽 5(기획 및 활성화)의 연도별 키워드 추이

www.kci.go.kr

〈그림 10〉과 〈그림 11〉은 각각 토픽 6 도서관 서비스와 토픽 9 도서관체계에 대한 연구 주제의 연도별 키워드 추이를 나타낸 것이다. 〈표 2〉에서 나타난 것처럼 5개년별 상위 키워드는 안정적이었으나, 그 추이를 살펴보면 〈그림 10〉에서 도서관 교육 및 정보리터러시에서는 학생(student)의 키워드가 타 키워드보다 매우 높은 수치와 지속적인 상승추세로 최근일수록 보다 견고한 연구가 수행 중임을 시사하였다. 이에 관련 토픽의 원 데이터의 초록을 살펴본바, 특히 2010년 토픽 6에 해당하는 초록데이터 103건 중 'mobile', 'social media' 등과 관련한 연구는 11건에 불과하였으나, 2020년의 경우, 해당 초록 데이터 209건 중 36건이었고, 소셜미디어 관련 정보리터러시 연구가 31건으로 나타나, 최근일수록 현대 사회현상을 반영한 소셜미디어나 디지털 기기 관련 정보 활용 교육 등의 연구가 과거에 비하여 보다 활발히 수행되고 있음을 확인할 수 있었다.



〈그림 10〉 토픽 6(도서관교육, 정보리터러시)의 연도별 키워드 추이



〈그림 11〉 토픽 9(도서관 시책 및 정책)의 연도별 키워드 추이

www.kci.go.kr

한편 〈그림 11〉의 도서관체계 주제 중 시책과 정책에 대한 연구 분야에서는 2007년을 기점으로 키워드 'model'과 'network'의 순위변화에서 대략 5년간의 과도기를 거쳐 완전히 전환되는 것으로 확인하면서 도서관 정책에 있어 도서관 모델의 정규화를 지향하는 연구에서 도서관과의 교류나 혹은 도시 및 타 기관 등의 관계나 협약 등을 아우르는 정부 정책 혹은 계획에 대한 프로젝트 수행이 진전되고 이에 따른 결과로서의 연구 수행이 나타남을 시사하고 있다.

Ⅴ. 결 론

본 연구는 2001년부터 2020년까지 Web of Science에 등재된 문헌정보학 분야 SSCI 학술지 85종에 수록된 학술논문을 대상으로 다이나믹토픽모델링을 진행하여 약 20년간 수행된 문헌정보학 분야의 연구 주제를 고찰하고 주제변화의 흐름을 파악하고자 하였다. 이에 연구 결과는 다음과 같다.

첫째, 최근 20년간의 대주제 분야는 주로 도서관경영, 정보학, 도서관 서비스, 도서관 체제 등으로 확인할 수 있었으며, 도서관경영은 도서관 기획, 지식경영, 장서 관리, 평가와 관련한 연구를, 정보학은 계량정보학, 이용자 이용행태 등의 평가, 정보시스템, 의료정보와 관련한 연구를, 도서관 서비스는 도서관교육과 정보리터러시, 도서관 체제는 도서관의 시책이나 정부 정책에 관련한 연구를 진행하는 것으로 나타났다.

둘째, 정보학 분야 중 의료정보 분야가 문헌정보학 하위주제영역으로 나타나, 학계 내 정보 활용 연구의 범위가 다소 광범위 됐음을 확인할 수 있었다.

셋째, 다이나믹토픽모델링에서 키워드의 시계열 추이를 확인한 바, 대분류별 하위주제영역이 시기에 따라 미세하게 변화가 있는 것으로 나타났다. 정보학 하위 계량정보학 주제영역에서 학술지 단위의 연구가 논문 단위연구로 이동하고 공동연구 관련 연구가 최근에서 수행되고 있었으며, 정보학 하위 정보이용행태 주제영역에서 컴퓨터와 인터넷 주제 연구보다 개인정보 관련 연구와 대중들의 실제 정보행태의 의도를 파악하는 연구가 증가하고 있었고, 도서관경영 하위 도서관 평가 주제영역에서 이용자의 의견과 감정과 관련한 연구가 나타났으며, 도서관체계 하위 시책 정책 주제영역에서 도서관 단위의 프로젝트가 지역, 시도, 국가 등 다양한 네트워크를 활용한 정책 개발의 결과를 연구 성과로 나타내고 있었다.

넷째, 도서관경영 하위 장서 관리 주제영역과 도서관 서비스 하위 도서관 교육 및 리터러시 주제영역은 최근까지도 급격한 연구 주제의 변화 없이 안정적인 행태를 나타내었다.

본 연구는 국외 학술지를 대상으로 영문초록만을 활용했다는 점에서 국내의 문헌정보학 분야의 주요 연구 주제와는 다소 상이할 수 있다. 그러므로 본 연구에서는 후속연구로 국내 문헌정보학 분야의 학술논문을 대상으로 한 연구나, 한국어 학습을 통한 기계학습 전개를 수행하는 연구 등과 더불어 컴퓨팅 환경의 개선과 한국어 코퍼스 수집 등 다양한 방법으로 실험의 성능과 효과의 상승을 위한 연구도 필요함을 제언하고자 한다.

www.kci.go.kr

본 연구의 결과는 국외 문헌정보학 분야의 주요 주제의 변화와 흐름을 전반적으로 파악할 수 있다는 점에서 그 의의가 있다. 더불어 학계 연구자들이 더 창의적이고 깊이 있는 연구 주제를 발굴하는 데에 기초적인 자료로서, 지식의 지평을 넓히는 데에 도모할 수 있을 것으로 기대한다.

참 고 문 헌

김동훈, 오찬희, 주영준 (2021). 다이나믹 토픽모델링 및 네트워크 분석 기법을 통한 블록체인 관련 국내 연구 동향 분석. 정보관리학회지, 38(3), 23-39. https://doi.org/10.3743/KOSIM.2021.38.3.023

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

박주섭, 홍순구, 김종원 (2017). 토픽모델링을 활용한 과학기술동향 및 예측에 관한 연구. 한국산업정보학회논문지, 22(4), 19-28.

박준형, 오효정 (2017). 국내 기록관리학 연구동향 분석을 위한 토픽모델링 기법 비교: lda와 hdp를 중심으로. 한국도서관·정보학회지, 48(4), 235-258. doi:10.16981/kliss.48.4.201712.235

손정표 (2003). 한국의 문헌정보학분야 연구동향 분석: 1957~2002. 한국도서관·정보학회지, 34(3), 9-32.

송정숙 (2010). 한국 문헌정보학의 연구 현황분석. 한국도서관·정보학회지, 41(4), 333-353.

양기덕, 김선욱, 이혜경 (2021). 국제 및 국내 문헌정보학 분야의 연구성과 비교 분석. 한국문헌정보학회지, 55(1), 365-392.

오세훈 (2005). 우리나라 문헌정보학 학술지 논문 및 인용문헌 분석을 통한 연구동향 연구. 정보관리학회지, 22(3), 379-408. https://doi.org/10.3743/KOSIM.2005.22.3.379

우창우, 이종연 (2020). 다이나믹 토픽 모델을 활용한 D(data)·N(network)·A(A.I) 중심의 연구동향 분석. 한국융합학회논문지, 11(9), 21-29.

이명희 (2005). 문헌정보학 분야의 질적 연구 동향과 과제. 한국비블리아학회지, 16(2), 177-201.

이재윤 (2015). 문헌동시인용 분석을 통한 한국 문헌정보학의 연구 전선 파악. 정보관리학회지, 32(4), 77-106.

이지영 (2020). 토픽모델링 분석을 활용한 학술지『리터러시연구』연구 동향 분석. 리터러시연구, 11(6), 537-565. https://doi.org/10.37736/kjlr.2020.12.11.6.18

장남경, 김민정 (2017). 토픽모델링을 이용한 국내 패션디자인 연구동향 분석. 디지털융복합연구, 15(6), 415-423.

www.kci.go.kr

진설아, 송민 (2016) 토픽모델링 기반 정보학 분야 학술지의 학제성 측정 연구. 정보관리학회지, 33(1), 7-32.

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

Blei, D. M., & Lafferty, J. D. (2006, June). Dynamic topic models. In Proceedings of the 23rd International Conference on Machine Learning, 113-120.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of Machine Learning Research, 3(4-5), 993-1022.

Deveci, T. (2019). Sentence length in education research articles: a comparison between Anglophone and Turkish authors. The Linguistics Journal, 14(1), 73-100.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998) An introduction to latent semantic analysis. Discourse Processes, 25, 259-284.

• 국한문 참고문헌의 영문 표기
(English translation / Romanization of references originally written in Korean)

Choi, Hyung Wook, Choi, Ye Jin, & Nam, So-Yeon (2018). Time series analysis of intellectual structure and research trend changes in the field of library and information science: 2003 to 2017. Journal of the Korean Society for Information Management, 35(2), 89-114. https://doi.org/10.3743/KOSIM.2018.35.2.089

Jang, Namkyung & Kim, Min-Jeong (2017). Research trend analysis in fashion design studies in Korea using topic modeling. Journal of Digital Convergence, 15(6), 415-423.

Jin, Seol A & Song, Min (2016). Topic modeling based interdisciplinarity measurement in the informatics related journals. Journal of the Korean Society for Information Management, 33(1), 7-32.

Kim, Donghun, Oh, Chanhee, & Zhu, Yongjun (2021). Analyzing research trends in Blockchain studies in South Korea using dynamic topic modeling and network analysis. Journal of the Korean Society for Information Management, 38(3), 23-39. https://doi.org/10.3743/KOSIM.2021.38.3.023

Lee, Jae Yun (2015). Identifying the research fronts in Korean library and information science by document co-citation analysis. Journal of the Korean Society for Information Management, 32(4), 77-106.

www.kci.go.kr

Lee, Jiyoung (2020). A topic modeling analysis of research trends in the Journal of Literacy. The Korean Journal of Literacy Research, 11(6), 537-565. https://doi.org/10.37736/kjlr.2020.12.11.6.18

Lee, Myeong-Hee (2005). Some trends and issues of qualitative research method in library and information science. Journal of the Korean Biblia Society for Library and Information Science, 16(2), 177-201.

Oh, Se-Hoon (2005). A study on the research trends of library & information science in Korea by analyzing journal articles and the cited literatures. Journal of the Korean Society for Information Management, 22(3), 379-408. https://doi.org/10.3743/KOSIM.2005.22.3.379

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. doi:10.3743/KOSIM.2013.30.1.007

Park, Ju Sep, Hong, Soon-goo, & Kim, Jong-Weon (2017). A study on science technology trend and prediction using topic modeling. Journal of the Korea Society Industrial Information Syste, 22(4), 19-28.

Park, JunHyeong & Oh, Hyo-Jung (2017). Comparison of topic modeling methods for analyzing research trends of archives management in Korea: focused on LDA and HDP. Journal of Korean Library and Information Science Society, 48(4), 235-258. doi:10.16981/kliss.48.4.201712.235

Sohn, Jungpyo (2003). An analytical study on research trends of library and information science in Korea: 1957~2002. Journal of Korean Library and Information Science Society, 34(3), 9-32.

Song, Jung-Sook (2010). Trends in library and information science research in Korea: focused on masters theses and doctoral dissertations from 2001 to 2010. Journal of Korean Library and Information Science Society, 41(4), 333-353.

Woo, Changwoo & Lee, JongYun (2020). Investigation of research trends in the D(Data) · N(Network) · A(A.I) field using the dynamic topic model. Journal of the Korea Convergence Society, 11(9), 21-29.

Yang, Kiduk, Kim, SeonWook, & Lee, HyeKyung (2021). Comparison of research performance between domestic and international library and information science scholars. Journal of the Korean Library and Information Science Society, 55(1), 365-392.

www.kci.go.kr