

텍스트마이닝을 활용한 대전시 공공도서관 이용자의 인식과 경험 연구*

- SNS와 온라인 뉴스 기사를 중심으로 -

A Study on the Perception and Experience of Daejeon Public Library Users Using Text Mining: Focusing on SNS and Online News Articles

최 지 원 (Jiwon Choi)**

곽 승 진 (Seung-jin Kwak)***

목 차

- | | |
|-----------|----------|
| 1. 서 론 | 4. 연구 결과 |
| 2. 이론적 배경 | 5. 결 론 |
| 3. 분석방법 | |

초 록

본 연구는 텍스트마이닝 기법을 중심으로 빅데이터 분석을 활용하여 대전시 공공도서관에 대한 이용자의 인식과 경험을 살펴보고자 수행되었다. 이를 위하여 첫째, 소셜미디어에 나타난 이용후기 데이터를 수집하여 대전시 공공도서관에 대한 이용자들의 전반적인 인식과 평가를 탐색하였다. 둘째, 온라인 뉴스 기사 분석을 통해 사회적으로 논의되고 있는 현안을 파악하였다. 분석 결과, 첫째로 어린이 동반 이용자 비중의 높다는 것과 다음으로 LDA 분석을 통한 토픽이 '문화행사/프로그램', '자료 이용', '물리적 환경 및 시설', '도서관 서비스'의 네 가지 분류로 나타난다는 것, 마지막으로 뉴스기사 데이터에 도서관 및 복합문화공간 추가 건립과 도서관 협력 체계 구축에 대한 키워드가 핵심적으로 등장한다는 것을 확인하였다. 이를 바탕으로 지역 균형을 고려한 공공도서관 건립과 육아 및 보육 기관과의 업무협약을 통한 사회적 육아공동체 네트워크 조성을 제안하였다. 본 연구를 활용하여 대전시 공공도서관의 정책적·사회적 흐름을 알아보고 지역사회 수요를 반영하는 공공도서관 운영을 데이터에 기반하여 실행할 수 있기를 기대한다.

ABSTRACT

This study was conducted to examine the user's experiences with the public library in Daejeon using big data analysis, focusing on the text mining technique. To know this, first, the overall evaluation and perception of users about the public library in Daejeon were explored by collecting data on social media. Second, through analysis using online news articles, the pending issues that are being discussed socially were identified. As a result of the analysis, the proportion of users with children was first high. Next, it was found that topics through LDA analysis appeared in four categories: 'cultural event/program', 'data use', 'physical environment and facilities', and 'library service'. Finally, it was confirmed that keywords for the additional construction of libraries and complex cultural spaces and the establishment of a library cooperation system appeared at the core in the news article data. Based on this, it was proposed to build a library in consideration of regional balance and to create a social parenting community network through business agreements with childcare and childcare institutions. This will contribute to identifying the policy and social trends of public libraries in Daejeon and implementing data-based public library operations that reflect local community demands.

키워드: 데이터, 공공도서관, 이용자 연구, SNS, 소셜미디어, 텍스트마이닝

Data, Public library, User study, SNS, Social Media, Text Mining

* 이 연구는 충남대학교에 의해 지원되었음.

** 충남대학교 문헌정보학과 대학원 석사과정(lib7209@cnu.ac.kr / ISNI 0000 0005 1449 7633) (제1저자)

*** 충남대학교 문헌정보학과 교수(sjkwak@cnu.ac.kr / ISNI 0000 0004 6812 0586) (교신저자)

논문접수일자: 2024년 5월 4일 최초심사일자: 2024년 5월 9일 게재확정일자: 2024년 5월 15일

한국문헌정보학회지, 58(2): 363-384, 2024. <http://dx.doi.org/10.4275/KSLIS.2024.58.2.363>

* 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. 서론

1.1 연구의 배경과 목적

조직 경영에 있어 데이터에 기반한 의사결정이 필수적인 시대를 향해가고 있다. 데이터 기반 사회 흐름에 도서관 경영 역시 예외일 수는 없다. 특히 사용자가 자발적으로 상호작용하고 자신의 의견과 경험을 공유할 수 있는 친숙한 매체인 소셜미디어는 이용자의 답론이 가장 활발한 데이터로, 정치적 성향 분석에서 제품 리뷰에 이르기까지 다양한 유형의 연구에서 소비자의 여론을 수집하기 위한 유용한 자료원으로 활용되고 있다. 디지털 대전환 시대를 맞이하며 산업, 조직, 문화, 시스템 등 사회 전반이 데이터를 기반으로 변화하는 지금, 도서관계에서 전략적 의사결정을 내리는 데 데이터에 기반한 시도를 이어가는 것이 중요한 시점이다.

이에 본 연구는 텍스트마이닝 기법을 중심으로 빅데이터 분석을 활용하여 대전시 공공도서관에 대한 이용자의 경험을 살펴보고자 한다. 연구는 두 단계로 진행하는데, 첫 번째 단계에서는 소셜미디어에 나타난 이용후기 데이터를 수집하여 대전시 공공도서관에 대한 이용자들의 전반적인 평가와 인식이 어떠한지를 탐색해보고자 한다. 두 번째 단계는 온라인 뉴스 기사를 이용한 키워드 분석으로, 이를 통해 대전시 공공도서관의 정책적·사회적 흐름을 알아보고자 한다.

SNS 데이터의 텍스트마이닝을 통한 대전시 공공도서관의 이용자 인식 및 평가를 위해 연구에서 설정한 연구문제는 다음과 같다. 첫째, 소셜미디어에서 나타난 대전시 공공도서관의

키워드는 무엇인가? 둘째, 소셜미디어에서 나타난 대전시 공공도서관의 키워드는 어떤 주제 군집을 형성하는가? 셋째, 온라인 뉴스를 통한 대전시 도서관의 사회적 논의는 어떻게 나타나는가?

이용 당사자인 시민의 요구를 충족시켜 이용을 활성화하는 것은 도서관의 필수과제이며 이에 이용자 인식과 평가를 살펴보기 위한 연구가 수차례 진행되었다. 그러나 기존의 이용자 면담과 설문을 활용한 연구가 연구자에 의해 설계된 문항으로 연구자의 요구에 의해 수집된다는 점을 고려할 때, 이용자가 직접 자신의 입장에서 자발적으로 작성한 이용 후기를 살펴볼 수 있는 소셜미디어는 매우 중요한 데이터라고 할 수 있다. 하지만 이를 활용한 도서관 이용자 인식 연구는 미미한 상황이었다. 따라서 본 연구는 인터넷 소셜미디어에 나타난 이용후기 데이터를 분석하여 기존의 연구에서 발견할 수 없었던 시사점을 도출하고 이용자의 실질적인 수요를 파악하여 공공도서관의 발전을 위한 방향을 제시하고자 하였다.

1.2 연구의 대상 및 범위

본 연구의 공간적 범위는 대전광역시로 설정하였고, 연구의 대상은 대전광역시의 공공도서관 총 26개관으로 선정하였다. 시간적 범위는 2019년 2월부터 2024년 2월 사이 생성된 데이터를 대상으로 설정하였다.

본 연구의 내용적 범위는 다음과 같다.

첫째, SNS 텍스트마이닝을 활용한 연구와 공공도서관 이용자 인식 및 평가 연구의 선행 연구를 각각 고찰하고 기존 연구의 의의와 한

계점을 고찰한다.

둘째, 대전광역시 공공도서관 26개의 도서관 명을 키워드로 하여 이용경험을 기반으로 한 블로그의 텍스트데이터를 수집한다.

셋째, 수집한 데이터의 빈도분석을 통해 핵심 키워드를 도출하고 LDA(Latent Dirichlet Allocation) 토픽모델링을 활용하여 주제별로 탐색한다.

넷째, 대전광역시 도서관에 관한 뉴스 기사를 수집하여 핵심 키워드를 도출함으로써 사회적 논의의 방향성을 살펴본다.

2. 이론적 배경

2.1 선행연구

선행연구를 위해 먼저, SNS 텍스트마이닝 분석을 통해 이용자나 고객의 요구를 파악하여 전략적 의사결정을 내리는 것과 관련된 연구를 국내와 국외 연구로 나누어 살펴보았다.

국내에서 이지원, 유기윤, 김지영(2019)은 소셜미디어 데이터를 이용하여 이용자 중심의 공공시설 선호요인을 도출하고 공급측면과 수요측면에서 정량화하여 공공시설 매력도지수를 산출하였으며 서울시 공공도서관 15개 기관에 대한 네이버 블로그를 수집하여 LDA를 통해 공공도서관 선호요인을 파악하였다. 서혜선(2023)은 SNS 텍스트 데이터를 기반으로 대전시 환경 관련 토픽 모델링 및 감성 분석을 연구하였고 환경과 관련된 SNS 콘텐츠의 주제와, 해당 게시물의 주제와 감성어와의 관계를 각각 파악하였다. 블로그, 카페, 웹 문서, 대전 지역 신문 등

의 데이터를 사용하였고 사람들이 환경과 함께 관심을 가지는 주제에 관해 탐색하였다. 이옥, 임혜원, 여하림, 황혜선(2021)은 인공지능 서비스에 대한 온라인뉴스, 소셜미디어, 소비자리뷰를 텍스트마이닝 및 LDA를 통해 분석하였다. 온라인뉴스와 소셜미디어 감성분석을 통해 소비자의 기능적, 정서적 인식을 파악하였으며 소비자 리뷰 분석을 통해 소비자의 활용영역과 불편 사항 등을 도출하였다. 이주경, 손용훈(2021)은 구글맵리뷰를 활용한 텍스트마이닝을 통해 공원 이용자의 인식 및 평가를 알아보고자 하였으며 TF-IDF를 이용한 단어 빈도분석과 중심성 분석(centrality analysis)을 활용하여 파악한 공원 이용자의 수요와 이용평가를 기반으로 공원 평가에 대한 새로운 방안을 제시하였다.

국외에서 Luo(2021)은 소셜미디어 텍스트마이닝을 활용한 성별에 따른 대중교통 서비스에 대한 인식 연구를 위해 웨이보(Weibo)의 마이크로블로그에서 데이터를 수집하고 LDA로 주제를 탐색하였다. 연구 결과 대중교통에 대한 소셜미디어 데이터의 74%가 여성이며 여성은 대중교통 환경의 편안함에 관심을 가지는 반면, 남성은 전자 결제 서비스 및 교통사고에 관심을 가지는 경향이 있음이 확인되었다. Dewinta, Irawan(2021)은 정부 식수 회사의 고객 불만 사항을 추출하기 위해 트위터(Twitter)의 데이터를 텍스트마이닝 하였다. LDA와 자기조직화지도(Self-Organizing Maps, SOM)를 사용하여 주제를 탐색하였고 배관 누수가 가장 주요한 요인임을 확인하였다. Nilashi et al.(2021)은 소셜미디어 온라인 리뷰를 통해 채식 레스토랑의 서비스 품질을 향상시키는 방법을 연구하였고 LDA와 자기조직화지도를 활용하여 고객의 평

가요소와 만족도를 탐색하는 모델을 제시하였다. Oyeboode et al.(2021)은 소셜미디어 댓글을 기반으로 한 코로나 19의 건강 및 심리사회적 문제 연구를 수행하였고 감성분석을 통해 도출된 부정적 이슈와 긍정적 이슈를 각각 찾아보며 정부의 의료계가 취할 수 있는 대응 방안을 제안하였다.

이와 같이 소셜미디어를 활용하여 이용자의 인식과 요구를 파악한 선행연구를 조사한 결과 다음과 같은 의의를 얻을 수 있었다. 첫 번째로 이용자의 자발적이고 주관적인 의견이 사용되어 전통적인 설문조사 데이터의 단점을 보완할 수 있어 서비스품질 향상에 유용하다는 견해가 공통적으로 등장하였다. 두 번째로 국내와 국외의 선행연구를 조사한 결과 특정한 주제 분야에 국한되지 않고 경영, 행정, 의학 등의 다양한 분야에서 서비스 개선을 위해 활발하게 사용되는 방법론임이 파악되었으나 도서관을 대상으로 한 연구는 미미한 것으로 나타났다. 마지막으로 텍스트마이닝 기법으로 LDA, 자기조직화지도, 감성분석, 중심성분석 등 다양한

분석 기법이 사용되었으나 그 중 LDA 토픽모델링 기법이 가장 빈번하게 사용되었음을 확인하였다.

다음으로, 최근 3년간 설문을 통해 공공도서관 이용자의 인식 및 평가를 조사한 연구와, 해당 연구에서 설문에 활용한 요인을 <표 1>과 같이 정리하였다. 송민선, 장인호, 황금숙(2022)은 안양시 도서관의 중장기 계획 수립을 위해 이용자와 전문가의 인식을 조사하기 위해 이용자를 대상으로 설문조사를 진행하였다. 전경선, 신동희(2022)는 양평군의 지역사회 특성을 반영하여 공공도서관을 활성화하기 위한 방안을 연구하였고 이를 위해 양평군민을 대상으로 설문을 통해 도서관 이용과 비이용 의도를 분석하였다. 김신영, 차성중(2022)은 인구감소지역의 공공도서관 서비스 개선을 위해 의성군을 중심으로 이용자 인식조사 연구를 수행하였고 주민들의 이용 현황과 서비스 만족도 등을 조사하였다. 이수영 외(2023)는 도봉구 공공도서관의 중장기 발전계획 수립을 위해 설문을 활용하여 지역주민의 인식과 요구를 파악하였다. 노

<표 1> 공공도서관 이용자 인식 및 평가에 관한 선행연구의 설문조사 문항

발행년도	2022	2022	2022	2023	2024	2024	2024
연구자명	송민선 외 2명	전경선, 신동희	김신영, 차성중	이수영 외 3명	노영희 외 2명	임성관	장인호, 권선희
연구대상 지역	안양시	경기 양평군	경북 의성군	서울 도봉구	경기 과천시	서울 강동구	포천
장서 및 자료 현황	○	○	○	○	○	○	○
인적자원	○		○	○		○	○
독서프로그램/문화행사	○	○	○	○	○	○	○
공간 및 시설	○	○	○	○	○	○	○
접근성	○		○			○	
도서관 서비스	○	○		○	○	○	
기타	○	○	○			○	

영희, 장인호, 박우정(2024)은 과천시에 건립 예정인 도서관 운영계획을 수립하기 위해 이용자의 인식 및 평가를 진행하였고 주민들의 도서관 주이용목적에 파악하였다. 임성관(2024)은 강동구 공공도서관에 대한 지역 주민과 도서관 관계자의 인식을 알아보기 위해 연구를 진행하였고 주민을 대상으로 도서관 요인별 만족도를 조사하였다. 마지막으로 장인호, 권선희(2024)는 포천시를 중심으로 공공도서관 이용자의 인식조사 연구를 수행하였고 이를 위해 포천시민을 대상으로 도서관 이용자의 만족도를 조사하였다.

논문의 특성에 따른 예외적인 질문을 제외하면 대부분 유사한 질문으로 구성되어 있으며, 이를 통해 최근 3년간 설문조사를 이용하여 이용자 인식 및 평가를 조사한 연구에서 장서 및 자료 현황, 도서관 서비스, 독서 프로그램 및 문화행사, 공간 및 시설, 접근성, 인적 자원 등의 요인을 사용하였음을 확인하였다. 이처럼 기존 연구에서 이용자에게 어떤 요인에 관한 의견을 수집하고 있는지 파악하는 과정을 통해 기존 연구에 대한 방법론적 보완 가능성을 탐색하였다.

이상의 선행연구 결과를 요약하면, LDA를 활용한 SNS 텍스트마이닝 분석으로 이용자 인식 및 평가를 수행한 연구가 유용성을 인정받아 국내외에서 활발히 이루어지고 있는 동향임에도 불구하고 도서관계에서 서울 이외에 지역을 대상으로 SNS 텍스트마이닝 분석 기법을 적용하여 이용자 인식 및 평가를 수행한 연구는 찾아볼 수 없었다.

이에 본 연구를 통하여 데이터 기반 사회에서 도서관 경영에 이용자가 생성한 실제적 경험 및 의견 데이터를 반영하기 위해 대전시 공

공도서관을 중심으로 탐색적 연구를 시도해 보고자 한다.

2.2 대전광역시 공공도서관 현황

대전광역시는 대한민국의 5대 광역시 중 하나로, 대한민국의 중심에 위치한 도시이다. 약 540km²의 면적에 인구가 약 144만명에 달하며 동구, 중구, 서구, 유성구, 대덕구의 5개 구와 79개의 동으로 이루어져 있다. 크기는 유성구>동구>서구>대덕구>중구 순이며 인구수는 서구>유성구>중구>동구>대덕구 순으로 이루어져 있다. <표 2>는 대전광역시에서 운영하는 '대전의 통계' 홈페이지에 2024년 4월 1일 게시된 2024년 3월말 주민등록 인구현황을 바탕으로 재구성하였다.

대전시는 인구수가 140만이 넘는 광역도시임에도 2024년 2월 현재 총 26개관의 공공도서관만이 운영되고 있어 대전의 공공도서관 부족 문제는 지속적으로 제기되고 있다. 안영웅, 임윤택(2023)은 대전 공공도서관의 접근성에 대한 공간적, 경제적 계층 간 형평성을 연구하였고 대전의 외곽에 위치한 대덕구와 유성구의 접근성이 상대적으로 낮았음을 도출하였다. 정유경, 박성희 외(2021)는 대전 시민을 대상으로 한 공공도서관 서비스 인식 및 요구를 분석하였고 공공도서관의 낮은 물리적, 시간적 접근성을 주요 원인으로 파악하였으며 도서관 재건축 및 증축을 제안하였다. 각 도서관 홈페이지의 시설현황에 공개된 정보를 참고하여 대전시 공공도서관의 현황을 <표 3>으로 정리하였으며 확인이 필요한 정보는 유선상의 문의를 통해 사서에게 직접 확인하여 작성하였다.

〈표 2〉 대전광역시 자치구별 규모 및 공공도서관 수

자치구	인구수(명)	면적(km)	공공도서관 수
동구	216,872	136.68	6
중구	223,076	62.18	3
서구	464,115	95.5	5
유성구	367,727	176.53	9
대덕구	169,899	68.69	3
합계	1,441,689	539.58	26

〈표 3〉 대전광역시 공공도서관 시설 현황

도서관명	개관 년도	연면적 (㎡)	회원수	공간운영		
				열람실	자료실	기타공간
가오도서관	1996	2,800	35,645	열람실	어린이자료실, 종합자료실, 유아방	휴게실, 동아리방, 시청각실, 문화교실
용운도서관	1994	1,872	51,772	성인열람실, 열람실	종합자료실, 신문·큰글자책 자료 실, 어린이 자료실	휴게실, 시청각실, 강좌실
판암도서관	2004	469	6,397	-	어린이자료실(유아방포함), 종합 자료실(미디어창작소), 디지털자 료실	-
무지개도서관	2010	511	1,836	-	종합자료실, 어린이·영어자료실	다목적실
자양도서관	2009	1,010	3,231	-	종합자료실, 유아자료실, 아동자 료실	세미나실
홍도도서관	2009	932	4,329	-	종합자료실, 어린이자료실	-
한밭도서관	1989	22,519	289,805	열람실, 성인열람실	자료실, 다문화자료실, 정기간행 물실, 어린이자료실, 시각장애인실, 청록실(노인자료실)	디지털 창작실, 진로상담실, 디지털정 보센터(디지털영상관, 정보검색실, 영상음향실, 멀티미디어연수실), 전 시실, 생활문화센터(동아리실, 강의 실, 준비실), 다목적홀, 창작실, 식당
학생교육문화원	1988	4,990	164,611	열람실	종합자료실, 어린이자료실, 유아 코너, 책숲(VOD, 간행물, 커뮤니 티룸)	아트강좌실, 강좌실, 어린이강좌실, 마루홀(영화상영), 미디어해봄, 1216 해봄(청소년 다목적공간)
산성어린이 도서관	2023	2,318	104,190	열람실	산성책뜰(종합자료실), 새싹누리 (유아자료실), 꿈누리(어린이자 료실), 영어누리(영어자료실)	누리홀(시청각실), 꿈터(강좌실), 상상누리(유아·어린이창작공간)
가수원도서관	2000	3,877	9,974	일반열람실	어린이도서실, 도서전자자료실	회의실, 문화행사실, 휴게실, 시청각실
갈마도서관	1994	2,706	10,217	일반열람실	도서정보실, 어린이도서실, 어린 이영어자료실, 임산부실, 전자정 보자료실	시청각실, 문화교실, 식당, 청소년진 로진학상담센터
둔산도서관	2007	2,367	12,108	일반열람실	어린이자료실, 도서전자실, 교구 놀이방	문화행사실, 휴게실, 시청각실, 전시실
월평도서관	2019	3,585	492,031	일반열람실	도서전자실, 어린이자료실, 교구 놀이방	시청각실, 회의실, 문화행사실, 휴게실
서구어린이 도서관	2008	560	2,871	가족열람실	그림책방, 어린이도서전자실	이야기방

도서관명	개관 년도	연면적 (㎡)	회원수	공간운영		
				열람실	자료실	기타공간
송촌도서관	2010	1,983	6,448	-	종합자료실, 어린이자료실, 유아 실, 노인·장애인실	정보화교육장, 예술강좌실, 평생학 습실, 세미나실, 북카페, 교양교실, 다목적실
신탄진도서관	2002	4,891	4,453	열람실	전자정보실, 노인장애인실, 어린 이 자료실, 유아실, 종합자료실, 보 존자료실	강의실, 노트북실, 식당, 전산교육 장, 다목적홀
안산도서관	1994	1,937	5,723	학습열람실	노인·장애인 열람실, 어린이자료 실, 종합자료실, 전자정보실,	그림책 이야기방, 시청각실, 학습동 아리방, 문화사랑방, 독서사랑방
관평도서관	2015	2,029	15,978	열람실	종합자료실, 어린이자료실	문화나눔공간, 휴게실, 대강당
구암도서관	2013	4,096	447,962	열람실	종합자료실, 어린이자료실	-
구죽도서관	2003	1,496	6,573	열람실	어린이실, 종합자료실	시청각실, 문화사랑방, 디지털휴게실
노은도서관	2010	3,355	34,773	열람실	종합자료실, 어린이자료실, 디지 털실	시청각실, 문화사랑방, 독서사랑방, 독서동아리방, 매점
아가랑도서관	2021	996	1,476	-	자료실	공동육아나눔터, 여성커뮤니티센터, 지혜나눔터, 활동나눔터
원신흥도서관	2019	3,888	11,941	열람실	종합자료실, 어린이자료실, 보존 서고	배움터, 지혜나눔터, 꿈나눔터, 다문 화커뮤니티센터, 강당, 갤러리북, 북 카페, 와이파이존, 유아실, 수유실, 어울림터(휴게실)
유성도서관	1994	2,466	42,263	열람실	디지털체험실, 종합자료실, 아동 자료실	강당, 독서사랑방, 스튜디오, 문화사 랑방, 유아실, 수유실, 매점
전민도서관	2023	2,791	448	-	어린이자료실, 임산부휴게실, 종합 자료실	북카페, 프로그램실, 다목적실, 마루 공간, 커뮤니티실, 밴드연습실, 개인 연습실, 어쿠스틱연습실, 창작실, 다 목적공방
진잠도서관	2010	2,591	9,914	열람실	디지털자료실, 종합자료실, 어린이 자료실	북나눔터, 스마트워크센터(원격근 무지원실), 동아리나눔터, K-도서 관(미디어창작공간), 시청각실, 휴 게실

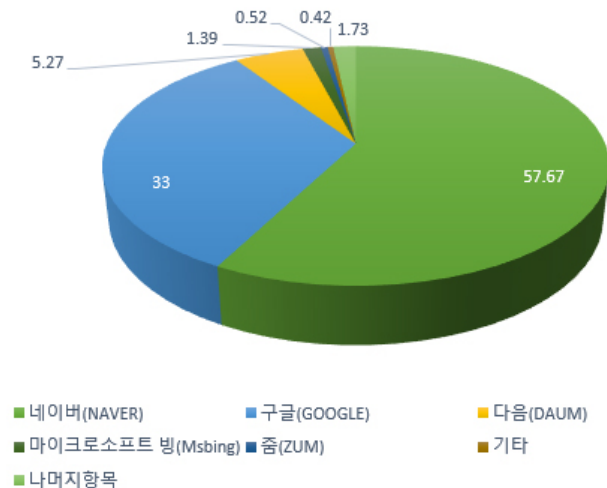
3. 분석방법

3.1 데이터 수집 및 정제

본 연구에서 대상으로 하는 대전시 공공도서관에 대한 SNS 데이터는 네이버(NAVER) 블로그를 대상으로 수집하였다. 이는 포털사이트의 온라인 점유율 통계를 제공하는 사이트인 Internet Trend에서 확인한 2019년 2월부터

2024년 2월 사이 네이버의 점유율이 57.67%로 과반수 이상이었고 전체 포털사이트 중 블로그와 카페 기능을 제공하는 다음(DAUM)의 점유율 5.27%와 비교하여 약 11배로 월등히 높았기 때문이다. <그림 1>은 Internet Trend를 통해 2019년 2월부터 2024년 2월 사이의 포털사이트별 점유율을 확인하여 표로 재구성한 모습이다.

분석 대상 데이터는 2019년 2월 1일부터 2024년 1월 31일까지 5년간 생성된 데이터를 기준



〈그림 1〉 포털사이트별 점유율

으로 수집하였다. 도서관은 빠르게 급변하는 정보매체 및 데이터 기술의 발달에 따른 환경적 변화에 적응해야 할 뿐만 아니라 5개년을 기준으로 하는 독서문화진흥기본계획에 맞추어 도서관의 정책적 방향성과 비전을 수립하기 때문에 수집된 데이터가 지나치게 오래됐을 경우 자료의 최신성이 낮아 현황과 동떨어진 연구결과를 얻을 우려가 있다. 이를 고려하여 적절한 데이터를 충분히 수집하기 위해 기간을 5년으로 설정하였다.

먼저 제목과 원문에 수집 키워드가 포함된 블로그를 대상으로 1차 수집을 진행하였다. 수집은 Python을 프로그래밍 언어로 사용하는 Selenium 패키지를 활용하여 진행하였으며, 수집 키워드로 현재 운영 중인 대전시 공공도서관 26개관의 명칭을 사용하여 각각 수집하였다. 하지만 공공도서관명으로 블로그 데이터를 수집 시 근처 맛집, 아파트 분양 등의 관련성 없는 광고성 데이터가 다량 수집되는 문제점이 있었다. 이를 방지하기 위해 사전에 ‘식당’, ‘카페’,

‘맛집’, ‘아파트’, ‘분양’ 등의 키워드를 제외한 블로그 데이터만 수집하였고, 총 8,403개의 데이터를 수집하였다.

데이터 수집 결과 제외 키워드로 걸러내지 못한 광고성 데이터, 동일한 지명의 도서관에 대한 데이터(안산 도서관으로 검색 시 경기도 안산시의 도서관이 도출), 실제 도서관 이용에 대한 후기가 아닌 데이터(도서관명 단순 언급 등) 같이 연구에 활용하기 어려운 데이터를 최종 제외하기 위한 2차 수집을 진행하였다. 데이터의 정확성을 위해 직접 블로그를 일일이 확인하는 작업을 거쳐 관련 없는 데이터를 모두 제외하였고 총 1,729개의 데이터가 분석에 적합한 데이터로 선정되었다.

다음으로 분석에 사용하기 위해 동일한 의미를 가졌으나 여러 가지 형태로 나타나는 단어들을 하나로 합치는 과정을 거쳤다. 예로 들어 ‘장애인’과 ‘장애자’, ‘아쉽다’와 ‘아쉬움’ 등 의미가 같은 단어들을 병합하였다. 다만 ‘어린이’, ‘아이’, ‘유아’, ‘아기’는 병합하지 않고 분석을

진행하였는데 이는 유아는 보통 만 2세부터 5세까지를, 어린이는 더욱 포괄적인 연령대를 의미하기 때문이며 이에 따라 가오도서관, 판암도서관, 산성어린이도서관, 송촌도서관 등의 도서관도 유아와 어린이의 공간을 분리하고 있는 것을 확인할 수 있다. 아이와 아기 또한 연령대에서 차이가 있으며 다른 단어에 비해 부모 및 보호자와의 관계적인 호칭으로 사용되기 때문에 이용자의 특성이나 이용목적 등을 보다 세밀하게 파악할 수 있을 것이라 판단하여 분리하여 분석을 진행하였다. 또한 형태소 분석기에 학습되어 있지 않은 단어('북큐레이'를 '북큐레이션'으로 대체)을 정제하였으며, 불용어와 Null 값 및 의미 없는 특수문자 등의 단어들을 제거하는 과정을 거쳤다. 추가로, 분석의 정확도를 높이기 위해 데이터 전체 등장 횟수가 5번 미만인 키워드는 제외하여 분석을 진행하였다.

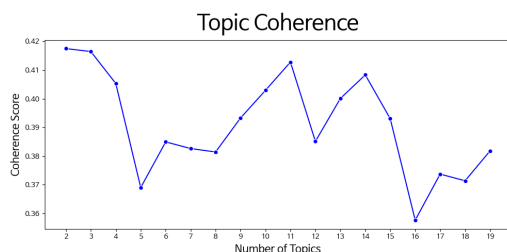
3.2 LDA 기반 토픽 모델링

토픽모델링은 문서에 등장하는 단어를 분류하고 각각의 토픽으로 묶어 데이터에 나타나는 주제에 대한 탐색과 분류를 수행하는 기법이다(Blei, 2012). 본 연구에서는 LDA 토픽모델링

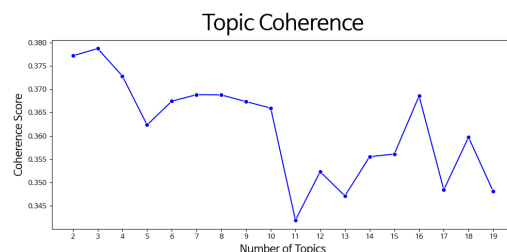
을 활용하여 총 1,729개의 블로그 텍스트 데이터 분석을 통해 이용자의 인식과 평가를 도출해보고자 하였다. 수집된 데이터는 동구 156건, 중구 570건, 서구 383건, 유성구 406건, 대덕구가 214건으로 조사되었다.

비지도 머신러닝의 특성상 사전에 물리적인 토픽 수(K)를 설정해야 하므로 LDA 모델링을 수행할 때 최적의 개수를 선택하는 것이 필수적이다. 그러나 토픽이 너무 적으면 제한적이고 일반적인 의미만 다를 위험이 있고, 토픽이 너무 많으면 각 토픽들이 서로 과도하게 많은 키워드를 공유하기 때문에 해석에 어려움을 겪을 수 있다. 주제 일관성(Topic Coherence)은 주제의 확률적 일관성을 평가하여 주제 모델의 품질을 정량화하기 위해 널리 채택되는 척도이며(Syed & Spruit, 2017), 이는 연구자가 최고의 결과를 보이는 최적의 토픽 수를 결정하는 데 유용하게 사용된다.

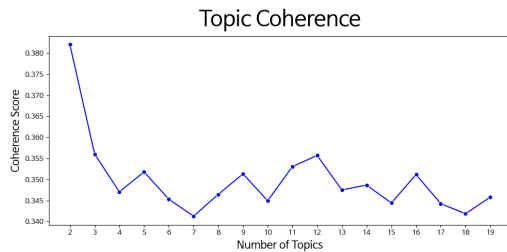
본 연구에서도 토픽 수의 타당성을 확보하기 위해 주제 일관성을 통해 군집 내의 일관성을 파악하였다. y축은 일관성 계수(Coherence Score)를 나타내며 1에 가까울수록 주제의 일관성이 높다는 것을 의미하고 x축은 K값을 나타낸다. 즉, 일관성 계수가 가장 높은 K값으로 토픽 수를 설정할 때 가장 적합한 결과를 얻을 수 있다.



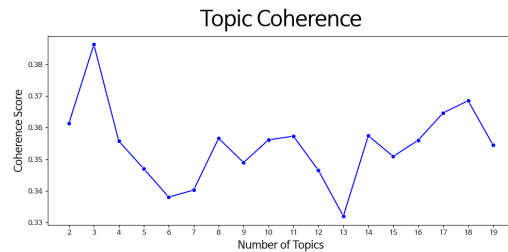
<그림 2-1> 주제 일관성(동구)



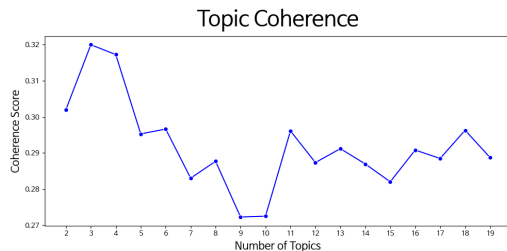
<그림 2-2> 주제 일관성(중구)



〈그림 2-3〉 주제 일관성(서구)



〈그림 2-4〉 주제 일관성(유성구)



〈그림 2-5〉 주제 일관성(대덕구)

이를 확인한 결과 중구, 유성구, 대덕구는 토픽이 세 개일 때 가장 주제 일관성이 높았으며 서구는 두 개일 때가 최적의 결과인 것으로 파악되었다. 마지막으로 동구는 2와 3이 아주 근소한 차이임이 확인되었기에 토픽을 세 개로 분류하였다.

4. 연구 결과

4.1 SNS 데이터 텍스트마이닝

전처리 등의 정제 과정을 거친 블로그 데이터의 키워드를 대상으로 빈도분석을 실시하였고 상위 100개 단어를 <표 4>와 같이 추출하였다. '어린이', '아이', '학생', '회원', '주민', '가족', '유아' 등의 이용 대상에 대한 단어와 '문화', '프

로그램', '행사', '수업', '강좌', '체험', '공연' 등 문화행사나 프로그램에 관련한 단어, '도서', '독서', '대출', '그림책', '반납' 등 도서 이용 관련 단어, '공간', '자료실', '홈페이지', '열람실', '주차', '교실' 등 온·오프라인 공간을 포함한 시설에 대한 단어와 같이 이용자, 방문 목적, 시설에 대한 정보가 포함된 이용 경험 데이터가 다양하게 수집되었음을 확인할 수 있었다.

또한, '어린이', '아이', '가족', '유아', '부모', '아기', '육아'와 같이 가정·보육에 관한 키워드와 '좋음', '가능', '다양', '행복' 등의 긍정적인 키워드도 다수 등장하였다. 이에 반해 부정적인 키워드는 발견되지 않았는데 이는 예상하지 못한 결과로, 타당성을 확인하기 위해 각 구별로 회원이 가장 많은 도서관을 선정하여 해당 기관의 사서에게 이용자 만족도 조사 여부와 결과에 대해 질의하였다. 그 결과 각 도서관이

〈표 4〉 SNS 데이터 빈도분석 결과

순위	단어	빈도	순위	단어	빈도	순위	단어	빈도	순위	단어	빈도
1	어린이	2839	26	영어	845	51	가족	549	76	제공	431
2	도서	2772	27	체험	834	52	통합	547	77	문의	428
3	독서	2495	28	참여	803	53	열람실	546	78	강연	427
4	좋음	2284	29	생각	801	54	유아	538	79	추천	427
5	아이	2253	30	활동	749	55	코로나	534	80	사용	424
6	교육	2036	31	휴관	732	56	내용	525	81	예약	422
7	대출	2028	32	공연	729	57	확인	509	82	대면	415
8	문화	1911	33	북스타트	722	58	부모	507	83	관심	411
9	신청	1873	34	회원	711	59	마음	505	84	평생	409
10	프로그램	1850	35	센터	707	60	위치	495	85	지원	404
11	학생	1492	36	정보	697	61	종합	491	86	창작	403
12	공간	1483	37	공공	684	62	친구	489	87	검색	399
13	자료실	1400	38	모집	667	63	아기	481	88	행복	399
14	가능	1373	39	반납	664	64	이동	478	89	육아	399
15	다양	1327	40	사람	661	65	작품	477	90	필요	398
16	안내	1184	41	강의	656	66	학년	468	91	선착순	391
17	행사	1178	42	접수	652	67	준비	465	92	놀이	390
18	작가	1066	43	방법	647	68	주제	461	93	목록	382
19	서비스	1036	44	이야기	621	69	다음	455	94	교실	380
20	홈페이지	955	45	주민	610	70	모습	452	95	마련	377
21	그림책	940	46	사진	608	71	특강	452	96	공지	368
22	지역	885	47	기간	566	72	무료	451	97	주차	365
23	대상	870	48	자료	565	73	그림	447	98	청소년	365
24	수업	870	49	가입	559	74	인문학	435	99	주말	364
25	강좌	848	50	학교	559	75	방학	433	100	뉴스	360

실시한 이용자 만족도 및 설문조사에서도 부정적인 의견이 거의 등장하지 않으며 대부분의 이용자가 만족하는 경향을 보인다는 답변이 공통적으로 수집되었다.

워드클라우드(Word Cloud)란 데이터 시각화의 한 방법으로 키워드 빈도수에 따라 크기에 차이를 두어 문서의 핵심 내용을 직관적으로 확인할 수 있는 유용한 도구이다(이재희, 배민철, 2021). 앞서 빈도분석 결과에서 확인했듯이 ‘어린이’, ‘도서’, ‘독서’, ‘좋음’, ‘아이’ 등 가장 높은 빈도로 검출된 단어들이 가시적으로 확인

되는 워드클라우드가 생성되었다. 이처럼 블로그 데이터에서 자주 등장하는 단어를 통해 어린이 연령대의 아이를 동반한 이용자와 도서대출이나 열람을 통한 독서가 주목적인 이용자가 데이터의 다수를 차지하고 있음을 확인할 수 있으며 그 영향으로 가정 및 보육과 관련한 키워드가 다수 등장한 것으로 판단된다.

4.2 SNS 데이터 토픽모델링

지금까지 SNS 데이터에 대한 키워드 분석



〈그림 3〉 SNS 데이터 워드클라우드(대전시 전체)

을 통해 대전시 공공도서관 이용자들의 인식과 주요 이용행태를 전반적으로 파악할 수 있었다. 연구에 적합하다고 판단한 데이터는 총 1,729건이며 그중 중구가 570건으로 가장 많은 수를 차지하였으며 그 다음으로 유성구 406건, 서구 383건, 대덕구 214건, 동구 156건이 수집되었다. 본 연구에서는 수집데이터에 대해 구별로 각각 토픽모델링을 실시하였다. 일관성계수로 파악한 결과 동구, 중구, 유성구, 대덕구는 3개의 토픽이, 서구는 2개의 토픽이 도출되었다. 구별 도서관의 구체적인 토픽은 〈표 5〉와 같다.

LDA의 알고리즘을 통해 모델이 자동적으로 도출한 토픽 내용을 검토한 후에, 토픽 간 유사성에 따라 네 개의 공통주제로 분류하였다. 먼저 분류 1은 독서 및 문화 프로그램과 문화행사에 관한 공통주제로 구성된 토픽들로서, '문화행사/프로그램'으로 명명하였다. 다음으로 분류 2는 도서 등 정보자원의 이용에 대한 공통주제로 묶였으며 '자료 이용'으로 명명하였다. 분류 3은 자료실, 열람실이나 도서관 시설 전반에

대한 공통주제를 가지고 있어 '물리적 환경 및 시설'로 명명하였다. 마지막으로 분류 4는 상호대차, 희망도서 바로대출 등 도서관에서 운영하고 있는 구체적 서비스에 관한 공통주제로, 이를 '도서관 서비스'로 판단하였다.

구체적으로 각 토픽을 내용의 유사성에 따라 분류한 결과는 다음과 같다. 먼저, 동구의 토픽 2(독서마라톤), 중구의 토픽 3(도서관 프로그램), 서구의 토픽 1(도서관 어린이 프로그램), 유성구의 토픽 2(육아 관련 프로그램)와 토픽 3(문화 행사), 대덕구의 토픽 1(가족 대상 프로그램)은 문화, 프로그램, 신청 등의 단어로 구성되어 있고 이는 도서관에서 진행하는 문화행사 및 프로그램에 관련한 단어들이기에 '문화행사/프로그램' 군집에 해당한다고 파악하였다. 다음으로 동구의 토픽 1(어린이 자료실 도서), 중구의 토픽 2(자료), 대덕구의 토픽 3(자료 이용)은 '도서', '독서', '자료', '자료실' 등의 유사한 단어를 포함하고 있어 이를 '자료 이용'으로 분류하였다. 또한 중구의 토픽 1(시설),

〈표 5〉 토픽모델링 결과

동구			중구			서구	
Topic 1 (어린이 자료실 도서)	Topic 2 (독서마라톤)	Topic 3 (북스타트)	Topic 1 (시설)	Topic 2 (자료)	Topic 3 (도서관 프로그램)	Topic 1 (도서관 어린이 프로그램)	Topic 2 (희망도서 동네 서점 바로대출)
아이	독서	북스타트	어린이	학생	프로그램	어린이	대출
독서	대면	신청	학생	교육	문화	신청	서점
어린이	지혜	좋은	독서	도서	다양	프로그램	도서
자료실	교육	독서	도서	자료	강좌	홈페이지	희망
영어	마라톤	접수	공간	그림책	체험	행사	반납
좋은	완주	참여	안내	장애인	창작	안내	동네
성교육	기록	그림책	자료실	안내	사진	가능	회원
과학	독서	꾸러미	서비스	다양	신청	문화	가능
가능	참가	아기	좋은	지원	접수	좋은	신청
도서	코스	부모	주차	정보	공연	다양	안내

유성구			대덕구		
Topic 1 (자료실 및 열람실)	Topic 2 (육아 관련 프로그램)	Topic 3 (문화 행사)	Topic 1 (가족 대상 프로그램)	Topic 2 (시설)	Topic 3 (자료 이용)
도서	신청	문화	문화	도서	도서
좋은	아이	신청	독서	어린이	대출
자료실	육아	도서	공연	자료실	독서
공간	아가	가능	어린이	다양	교육
생각	어린이	회원	좋은	좋은	자료실
어린이	출산	작가	행사	아이	반납
열람실	부모	홈페이지	아이	공간	신청
독서	수업	접수	가족	종합	가능
대출	보육	행사	프로그램	열람실	다양
아이	좋은	공연	신청	홈페이지	좋은

유성구의 토픽 1(자료실 및 열람실), 대덕구의 토픽 2(시설)는 ‘자료실’, ‘열람실’, ‘공간’ 등의 동일한 단어로 구성되었고 ‘자료 이용’ 분류와는 다르게 자료실에 국한된 단어가 아닌 도서관 공간에 대한 주제를 다루고 있어 ‘물리적 환경 및 시설’로 분류하였다. 마지막으로 동구의 토픽 3(북스타트), 서구의 토픽 2(희망도서 동네서점 바로대출)는 ‘북스타트’와 ‘희망’, ‘도서’, ‘동네’, ‘서점’으로 각각 특정한 도서관 서비스에 대한 단어로 이루어져 있기 때문에 ‘도서관 서비스’라고 분류하였다.

위와 같은 분석을 통해 SNS 상에서 대전시 공공도서관 이용자의 인식 및 평가가 문화행사/프로그램, 자료 이용, 물리적 환경 및 시설, 도서관 서비스 등의 요인에서 이뤄지고 있음이 도출되었다. 또한 이는 앞서 설문을 통해 이용자의 인식 및 평가를 조사한 선행연구에서 사용한 문항들과 장서 및 자료 현황, 도서관 서비스, 독서 프로그램 및 문화행사, 공간 및 시설의 요인이 유사하며, 따라서 본 연구가 기존 선행연구의 연구 목적을 새로운 방법론으로 탐색하고 있음을 보여준다.

동구는 이질적인 키워드인 '대면'과 '지혜'가 높은 빈도로 등장하였는데, 이는 동구의 공공도서관이 통합하여 진행한 행사인 '동구 행BOOK한 독서마라톤대회'의 영향으로 보인다. 독서마라톤이란 독서활동을 마라톤에 적용하여 책 1쪽을 2m로 환산해 독서량으로 완주가 인정되는 행사로, 코로나 이후 대면으로 진행되었고 가장 긴 코스인 지혜 코스에 대한 관심도가 높았던 것으로 해석된다.

중구는 '어린이'와 '아이' 등이 주로 등장하는 자치구들과 다르게 '학생'이 가장 높은 빈도로 등장하였으며, 이는 함께 나타난 '교육' 키워드와 조합하여 판단하였을 때 중구에 위치한 학생 및 학교 중심의 교육 문화 활동을 지원하는 도서관인 학생교육문화원에서 파악된 것으로 보인다.

서구는 자치구들 중 '어린이'가 가장 가시적으로 도출되었으며 이는 서구어린이도서관의 영향도 있으나, 2024년 3월 기준 5세에서 9세 사이 인구가 18,614명으로 서구의 어린이 인구가 가장 높은 것이 원인으로 파악된다.

유성구는 키워드의 빈도가 가장 고르게 분산되어 있는 것이 확인되며 이는 공공도서관 수가 9개로 자치구 중 가장 많은 도서관을 보유하고 있어 키워드가 다양하게 나타난 것으로 해석된다. 추가적으로 중구의 데이터 570건 중 한밭도서관이 약 56%인 322건을 차지한 것과 다르게 유성구는 406건 중 가장 많은 도서관의 데이터가 약 21%(86건)으로 도서관별 데이터 수의 차이가 상대적으로 적은 것으로 나타났다.

마지막으로 대덕구는 '독서'와 '문화' 키워드의 빈도가 가장 높은 것으로 파악되었으며 이

는 대덕구 도서관의 문화학교, 문화가 있는 날, 문화강좌 등 문화 프로그램에 대한 시민의 관심이 집중된 결과로 보인다. 또한 도서관과 유사한 기능을 제공하는 기관인 대덕문화원과 대덕구 중리근린공원 복합문화센터가 함께 언급되었을 가능성도 추측될 수 있다.

4.3 뉴스 기사 텍스트마이닝

앞선 소셜미디어에서 대전시 공공도서관 이용자의 이용행태를 비롯한 인식을 알아보았다면 이와 연계하여 대전시 도서관에 대한 사회적 논의와 새로운 이슈를 파악할 필요가 있다. 따라서 온라인뉴스 기사라는 매체를 활용하여 뉴스기사의 텍스트데이터를 수집하고 그 키워드로 빈도분석을 실시하였다. 수집은 한국언론진흥재단에서 운영하는 뉴스 데이터베이스 및 뉴스 분석 시스템 빅카인즈를 활용하여 진행하였다. 공공도서관에 국한되지 않은 포괄적인 대전시의 도서관 관련 이슈에 대해 살펴보고자 '대전 AND 도서관' 검색식을 사용하여 종합적인 데이터를 얻고자 하였고 그 결과 총 1,801건이 색인되었으며 총 348건의 문서가 주제와 적합하다고 판단되었다.

대전시 도서관에 대한 뉴스기사 분석 결과 첫째로 블로그 데이터의 빈도분석 결과와 유사한 키워드가 상위 빈도로 도출되었다. 가장 빈번하게 나타난 '독서', '문화', '학생', '교육', '프로그램', '도서', '공간' 등의 키워드를 통해 이용자가 실제로 인식하고 이용하는 분야와 뉴스를 통해 논의되고 있는 분야가 일부 일치함을 확인할 수 있었다.

두 번째로 '사업', '조성', '진행', '지원', '건립',

‘추진’ 등 도서관 사업에 관한 키워드가 상위 키워드로 도출되었음을 확인할 수 있었다. 앞선 선행연구 결과 대전시의 도서관 수 부족이 지속적으로 제기되었음을 확인하였고 이에 따라 도서관의 추가 건립 예정이 핵심적인 현안으로 논의된 것으로 파악된다. 또한 건립 예정, 또는 건립이 논의되고 있는 추가 기관에 대한 키워드가 ‘국회도서관’, ‘동대전시립도서관’, ‘제2대전문학관’ 등으로 발견되었다. ‘국회도서관’ 키워

드의 정식명칭은 ‘국회통합디지털센터’로 유성구 갑천호수공원에 건립이 예정되어있는 디지털 도서관 겸 디지털 체험 및 전시 등이 이뤄질 시민개방 복합문화공간이다. ‘동대전시립도서관’은 동구 가양동의 기존 가양도서관 부지에 조성될 대전의 두 번째 시립도서관으로 2024년 9월 준공을 앞두고 있다. ‘제2대전문학관’은 동구 용전동에 위치한 기존 대전문학관의 수장고 부족 문제로 촉발되어 추진된 사업으로, 자료 수

〈표 6〉 뉴스 기사 빈도분석 결과

순위	단어	빈도	순위	단어	빈도	순위	단어	빈도	순위	단어	빈도
1	독서	781	26	시민	142	51	정책	91	76	문화원	70
2	문화	538	27	홈페이지	137	52	주제	90	77	강좌	70
3	학생	418	28	계획	133	53	규모	89	78	청소년	70
4	교육	395	29	전시	126	54	개최	88	79	확대	69
5	프로그램	363	30	국회도서관	124	55	평가	85	80	문의	69
6	도서	296	31	신청	117	56	시작	84	81	관심	69
7	공간	290	32	디지털	116	57	그림책	84	82	사회	69
8	학교	278	33	가능	115	58	가족	84	83	서점	69
9	사업	276	34	자료	114	59	학습	81	84	의원	67
10	조성	274	35	전국	111	60	중심	80	85	방문	66
11	진행	258	36	사진	109	61	노력	79	86	인문학	66
12	지원	253	37	수업	105	62	예술	79	87	산성	65
13	건립	244	38	마련	102	63	동대전 시립도서관	79	88	관리	65
14	공공도서관	242	39	대표	101	64	미래	77	89	부지	65
15	어린이	205	40	주민	101	65	모집	77	90	가을	65
16	추진	202	41	대출	101	66	문학	75	91	생각	64
17	제공	192	42	환경	100	67	연계	75	92	활성	63
18	작가	179	43	한밭도서관	99	68	대전교육청	74	93	콘텐츠	62
19	체험	175	44	예산	99	69	협력	73	94	시민들	62
20	예정	169	45	전입	98	70	공모	73	95	준비	62
21	활용	162	46	영어	97	71	구성	72	96	접수	61
22	작은도서관	153	47	통합	95	72	개관	72	97	작품	60
23	서비스	152	48	정보	94	73	복지	72	98	대전학생 교육문화원	60
24	복합	151	49	공연	93	74	행정	71	99	행복	60
25	시설	148	50	구축	92	75	제2대 전문학관	70	100	인프라	60

장뿐이 아닌 라키비움 및 시민복합문화공간으로 중구 대흥동에 개관될 예정이다. 이를 통해 도서관 추가 건립에 대한 논의가 지속되어 왔으며 필수적인 사업으로 인식되고 있음을 알 수 있다.

세 번째로 주목할 만한 키워드는 ‘교육’, ‘공간’, ‘복합’, ‘문화’, ‘시설’, ‘복지’, ‘행정’ 등으로, 사회적으로 논의되고 인식되는 도서관의 기능에 대한 키워드가 확장되고 있음을 살펴볼 수 있다. 또한 추가로 등장하는 ‘연계’, ‘협력’ 등의 키워드를 통해 도서관 역할 및 기능의 확대를 위한 도서관 협력 및 연계 체계가 논의되고 있음을 파악할 수 있다.

대전시 공공도서관에 대한 뉴스 기사 텍스트 데이터를 활용하여 워드클라우드를 생성한 결과물이 <그림 5>와 같이 도출되었다. 앞서 언급하였듯 SNS 데이터에 비해 상대적으로 다양한 키워드가 직관적으로 확인되며, 특히 공공도서관의 추가 건립 사업에 대한 키워드와 협력 체계가 사회적 논의의 주요 과제로 논의되고 있음을 파악할 수 있다.

5. 결 론

이용자 친화적인 도서관을 조성하기 위해 필요한 것은 이용 당사자의 수요에 맞는 서비스를 제공하는 것이다. 이용자가 직접 이용 후 자신의 입장에서 작성한 SNS 데이터를 바탕으로 이용자의 인식과 평가를 분석하는 연구 방법론은 이미 각계의 분야에서 활용되고 있지만 이를 도서관에 접목시킨 연구는 미미하다는 것을 확인하였다. 이러한 문제의식 하에 본 연구에서는 도서관 이용 후기가 포함된 SNS 텍스트와 뉴스 기사 데이터를 분석하여 도서관 전반에 대한 인식과 평가를 알아보고자 연구를 수행하였다. 분석 결과를 요약하면 다음과 같다.

첫째, SNS에서 아이를 동반한 이용자와 도서 대출이나 열람을 통한 독서가 주목적인 이용자의 데이터가 비중 있게 등장하였다. 둘째, LDA 토픽모델링 결과 토픽들은 크게 ‘문화행사/프로그램’, ‘자료 이용’, ‘물리적 환경 및 시설’, ‘도서관 서비스’로 나타났으며 이는 선행연구에서 이용자 인식 및 평가를 위해 수집한 요



<그림 5> 뉴스 기사 워드클라우드

인과 유사하다. 셋째, 뉴스 기사 데이터에서 도서관 및 복합문화공간 추가 건립과 도서관 협력체계 키워드가 핵심적으로 도출되었다.

이러한 결과를 종합하여 대전시 공공도서관의 양적, 질적 향상을 위한 구체적인 내용을 제안하고자 한다. 첫째, 육아 및 보육 사업을 수행하는 여러 기관들과의 업무 협약을 통해 사회적 육아공동체 네트워크를 조성하여야 한다. SNS 텍스트마이닝 결과 어린이 동반 이용자의 양적 비중이 높다는 것이 확인되었고, 가족 프로그램, 육아 프로그램, 어린이 프로그램 등 가정과 보육에 대한 프로그램에 관한 선호 또한 높게 나타났다. 이에 덧붙여 뉴스 기사 분석을 통해 도서관의 기능 확대와 협력 체계에 대한 키워드가 논의되고 있음을 도출하였다. 이는 '교류 협력의 플랫폼 기능 강화', '경계를 넘는 서비스의 연계', '도서관 협력체계의 강화'를 핵심과제로 포함하고 있는 제3차 도서관발전종합계획(2019-2023)을 바탕으로 도서관 정책이 수립된 결과로 파악된다. 따라서 영유아를 둔 부모의 육아 및 보육 정보에 대한 접근성을 강화하고 후속세대인 자녀들의 독서입문을 위한 생애주기별 맞춤형 도서관 서비스가 가능하도록 유치원, 학교 등 보육 및 교육기관과의 협력 및 연계가 필요하다.

둘째, 지역 균형을 고려한 도서관 건립이 계획되어야 한다. 대전시는 분석을 통해 키워드로 도출된 기관 외에도 서구 관저동에 제3 시립

도서관의 건립을 예정하고 있다. 대전시가 추가 건립 예정인 도서관 및 복합문화공간의 양적 증가는 긍정적이지만 이는 지속적으로 제기되었던 중구와 대덕구의 상대적 소외 문제가 심화되는 결과를 야기한다. 도서관 현황 조사에서 중구와 대덕구에 가장 적은 수의 도서관이 설립되어 있음이 확인되었다. 따라서 추후 공공도서관의 건립 위치와 입지조건을 선정하는 과정에서 지역의 균형적인 발전을 고려하여 소외 지역 없이 고른 인프라 확충이 계획되어야 할 필요가 있다.

본 연구는 SNS와 뉴스기사 데이터를 활용한 텍스트마이닝을 통해 기존의 설문조사를 사용한 이용자 인식 및 평가의 한계점을 보완하고 새로운 방법론을 제안하였다는 의의가 있다. 그러나 네이버 블로그로 SNS 데이터를 한정하여 다양한 매체를 활용하지 못했다는 제한점이 있으며 대전시에 한정된 연구이기에 광범위하게 적용하기엔 어렵다는 아쉬움이 있다. 하지만 이는 후속 연구에 사용할 수집 데이터의 채널을 다양화하고 연구범위를 확장함으로써 해결할 수 있을 것이다. 추후에는 SNS 데이터 중 텍스트를 활용한 분석뿐 아니라 시장 점유율이 높은 이미지와 영상 기반의 SNS를 분석에 활용하거나 SNS의 '좋아요'와 리그램 등의 상호작용을 통한 데이터를 포함한다면, 더 사실적이고 풍부한 연구가 가능할 것으로 보인다.

참 고 문 헌

- 김신영, 차성중 (2022). 인구감소지역 공공도서관 서비스를 위한 이용자 인식조사 연구: 경북 의성군을 중심으로. 한국도서관·정보학회지, 53(3), 95-117.
<https://doi.org/10.16981/kliss.53.3.202209.95>
- 김영석 (2021). 우리나라 공공도서관의 이용변화 추이 분석 및 대응방안 연구. 한국도서관·정보학회지, 52(2), 379-400. <https://doi.org/10.16981/KLISS.52.2.202106.379>
- 노영희, 장인호, 박우정 (2024). 과천시 도서관의 효과적인 운영 방안 수립을 위한 이용자 인식조사 연구. 한국비블리아학회지, 35(1), 73-96. <https://doi.org/10.14699/kbiblia.2024.35.1.073>
- 대전광역시 (2024. 04. 01). 2024년 3월말 주민등록 인구현황. 대전의 통계.
출처: https://www.daejeon.go.kr/sta/StaStatisticsFldView.do?ntatcSeq=1452921305&menuSeq=180&column1Cont=&column2Cont=&boardId=normal_0009&pageIndex=1
- 서혜선 (2023). SNS 텍스트 데이터 기반 대전시 환경 관련 토픽 모델링 및 감성 분석. 한국자료분석학회, 25(3), 949-958. <http://dx.doi.org/10.37727/jkdas.2023.25.3.949>
- 안영웅, 임운택 (2023). 대전광역시 공공도서관 입지의 공간적 형평성 분석. 한국콘텐츠학회논문지, 23(5), 607-617. <https://doi.org/10.5392/jkca.2023.23.05.607>
- 이수영, 양지안, 남재우, 송민선 (2023). 도봉구 공공도서관 중장기 발전계획 수립을 위한 지역주민 인식 연구. 한국비블리아학회지, 34(4), 183-205.
- 이욱, 임혜원, 여하림, 황혜선 (2021). 인공지능 서비스에 대한 온라인뉴스, 소셜미디어, 소비자리뷰 텍스트마이닝. Human Ecology Research, 59(1), 23-43. <https://doi.org/10.6115/fer.2021.003>
- 이재희, 배민철 (2021). 워드클라우드를 활용한 도시재생사업 인식변화 트렌드 분석. 차세대융합기술학회논문지, 5(2), 238-245. <http://dx.doi.org/10.33097/JNCTA.2021.05.02.238>
- 이주경, 손용훈 (2021). 구글맵리뷰 텍스트마이닝을 활용한 공원 이용자의 인식 및 평가: 서울숲, 보라매공원, 올림픽공원을 대상으로. 한국조경학회, 49(4), 15-29.
<https://doi.org/10.9715/kila.2021.49.4.015>
- 이지원, 유기윤, 김지영 (2019). SNS 데이터를 이용한 공공시설 매력도지수에 따른 접근성 분석기법. 한국측량학회지, 37(1), 29-42. <https://doi.org/10.7848/ksgpc.2019.37.1.29>
- 임성관 (2024). 강동구 도서관에 대한 지역 주민과 도서관 관계자의 인식 조사 연구. 한국문헌정보학회지, 58(1), 53-82. <https://doi.org/10.4275/KSLIS.2024.58.1.053>
- 장인호, 권선희 (2024). 공공도서관 이용자의 인식조사 연구. 한국휴면이미지디자인, 6(1), 71-91.
- 장인호, 황금숙, 송민선 (2022). 안양시 공공도서관 중장기 발전 계획 수립을 위한 현황 분석 및 운영 전략 연구. 정보관리학회지, 39(1), 145-170. <https://doi.org/10.3743/KOSIM.2022.39.1.145>

- 전경선, 신동희 (2022). 지역사회 특성을 반영한 공공도서관 활성화 방안 연구: 경기도 양평군을 중심으로. 한국비블리아학회지, 33(1), 403-423. <https://doi.org/10.14699/kbiblia.2022.33.1.403>
- 정유경, 박성희, 권선영, 김보일, 구정화 (2021). 공공도서관에 대한 지역주민들의 인식과 요구 분석: 대전지역 공공도서관 이용자 및 비이용자를 중심으로. 한국문헌정보학회지, 55(3), 239-258. <https://doi.org/10.4275/KSLIS.2021.55.3.239>
- 행정안전부 (2024. 04. 02). 대전광역시_행정구역 현황. 공공데이터포털.
출처: <https://www.data.go.kr/data/15062128/fileData.do>
- Blei, D. M. (2012). Probabilistic topic models. Communications of the ACM, 55(4), 77-84. 10.1145/2133806.2133826
- Dewinta, A. & Irawan, M. I. (2021). Customer complaints clusterization of government drinking water company on social media twitter using text mining. 2021 3rd East Indonesia Conference on Computer and Information Technology, 338-342.
- Luo, S. & He, S. Y. (2021). Understanding gender difference in perceptions toward transit services across space and time: a social media mining approach. Transport Policy, 111, 63-73. <https://doi.org/10.1016/j.tranpol.2021.07.018>
- Nilashi, M., Ahmadi, H., Arji, G., Alsalem, K. O., Samad, S., Ghabban, F., Alzahrani, A. O., Ahani, A., & Alarood, A. A. (2021). Big social data and customer decision making in vegetarian restaurants: A combined machine learning method. Journal of Retailing and Consumer Services, 62, 102630. <https://doi.org/10.1016/j.jretconser.2021.102630>
- Oyebode, O., Ndulue, C., Adib, A., Mulchandani, D., Suruliraj, B., Orji, F. A., Chambers, C. T., Meier, S., & Orji, R. (2021). Health, psychosocial, and social issues emanating from the COVID-19 pandemic based on social media comments: text mining and thematic analysis approach. JMIR medical informatics, 9(4), e22734.
- Syed, S. & Spruit, M. (2017). Full-text or abstract? Examining topic coherence scores using latent dirichlet allocation. 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Tokyo, Japan, 165-174.

• 국문 참고자료의 영어 표기

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

- Ahn, Yeoung-Woong & Leem, Yountaik (2023, May 31). Analysis of spatial equity in the location of public library in Daejeon metropolitan city. Journal of the Korea Contents Association, 23(5), 607-617. <https://doi.org/10.5392/jkca.2023.23.05.607>

- Chang, Inho & Kwon, Sunhee (2024). A study on user perception survey for public library. *The Korean Human Image Design*, 6(1), 71-91.
- Chang, Inho, Hoang, Gum-sook, & Song, Min Sun (2022). A study on the current status analysis and operation strategies for establishing the mid- to long-term development plans for Anyang municipal libraries. *Journal of the Korean Society for Information Management*, 39(1), 145-170. <https://doi.org/10.3743/KOSIM.2022.39.1.145>
- Daejeon Metropolitan City (2024, 04. 01). Resident registration population status at the end of March 2024. Statistics of Daejeon. Available: https://www.daejeon.go.kr/sta/StaStatisticsFldView.do?ntatcSeq=1452921305&menuSeq=180&column1Cont=&column2Cont=&boardId=normal_0009&pageIndex=1
- Jeon, Kyungsun & Shin, Donghee (2022). Research on revitalizing public libraries reflecting community characteristics in Yangpyeong-gun. *Journal of the Korean Biblia Society for Library and Information Science*, 33(1), 403-423. <https://doi.org/10.14699/kbiblia.2022.33.1.403>
- Jeong, Yoo-Kyung, Park, Sung-Hee, Kwon, Sunyoung, Kim, Bo Il, & Koo, Joung-Hwa (2021). Citizens' perception, demands and expectations of their public library services in Daejeon. *Journal of the Korean Society for Library and Information Science*, 55(3), 239-258. <https://doi.org/10.4275/KSLIS.2021.55.3.239>
- Kim, Sin-Young & Cha, Sung-Jong (2022) A study on the users perception of public library services in depopulation areas: focusing on Uiseong-gun. *Journal of Korean Library and Information Science Society*, 53(3), 95-117. <https://doi.org/10.16981/kliss.53.3.202209.95>
- Kim, Young-Seok (2021). A study on the changes in the use of public libraries in Korea and countermeasures. *Journal of Korean Library and Information Science Society*, 52(2), 379-400. <https://doi.org/10.16981/KLISS.52.2.202106.379>
- Lee, Jaehee & Bae, Mincheul (2021). A study on trend analysis of change in perception of urban regeneration projects applying word cloud. *The Journal of Next-generation Convergence Technology Association. International Next-Generation Convergence Technology Association*. <https://doi.org/10.33097/jncta.2021.05.02.238>
- Lee, ji won, Yu, Kiyun, & Kim, Jiyoung (2019). Accessibility analysis method based on public facility attraction index using SNS data. *Journal of the Korean Society of Survey, Geodesy, Photogrammetry, and Cartography*, 37(1), 29-42. <https://doi.org/10.7848/ksgpc.2019.37.1.29>
- Lee, Ju-Kyung & Son, Yong-Hoon (2021). Perception and appraisal of urban park users using

- text mining of Google maps review: cases of Seoul forest, Boramae park, Olympic park. *Journal of the Korean Institute of Landscape Architecture*. Korean Institute of Landscape Architecture, 49(4), 15-29. <https://doi.org/10.9715/kila.2021.49.4.015>
- Lee, Su Young, Yang, Ji-Ann, Nam, Jaewoo, & Song, Min Sun (2023). A study on perceptions of users for establishing mid-to long term development plan for public libraries in Dobong-gu. *Journal of the Korean Biblia Society for Library and Information Science*, 34(4), 183-205.
- Li, Xu, Lim, Hyewon, Yeo, Harim, & Hwang, Hyesun (2021). Text mining of online news, social media, and consumer review on artificial intelligence service. *Human Ecology Research*, 59(1), 23-43. <https://doi.org/10.6115/fer.2021.003>
- Lim, Seong Kwan (2024). A study on the perception survey of local residents and library officials about Gangdong-gu libraries. *Journal of the Korean Society for Library and Information Science*, 58(1), 53-82. <https://doi.org/10.4275/KSLIS.2024.58.1.053>
- Ministry of the Interior and Safety (2024.02) Current Status of Daejeon Metropolitan City_Administrative District. Public Data Portal. Available: <https://www.data.go.kr/data/15062128/fileData.do>
- Noh, Younghee, Chang Inho, & Kwak, Woojung (2024). A study on user perception survey for establishing effective operational strategies of the Gwacheon city library. *Journal of the Korean Biblia Society for Library and Information Science*, 35(1), 73-96. <https://doi.org/10.14699/kbiblia.2024.35.1.073>
- Suh, Hyesun (2023). Topic modeling and sentiment analysis based on SNS text data related to the environment in Daejeon. *Journal of the Korean Data Analysis Society*, 25(3), 949-958. <http://dx.doi.org/10.37727/jkdas.2023.25.3.949>