

공공도서관 미대출 도서 추천시스템 구현 : 대구 D도서관을 중심으로

진민하¹, 정승연², 조은지³, 이명훈⁴, 김건욱^{5*}

¹한동대학교 경영학/데이터사이언스학과 학부과정, ²경북대학교 통계학과 학사졸업, ³영남대학교 컴퓨터공학과 학사졸업
⁴경북대학교 수학과 학사졸업, ⁵대구디지털산업진흥원 빅데이터활용센터 센터장

Implementation of the Unborrowed Book Recommendation System for Public Libraries: Based on Daegu D Library

Min-Ha Jin¹, Seung-Yeon Jeong², Eun-Ji Cho³, Myoung-Hun Lee⁴, Keun-Wook Kim^{5*}

¹Undergraduate, School of Management/Data Science, Handong Global University

²Bachelor's degree, Department of Statistics, Kyungpook National University

³Bachelor's degree, Department of Computer Science, Yeungnam University

⁴Bachelor's degree, Department of Mathematics, Kyungpook National University

⁵Director, Big Data Center, Daegu Digital Industry Promotion Agency

요 약 국내 공공도서관의 역할과 기능은 다양해지고 있는 반면, 내부적으로는 편향된 도서 대출로 다양한 문제들이 나타나고 있다. 또한 최근 4차 산업혁명으로 공공도서관에서 인기도서 위주의 도서 추천시스템이 도입되고 있으나, 이용자가 접할 수 있는 도서의 다양성은 제한되고 있다. 이에 본 연구에서는 공공도서관 이용자의 만족을 제고하기 위해 공간적으로는 대구시 두류도서관으로 한정하여 대출이력 자료(213,093건), 회원정보(35,561명) 등을 활용하여 군집분석과 토픽 모델링, 콘텐츠 기반 필터링 추천 알고리즘으로 공공도서관 미대출 도서 추천시스템을 구현하였으며, 이에 대한 실제 사용자들의 만족도 설문조사를 실시하여 미대출 도서 추천시스템의 가능성과 시사점을 제시하였다. 분석 결과 대다수의 이용자들이 높은 만족도로 응답하였으며, 특정 성·연령대, 직업, 평소 독서량 등으로 분류된 계층에서 만족도가 상대적으로 높게 나타난 것을 확인할 수 있었다. 본 연구결과를 통해 공공도서관의 편향된 도서 대출, 운영 효율성 저하 등의 문제를 일부 개선할 수 있을 것으로 기대하며, 연구의 한계점 또한 제시하였다.

주제어 : 공공도서관, 미대출 도서, 추천시스템, 콘텐츠 기반 필터링, 융합, 토픽 모델링

Abstract The roles and functions of domestic public libraries are diversifying, but various problems have emerged due to internally biased book lending. In addition, due to the 4th Industrial Revolution, public libraries have introduced a book recommendation system focusing on popular books, but the variety of books that users can access is limited. Therefore, in this study, the public library unborrowed book recommendation system was implemented limiting its spatial scope to Duryu Library in Daegu City to enhance the satisfaction of public library users, by using the loan records data (213,093 cases), user information (35,561 people), etc. and utilizing methods like cluster analysis, topic modeling, content-based filtering recommendation algorithm, and conducted a survey on actual users' satisfaction to present the possibility and implications of the unborrowed book recommendation system. As a result of the analysis, the majority of users responded with high satisfaction, and was able to find the satisfaction was relatively high in the class classified by specific gender, age, occupation, and usual reading. Through the results of this study, it is expected that some problems such as biased book lending and reduced operational efficiency of public libraries can be improved, and limitations of the study was also presented.

Key Words : Public Library, Unborrowed book, Recommendation System, Content-based filtering, Convergence, Topic Modeling

*Corresponding Author : Keun-Wook Kim(aut7767@dip.or.kr)

Received February 3, 2021

Accepted May 20, 2021

Revised March 12, 2021

Published May 28, 2021

1. 서론

국내 공공도서관은 1931년 ‘도서관 법’ 제정 이후 양적인 발전을 이루는 계기를 가지게 되었으며, 이후 지속적으로 증가하여 지난 40년간 약 27배 증가하였다. 하지만 해외 선진국들과 비교 시 아직도 절대적인 수준에는 미치지 못하고 있어, 정부 차원에서 공공도서관 수 확대를 앞으로도 지속적으로 추진할 것으로 판단된다[1].

또한 최근 4차 산업혁명으로 IT기술의 발전과 공공도서관 이용자들의 다양한 수요가 증가하여 기존의 공공도서관의 역할과 기능이 확대되어가고 있다. 이는 기존의 공공도서관의 역할인 도서 대출의 기능을 넘어서 문화센터, 전시회, 영화관람, 지역 커뮤니티 등의 이용행위가 이루어지면서 거주민들의 삶과 밀접한 관계를 가지게 되었다[2].

이러한 공공도서관의 양적성장과 역할의 다양성 증가로 중요성은 증가하고 있지만, 내부적으로 편향된 도서 대출 등의 문제가 존재하고 있으며, 이로 인해 특정 인기 도서의 대출 집중, 도서관 이용 활성화 저하, 장서 포화 등의 문제가 나타나고 있다[2]. 이를 해결하기 위해 공공도서관에서는 정기적으로 큐레이터를 활용한 도서 추천과 담당자 주관에 의한 이용 활성화 정책수립을 하고 있으나 많은 한계점이 존재하고 있다.

빅데이터, 인공지능 분야에서는 가장 성공한 분야로 추천시스템을 많은 전문가들이 언급하고 있으며, 이는 특정회원과 유사한 성향을 가진 회원을 찾아 아직 경험하지 못한 서비스를 제공함으로써 시장에서 높은 가치 창출을 하고 있다. 이를 서비스화 하고 있는 기업으로는 유튜브, 아마존, 교보문고 등이 있으며[3], 이러한 기술의 발전은 공공도서관에서도 점차 도입이 될 것으로 판단된다.

이러한 배경 하에, 본 연구에서는 공공도서관에서 주된 문제인 편향된 도서 대출의 문제를 빅데이터, 인공지능 분야에서 각광받고 있는 추천시스템 알고리즘을 활용하여 접근하고자 한다. 이를 위해 2019년 1월부터 2020년 7월까지의 공공도서관의 장서/대출목록 자료(272,652건)와 대출이력 자료(213,093건), 회원정보(35,561명) 그리고 도서관 정보나루 내 도서 상세정보 자료(272,652건)를 결합하여 공공도서관 미대출 도서 추천시스템을 구현하고자 한다.

개인 맞춤형 도서 추천시스템을 구현하기 위해 다양한 선행 연구들이 진행되어 왔으며, 대다수 선행 연구에서 협업 필터링 알고리즘, 콘텐츠 기반 필터링 알고리즘, 하이브리드 알고리즘 등을 활용하여 구현하였다. 이는 개

인별 도서 대출이력 자료를 활용하여 머신러닝 기반의 과학적인 추천시스템을 구현한 장점은 있으나, 대다수 대출빈도가 높은 도서를 대상으로 추천을 하여 다양한 도서를 추천하기에는 한계를 가지고 있다.

따라서 본 연구에서는 공공도서관 도서 대출이력을 기반으로 Spherical K-means 클러스터링과 LDA 기반의 토픽모델링을 이용하여 성·연령대별 인기 카테고리를 추출한 후, 콘텐츠 기반 필터링에서의 일대다 대응을 통해 해당 카테고리 내 인기도서와 유사한 특성을 지닌 미대출 도서를 추천해주는 시스템을 구현하고, 이에 대한 만족도를 평가함으로써 미대출 도서 추천시스템의 가능성을 살펴본 점이 선행 연구와의 차별성이라 할 수 있다.

이러한 분석 결과는 공공도서관을 이용하는 거주민들이 다양한 지식과 이념을 학습하는데 기여할 것으로 판단되며, 공공도서관 내 편향된 도서 대출의 완화와 이용 활성화에도 기여할 수 있을 것으로 보인다.

2. 선행 연구 고찰

2.1 도서 추천시스템 관련 선행 연구

국내 도서 추천시스템과 관련된 연구들은 대부분 협업 필터링 또는 콘텐츠 기반 필터링의 도서 추천시스템을 주제로 진행되어 왔으며[4-8], 국외 연구들의 경우에는 콘텐츠 기반 필터링 또는 하이브리드 도서 추천시스템을 주제로 대다수의 연구가 진행되어 왔다[9,10].

이를 세부적으로 살펴보면 조현양(2017)은 국립어린이청소년도서관 내 어린이·청소년을 위한 추천 도서의 도서 소개에 대하여 LIBLINEAR, LibSVM 모델을 활용하여 서평 기반의 자동분류시스템을 개발하였으며, 이를 기반으로 이용자의 개인별 성향이나 상황에 적합한 추천 시스템을 제안하였다[4].

정영진, 조윤호(2017)는 국내 대형 온라인 서점의 고객 트랜잭션 데이터 및 도서 소개, 목차 정보에 대하여 토픽 모델링 기법을 활용하여 토픽 주제에 대한 선호도, 도서구입의 최신성과 가격 수용도 등의 사용자 선호를 중심으로 하는 협업 필터링 기반의 도서 추천시스템을 구현하였다[5].

정윤서(2018)는 공공도서관 장서/대출 데이터, 인기 대출도서를 키워드 분석, 네트워크 분석을 활용하여 분석하였으며, 분석 결과 개인의 독서경향을 파악함으로써 평소에 읽지 않던 분야 내 인기 대출도서를 추천하는 시스템을 제안하였다[6].

최인수, 박수견, 조영환, 이채연, 김아원, 김우창 (2020)은 한국십진분류법 내 인문 분야 도서 중 한국어로 번역된 도서의 내용 및 리뷰 정보에 대하여 군집 분석, LDA, 인공지능망 기법을 활용하여 사용자 군집별 예상 도서 만족도를 산출하였으며, 이를 통해 도서 정보 및 리뷰 기반의 도서 추천시스템을 개발하였다[7].

그 외 선행 연구에서는 빅데이터 기반의 이용자 중심 도서 추천시스템 설계를 통해 추천시스템 사용에 있어서의 편리성을 증대시키고자 하였다[8].

국외 연구로는 Keita Tsuji, Nobuya Takizawa, Sho Sato, Ui Ikeuchi, Atsushi Ikeuchi, Fuyuki Yoshikane and Hiroshi Itsumura(2014)는 대학 도서 대출 기록 및 서지 데이터에 대하여 SVM 기법과 연관 규칙을 활용하여 분석을 수행함으로써 자동분류기반의 도서 추천시스템을 개발하였다[9].

Yonghong Tian, Bing Zheng, Yanfang Wang, Yue Zhang, Qi Wu(2019)는 대학 내 기술 분야 도서관의 도서대출 기록 및 회원정보 데이터에 대하여 하이브리드 알고리즘 기법을 활용하여 개인 맞춤형 도서 추천 시스템을 구현하였다[10].

이상의 선행 연구들은 빅데이터 기반 사용자 맞춤형 도서 추천시스템을 제공함으로써 4차 산업혁명 시대에 부합한 도서 관련 서비스를 제안하고 이용자의 만족도를 제고한 장점이 있으며, Table 1에서 확인할 수 있다. 그러나 추천 도서의 범위가 대출빈도가 높은 인기도서로 한정되어 있어 도서관 이용자가 추천시스템을 통해 접할 수 있는 도서의 다양성이 제한된다는 한계가 있다.

Table 1. Book Recommendation System previous study

Author (Year)	Purpose of study	Analysis materials		Method of analysis
		material	period	
H.Y. Cho (2017)	A Experimental Study on the Development of a Book Recommendation System Using Automatic Classification, Based on the Personality Type	book introduction	2017	LIBLINEAR, LibSVM model
Y.J. Jung (2017)	Topic Modeling-based Book Recommendations Considering Online Purchase Behavior	customer transaction, book introduction, book contents	2014	Topic modeling (LDA)
Y.S. Jeong (2018)	A study on the Book Recommendation Model Using Big Data	books/loan list, popular loaned books list	2018	social network analysis

D.W. Park (2020)	A study on Architecture for BigData-based Book Curation System	user information, loan records	2020	BDCM model
I.S. Choi (2020)	Developing a Book Recommendation Model Based on Reading Level Using Book Information and Book Review Data	book contents, book reviews	2019	cluster analysis, neural network, Topic modeling (LDA)
Keita Tsuji (2014)	Book Recommendation Based on Library Loan Records and Bibliographic Information	loan records, bibliographic information	2006 ~ 2012	SVM, association rule
Yonghong Tian (2019)	College Library Personalized Recommendation System Based on Hybrid Recommendation Algorithm	loan records, user information	2014~ 2016	Hybrid Algorithm

2.2 이론적 고찰

추천시스템은 사용자의 특성, 취향을 고려하여 방대한 데이터로부터 선별된 정보를 제공하는 개인화된 정보 탐색 도구이다[11,12]. 유튜브, 넷플릭스에서 제공하는 콘텐츠 기반 필터링 추천시스템은 이용자의 편리성을 증대시키고 플랫폼 이용 활성화의 효과를 창출하는 등 추천시스템의 대표적인 성공 사례라 할 수 있으며 그 중요성은 커지고 있다[13,14].

추천시스템의 알고리즘은 대표적으로 협업 필터링과 콘텐츠 기반 필터링이 널리 활용되고 있다. 협업 필터링(Collaborative Filtering)은 사용자와 유사한 선호도를 가진 집단에 대한 축적된 데이터를 기반으로 사용자가 관심 있어 할 정보를 제공하는 기법이다[15,16]. 반면에 콘텐츠 기반 필터링(Content-based Filtering)은 아이템을 분석한 프로파일과 사용자의 선호도 프로파일을 통해 아이템들의 속성 간 유사성을 기반으로 추천을 제공하는 기법이다[17,18]. 최근에는 협업 필터링과 콘텐츠 기반 필터링을 적절하게 결합하여 보다 우수한 성능으로 사용자가 필요로 하는 아이템을 추천하는 하이브리드 알고리즘 추천시스템과 관련된 연구도 다수 진행되고 있다[19,20].

본 연구에서는 공공도서관의 대출이력과 회원정보를 기반으로 사용자 집단의 선호를 반영하여 미래출 도서를 추천하는 콘텐츠 기반 필터링 도서 추천시스템을 구현하였다. 협업 필터링의 경우, 기존의 데이터가 축적되지 않은 신규 사용자에게 어떠한 아이템도 추천해줄 수 없는

콜드 스타트¹⁾의 문제를 가지고 있기에 콘텐츠 기반 필터링을 채택하였다. 군집 분석과 토픽 모델링 내 LDA 기법을 통해 사용자 집단별 인기 카테고리 내 인기도서를 추출한 후 이와 유사한 미대출 도서를 코사인 유사도 계산을 통해 추천하는 방식으로 구현하였다.

3. 성·연령대별 인기 카테고리 분석

3.1 데이터 수집

대구광역시의 전체 43개 공공도서관이 보유한 평균 도서 수인 105,893권을 기준으로, 그 이상의 도서를 보유한 대구시 내 13개의 공공도서관을 대상으로 미대출 도서의 빈도를 파악하였다. 두류도서관의 미대출 도서 수가 69,752권으로 가장 높게 나타났으며, 다음으로 2.28기념학생도서관(46,498권), 범어도서관(28,825권) 순으로 나타났다.

이에 본 연구에서는 미대출 도서 추천시스템을 구현하기 위해 공간적 범위는 대구시 두류도서관, 시간적 범위는 2019년 1월부터 2020년 7월까지에 해당하는 데이터를 수집하였으며, 이에 대한 구성요소는 Table 2에서 확인할 수 있다. 이를 세부적으로 살펴보면 장서/대출목록 자료(272,652건)와 대출이력 자료(213,093건), 회원정보(35,561명)로 구성되어 있다. 장서/대출목록 자료의 경우 도서명, 저자, 출판사, 발행 연도, ISBN²⁾, 세트 ISBN, 부가기호³⁾, 주제분류번호, 도서권수, 대출 건수, 등록일자 등이 포함되어 있으며, 대출이력 자료의 경우 대출 건별로 사용자 번호, 도서 번호, KDC⁴⁾, 도서명, 저자, 출판사, 대출일, 반납일, 미디어 형태 등이 포함되어 있다. 회원정보에는 사용자 번호, 등록일, 생년월일, 성별, 우편번호, 도서 대여 횟수, 연체 횟수 등이 포함되어 있다.

이상의 도서 자료만으로는 고도화된 추천시스템 구현이 어려울 것으로 판단되어, 도서관 정보나루에서 제공

하는 Open API를 활용하여 대구시 두류도서관에서 보유하고 있는 도서의 ISBN을 기준으로 도서 상세정보 자료(272,652건)를 수집하였다. 해당 자료에는 도서명, KDC, 출판일, 저자, 출판사, 도서 이미지 URL, 도서 소개 등이 포함되어 있다. 이후 파이썬(Python)을 활용하여 XML, JSON 형태의 반정형 데이터인 도서 상세정보 자료를 정형 데이터로 변환하였다.

Table 2. Components of the Data

Data	Components
Book/loan list (272,652 cases)	Title, Author, Publisher, Year of publication, ISBN, Set ISBN, Additional code, Subject classification number, Number of books, Number of loans, Registration date and etc
Loan Records (213,093 cases)	User key, Book key, KDC, Title, Author, Publisher, Loan date, Return date, Media type and etc
User Information (35,561 people)	User key, Registration date, Date of birth, Gender, Zipcode, Number of loans, Number of overdue and etc
Book Details (272,652 cases)	Title, KDC, Date of publication, Author, Publisher, Book Image URL, Book Outline and etc

3.2 데이터 전처리

데이터의 전처리는 통합된 하나의 데이터셋 구축을 위한 전처리와 텍스트 마이닝을 위한 전처리로 구분하여 진행하였으며, 이를 세부적으로 살펴보면 다음과 같다.

첫째, 데이터셋 구축을 위한 전처리의 경우 도서 상세정보 자료에 도서 소개가 공백이거나 도서명이 동일한 도서가 존재하여 R을 통해 공백인 행은 제거하고 도서명이 동일한 도서의 도서 소개는 통합하였다. 대출이력 자료 및 회원정보의 경우 성·연령대별 인기 카테고리 기반의 도서 추천을 위해 사용자 번호(User key)를 기준으로 결합하였고, 분석에 필요한 사용자 번호, 등록일, 생년월일, 성별을 제외한 데이터는 제거하였다. 또한 이상치와 결측치를 제거하기 위해 출생연도가 2020년 이후인 데이터와 누락된 데이터를 일괄적으로 제거하였으며, 이산형의 생년월일 데이터를 범주형으로 변환하여 연령대 변수를 생성하였다. 또한 장서/대출목록 자료를 도서명 기준으로 병합하여 ISBN 열을 추가하였다. 이상의 전처리를 통해 대출 정보 데이터셋을 구축하였다.

콘텐츠 기반 필터링 도서 추천을 위해 대출 정보 데이터셋에 도서 상세정보 자료를 ISBN 또는 도서명을 기준으로 병합하여 도서 소개 열을 추가하였다. 대출 정보 데

1) 콜드스타트(Cold Start) : 데이터가 충분하지 못한 경우 추천 시스템 구현이 어려운 상황
 2) ISBN(International Standard Book Number) : 도서를 서점에 유통시키기 위해 발급받는 도서번호(국제표준도서번호)
 3) 부가기호 : 도서의 독자대상, 발행형태, 내용 등의 분류를 위해 지정한 기호
 4) KDC(Korean Decimal Classification) : 한국십진분류법에 따른 도서 분류기

이터셋의 ISBN이 결측치, 이상치 또는 10자리인 경우가 있어, 10자리인 경우 13자리로 변환한 후 ISBN을 기준으로 병합하였다. ISBN이 결측치 또는 이상치인 경우 SequenceMatcher⁵⁾ 라이브러리를 활용하여 대출 정보 데이터셋과 도서 상세정보 자료 간 도서명의 유사도를 산출하였다. 유사도가 90% 초과인 도서는 도서 상세정보 자료의 도서명으로 통일하여 병합하고 그렇지 않은 도서는 제거하였다. 이와 같은 과정을 통해 최종 대출 정보 데이터셋을 구축하였다.

둘째, 텍스트 마이닝을 위한 전처리의 경우 도서 상세정보 자료의 도서명, 도서 소개와 같은 텍스트 데이터에서 숫자, 특수문자 등을 제거하는 텍스트 클렌징을 수행하였다. 이후 텍스트를 분석하기 위해 KoNLPy 라이브러리를 활용하여 한글을 형태소 단위로 분리하는 토큰화(Tokenization) 작업을 진행하였고, 품사가 명사인 단어만 추출하였다. 또한 유의미한 단어만을 추출하기 위해 KoNLPy 라이브러리에서 제공하는 불용어 사전만을 적용하는 것은 한계가 있어 자체적으로 불용어 사전을 생성하여 불용어를 제거하였다. 이상의 수행한 데이터 전처리에 관한 내용은 Table 3에서 확인할 수 있다.

Table 3. Types of Preprocessing

Purpose of Preprocessing	Content
Building of dataset	removal of missing values and outliers, conversion of data type, addition of required data, calculation of similarities between strings, sequential combination of 4 materials
Text mining	text cleansing, tokenization, noun extraction, stopwords removal(286 cases)

3.3 도서 군집별 카테고리 분석

본 연구에서는 실질적인 도서의 분류를 위해 Spherical K-means 클러스터링⁶⁾과 LDA⁷⁾ 기반의 토픽모델링을 활용하여 수집한 모든 도서를 군집으로 분류하고, 각각의 군집이 의미하는 토픽을 분석하였다. 먼저 전체 159,132 권의 도서가 포함하고 있는 도서명, 도서 소개에서 단어

5) SequenceMatcher : 두 개의 문자열에 대한 상호 유사성을 수치화하는 기능을 제공하는 파이썬 표준 라이브러리

6) Spherical K-means 클러스터링 : 문서 집합과 같은 고차원 데이터의 군집 분석에 적합한 머신러닝 기법

7) LDA(Latent Dirichlet Allocation, 잠재 디리클레 할당) : 주어진 문서에 대해 각 문서에 어떤 토픽들이 존재하는지를 서술하는 확률적 토픽 모델 기법

토큰화로 추출된 명사를 활용하여 Spherical K-means 클러스터링을 수행함으로써 모든 도서를 20개의 군집으로 분류하였다. 군집의 개수는 토픽 모델링 관련 선행 연구 내 토픽 13개~27개의 범위를 준용하였으며[21, 22], 토픽의 개수가 너무 많으면 해석에 어려움이 존재하기에 20개로 지정하였다. 해당 분야의 정책적 전문가 1명과 연구원 1명의 자문 또한 반영하였다. 군집화를 통해 분류된 20개의 군집에 대하여 LDA 기반의 토픽 모델링을 활용하여 개별적으로 내포하고 있는 토픽을 분석하였으며, 분석 결과 토픽에 포함되는 키워드를 통해 20개의 도서 카테고리를 추출하였다.

이에 해당하는 도서 카테고리는 다음과 같다. 문학·여행, 사회문제, 심리·정치, 종교·철학, 건강·요리, 외교·세계정치, 청소년교육, 취미·생활·IT, 외국어, 패션·디자인, 의학·세계소설, 자기계발·창업, 소설, 세계사·역사, 문화·예술, 교육, 산문·시집, 과학, 만화·오락, 한국수필이다.

전체 도서에 대한 20개의 토픽과 토픽별 키워드는 Table 4와 같으며 미래출 도서의 콘텐츠 기반 필터링 추천을 위해 추출한 도서 카테고리를 전체 도서 목록에 추가하여 각각의 도서에 해당하는 토픽별 의미를 부여하였다.

Table 4. Keyword of Topics

Topic (Topic Name)	Major Keyword
Topic-1 (Literature/Travel)	Korean, Joseon Dynasty, literature, history culture, travelogue, big river
Topic-2 (Social problem)	modern times, solutions, communication, analysis, justice, matter, criticism, society
Topic-3 (Psychology/Politics)	cognition, nation, the West, cure, meeting, family, soul, situation
Topic-4 (Religion/Philosophy)	communication, mystery, principle, rule, world, faith, religion, Bible
Topic-5 (Health/Cooking)	human, making, tree, village, miracle
Topic-6 (Diplomacy/World politics)	consideration, situation, France, peace, Japan, freedom, history culture, world history, the world
Topic-7 (Youth education)	music, Korean, lesson, observation
Topic-8 (Hobby/Life Style/IT)	photograph, secret, making, city, diary, modern people, principle, design
Topic-9 (Foreign language)	creation, meeting, program, friend, concentration, big, country, refinement, ability
Topic-10 (Fashion/Design)	architecture, empathy, sport, company, try, essay, peace
Topic-11 (Medical science/World novel)	medicine, the world, alternative, creation, romance, mystery, research, presentation

Topic-12 (Self-improvement/ Startup)	birth, method, secret, principle, diary, passion, problem, analysis
Topic-13 (Novel)	mystery, popularity, citizen, sport
Topic-14 (World history/History)	humanities, country, the West, Europe, solution, theory, leader
Topic-15 (Culture/Art)	short story, style, English, Russia, North Korea, classic, religion, period, art
Topic-16 (Education)	preparation, baby, children, internet, birth, adult, secret, child, way
Topic-17 (Prose/Poems)	scenery, tree, fun, diary
Topic-18 (Science)	mystery, indeed, experiment, type, analysis, birth, development
Topic-19 (Comic/Entertainment)	strategy, learning, Greece, mystery, game, internet
Topic-20 (Korean essay)	insight, story, situation, language, friend

3.4 성·연령대별 인기 카테고리 분석

본 연구에서는 성·연령대별 인기 카테고리 기반의 미대출 도서 추천을 위해 전처리된 전체 도서 목록에서 대출 도서 목록과 미대출 도서 목록을 생성하였다. 이후 대출 도서 목록을 활용하여 성·연령대별로 대출 건수가 가장 높은 상위 3개의 인기 카테고리를 추출하였다.

Table 5. Topics Popular by Gender and Age

Gender	Age	Top 3 Popular Topics		
		1st	2nd	3rd
Man	0~19	Youth education	Comic/Entertainment	Science
	20~29	Self-improvement/Startup	Literature/Travel	Hobby/Life Style/IT
	30~39	Self-improvement/Startup	Literature/Travel	Hobby/Life Style/IT
	40~49	Self-improvement/Startup	Youth education	Comic/Entertainment
	50~59	Literature/Travel	Self-improvement/Startup	Hobby/Life Style/IT
	60~	Literature/Travel	Self-improvement/Startup	World history/History
Woman	0~19	Youth education	Comic/Entertainment	Science
	20~29	Literature/Travel	Self-improvement/Startup	Novel
	30~39	Youth education	Literature/Travel	Comic/Entertainment
	40~49	Literature/Travel	Youth education	Comic/Entertainment
	50~59	Literature/Travel	Self-improvement/Startup	Novel
	60~	Literature/Travel	Comic/Entertainment	Youth education

성별은 남성과 여성 2가지 유형, 연령대는 각 성별마다 10대 이하, 20대, 30대, 40대, 50대, 60대 이상의 6가지 유형으로 구분하여 총 12개의 사용자 집단에 대한 상위 3개의 인기 도서 카테고리를 추출하였고, 그 결과는 Table 5와 같다.

4. 미대출 도서 추천시스템 구현

4.1 추천시스템 구현

본 연구에서는 콘텐츠 기반 필터링 추천시스템 구현을 위해 불용어 처리 및 명사 추출을 완료한 전체 도서 목록에 대하여 TF-IDF 기반의 피쳐 벡터화를 수행하였다. 피쳐 벡터화란 텍스트를 벡터 값을 가진 피쳐로 변환하는 수치화 방식으로 그 중 TF-IDF(Term Frequency Inverse Document Frequency) 기법은 다중 문서로 이루어진 문서 집합에서 특정 문서 내 각 단어의 중요도를 산출하는 기법이다. TF-IDF는 특정 문서 내 단어 빈도수가 높을수록 가산점을 부여할 뿐만 아니라, 여러 문서에서 많이 등장하는 단어에는 패널티를 부여함으로써 각 단어의 특성을 보다 잘 반영한다는 장점이 있다.

이에 본 연구에서는 TF-IDF 기법을 사용하여 피쳐 벡터화를 수행하였으며 TF-IDF 값을 산출하는 수식은 다음과 같다.

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$tf_{i,j}$: j 문서에서 단어 i의 빈도 수

df_i : 단어 i를 포함한 문서의 수

N : 전체 문서의 수

다음으로 피쳐 벡터화를 수행한 전체 도서 목록에 대하여 동일 카테고리 내 대출 도서와 미대출 도서 간의 코사인 유사도(Cosine Similarity)를 계산하였다. 코사인 유사도란 두 벡터 사이의 각도에 해당하는 코사인 값을 통해 산출된 벡터 간의 유사성에 대한 수치를 의미한다. 두 벡터의 사잇각이 0°일 경우 1, 90°일 경우 0, 그리고 180°일 경우 -1의 값을 가지며 이에 대한 계산식은 아래와 같다.

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

코사인 유사도는 차원의 제한을 받지 않는 특성으로 인해 다차원 공간에서의 벡터 간 유사도 측정에 적합하

다는 장점이 있다. 이에 본 연구에서는 각각의 미대출 도서와 모든 대출 도서 간 코사인 유사도의 평균을 계산하였으며, 이를 내림차순으로 정렬하여 미대출 도서별 추천 순위를 산정하였다. 추천시스템 구현을 위해 최종적으로 생성한 성·연령대별 인기 카테고리에 해당하는 미대출 도서 추천 목록 예시는 Table 6과 같다.

Table 6. Example of Unborrowed Book Recommendation List by Gender and Age

Age/Gender	Popular Category	Title	Rank
0~19 woman	Youth education	적성따라 공부하기	3
		⋮	⋮
		새로운 말 따라 하기	95
	Comic/Entertainment	(오은희의)만화 토피아	8
		⋮	⋮
		식객	91
	Science	과학공화국 지구법정	4
		⋮	⋮
은몸이 물리천지		98	
50~59 man	Literature/Travel	은비행	4
		⋮	⋮
		TV 드라마 & 영화 촬영지 여행	71
	Self-improvement/Startup	(잡아라!)부동산 대박	2
		⋮	⋮
		젊은 부자가 되는 9가지 비밀	90
	Hobby/Life Style/IT	시네마4D: 곧바로 활용하기	9
		⋮	⋮
엑셀실무 활용기술77		100	

4.2 실험 및 평가

앞서 산출된 분석 결과물을 활용하여 Fig. 1과 같이 미대출 도서를 추천해주는 웹 사이트를 구현하였으며, 특정 성·연령을 선택하면 해당 범주의 인기 도서 카테고리 내 인기도서와 유사한 내용을 포함하는 미대출 도서가 유사도가 높은 순으로 도출된다.

앞서 구현된 미대출 도서 추천시스템의 만족도를 조사하기 위해 설문조사를 실시하였으며, 설문문항은 성별, 연령대, 직업, 평소 독서량, 인기도서 위주의 추천시스템에 대한 불만 여부 등의 개인적 특성과 구현된 미대출 도서 추천시스템의 만족도와 기존 추천시스템과의 선호도로 구성하였다. 이를 시간적으로는 2021년 1월 29일부터 31일까지 3일간 온라인 설문조사를 실시하였으며, 공간적으로는 대구시 공공도서관 이용자들을 대상으로 하여 최종적으로 250명이 설문조사에 응답하였으며, 세부적인 현황은 Table 7과 같다.

설문조사 진행 방식은 온라인 환경에서 구현된 미대출 도서 추천시스템을 이용한 후 만족도와 기존 추천시스템과의 비교 선호도 문항에 대해 최소 1점에서 최대 5점까지의 점수를 부여하도록 하였다. 1점은 매우 불만족, 3점은 만족, 5점은 매우 만족을 의미한다. 1점의 경우 추천시스템 이용에 있어 큰 불편을 초래하는 만족도의 지표를 의미하며, 3점의 경우 통상적으로 이용에 있어서 불편



Fig. 1. Unborrowed Books Recommendation System Dashboard

함이 없으며 추천받은 도서와 카테고리에 대하여 보통 수준의 만족을 의미한다. 마지막으로 5점의 경우, 이용하기에 매우 편리하며 추천받은 도서와 카테고리에 대하여 매우 큰 만족을 나타내는 기준으로 판단된다.

Table 7. General Characteristics of the Subjects

Variables		Persons	Ratio (%)
Gender	Man	144	57.6
	Woman	106	42.4
Age	0~19	10	4.0
	20~29	69	27.6
	30~39	56	22.4
	40~49	23	9.2
	50~59	32	12.8
	60~	60	24.0
Occupation	Office worker	69	27.6
	Professions	23	9.2
	Self-employed	26	10.4
	Public officer/Teacher	21	8.4
	Housewives	20	8.0
	Students	48	19.2
	Unemployed	35	14.0
	Etc.	8	3.2
Number of books read (per year)	0	30	12.0
	1~3	101	40.4
	4~6	55	22.0
	7~9	19	7.6
	10~	45	18.0
Complaints Existence (for existing book system)	yes	71	28.4
	no	179	71.6
Total		250	100.0

미대출 도서 추천시스템의 만족도 지표 관련 설문결과는 Table 8에서 확인할 수 있으며, 이를 세부적으로 살펴보면 본인의 관심사에 적합한 도서 카테고리의 추천이 되었는지에 대한 질문은 평균 3.484점으로 평균 만족도 이상의 점수로 나타났다. 그리고 추천 받은 도서를 읽을 것인지의 질문에는 3.78점으로 보통 수준 이상의 높은 만족도를 보였음을 알 수 있다.

Table 8. Survey Result on satisfaction with unborrowed book recommendation system

Question	Avg	Std
Have you been recommended a book category suitable for your interests?	3.484	1.007
Are you willing to read recommended books?	3.780	1.016

미대출 도서 추천시스템의 만족도 관련 문항 2개에 대한 평균 점수를 성·연령대별로 산정하였고, 분석 결과는 Table 9에서 확인할 수 있다. 성별의 경우 여성이 평균 3.684점, 남성이 3.594점으로 여성이 남성에 비해 높게 나타났으며, 연령대의 경우 40대가 3.957점으로 가장 높게 나타났으며, 20대가 3.761점, 50대는 3.625점 순으로 나타났다.

Table 9. Survey Result on satisfaction based on General Characteristics of the Subjects

Variables		Avg
Gender	Man	3.594
	Woman	3.684
Age	0~19	3.600
	20~29	3.761
	30~39	3.545
	40~49	3.957
	50~59	3.625
	60~	3.450
Occupation	Office worker	3.645
	Professions	3.326
	Self-employed	3.673
	Public officer/Teacher	3.762
	Housewives	3.500
	Students	3.750
	Unemployed	3.457
	Etc.	4.313
Number of books read (per year)	0	3.300
	1~3	3.564
	4~6	3.818
	7~9	3.763
	10~	3.722
Complaints Existence (for existing book system)	yes	3.683
	no	3.612

또한 사회적 특성에 따른 미대출 도서 추천시스템에 대한 만족도를 분석한 결과 직업의 경우 공무원/교사가 평균 3.762점으로 가장 높은 점수로 나타나며, 다음으로 학생 3.750점, 자영업자 3.673점 순으로 높게 나타났다. 평소 독서량의 경우 0권은 3.300점, 1~3권은 3.564점으로 나타난 반면에, 4~6권은 3.818점, 7~9권은 3.763점, 10권 이상은 3.722점으로 다독자 계층에서 만족도가 상대적으로 높게 나타났다.

다음으로 기존의 인기도서 위주의 도서 추천시스템과 미대출 도서 추천시스템의 추천 방식에 대한 선호도 관

런 설문결과는 Table 10과 같으며, 기존의 인기도서 위주의 추천시스템 중 협업필터링 방식은 평균 3.996점, 콘텐츠 기반 필터링 방식은 4.176점, 하이브리드 방식은 4.128점, 마지막으로 미대출 도서 추천시스템은 3.772점의 선호도로 나타났다.

Table 10. Survey Result on preference with various book recommendation system

Question	Avg	Std
Recommend books that I might like based on loan records of a user with a similar taste as me	3.996	0.898
Recommend books that I might like based on contents of my favorite books	4.176	0.879
Recommend books as Hybrid-style that combines the first and second recommendations	4.128	0.918
Recommend unborrowed books based on borrowed books in the popular book category of the gender and age group to which I belong	3.772	1.022

미대출 도서 추천시스템이 다른 3개의 기존 추천시스템에 비해 상대적으로 낮은 선호도를 보인다고 할 수 있으나, 절대적으로 평가한다면 3.772점의 선호도는 실제 도서관 이용자들의 자발적인 이용을 기대할 수 있는 유의미한 수치로 판단된다. 반면 표준편차의 경우 미대출 도서 추천시스템이 1.002로 기존 추천시스템에 비해 상대적으로 높게 나타났으며, 응답자들의 분포가 상대적으로 분산되어 있음을 확인할 수 있다.

5. 결론

5.1 분석 결과 요약

본 연구에서는 공공도서관 도서대출 편향의 문제를 빅데이터, 인공지능 분야에서 각광받고 있는 추천시스템을 활용하여 접근하였으며, 대출빈도가 높은 인기도서 위주의 도서 추천이 아닌 인기도서와 유사한 도서 내용을 포함하고 있는 미대출 도서를 추천한다는 점에서 선행 연구와 차별성을 가진다.

이를 위해 대구시에서 미대출 도서의 빈도가 가장 높은 두류도서관을 대상으로 Spherical K-means 클러스터링과 LDA 기반의 토픽모델링을 활용하여 성·연령대별 인기 도서 카테고리를 추출하였다. 그리고 카테고리별 인기도서와 유사한 도서 내용을 포함하는 미대출 도서와의 유사도 계산으로 미대출 도서 추천시스템을 구현하

고, 시스템 이용자들의 만족도 조사를 수행함으로써 미대출 도서 추천의 가능성을 살펴보았으며, 분석한 결과를 요약하면 다음과 같다.

첫째, Spherical K-means 클러스터링과 LDA 기반의 토픽모델링으로 성·연령대별 인기 도서 카테고리를 집단별로 3개씩 도출하였으며, 세부적으로 살펴보면 10대 이하 여성은 청소년교육, 만화오락, 과학 카테고리 순으로 도서 선호가 높은 것으로 나타났으며, 50대 남성의 경우 문학여행, 자기계발·창업, 취미생활·IT 순으로 분석되었다.

둘째, 인기 도서 카테고리 내 인기도서와 미대출 도서 간 코사인 유사도를 계산하였으며, 그 중에서 문학·여행 카테고리가 가장 높게 나타났으며, 세부적으로는 20대 남성, 30대 여성, 20대 여성 순으로 유사도가 높게 분석되었다. 다음으로 과학, 자기계발·창업 카테고리 순으로 높게 나타났으며, 세부적으로는 과학의 경우 10대 이하 여성, 10대 이하 남성 순으로 높게 나타났으며, 자기계발·창업을의 경우 50대 남성과 여성, 40대 남성 순으로 인기도서와 미대출 도서간 코사인 유사도가 높은 것으로 분석되었다.

셋째, 미대출 도서 추천시스템의 이용자들을 대상으로 실시한 설문조사에서 본인의 관심사에 적합한 도서가 추천되었는지에 대한 질문에는 3.484점, 추천된 미대출 도서를 읽을 것인지에 대한 질문에는 3.78점으로 평균 만족도인 3점을 상회하는 것으로 분석되었다.

넷째, 설문조사 결과를 통해 성·연령대별 만족도를 세부적으로 살펴보면 성별은 여성의 경우 남성보다 높게 나타났으며, 연령대는 40대, 20대, 50대 순으로 만족도가 높게 나타났다.

다섯째, 사회적 특성에 따른 만족도를 살펴보면 직업의 경우 공무원/교사가 가장 높게 나타났으며, 학생, 자영업자가 그 뒤를 이었다. 그리고 평소 독서량이 3권 이하에 비해 4권 이상의 다독자 계층에서 만족도가 상대적으로 높게 나타났으며, 기존 인기도서 위주의 추천시스템에 대한 불만이 있는 집단이 상대적으로 만족도가 높은 것으로 분석되었다.

마지막으로 알고리즘별 도서 추천시스템의 선호도 조사 결과 협업 필터링 3.996점, 콘텐츠 기반 필터링 4.176점, 하이브리드 4.128점으로 나타났으며, 미대출 도서 추천시스템은 3.772점의 선호도로 나타났다.

5.2 정책적 시사점 및 연구의 한계

본 연구의 분석 결과를 통해 정책적 시사점을 도출하

면 다음과 같다.

첫째, 문학과 여행, 과학, 자기계발·창업 카테고리 분야에 있어 인기도서와 미대출 도서의 유사도가 높게 나타났다으며, 해당 분야의 도서 추천이 유사한 내용의 도서를 추천한다는 측면에서 효과적인 것으로 보인다.

둘째, 미대출 도서 추천시스템의 만족도는 연령대별로 살펴보면 40대, 20대, 50대 순으로 높게 나타났으며, 직업의 경우 공무원/교사, 학생, 자영업자 순으로 만족도가 높은 것으로 분석되었다. 또한 평소 독서량이 3권 이하에 비해 4권 이상의 다독자 계층일수록 미대출 도서 추천시스템을 선호하는 것으로 확인되었다. 따라서 해당 계층을 대상으로 미대출 도서 추천시스템을 구현할 경우 만족도 높은 도서 추천이 가능할 것으로 판단되며, 공공도서관의 편향된 도서 대출, 운영 효율성 저하 등의 문제를 일부 개선할 수 있을 것으로 판단된다.

셋째, 알고리즘별 도서 추천시스템의 선호도 조사 결과 기존 인기도서 위주의 추천시스템에 비해 상대적으로 낮은 선호도 점수로 나타났으나, 평균 이상의 만족도와 응답자의 분포가 분산되어 있음을 감안할 때 특정 계층에 있어서 미대출 도서 추천시스템의 활성화에 대한 가능성이 존재할 것으로 판단된다.

넷째, 미대출 도서 추천시스템으로 기존 출판사의 마케팅과 환경적 요인 등으로 인기도서에 비해 상대적으로 조명 받지 못한 양질의 도서를 발굴함으로써 향후 도서 산업 발전에 일부 기여할 수 있을 것으로 보인다.

다섯째, 공공도서관 이외의 민간 도서 분야에서도 추천시스템 구현 시 판매 빈도가 높은 인기도서 이외에도 특성이 유사한 비인기도서를 함께 추천함으로써 고도화된 추천시스템 구현이 가능할 것으로 판단된다.

마지막으로 본 연구는 대구시 두류도서관을 중심으로 분석이 수행되었으며, 타 지자체의 공공도서관에도 적용한다면 공공도서관 운영 활성화에 기여할 것으로 판단되며, 거주민들의 다양한 지식과 이념을 학습하는데 도움이 될 것으로 기대한다.

본 연구의 학술적 함의는 다음과 같다.

본 연구에서는 텍스트 마이닝에 적합한 Spherical K-means 클러스터링을 통해 1차적으로 도서를 군집화하고, 2차적으로 LDA 기반의 토픽 모델링을 활용하여 토픽별 키워드를 확인함으로써 도서 카테고리를 분류하였다. 도서 카테고리 별로 포함하고 있는 도서가 서로 중복되지 않아야 하는 점에 착안하여 이상의 기법들을 통해 우수하게 분류할 수 있었다.

그리고 기존의 인기도서 위주 추천시스템에서 사용자

또는 아이템 간에 일대일 대응을 통해 코사인 유사도를 계산한 것과 달리, 본 연구에서는 성·연령대별 인기 카테고리 내의 인기도서들을 기반으로 아이템 간에 일대일 대응을 통해 코사인 유사도로 미대출 도서를 추천하였다. 또한 미대출 도서 추천시스템의 이용자들의 만족도를 조사한 결과 보통 이상의 높은 만족도가 나타났으며, 미대출 도서 추천시스템 후속 연구의 기초 자료로 활용될 수 있을 것으로 기대한다.

한편, 본 연구의 한계는 다음과 같다.

첫째, 콘텐츠 기반 필터링 알고리즘의 경우 도서 내용을 기반으로 도서를 추천하는 것이기에 때문에 추후 사용자와 아이템 기반의 하이브리드 알고리즘을 적용한다면 고도화된 추천시스템 구현이 가능할 것으로 보인다.

둘째, 카테고리 군집의 개수를 산정함에 있어 정량적 지표를 적용하지 않고, 전문가 자문을 활용한 점에 있어 연구자의 주관이 개입될 수 있는 한계점이 존재한다.

셋째, 미대출 도서 추천시스템 이용자들의 만족도 설문조사를 수행함에 있어 표본의 선정에 층화추출법, 군집추출법 등이 아닌 무작위추출법을 적용하여 표본의 대표성 문제가 일부 존재할 것으로 판단된다.

넷째, 인기 도서 카테고리 내 대출 도서와 유사한 도서 내용이 포함된 미대출 도서를 이용자들이 선호한다는 가정에 대한 객관적인 논리가 일부 부족한 것으로 보인다.

마지막으로 공간적 범위를 대구시 두류도서관으로 한정하여 분석하였는데, 이용자들의 특성이 시·공간적으로 다양하게 변화됨을 감안할 때 관련된 후속 연구들이 활발히 수행될 것으로 기대된다.

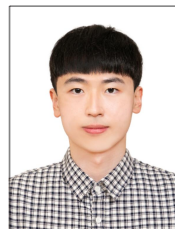
REFERENCES

- [1] Y. S. Kim. (2007). A Study on the Effective Enhancement of the Public Library Infrastructure in Korea: Focusing on the Operation of Small Public Libraries in the UK. *Journal of Korean Library and Information Science Society*, 38(1), 29-48.
DOI : 10.16981/kliss.38.1.200703.29
- [2] Y. J. Shin & Y. H. Noh. (2020). A Study on the Program Structure and Activation Plan of the Complex Cultural Space in Public Library. *Journal of the Korean Society for Library and Information Science*, 54(1), 89-114.
DOI : 10.4275/KSLIS.2020.54.1.089
- [3] W. S. Kim. (2019). Content Recommendation Techniques for Personalized Software Education. *Journal of Digital Convergence*, 17(8), 95-104.
DOI : 10.14400/JDC.2019.17.8.095

- [4] H. Y. Cho. (2017). A Experimental Study on the Development of a Book Recommendation System Using Automatic Classification, Based on the Personality Type. *Journal of Korean Library and Information Science Society*, 48(2), 215–236. DOI : 10.16981/kliss.48.201706.215
- [5] Y. J. Jung & Y. H. Cho. (2017). Topic Modeling-based Book Recommendations Considering Online Purchase Behavior. *Knowledge Management Research*, 18(4), 97–118. DOI : 10.15813/kmr.2017.18.4.004
- [6] Y. S. Jeong. (2018). A study on the Book Recommendation Model Using Big Data.
- [7] I. S. Choi & S. K. Park & Y. H. Cho & C. Y. Lee & A. W. Kim & W. C. Kim. (2020). Developing a Book Recommendation Model Based on Reading Level Using Book Information and Book Review Data. *Journal of the Korean Institute of Industrial Engineers*, 46(3), 179–189. DOI : 10.7232/JKIIE.2020.46.3.179
- [8] D. W. Park. (2020). A Study on Architecture for BigData-based Book Curation System
- [9] Keita Tsuji et al. (2014). Book Recommendation Based on Library Loan Records and Bibliographic Information. *Journal of Procedia - Social and Behavioral Sciences*, 147, 478–486. DOI : 10.1016/j.sbspro.2014.07.142
- [10] Yonghong Tian & Bing Zheng & Yanfang Wang & Yue Zhang & Qi Wu. (2019). College Library Personalized Recommendation System Based on Hybrid Recommendation Algorithm. *Journal of Procedia CIRP*, 83, 490–494. DOI : 10.1016/j.procir.2019.04.126
- [11] B. H. Oh & J. H. Yang & H. J. Lee. (2014). A Hybrid Recommender System based on Collaborative Filtering with Selective Utilization of Content-based Predicted Ratings. *Journal of KISS : Software and Applications*, 41(4), 289–294. DOI : 10.1016/j.procir.2019.04.126
- [12] K. J. Kim & H. C. Ahn. (2009). User-Item Matrix Reduction Technique for Personalized Recommender Systems. *Journal of Information Technology Applications & Management*, 16(1), 97–113.
- [13] Y. J. Lee & C. H. Lee. (2020). What Do The Algorithms of The Online Video Platform Recommend: Focusing on Youtube K-pop Music Video. *Journal of the Korea Contents Association*, 20(4), 1–13. DOI : 10.5392/JKCA.2020.20.04.001
- [14] Airoidi, M & Beraldo, D & Gandini, A. (2016). Follow the algorithm : An exploratory investigation of music on Youtube. *Journal of Poetics*, 57, 1–13. DOI : 10.1016/j.poetic.2016.05.001
- [15] S. Y. Yun & S. D. Yoon. (2020). Item-Based Collaborative Filtering Recommendation Technique Using Product Review Sentiment Analysis. *Journal of the Korea Institute of Information and Communication Engineering*, 24(8), 970–977. DOI : 10.6109/jkiice.2020.24.8.970
- [16] J. Y. Kang & H. S. Lim. (2019). Proposal of Content Recommend System on Insurance Company Web Site Using Collaborative Filtering. *Journal of Digital Convergence*, 17(11), 201–206. DOI : 10.14400/JDC.2019.17.11.201
- [17] H. W. An & N. M. Moon. (2019). Influential Factor Based Hybrid Recommendation System with Deep Neural Network-Based Data Supplement. *Journal of Broadcast Engineering*, 24(3), 515–526. DOI : 10.5909/JBE.2019.24.3.515
- [18] E. Y. Bae & S. J. Yu. (2018). Keyword-based Recommender System Dataset Construction and Analysis. *Journal of Korean Institute of Information Technology*, 16(6), 91–99. DOI : 10.14801/jkiit.2018.16.6.91
- [19] J. E. Son & S. B. Kim & H. J. Kim & S. Z. Cho. (2015). Review and Analysis of Recommender Systems. *Journal of the Korean Institute of Industrial Engineers*, 41(2), 185–208. DOI : 10.7232/JKIIE.2015.41.2.185
- [20] Barragans-Martinez, A. B. & Costa-Montenegro, E. & Burguillo, J. C. & Rey-Lopez, M. & Mikic-Fonte, F. A. & Peleteiro, A. (2010). A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition. *Information sciences*, 180(22), 4290–4311. DOI : 10.1016/j.ins.2010.07.024
- [21] Y. J. Jung & H. J. Kim. (2020). A Study on the School Library Research Trends Using Topic Modeling. *Journal of Korean Library and Information Science Society*, 51(3), 103–121. DOI : 10.16981/kliss.51.3.202009.103
- [22] H. S. Kim & B. R. Kang. (2019). Analysis of Research Topics among Library, Archives and Museums using Topic Modeling. *Journal of Korean Library and Information Science Society*, 50(4), 339–358. DOI : 10.16981/kliss.50.201912.339

진 민 하(Minha Jin)

[학생회원]



- 2015년 3월 ~ 현재 : 한동대학교 경영학/데이터사이언스학과
- 2020년 10월 ~ 2021년 2월 : 대구디지털산업진흥원 연구인턴
- 관심분야 : 빅데이터, 인공지능, 기술경영
- E-Mail : jk03134@naver.com

정 승 연(Seungyeon Jeong) [학생회원]



- 2020년 8월 : 경북대학교 통계학, 빅데이터 전공 (이학사, 공학사)
- 2020년 10월 ~ 12월 : 대구디지털산업진흥원 공공 빅데이터 인턴
- 관심분야 : 빅데이터, 인공지능, 추천시스템, 텍스트마이닝
- E-Mail : crowsy01@gmail.com

조 은 지(Eunji Cho) [학생회원]



- 2021년 2월 : 영남대학교 컴퓨터공학 (공학사)
- 2020년 10월 ~ 12월 : 대구디지털산업진흥원 공공 빅데이터 인턴
- 관심분야 : 빅데이터, 인공지능, 웹, 소프트웨어공학
- E-Mail : dmswl_0311@naver.com

이 명 훈(Myounghun Lee) [학생회원]



- 2020년 8월 : 경북대학교 수학, 빅데이터 전공 (이학사, 공학사)
- 2020년 10월 ~ 12월 : 대구디지털산업진흥원 공공 빅데이터 인턴
- 관심분야 : 빅데이터, 인공지능
- E-Mail : audgnsdl115@naver.com

김 건 욱(Keunwook Kim) [장회원]



- 2009년 2월 : 영남대학교 도시공학(공학사)
- 2011년 8월 : 아주대학교 교통공학(공학석사-교통모델링)
- 2019년 7월 ~ 현재 : 대구디지털산업진흥원 빅데이터활용센터 센터장
- 관심분야 : 도시데이터분석, 빅데이터, 인공지능, 텍스트마이닝
- E-Mail : aut7767@dip.or.kr