

리뷰 정보를 활용한 이용자의 선호요인 식별에 관한 연구*

Identification of User Preference Factor Using Review Information

송성전 (Sungjeon Song)**

심지영 (Jiyoung Shim)***

초 록

본 연구는 도서관 정보서비스 환경에서 도서 이용자의 도서추천에 영향을 미치는 선호요인을 파악하기 위해 전 세계 도서 이용자의 참여로 이루어지는 사회적 목록 서비스인 Goodreads 리뷰 데이터를 대상으로 내용분석하였다. 이용자 선호의 내용을 보다 세부적인 관점에서 파악하기 위해 샘플 선정 과정에서 평점 그룹별, 도서별, 이용자별 하위 데이터 집합을 구성하였으며, 다양한 토픽을 고루 반영하기 위해 리뷰 텍스트의 토픽모델링 결과에 기반하여 층화 샘플링을 수행하였다. 그 결과, '내용', '캐릭터', '글쓰기', '읽기', '작가', '스토리', '형식'의 7개 범주에 속하는 총 90개 선호요인 관련 개념을 식별하는 한편, 평점에 따라 드러나는 일반적인 선호요인은 물론 호불호가 분명한 도서와 이용자에서 드러나는 선호요인의 양상을 파악하였다. 본 연구의 결과는 이용자 선호요인의 구체적 양상을 파악하여 향후 추천시스템 등에서 보다 정교한 추천에 기여할 수 있을 것으로 보인다.

ABSTRACT

This study analyzed the contents of Goodreads review data, which is a social cataloging service with the participation of book users around the world, to identify the preference factors that affect book users' book recommendations in the library information service environment. To understand user preferences from a more detailed point of view, sub-datasets for each rating group, each book, and each user were constructed in the sample selection process. Stratified sampling was also performed based on the result of topic modeling of review text data to include various topics. As a result, a total of 90 preference factors belonging to 7 categories('Content', 'Character', 'Writing', 'Reading', 'Author', 'Story', 'Form') were identified. Also, the general preference factors revealed according to the ratings, as well as the patterns of preference factors revealed in books and users with clear likes and dislikes were identified. The results of this study are expected to contribute to more sophisticated recommendations in future recommendation systems by identifying specific aspects of user preference factors.

키워드: 이용자 선호요인, 리뷰 데이터, 자연어 처리, 토픽모델링, 내용분석

user preference, review data, natural language processing, topic modeling, content analysis

* 이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2019S1A5B5A07111984).

** 독립연구자(timothy.dim12@gmail.com) (제1저자)

*** 연세대학교 대학도서관발전연구소 연구교수(jiyoung_shim@yonsei.ac.kr) (교신저자)

■ 논문접수일자: 2022년 8월 21일 ■ 최초심사일자: 2022년 9월 4일 ■ 게재확정일자: 2022년 9월 8일

■ 정보관리학회지, 39(3), 311-336, 2022. <http://dx.doi.org/10.3743/KOSIM.2022.39.3.311>

※ Copyright © 2022 Korean Society for Information Management.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

실시간으로 생산되는 정보의 양이 기하급수적으로 증가하고, 정보 선별의 중요성 역시 높아짐에 따라 추천 및 큐레이션 서비스에 대한 필요성이 커지고 있다. 도서관 정보서비스 분야에서는 Goodreads, LibraryThing과 같은 도서 이용자의 참여로 이루어지는 사회적 목록(social cataloging) 서비스의 이용도 활발히 이루어지고 있으며, 이들 서비스를 통해 이용자의 방대한 리뷰 데이터가 손쉽게 생성되고 축적되고 있다. 한편, 국내에서는 각종 도서 추천 앱 서비스를 통해 도서관과 서점을 중심으로 도서에 대한 이용자의 리뷰 데이터가 축적되고 있으며, 이를 활용한 연구가 진행되어 왔다(유다빈 외, 2015; 최지은 외, 2016).

이와 같은 사회적 흐름에 맞춰 오피니언 마이닝 분야를 중심으로 리뷰 데이터를 이용한 도서 추천과 관련된 다양한 연구가 이루어지고 있다(유다빈 외, 2015; 최지은 외, 2016; Liu et al., 2013; Sohail, Siddiqui, & Ali, 2013; Tubishat, Idris, & Abushariah, 2018). 이러한 연구들은 다양한 추천 시스템의 기법을 개발하고 적용하여 추천 시스템을 향상시키는 데 기여를 해왔지만, 이용자로부터 선호요인(preference)을 파악하여 이용자의 구체적 요구를 반영하는 데는 제한적이다. 추천 시스템을 개선하기 위해 이용자의 선호 속성을 반영하고자 한 연구들 또한 도서의 속성 정보와 이용자의 대출정보에 국한된 도서대출 데이터를 이용한 연구(Mariana et al., 2017; Tu, Chang, & Hawang, 2021; Wu, Hu, & Wang, 2017)로 이용자의 제한적 속성을 반영하고 있다. 또한 사회적 목록

서비스로부터 수집된 태그 혹은 리뷰에 대한 분석도 오피니언 마이닝 연구와 같은 양적 연구에 주로 한정되어 있어 결과로 도출된 이용자의 선호 속성의 구체적 맥락을 파악하기 어렵다.

이에 본 연구에서는 이용자 선호요인을 파악하기 위해 도서 이용자가 생성하고, 이용자의 다양한 의견과 관점이 명시적으로 표현된 리뷰 데이터를 대상으로 질적 내용분석을 수행하고자 한다. 이용자 선호의 내용을 보다 세부적인 관점에서 파악하기 위해 샘플 선정 과정에서 평점 그룹별, 도서별, 이용자별 하위 데이터 집합을 구성하는 한편, 토픽모델링을 통한 다양한 토픽을 고루 반영하여 샘플링을 수행하고자 한다. 다양한 속성과 토픽이 반영된 이용자 리뷰를 대상으로 내용분석하여 궁극적으로 도서에 대한 이용자의 선호요인과 속성 그룹별로 드러나는 선호 양상의 특성을 파악하여 향후 추천 시스템에 적용하기 위한 보다 세부적인 이용자 선호요인 및 맥락을 제시하고자 한다.

2. 이론적 배경

2.1 이용자 선호요인 관련 연구

이용자 선호요인은 이용자의 특정 아이템 혹은 서비스 이용과 관련된 의사결정에 영향을 미치는 구체적인 요인으로, 주로 비즈니스, 마케팅 분야에서 소비자가 제품이나 서비스의 어떤 요소에 가치를 두는지 파악하기 위한 목적으로 컨조인트 분석 기법 등을 활용하여 연구

되어 왔다(Netzer et al., 2008). 이들 연구에서는 이용자 선호에 영향을 미치는 주요 속성을 조합하여 설문 응답을 통해 선호도에 영향을 미치는 중요도를 파악하는 데 초점을 두고 있다. 예를 들면, 최신 영화 VOD의 이용자 선호도에 미치는 속성으로 요금, 이용기기, 최근성, 생산국을 이용자에게 속성변인으로 제시하고, 상대적 중요도를 파악하거나(임정수, 2013), 헬스케어 서비스 대상 환자들에게 대기시간, 진단을 위한 장소, 치료를 위한 장소와 같은 속성 중 어떠한 속성이 더 중요한지를 파악하는 형태를 보인다(Ryan & Farrar, 2000). 이러한 이용자 선호요인 분석을 위해서는 무엇보다 이용자에게 영향을 미치는 주요 속성을 사전에 파악하는 것이 전제되어야 하며, 기존 연구에서는 문헌리뷰, 포커스 그룹 인터뷰, 개별 인터뷰 방식이 주로 사용되어 왔다(Green & Srinivasan, 1990).

도서관 정보서비스와 관련한 이용자 선호요인을 파악하고자 한 연구는 주로 도서관 환경에서 새로운 서비스 도입과 관련한 이용자들의 요구를 파악하거나 서비스 평가와 맞물려 연구되어 왔다. 예를 들면, 디지털 도서관에서 통합 원문 DB를 탐색하는 이용자들의 선호요인을 파악하기 위해 탐색실험 및 인터뷰 내용을 내용분석한 연구(Park, 2000), 대출도서 분석과 이용자 설문을 사용하여 인쇄도서와 e-book 선호요인을 비교 분석한 연구(Wilson et al., 2021)는 이용자 요구를 파악하는 데 초점을 맞추고 있다. 한편, 설문을 통해 전자자원에 대한 이용자 선호도를 평가한 연구(Bhat & Ganai, 2018), e-book 장서개발과 관련하여 이용자 선호요인을 파악한 연구(Reyes & Devlin, 2021)는 서

비스 평가에 초점을 맞춘 연구로 간주될 수 있다. 그 외, 학습과 도서관 홍보와 관련하여 인스타그램에 대한 이용자 선호요인을 파악하고자 한 연구(Lam, Ho, & Chiu, 2022), 이용자 로그 데이터를 사용하여 도서관 공간 관련 선호요인을 파악한 연구(손은정, 박태연, 오효정, 2020)도 이루어진 바 있다.

한편, 각종 온라인 서비스 플랫폼의 등장과 탐색 시스템이 보편화 된 환경에서 이용자 선호요인 파악은 추천 시스템에서 선행되어야 할 과제로 여겨져 왔다(McNee, Riedl, & Konstan, 2006). 추천 시스템 성능을 개선하기 위해 이용자 선호의 구체적 요소를 파악하고자 한 연구들이 이루어져 왔으며, 특히 이용자의 의견과 반응이 포함된 리뷰 데이터는 이용자 선호와 관련된 풍부한 정보를 갖는 한편, 상품 및 정보 서비스 추천 과정에서 다수 사용자들의 높은 신뢰를 받는 데이터로 드러나며(Xiao, Wei, & Dong, 2016), 선호요인을 식별하는 대상으로 유용하게 간주되어 사용되어 왔다(Chen, Chen, & Wang, 2015; Liu et al., 2013).

2.2 리뷰 데이터 활용 연구

도서관 정보서비스 환경에서 리뷰 데이터를 활용한 연구는 도서관 이용자 리뷰를 분석하여 도서관 서비스 평가에 활용한 연구, 도서 리뷰를 사용하여 도서추천에 활용한 연구로 요약될 수 있다. 도서관 서비스 평가와 관련해서는 LibQUAL+ 이용자 코멘트를 내용분석하여 대학도서관 서비스 평가에 이용한 바 있다(변제연, 심원식, 2013; Begay et al., 2004). 도서 추천과 관련해서는 리뷰 텍스트로부터 의견을

표출한 부분과 그 대상을 자동으로 추출하는 오피니언 마이닝 기법이 주로 적용되어 왔으며, 이를 통해 추출된 의견어 등에 기반하여 이용자의 선호를 예측하는 형태로 활용되어 왔다(유다빈 외, 2015; 최지은 외, 2016; Liu et al., 2013; Sohail, Siddiqui, & Ali, 2013; Tubishat, Idris, & Abushariah, 2018).

특히 도서 리뷰를 활용한 연구는 온라인 서점은 물론 도서 관련 모바일 플랫폼을 통해 축적된 데이터에 기반하여 도서추천을 목적으로 활발히 수행되고 있다. 예를 들면, Sohail, Siddiqui, Ali (2013)는 웹 검색을 통해 도서의 리뷰를 수집하고, 구성된 의견어 목록을 바탕으로 속성 추출 및 감성분석을 통해 도서를 순위화하여 추천에 적용하였다. 최지은 외(2016)는 온라인 서점의 이용자 리뷰로부터 각 도서에 대한 테마 핵심어를 추출하고 리뷰에 대한 감성분석을 실시하여, 이용자에게 도서에 대한 이용자 반응과 주요 테마 핵심어를 함께 제공하는 시스템을 구현한 바 있다. 한편, Goordreads 리뷰 데이터는 전 세계 독자들의 서평과 추천이 공유되며, WorldCat의 서지레코드에 외부 정보원으로 연결되어 있으며(윤정옥, 2012), 2022년 9월 현재 85개 공공도서관 그룹과 협력하며 도서관 목록에 리뷰 서비스를 연계하고 있는 등(Goordreads, 2022), 도서관 정보서비스 분야에서 도서이용의 선호를 파악하는 도구로 간주되어 활용되어 왔다(Thelwall & Kousha, 2017). Goordreads를 이용한 도서추천과 평가(Kadiresan, Singson, & Thiyagarajan, 2021; Kousha, Thelwall, & Abdoli, 2017)는 물론 독자자문 서비스와 연관되어 리뷰의 활용이 적극적으로 논의되어 왔다(Naik & Trott, 2012).

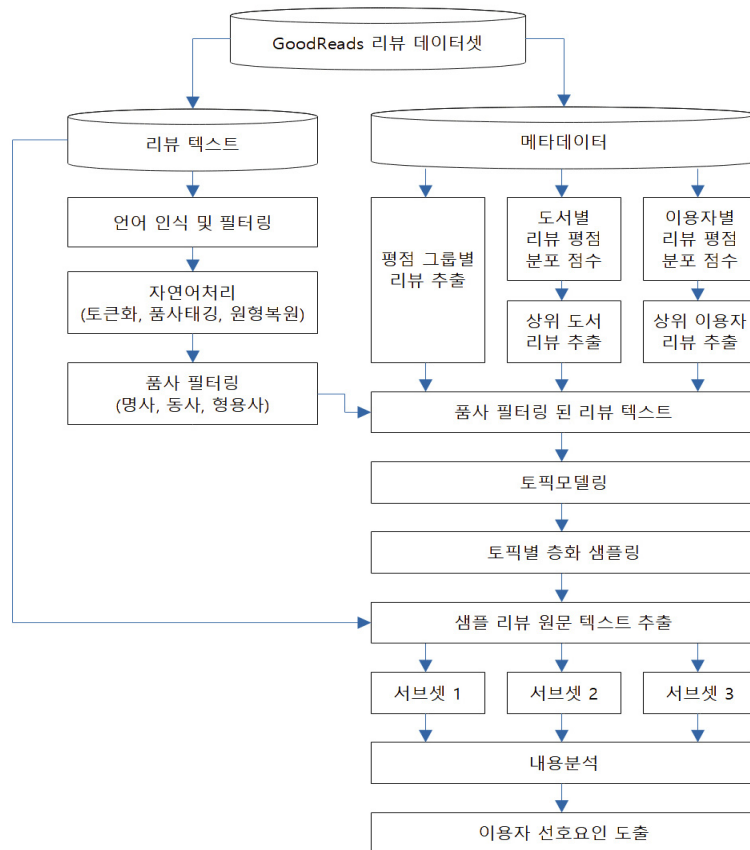
또한, 리뷰 데이터는 주제기반 오피니언 마이닝 기법인 토픽모델링을 적용하여 리뷰 데이터의 내용에 따른 주제의 종류를 파악하는 연구가 진행되어 왔다. Sun et al.(2015)은 의견어를 반영한 토픽모델링을 제안했고, 이를 통해 의견어와 함께 출현하는 키워드 중심으로 토픽이 구성될 수 있도록 하였다. Korfiatis et al.(2019)은 TripAdvisor 사이트에서 비행기 승객들의 리뷰를 수집해 STM 토픽모델링 기법을 통해 토픽과 키워드를 추출해 각 요소에 따른 서비스 품질을 측정하는 연구를 수행하였다. Shah et al.(2021)은 환자들의 리뷰를 수집하고, 토픽모델링과 SentiNet을 통해 토픽 키워드와 그에 대한 감성분석을 수행하였다. 이를 통해, 환자들이 만족하는 요소와 불만족하는 요소를 식별하였다. 이처럼 토픽모델링은 방대한 텍스트에서 주제에 따른 키워드를 추출하는데 주로 이용되며 전체 데이터의 토픽을 파악하는데 적용되어 왔다. 이와 같이 리뷰 데이터는 주로 오피니언 마이닝을 통해 속성 및 의견어를 자동 추출되어 활용되어 왔지만, 추출된 단어들 간의 관계 및 연관성은 구조적으로 파악하기 어렵다는 한계가 있다.

이에 본 연구에서는 이용자의 다양한 의견과 관점이 명시적으로 표현된 리뷰 데이터를 대상으로 선호요인을 파악하기 위해 이용자의 관점을 파악하는데 유용한 질적 내용분석을 통해 다양한 속성과 개념을 식별하고자 한다. 이와 같은 질적 연구방법은 최근 도서추천과 관련하여 도서 이용자의 세부적인 관점을 파악하는데 사용되어 온 방법이기도 하다(김성훈, 노윤주, 김미령, 2021; 심지영, 2022). 본 연구에서는 이용자 선호의 보다 세부적인 관점을 파악

하기 위해 샘플 선정 과정에서 평점 그룹별, 도서별, 이용자별 하위 데이터 집합을 구성하는 한편, 토픽모델링을 통해 다양한 토픽을 고루 반영하여 샘플링을 수행하여 기존 연구에서 토픽모델링이 지닌 장점을 적용하고자 한다. 궁극적으로 다양한 속성과 토픽이 반영된 이용자 리뷰를 대상으로 내용분석하여 도서에 대한 이용자의 선호요인과 속성 그룹별로 드러나는 선호 양상의 특성을 파악하여 향후 추천 시스템에 적용하기 위한 세부적인 이용자 선호요인을 제시하고자 한다.

3. 연구방법

본 연구는 도서에 대한 이용자 선호요인을 식별하기 위해 이용자가 생성한 리뷰 데이터를 대상으로 내용분석하였다. 이용자 선호의 내용을 보다 세부적인 관점에서 파악하기 위해 샘플 선정 과정에서 세 가지-평점 그룹별, 도서별, 이용자별- 하위 데이터 집합을 구성하였으며, 다양한 토픽을 고루 반영하기 위해 리뷰 텍스트를 대상으로 자연어처리 과정과 토픽모델링을 통하여 토픽별 층화 샘플링을 수행하였다. 본 연구의 연구설계는 <그림 1>과 같다.



<그림 1> 연구 설계 개요

3.1 리뷰 데이터 수집

리뷰 데이터 수집은 UCSD Book Graph의 Goodreads Book Reviews 데이터셋(Wan et al., 2019)을 대상으로 하였다. 본 연구에서 사용한 리뷰 데이터셋은 UCSD(University of California San Diego)에서 2017년 후반 Goodreads 사이트의 도서 리뷰 데이터를 수집한 것으로, 약 45만 명의 도서 이용자가 작성한 약 1,500만 개의 다양한 도서 장르(아동(Children), 코믹/그래픽(Comics & Graphic), 환상/초자연(Fantasy & Paranormal), 역사/전기(History & Biography), 미스터리/스릴러/범죄(Mystery, Thriller, & Crime), 시(Poetry), 로맨스(Romance), 청소년(Young Adult))의 리뷰 데이터를 포함하고 있다. 각 리뷰 데이터는 사용자 식별정보(user_id), 도서 식별정보(book_id), 리뷰 식별정보(review_id), 평점(rating), 리뷰 텍스트(review_text), 작성일자(date) 등을 포함하고 있으며, 이를 통해 특정 이용자가 작성한 리뷰를 추출하거나, 특정 도서에 관한 리뷰 데이터를 식별하여 추출할 수 있었다.

3.2 샘플 선정

본 연구에서는 이용자 선호의 내용을 보다 세부적인 관점에서 파악하기 위해 샘플 선정에서 세 가지-평점 그룹별, 도서별, 이용자별 하위 데이터 집합을 구성하는 한편, 다양한 토픽을 고루 반영하기 위해 리뷰 텍스트를 대상으로 자연어처리 과정과 토픽 모델링을 통하여

토픽별 층화 샘플링을 수행하였다.

먼저 이용자의 세부 선호요인을 파악하기 위한 하위 데이터 집합은 평점 그룹별, 도서별, 이용자별의 3가지로 구성하였고, 각 하위 데이터 집합은 평점이 높은(5점 척도 중 4, 5) 그룹과 낮은(5점 척도 중 1, 2) 그룹이 균등하게 분포하도록 구성하였다. 첫째, 도서 혹은 이용자와 무관하게 단순히 평점이 높고 낮음에 따른 일반적인 이용자 선호 양상을 볼 수 있는 하위 데이터 집합은 샘플링 프레임인 UCSD Book Graph의 Goodreads Book Reviews(Wan et al., 2019)의 모든 리뷰 데이터를 대상으로 평점이 높고 낮음으로만 구분하여 생성하였다. 둘째, 동일한 도서에 대해서도 낮은 평점과 높은 평점이 나타나는 논쟁적인 상황 등에서 나타나는 도서별 선호요인을 파악할 수 있는 하위 데이터 집합은 도서에 대한 리뷰 수가 높은 평점과 낮은 평점 최소 각 10개 이상인 경우를 대상으로 평점이 높고 낮은 리뷰 수의 비율을 균등하게 조정하여(공식1)¹⁾ 순위화한 결과 상위 100개의 도서 및 리뷰 데이터를 선정하여 구성하였다. 셋째, 특정 이용자를 기준으로 도서에 대한 선호요인의 양상을 파악하기 위한 하위 데이터 집합은 리뷰 수가 최소 10개 이상인 이용자를 대상으로 공식 1에 의하여 높은 평점과 낮은 평점의 리뷰 수의 비율을 균등하게 조정하여 상위 100개 이용자 및 리뷰 데이터를 선정하여 구성하였다.

다음으로 샘플 선정에서 다양한 토픽을 고루 반영하기 위해, 앞서 구성된 세 가지-평점 그룹별, 도서별, 이용자별- 하위 데이터 집합 리뷰

$$1) \text{ Score} = \left| \frac{n(\text{review}_{low}) - n(\text{review}_{high})}{n(\text{review}_{all})} \right| \quad (\text{공식 1})$$

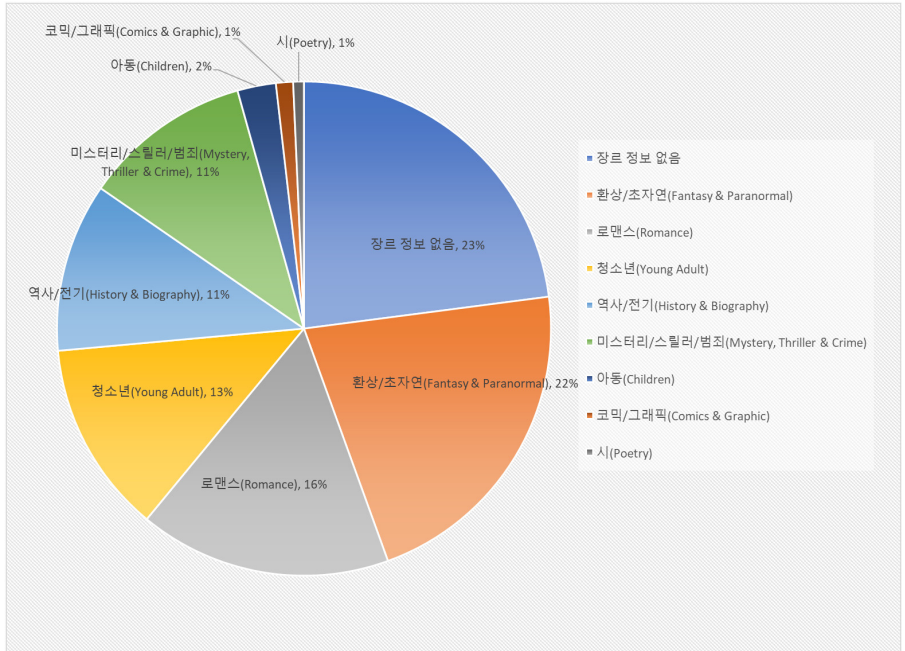
텍스트를 대상으로 자연어처리 과정과 토픽모델링을 통하여 토픽별 층화 샘플링을 수행하였다. 샘플링을 위한 토픽모델링 결과를 향상시키기 위해 리뷰 텍스트에 대한 전처리를 언어인식(language identification), 품사태깅 필터링(part-of-speech tagging filtering), 원형복원(lemmatization) 등의 과정을 거쳤다. 이를 통해, 전체 리뷰 중 영어로 된 리뷰만 선정하였고, 리뷰 텍스트 중 명사, 형용사, 동사 품사에 해당하는 단어만으로 필터링하고, 원형복원된 형태로 토픽모델링을 수행하였다. 언어인식은 fastText 파이썬 모듈을 통해 이루어졌고(Joulin et al., 2017), 품사태깅과 원형복원은 Stanza 파이썬 모듈로 수행하였다(Qi et al., 2020).

최종적으로 자연어 처리를 거친 세 가지-평점 그룹별, 도서별, 이용자별- 하위 데이터 집합을 1) 전체 리뷰를 평점에 따른 구분만 한 일반 그룹, 2) 도서를 중심으로 호불호가 나뉘는 리뷰로 구성된 도서 중심 그룹, 3) 이용자를 중심으로 호불호가 나뉘는 리뷰로 구성된 이용자 중심 그룹으로 명명하고, 각 리뷰 데이터 집합 내에서 평점이 낮은 그룹과 높은 그룹으로 각

각 구분하여 리뷰 텍스트를 대상으로 토픽모델링을 통한 토픽별 층화 샘플링을 수행하였다. 이를 통해 리뷰 텍스트에 포함된 토픽별 주요 키워드를 식별하고 그룹화하여 다양한 주제와 관점의 리뷰가 샘플에 포함되도록 하였으며 각 토픽별 10개의 리뷰를 샘플로 선정하였다. 예를 들면, <표 1>과 같이 일반 그룹 중 낮은 평점 그룹의 토픽모델링 결과 11개 토픽 그룹에서 상위 10개의 토픽에 대응되는 총 110개의 리뷰를 선정하였으며, 마찬가지로 높은 평점 그룹에서도 토픽 11개에 대한 110개의 리뷰 샘플을 선정하여 총 220개의 리뷰 샘플을 선정하였다. 마찬가지로의 방식으로 도서를 중심으로 호불호가 나뉘는 리뷰로 구성된 도서 중심 그룹에서는 10개의 토픽에서 낮은 평점 그룹과 높은 평점 그룹에서 각 100개의 총 200개 리뷰 샘플을 선정하였으며, 이용자를 중심으로 한 호불호가 나뉘는 리뷰로 구성된 이용자 중심 그룹에서는 10개의 토픽에서 낮은 평점과 높은 평점 그룹에서 각 100개, 총 200개 리뷰 샘플을 선정하여 최종적으로 세 개의 하위 집합 내에서 평점이 높고 낮음을 구분하여 총 620개 리뷰 샘플을 선

<표 1> 토픽모델링 결과 예시(일반 그룹 중 낮은 평점 그룹)

토픽1	토픽2	토픽3	토픽4	토픽5	토픽6	토픽7	토픽8	토픽9	토픽10	토픽11
history	character	character	story	story	people	sex	story	character	world	time
author	girl	story	character	comic	author	love	year	series	people	finish
time	thing	life	plot	art	work	woman	family	end	power	page
life	friend	woman	author	movie	way	man	life	story	fantasy	story
people	time	end	time	work	life	thing	mother	love	time	star
work	end	time	end	fan	thing	time	friend	time	vampire	author
war	way	plot	page	collection	world	heroine	child	start	character	review
man	people	reader	thing	volume	idea	guy	father	thing	plot	thing
fact	school	way	way	film	need	scene	time	plot	fight	love
page	guy	author	review	writer	time	way	work	finish	series	people



〈그림 2〉 도서 리뷰 샘플의 장르 구성(N=620)(도서 장르는 중복 분류됨)

정하였다. 샘플 리뷰 텍스트의 평균 길이는 804자로, 가장 짧은 리뷰는 31자, 가장 긴 리뷰는 7,952자로 나타났으며, 도서 장르 구성은 〈그림 2〉와 같다.

3.3 내용분석

평점 그룹별, 도서별, 이용자별로 구분한 세 가지 유형의 영어 도서 리뷰 샘플(N=620)을 대상으로 이용자 선호요인 식별을 위한 내용분석을 수행하였다. 내용분석은 두 명의 코더에 의해 오픈코딩 방식으로 2회에 걸쳐 반복 수행하였다. 리뷰 텍스트에서 이용자의 명시적 선호 반응이 드러난 문장(예, I particularly liked the twist on the wolf in Red Riding Hood.)을 대상으로 개념 코드(예, Story-plot twist)

를 부여하는 한편, 본 연구에서는 선호 반응에서 나타날 수 있는 속성이 불호(예, The ending was meh.)의 형태로 드러난 문장의 경우(예, The ending was meh.)도 개념 코드(예, Story-ending)를 부여하며 코딩 체계를 수립하였다. 선호요인을 파악하는 과정에서 낮은 평점 그룹을 비선호요인으로 분리하지 않고 동시에 고려한 이유는 이용자 개개인의 취향의 양면성을 고려하여 특정 도서 혹은 이용자에 따라 드러나는 선호요인의 양상을 복합적으로 파악하고자 함이다. 본 연구에서는 선호요인을 기준으로 해당 요인이 긍정과 부정의 형태로 어떤 양상을 보이는지를 파악하고자 하였다. 내용분석 과정에서 하나의 quotation에 두 개 이상의 개념이 나타나는 경우 중복 코딩을 허용하였다.

코딩 체계 수립 과정에서 김리연, 박은경(2011)과 박동진(2010)의 연구에서 제시한 독후감의 구성요소 가운데 도서에 대한 이해와 연관된 요소(예, 줄거리, 구성, 주제, 형식적 정보, 문체, 장르)와 도서의 내용적 정보요소(예, 줄거리, 구성/문체 등)를 참고하였다. 한편, 대량의 데이터를 일관되게 처리하기 위하여 질적 내용분석 소프트웨어인 ATLAS.ti 7.5.4를 이용하였다. 내용 분석을 통해 최종적으로 내용(Content), 캐릭터(Character), 글쓰기(Writing), 읽기(Reading), 작가(Author), 스토리(Story), 형식(Form)의 7개 범주에 속하는 총 90개 선호요인 관련 개념을 식별하였다.

4. 연구결과

본 연구는 도서 이용자 선호요인을 식별하기 위해 Goodreads 이용자가 생성한 도서 리뷰 데이터를 대상으로 분석하였다. 이용자 선호의 내용을 세부적인 관점에서 파악하기 위해 세 가지 속성(평점 그룹별, 도서별, 이용자별)으로 구분하는 한편, 토픽모델링 결과를 반영하여 다양한 토픽으로 구성된 620개의 리뷰 샘플을 대상으로 내용분석하였다. 그 결과, 내용(Content), 캐릭터(Character), 글쓰기(Writing), 읽기(Reading), 작가(Author), 스토리(Story), 형식(Form)의 7개 범주에 속하는 총 90개 선호요인 관련 개념을 식별하였다(〈표 2〉~〈표 8〉 참고).

4.1 도서 이용자의 선호요인

이용자 리뷰 데이터의 내용분석을 통해 식별

된 도서 이용자의 선호요인으로는 7가지 범주로, 이 가운데 도서 이용자들의 도서 이해 및 내용적 측면과 관련된 범주는 스토리(Story), 캐릭터(Character), 내용(Content)으로 드러났으며, 도서 창작자와 관련된 범주로 글쓰기(Writing)와 작가(Author), 도서 이용자의 행위와 관련된 범주로 읽기(Reading) 관련 개념들이 식별되었다. 그 외, 도서와 관련된 외형적 측면과 관련된 형식(Form)까지 총 7가지 범주로 선호요인 관련 개념들이 나타났다.

리뷰 데이터의 분석 결과, 도서 이용자들의 선호요인은 도서 이해 및 내용적 측면과 관련된 범주인 스토리(Story), 캐릭터(Character), 내용(Content)에서 각각 23.5%, 17.3%, 16.8%로 절반 이상(57.6%)의 비중을 차지하고 있어 이용자들의 관심이 집중되는 것을 볼 수 있었다. 본 연구에서 스토리(Story)와 캐릭터(Character)를 도서 내용(Content)과 구분한 이유는, 도서 내용과 별도로 스토리와 캐릭터 관련 하위 개념들이 다양한 양상으로 드러났으며 많은 비중을 차지했기 때문이다. 또한, 스토리와 캐릭터는 소설, 희곡 같이 이야기 구조를 가지는 장르에 국한되는 반면, 도서 내용은 사실정보를 다루는 논픽션 등 다양한 장르를 포괄하며 도서 내용의 소재, 깊이, 설득력과 연관되어 구분되는 양상으로 나타났다기 때문이다.

〈표 2〉에서 보이는 바와 같이 도서 이용자의 스토리 관련 선호요인은 총 18개의 개념이 드러났으며, 특히, 스토리 전개(development), 스토리 전반(general), 스토리에 대한 흥미, 관심사(interest)와 같이 이야기의 특정 부분을 의미하기보다 전체적인 구조 혹은 그에 대한 인상과 관련한 선호가 두드러지게 드러났다. 예를

〈표 2〉 도서 이용자의 스토리(Story) 관련 선호요인

선호요인	정의
Story	이야기 구조를 가지는 도서에서 나타나는 선호요인
general	스토리 관련 전반
coherent	스토리의 응집력, 이야기 구조 내에서 의미가 연결되는 정도
compelling plot	플롯의 설득력
consistency	이야기 전체가 일관된 어조를 가지는지, 모순은 없는지 정도
detail	이야기의 서술에서 세부적인 내용들이 제대로 표현되는 정도
development	스토리의 전개 과정
empathy	스토리에 대한 공감
ending	스토리의 결말
interest	스토리에 대한 흥미, 관심사
narrative	실화나 허구의 사건들을 묘사하고 표현하는 구조적 형식
pace	스토리 전개의 속도
plot twist	스토리의 반전
premise	스토리에 전제된 설정(예, 인물의 직업, 환경, 경험, 연령 등)
reality	스토리의 현실성
scene	스토리에 묘사된 장면
setting	스토리의 배경
tension	스토리가 주는 긴장감
theme	스토리의 테마, 주제

들면, 스토리 전개와 관련하여, '이 책의 도입부는 매우 강렬하다', '이 이야기는 가장 중요한 플롯을 미묘하게 진행하면서 충분히 다양한 전개를 보인다'와 같은 내용이 언급되었으며, 스토리 전반과 관련해서는 '나는 이 이야기를 좋아한다', '이것은 훌륭한 이야기꾼이 들려주는 훌륭한 이야기이다'와 같이 이야기 전반에 대한 선호를 드러낸 바 있다. 스토리에 대한 흥미, 관심사와 관련해서는 '이야기에 많은 액션이 있어서 좋았다'와 같은 내용이 선호하는 내용으로써 언급되었다.

스토리의 결말(ending), 배경(setting), 인물의 직업, 환경 등 스토리에 전제된 설정(premise), 스토리의 주제(theme) 등과 같이 이야기의 구성 성분에 대한 선호도 드러났다. 예를 들면, 스

토리의 결말과 관련해서는 '아무리 끔찍한 일을 겪더라도 삶이 희망으로 가득차 있다는 것으로 끝나는 이 엔딩을 좋아한다'와 같이 결말 부분에 집중해서 좋은 반응을 드러낸 바 있으며, 'Pendurra, Cornwall은 매우 매력적인 배경이었다'와 같이 스토리의 배경에 집중한 선호 개념도 나타났다. 스토리에 전제된 설정과 관련해서는 '엘리자베스는 1896년 14세의 나이로 11분 동안 죽어있는 동안 수백 년이 흐른다... 이런 컨셉이 너무 마음에 들었고...'와 같이 스토리와 관련해서 작가가 부여한 설정과 관련한 선호를 드러냈으며, 스토리의 주제와 관련해서는 '나는 이야기에서 예술과 구원을 연결하는 테마를 좋아한다'와 같이 스토리가 있는 도서에서의 주제에 대한 선호 개념이 드러났다.

한편, 본 연구에서는 선호 반응에서 나타날 수 있는 속성(예, Story-ending)을 기준으로 동시에 불호(예, The ending was meh.)의 형태로 드러난 양상(예, The ending was meh.)을 식별하여 이용자 취향에 따라 선호요인의 긍, 부정 양상의 정도를 파악하고자 하였다. 그 결과, 플롯의 설득력(compelling plot), 스토리의 일관성(consistency), 스토리가 주는 긴장감(tension) 개념은 주로 높은 평점보다 낮은 평점의 리뷰에서 불호와 관련된 내용으로 식별되었으며, 특히 스토리의 일관성은 불호의 형태로만 드러나고 있어 선호요인보다 비선호요인으로 작용하는 특성을 드러냈다(〈표 9〉참고). 예를 들면, '이 이야기의 로맨스는 따분하고 믿기 어렵다'와 같이 플롯의 설득력과 관련된 개념이나, '긴장감이 없으면 Helprin(작가)의 소설을 더 이상 읽을 이유가 없는데, 이 이야기는 긴장감이 전혀 없다', '스토리 전반에 걸쳐 일관성이 더 있어야 한다고 생각했다'와 같이 스토리의 긴장감과 일관성은 선호요인보다 비선호요인으로 나타나는 양상을 드러냈다.

도서 이용자들의 선호요인 중 도서 이해 및 내용적 측면과 관련되어 스토리(Story)와 더불어 많은 비중을 차지하는 개념이었던 캐릭터(Character) 관련 선호요인은 총 10개의 개념으로 드러났다(〈표 3〉참고). 도서 이용자들은 이야기에 등장하는 인물들의 행동(behavior), 성격(personality), 캐릭터 간 관계 혹은 상호작용(relationship), 캐릭터가 작품에서 나누는 대화(dialogue), 생각(thought)까지도 선호요인으로 드러내고 있었다. 가령, 사려 깊고 배려심 많은 캐릭터에 대한 선호를 드러내는가 하면(예, '나는 Willowdean(캐릭터)이 인간의 본성에 대해 지니고 있는 기발하고 사려 깊고 미묘하고 진정한 통찰력을 좋아했다'), 자신의 인간에 대한 취향을 반영한 캐릭터에 호감을 표명하기도 했다(예, 'Samantha(캐릭터)는 똑똑하고 기발한 캐릭터로 내가 평소 좋아하는 성격을 지녀서 좋다').

캐릭터의 성격, 행동과 같이 개인의 취향과 관련된 선호요인 외에도 캐릭터에 대한 도서 이용자의 평가가 반영된 선호요인으로 캐릭터가 이

〈표 3〉 도서 이용자의 캐릭터(Character) 관련 선호요인

선호요인	정의
Character	소설, 희곡 등의 장르에서 나타나는 등장인물에 대한 이용자의 선호요인
general	캐릭터 전반
behavior	캐릭터의 행동
development	캐릭터가 작품에서 형성되어 가는 과정
dialogue	캐릭터가 작품에서 나누는 대화
empathy	캐릭터에 대한 공감, 감정이입
personality	캐릭터가 지니는 성격, 인성
propensity	캐릭터가 작품에서 드러내는 경향, 성향
reality	캐릭터의 현실성
relationship	캐릭터 간 설정된 관계나 상호작용
thought	캐릭터가 작품에서 생각하는 내용, 내부 감정

야기 속에서 잘 형성되었는지와 관련되는 캐릭터 형성(development)(예, '그녀의 캐릭터 형성은 다양하고 멋지게 개발되었다'), 캐릭터의 현실성(reality)(예, '유머와 재치있는 캐릭터가 현실감 있어 책을 멈추지 않고 읽게 되었다')과 같은 선호요인도 드러났으며, 드물지만 캐릭터의 성향(prospensity)(예, '딸이 결혼하는 데 집착하는 엄마의 뻔한 클리셰')도 나타났다.

내용(Content) 관련 선호요인은 도서 내용의 전반적인 요인들로 이야기 구조를 가진 소설, 희곡 장르와 구분되며 주로 사실정보를 다루는 논픽션 등 다양한 장르의 도서 내용의 소재, 장르, 내용의 깊이 등을 포함하며 총 20개의 개념

을 드러내고 있다(〈표 4〉 참고). 내용 관련 선호요인은 도서 내용 관련 정보적 성격을 지닌 개념들(예, 주제(subject), 소재(material), 도서 장르(genre), 길이(length), 도서에서 제공되는 정보적 내용(information), 도서에서 언급된 사실정보(fact) 등)과 이용자의 도서 내용 자체에 대한 평가적 성격을 지닌 개념들(예, 도서 내용의 독창성(originality), 도서 내용의 정확성(accuracy), 도서 내용의 설득력(convincing), 도서 내용의 밀도, 주제 응집성(cohesive))도 함께 드러났다.

한편, 내용 관련 선호요인에서 도서에서 제공되는 정보적 내용과 관련된 소재(material), 장

〈표 4〉 도서 이용자의 내용(Content) 관련 선호요인

선호요인	정의
<i>Content</i>	도서 내용의 전반적인 선호요인
general	도서의 내용 전반
accuracy	도서 내용의 정확성
cohesive	도서 내용의 밀도, 주제 응집성
convincing	도서 내용의 설득력
depth	도서 내용의 깊이
detail	도서 내용의 디테일, 충분한 기술, 설명
evidence	도서 내용 서술의 근거
fact	도서에서 언급된 사실정보
focus	도서 내용의 요지, 골자, 초점
genre	도서 장르
harmony	도서 내용 기술의 전반적인 조화
information	도서에서 제공되는 정보적인 내용
length	도서 내용의 길이
material	도서 내용의 소재(예, 음악, 인종 등)
originality	도서 내용의 참신성, 독창성
quality	도서 내용의 품질
readers' interest	독자의 관심사와 관련된 도서 내용
reference	도서 내용에 대한 레퍼런스
structure	도서 내용의 전체적인 구조
subject	도서 내용이 다루는 주제

르(genre), 주제(subject)는 주로 높은 평점의 리뷰에서 선호의 형태로 드러난 반면, 이용자의 도서 내용 자체에 대한 평가와 관련되는 선호요인인 도서 내용의 독창성(originality), 정확성(accuracy)은 낮은 평점의 리뷰에서 불호의 형태로 식별되는 경향을 보였다. 예를 들면, 도서의 내용 속에 특정 소재(예, '오드리 헵번의 삶과 라이프 스타일에 대한 내용으로 빠르게 읽을 수 있었다')가 나타났을 때 그에 대한 선호를 드러낸 반면, 도서의 독창성 선호요인(예, '끔찍한 내러티브가 있는 또 다른 진부한 사랑 이야기이다')과 도서의 정확성 선호요인(예, '나는 이 책의 일부 섹션이 너무 노골적으로 부정확했기 때문에 이 책의 일부를 읽은 후 소름이 돋았다')과 관련해서는 주로 불호의 요인과 연관 짓는 양상을 보였다.

앞서, 도서 이용자들의 도서 이해 및 내용적 측면과 관련된 선호요인 범주로 스토리(Story), 캐릭터(Character), 내용(Content)이 드러났다면, 읽기 관련 선호요인은 도서 이용자의 읽기 행위와 관련된 요인들로 이용자가 도서와 상호작

용하는 과정에서 나타나는 다양한 사고과정이나 감정적 반응과 주로 연관되며, 전체 선호요인 개념들 가운데 약 5분의 1(21.5%)을 차지하고 있는 것으로 나타났다(〈표 5〉, 〈표 9〉 참고). 예를 들면, 도서 이용자에게 즐거움, 기쁨을 제공(pleasure)하거나, 읽는 과정에서 몰입을 경험(absorption)하게 하거나, 삶의 지침, 통찰력 등 영감을 주거나(inspiring), 감정을 자극하는 것(touching)과 관련하여 선호를 드러내고 있었다. 읽기 과정에서 얼마나 수월하게 읽히느냐와 관련된 개념(readability)도 선호요인으로 드러나고 있지만, 이용자의 감정 반응과 관련한 개념들이 더 다양하게 선호요인으로 나타났다.

흥미로운 것은 읽기 과정에서 도서 이용자의 경험 혹은 사전 지식이 선호요인에 영향을 미치는 것으로 드러났다. 예를 들면, 관련되는 작품에 대한 이용자의 관심(related works)과 해당 작품 혹은 관련 작품을 읽었던 이용자의 이전 경험(previous experience)으로, 이용자의 과거 경험이 현재 도서의 선호에 영향을 미치

〈표 5〉 도서 이용자의 읽기(Reading) 관련 선호요인

선호요인	정의
Reading	도서 이용자의 읽기 행위와 관련된 선호요인
absorption	이용자의 읽기 과정에서의 몰두, 몰입
inspiring	이용자에게 통찰력, 삶의 지침, 영감을 제공
memorable	이용자에게 인상적인 기억을 제공
pleasure	이용자에게 즐거움, 기쁨을 제공
previous experience	해당 작품 혹은 관련 작품을 읽었던 이용자의 이전 경험
readability	이용자가 읽기 쉬운 정도
recommendation	이용자가 다른 사람에게 해당 작품을 추천하고자 하는 의향
related works	관련되는 작품에 대한 이용자의 관심
situation	이용자의 상황(예, 여름에 읽기 좋은 책 등)
stimulating	이용자의 지적인 호기심 자극
touching	이용자의 감정 자극
worth	이용자에게 읽을 가치 제공

거나 현재의 도서의 경험이 관련 저작들에까지 영향을 미치는 것을 보여준다. 예컨대, '나는 이 시리즈를 너무 좋아한다'와 같이 과거 경험이 현재의 도서의 선호에 영향을 미치기도 하고, '이 작가의 다른 작품도 보고 싶다'와 같이 관련 저작에 대한 선호로 이어지기도 했다. 그와 반대로 이전의 작품이나 작가에 대한 좋지 않은 경험이 관련되는 다른 작품을 읽기 목록에서 제외하게 만드는 요인으로도 드러났다.

도서 창작자와 관련된 도서 이용자의 선호요인으로는 글쓰기(Writing)와 작가(Author)로 나타났다. 이 가운데 글쓰기에 관한 이용자의 선호요인은 작품에 나타난 화자의 시점, 등장 인물의 성격묘사, 문체, 표현, 어조, 용어의 사용과 같이 구체적인 글쓰기 요소와 관련되어 호감을 드러내는 경우로 총 13개 관련 요인이 식별되었다(〈표 6〉 참고).

글쓰기 선호요인에서는 특히 화자의 시점(POV), 캐릭터 형성(character formation), 캐

릭터의 성격묘사(characterization)와 같이 소설 장르 관련 선호요인들이 주로 나타나는 가운데, 글쓰기에서의 용어 사용(use of terms)과 같이 전문적인 기술방식과 관련한 선호요인도 식별되었다. 예를 들면, 화자의 시점과 관련해서는 '나는 특히 스토리를 더 풍부하게 하고 주인공 시점에서만 볼 수 있는 내용을 엿볼 수 있어 좋았다'와 같은 내용이 언급되는가 하면 '그녀는 글에서 '내면 아이'와 같은 진부한 심리학 용어를 두 번 이상 사용한다'와 같은 경우처럼 용어 사용과 관련하여 부적절했을 때 불호를 드러내거나, 전문용어를 적절히 사용할 경우(예, '전문용어가 거슬리지 않게 기술되어 좋았다')에 대한 선호를 드러낸 경우도 나타났다.

다음으로 도서 창작자와 관련된 도서 이용자의 선호요인으로는 작가(Author)로, 작가의 저술에 직접적으로 영향을 미치는 작가의 생각(idea), 견해, 관점(perspective), 정치적 올바름에 대한 견해(perspective-political correctness)와

〈표 6〉 도서 이용자의 글쓰기(Writing) 관련 선호요인

선호요인	정의
Writing	작가의 글쓰기 방식과 관련된 선호요인
general	도서의 글쓰기 전반
character formation	글쓰기에서의 캐릭터 형성
characterization	캐릭터의 성격묘사, 인물묘사와 관련된 글쓰기
conveying context	글쓰기에서의 문맥, 맥락 전달
description	글쓰기에서의 서술방식
expression	글쓰기에서의 표현방식, 표현된 내용
language	글쓰기에서 사용된 언어사용(외국어 포함)
narration	글쓰기에서 이야기를 진행하고, 해설하는 방식
POV(point of view)	글쓰기에서의 화자의 시점(예, 1인칭 주인공 시점, 3인칭 전지적 작가 시점 등)
style	글쓰기에서의 문체
target audience	글쓰기에서의 대상 독자
tone	글쓰기에서의 어조
use of terms	글쓰기에서 용어, 특히 전문용어의 사용

같이 작가의 가치체계, 세계관과 관련된 선호요인이 나타났다(〈표 7〉 참고). 예를 들면, 작가의 관점과 관련하여 작가가 드러낸 태도를 좋아하거나(예, 'Marquez(작가)가 취하는 저널리즘적 태도는 감탄스럽다'), 정치적 올바름과 관련된 견해와 관련하여 불호를 드러낸 경우(예, 소설을 읽으면 당혹스럽고 최악의 경우 여성혐오적이라는 인상을 준다)의 경우다.

한편, 이용자가 작가에 대한 이미 지니고 있는 정보가 선호요인에 작용하는 경우로 특정 작가에 대한 좋은 경험(good experience with the author)이나 작가가 지니는 명성(reputation), 작가의 삶(life)과 같은 개념도 선호요인으로 식별되었다. 작가에 대한 좋은 경험(예, '이것은 내가 지난 몇 주 동안 읽은 Gayle Forman(작가)의 세 번째 책이며, 그녀는 내가 가장 좋아하는 청소년 현대 작가 중 한 명이다')의 경우, 작가의 전작에 대한 읽기 경험의 만족도와 관련되는 것으로 작가에 대한 선호가 해당 작가의 도서의 선호로 이어지는 경우로 해석될 수

있다. 작가의 명성과 관련해서는 '이 책은 Jeff Ashton(작가)에 의해 쓰여졌기 때문에 나는 그것을 읽기로 결정했다'나 'Atwood(작가)는 캐나다의 전설이며 이 책은 그 이유를 쉽게 설명한다'와 같은 형태로 선호를 드러냈으며, 작가의 삶과 관련해서는 작가의 삶의 면모(예, '이 이야기는 작가의 부모와 조부모의 실화로 나는 애국심이 강한 그들의 이야기를 좋아한다')와 주로 연관되어 이용자의 선호가 나타났다. 그 외, 작가가 지닌 역사, 통계에 대한 지식 등 작가의 능력(ability)에 대한 개념과 작가가 주제나 소재를 해석하는 방식(interpretation of the subject matter) 또한 작가 관련 선호요인으로 식별되었다.

마지막으로 도서 이용자의 형식(form) 관련 선호요인으로는 도서의 편집 상태(editing), 번역(translation), 삽화(illustration), 도서의 표지(book cover), 오디오북(audiobook) 등 주로 도서의 형식적인 요소와 관련되는 것으로 나타났다(〈표 8〉 참고). 편집과 관련해서는 문법적 오

〈표 7〉 도서 이용자의 작가(Author) 관련 선호요인

항목	정의
<i>Author</i>	도서 이용자의 작가에 대한 선호요인
general	작가에 대한 이용자의 정보 혹은 생각 전반
ability	작가의 능력(예, 작가의 역사, 통계에 대한 지식, 단어를 적절하게 사용하는 능력 등)
idea	작가의 생각, 사상
interpretation of the subject matter	작가의 주제, 소재에 대한 해석하는 방식
life	작가의 삶
perspective	작가의 견해, 관점
perspective-political correctness	작가의 말의 표현이나 용어의 사용에서, 인종·민족·언어·종교·성차별 등의 편견을 다루는 방식
good experience with the author	작가에 대한 좋은 경험(전작에 대한 호감 등)
reputation	작가가 지니는 명성

〈표 8〉 도서 이용자의 형식(Form) 관련 선호요인

항목	정의
Form	도서의 형식, 형태적 특성과 연관된 선호요인
audiobook	오디오북 형태의 도서
award	도서의 수상 여부
book cover	도서의 표지
editing	도서의 편집 상태
illustration	도서 내용에 포함된 그림, 삽화
publisher	도서를 출간한 출판사
title	도서의 제목
translation	도서의 번역과 관련한 내용

류나 오자, 탈자는 물론 레이아웃 등이 이용자들의 선호와 관련하여 언급되었다. 한편, 오디오북(audiobook)과 관련해서는 목소리 톤(예, '내레이터 톤으로서는 끔찍했다')과 연기력(예, '연기가 훌륭했다') 등도 선호에 영향을 미치는 것으로 나타났다.

4.2 샘플 그룹별 특성에 따른 선호요인

본 연구는 일반적인 도서 평점별로 리뷰를 수집한 그룹(일반 그룹)과 동일 도서에 대한 리뷰의 평점이 양분되는 리뷰를 수집한 그룹(도서 중심 그룹), 동일 이용자가 작성한 리뷰의 평점이 양분되는 리뷰를 수집한 그룹(이용자 중심 그룹)으로 샘플을 구분하여 평점별, 도서별, 이용자별 특성이 선호요인에 어떠한 양상으로 나타나는가를 살펴보았다(〈표 9〉 참고).

먼저 전체 리뷰 텍스트를 대상으로 평점 그룹에 따른 구분만 한 일반 그룹은 평점 그룹에 따른 전체적인 양상을 보여준다. 〈표 9〉에서 일반 그룹을 기준으로 전체 선호요인 범주 7가지 가운데 이용자 리뷰에서 긍정적인 측면의 비율이 부정적인 측면의 비율보다 우세했던 경우는 읽기

(Reading), 캐릭터(Character), 작가(Author)의 경우로 드러났으며, 특히 읽기(Reading)의 경우는 도서 혹은 이용자를 기준으로 호불호가 나뉘는 리뷰 샘플을 대상으로 한 두 그룹(도서 중심 그룹, 이용자 중심 그룹)에서도 긍정 비율이 우세하게 드러났다. 즉, 도서 이용자들은 소설, 혹은 희곡 등의 장르에서 나타나는 캐릭터(Character), 도서의 창작자인 작가(Author), 독서 과정에서 이루어지는 사고과정 및 감정 반응과 연관되는 읽기(Reading)에서 더 긍정적인 요인과 연결시키는 경향이 있다고 해석할 수 있다.

반면, 스토리(Story), 내용(Content), 글쓰기(Writing), 형식(Form) 범주와 관련된 선호요인은 전반적으로 부정적인 측면으로 작용한 경우가 긍정 비율보다 높게 나타났으며, 특히 긍정, 부정 비율의 편차가 가장 컸던 스토리(Story)를 비롯하여, 내용(Content), 형식(Form) 범주는 본 연구에서 구분하여 살펴본 일반 그룹, 도서 중심 그룹, 이용자 중심 그룹 모두에서 전반적으로 비선호요인으로 작용하는 비율이 우세하게 드러났다. 흥미로운 것은 작가나 독서 행위와 같이 개인적 취향이나 감정 반응과 밀접하

〈표 9〉 샘플 그룹별 특성에 따른 도서 이용자의 선호요인

선호요인	합계(비율)	일반 그룹		도서 중심 그룹		이용자 중심 그룹	
		낮은 평점	높은 평점	낮은 평점	높은 평점	낮은 평점	높은 평점
<i>Story</i>	344 (23.5%)	74 (28.9%)	59 (19.3%)	56 (25.1%)	68 (23.0%)	56 (28.9%)	31 (16.1%)
development	36 (2.5%)	12 (4.7%)	2 (0.7%)	5 (2.2%)	8 (2.7%)	7 (3.6%)	2 (1.0%)
general	33 (2.3%)	2 (0.8%)	9 (3.0%)	6 (2.7%)	5 (1.7%)	6 (3.1%)	5 (2.6%)
interest	32 (2.2%)	1 (0.4%)	5 (1.6%)	2 (0.9%)	21 (7.1%)	3 (1.5%)	-
ending	27 (1.8%)	8 (3.1%)	7 (2.3%)	1 (0.4%)	7 (2.4%)	2 (1.0%)	2 (1.0%)
narrative	25 (1.7%)	1 (0.4%)	1 (0.3%)	12 (5.4%)	10 (3.4%)	1 (0.5%)	-
setting	25 (1.7%)	7 (2.7%)	7 (2.3%)	1 (0.4%)	2 (0.7%)	3 (1.5%)	5 (2.6%)
pace	24 (1.6%)	7 (2.7%)	6 (2.0%)	2 (0.9%)	2 (0.7%)	5 (2.6%)	2 (1.0%)
coherent	23 (1.6%)	9 (3.5%)	3 (1.0%)	-	-	10 (5.2%)	1 (0.5%)
detail	20 (1.4%)	4 (1.6%)	1 (0.3%)	5 (2.2%)	2 (0.7%)	8 (4.1%)	-
empathy	18 (1.2%)	1 (0.4%)	2 (0.7%)	11 (4.9%)	4 (1.4%)	-	-
reality	16 (1.1%)	3 (1.2%)	6 (2.0%)	-	6 (2.0%)	-	1 (0.5%)
premise	15 (1.0%)	4 (1.6%)	4 (1.3%)	-	-	1 (0.5%)	6 (3.1%)
compelling plot	14 (1.0%)	4 (1.6%)	-	5 (2.2%)	-	4 (2.1%)	1 (0.5%)
theme	11 (0.8%)	3 (1.2%)	4 (1.3%)	-	-	-	4 (2.1%)
consistency	10 (0.7%)	5 (2.0%)	-	2 (0.9%)	-	3 (1.5%)	-
tension	7 (0.5%)	1 (0.4%)	-	4 (1.8%)	1 (0.3%)	-	1 (0.5%)
plot twist	4 (0.3%)	-	1 (0.3%)	-	-	3 (1.5%)	-
scene	4 (0.3%)	2 (0.8%)	1 (0.3%)	-	-	-	1 (0.5%)
<i>Reading</i>	315 (21.5%)	30 (11.7%)	77 (25.2%)	25 (11.2%)	100 (33.8%)	18 (9.3%)	65 (33.9%)
pleasure	58 (4.0%)	5 (2.0%)	19 (6.2%)	1 (0.4%)	22 (7.4%)	-	11 (5.7%)
recommendation	49 (3.3%)	11 (4.3%)	16 (5.2%)	1 (0.4%)	8 (2.7%)	3 (1.5%)	10 (5.2%)
related works	41 (2.8%)	3 (1.2%)	16 (5.2%)	-	8 (2.7%)	1 (0.5%)	13 (6.8%)
absorption	37 (2.5%)	6 (2.3%)	7 (2.3%)	4 (1.8%)	9 (3.0%)	5 (2.6%)	6 (3.1%)
previous experience	34 (2.3%)	-	5 (1.6%)	-	17 (5.7%)	7 (3.6%)	5 (2.6%)
inspiring	29 (2.0%)	-	7 (2.3%)	-	12 (4.1%)	-	10 (5.2%)
worth	28 (1.9%)	1 (0.4%)	1 (0.3%)	10 (4.5%)	9 (3.0%)	2 (1.0%)	5 (2.6%)
readability	21 (1.4%)	4 (1.6%)	1 (0.3%)	9 (4.0%)	7 (2.4%)	-	-
touching	11 (0.8%)	-	5 (1.6%)	-	3 (1.0%)	-	3 (1.6%)
memorable	3 (0.2%)	-	-	-	2 (0.7%)	-	1 (0.5%)
situation	3 (0.2%)	-	-	-	2 (0.7%)	-	1 (0.5%)
stimulating	1 (0.1%)	-	-	-	1 (0.3%)	-	-
<i>Character</i>	253 (17.3%)	41 (16.0%)	56 (18.4%)	46 (20.6%)	38 (12.8%)	43 (22.2%)	29 (15.1%)
general	72 (4.9%)	19 (7.4%)	9 (3.0%)	5 (2.2%)	15 (5.1%)	12 (6.2%)	12 (6.2%)
behavior	40 (2.7%)	2 (0.8%)	9 (3.0%)	21 (9.4%)	4 (1.4%)	3 (1.5%)	1 (0.5%)
personality	32 (2.2%)	5 (2.0%)	9 (3.0%)	9 (4.0%)	4 (1.4%)	3 (1.5%)	2 (1.0%)
empathy	31 (2.1%)	2 (0.8%)	8 (2.6%)	6 (2.7%)	4 (1.4%)	7 (3.6%)	4 (2.1%)
development	25 (1.7%)	5 (2.0%)	9 (3.0%)	1 (0.4%)	1 (0.3%)	7 (3.6%)	2 (1.0%)
relationship	19 (1.3%)	1 (0.4%)	7 (2.3%)	3 (1.3%)	5 (1.7%)	1 (0.5%)	2 (1.0%)
reality	16 (1.1%)	-	3 (1.0%)	1 (0.4%)	5 (1.7%)	5 (2.6%)	2 (1.0%)
dialogue	11 (0.8%)	3 (1.2%)	2 (0.7%)	-	-	3 (1.5%)	3 (1.6%)

선호요인	합계(비율)	일반 그룹		도서 중심 그룹		이용자 중심 그룹	
		낮은 평점	높은 평점	낮은 평점	높은 평점	낮은 평점	높은 평점
thought	4 (0.3%)	3 (1.2%)	-	-	-	-	1 (0.5%)
propensity	3 (0.2%)	1 (0.4%)	-	-	-	2 (1.0%)	-
<i>Content</i>	246 (16.8%)	56 (21.9%)	53 (17.4%)	30 (13.5%)	36 (12.2%)	53 (27.3%)	18 (9.4%)
material	57 (3.9%)	9 (3.5%)	18 (5.9%)	1 (0.4%)	12 (4.1%)	7 (3.6%)	10 (5.2%)
originality	43 (2.9%)	11 (4.3%)	7 (2.3%)	6 (2.7%)	3 (1.0%)	13 (6.7%)	3 (1.6%)
genre	25 (1.7%)	4 (1.6%)	11 (3.6%)	1 (0.4%)	8 (2.7%)	1 (0.5%)	-
length	24 (1.6%)	6 (2.3%)	2 (0.7%)	2 (0.9%)	-	14 (7.2%)	-
readers' interest	15 (1.0%)	3 (1.2%)	-	2 (0.9%)	1 (0.3%)	6 (3.1%)	3 (1.6%)
subject	14 (1.0%)	1 (0.4%)	5 (1.6%)	1 (0.4%)	5 (1.7%)	1 (0.5%)	1 (0.5%)
focus	10 (0.7%)	8 (3.1%)	1 (0.3%)	1 (0.4%)	-	-	-
accuracy	9 (0.6%)	5 (2.0%)	-	-	-	4 (2.1%)	-
information	8 (0.5%)	1 (0.4%)	3 (1.0%)	1 (0.4%)	3 (1.0%)	-	-
quality	8 (0.5%)	1 (0.4%)	-	5 (2.2%)	1 (0.3%)	1 (0.5%)	-
general	7 (0.5%)	1 (0.4%)	3 (1.0%)	3 (1.3%)	-	-	-
cohesive	6 (0.4%)	2 (0.8%)	1 (0.3%)	-	-	3 (1.5%)	-
fact	4 (0.3%)	-	-	2 (0.9%)	2 (0.7%)	-	-
structure	4 (0.3%)	-	1 (0.3%)	1 (0.4%)	-	1 (0.5%)	1 (0.5%)
convincing	3 (0.2%)	2 (0.8%)	-	-	-	1 (0.5%)	-
evidence	3 (0.2%)	-	-	2 (0.9%)	1 (0.3%)	-	-
detail	2 (0.1%)	1 (0.4%)	-	-	-	1 (0.5%)	-
reference	2 (0.1%)	1 (0.4%)	1 (0.3%)	-	-	-	-
depth	1 (0.1%)	-	-	1 (0.4%)	-	-	-
harmony	1 (0.1%)	-	-	1 (0.4%)	-	-	-
<i>Writing</i>	185 (12.6%)	33 (12.9%)	31 (10.2%)	39 (17.5%)	35 (11.8%)	18 (9.3%)	29 (15.1%)
description	39 (2.7%)	4 (1.6%)	2 (0.7%)	15 (6.7%)	9 (3.0%)	3 (1.5%)	6 (3.1%)
general	39 (2.7%)	7 (2.7%)	10 (3.3%)	1 (0.4%)	3 (1.0%)	7 (3.6%)	11 (5.7%)
style	28 (1.9%)	6 (2.3%)	6 (2.0%)	4 (1.8%)	10 (3.4%)	1 (0.5%)	1 (0.5%)
characterization	17 (1.2%)	5 (2.0%)	5 (1.6%)	-	-	4 (2.1%)	3 (1.6%)
use of terms	17 (1.2%)	3 (1.2%)	-	10 (4.5%)	1 (0.3%)	1 (0.5%)	2 (1.0%)
character formation	9 (0.6%)	-	-	2 (0.9%)	7 (2.4%)	-	-
narration	8 (0.5%)	3 (1.2%)	1 (0.3%)	-	-	1 (0.5%)	3 (1.6%)
target audience	8 (0.5%)	1 (0.4%)	-	3 (1.3%)	4 (1.4%)	-	-
POV(point of view)	6 (0.4%)	1 (0.4%)	4 (1.3%)	-	-	-	1 (0.5%)
tone	6 (0.4%)	1 (0.4%)	1 (0.3%)	1 (0.4%)	-	1 (0.5%)	2 (1.0%)
conveying context	4 (0.3%)	-	-	3 (1.3%)	1 (0.3%)	-	-
expression	2 (0.1%)	2 (0.8%)	-	-	-	-	-
language	2 (0.1%)	-	2 (0.7%)	-	-	-	-
<i>Author</i>	84 (5.7%)	6 (2.3%)	22 (7.2%)	20 (9.0%)	17 (5.7%)	2 (1.0%)	17 (8.9%)
perspective	28 (1.9%)	2 (0.8%)	5 (1.6%)	13 (5.8%)	3 (1.0%)	1 (0.5%)	4 (2.1%)
good experience with the author	24 (1.6%)	-	9 (3.0%)	-	8 (2.7%)	-	7 (3.6%)
ability	8 (0.5%)	2 (0.8%)	5 (1.6%)	1 (0.4%)	-	-	-

선호요인	합계(비율)	일반 그룹		도서 중심 그룹		이용자 중심 그룹	
		낮은 평점	높은 평점	낮은 평점	높은 평점	낮은 평점	높은 평점
idea	7 (0.5%)	2 (0.8%)	-	-	3 (1.0%)	-	2 (1.0%)
reputation	7 (0.5%)	-	3 (1.0%)	-	-	-	4 (2.1%)
perspective-political correctness	4 (0.3%)	-	-	4 (1.8%)	-	-	-
life	3 (0.2%)	-	-	-	3 (1.0%)	-	-
interpretation of the subject matter	2 (0.1%)	-	-	2 (0.9%)	-	-	-
general	1 (0.1%)	-	-	-	-	1 (0.5%)	-
<i>Form</i>	39 (2.7%)	16 (6.2%)	7 (2.3%)	7 (3.1%)	2 (0.7%)	4 (2.1%)	3 (1.6%)
editing	21 (1.4%)	8 (3.1%)	3 (1.0%)	7 (3.1%)	-	1 (0.5%)	2 (1.0%)
translation	5 (0.3%)	3 (1.2%)	-	-	-	1 (0.5%)	1 (0.5%)
award	3 (0.2%)	2 (0.8%)	-	-	-	1 (0.5%)	-
illustration	3 (0.2%)	-	3 (1.0%)	-	-	-	-
audiobook	2 (0.1%)	1 (0.4%)	1 (0.3%)	-	-	-	-
book cover	2 (0.1%)	1 (0.4%)	-	-	1 (0.3%)	-	-
title	2 (0.1%)	1 (0.4%)	-	-	-	1 (0.5%)	-
publisher	1 (0.1%)	-	-	-	1 (0.3%)	-	-
합계	1466 (100%)	256 (100%)	305 (100%)	223 (100%)	296 (100%)	194 (100%)	192 (100%)

게 관련된 선호요인에서는 긍정, 즉 선호요인으로써 작용하는 경향을 보였다면, 스토리, 내용, 글쓰기, 형식과 같이 도서 콘텐츠의 객관적인 품질과 관련된 선호요인에서는 부정, 즉 비선호요인으로 나타나는 경향이 나타났다.

다음으로 도서를 기준으로 호불호가 나뉘는 리뷰 샘플을 대상으로 한 도서 중심 그룹은 평가가 상이하게 나뉘는 도서에서 이용자들의 선호요인의 양상이 어떻게 나타나는가를 보여준다(〈표 9〉 참고). 도서 중심 그룹이 단순 평점으로 구성된 일반 그룹과 구분되는 특징은 캐릭터(Character)와 도서의 창작자인 작가(Author)와 관련된 선호요인이 차별되는 양상으로 나타났다. 호불호가 나뉘는 도서에 의해 통제되지 않은 일반 그룹에서는 캐릭터와 작가에 대한 선호요인이 긍정 리뷰에서 더 언급되며 우세하게 나타났던 반면, 논쟁적인 도서의 경우

캐릭터와 작가 요소가 부정 리뷰에서 더 언급되며 비선호요인으로 작용되는 경향을 나타내고 있었다. 특히 캐릭터의 행동(behavior)과 작가의 관점(perspective) 요인이 긍정과 부정간 가장 큰 편차를 보이며 도서를 중심으로 한 이용자 선호요인에서 특징적인 면모를 보였다.

마지막으로 이용자를 기준으로 호불호가 나뉘는 리뷰 샘플을 대상으로 한 이용자 중심 그룹은 평가가 상이하게 나뉘는 도서에서 이용자들의 선호요인의 양상이 어떻게 나타나는가를 보여준다(〈표 9〉 참고). 도서 중심 그룹이 단순 평점으로 구성된 일반 그룹과 구분되는 특징은 캐릭터(Character)와 글쓰기(Writing) 관련 선호요인에서 보였다. 도서를 기준으로 호불호가 나뉘는 리뷰 샘플에서는 캐릭터의 행동(behavior)이 가장 큰 호불호의 편차를 보였던 반면, 이용자 중심 그룹에서는 작가의

글쓰기(Writing)와도 관련되는 캐릭터 형성 과정(development), 캐릭터의 현실성(reality)이 일반 리뷰 그룹과 다르게 선호요소보다 비선호요소로써 두드러지게 나타나는 양상을 드러냈다. 특히, 이용자를 기준으로 도서에 대해 호불호가 나뉘는 그룹은 글쓰기(Writing) 관련 선호요인에서 나머지 두 그룹(일반 그룹과 도서 중심 그룹)과 비교하여 긍정 리뷰가 두드러지게 우세하며 차별되는 양상을 드러냈다. 이는 호불호가 분명한 이용자 그룹에서 주요하게 생각하는 요인이 글쓰기와 연관되며, 글쓰기 요소가 일반적인 그룹보다 더 중요하게 고려된다고 해석될 수 있다. 이와 같이 일반적인 평점뿐만 아니라, 도서 및 이용자를 중심으로 한 특정 그룹을 고려하여 선호요인을 파악한다면 보다 세부적인 추천 시스템을 설계하는 데 도움이 되리라 본다.

5. 결론 및 제언

본 연구는 기존의 추천 관련 연구가 이용자의 구체적인 선호요인을 반영하는 질적 연구가 부족하다는 인식하에, 이용자의 다양한 의견과 관점이 명시적으로 표현된 도서 리뷰 데이터를 대상으로 질적 내용분석을 통해 다양한 관점의 선호요인을 파악하였다. 특히, 기존 주제 기반 오피니언 마이닝 연구에서 사용되던 토픽모델링을 샘플 선정 과정에서 적용하여 다양한 토픽이 적용된 샘플 집합을 구성하였으며, 선호요인 가운데서도 다양한 속성별로 나타나는 양상을 파악하기 위해 평점 그룹별, 도서별, 이용자별 속성을 중심으로 하위 데이터 집합으로

구분하여 속성별 선호요인이 드러나는 양상을 파악하였다.

그 결과, 도서 이용자의 선호요인 관련 7개 범주가 식별되었으며, 이 가운데 도서 이해 및 내용적 측면과 관련되는 스토리(Story), 캐릭터(Character), 내용(Content), 도서 창작자와 관련되는 글쓰기(Writing), 작가(Author), 그 외, 도서와 관련된 외형적 측면과 관련되는 형식(Form) 관련 총 90개 개념이 선호요인으로 식별되었다. 한편, 도서 평점별로 리뷰를 수집한 그룹(일반 그룹)과 동일 도서에 대한 리뷰의 평점이 양분되는 리뷰를 수집한 그룹(도서 중심 그룹), 동일 이용자가 작성한 리뷰의 평점이 양분되는 리뷰를 수집한 그룹(이용자 중심 그룹)으로 속성을 나누어 분석한 결과, 각 그룹별 특징적인 이용자 선호요인 양상이 발견되었다. 일반 그룹에서는 특히 읽기(Reading), 캐릭터(Character), 작가(Author) 관련 선호요인에서 긍정 비율이 우세하게 드러났으며, 반면 스토리(Story), 내용(Content), 글쓰기(Writing), 형식(Form) 범주와 관련된 선호요인은 전반적으로 부정적인 측면으로 작용한 경우가 긍정 비율보다 높게 나타났다. 이는 작가와 같은 독자의 취향이 반영될 여지가 있거나 독자의 개인적 감정 반응인 독서 행위 관련 개인적 요소가 이용자들에게 긍정 요인으로 주로 발현된다면, 도서 콘텐츠의 객관적인 품질과 연관된 스토리, 내용, 글쓰기, 형식의 경우 부정 요인에서 주로 발현되는 것으로 파악될 수 있다. 다음으로 도서를 기준으로 호불호가 나뉘는 리뷰 샘플을 대상으로 한 도서 중심 그룹은 호불호가 나뉘는 도서에 의해 통제되지 않은 일반 그룹과 비교하여 캐릭터(Character)와 작가(Author) 관

런 선호요인이 부정 리뷰에서 더 많이 언급되는 경향을 드러냈다. 특히 캐릭터의 행동(behavior)과 작가의 관점(perspective) 요인이 긍정과 부정간 가장 큰 편차를 보이며 도서를 중심으로 한 이용자 선호요인에서 특징적인 면모를 드러냈다. 이용자를 기준으로 도서에 대해 호불호가 나뉘는 그룹은 글쓰기(Writing) 관련 선호요인에서 나머지 두 그룹(일반 그룹과 도서 중심 그룹)과 긍정 리뷰가 두드러지게 우세하며 차별되는 양상을 드러냈다. 이는 호불호가 분명한 이용자 그룹에서 주요하게 생각하는 요인이 글쓰기와 연관되며, 글쓰기 요소가 일반적인 그룹보다 더 중요하게 고려된다고 해석될 수 있다.

이와 같이 본 연구에서는 일반적인 평점뿐만 아니라, 도서 및 이용자를 중심으로 한 특정 그룹을 고려하여 선호요인의 세부적 내용을 파악하였다. 본 연구에서 도서 이용자 리뷰로부터 평점에 따른 일반 그룹뿐만 아니라 도서의 특성, 이용자의 특성을 반영하여 파악한 선호요인 및 선호요인별로 드러난 다양한 양상은 향후 도서관 독자자문 서비스의 도서추천 상황이나 도서추천 시스템 설계에서 보다 세분화된 이용자 서비스를 하는 데 유용한 자료로 사용될 수 있다. 예를 들면, 본 연구에서 파악된 도서 중심 그룹에서 밝혀진 선호요인 양상에 의하면, 이용자의 호불호가 극명하게 나뉘는 도서를 이용자에게 제공할 경우, 캐릭터와 작가, 특히 캐릭터의 행동과 작가의 관점을 최우선으로 고려하여 도서추천을 하거나, 도서추천 시스템의 패킷 적용이나 추천 알고리즘에서 가중치를 높게 주는 방안을 고려할 수 있다. 또한, 이용자가 캐릭터나 작가의 긍정 요소보다 부정 요소에 보다 민감하게

반응한다는 점도 세부적으로 반영할 수 있다. 마찬가지로 호불호 취향이 명확한 이용자 중심 그룹의 경우, 글쓰기 요소가 상대적으로 최우선시된다는 점을 고려하여, 도서추천을 하거나 추천 시스템 설계에 반영할 수 있을 것이다. 한편, 도서나 이용자 특유의 속성을 배제하고 보편적 특성을 가진 일반 그룹의 결과는 작가나 독서행위와 관련된 개인적 요소를 상대적으로 주요한 선호요인으로 고려하는 한편, 도서 콘텐츠의 객관적인 품질과 연관된 스토리, 내용, 글쓰기, 형식과 관련된 요인은 선호보다 비선호요인의 측면에서 이용에 영향을 미칠 수 있다는 내용을 반영하여 도서추천 상황에서 세분화된 서비스 제공을 도모할 수 있을 것이다.

한편, 본 연구에서 소셜 빅데이터를 대상으로 양적 연구뿐만 아니라 질적 연구를 통합 체계화한 방법론은 도서 추천 상황뿐만 아니라 다양한 형태의 빅데이터를 대상으로 적용될 수 있으리라 본다. 본 연구의 한계는 분석 대상 샘플수(N=620)로 인해 통계적 분석 결과를 일반화하기 어렵고 선호요인의 전반적인 패턴을 파악하는 정도에 그쳤다는 것이며, 향후 본 연구에서 식별된 선호요인을 기반으로 설문조사 기법 등을 통한 통계적 검증에 관한 후속 연구가 필요하리라 본다. 또한, 본 연구에서 사용한 Goordreads 리뷰 데이터는 전 세계 이용자에 의해 생성되고, WorldCat의 서지레코드에 연계되는 외부정보원으로 사용되며, 여러 도서추천 및 평가 연구에서 활용되어 온 데이터임에도 불구하고, 영미권 이용자 비중이 크며 본 연구에서 사용한 언어도 영어 리뷰만을 대상으로 한 만큼 국내 이용자 상황을 온전히 반영하기 어렵다는 한계를 지닌다. 그럼에도 불구하고

고 기존 많은 연구에서 검증된 리뷰 데이터를 대상으로 한 분석은 도서 이용자의 보편적인 선호 패턴을 파악하여 국내 도서 이용자를 위한 서비스 개선에 도움이 되리라 본다. 실례로 영어와 중국어 이용자 태그를 대상으로 언어적, 국가적 차이가 있는지 비교분석한 연구(Wu et

al., 2013)에 따르면, 이용자 태그의 양상이 언어, 국가에 관계없이 비슷하게 분포하는 내용의 연구결과도 나온 바 있다. 향후, 국내 이용자 리뷰를 영미권 이용자 리뷰와 결합하여 비교 검증하는 방식의 후속 연구 또한 필요하리라 본다.

참 고 문 헌

- 김라연, 박은경 (2011). 독서 감상문의 구성 요소 분석 연구. *교육과학연구*, 17, 17-30.
- 김성훈, 노윤주, 김미령 (2021). 연관성분석 기반 도서추천서비스의 이용자 만족에 관한 내러티브 연구. *한국도서관·정보학회지*, 52(3), 287-311. <https://doi.org/10.16981/kliss.52.3.202109.287>
- 박동진 (2010). 국어교육: 독서 감상문 쓰기의 실태: 구성 요소 분석을 중심으로. *새국어교육*, 84, 109-125.
- 변제연, 심원식 (2013). 도서관 이용자 코멘트 데이터 분석도구 리뷰 및 적용: LibQUAL+ 설문 데이터를 중심으로. *정보관리학회지*, 30(3), 157-181. <https://doi.org/10.3743/KOSIM.2013.30.3.157>
- 손은정, 박태연, 오희정 (2020). 이용자 로그데이터 기반 대학도서관일반열람실 활용도 및 선호요인 분석. *한국문헌정보학회지*, 54(2), 375-398. <https://doi.org/10.4275/KSLIS.2020.54.2.375>
- 심지영 (2022). 도서추천 시스템 개선을 위한 도서이용 맥락 요소 탐색. *정보관리학회지*, 39(2), 299-324. <https://doi.org/10.3743/KOSIM.2022.39.2.299>
- 유다빈, 유혜진, 김나라, 김윤희 (2015). 통합 리뷰 감정 분석을 통한 맞춤형 도서 비교 및 평가 시스템 설계. *한국정보처리학회 학술대회논문집*, 22(1), 108-111. <https://doi.org/10.3745/PKIPS.y2015m04a.108>
- 윤정옥 (2012). 차세대 도서관 목록의 이용자 서평에 관한 고찰. *한국도서관·정보학회지*, 46(2), 115-132. <https://doi.org/10.4275/KSLIS.2012.46.2.115>
- 임정수 (2013). 최신 영화 VOD 이용자의 선호도에 대한 컨조인트 분석. *한국콘텐츠학회논문지*, 13(5), 191-198. <https://doi.org/10.5392/JKCA.2013.13.05.191>
- 최지은, 유혜진, 유다빈, 김나라, 김윤희 (2016). 리뷰 감정 분석을 통한 전자상거래 상품 분석 및 평가 시스템 설계. *정보과학회 컴퓨팅의 실제 논문지*, 22(5), 209-217.
- Begay, W., Lee, D. R., Martin, J., & Ray, M. (2004). Quantifying qualitative data: using LibQUAL+ (TM) comments for library-wide planning activities at the university of arizona.

- Journal of Library Administration, 40(3-4), 111-119.
https://doi.org/10.1300/j111v40n03_09
- Bhat, N. A. & Ganai, S. A. (2018). Assessment of user preference to information resources in agricultural libraries in north India. *Annals of Library and Information Studies*, 65(2), 96-99. <https://doi.org/10.56042/alis.v65i2.17377>
- Chen, L., Chen, G., & Wang, F. (2015). Recommender systems based on user reviews: the state of the art. *User Modeling and User-Adapted Interaction*, 25(2), 99-154.
<https://doi.org/10.1007/s11257-015-9155-5>
- Goodreads (2022, September 5). Public Library Groups. Available:
https://www.goodreads.com/group/show_tag/public-library
- Green, P. E. & Srinivasan, V. (1990). Conjoint analysis in marketing research: new developments and directions. *Journal of Marketing*, 54(4), 3-19. <https://doi.org/10.2307/1251756>
- Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2017). Bag of tricks for efficient text classification. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, 2, 427-431. <https://doi.org/10.48550/arXiv.1607.01759>
- Kadiresan, N., Singson, M., & Thiyagarajan, S. (2021). Examining the relationship between academic book citations and Goodreads reader opinion and rating. *Annals of Library and Information Studies*, 67(4), 215-221. <https://doi.org/10.56042/alis.v67i4.32597>
- Korfiatis, N., Stamolampros, P., Kourouthanassis, P., & Sagiadinos, V. (2019). Measuring service quality from unstructured data: a topic modeling application on airline passengers' online reviews. *Expert Systems with Applications*, 116, 472-486.
<https://doi.org/10.1016/j.eswa.2018.09.037>
- Kousha, K., Thelwall, M., & Abdoli, M. (2017). Goodreads reviews to assess the wider impacts of books. *Journal of the Association for Information Science and Technology*, 68(8), 2004-2016.
<https://doi.org/10.1002/asi.23805>
- Lam, A. H. C., Ho, K. K. W., & Chiu, D. K. W. (2022). Instagram for student learning and library promotions: a quantitative study using the 5E Instructional Model. *Aslib Journal of Information Management*(ahead-of-print). <https://doi.org/10.1108/ajim-12-2021-0389>
- Liu, H., He, J., Wang, T., Song, W., & Du, X. (2013). Combining user preferences and user opinions for accurate recommendation. *Electronic Commerce Research and Applications*, 12(1), 14-23. <https://doi.org/10.1016/j.elerap.2012.05.002>.
- Mariana, S., Surjandari, I., Dhini, A., Rosyidah, A., & Prameswari, P. (2017). Association rule mining for building book recommendation system in online public access catalog. *Proceedings*

- of the 2017 3rd International Conference on Science in Information Technology, 246-250.
<https://doi.org/10.1109/icsitech.2017.8257119>
- McNee, S. M., Riedl, J., & Konstan, J. A. (2006, April 22-27). Being accurate is not enough: how accuracy metrics have hurt recommender systems. Paper presented at the CHI'06 extended abstracts on Human factors in computing systems, Montréal Québec Canada, 1097-1101. <https://doi.org/10.1145/1125451.1125659>
- Naik, Y. & Trott, B. (2012). Finding good reads on Goodreads: Readers take RA into their own hands. *Reference and User Services Quarterly*, 51(4), 319-323.
- Netzer, O., Toubia, O., Bradlow, E. T., Dahan, E., Evgeniou, T., Feinberg, F. M., Feit, E. M., Hui, S. K., Johnson, J., Liechty, J., Orlin, J. B., & Rao, V. R. (2008). Beyond Conjoint Analysis: Advances in Preference Measurement. *Marketing Letters*, 19(3/4), 337-354. <https://doi.org/10.1007/s11002-008-9046-1>
- Park, S. (2000). Usability, user preferences, effectiveness, and user behaviors when searching individual and integrated full-text databases: implications for digital libraries. *Journal of the American Society for Information Science*, 51(5), 456-468.
[https://doi.org/10.1002/\(SICI\)1097-4571\(2000\)51:5<456::AID-ASI6>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1097-4571(2000)51:5<456::AID-ASI6>3.0.CO;2-O)
- Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: a python natural language processing toolkit for many human languages. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 101-108. <https://doi.org/10.48550/arXiv.2003.07082>
- Reyes, B. M. & Devlin, F. A. (2021). An assessment of e-book collection development practices among romance language librarians. *Collection and Curation*, 40(1), 24-30.
<https://doi.org/10.1108/cc-12-2019-0047>
- Ryan, M. & Farrar, S. (2000). Using conjoint analysis to elicit preferences for health care. *Bmj*, 320(7248), 1530-1533. <https://doi.org/10.1136/bmj.320.7248.1530>
- Shah, A. M., Yan, X., Tariq, S., & Ali, M. (2021). What patients like or dislike in physicians: analyzing drivers of patient satisfaction and dissatisfaction using a digital topic modeling approach. *Information Processing & Management*, 58(3), 102516.
<https://doi.org/10.1016/j.ipm.2021.102516>
- Sohail, S. S., Siddiqui, J., & Ali, R. (2013, August 22-25). Book Recommendation System Using Opinion Mining Technique. Paper presented at the 2013 International Conference on Advances in Computing, Communications and Informatics, Mysore, India, 1609-1614.
<https://doi.org/10.1109/icacci.2013.6637421>

- Sun, L., Chen, J., Li, J., & Peng, Y. (2015). Joint topic-opinion model for implicit feature extracting. *Proceedings of 2015 10th International Conference on Intelligent Systems and Knowledge Engineering*, 208-213. <https://doi.org/10.1109/iske.2015.17>
- Thelwall, M. & Kousha, K. (2017). Goodreads: A social network site for book readers. *Journal of the Association for Information Science and Technology*, 68(4), 972-983. <https://doi.org/10.1002/asi.23733>
- Tu, Y. F., Chang, S. C., & Hwang, G. J. (2021). Analysing reader behaviours in self-service library stations using a bibliomining approach. *Electronic Library*, 39(1), 1-16. <https://doi.org/10.1108/el-01-2020-0004>
- Tubishat, M., Idris, N., & Abushariah, M. A. (2018). Implicit aspect extraction in sentiment analysis: review, taxonomy, opportunities, and open challenges. *Information Processing & Management*, 54(4), 545-563. <https://doi.org/10.1016/j.ipm.2018.03.008>
- Wan, M., Misra, R., Nakashole, N., & McAuley, J. (2019). Fine-grained spoiler detection from large-scale review corpora. *Proceedings of the 2019 57th Conference of the Association for Computational Linguistics*, 2605-2610. <https://doi.org/10.48550/arXiv.1905.13416>
- Wilson, K. M., Hooper, R., Simpson, J., & Slay, J. (2021). Comparing print and ebook usage to meet patron needs. *Collection Management*, 46(2), 91-106. <https://doi.org/10.1080/01462679.2020.1833802>
- Wu, D., He, D., Qiu, J., Lin, R., & Liu, Y. (2013). Comparing social tags with subject headings on annotating books: a study comparing the information science domain in English and Chinese. *Journal of Information Science*, 39(2), 169-187. <https://doi.org/10.1177/0165551512451808>
- Wu, F., Hu, Y. H., & Wang, P. R. (2017). Developing a novel recommender network-based ranking mechanism for library book acquisition. *Electronic Library*, 35(1), 50-68. <https://doi.org/10.1108/el-06-2015-0094>
- Xiao, S., Wei, C. P., & Dong, M. (2016). Crowd intelligence: analyzing online product reviews for preference measurement. *Information & Management*, 53(2), 169-182. <https://doi.org/10.1016/j.im.2015.09.010>

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

Byun, Jaeyeon & Shim, Wonsik (2013). A review and application of library user comments data

- analysis tool: focused on the LibQUAL+ survey comments. *Journal of the Korean Society for Information Management*, 30(3), 157-181. <https://doi.org/10.3743/KOSIM.2013.30.3.157>
- Choi, Jieun, Ryu, Hyejin, Yu, Dabeen, Kim, Nara, & Kim, Yoonhee (2016). System design for analysis and evaluation of e-commerce products using review sentiment word analysis. *Korean Institute of Information Scientists and Engineers Transactions on Computing Practices*, 22(5), 209-217.
- Kim, Ra Yeon & Park, Eun Gyung (2011). A study on the analysis of essential ingredient for Book Reports. *Education Research Studies*, 17, 17-30.
- Kim, Seonghun, Roh, Yoonju, & Kim, Mi Ryung (2021). A narrative study on user satisfaction of book recommendation service based on association analysis. *Journal of Korean Library and Information Science Society*, 52(3), 287-311. <https://doi.org/10.16981/kliss.52.3.202109.287>
- Park, Dong Jin (2010). Condition of writing book reports. *Korean Education*, 84, 109-125.
- Shim, Jiyoung (2020). Exploring the contextual elements of book use to improve book recommender systems. *Journal of the Korean Society for Information Management*, 39(2), 299-324. <https://doi.org/10.3743/KOSIM.2022.39.2.299>
- Son, Eun-Jeong, Park, Tae-yeon, & Oh, Hyo-Jung (2020). Analysis of utilization status and preference factors of reading room in university library based on user log data: focusing on the case of "J" university. *Journal of the Korean Society for Library and Information Science*, 54(2), 375-398. <https://doi.org/10.4275/KSLIS.2020.54.2.375>
- Yim, Jungsu (2013). The conjoint analysis of users' preference on the VODs of the newly-released movies. *The Journal of the Korea Contents Association*, 13(5), 191-198. <https://doi.org/10.5392/JKCA.2013.13.05.191>
- Yoon, Cheong Ok (2012). A study on the user-contributed reviews for the next generation library catalogs. *Journal of the Korean Society for Library and Information Science*, 46(2), 115-132. <https://doi.org/10.4275/KSLIS.2012.46.2.115>
- Yu, Da-bin, Ryu, Hye-jin, Kim, Na-ra, & Kim, Yoon-hee (2015). A design of a system for customized comparison and evaluation of books using integrated review emotion words analysis. *Proceedings of the Korea Information Processing Society Conference*, 22(1), 108-111. <https://doi.org/10.3745/PKIPS.y2015m04a.108>