

이용자 생성 도서정보 태그에 기반한 소설 검색의 패싯 유형 개발*

Developing Facets for Fiction Retrieval Based on User-generated Book Tags

심지영 (Jiyoung Shim)**

초 록

본 연구는 소설 검색 환경을 개선하기 위해, 도서태그로부터 소설 이용자가 소설 탐색 상황에서 요구하는 다양한 패싯 요소를 식별하고 체계화하는 것을 목적으로 한다. 소설의 기본 패싯 체계를 랭가나단의 PMEST 기본 패싯에 기반하여, 1) 소설 자료를 형성하는 주체, 2) 소설을 구성하는 내용적, 외형적 성질, 3) 독자가 책과 상호작용하는 행위, 4) 소설 및 독서활동과 관련된 공간 정보, 5) 소설 및 독서활동과 관련된 시간 정보로 정의하고, 소설 7,174건에 부여된 약 31만 건의 태그 중 핵심 태그 3,730건을 선별하여 내용분석하였다. 그 결과, 소설 패싯의 상위범주 25개를 중심으로 다양한 속성을 체계화하였다. 본 연구의 결과는 향후 도서관 OPAC이나 소설 DB에 패싯 내비게이션 형태로 적용될 수 있을 것으로 기대된다.

ABSTRACT

The purpose of this study is to identify and systematize various facet elements required by users in fiction search situations from book tags to improve the fiction search environment. Based on the Ranganathan's PMEST formula, the basic facet system of the fiction was defined as 1) the personality that forms the fiction material, 2) the content and external characteristics that compose the fiction, 3) the reader interaction with books, 4) spatial information related to fiction and reading activities, and 5) time information related to fiction and reading activities. Out of approximately 310,000 tags assigned to 7,174 fiction, 3,730 core tags were selected and content-analyzed. As a result, various attributes were systematized around the top 25 categories of the fiction facets. The results of this study can be applied to facet navigation of OPAC and fiction DB in the future.

키워드: 소설 태깅, 소설 검색, 라이브러리팅, 내용분석, 패싯
social tagging, fiction retrieval, LibraryThing, content analysis, facet

* 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임
(NRF-2017S1A5B5A07061692)

** 독립 연구자(jiyoung_shim@yonsei.ac.kr)

■ 논문접수일자: 2020년 5월 26일 ■ 최초심사일자: 2020년 6월 15일 ■ 게재확정일자: 2020년 6월 22일
■ 정보관리학회지, 37(2), 225-249, 2020. <http://dx.doi.org/10.3743/KOSIM.2020.37.2.225>

※ Copyright © 2020 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

웹 2.0 환경에서 이용자가 생성한 도서정보 태그(이하 도서태그)는 도서 자체에 대한 기술과 묘사는 물론, 전통적인 도서조직 체계에서 표현할 수 없었던 도서와 독자와의 상호작용 내용 표현이 가능한 것으로 여겨졌다. 또한, 도서태그의 주제 표현 기능도 여러 연구를 통해 주제명표목표를 보완하는 도구로 기능할 수 있음을 시사한 바 있다(Wu, He, Qiu, Lin, & Liu, 2013).

그럼에도 불구하고, 도서검색 환경 개선을 위해 도서태그를 연계하는 방안에 관한 논의는 여전히 제한적이다. 지금까지 이루어진 연구는 이용자 생성 태그가 갖는 본질적인 한계인 의미의 불확실성과 비일관성 등 의미표현의 한계를 해결하는 데 집중되어왔다. 태그 내에서의 관계를 설정해 주어 태그의 의미를 명확하게 해주는 방법과 통제어휘 체계를 사용하여 태그의 의미적 제어를 하는 방법 등이 제안된 바 있다. 전자의 경우, 도서태그가 생성되는 맥락을 반영하여, 태그 간 의미 관계의 유형을 파악할 수 있다는 데서 의미가 있지만, 태그 용어의 비일관성 문제를 해결하기 어렵다는 한계를 갖고 있다. 후자의 경우, 용어 사이의 구조가 명확한 통제어휘 체계를 연계하여 태그 용어 간의 의미 관계를 제어해 줄 수 있으나, 태그 집합의 성격에 부합되는 시소러스나 사전 구축이 선행되어야 하며, 통제어휘 체계를 벗어나는 자료 고유의 특성이나 이용자와 관련된 맥락 정보 표현은 어렵다는 한계가 여전히 남아있다.

본 연구는 도서검색 환경에서 이용자 요구 맥

락을 반영한 지적 접근점을 제공하기 위해, 도서에 대한 이용자의 다양한 요구가 표현된 도서태그를 기반으로 소셜 고유의 ‘패싯(facet)’ 유형을 개발하는 것을 목표로 한다. 특히, 도서태그 관련 연구의 주요 대상으로 여겨져 왔으며, 공공도서관의 이용요구에서도 높은 비중을 차지하고 있는 소셜 장르로 한정하여, 소셜검색에 특화된 패싯 유형을 개발하고자 한다. 웹 정보 검색 환경에서 지적 접근점을 제공하는 데 유용한 개념으로 주목받아 온 랭가나단의 PMEST 기본 패싯을 기반으로, 내용분석을 통해 소셜 고유의 패싯 속성을 추출하고 의미 관계를 체계화하고자 한다. 한편, 대량의 도서태그 컬렉션으로부터 용어의 일관성을 유지하고 각 태그의 의미를 보다 효율적으로 분류하기 위해, 특정 속성(예, 인물, 지리적 위치, 시간, 조직, 장소 등)에 대응되는 태그를 자동으로 감지하고 분류할 수 있는 개체명 인식(Named-Entity Recognition, 이하 NER) 기법 적용을 시도한다. 궁극적으로 소셜 독자들이 자주 사용하는 전형적인 요구 유형을 식별하고, 검색에 주로 관여하는 요소를 파악하여 체계적으로 제공하는 것을 목적으로 한다.

2. 이론적 배경

2.1 소셜 색인

이용자의 소셜 탐색 환경을 개선하기 위한 가장 근본적인 노력은 소셜 색인이다. 소셜 색인은 정보조직과 검색 연구에서 꾸준히 논의되어 온 주제로(Saarti, 2019), 전통적인 정보조직

체계에서 소설 콘텐츠 자체가 지닌 요소(예, 주제, 플롯, 등장인물, 시간적 배경 등)에 초점을 맞추었던 것과 더불어 다양한 이용자 연구를 통해 소설 이용자의 다양한 해석과 관계된 요소(예, 저자의 의도, 내용을 읽고 이해하기 쉬운 정도(readability), 내용에 대한 감상 등)를 파악하고 체계화하는 것으로 발전하여왔다. 소설 콘텐츠의 요소를 파악하고 체계화하는 방식은 기존 분류체계를 기반으로 소설 콘텐츠 고유의 요소를 구성하거나 소설 자료에 특화된 분류체계를 만드는 형태로 이루어져 왔다(Pejtersen, 1997). 또한, 독자와의 상호작용의 영향을 많이 받는 소설 자료의 특성을 반영하기 위해, 소설 탐색 상황에서 이용자 요구를 분석하고 식별하기 위한 연구들이 주로 이루어져 왔다.

특히, 소설 검색 상황에서의 이용자 요구 분석은 전문가들에 의해 연역적으로 구성된 분류체계가 실제 이용 상황의 요구를 완전히 반영하지 못한다는 문제 인식에서 비롯되었다. 소설 색인에서 분류체계의 한계를 보완하기 위해, 독자들이 자주 사용하는 전형적인 요구 유형을 식별하고, 검색에 주로 관여하는 요소를 파악하여 이용자 중심적인 접근점을 제공하기 위한 노력이 꾸준히 이어져 왔다. 소설 이용자 요구 분석과 관련한 초기 연구는 검색 상황에 직면한 소설 자료의 이용자들이 자신들의 요구를 어떻게 형성하고 있는지를 분석하는 형태로 나타났다(Morbhead, Pejtersen, & Rouse, 1984; Pejtersen & Austin, 1983, 1984). 이러한 연구들은 소설 이용자의 검색 질의를 기반으로 정보요구를 분석하였으며, 이용자가 OPAC 등 정보검색 시스템에 입력한 내용(Pejtersen & Austin, 1983, 1984; Solomon, 1997; Yu &

O'Brien, 1997)이나 사서가 이용자의 정보요구를 파악하기 위한 면담 내용(Pejtersen, 1980) 등을 분석대상으로 한다. 대표적 예로, 여러 차례의 이용자 연구로부터 도출된 Pejtersen(1980)의 소설 분류체계인 'AMP 분류체계'는 네 개의 상위 범주(주제, 구조, 저자의 의도, 접근성) 아래 기존 분류체계의 요소(예, 플롯, 시공간적 배경 등)뿐만 아니라, 이용자의 반응과 해석이 결부된 요소들(예, 감정적 경험, 내용을 읽고 이해하기 쉬운 정도 등)을 추출하고, 체계화하여 이용자 중심적인 소설 접근점을 제공하는 데 기여했다.

소설 색인에서 이용자 연구는 웹 2.0 환경의 다양한 도서정보 서비스 플랫폼에서 이용자가 색인 작성에 참여한 태그 데이터가 활발히 생성됨에 따라 다양한 형태로 이루어져 왔다. 이용자 태그는 이용자의 자연스러운 행동이 반영된 데이터로 도서관 OPAC, 사회적 목록 서비스, 온라인 서점 등을 통해 이용자가 소설 이용과 관련하여 제출한 정보가 분석대상이 된다. 다수의 도서태그 분석 연구들이 이용자와 소설 자료의 상호작용 과정에 초점을 맞춰 소설 독자가 도서에 대해 표현하는 다양한 요소들이 존재하고 있음을 파악하였다. 예를 들어, Lawson(2009)의 연구는 소설 이용자가 소설 제목의 내용을 설명하는 단어나 구뿐만 아니라, 이용자의 관점, 개인적인 표식 등을 도서태그에 표현하고 있음을 파악하였다. 또한, Desrochers, Laplante, Martin, Quan-Haase, Spiteri(2016)의 연구에서도 '독서 경험', '독자에게 어필하는 요소', '독자 반응' 등과 같이 이용자가 책과 상호작용하는 태그 유형들을 식별하였다. 한편, 소설 정보를 제공하는 다양한 접근점이 실제 정보

서비스에서 어떻게 제공되는가에 관해 비교 분석한 연구도 등장한다. Adkins와 Bossaller (2007)는 온라인 서점, 독자자문 데이터베이스(readers' advisory databases), 공공도서관 OPAC 내에서 소설 검색 이용자에게 제공되는 요소 26개(예, 플롯의 전개, 실제 사건, 등장인물의 직업, 등장인물의 관계, 대상 독자, 독자 리뷰, 전문가 리뷰 등)를 식별하고, 각 서비스 플랫폼에 따라 해당 요소가 얼마나 충실하게 제공되는지를 양적 내용분석을 통해 파악한 바 있다.

본 연구에서는 기존의 소설 색인연구에서 주목해온 소설 콘텐츠와 이용자의 요구에 기반한 소설 고유의 요소 식별에 주목한다. 특히, 웹 2.0 환경 이후 등장한 도서태그 데이터 분석 연구에서 한 단계 더 나아가, 단순한 요소 식별이 아닌 요소 간의 관계를 체계화하는 데 초점을 두고자 한다.

2.2 도서태그 분석

소설 색인과 관련하여, 최근 10년간 이용자가 소설 탐색 환경에서 생성한 도서태그의 색인어로서의 가치에 주목하며, 도서태그를 대상으로 한 다양한 연구들이 이루어져 왔다. 도서태그 분석 연구는 이용자가 생성한 도서태그가 전통적인 색인 기능을 적절히 수행할 수 있는가와 기존의 색인과 어떤 차별화된 속성을 지니고 있는가를 파악하는 데 초점을 맞춰왔다. 색인어로서의 도서태그의 가치를 살펴본 실증 연구들은 도서태그가 도서 자체에 대한 상세한 기술은 물론, 도서와 상호작용하는 이용자의 개인적, 사회적 맥락을 잘 표현해 주고 있다는

사실을 보여준다. 예를 들어, 사회적 목록 서비스인 LibraryThing 이용자의 도서태그를 분석한 Bartley(2009)의 연구에 따르면, 도서태그에 사실 정보 유형(예, '문화비평', '대통령'), 의견 유형(예, '끔찍한', '권장하지 않음'), 개인 메모 유형(예, '보유하지 않은 책', '앞으로 읽을 책')과 같이 도서에 대한 주제 이외의 다양한 관점이 포함되는 것으로 드러난다. Lawson(2009)은 도서태그의 속성을 도서에 대한 주제 정보를 표현하는 객관적 태그와 책과 상호작용하는 이용자 고유의 특성이 반영된 주관적 태그로 구분하여 각각의 비율과 특징을 분석하였다. 그녀의 연구는 이용자들이 독서와 관련된 개인적 경험과 관련하여 특정 도서에 대해 부여한 주관적 태그의 다양한 속성 유형을 식별하여, 태그 고유의 강점을 드러내고 있다. 김동숙과 정연경(2010)은 도서태그의 속성을 파악하는데 있어, 소설의 고유요소인 소설의 3요소(주제, 구성, 문체)와 소설구성의 3요소(인물, 사건, 배경)를 패킷으로 설정하고, 해당 속성에 대응되는 태그를 파악하여 기존 색인어를 보완하고자 하였다.

한편, 도서태그가 특정 도서의 주제 표현을 얼마나 잘 하고 있는지에 초점을 두고, 도서태그의 색인어로서의 성능을 파악하기 위한 연구들도 도서태그 분석 연구의 주요 대상으로 다루어져 왔다. 이들 연구는 다양한 유형의 도서를 대상으로 미국 의회도서관 주제명표목표(Library of Congress Subject Headings List, 이하 LCSH) 등의 통제어와 도서태그를 비교 분석하는 형태를 띤다. 예를 들어, Smith(2007)는 다양한 장르(예, 판타지 소설, 현대 소설, 그래픽 노블, 유머집 등)의 도서 샘플을 대상으로

LibraryThing 이용자들이 해당 도서에 부여한 도서태그와 LCSH의 통제어휘를 비교하여, 도서태그의 주제 표현 가능성을 시사한 바 있다. Smith(2007)의 연구가 소수의 도서 샘플을 대상으로 내용분석에 집중하였다면, 대량의 데이터를 대상으로 양적 분석하여, 도서태그의 주제 표현 정도를 통계적으로 파악한 연구들도 이루어져 왔다. Lu, Park, & Hu(2010)는 약 8,500권의 도서에 부여된 LibraryThing 도서태그와 LCSH의 통제어휘를 수집하고, 두 어휘 간의 중복율 산정을 통해 도서태그가 LCSH의 절반 이상(50.1%)의 주제를 포함하고 있음을 파악하였다. 또한, Wu 외(2013)는 다수 이용자가 부여한 고빈도 도서태그의 경우, 언어적 차이(예, 영어와 중국어)와 무관하게 주제명표목표의 통제어휘와 비슷한 패턴으로 대응된다는 사실을 밝혀냈다. 이와 같이, 도서태그의 색인어로서의 가치와 주제 표현 기능에 대한 실증적 검증이 이루어진 바 있다.

도서태그 분석과 관련한 연구들이 다양한 형태로 이루어졌음에도 불구하고, 태그를 이용하여 이용자의 검색 상황 개선을 도모한 연구는 드물며, 제한적으로 이루어져 왔다. 태그 내에서의 관계를 설정해 주어 태그의 의미를 명확하게 해주는 방법(예, 김동숙, 정연경, 2010)과 통제어휘 체계를 사용하여 태그의 의미적 제어를 하는 방법(예, 이승민, 2011) 등이 제안된 바 있다. 본 연구는 자료 고유의 특성과 이용자와 관련된 맥락 정보 표현에 유용한 전자의 방식을 기본으로 하되, 대량의 도서태그 컬렉션으로부터 용어의 일관성을 유지하고 각 태그의 의미를 보다 효율적으로 분류하기 위해, 특정 속성(예, 인물, 지리적 위치, 시간, 조직, 장소

등)에 대응되는 태그를 자동으로 감지하고 분류할 수 있는 NER 기법을 적용하여 후자의 방식도 절충하는 형태로 진행하고자 한다.

2.3 패킷분석

패킷분석은 웹 정보 검색 환경에서 지적 접근점을 제공하는 데 유용한 것으로 주목받아 오며(Broughton, 2006), 도서는 물론, 이미지 자료, 방송자료, 금융상품 등 다양한 유형의 자료를 대상으로 광범위하게 적용되어왔다. ‘패킷(facet)’은 어떤 대상에 대한 관점이나 특성, 속성으로 표현되고 접근할 수 있는 개념으로, 패킷 분석을 통해 대상 자료는 하나 이상의 관점으로 표현되고 접근될 수 있다. 예를 들면, ‘레몬을 곁들인 구운 붉은 감자’라는 음식 메뉴는 패킷 분석을 통해 조리법, 재료와 같이 메뉴명을 들었을 때 바로 알 수 있는 속성 이외에도, 채식주의자와 아침 메뉴와 같은 식이요법, 식사코스과 같은 속성으로도 접근할 수 있어, 이용자가 검색 상황에서 다면적인 관점과 풍부한 정보를 제공할 수 있다.

특히, Ranganathan(1967)의 PMEST 기본 패킷은 패킷 분석 연구에서 널리 사용되어 온 구조로, 지식 영역에서 반복되는 규칙성을 가진 속성에 기반하여, PMEST 패킷을 구성하고 있다. PMEST 패킷은 해당 주제를 형성하는 대상이 되는 개체(Personality, 이하 P) 패킷을 중심으로, 개체가 지닌 고유한 성질이나 처리방법 혹은 재료를 의미하는 재료((Matter, 이하 M) 패킷, 행위주체가 어떠한 힘을 발산하거나 어떠한 방향으로 작용하는 것과 관련이 있는 에너지(Energy, 이하 E) 패킷, 지리적 속성

을 표현하는 공간(Space, 이하 S) 패킷, 시간적 속성을 표현하는 시간(Time, 이하 T) 패킷으로 구성된다. 이와 같은 PMEST 패킷은 단순하지만, 보도 기사 작성의 기본원칙인 육하원칙의 다섯 가지 주요소와도 대응되며(예, 누가(P), 언제(T), 어디서(S), 무엇을(M), 어떻게(E)), 대상의 속성을 압축적으로 표현하는 데 유용한 구조로 널리 활용되어 왔다(Satija, 2017).

본 연구에서는 랑가나단의 PMEST 패킷을 인물, 사건, 배경을 중심으로 서사구조를 지닌 소설의 속성에 잘 부합한다고 보고, 소설 검색의 패킷 유형을 개발하는 데 기본 구조로 삼았다(〈표 1〉참고). 개체(P) 패킷을 소설 자료를 형성하는 주체로 해석하여, 소설 타이틀 요소로 간주하였으며, 재료(M) 패킷의 경우, 소설을 구성하는 내용적 성질(예, 저자, 주제, 언어, 장르, 등장인물 등)과 매체 형식 등의 외형적 성질로 해석하였다. 에너지(E) 패킷은 독자가 소설과 상호작용하는 행위와 관련된 패킷으로 독자 반응, 독자 리뷰, 대상연령 등의 요소로 해

석하였고, 공간(S) 패킷은 소설의 공간적 배경과 독서활동과 관련된 공간 정보 속성을 내포하는 패킷으로, 시간(T) 패킷은 소설의 시간적 배경, 독서활동과 관련된 시간 정보 속성을 의미하는 패킷으로 규정하였다. 한편, 소설의 패킷 요소를 추출하는 데 있어, 온라인 서점과 공공도서관 OPAC, 독자자문 데이터베이스로부터 소설 접근점을 식별한 Adkins와 Bossaller(2007)의 연구도 참고하였다.

3. 연구방법

소설 색인 작성 과정에서 이용자 요구를 반영하고자 하는 여러 노력에도 불구하고, 소설 검색 상황의 이용자 행태 분석과 도서태그를 통한 이용자 요구 분석 연구들은 몇 가지 한계를 지닌다. 전자의 경우 실험 샘플의 규모가 제한적이며, 색인자의 주관적 해석이 개입된다는 문제를 내포하고 있다(Saarti, 1999). 후자의 경우, 태그가

〈표 1〉 소설의 기본 패킷 체계 정의

PMEST 패킷	Ranganathan(1967)의 정의	본 연구의 소설 패킷 정의	Adkins & Bossaller(2007)의 소설 접근점 예시
P (Personality)	주제를 형상화하는 본질적 요소로서, 주제를 형성하는 주체	소설 자료를 형성하는 주체	제목
M (Material)	개체가 지닌 고유한 성질이나 처리방법 혹은 재료	소설을 구성하는 내용적, 외형적 성질과 관련된 내용	저자, 주제, 언어, 장르, 매체 형식, 판, 등장인물-소설 속 캐릭터/직업/관계, 출판사, 사실 정보
E (Energy)	행위 주체가 어떠한 힘을 발산하거나 특정 방향으로 작용하는 것과 관련된 (예, 처리 과정이나 행위, 반응 등)	독자가 소설과 상호작용하는 행위와 관련된 내용	독자 반응, 독자 리뷰, 전문가 리뷰, 대상연령
S (Space)	지리적 속성	소설의 공간적 배경, 독서활동과 관련된 공간 정보	공간적 배경
T (Time)	연대나 시대적 특성	소설의 시간적 배경, 독서활동과 관련된 시간 정보	출판일, 시대적 배경

갖는 의미의 불확실성과 비일관성 문제라는 한계를 지닌다(Golder & Huberman, 2006).

본 연구는 소수 이용자 샘플로 인한 한계를 해결하고, 색인자의 주관을 가능한 배제하기 위해, 다수의 이용자 관점이 자연스럽게 반영된 대량의 도서태그 데이터를 사용한다. 또한, 태그 용어가 지닌 모호성을 해결하고, 이용자의 이용 맥락을 추출하기 위해 자료의 속성을 표현하는 데 효과적이며(Broughton, 2006), 주제 접근성을 높이는 데 유용한 것으로 여겨져 온(Chung, 2014) 패킷 분석 방법을 적용하여 이용자 태그의 의미를 식별하고 구조화한다. 한편, 대량의 태그 데이터의 의미를 효율적으로 식별하고 일관적으로 분류하기 위해, NER 기법을 사용하여 자동으로 초별 분류한다(Nadeau & Sekine, 2007). 또한, 기계학습만으로 감별되지 않는 의미구조를 파악하기 위해 내용분석을 수행하고, 내용분석 결과는 코더간 신뢰도 검증을 통해 검증한다.

3.1 데이터 수집

본 연구는 소설검색에서의 이용자들의 요구를 도서태그 데이터를 통해 식별하기 위해, 소설 관련 도서태그 수집을 기본으로 한다. 다수의 독자를 지닌 소설을 선별하는 한편, 도서 태그 활동이 활발한 영어권 사회적 목록 플랫폼을 대상으로 하기 위해, 소설 저작의 표집틀(sampling frame)을 영미권 고전과 현대 소설의 정선된 목록인 Fiction Core Collection(Wilson, 2016)을 대상으로 하였다. Fiction Core Collection(18판, 2016)에 포함된 소설 목록 및 서지정보(예, 제목, 저자, 출판사, 키워드, ISBN)를 EBSCOhost

를 통해 2017년 3월 수집하였으며, 총 7,277건의 소설을 샘플로 선정하였다. 각 소설의 서지정보 가운데 제목과 ISBN 정보는 도서태그 데이터 수집 과정에서 샘플 도서를 식별하기 위함이며, 저자와 출판사 정보는 데이터 분석 과정에서 초기 분류에 필요한 전거 사전을 구성하기 위한 용도로 수집되었다. 또한, 키워드는 본 연구에서 식별된 소설 이용자 태그와 비교 분석하기 위함이다.

도서태그 데이터 수집은 도서태그 관련 선행 연구에서 널리 사용되어 온 사회적 목록 서비스인 LibraryThing으로부터 하였다. 7,277건의 소설 데이터가 LibraryThing 내에 존재하는지를 식별하기 위해, EBSCOhost에서 수집한 각 소설의 서지정보를 이용하여 검색을 수행하였다. 검색에서 먼저 사용한 질의는 도서의 고유식별기호인 ISBN이며, ISBN 정보가 없는 저작들(1,449건, 19.9%)의 경우, 소설의 제목과 저자를 AND 연산자로 조합하여 해당 저작을 식별하였다. 그 결과 샘플 소설의 98.8%를 차지하는 총 7,190건의 도서를 LibraryThing으로부터 식별하였다(2018년 3월 기준). LibraryThing에서 식별된 샘플 소설의 도서태그 데이터는 각 저작의 고유 ID(<그림 1>의 URL을 구성하는 숫자 286,489에 해당)를 포함한 URL로 접근하여 수집하였다(<그림 1> 참고). 이 가운데, 태그가 전혀 할당되지 않은 저작들(16건)은 제외하였으며, 최종적으로 7,174개 소설로부터 총 316,844개 단어(구) 유형의 도서태그 데이터를 수집하였다. 도서태그 수집 과정은 Python 프로그래밍 언어를 통해 자동화된 방법으로 수행하였다.

A Close Run Thing
by Allan Mallinson

Series: Matthew Hervey stories (1)

Members	Reviews	Popularity	Average rating	Mentions
236	6	77,205	☆☆☆ (3.21)	7

As the war against Bonaparte rages to its bloody end upon the field of Waterloo, a young officer goes about his duty in the ranks of Wellington's army. He is Cornet Matthew Hervey of the Sixth Light Dragoons - a soldier, gentleman and man of honour who suddenly finds himself allotted a hero's role.

Recently added by TonySteele, Brodysatva, Drumclogmom, ralphruff, rcalf, Stephen_McGarry, Itchyfeetreader, barten

Tags: 1815, 19th century, adventure, audio, Battle of Waterloo, Britain, British, British Army, Cavalry, fiction, Hervey, historical, historical fiction, historical novel, history, Matthew Hervey, Matthew Hervey Series, military, military fiction, military history, Napoleonic, Napoleonic Wars, novel, paperback, read to-read, war, Waterloo, Wellington

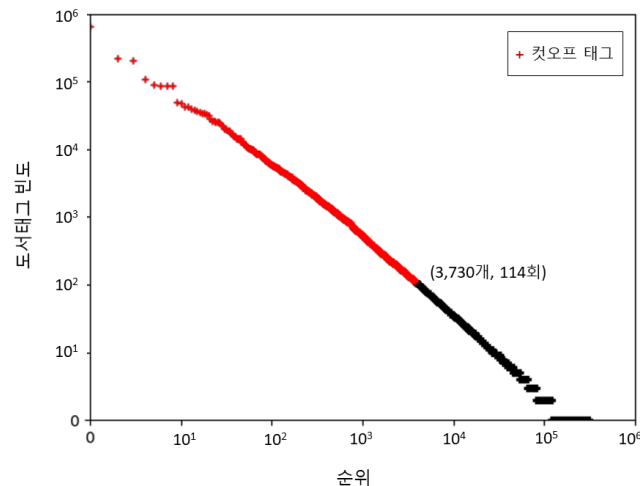
〈그림 1〉 LibraryThing의 Work ID 286489에 해당하는 소설에 관한 정보 및 도서태그 데이터를 제공하는 웹 페이지 화면

3.2 핵심 도서태그 선정

본 연구에서는 7,174개의 소설에 부여된 전체 31만여 건에 달하는 자연어 상태의 도서태그 데이터를 의미 단위로 결합하고, 방대한 태그 집합 내에서 상대적으로 많은 소설 이용자들에게 부여받은 핵심 도서태그를 선정하는 작업을 수행하여 작업의 효율성을 꾀하였다. 먼저 자연어 상태의 도서태그 데이터들을 의미 단위로 결합하기 위해, 자연어 처리에서 많이 사용하는 두 가지 자연어 처리 기법인 대소문자 통일(case folding)과 원형복원(lemmatization)을 사용하였다. 전자는 대문자와 소문자로 분리된 도서 태그들(예, 'fiction', 'Fiction')을 하나의 표제어로 결합하는 것이고, 후자는 동일 개념이지만, 다양한 활용형을 가진 도서태그를 하나의 표제어로 묶어주는 방식이다. 예를 들면,

'read in 2010(7,387회)' 태그와 'Reading in 2010(1회)' 태그는 동일 개념으로 묶여 하나의 표제어로 카운트되었으며, 이 가운데 표제어는 가장 많은 이용자의 선택을 받은 것을 기준으로 빈도수가 가장 높은 태그(예, 'read in 2010')를 표제어로 선정하는 방식을 적용하였다. 대소문자 통일은 Python 프로그래밍 언어로 일괄 변환하였고, 원형복원은 Stanford CoreNLP toolkit (Manning, Surdeanu, Bauer, Finkel, Bethard, & McClosky, 2014)을 사용하여 처리하였다. 그 결과, 총 309,313개의 고유 태그를 식별하였다.

다음으로 유사 개념들을 통제된 고유 태그들을 대상으로, 장서 빈도의 80%를 커버하는 상위 순위의 도서태그들을 핵심 도서태그로 선정하였다. 본 연구에서 식별된 고유 태그 309,313개는 〈그림 2〉와 같이 빈도와 순위에 로그 스케일을 적용했을 때 회귀 계수가 -1에 근접한



〈그림 2〉 고유 태그의 분포 및 컷오프를 통한 핵심 도서태그 선정

-1.1로 Zipf 법칙의 분포를 따르며, 전체 빈도가 소수의 고빈도 태그에 집중되는 것으로 드러났다. 〈그림 2〉에서 핵심 고유 태그 309,313개의 장서 빈도는 6,175,174회로, 장서 빈도의 80%(4,942,024회)로 컷오프한 핵심 도서태그 수는 붉은 선과 검은 선의 마지막 경계점인 114회의 출현빈도를 가지는 태그까지 총 3,730개로, 이는 전체 고유 태그의 1.2%에 해당한다.

3.3 NER을 통한 초벌 분류

데이터 분석은 핵심 도서태그 데이터로 선정된 3,730개의 도서태그를 대상으로 NER 기법을 사용하여 도서태그들의 개체명 유형을 자동으로 식별하는 초벌 분류를 먼저 수행하였다. 이는 대량의 도서태그 컬렉션으로부터 특정 속성에 대응되는 태그를 자동으로 감지하여 보다 효율적으로 분류하기 위함이다. NER을 통한 초벌 분류 과정에서는 두 가지 방식을 결합하였다. 그 하나는 Stanford CoreNLP toolkit

(version 3.9.1)에 포함된 Stanford Named Entity Recognizer(NER)를 사용하여 Stanford NER에서 정의된 일반적인 개체명 유형(예, 인물, 날짜, 국적 등)을 인식하는 방식이고, 다른 하나는 Fiction Core Collection에서 제공하는 서지정보를 활용하여 소설과 연관된 개체명 유형(제목, 저자, 출판사)을 인식하는 방식이다.

Stanford NER은 영어 소설, 어린이 소설 등을 대상으로 고유명사인 인물명을 자동으로 추출하기 위해 사용되어왔으며(Elson & McKeown, 2010; Iosif & Mishra, 2014), 개체명을 인식하는 성능도 상대적으로 좋은 편으로 보고된 바 있다. 위키피디아의 전기 문서를 대상으로 Stanford NER을 비롯하여, Illinois NET, Open Calais NER WS, Alias-i LingPipe의 네 가지 NER의 성능을 평가한 연구에서, Stanford NER의 성능은 카테고리별 정확률이 최소 0.83, 최대 0.93, 재현율은 최소 0.91, 최대 0.97로 가장 높은 것으로 드러났다(Atdağ & Labatut, 2013). 본 연구에서는 Stanford NER에서 카테고리의

정의 자체가 포괄적이고 모호한 기타(MISC)를 제외하고, 분류 성능을 평가한 결과, 마이크로 평균 정확률은 0.88, 마이크로 평균 재현율은 0.79, 마이크로 평균 F1 척도는 0.83에 해당하는 분류 성능을 보여주었다. 단, 숫자, 지리, 날짜, URL 등 범위가 명확한 범주의 개체명은 높은 성능을 보인 반면, 통화의 경우, #, \$ 기호가 들어간 경우를 통화로 인식하여, 매우 낮은 성능(평균 F1 척도 0.33)을 보였다.

핵심 도서태그 데이터 3,730개의 도서태그 중 Stanford NER을 통해 개체명 유형이 식별된 태그 수는 1,496개로 전체의 40.1%에 해당하며, 총 20개 유형의 개체명이 인식되었다(〈표 2〉 참고). 단, 하나의 태그에 두 개 이상의 개체명 유형이 인식되는 경우(예, British(국적), author(직책))가 포함되어, 개체명 유형별로 할당된 총 태그 수는 1,609개로 나타났다. 각 도서태그는 Stanford NER에서 정의한 개체명

〈표 2〉 Stanford NER을 통해 식별된 도서태그의 개체명 유형 및 분포(N=3,730)

개체명	태그 예시*	태그 수(개)	비율(%)
인물(PERSON)	<i>Stephen King, Sherlock Holmes</i>	457	12.3
날짜(DATE)	read in <i>2014, 18th century, 21st century</i> fiction	222	6.0
국적(NATIONALITY)	<i>Irish, British</i> author, <i>German</i> literature	167	4.5
숫자(NUMBER)	book <i>1</i> , Box <i>3</i> , Shelf <i>1</i> , Average Rating <i>4.0, 813</i>	118	3.2
국가(COUNTRY)	<i>Hungary, Kenya, Philippines</i>	109	2.9
직책(TITLE)	<i>Rabbi</i> Small, British <i>author, Inspector</i> Lynley	90	2.4
주(STATE_OR_PROVINCE)	<i>New York, California, Florida, Maine</i>	75	2.0
기타(MISC)**	<i>Arthurian</i> , American <i>West</i> , African-American <i>Literature</i>	70	1.9
도시(CITY)	<i>Venice, Oxford, Seattle, Washington DC</i>	68	1.8
사인(CAUSE_OF_DEATH)	school <i>shooting, AIDS</i> , American Civil <i>War, poison</i>	38	1.0
기관(ORGANIZATION)	<i>Amazon Vine, Heritage Press, Limited Editions Club, Harvard</i>	37	1.0
지역(LOCATION)	<i>Africa, Europe, Latin</i> America, <i>South</i> America, <i>Middle East</i>	33	0.9
형사 고발(CRIMINAL_CHARGE)	<i>murder, espionage, serial killer, terrorism, kidnapping</i>	29	0.8
서수(ORDINAL)	<i>First</i> Edition, <i>first</i> in series, <i>first</i> printing, <i>first</i> novel, <i>2nd</i> Copy, <i>20th</i> c author, <i>21st</i> C. American Literature	27	0.7
종교(RELIGION)	<i>Puritans, Christian</i> Living	25	0.7
배경(SET)	<i>1810s, 1930s, 2000s</i>	14	0.4
이데올로기(IDEOLOGY)	<i>totalitarianism, capitalism, socialism</i>	13	0.3
기간(DURATION)	<i>Hundred Years</i> War, modern <i>day, might-read-one-day, Millennium</i>	11	0.3
통화(MONEY)**	<i>#3, \$z-new_tags_done</i>	3	0.1
URL(URL)	Purchased from <i>Audible.com, amazon.com</i>	3	0.1

* 태그 예시 항목의 도서태그에서 이탤릭, 볼드체로 표시된 단어(구)에 의해 해당 개체명으로 인식됨

** 도서태그의 NER 결과에서 의미 있는 패턴을 찾기 어려운 개체명으로 드러남

유형에 자동으로 할당됨으로써 대량의 도서태그로부터 초별 분류의 기능을 수행하였다. 예를 들면, 가장 많은 비중을 보인 인물(PERSON) 개체명에 할당된 도서태그 그룹은 소설의 작가 이름, 캐릭터 이름과 관련한 태그들로 드러났다. 국적(NATIONALITY)의 경우 작가의 국적과 작품의 국적을 나타내는 태그들이 분류되었으며, 작중 캐릭터의 직책을 파악하는 단서를 제공하는 태그들은 직책(TITLE)으로 분류되었다.

수와 관련된 개체명으로 날짜(DATE), 숫자(NUMBER), 서수(ORDINAL)가 있었는데, 날짜(DATE)로 인식된 태그들의 경우, 독서일, 시대적 배경, 문학의 시기 구분 등이 나타났으며, 숫자(NUMBER)의 경우, 도서/DVD의 시리즈, 서가의 위치, 도서 평점, DDC 분류 번호(예, 미국소설을 의미하는 813) 등이, 서수(ORDINAL)의 경우, 도서의 판과 책, 시리즈, 복본, 작가의 활동 시기, 작품의 창작 시기 등과 관련된 태그들이 분류되었다. 기간(DURATION)의 경우, 도서의 시대적 상황과 관련한 기간이나 독서 기간과 관련된 태그 등이 식별되었다. 기관(ORGANIZATION)은 주로 도서 유통과 관련한 출판사, 서점 등의 도서 서비스 기관과 소설의 공간적 배경과 관련된 기관(예, 하버드 대학 등)이 나타났다. 한 가지 유형으로 대응되어 나타났던 개체명 유형으로는 국가(COUNTRY), 주(STATE_OF_PROVINCE), 도시(CITY), 지역(LOCATION), 소설의 지리적 배경과 시대적 배경을 나타냈던 배경(SET), 소설에 담긴 사상과 관련된 이데올로기(IDEOLOGY), 소설 캐릭터의 종교와 관련된 종교(RELIGION), 도서 입수 경로와 대응되는 URL(URL)이었다.

그 외, 기타(MISC)와 통화(MONEY) NER 결과로부터는 도서태그 내 의미 있는 유형을 파악할 수 없었다.

한편, 본 연구는 소설의 서지정보로부터 소설의 제목, 저자, 출판사 사전을 직접 구성하고, 각 개체 유형에 대한 매칭 규칙을 마련하여 소설과 연관된 개체명을 인식하였다. 소설의 서지정보로부터 소설과 연관된 개체명(제목, 저자, 출판사)과 관련된 사전을 직접 구성하여 개체명을 인식한 결과, 저자로 인식된 태그 수는 88개, 제목으로 인식된 태그 수는 64개, 출판사로 인식된 태그 수는 19개로 각각 나타났다. 사전을 통한 NER 결과를 Stanford NER 결과와 통합한 결과, 최종적으로 개체명으로 인식된 태그의 개수는 1,541개로, 핵심 도서태그(N=3,730) 가운데 NER을 통해 자동으로 초별 분류된 비율은 41.3%로 드러났다.

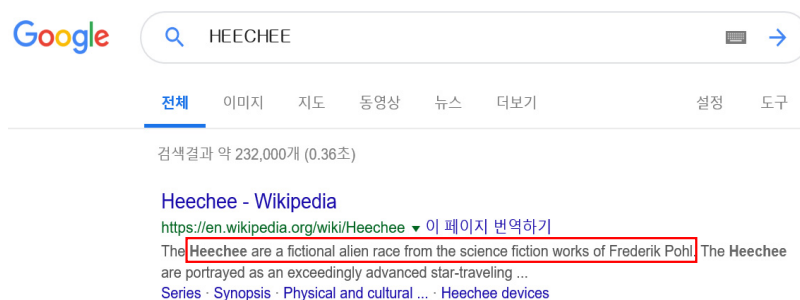
3.4 내용분석을 통한 패킷 유형 개발

본 연구의 분석 대상인 소설에 대한 핵심 도서태그 3,730개의 초별 분류 결과(〈표 2〉 참고)를 바탕으로, 내용분석을 수행하였다. 내용분석 과정은 코딩 체계(coding scheme)를 개발하고, 코더간 신뢰도 검증(inter-coder reliability)을 수행하고, 전체 데이터를 대상으로 최종 확립된 코딩 체계를 적용하는 방식으로 이루어졌다(Flick, 2019). 코딩 체계 개발은 Ranganathan(1967)의 PMEST 기본 패킷을 구조로 하는 한편, Adkins와 Bossaller(2007)의 연구에서 온라인 서점과 공공도서관 OPAC, 독자자문 데이터베이스로부터 식별된 소설의 접근점 내용을 참고하였다(〈표 1〉 참고).

핵심 도서태그 데이터(N=3,370)에 대한 코딩은 총 3회에 걸쳐 이루어졌으며, 두 개 이상의 개념이 존재하는 경우 중복 코딩을 허용하였다. 대량의 데이터를 대상으로 일관성을 확보하고, 코딩 후 결과 데이터를 손쉽게 처리하기 위해 질적 내용분석 소프트웨어인 ATLAS.ti 7.5.4를 이용하였다. 한편, 내용분석 과정에서 태그 명으로 정보를 유추하기 힘든 경우, 구글 검색을 통해 위키피디아 정보 등 부가적인 정보를 참고하였다(〈그림 3〉과 〈그림 4〉 참고).

코더간 신뢰도 검증은 내용분석의 일관성과 타당성을 확보하기 위해, 전체 도서태그 데이터(N=3,730)의 10%에 해당하는 370개 태그

를 계통 표집(systematic sampling)하여, 제2 코더에 의한 분석을 수행하였다. 두 명의 코더에 의한 분석은 홀스티 공식(Holsti, 1969, p. 137)을 사용하여 코더간 일치도가 80% 이상 도달할 때까지 반복적으로 재수행하며, 코딩 체계를 수립하였다. 최종적으로 두 명의 코더간 일치도는 89.3%였으며, 최종적으로 확립된 코딩 체계를 전체 데이터에 적용하였다. 한편, 소설 검색의 패킷 유형에 대한 결과는 자료 조직 관련 연구 경험이 있는 박사급 연구자 1인으로부터 동료 검증(peer debriefing)을 통한 전체적 리뷰를 거쳤다.



〈그림 3〉 구글 검색을 통한 캐릭터 정보 수집 결과 화면



〈그림 4〉 위키피디아를 통한 장르 정보 수집 결과 화면

4. 연구결과

4.1 도서태그에 기반한 소설 패킷 체계의 구조

본 연구에서는 이용자가 생성한 도서태그를 기반으로 하여 소설 이용자의 요구를 분석하고, 소설 검색 환경에서 활용할 수 있는 의미 있는 패킷 유형을 파악하였다. Fiction Core Collection에 수록된 소설 7,174건에 대해 이용자가 부여한 도서태그를 LibraryThing으로부터 수집하고, 수집된 도서태그 약 31만 개 중 핵심 도서태그 3,730개를 대상으로 NER 기법과 내용분석을 통해 소설 검색의 패킷 체계를 도출하였다. 패킷 구조는 랑가난단의 PMEST 기본 패킷을 토대로, 본 연구에서 정의한 기본 패킷을 토대로 하였다(〈표 1〉 참고). 미분류된 도서태그 188개(식별력이 없는 fiction과 의미 불분명 태그 187개)를 제외하고, 핵심 도서태그의 95%로부터 소설 패킷의 상위범주 25개가 식별되었다(〈표 3〉 참고).

소설 이용자가 가장 많이 언급했던 패킷 유형은 소설의 내용적, 외형적 성질과 관련된 재료(M) 패킷으로, 상위범주의 절반 이상(13개 범주)이 식별되었다. 흥미로운 것은 소설 이용자들이 전통적인 도서관 분류체계에서 제공되는 주제(11.4%), 저자(6.0%), 매체 형식(3.8%), 시리즈 정보(3.8%), 언어(2.2%), 형태 사항(1.7%), 판 사항(1.3%), 출판사(0.8%) 정보보다 소설의 장르(14.6%)나 등장인물(14.6%) 정보를 통한 소설 접근을 선호한다는 것이다. 비록 적은 비중이지만, 플롯/줄거리(1.8%)와 문학 기법(0.5%)을 소설의 접근점으로 고려하는 이용

자도 눈에 띄었다.

다음으로 많은 비중을 차지하고 있었던 패킷 유형은 독자와 책의 상호작용 행위와 관련된 에너지(E) 패킷으로, 독서활동(20.2%) 및 독서활동에 유용한 정보(4.1%)를 제공하는 속성을 드러냈다. 특히, 독서활동 과정에서 수반되는 독서활동 요소(예, 도서 입수 경로, 독서 상태 등)는 전통적인 도서관 분류체계에서 거의 다루어지지 않던 요소이지만, 의미 있는 비중을 차지하고 있었다. 소설 및 독서활동의 공간 정보 요소인 장소(S) 패킷으로는 소설의 공간적 배경(14.0%), 도서 입수 장소(3.2%), 독서 장소(1.3%), 출판 장소(0.1%)가 식별되었으며, 소설 작품과 독서활동과 관련된 시간 정보 요소인 시간(T) 패킷은 소설의 시간적 배경(4.5%), 독서 시기(3.6%), 문예 사조(1.4%), 출판일(1.0%), 도서 입수 시기(0.6%) 순으로 드러났다.

한편, 소설 자료를 형성하는 주제와 관련된 개체(P) 패킷 유형의 소설 제목은 상대적으로 가장 적은 비중(3.6%)을 차지했는데, 이는 사전을 통한 NER 과정에서 제목을 식별할 때, 3개 이하의 단어로 구성된 태그의 경우 미리 구성된 사전과 완전히 일치하는 경우에만 타이틀로 판정하는 등 비교적 엄격한 규칙을 적용한 결과로 보인다. 또한, 이용자 태그의 경우 태그 길이가 긴 것을 선호하지 않는 데서 비롯된 결과로 추정할 수 있다. 그러나, 이용자가 책의 타이틀을 인식할 때 타이틀의 일부보다 주로 전체로 인식하고, 제목의 경우 '알고 있는 자료 탐색(known-item search)' 요구에서 주로 발생하게 된다는 점을 고려할 때, 소설 검색의 이용자 요구 상황에서 제목이 상대적으로 이용도가 낮다고 해석될 수도 있다고 본다.

〈표 3〉 도서태그로부터 식별된 소설 검색의 패킷 체계의 상위범주 및 분포(N=3,730)*

PMEST 패킷	소설 패킷 상위범주	태그 수(개)	비율(%)
P (개체)	제목	133	3.6
M (재료)	장르	546	14.6
	등장인물	544	14.6
	주제	425	11.4
	저자	223	6.0
	사실 정보(예, 역사적 인물, 사건 등)	157	4.2
	매체 형식	142	3.8
	시리즈	142	3.8
	언어	82	2.2
	플롯/줄거리	68	1.8
	형태 사항(도서의 물리적 외형)	64	1.7
	판 사항	49	1.3
	출판사	29	0.8
	문학적 장치/기법	18	0.5
	소계	2,489	66.7
E (에너지)	독서활동	754	20.2
	독서활동에 유용한 정보(예, 추천도서목록, 대상 독자 등)	154	4.1
	소계	908	24.3
S (장소)	지리적 배경	524	14.0
	도서 입수 장소	120	3.2
	독서 장소	48	1.3
	출판 장소	3	0.1
	소계	695	18.6
T (시간)	시간적 배경	167	4.5
	독서 시기	135	3.6
	문예 사조	51	1.4
	출판일	39	1.0
	도서 입수 시기	22	0.6
	소계	414	11.1
미분류		188	5.0

* Fiction Core Collection에 수록된 소설 7,174건에 대해 LibraryThing으로부터 수집된 이용자 태그 약 31만 개 중 핵심 도서태그 3,730개를 대상으로 중복 코딩을 허용하여 내용분석한 결과

4.2 이용자로부터 도출된 소설 패킷 요소의 내용

이용자 생성 도서태그로부터 도출된 소설의 패킷 요소는 랭가나단의 PMEST 기본 패킷에서 소설을 형성하는 주체로 소설 자체를 대표

하는 제목인 개체(P) 패킷을 제외하고, 〈표 4〉에서부터 〈표 7〉에서 보이는 바와 같이 다양한 양상으로 드러나고 있었다.

먼저, 가장 다양한 요소가 식별된 소설의 재료(M) 패킷은 소설의 내용이 어떻게 이루어지고, 물리적인 외형은 어떻게 구성되었는지와

관련된 속성으로, 도서태그는 이용자가 소설에 내용적, 외형적으로 어떤 측면에 초점을 맞추는지와 관련해서 다양한 관점을 보여주고 있다(〈표 4〉 참고). 특히, 소설 자료와 관련된 외형적 성질(예, 매체 형식, 시리즈, 언어, 형태 사항, 판 사항, 출판사)보다 내용적 측면에 대한 다양한 접근점이 소설 이용자로부터 식별되었으며, 소설 이용자들은 장르와 등장인물과 관련된 속성을 매우 세부적으로 구분하여 탐색하고 있는 것으로 드러났다. 예를 들면, 장르 패킷의 하위 속성으로 주제, 형식, 지역, 시기, 이용자 연령으로 구분하여 탐색하고 있었으며, 장르의 주제인 판타지 소설 내에서도, 'steampunk(역사적 배경에 공상과학이나 판타지 요소를 적용하고 전자제품 대신 증기로 작동하는 기계가 등장하는 문학 장르)'나 'sword and sorcery(검 마 소설, 판타지 장르의 하위 장르)'와 같이 매우 세부적인 구분으로 접근하는 것으로 드러났다. 또한, 등장인물 패킷에서도 인간과 인간 이외의 개체로 구분하고 있으며, 인간 캐릭터의 경우 '미스 마플', '아서 왕'과 같이 이름은 물론, '경찰', '탐정'과 같은 직업, '나바호족'과 같은 인종으로도 구분하여 접근하고 있었다. 그 외, 등장인물의 성별, 역할, 관계, 성격, 연령, 종교 등 매우 세부적인 속성으로 구분하고 있었다.

한편, 전통적인 정보조직 체계에서 소설의 내용적 접근점으로 가장 일반적으로 사용하는 주제의 경우, 특정한 가치나 학문 분야와 관련된 속성도 나타났지만, 인간의 행동과 일상생활과 관련된 속성이 좀 더 두드러지게 나타났다. 인간의 행동과 관련된 속성은 크게 사회적인 측면(예, 중매결혼, 입양, 장기기증 등)과 반사회적인 측면(예, 방화, 유괴 등)으로 구분될

수 있으며, 일상생활과 관련된 속성은 장소, 음식, 미디어, 교통 등 우리 삶을 둘러싼 소재와 관련된다. 흥미로운 것은 이용자들이 소설 속에 등장하는 사실 정보에 관련된 접근점도 상당수 사용하고 있었다는 것이다. 예를 들면 역사적 인물(예, 레오나르도 다 빈치, 올리버 크롬웰, 에이브러햄 링컨 등), 사상(예, 식민지주의, 페미니즘 등), 역사적 사건(예, 제2차 세계대전, 스페인 내전, 마그나 카르타 서명 등)을 들 수 있다. 그 외, 소설 저자의 경우 저자명으로 주로 접근하고 있었으며, 저자의 국적도 주요한 접근점으로 사용하고 있는 것으로 드러났다.

다음으로, 독자와 소설과의 상호작용 행위에 초점이 맞춰진 에너지 패킷(E)은 독서활동과 독서활동에 유용한 정보로 드러났다(〈표 5〉 참고). 독서활동 패킷의 경우, 독자가 해당 도서를 어떻게 얻을 수 있는지와 관련된 속성(도서 입수 경로)이 가장 큰 비중을 차지하고 있었다. 예를 들면, 해당 책을 아마존 서점에서 구매했거나(purchased from Amazon), 도서관으로부터 대출했다는 것(borrowed from library)이 이러한 속성에 해당하는 내용이다. 또한, 해당 소설이 다양한 포맷으로 존재하는 경우(예, 영화 또는 TV로 각색된 경우 등)에 대한 정보를 소설 이용자들이 주목하고 있는 것으로 드러났으며, 독자 개인의 독서 상태(예, Already read, to-read, Not Started 등)나 언제 읽었는지(예, read in 2002)와 같은 맥락 정보도 독서활동 속성과 관련하여 비중 있게 나타났다. 그 외, 소설을 읽으면서 느꼈던 감상(독자 반응), 읽고 난 후의 평가(독자 리뷰), 독서를 원활하게 하기 위한 다양한 활동 관련 속성(예, 개인 독서 리스트, 독서 프로그램, 독서토론 동호회)

〈표 4〉 소설의 재료(M) 패킷의 상하위 범주 및 내용(n=2,489)

소설의 재료(M) 패킷			도서태그 예시	태그 수(개)
장르	주제		chick lit, steampunk, sword and sorcery	230
	형식		short story, epistolary novel, epic	122
	지역		UK fiction, Russian literature	118
	시기		literature-2000s, modern fiction, gothic	58
	연령		young adult literature, children's books	18
등장인물	인간	이름	Miss marple, King Arthur, Alex Delaware	163
		직업	police, detective, guru, professor	105
		인종	Navajo Indians, Chinese, Aborigines	74
		성별	female protagonist, male protagonist	39
		역할	guardian, prisoners of war, survivors	37
		관계	father-son relationship, friends, sisters	35
		성격	strong woman, geek, sharks	28
		연령	children, teenagers, teenage girls	12
		종교	Christian, Puritans, Catholic, Muslims	9
	인간 이외	개체	animals, monsters, vampires, spirits	37
		이름	Bilbo Baggins, Moby Dick, Gollum	5
주제	인간의 행동	사회적 행동	organ donation, arranged marriage, adoption	73
		반사회적 행동	arson, kidnapping, crime, caper	69
	일상생활	장소	hospital, restaurants, salon, hotel	38
		음식	chocolate, recipes, wine, spice, rice	8
		미디어	newspaper, letters, music, jazz, maps	8
		교통	cars, ships, trains, boats, spaceships	8
	가치		civil rights, freedom, humanity, religion	60
	재난		earthquake, fire, war, disease, cancer	59
	학문 분야		art, psychology, artificial intelligence	46
	공상		time travel, space travel, supernatural	25
	자연		snow, wilderness, lake, desert, ocean	18
	감정		guilty pleasure, jealousy, hope	13
저자	인명		J.K. Rowling, Shakespeare, Mark Twain	205
	국적		US author, UK author, Japanese Author	18
사실 정보	역사적 인물		da Vinci, Cromwell, Abraham Lincoln	47
	사상		colonialism, feminism	45
	역사적 사건		WWII, Spanish Civil War, the signing of the Magna Carta	32
	사회제도		slavery, monarchy, caste system	23
	기관		Church of England, IRA	10
매체 형식			PDF, DVD, mp3, audiobook, e-book	142
시리즈			Harry Potter Series, book1, 2nd in series	142
언어			English fiction, in Finnish	82
플롯/줄거리			love triangle, coming of age, missing children	68
형태 사항			hard copy, Leather Bound, slipcase	64
판 사항			first edition, UK edition, First published	49
출판사			Bantam, HaperCollins, Penguin books	29
문학적 장치/기법			third person, irony, picaresque	18

〈표 5〉 소설의 에너지(E) 패킷의 상하위 범주 및 내용(n=908)

소설의 에너지(E) 패킷		도서태그 예시	태그 수(개)
독서활동	도서 입수 경로	library book, borrowed from library, purchased from Amazon	182
	다양한 포맷 이용	saw the movie, on-kindle, movie tie-in, TV Series	144
	독서 상태	read in 2002, Already read, to-read, Not Started	139
	독자 반응	awesome, scary, funny, boring, sad, creepy	85
	개인 독서 리스트	in_my_library, bookmarks, keep	79
	독자 리뷰	5-stars, must read, my-favorites, disappointing	69
	독서 프로그램	1010 Challenge, storytelling, SantaThing 2014	38
	독서토론 동호회	Oprah's Book Club, book club, Book Club Read	18
독서활동에 유용한 정보	추천도서리스트	Easton Press 100 Greatest Books, Need to Read	58
	문학상	Hugo Award, Booker Prize, Pulitzer Prize	42
	대상 독자	adult fiction, juvenile fiction, children's literature	37
	독서 수준	grade 5, Age Group 9-12, 3-read, BR	11
	전문가 리뷰	New York Times Review of Books, guardian	6

도 소설과 관련하여 이용자들이 의미 있게 고려하는 것으로 드러났다. 한편, 추천도서 리스트, 문학상, 대상 독자, 독서 수준, 전문가 리뷰와 같은 독서활동에 유용한 다양한 정보와 관련된 속성도 드러났다. 단, 여기서 ‘대상 독자’나 ‘독서 수준’은 도서 자체가 지니는 성질로 해석하여 재료(M) 패킷으로 간주할 수도 있으나, 본 연구에서는 해당 속성이 도서 자체에 초점이 있는 것이 아니라, 이용자와 연관되어 해석되었다는 점에 방점을 두고, 에너지(E) 패킷으로 해석하여 분석하였다.

소설의 장소(S) 패킷은 소설의 공간적 배경, 독서활동과 관련된 공간 정보와 관련되며, 특히, 소설의 지리적 배경과 관련된 하위 속성이 다양하게 드러나고 있었다(〈표 6〉 참고). 지리적 배경은 주로 국가 단위로 구분하고 있었지만, ‘발칸 반도’, ‘카리브해’, ‘애팔래치아 산맥’ 등과 같은 자연적인 경계 혹은 ‘87지구’ 등과 같은 행정 구역으로 지역 구분의 경우도 상대적으로 높은 비중을 차지했다. 또한, ‘카멜롯’과 같은 가상의 장소도 소설과 관련된 접근점으로

사용하고 있었다. 한편, 소설의 내용과 별개로, 독자가 책을 어디에서 입수했는지, 어디에서 읽었는지와 같이 독서활동과 관련된 장소 속성 두 가지(도서 입수 장소, 독서 장소)가 식별되었다. 도서 입수 장소와 관련한 하위 속성으로는 책을 구매하거나 빌린 장소와 서가 상의 위치가 드러났다. 그 외, 소설의 외형적 성질과 관련해서 출판 장소와 관련된 속성도 발견되었다.

소설 작품과 독서활동과 관련된 시간 정보 요소인 시간(T) 패킷은 소설의 시간적 배경 속성이 가장 두드러지게 드러나며, 일반적인 시기 구분(예, 11세기, 1830년대, 선사시대 등)과 역사적 시기 구분(예, 빅토리아시대, 에드워드 7세 시대의, 나폴레옹 시대 등) 속성으로 구분되었다(〈표 7〉 참고). 또한, 소설 이용자들은 독서활동과 관련하여 독서 시기(예, 1990년대에 읽음, 2012년에 읽은 책 등)와 책 입수 시기(예, 2007년에 입수함, 2010년에 구매 등)를 의미 있게 구분하고 있었다. 한편, 소설의 내용적 성질과 관련해서 문예 사조(예, 고딕, 포스트모더니즘, 초현실주의 등)와 출판일(예, 2000년대

〈표 6〉 소설의 장소(S) 패킷의 상하위 범주 및 내용(n=695)

소설의 장소(S) 패킷		도서태그 예시	태그 수(개)
공간적 배경	국가	Australia, Belgium, England, France	200
	지역	Balkans, Caribbean, Appalachia, 87th Precinct	90
	도시	Berkeley, Berlin, Dublin, New York City, Vienna	77
	주	Alabama, Kansas, Kentucky, Michigan	60
	시설	bookstores, Louvre, court, loft	40
	단체(institution)	Scotland Yard, Opus Dei, museum, FBI	31
	가상의 장소	Ankh-Morpork, Barsetshire, Pern, Camelot	19
	섬	Channel Islands, Martha's Vineyard, Guernsey	7
도서 입수 장소	구입 혹은 대출 장소	Purchased from Audible.com, borrowed from library	68
	서가 위치	800 LITERATURE: FICTION, location: bookcase.right.6	52
독서 장소		read in Pittsburgh, read in high school, @home	48
출판 장소		US 1st Edition, First American Edition, US 1st ed	3

〈표 7〉 소설의 시간(T) 패킷의 상하위 범주 및 내용(n=414)

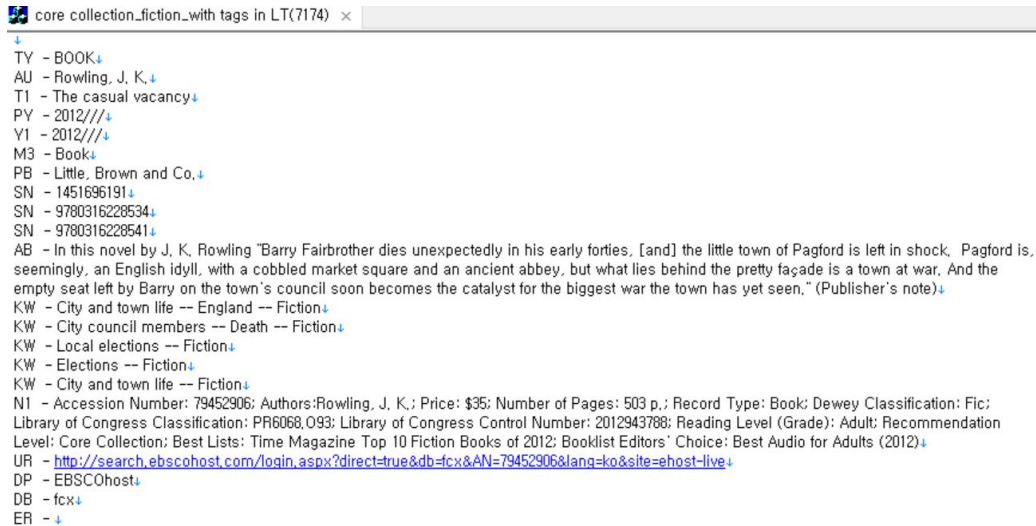
소설의 시간(T) 패킷		도서태그 예시	태그 수(개)
시간적 배경	일반적 시기구분	11th century, 1830s, prehistory	124
	역사적 시기구분	Victorian Era, Napoleonic Era, Edwardian	43
독서 시기		read in 2002, read 1990s, books-read-in-2012	135
문예 사조		gothic, postmodern, surrealism, modernism	51
출판일		Published in 2000s, 1999, 2001	39
도서 입수 시기		acquired in 2007, bought 2008, purchased 2010	22

출판 등) 속성이 드러났다. ‘문예 사조’의 경우, 도서 자체가 지니는 성질로 해석하여 재료(M) 패킷으로 간주할 수도 있으나, 본 연구에서는 문예 사조가 시대적인 사상의 흐름을 반영한다는 측면에 방점을 두고, 시간(T) 패킷으로 해석하여 분석하였다.

4.3 도서태그 요소의 차별점

소설에 대해 일반 이용자들이 부여한 도서태그가 소설 정보를 전문적으로 제공하는 검색 서비스에서 제공하는 요소들과 어떠한 차별점을 드러내는가는 본 연구에서 소설 저작의 표집틀로 사용한 Fiction Core Collection의 서지정

보 요소와의 비교를 통해 이루어졌다. Fiction Core Collection과 같은 독자자문 DB는 기존에 도서관 OPAC과 온라인 서점에서 제공하는 요소들을 포함하고 있을 뿐만 아니라, 상대적으로 상세하고 다양한 소설 접근점을 제공하고 있는 것으로 드러났기 때문에(Adkins & Bossaller, 2007), 비교 대상으로 의미 있다고 여겼다. Fiction Core Collection에서는 〈그림 5〉와 같이 저자(AU), 타이틀(T1), 출판연도(PY), 출판사(PB), ISBN(SN), 초록(AB), 키워드(KW), 주기(N1) 등을 제공하고 있으며, 주기에는 가격정보, 페이지 수, LC 분류번호 및 제어번호, 독서 수준, 추천도서리스트, 전문가 리뷰 등의 정보를 포함하고 있다. 독자자문 DB에서 제공되는 서지



〈그림 5〉 Fiction Core Collection에서 제공하는 소설에 대한 서지정보 예시

정보 요소는 키워드와 주기의 일부 요소(예, 독서 수준, 추천도서리스트, 전문가 리뷰)를 제외하고 소설의 개체(P)와 재료(M) 패킷, 특히 외형적 성질과 관련되는 것을 볼 수 있다.

또한, Fiction Core Collection에서 제공되는 키워드는 재료(M) 패킷의 내용적 속성을 주로 드러내며, 장르와 주제 요소가 주류를 이루고 있었다. 그 외, 문학적 장치/기법(예, 풍자, 풍유법 등), 등장인물의 직업(예, 변호사, 교사, 기자 등), 성별(예, 독신 여성 등), 인종(북미 인디언 등), 등장인물의 관계(예, 어머니와 딸, 아버지와 아들 등), 플롯(예, 복수, 결혼문제 등) 요소 등이 표현되어 있었다(〈표 8〉 참고).

소설 7,174건에 대해 사회적 목록 서비스인 LibraryThing에서 식별한 도서태그와 소설 검색 전문 DB(이하 소설 DB)인 Fiction Core Collection에서 추출한 키워드 상위 100개를 비교해 본 결과, 소설 DB의 키워드의 경우, 주로 소설의 주제와 장르와 관련된 내용이 집중되는

것을 볼 수 있다. 그에 반해, 도서태그는 이용자의 독서활동과 관련된 내용이 다수 포함되어 있었다. 예를 들면 ‘앞으로 읽을 책’, ‘아직 읽지 않은 책’, ‘읽고 싶은 책 리스트’와 같이 독자가 소설과 어떻게 상호작용하는 지가 드러난 에너지(E) 패킷의 독서 상태가 주요한 속성으로 드러났다. 또한, 도서태그는 매체 형식(예, 전자책, 킨들, 오디오북 등)이나 독서 시기(예, 2011년에 읽음)와 관련된 태그도 비교적 상위에 분포되어 있었으며, ‘성인이 되는 과정(coming of age)’과 같은 플롯 정보 등 소설의 내용적 성질과 관련한 재료(M) 패킷에서도 풍부한 속성을 지니고 있었다(〈표 4〉 참고).

이러한 결과는 소설 색인 연구에서도 드러났듯이, 소설검색 전문 DB의 경우 소설 콘텐츠 자체가 지닌 요소(예, 주제, 플롯, 등장인물, 시간적 배경 등)에 초점을 맞추는 양상으로 드러난다. 반면, 도서태그의 경우 소설 탐색 상황에서 이용자의 다양한 요구 표현을 보여준다.

〈표 8〉 소설에 대한 도서태그와 소설검색 전문 DB의 키워드 Top 100 비교*

LibraryThing 도서태그				Fiction Core Collection 키워드			
1	to-read	51	English literature	1	Detectives	51	Terrorism
2	mystery	52	2010	2	Mystery and detective stories -- United States	52	African Americans
3	novel	53	21st century	3	Suspense novels	53	Boys
4	historical fiction	54	British literature	4	Love stories	54	Trials
5	fantasy	55	war	5	Short stories	55	Supernatural phenomena
6	science fiction	56	wishlist	6	Psychological novels	56	Physicians
7	read	57	adult	7	Murder stories	57	Adventure fiction
8	thriller	58	love	8	Science fiction	58	Brothers and sisters
9	literature	59	2011	9	Mystery fiction	59	Sisters
10	crime	60	vampires	10	Fantasies	60	Fathers and daughters
11	own	61	literary fiction	11	Family life	61	Violence
12	ebook	62	2012	12	Mystery and detective stories -- England	62	Poets
13	unread	63	goodreads	13	Historical fiction	63	Kidnapping
14	horror	64	book club	14	Satire	64	Women -- Social conditions
15	classic	65	France	15	Authors	65	Missing persons
16	romance	66	2009	16	Marriage problems	66	Secret service
17	England	67	2013	17	Adventure	67	Wealth
18	short stories	68	time travel	18	Love affairs	68	Death
19	series	69	American fiction	19	Small town life	69	Missing persons -- Fiction
20	20th century	70	relationships	20	Humor	70	Brothers
21	Kindle	71	women	21	Adolescence	71	Novelists
22	humor	72	2008	22	Crime and criminals	72	Autobiographical stories
23	sf	73	Adult Fiction	23	Law and lawyers	73	Mothers and sons
24	suspense	74	religion	24	Mothers and daughters	74	Homosexuality
25	historical	75	magic	25	Women -- Psychology	75	Eccentrics and eccentricities
26	American literature	76	London	26	Fathers and sons	76	Sea stories
27	American	77	read in 2011	27	Fantasy fiction	77	Family life -- Fiction
28	British	78	coming of age	28	Teachers	78	Women -- Relation to other women
29	paperback	79	espionage	29	Friendship	79	Life on other planets
30	crime fiction	80	read in 2010	30	Future	80	Indians of North America
31	audiobook	81	chick lit	31	Journalists	81	Women lawyers
32	hardcover	82	books	32	Family chronicles	82	Kings
33	family	83	ARC	33	Homicide -- Fiction	83	Old age
34	detective	84	India	34	Novelettes	84	Murderers
35	Roman	85	2014	35	Girls	85	Suicide
36	adventure	86	read in 2012	36	Parent and child	86	Ghost stories
37	favorites	87	pb	37	Horror stories	87	Experimental stories
38	murder	88	read in 2009	38	Single women	88	Family secrets -- Fiction
39	library	89	New York	39	Suspense stories	89	Bereavement
40	contemporary fiction	90	magical realism	40	Philosophical novels	90	Immigrants
41	WWII	91	speculative fiction	41	Marriage	91	Allegories
42	19th century	92	urban fantasy	42	Conspiracies	92	Picaresque novels
43	history	93	read in 2008	43	Widows	93	Ex-convicts
44	USA	94	friendship	44	Spies	94	United States -- Federal Bureau of Investigation
45	First Edition	95	read in 2013	45	Sex	95	Magic -- Fiction
46	sff	96	General Fiction	46	Women	96	Women authors
47	signed	97	2007	47	Western stories	97	Divorced persons
48	contemporary	98	borrowed	48	Revenge	98	Horror fiction
49	English	99	dystopia	49	International intrigue	99	Guilt
50	audio	100	death	50	Murder -- Investigation -- Fiction	100	Assassination

* 단, 도서태그에서 상위 1위인 fiction 태그는 식별력이 없다고 판단하여 제외

또한, 전문가들에 의해 연역적으로 구성된 분류 체계가 콘텐츠에 대한 해석을 통해 해당 자료를 가장 잘 드러낼 수 있는 색인어로 표현하여 제공되었다면, 도서태그의 경우, 다수 이용자의 다양한 관점이 원형 그대로 반영되어 드러나는 것으로 해석할 수 있다. 본 연구의 결과는 실제 이용 상황의 요구를 완전히 반영하지 못한다는 전문가 색인의 한계를 보완하기 위해, 독자들이 자주 사용하는 전형적인 요구 유형을 식별하고, 검색에 주로 관여하는 요소를 식별하고자 하는 근래 소설 색인 연구의 요구에도 부응한다.

한편, 상위 키워드에서는 드러나지 않았지만, 도서태그의 경우 독자 리뷰, 독자 반응, 독서 프로그램, 문학상 등 이용자의 독서활동과 소설 이용자의 다양한 해석과 관계된 요소(예, 저자의 의도, 내용을 읽고 이해하기 쉬운 정도(readability), 내용에 대한 감상 등) 등 독서활동과 관계된 동적 요소인 에너지(E) 패킷 측면에서 차별화된 속성이 두드러진다. 이러한 속성은 오늘날 전자책을 통한 소설 리딩(social reading) 방식 등 개인 단위의 독서활동뿐만 아니라, 타인과의 독서활동을 공유하는 사회적 독서가 장려되는 독서환경에서 이용자들에게 의미 있는 접근점으로 제공될 수 있으리라 본다.

5. 결론 및 제언

본 연구는 소설 검색 환경을 개선하기 위해, 도서태그로부터 소설 이용자가 소설 탐색 상황에서 요구하는 다양한 패킷 요소를 식별하고 체계화하는 것을 목적으로 한다. 특히, 본 연구에서는 자료의 속성을 표현하고, 구조화하는 데

효과적인 랑가나단의 패킷 개념을 소설의 특성과 관련지어 해석하여 도서태그 분석에 적용하였다. 도서태그를 대상으로 소설 고유의 요소(소설의 3요소, 소설 구성의 3요소)를 패킷 유형으로 규정한 연구(김동숙, 정연경, 2010)가 기준에 정의된 요소를 패킷으로 사용하였다면, 본 연구에서는 랑가나단의 PMEST 기본 패킷 체계를 응용하여 소설 자료의 이용자가 제시한 용어들에 기반한 패킷 유형을 개발하였다. 이처럼 현상에 기반한 방법은 이용자의 독특한 맥락을 발견하는 데 유용할 것이라 여겨진다. 일례로 호주 국립도서관 OPAC에서는 ‘직업’ 패킷이라는 독특한 유형의 접근점을 제공하여, ‘캥디드’라는 볼테르의 소설을 검색하면, 정치가, 변호사, 농부 등 소설에 등장했던 인물의 직업을 검색 접근점으로 제공해주고 있다. 이는 ‘직업’ 패킷이 유달리 호주 문화권의 이용자에 의해 탐색 접근점으로써 의미가 있음을 추정해 볼 수 있는 내용으로 검색 상황에서 이용자 관점을 분석하고 적용할 필요성을 역설하는 예이다. 이에 본 연구에서는 많은 이용자로부터부터 선택 받은 용어를 소설의 구조적인 속성을 추출하는데 사용하여 현상에 기반한 지적 접근점을 구현하려는 노력을 하였다.

본 연구를 통해 파악된 소설 검색의 패킷 요소는 전통적인 소설 색인에서 드러나지 않았던 요소는 물론, 기존에 색인 요소를 보다 세분화하는 요소들을 식별해 내었다. 특히 소설 자료와 독자와의 상호작용 행위와 관련되는 에너지(E) 패킷의 요소가 독서활동 요소(예, 도서 입수 경로, 독서 상태, 독자 반응, 독자 리뷰, 독서 프로그램 등)와 독서활동에 유용한 정보 요소(예, 추천도서리스트, 문학상, 독서 수준 등)로

식별되었고, 소설을 구성하는 내용적, 외형적 성질을 의미하는 재료(M) 패킷과 관련해서도 이용자들이 소설에 대해 매우 세부적으로 접근하는 양상을 파악할 수 있었다. 예를 들면, 등장인물 요소와 관련해서도 등장인물을 인간과 동물이나 신 등 인간 이외의 요소로 구분하고 있으며, 인간 가운데서도 이름뿐만 아니라, 직업, 인종, 성별, 역할, 관계, 성격, 연령, 종교와 같은 세부적인 속성으로 구분하여 탐색하고 있는 것을 파악할 수 있었다. 소설 속에서 다루지는 사실 정보와 관련해서도 역사적 인물, 사상, 역사적 사건, 기관, 사회제도 등 다양한 접근을 하고 있음을 드러냈다. 한편, 문학적 장치나 기법과 같이 전문적인 요소도 소설 탐색의 접근점으로 이용되고 있음을 파악하였다.

본 연구에서는 방대한 태그 데이터를 일관되고 신속하게 처리하기 위해, 전체 태그 데이터 집합에서 핵심 태그 집합을 선별하였으며, 소설 검색의 패킷 유형을 구분하는 내용분석을 하기 이전에, NER 기법을 적용하여 자동으로 유사한 패턴이 발견되는 태그들을 초별 분류하였다. 이러한 방식은 향후 유사한 연구에도 적용될 수 있으리라 본다. 다만, 본 연구의 샘플은 도서태그 활동이 활발한 영미권의 사회적 목록을 대상으로 하였기 때문에 영미권 소설 이용

자의 요구로 한정되며, 분석 결과에 문화적 차이가 내포될 수 있다. 가령, 목록규칙에서 동양권에서는 서명 기본기입을 원칙으로 하고, 영미권에서는 저자명 기본기입을 원칙으로 하듯이, 영미권의 저자 접근점 중시 경향이 반영되었을 수 있다. 또한, 반복적인 코딩 과정과 제2 코더에 의한 코더간 신뢰도 검증, 동료 검증의 프로세스에도 불구하고, 내용분석 과정에서의 연구자의 바이어스도 배제할 수 없다. 본 연구의 한계는 다양한 샘플을 대상으로 유사한 연구가 반복적으로 진행되는 것을 통해 보완되고 검증될 수 있을 것으로 보인다.

본 연구의 결과로부터 도출된 패킷 요소는 향후 소설 색인 등 메타데이터 작성에서 참고할 수 있을 것으로 보이며, 도서관 OPAC이나 소설 DB에 패킷 내비게이션 형태로 적용하여 이용자의 탐색 접근점을 확장할 수 있을 것으로 기대된다. 한편, 소설 이용의 접근점 개선을 통해, 최근 인문학 열풍과 더불어 독서문화를 장려하는 사회적 분위기와 소설 리딩(social reading)과 같은 변화된 환경의 기대에 부응할 수 있다는 점에서 또한 의미가 있다. 도서정보 이용자에게 소설 장르에 특화된 다양한 속성을 탐색 접근점으로 제시함으로써, 보다 세분화된 정보요구를 충족시킬 수 있을 것으로 기대된다.

참 고 문 헌

- 김동숙, 정연경 (2010). 저작권 관련 요소분석을 통한 폭소노미 태그의 활용 방안에 관한 연구. 정보관리학회지, 27(1), 41-60. <https://doi:10.3743/KOSIM.2010.27.1.041>
- 이승민 (2011). 서지레코드와의 연계를 통한 폭소노미 태그 프레임워크 구축. 한국문헌정보학회지,

- 45(2), 185-207. <https://doi:10.4275/KSLIS.2011.45.2.185>
- Adkins, D., & Bossaller, J. E. (2007). Fiction access points across computer-mediated book information sources: A comparison of online bookstores, reader advisory databases, and public library catalogs. *Library & Information Science Research*, 29(3), 354-368. <https://doi:10.1016/j.lisr.2007.03.004>
- Atdağ, S., & Labatut, V. (2013). A comparison of named entity recognition tools applied to biographical texts. Paper presented at the 2nd International Conference on Systems and Computer Science, Villeneuve d'Ascq, France. <https://doi.org/10.1109/icconscs.2013.6632052>
- Bartley, P. (2009). Book tagging on LibraryThing: How, why, and what are in the tags? *Proceedings of the American Society for Information Science and Technology*, 46(1), 1-22. <https://doi:10.1002/meet.2009.1450460228>
- Broughton, V. (2006). The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings*, 58(1-2), 49-72. <https://doi:10.1108/00012530610648671>
- Chung, Y. K. (2014). A study on varieties of subject access and usabilities of the national library of Korea subject headings. *Journal of the Korean Biblia Society for Library and Information Science*, 25(4), 171-185. <https://doi:10.14699/kbiblia.2014.25.4.171>
- Desrochers, N., Laplante, A., Martin, K., Quan-Haase, A., & Spiteri, L. (2016). Illusions of a "Bond": Tagging cultural products across online platforms. *Journal of Documentation*, 72(6), 1027-1051. <https://doi:10.1108/jd-09-2015-0110>
- Elson, D. K., & McKeown, K. R. (2010). Automatic attribution of quoted speech in literary narrative. Paper presented at the Twenty-Fourth AAAI Conference on Artificial Intelligence, Atlanta, Georgia, USA.
- Flick, U. (2019). *An introduction to qualitative research* (18th ed.). Los Angeles: Sage.
- Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2), 198-208. <https://doi:10.1177/0165551506062337>
- Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. MA: Addison-Wesley Pub. Co.
- Iosif, E., & Mishra, T. (2014). From speaker identification to affective analysis: A multi-step system for analyzing children's stories. Paper presented at the Proceedings of the 3rd Workshop on Computational Linguistics for Literature, Gothenburg, Sweden. <https://doi.org/10.3115/v1/w14-0906>
- Lawson, K. G. (2009). Mining social tagging data for enhanced subject access for readers and

- researchers. *Journal of Academic Librarianship*, 35(6), 574-582.
<https://doi.org/10.1016/j.acalib.2009.08.020>
- Lu, C. M., Park, J. R., & Hu, X. H. (2010). User tags versus expert-assigned subject terms: A comparison of LibraryThing tags and Library of Congress Subject Headings. *Journal of Information Science*, 36(6), 763-779. <https://doi.org/10.1177/0165551510386173>
- Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D. (2014, June). The Stanford CoreNLP natural language processing toolkit. Paper presented at the 52nd annual meeting of the Association for Computational Linguistics, System Demonstrations, Stroudsburg, PA. <https://doi.org/10.3115/v1/p14-5010>
- Morbhead, D. R., Pejtersen, A. M., & Rouse, W. B. (1984). The value of information and computer-aided information seeking: Problem formulation and application to fiction retrieval. *Information processing & management*, 20(5-6), 583-601.
[https://doi.org/10.1016/0306-4573\(84\)90075-X](https://doi.org/10.1016/0306-4573(84)90075-X)
- Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. *Linguisticae Investigationes*, 30(1), 3-26. <https://doi.org/10.1075/li.30.1.03nad>
- Pejtersen, A. M. (1980). Design of a classification scheme for fiction based on an analysis of actual user-librarian communication, and use of the scheme for control of librarian's search strategies. In O. Harboe & L. Kajberg (Eds.), *Theory and Application of Information Research* (pp. 146-159). London: Mansell.
- Pejtersen, A. M. (1997). *Subject access to Scandinavian fiction literature: Index methods and OPAC development*. Copenhagen: Nordic Council of Ministers.
- Pejtersen, A. M., & Austin, J. (1983). Fiction retrieval: experimental design and evaluation of a search system based on users' value criteria (part 1). *Journal of Documentation*, 39(4), 230-246. <https://doi.org/10.1108/eb026750>
- Pejtersen, A. M., & Austin, J. (1984). Fiction retrieval: experimental design and evaluation of a search system based on users' value criteria (part 2). *Journal of Documentation*, 40(1), 25-35. <https://doi.org/10.1108/eb026755>
- Ranganathan, S. R. (1967). *Prolegomena to library classification* (3rd ed.). Bombay; New York: Asia Publishing House.
- Saarti, J. (1999). Fiction indexing and the development of fiction thesauri. *Journal of librarianship and information science*, 31(2), 85-92. <https://doi.org/10.1177/096100069903100203>
- Saarti, J. (2019). Fictional literature, classification and indexing. *Knowledge Organization*, 46(4), 320-332. <https://doi.org/10.5771/0943-7444-2019-4-320>

- Satija, M. (2017). Reviews of Concepts in Knowledge Organization. *Knowledge Organization*, 44(4), 291-307. <https://doi:10.5771/0943-7444-2017-4-291>
- Smith, T. (2007). Cataloging and you: Measuring the efficacy of a folksonomy for subject analysis. Paper presented at the 18th workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research, Milwaukee, WI.
- Solomon, P. (1997). Access to fiction for children: A user-based assessment of options and opportunities. *Information Services & Use*, 17(2), 139-146. <https://doi.org/10.3233/isu-1997-172-308>
- Wilson, H. W. (2016). *Fiction Core Collection* (18th ed.). Ipswich, Massachusetts: Grey House Publishing, Inc.
- Wu, D., He, D. Q., Qiu, J., Lin, R. N., & 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:10.1177/0165551512451808>
- Yu, L., & O'Brien, A. (1997). Constructing references from the book to the reader in fiction searching, an experiment on the construction of information cues from the reading context approach. *Information Services & Use*, 17(2-3), 187-199. <https://doi.org/10.3233/isu-1997-172-315>

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

- Kim, Dong-Suk, & Chung, Yeon-Kyoung (2010). A study on the application of librarything folksonomy tags through the analysis of elements related with work. *Journal of the Korean Society for Information Management*, 27(1), 41-60. <https://doi:10.3743/KOSIM.2010.27.1.041>
- Lee, Seungmin (2011). Construction of folksonomy tag framework using bibliographic record. *Journal of the Korean Society for Library and Information Science*, 45(2), 185-207. <https://doi:10.4275/KSLIS.2011.45.2.185>

