

# 주제명 활용 분석을 통한 국립중앙도서관 주제명표목표의 현황 연구\*

## A Study on the Current Status of National Library of Korea Subject Headings List through Utilization Analysis of Subject Headings

이혜경 (HyeKyung Lee)\*\*

이용구 (Yong-Gu Lee)\*\*\*

### 초 록

이 연구는 국립중앙도서관에서 개발한 주제명표목표의 구성 현황과 2003년부터 2022년까지 국가서지 1,218,867건에 부여된 주제명의 현황을 분석하였다. 그 결과 첫째, 주제명표목표의 전체 주제명 중 우선어는 257,103개로 전체 용어의 50.2% 이상이었으며, 외국어가 169,466개(33.1%), 비우선어가 61,442개(12.0%) 등으로 구성되었다. 우선어 중에 활용된 주제명은 57,312종으로 22.3%에 해당하며 이 중 54.7%인 31,351종의 주제명은 부여횟수 5회 미만으로, 전체 중에서 적은 수의 주제명만을 활용하고 있음을 파악하였다. 둘째, 관계지시기호의 빈도는 RT, BT, NT 순으로 나타났으며, 최상위 주제명은 12,602종, 최하위 주제명은 143,704종이었고 최대 심도는 17수준이었다. 셋째, 서지 당 평균 1.72회의 주제명이 부여되었으며, 자료의 내용이 특정적일수록 주제명 부여 횟수가 많아지고 심도도 깊어지는 경향을 보였다. 그리고 최근에 입력한 서지일수록 부여된 주제명 수가 증가하였고 심도도 깊어졌으나, KDC 주류에 따라 서지 당 부여된 주제명 수는 편차가 있는 것으로 나타났다. 현황 분석을 통해 결과적으로 주제명표목표가 수록한 용어의 범위에 대한 평가와 주제명의 계층 관계 및 심도의 재정비가 요구되며, 주제명표목표의 세목 개발을 통한 개선이 필요한 것으로 판단하였다.

### ABSTRACT

This study analyzed the structure and utilization of subject headings in the National Library of Korea Subject Headings List (NLSH) based on an analysis of subject headings assigned to 1,218,867 national bibliographies from 2003 to 2022. The findings of the study are as follows: Firstly, among all subject headings in the NLSH, there were 257,103 preferred terms, accounting for 50.2% of the total terms. Foreign language terms constituted 33% (169,466), while non-preferred terms comprised 12% (61,442). Among the preferred terms, 57,312 subject headings were used, accounting for 22.3%. However, it was observed that 54.7% (31,351) of these subject headings were assigned less than 5 times, indicating that only a small number of subject headings were frequently utilized. Secondly, the frequency of relationship indicators appeared in the order of RT, BT, and NT. The NLSH consisted of 12,602 top-level subject headings and 143,704 lowest-level subject headings, with a maximum depth of 17 levels. Thirdly, on average, 1.72 subject headings were assigned per bibliographic record. The number of subject headings assigned and the depth of the hierarchy increased for materials with more specific contents. Recent bibliographic records have been assigned more subject headings and deeper into the hierarchy of the NLSH. It was also found that the number of subject headings assigned per bibliography varied depending on the main class of KDC. Based on the findings, it is recommended to evaluate the coverage of terms in the NLSH, reorganize hierarchical relationships and depth of subject headings, and enhance the development of subdivisions within the NLSH.

키워드: 주제명표목표, 국립중앙도서관 주제명표목표, 활용 분석, 국가서지, 정보 검색  
subject headings list, National Library of Korea Subject Headings List, utilization analysis, national bibliography, information retrieval

\* 본 연구는 2022년도 국립중앙도서관이 지원한 연구과제 “국가서지를 활용한 주제명 자동 분류 적용방안 연구”의 일부 내용을 수정·보완하였음.

\*\* 경북대학교 문헌정보학과 강사(keilee@knu.ac.kr) (제1저자)

\*\*\* 경북대학교 문헌정보학과 부교수(yglee@knu.ac.kr) (교신저자)

■ 논문접수일자: 2023년 5월 15일 ■ 최초심사일자: 2023년 6월 8일 ■ 게재확정일자: 2023년 6월 20일

■ 정보관리학회지, 40(2), 157-182, 2023. <http://dx.doi.org/10.3743/KOSIM.2023.40.2.157>

※ Copyright © 2023 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

정보 검색 관련하여 문헌정보학 분야를 비롯하여 도서관은 이용자에게 양질의 정보 혹은 콘텐츠를 제공하기 위해 그 대상이 되는 자료의 수집과 축적에 많은 노력을 기울이기도 하며, 효과적이고 효율적인 검색이 되게 하도록 다양한 장치와 방법을 제공하고 있다. 이러한 도서관만의 대표적인 장치 중 하나인 주제 전거(subject authority)는 하나의 저작에 대한 주제적 접근을 위해 편목 과정에서 수행하며 그 결과로 저작에 주제명이 부여된다. 현재 주제명 부여는 기계에 의한 처리되는 자동색인과 대비되어 아직까지 대부분 수작업으로 이루어지고 있으며, 통제된 형태로 문헌의 주제를 나타내는 용어인 주제명 또는 디스크립터로 구성된 주제명표목표나 시소러스와 같은 통제어휘집이 도구로서 이용되고 있다.

월드 와이드 웹(World Wide Web)상의 검색엔진 대부분은 문헌에 나타난 형태 그대로인 자연어를 키워드로 하여 색인어를 채택하는 방식을 취하고 있다. 이용자 또한 이러한 키워드 중심의 검색 방식에 익숙하여 자연어에 해당하는 일반적인 용어를 사용하여 자신의 정보 요구를 표현하고 질의를 만들어 검색엔진에 입력한다. 자연어를 색인과 검색에서 적용할 때 나타날 수 있는 문제로는 동의어, 동형이의어, 이형이나 단·복수형 등을 들 수 있다. 동의어는 같은 의미이지만 어형이 달라 자연어 색인에서는 다른 용어로 인식되며, 어형이나 발음이 같지만 의미가 다른 동형이의어나 다의어 경우에는 이와 반대이다. 이용자가 이들 단어를 질의어로 검색할 때 의도하지 않은 다른 의미로 쓰인 문헌이

검색되거나 다른 형태의 동의어가 쓰인 문헌이 누락되므로 검색 성능의 저하를 초래한다. 다만 최근의 검색엔진은 일부 비교적 간단한 동의어 정도의 문제를 해결하기도 한다. 예를 들어 구글의 경우 'pc'로 검색하면 'computer'가 들어 있는 문헌을 찾아주기도 한다.

색인과 검색에 통제어휘집을 적용하면 이러한 문제는 하나의 개념이 하나의 용어로 매칭되어 많은 부분 온전하게 해결하게 된다. 다만 이들 통제어휘집의 사용도 단점을 가지는데, 일부 연구(정연경, 2010)에서 지적하듯이 통제어휘집의 구축과 갱신에 비용이 많이 들고 노동 집약적이며 높은 정확성을 요구하는 검색에서는 효율적이지 않다. 그럼에도 불구하고 현재의 도서관들은 서지데이터를 기술할 때 디스크립터나 주제명을 함께 부여하고 있다. 그 이유는 앞서 설명하였듯이 이러한 작업이 이용자 검색에 도움이 되기 때문이다. 하지만, 노지현(2015)은 서지 레코드마다 주제명이 부여 되어있지만 실제 주제를 표현한 주제명보다 외형이나 표제에서 추출한 주제명을 부여하여 해당 자료가 지닌 주제를 제대로 표현하지 못하고 있어 이용자의 주제 검색에 주제명을 활용하기에 다소 문제가 있음을 지적하였으며, Julien et al.(2013)은 LCSH(Library Congress of Subject Headings)의 주제명이 계층 구조와 연관 관계가 체계적으로 확립되어 있지 않아, 이용자가 검색할 때 브라우징이 매우 복잡해지는 결과를 초래할 수 있음을 주장하였다. 이렇듯 통제어휘 도구인 주제명표목표 자체나 부여된 주제명 데이터의 품질로 인한 문제와 관련된 선행연구를 보면, 통제어 색인이 이론적으로 이상적이나 현실은 그렇지 못한 것으로 보인다.

이에 본 연구는 주제명표목표 관련 선행연구를 살펴보고 국립중앙도서관에서 개발한 주제명표목표(National Library of Korea Subject Headings List, 이하 NLSH) 자체에 대한 다양한 분석과 이를 이용하여 2003년부터 국립중앙도서관이 국가서지에 주제명표목을 부여한 현황을 분석하고자 한다. 이를 통하여 국립중앙도서관의 주제명표목표에 대한 개선이나 목록 데이터에서 주제 접근을 통한 검색에서 향상 방안 등을 위한 연구의 토대가 되었으면 한다.

## 2. 이론적 배경

### 2.1 주제 전거와 접근점

2010년 IFLA(International Federation of Library Associations and Institutions)에 발표한 주제 전거 데이터의 기능 요건(Functional Requirements for Subject Authority Data, FRSAD)은 서지 개체를 다룬 FRBR(Functional Requirements for Bibliographic Records)이나 인명 전거를 다룬 FRAD(Functional Requirements for Authority Data)에 이은 세 번째 시리즈로 저작의 주제 전거와 관련성에 대하여 논의한다. 이들 시리즈에서 전거 제어의 본질은 정보검색을 위한 접근점이 되는 인명이나 지명, 주제명을 일관되게 표현하기 위함이다(IFLA Study Group on the FRBR, 1998). 정옥경(2002)은 앞서 IFLA에서 제정한 '전거저록과 참조저록을 위한 지침(Guidelines for Authority and Reference Entries, GARE)'이 제시하는 '전거 통제(제어)'의 정의는 서지 레코드에 대해 '통제된 표목'을

부여하는 것으로 이때 일관성 있게 그 표목이 채택되도록 하는 방법을 말하며, 일반 용어를 부여하고 통제하는 것이 아닌 것으로 해석하였다. 즉, FRBR 시리즈에 등장하는 전거 데이터라고 다루어지는 개체는 통제어휘를 활용하는 것으로, 마찬가지로 FRSAD 역시 LCSH에 의한 전거 제어에 대한 예로써 이를 설명하고 있다.

접근점은 전거형 접근점(authorized access point)과 이형 접근점(variant access point)으로 나눌 수 있으며, 이중 전거형 접근점은 자료에 대한 표준화된 접근점을 말한다. 특히 주제 전거를 위한 접근점의 부여는 주제명표목표의 주제명표목을 적용하며, 이때의 주제명표목이란 저작의 주제를 표현하는 표목(올림말)이 되는 단어나 구를 말하고, 주제명표목표란 주제명표목을 사용하기 위해 정해진 용어표를 말한다(한국도서관협회, 2010). 원칙적으로는 주제명으로 사용할 어휘의 범주나 어휘집들이 지정되어있는 것이 아니므로, 주제명표목은 기관이나 개인마다 정보의 주제를 파악하여 부여의 주체가 해당 정보에 임의로 부여할 수 있지만 주제명표목으로 사용할 어휘에 대하여 국내의 경우 국립중앙도서관의 NLSH, 국외의 경우 미국의회도서관의 LCSH를 고려하는 것이 일반적일 것이다. 이는 앞서 FRSAD의 관점에서 살펴보면 전거 데이터가 '통제된 표목'을 사용했을 때 다양한 정보에서의 동일한 주제에 일관적인 주제명표목이 부여되어 정보 검색의 접근점을 가질 수 있기 때문이다.

한편 국립중앙도서관에서는 주제명표목표에 대하여 주제명을 명시하고, 이들 간의 관계를 체계화한 어휘집으로 목록 담당 사서와 이용자가

각각 색인과 탐색을 수행함에 있어 서로가 의도한 바를 모호함 없이 이해할 수 있도록 증가하는 역할을 함과 더불어 특정 형식의 문헌이나 유사 주제 문헌에 대하여 군집화가 가능하고 나아가 용어 제어 장치로서의 활용으로 그 기능을 언급하고 있다(국립중앙도서관, 2021).

또 하나의 대표적인 주제명표목표로 알려진 LCSH는 1895년 자관의 소장목록에 대하여 체계적인 분류 및 접근을 위하여 'the List of Subject headings for use in dictionary catalog' 우선 발간하고 이후 1987년 미국의회도서관분류법(LCC)이 제정되고 나서 1898년 공식 제정되었다. Drabenstott와 Vizine-Goetz(1994)는 온라인 정보 검색에서도 LCSH가 꾸준히 사용될 수밖에 없는 이유로 첫째, 다양한 지식 수준의 이용자가 LCSH를 사용한다면 자신의 수준에 알맞은 주제어를 선정해 정보를 찾을 수 있다는 점, 둘째, 현재까지 주제 검색 측면에서 LCSH보다 정교하게 주제를 분류하여 정보를 제공하는 시스템이 등장하지 않았다는 점, 셋째, LCSH가 가진 주제 계층성에 대해 장점으로 작용하는 점 그리고 LCSH가 적용된 온라인 서지 레코드가 이미 수천만 건에 달하며 이용자는 이러한 서지 레코드를 LCSH를 통해 접근 가능하다는 점을 들었다.

최근의 정보검색시스템은 자연어를 통한 정보 검색이 가능하지만 살펴본 것처럼 국립중앙도서관에서도 주제명표목표를 '용어 제어 장치로서의 활용'이라는 기능을 제시하는바, 주제명표목표를 통하여 자연어를 통제어로 변환하거나 우선어로서의 연관 등을 통해 검색 결과의 주제 일관성을 확보할 수 있음을 내포한다.

## 2.2 주제명표목표의 구성과 계층 구조

FRSAD는 이용자의 과업으로 첫째, 이용자 요구에 부응하는 개체들의 명칭을 발견하는 것, 둘째, 속성과 관계를 토대로 개체와 주제를 식별하는 것, 셋째, 이용자의 요구에 적합한 주제나 그렇지 못한 것을 제외하거나 선정하는 것, 넷째, 주제와 개체의 명칭 관계를 탐험하는 것으로 정의하였고, 실제 이용자들은 주제 전거 데이터를 목록이나 메타데이터 작성 시에 관계를 탐험하기 위하여 사용한다고 응답했다(박지영, 2011). 이는 정연경(2010)의 연구에서 LCSH의 장점이 해당 주제어를 넓히거나 좁히거나 혹은 연관 주제어를 브라우징하는 것이라는 미국 대학도서관 사서들의 면담 결과와도 일맥상통한 것으로 확인된다. 이런 점에서 통제어휘집에서의 용어와 용어 간의 관계는 매우 중요하다고 여겨질 수 있으며, 대표적인 주제명표목표인 미국 의회도서관의 LCSH나 국내의 국립중앙도서관의 NLSH 역시 주제어와의 관계를 기술한 시소러스 형태를 보인다.

국립중앙도서관의 NLSH는 전 주제를 망라하는 수록 범위, 전문용어를 포함한 상세한 용어 수준으로 구축된다. 또한 주제명표목표로서 지데이터에 부여되는 용어는 우선어의 범주이며 이외는 모두 비우선어이다(국립중앙도서관, 2021). LCSH에서도 NLSH와 유사하게 어휘를 구축하고 있으며, 2022년 3월 현재 LCSH 제 44판은 총 382,713건의 전거 레코드를 수록하고 있다. 특히 LCSH는 SACO(Subject Authority Cooperative Program)에서 승인과정을 거쳐 매월 새로운 용어와 관계를 추가하여 온라인을 통해 배포하고 있는데, 신설된 주제나 사회

상황에 의해 조정되거나 변경된 주제, 참조 범위나 주기 등의 정보를 지속적으로 이용자에게 제공한다(Library of Congress, 2022). 더불어 LCSH는 주제의 세부화나 특징적인 주제를 표현할 수 있도록 고유 주제명에 조합할 수 있는 주제별 세목, 형식 세목, 연대 세목, 지리 세목뿐만 아니라 일반적 지리 세목과는 별개로 주제에 따라 구분된 지명의 지리 세목 그리고 미확정 세목(free-float subdivision)까지 총 6가지의 세목이 존재하며, 이는 NLSH와는 차별되는 LCSH만의 특징적인 요소라 할 수 있다.

한편, 주제명표목표는 시소러스와 같이 주제명 간의 관계가 표현된 것이 대표적인 특징(Julien et al., 2013)으로 각 관계는 동등 관계인 USE와 UF, 계층 관계인 BT와 NT, 연관 관계인 RT, 그리고 일반 참조 SA(Such As)로 표현한다. 그렇기에 주제명표목표 구성이라 함은 주제명표목표의 요소뿐만 아니라, 계층에 따른 관계도 함께 포함하며, 주제명표목의 수집을 대상으로 하는 연구들은 주로 실제의 주제명표목표를 수집하거나, 서지데이터에서의 주제명표목부출 필드인 600 태그 군을 추출하여 진행한다.

주제명표목표의 자체 구조적 분석을 수행한 Yi와 Chan(2010)에 따르면, 2009년 6월 5일 개정된 LCSH의 전체 342,684개의 주제명에서 이동문항용 939개의 주제명을 제외한 341,745개의 주제명 중 약 32.3%인 110,257개 주제명이 용어 간의 하위관계를 갖고 있었으며, 67.7%인 231,488개 주제명은 하위관계가 없으면서 동의어나 연관 관계를 갖는 것을 확인하였다. 결과적으로 실험 대상에 해당하는 전체 주제명

이 어떠한 유형이든 다른 주제명과의 관계가 나타나고 있다는 것이 밝혀졌다.

Julien et al.(2012)은 McGill 대학교 도서관에서 제공한 204,430건의 주제명과 Schulich 과학 및 공학 도서관에서 수집한 130,940건의 서지데이터를 분석한 결과, 한 레코드 당 대략 1.94개의 LCSH 주제명표목이 부여되었다. 실험 대상에서 부여된 주제명이 가지는 평균 심도는 17이며 최대 심도는 46이었으며, 최상위 주제명부터 내려온 심도는 평균 20으로 나타났다. O'Neill과 Aluri(1981)는 OCLC 단행본 데이터 33,455건에 대하여 1981년 당시 레코드 당 1.41건의 LC 주제명이 부여되고 있음을 밝혔으며, Frost와 Dede(1988)는 OCLC 데이터에서 650 태그와 지명 관련한 주제명 부출 태그인 651 태그에서 4,384건의 데이터를 수집해, 레코드 당 평균 1.8건의 주제명과 1.4건의 세목이 부여되고 있음을 확인하였다.

## 2.3 국내 주요 선행연구

주제명표목표에 관련한 국내 연구는 주로 국립중앙도서관 주제명표목표의 자체 품질 파악하거나 개선을 위한 연구, 주제명표목표 내의 주제명의 전개 현황을 파악하는 연구, 타 국가의 주제명표목표와의 비교를 통해 시사점을 밝히는 연구 등으로 구분할 수 있었으며, 주제명표목표와 관련하여 최근 수행된 연구는 김해인과 정연경(2022)의 구술기록 데이터에서의 주제명표목표 개발을 제안하는 것이고, 국립중앙도서관 주제명표목표에 한정된 연구 중 가장 최신의 연구는 노지현(2015)의 연구로 확인되었다. 이에 본 연구는 2022년 국립중

양도서관 주제명표목표와 국가서지 데이터를 기반으로 연구를 수행한 것으로, 관련 분야 연구의 최신성과 선행연구들의 결과와의 비교를 통해 과거에서부터 주제명표목표의 성과나 발전 등을 전반적으로 파악할 수 있을 것으로 사료된다. 선행연구를 세부적으로 고찰하면 다음과 같다.

김정현과 문지현(2009)은 미의회도서관 LCSH의 초판부터 2007년 개정된 30판까지 등장한 한국 관련 주제명의 변천을 전반적으로 고찰하면서, 한국 관련 주제명의 전개나 누락에서 나타나는 불균형을 해소하기 위해서 최근의 연구 성과나 문헌 등을 적극적으로 미의회도서관이 반영할 수 있도록 노력해야 함을 강조하였다. 이후에서도 LCSH의 초판부터 31판까지의 한국, 중국, 일본 관련 주제명의 변천 과정을 분석한 김정현(2010)의 연구에서도 과거 연구와 동일한 결과를 피력하였다. 한편 문지현과 김정현(2011)은 일본국회도서관 주제명표목표에서의 한국과 중국 관련 주제명표목표의 전개 현황도 파악하였는데, 일본의 주제명표목표에서도 중국에 비해 한국의 주제명은 매우 한정적으로 등장하였으며, 특히 국가 간의 논란 소지의 용어에 대하여 중국에 우세한 용어를 선택하는 등의 문제로 객관성에 의구심을 피력하였다.

정연경(2010)은 북미 도서관 사서들을 대상으로 웹 서베이나 온라인 면담 등을 통해 정보의 주제 접근 도구로의 주제명표목표의 유용성을 확인하기 위해 LCSH의 사용성과 장단점, 개선사항 등을 파악하였다. 그 결과 LCSH는 그 구성이 자연어에 비해 매우 엄격하고 한정적이어서 이용자 요구나 키워드 탐색에 취약할

수 있으므로 보완하는 것이 필요하며, 주제명간의 계층 구조와 관계를 이용하여 포괄적으로 관련 주제어를 파악할 수 있다는 강력한 장점을 부각하여 효과적인 주제 접근 도구로의 역할을 수행할 수 있도록 개선해야 함을 주장하였다.

최윤경과 정연경(2014)은 2013년까지 등록된 국립중앙도서관 주제명표목표의 주제명과 관계데이터에 대하여 주제명의 속성과 관계, 비활용 주제명 및 신규 신청 주제명의 특성, 비우선어와 미등록 용어 등의 현황을 파악하고 분석하였다. 그 결과 주제명표목표 매뉴얼과는 상이하게 주제명 기술이나 계층 관계의 비밀관성 등이 발견되었고 용어의 의미가 모호하거나 불필요하거나 불확실한 관계 등을 문제점으로 지적하면서 용어나 관계 면에서 정비가 요구되며 더불어 국립중앙도서관의 서지데이터의 정비도 필요함을 피력하였다. 또한 주제명표목표의 안정적인 유지관리를 위해서 전문가 자문 및 활용을 개선 방안으로 내세우는 등의 주제명표목표의 고품질화를 위한 전반적인 전략을 제안하였다. 그리고 백지원과 정연경(2014)은 국립중앙도서관 주제명표목표의 검색시스템에 대해서도 가시성이 부족하며, 주제명표목표의 구조를 명확히 나타내지 못하는 부실한 검색시스템의 접근성, 이용자 편의에 따른 검색시스템의 부재 등의 문제점을 밝혀내 주제명표목표를 이용한 검색시스템의 개선 방향을 나타냈다.

노지현(2015)은 주제명 데이터의 양과 속성 용어 수준 등에서 국립중앙도서관의 주제명 데이터의 품질 검증을 수행하였고, 그 결과 국립중앙도서관은 비교군이었던 미의회도서관 및

대학도서관보다 상대적으로 개별자료에 적은 양의 주제명 데이터를 부여하고 있는 것으로 나타났다. 더불어 국립중앙도서관의 주제명 데이터의 문제로 표현의 특정성과 일관성을 지적하고 있는데, 특히 사회과학 및 문학의 경우, 상 하위 계층 관계를 고려하지 않은 주제명 부여나, 장르를 무시하고 형식에 의존한 주제명 부여 등을 그 사례로 제시하면서, 데이터의 품질 향상을 위해서는 이러한 문제의 심각성을 인지하고 자료조직 실무를 개선할 것을 촉구하였다.

여지숙 외(2022)는 국립중앙도서관 주제명표목표에서 한인디아스포라 관련 주제명표목을 파악하고자 KCI(Korea Citation Index)에서 디아스포라 관련 학술논문에서 한인과 관련 용어를 추출하고 이 용어들의 국립중앙도서관 주제명표목표와 미국의 LCSH, 일본 국립국회도서관 주제명표목표 현황을 확인하였다. 그 결과 타 국가와 상대적으로 한인디아스포라 관련 용어의 계층이 부정확했으며, 용어 간의 우선어 혼용 등의 문제를 밝히면서 각 용어 간의 우선어나 계층성에 대한 재검토와 개선할 것을 권고하였다.

김혜인과 정연경(2022)은 구술기록 데이터의 효과적인 주제 검색을 가능하게 하기 위하여 실제의 녹취록을 분석하고, PMEST 패킷에 의거한 총 540개의 주제어를 도출해 주제명표목표를 개발하였다. 이는 도서관 서지정보에만 국한한 주제명표목표의 개념을 타 분야에 적용한 것으로 다양한 데이터의 주제 접근을 위한 주제명표목표의 개발 가능성을 시사하였다.

### 3. 연구방법 및 데이터 처리

국가서지는 일반적으로 한 국가에서 발간된 모든 출판물의 서지정보를 망라적으로 수록한 목록이다. 우리나라에서는 1964년부터 소급 조사하여 현재까지 국가서지를 발행하고 있는데, 처음에는 책자 형태였으나 CD-ROM, DVD-ROM 형태를 거쳐 현재는 국립중앙도서관에서 웹 서비스로 제공하고 있다. 국가서지는 기본적으로 한 나라에서 발행된 출판물의 총량을 알 수 있을 뿐만 아니라 그 나라의 문화, 현재의 관심사나 특성을 반영하기에 중요하다 할 수 있다.

이 연구는 국립중앙도서관에서 개발한 NLSH의 분석과 2003년부터 국립중앙도서관이 이 주제명표목표를 이용하여 국가서지에 부여한 주제명표목의 현황을 분석하고자 하였는데, 구체적으로 전자는 주제명표목표의 구성 현황으로 용어, 관계 기호, 계층과 심도, 용어의 범주 유형, 활용 현황 등을 분석하였다. 후자의 경우 서지레코드가 가지는 정보를 이용하여 주제명표목에 대해 한국십진분류법(KDC) 주류 및 시계열 추이를 분석하여, 국가서지에서 주제명표목의 부여 및 활용 현황을 파악하고자 하였다.

이 연구의 주된 분석 대상 데이터는 오프라인 중심의 단행본 도서의 서지데이터이며, 이 데이터에는 분류와 목록과정에서 NLSH를 이용하여 부여한 주제명표목이 포함되어 있다. 개체-관계 모형(Entity-Relation model) 관점에서 보면 이러한 데이터는 크게 두 개의 개체로 볼 수 있다. 하나는 서지 정보 또는 서지레코드가 나타내는 도서이며, 다른 하나는 특정 도서에 담긴 주제를 표현하는 주제명표목과 이의 원천인 주제명표목표이다. 이 둘은 다대다(N:M) 관계

에 해당하는데 하나의 도서(또는 서지레코드)는 여러 개의 주제(또는 주제명표목)를 가질 수 있으며, 반대로 하나의 주제명표목은 다수의 서지레코드에 부여될 수 있기 때문이다. 이 두 개체의 다대다 관계로부터 주제명의 미부여나 활용 정도를 파악할 수 있다. 아울러 각각의 개체는 이 연구에 필요한 주요 속성을 가지는데 서지레코드의 경우 KDC 분류기호, 입력일자 등이 있으며 주제명표목표는 우선어와 비우선어와 같은 용어분류, 관계 기호, 계층 및 심도, 용어의 범주 유형 등이 있다.

주요 분석에 사용된 이 속성들을 대상으로 데이터 전처리를 적용하였다. 우선 KDC 주류에 따른 주제명의 부여 정도를 파악할 필요가 있는데 이때 서지레코드에는 1개 이상의 KDC 분류기호가 부여될 수 있다. 2개 이상 부여된 서지레코드의 경우 먼저 부여된 분류기호를 대상으로 하였으며, 소수지만 분류 작업이 완료되지 않은 작성 중에 있는 서지레코드나 주제명이 미기입된 37건의 서지레코드는 실험에서 제외하였다. MARC 데이터로부터 008필드에서 2자리의 입력연도를 추출하여 시계열에 따른 추이 분석에 사용하였다. 그리고 NLSH에서 주제명의 범주 정보로부터 최상위 범주 유형을 추출하였으며, 계층 관계에 따라 최상위어로부터 심도를 계산하였다. 이러한 전처리와 데이터 분석을 위해 파이썬의 pandas 패키지를 사용하였다.

분석을 위한 데이터의 현황을 살펴보면, 주제명표목표에서 부여 가능한 우선어는 257,103종이며 이중 57,312종이 한번 이상 활용되었다. 주제명이 부여된 서지데이터 총 1,218,867건이며 주제명의 부여 및 활용 횟수는 총 2,097,682

회에 해당한다.

## 4. 주제명표목표 구성 및 주제명 부여 현황 분석

### 4.1 주제명표목표 구성 현황

#### 4.1.1 전체 용어 현황

국립중앙도서관은 2002년부터 『국립중앙도서관 주제명표목표』를 개발하고 2013년 주제명표목표의 고품질화 연구를 통해 주제명표목표 업무지침을 제정하였다.

국립중앙도서관 주제명표목표의 용어는 총 511,980개로, 우선어가 257,103개이며 비율이 50.2%로 주제명표목표 중 절반 정도의 비중을 차지하고, 이어 외국어 169,466개로 33.1%, 비우선어 61,442개로 12.0%, 분류어 23,969개로 4.7% 순으로 나타났다(〈표 1〉 참조). 여기서 비우선어의 경우 우선어와의 USE/UF 관계에 해당하는 용어를 말한다. 외국어의 경우 한국어에 대응하는 영어 또는 중국어 등과 같이 다른 국가의 언어를 말하며, 분류어의 경우 특정 주제명의 대한 KDC(6판) 분류표상의 분류 기호를 의미한다. 한편, 『국립중앙도서관 주제명표목 업무지침』(국립중앙도서관, 2021)에 의해 우선어를 주제명으로 서지데이터에 부여하고 있으므로 우선어를 제외한 나머지는 활용하지 않는다.

이에 우선어를 대상으로 각 활용 현황을 살펴보면 〈표 2〉와 같다. 전혀 부여되지 않은 비활용 주제명은 199,791종으로 우선어 중 77.7%를 차지하였고, 한 번이라도 서지 레코드에 부



〈표 1〉 주제명표목표 용어별 종수 현황

용어 분류	종수 (비율)
우선어	257,103 (50.22)
비우선어	61,442 (12.00)
외국어	169,466 (33.10)
분류어	23,969 (4.68)
소계	511,980 (100.0)

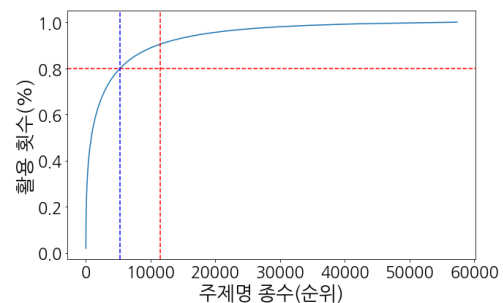
〈표 2〉 활용 수준별 주제명 종수

활용 수준	종수(비율, %)	수준별 합계(비율, %)
미 부여	199,791 (77.7)	199,791 (77.7)
1회 ~ 30회	48,499 (18.9)	57,312 (22.3)
31회 ~ 50회	2,657 (1.0)	
51회 ~ 70회	1,457 (0.6)	
71회 ~ 99회	1,194 (0.5)	
100회 이상	3,505 (1.4)	
합계	257,103 (100.0)	257,103 (100.0)

여되어 활용된 주제명은 57,312종으로 나타났다. 이 중에서 100회 이상 활용된 경우는 6.1%로 3,505종에 불과하였으며 이는 주제명표목표의 전체 우선어 중 1.4%에 해당한다. 특히 30회 이하로 활용된 48,499개의 주제명은 활용 주제명 57,312종의 84.6%이며 전체 우선어의 18.9%로 나타났다.

〈표 2〉에서 한번 이상 활용된 주제명 57,312종에 대해 활용 횟수 순위에 따라 고빈도부터 누적하여 그래프로 나타내면 〈그림 1〉과 같다. 주제명 활용 횟수에 대한 파레토 법칙(또는 80:20 법칙)을 적용한다면, 그 의미는 고빈도 주제명 20%가 전체 활용 횟수의 80%를 차지함을 뜻한다고 볼 수 있다. 이 그림에서 주제명 활용이 파레토 법칙에 따르면 20%에 해당하는 11,462종의 주제명에서 활용 횟수 80%의 누적이 나타나야 하나(붉은 점선), 11,462종의 실제

누적 현황은 이미 80%를 넘어선 90.4%에 해당하였으며(세로 붉은 점선과 곡선의 교차점), 활용 횟수의 80%에 해당하는 주제명은 5,312종으로 전체 주제명의 9.3%에 해당하였다(파란 점선). 이는 활용 주제명 중 현저하게 적은 주제명이 매우 많이 활용되고 있음을 보다 극명하게 보여준다.



〈그림 1〉 주제명의 활용 횟수 순위에 따른 누적 현황

한편 <표 2>에서 1회 이상 30회 이하 활용 주제명의 종수를 횡수에 따라 구체적으로 살펴보면 <표 3>과 같다. 1회 활용된 주제명의 종수는 16,376종으로 한 번 이상 활용된 주제명 57,312종의 28.6%를 차지하였으며, 2회 활용된 주제명 종수는 7,429종으로 13.0%를 차지하였다. 이렇게 활용 횡수 4회까지의 누적 주제명 종수와 비율은 31,351종으로 54.7%였으며, 10회 이하의 활용 주제명은 총 40,292종으로 이는 한 번 이상 활용된 주제명 57,312종 중 70.3%에 달하

여, 대체로 주제명의 전반이 10회 이하의 적은 횡수로 사용되고 있고, 이들 중 절반 이상이 5회 미만으로 사용되고 있음을 확인할 수 있었다.

<표 4>는 활용 주제명 중 활용 횡수 상위 30위의 주제명 현황을 보여주는 것으로, 가장 많이 부여된 주제명은 41,147회의 “한국 현대 시”였으며, 32,590회의 “한국 현대 소설”, 28,681회의 “한국 가요”가 그 뒤를 이었다. 그리고 문학 장르나 문학을 표현하는 주제명이 상위 30위 중 13종으로 약 43%를 차지하고 있었으며, “한국 가

<표 3> 저빈도 활용 횡수에 따른 주제명 종수(1~30회)

활용 횡수	종수	비율(%)	활용 횡수	종수	비율(%)	활용 횡수	종수	비율(%)
1	16,376	28.57	11	836	1.46	21	330	0.58
2	7,429	12.96	12	790	1.38	22	321	0.56
3	4,493	7.84	13	690	1.20	23	272	0.47
4	3,053	5.33	14	605	1.06	24	289	0.50
5	2,315	4.04	15	552	0.96	25	266	0.46
6	1,844	3.22	16	500	0.87	26	264	0.46
7	1,476	2.58	17	447	0.78	27	256	0.45
8	1,324	2.31	18	401	0.70	28	226	0.39
9	1,047	1.83	19	406	0.71	30	192	0.33
10	935	1.63	20	381	0.66	-	-	-
1-10	40,292	70.30	11-20	5,608	9.79	21-30	2,416	4.22

<표 4> 고빈도 활용 주제명(상위 30위)

주제명	활용 횡수	주제명	활용 횡수	주제명	활용 횡수
한국 현대 시	41,177	유아 교육	8,009	인디 음악	5,442
한국 현대 소설	32,590	수기(글)	8,007	영어 학습	5,324
한국 가요	28,681	임상 실습	7,954	한국 문학	5,170
대중 가요	24,537	교사 연수	7,264	영미 문학	5,114
고문서	15,877	한국(국명)	6,721	미국 영화	5,011
공무원 시험	12,421	한국사	6,613	전시 도록	4,893
기독교	10,693	일본 현대 소설	6,032	인생훈	4,892
한국 현대 수필	9,454	판타지 소설	5,931	창작 동화	4,828
창작 그림책	8,823	애정 소설	5,908	유적 발굴 조사	4,696
한국 현대 문학	8,629	발라드	5,871	간호학	4,657

요”, “대중 가요”, “발라드”와 같이 음악의 장르를 표현하는 주제명도 나타났다. 한편, 상위 30위 중 활용 횟수가 1만 회 이상인 주제명은 상위 7위까지였으며, 30위 주제명은 4,657회 부여된 “간호학”이었다. 이처럼 실제 주제명에서 장르와 같은 형식이 주제명으로 쓰이는 경우가 상위에 분포함을 볼 수 있으며, “한국[국명]” 주제명과 같은 지명도 6,721회 활용된 것을 확인할 수 있었다.

이에 따라 활용 횟수가 높은 주제명 상위 300위까지 확대하여 LCSH 세목(subdivision)의 성격을 띠는 형식(form)이나 지명을 나타내고 있는 주제명과 그렇지 않은 것으로 구분하여 그 활용 횟수를 확인하면 <표 5>와 같다. 300종 중에서 형식을 표현하는 주제명은 94종, 지명은 17종이었으며, 이들은 모두 300종 중 111종으로 약 37%에 해당하고, 300종의 총 활용 횟수 752,357회의 51%인 383,958회에 달했으며, 이는 전체 국가서지의 주제명 활용 횟수 중 18.3%에 해당하였다. 한편, 300종 중 일반 주제를 표현한 주제명은 199종이었으며, 300종의 총 활용 횟수 중 49%를 차지하였다. 즉, 상위 300위 활용 주제명에서 일반 주제명의 분포는 많았지만, 실제

많이 활용된 주제명은 자료의 주제를 표현한 주제명보다 형식이나 지명을 나타내는 주제명임을 시사한다.

주제명표목표에 사용된 주제명의 범주 유형은 <표 6>과 같이 총 14개 범주로 확인되었으며, 가장 많은 주제명 범주는 갈래 없는 일반 ‘주제어’로 90.4% 정도를 차지하였다. 이어 ‘지명’이 3.4%에 이은 비율을 보였다. 그러나 범주 유형별로 실제 부여된 주제명을 확인하면 일반 주제어의 경우, 53,319개에 불과하였으며, 지명은 1,098개였다. 동물 주제어로 범주화된 주제명은 6,411개로 식물 주제어는 4,173개보다 2,238개가 많지만, 실제 부여된 주제명으로는 식물 주제어가 동물 주제어 보다 약 100개 정도 더 많이 쓰이는 것을 확인할 수 있었다.

<표 7>은 주제명의 범주 유형과 활용 횟수에 따른 주제명 종수를 나타낸 것으로, 1,000회 이상 활용된 주제명을 보유한 범주는 일반 주제어와 지명, 법률명, 통일서명, 국명, 상품명, 행정구역이었다. 대부분의 범주별 주제명은 2-15회 정도 활용되는 것을 확인하였으며, 한 번만 활용된 주제명이 가장 많은 범주는 식물, 국보·보물과 회의명이었다.

<표 5> 고빈도 활용 주제명의 세목 현황(상위 300위)

구분	유형		종수	활용 횟수	비율 (상위 300)	비율 (전체 활용횟수)
세목	형식	문학	68	243,443	32.4%	11.6%
		음악	16	89,863	11.9%	4.3%
		영화	10	17,993	2.4%	0.9%
	지명	지역명	12	19,045	2.5%	0.9%
		국명	5	13,614	1.8%	0.6%
	소계		111	383,958	51.0%	18.3%
일반 주제명			199	368,399	49.0%	17.6%
합계			300	752,357	100.00%	35.90%

〈표 6〉 주제명 범주 유형별 현황

주제명 범주 유형	전체 종수	비율(%)	활용	비율(%)	비활용	비율(%)
주제어	232,524	90.44	53,320	93.03	179,204	89.70
지명	8,836	3.44	1,098	1.92	7,738	3.87
동물<생물<주제어	6,411	2.49	581	1.01	5,830	2.92
식물<생물<주제어	4,173	1.62	689	1.20	3,484	1.74
법률명<주제어	1,736	0.68	514	0.90	1,222	0.61
통일서명	1,501	0.58	580	1.01	921	0.46
국보·보물<기념물<주제어	745	0.29	128	0.22	617	0.31
기념물<주제어	587	0.23	95	0.17	492	0.25
국명<지명	356	0.14	231	0.40	125	0.06
상품명<주제어	97	0.04	40	0.07	57	0.03
회의명	71	0.03	11	0.02	60	0.03
행정구역<지명	60	0.02	22	0.04	38	0.02
국내기관·단체명<기관·단체명	4	0	2	0	2	0
기타	1	0	1	0	0	0
소계	257,102*	100	57,312	100	199,790	100

\* 전체 우선어는 257,103종이나 ID가 누락된 주제명을 제외하여 처리함.

〈표 7〉 범주 유형과 활용 횟수에 따른 주제명 종수

주제어 종류	활용 횟수	1	2 ~ 15	16 ~ 30	31 ~ 50	51 ~ 100	101 ~ 500	501 ~ 1000	1001 ~ 10000	10001 ~	소계
주제어		14,983	25,652	4,465	2,493	2,491	2,677	326	226	7	53,320
지명		283	480	85	49	100	85	10	6	0	1,098
동물<생물<주제어		255	257	27	18	11	12	1	0	0	581
식물<생물<주제어		357	272	27	17	11	5	0	0	0	689
법률명<주제어		146	248	42	27	24	24	1	2	0	514
통일서명		172	281	44	27	22	29	3	2	0	580
국보·보물<기념물<주제어		93	27	5	1	1	1	0	0	0	128
기념물<주제어		36	50	5	2	0	1	1	0	0	95
국명<지명		30	97	25	19	19	32	4	5	0	231
상품명<주제어		6	12	8	4	3	5	1	1	0	40
회의명		7	4	0	0	0	0	0	0	0	11
행정구역<지명		6	8	1	0	0	1	1	5	0	22
국내기관·단체명<기관·단체명		1	1	0	0	0	0	0	0	0	2
기타		1	0	0	0	0	0	0	0	0	1
소계		16,376	27,389	4,734	2,657	2,682	2,872	348	247	7	57,312

4.1.2 관계 현황

주제명표목표나 시소러스는 수록하고 있는 용어 간의 관계를 구성하여 제시하는데, 국립중앙도서관 주제명표목표에서 이들 용어인 주제명과 주제명 간의 관계를 표현하는 관계지시기호는 총 80개의 유형이 존재하였으며, 이들 유형을 사용하여 511,980개의 용어 사이의 구성된 관계의 빈도는 1,308,367회로 파악되었다. 관계지시기호의 유형별로 상위 30개의 구성 빈도를 확인하면, <표 8>과 같다. 상위 30개 유형을 구성하는데 사용한 빈도의 합은 총 1,304,603회로 관계지시기호의 전체 구성 빈도의 99.7%에 해당하였으며, 이는 현재 주제명표목표에서 구축한 관계지시기호 80개 유형 중 실제로 사용한 관계지시기호는 대부분 30개 이내의 유형임을 나타낸다.

전체 관계지시기호 중 최상위 빈도인 관계지시기호는 <표 8>에서처럼 연관 관계를 나타내는 RT가 406,666회 사용된 것으로 나타났으며, 이어 계층 관계 중 상위어의 BT 191,377회, 하위어의 NT 191,374회 순으로 사용되고 있음을

파악하였다. 그리고 동등 관계에서 USE, UF가 각각 약 59,100건 이상으로 나타났다. 외국어 관계를 표현하기 위해 구성된 가장 높은 빈도의 관계 지시기호는 한국어와 영어의 관계를 서로 표현하는 KEN, ENG, 그리고 일본어와의 관계를 나타내는 JPN로 나타났는데, KEN와 ENG는 각각 125,000회 이상이었으나, 이후 3순위의 JPN은 18,039회로 8% 이상 사용 차이가 나타나는 것을 확인하였다.

한편, <표 9>는 서지데이터에 100회 이상 부여되어 활용된 3,505개의 주제명들이 가지는 관계지시기호의 고유 빈도와 총 빈도를 파악한 것이다. 여기서 고유 빈도는 하나의 주제명이 특정 관계지시기호를 중복하여 다수를 갖더라도 하나로 계수하는 것을 말한다. 이들 주제명이 가진 관계지시기호의 고유 빈도는 13,726개이고 총 빈도는 68,688개이었으며 사용한 관계지시기호의 유형은 20개인 것으로 나타났다. 구체적으로 연관 관계 RT는 전체 3,505개 주제명 중에서 3,198개의 주제명이 가지며 전체 관계지시기호 68,688개 중의 33,066개(48.1%)

<표 8> 관계지시기호의 유형별 구성 빈도(상위 30개)

관계 지시기호	횟수	비율	관계 지시기호	횟수	비율	관계 지시기호	횟수	비율
RT	406,666	31.08	GER	17,780	1.36	SK	1,645	0.13
BT	191,377	14.63	KJA	15,525	1.19	KDC6	1,302	0.10
NT	191,374	14.63	KFE	9,879	0.76	ORG	1,142	0.09
KEN	125,205	9.57	FRA	9,878	0.75	KES	1,069	0.08
ENG	125,158	9.57	SNN	8,619	0.66	ESP	1,067	0.08
USE	59,189	4.52	KSN	8,618	0.66	ITA	387	0.03
UF	59,186	4.52	KCH	3,009	0.23	KIT	387	0.03
UNS	21,825	1.67	CHI	2,981	0.23	KRU	384	0.03
JPN	18,039	1.38	KJP	2,786	0.21	RUS	379	0.03
KGE	17,781	1.36	NK	1,647	0.13	LT	319	0.02

〈표 9〉 고빈도 주제명의 관계지시기호 비율(100회 이상, 3,505개)

관계 지시기호	고유 빈도 (비율)	총 빈도 (비율)	관계 지시기호	고유 빈도 (비율)	총 빈도 (비율)
RT	3,198 (23.30)	33,066 (48.14)	KDC6	134 (0.98)	152 (0.22)
NT	1,883 (13.72)	22,764 (33.14)	ORG	141 (1.03)	141 (0.21)
BT	2,578 (18.78)	3,527 (5.13)	NK	114 (0.83)	126 (0.18)
UF	1,492 (1.87)	3,136 (4.57)	ROM	51 (0.37)	99 (0.14)
ENG	2,102 (15.31)	3,093 (4.50)	ESP	58 (0.42)	62 (0.09)
GER	489 (3.56)	753 (1.10)	PT	16 (0.12)	27 (0.04)
JPN	583 (4.25)	719 (1.05)	SNN	14 (0.10)	14 (0.02)
FRA	386 (2.81)	453 (0.66)	LT	7 (0.05)	9 (0.01)
CHI	276 (2.01)	335 (0.49)	LAT	7 (0.05)	7 (0.01)
UNS	195 (1.42)	203 (0.30)	ITA	2 (0.01)	2 (0.00)

로 가장 많이 구성하고, 이어 NT(33.14%) 및 BT(5.13%)의 상·하위 관계, 동등 관계 UF(4.57%), 외국어 ENG(4.5%) 순서로 관계지시기호를 구성하고 있었다. 〈표 8〉의 비율과 비교했을 때 이들 고빈도 활용 주제명이 가지는 관계지시기호 RT와 NT가 매우 많은 것을 알 수 있다.

#### 4.1.3 계층 현황

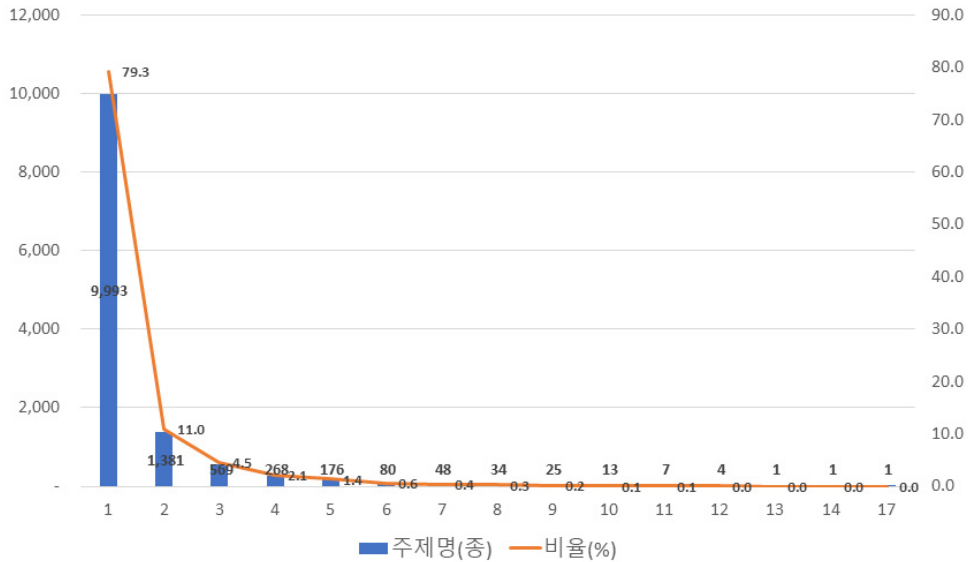
앞서 살펴본바, 주제명표목표는 여러 지시기호를 통하여 각각의 용어들과 관계를 이루어나가고 있다. 지시기호 중에서 BT와 NT는 두 용어 간의 상하의 관계를 나타내며 이는 두 주제명이 지니고 있는 개념 중 하나는 보다 넓은 또는 상위의 개념에 해당하며 다른 하나는 보다 좁은 또는 하위의 개념을 의미한다. 이러한 상·하위의 계층 구조는 이용자가 질의어를 표현할 때 계층 정도에 따라 자신의 정보요구에 적절한 주제명을 선택할 수 있으므로 통제어휘집이 가지는 중요한 측면으로 여길 수 있다. 국립중앙도서관 주제명표목표의 계층 구조를 알아보기 위해 우선어의 상·하위 관계를 확인하여 그 현황을 나

타내면, 〈표 10〉과 같다. 주제명 중 NT 관계를 지닌 최상위어는 12,602종으로 나타났고, 이들 주제명에서 NT 관계를 따라 끝까지 내려가면 도달하는 최하위에 해당하는 주제명은 143,704종으로 파악되었다. 또한 상하 계층 관계가 설정되어있지 않은 용어, 즉 주제명이 BT나 NT 관계를 가지지 않는 경우인 단독 주제명이 73,319종으로 전체 우선어(257,103종)에서 28.5%에 해당하는 것으로 확인되었다.

〈표 10〉 용어 갈래에 따른 주제명 종수

용어 갈래	종수
최상위 주제명	12,602
최하위 주제명	143,704
계층 관계 미존재 주제명	73,319
계층 연결 경로 총수	400,330

최상위 주제명으로부터 계층 구조의 심도(깊이)를 계산하면 국립중앙도서관 주제명표목표의 최대 심도는 17단계였으며, 최상위 주제명 12,602종에 따른 최대 하위 심도를 횡수와 비율로 나타내면 〈그림 2〉와 같다. BT 관계는 없이



〈그림 2〉 최상위 주제명의 계층 심도(비율)

오직 NT 관계를 하나만 가지며 더 이상의 하위 주제명이 없는 최상위 주제명을 의미하는 심도 1의 종수는 9,993개이며 그 비율은 최상위 주제명 12,602종의 79.3%에 해당하였다. 이로써 대부분의 주제명이 하위로 상세하게 혹은 세부적으로 구축되어 있지 않고 있음을 추측할 수 있다.

한편, 국립중앙도서관 주제명표목표 내에서 전체 우선어의 심도 현황과 활용 횟수에 따른 주제명에 대한 심도 현황을 분석한 결과, 〈표 11〉과 같다. 주제명표목표에 제시된 각각의 우선어에 대해 NT 관계를 이용하여 심도를 계산하고

이를 전체 우선어 수로 나누어 평균 심도를 구하여 2.92를 얻었다. 즉 이는 우선어가 평균적으로 2.92 단계의 깊이에 위치하고 있음을 의미한다. 구체적으로 활용 횟수에 따라 심도 현황을 살펴보면 최소 1회 활용된 모든 주제명의 경우 2.32, 100회 이상 고빈도로 활용된 주제명은 2.59, 여기에서 문학 분야(KDC 800)에 부여된 경우를 제외한 100회 이상 활용된 주제명은 2.52의 심도를 갖는 것으로 나타났다. 최소 한 번 이상 활용된 주제명의 경우 심도가 전체 우선어의 평균 심도(2.92)보다 얕은, 즉 매우 일반적이고 광의의

〈표 11〉 전체 우선어와 활용 주제명의 심도 현황

	전체 우선어	1회 이상 활용	100회 이상 활용	100회 이상 활용 (문학류 제외)
개수	257,102	57,312	3,505	3,259
평균	2.92	2.32	2.59	2.52
편차	3.26	2.81	2.58	2.57
최대	17	17	15	15

주제명들이 주로 활용되고 있음을 시사한다. 또한 100회 이상 부여되는 주제명은 최소 1회 이상 부여되는 주제명보다 심도가 상대적으로 깊은 듯하나 주제명표목표 전체 우선어의 평균 심도보다는 얕아서, 주로 서지 레코드에 주제명을 부여할 때 넓은 개념에 해당하는 상위의 용어를 사용하는 것으로 판단된다.

국가서지에서 1회 이상 활용된 주제명이 계층 관계가 존재하지 않는 주제명(〈표 10〉 참조), 즉 심도가 0인 주제명 중에 얼마나 포함되는지 분석하면 전체 57,312종에서 15,061종이 해당하며 그 비율은 26.3%이다. 이들 주제명의 활용 횟수는 280,935회로 전체 주제명 부여횟수 대비 13.4%에 해당하며 종수 비율인 26.3%에 비해서는 상대적으로 낮은 것을 알 수 있다. 즉 비교적 대다수의 고빈도 주제명은 여기에 해당하지 않는다고 볼 수 있다. 더불어 1회 이상 활용된 주제명 중에 0를 포함하여 심도가 1 이하인 경우는 27,496종(48.0%)이고 활용 횟수로는 582,148회(27.8%)였다. 이는 비록 활용 횟수가 상대적으로 낮지만 서지데이터에 활용된 주제명 중에서 약 절반 정도에 해당한다고 볼 수 있다. 이러한 모습은 검색 측면에서 주제명을 통해 이용자가 찾고자 하는 대상의 주제를 넓히거나 좁히는 탐색 또는 브라우징을 어렵게 할 수 있다.

이와 같이 심도가 0 또는 1과 같이 낮은 주제명은 주제명표목표에서 계층 구조의 검토 및 체계화를 통해 재정립이 필요하다. 실제 사례를 중심으로 보면 〈표 4〉의 상위 30개의 고빈도 활용 주제명에서 계층 구조에 따른 심도를 계산하면 “창작 그림책”, “임상 실습”, “교사 연수”, “발라드”, “인생훈”은 심도가 0이며 “간호학”

은 심도가 1, “고문서”, “수기(글)”은 심도 2를 갖는다. 여기서 심도 0의 주제명 대부분은 최상위 개념어라고 보긴 어려워 국립중앙도서관 주제명표목표에서 계층 관계의 정리가 필요한 것으로 보인다. 구체적으로 “창작 그림책”의 경우 이미 상위어로 볼 수 있는 “그림책[--冊]” 주제명이 주제명표목표에 존재하며, “발라드”의 경우도 마찬가지로 “대중 가요[大衆歌謠]” 주제명이 존재한다. 사실 주제명표목표가 현실 세계나 지식 세계를 반영하는 인공적인 지식 체계를 표현하는 것이어서 그 구축 목적이나 의도에 따라 다르게 가공될 수 있지만, 여기 예시로 제시된 주제명만 보더라도 현실 세계나 지식 세계의 실제 대비 주제명표목표의 계층 구조가 맞지 않거나 파편화 되어 있어 빈약하다고 볼 수 있다.

100회 이상 활용된 주제명 중에서 비교적 깊은 심도를 갖는 사례를 중심으로 살펴보면, 〈표 12〉와 같다. 예를 들어 이 표에서 “언어학[言語學]”의 최상위어는 “과학(학문)[科學]”인데 이 주제명의 NT 관계를 따라 구성된 계층 구조는 최대 심도 11단계까지 내려가게 되며 “언어학[言語學]”은 이 계층에서 세 번째에 위치하여 심도 2에 해당한다. 이러한 정보를 바탕으로 표에 나타난 사례를 살펴보면 실제 활용된 주제명의 상하관계에서 최대 심도는 9에서 14까지 나타나지만, 활용 횟수가 높아 비중 있게 사용되는 주제명들인 “수(숫자)[數]”, “예술[藝術]”, “금융 상품[金融商品]”과 같이 최상위(심도 0)이거나 “언어학[言語學]”, “통계 분석[統計分析]”, “철학(사상)[哲學]”처럼 차차상위(심도 2) 개념어에 해당하는 것을 볼 수 있다.



〈표 12〉 고빈도 활용 주제명의 심도 사례(100회 이상, 일부)

주제명	최대 심도	주제명의 심도	활용 횟수
수(숫자)[數]	10	0	272
예술[藝術]	11	0	631
언어학[言語學]	11	2	484
금융 상품[金融商品]	8	0	180
통계 분석[統計分析]	9	2	762
철학(사상)[哲學]	14	2	1,850

〈표 13〉 범주 유형에 따른 주제명의 평균 심도

주제어 범주	평균 심도	주제어 범주	평균 심도
동물<생물<주제어	11.17	국명<지명	4.05
행정구역<지명	8.68	국내기관·단체명<기관·단체명	4.00
식물<생물<주제어	7.38	통일서명	3.90
지명	6.63	주제어	2.47
상품명<주제어	5.37	기타	1.00
기념물<주제어	4.96	회의명	0.97
국보·보물<기념물<주제어	4.36	법률명<주제어	0.49

범주 유형에 따른 주제명의 심도 평균을 살펴보면 〈표 13〉과 같다. 법률명 범주의 주제명이 0.49, 일반 주제어는 2.47로 대체로 상하관계의 주제명 형성이 상대적으로 적거나 차상위까지 구성되어 있음이 나타났다. 반면 동물과 식물 등의 유강목처럼 정확한 주제 분야나 행정구역 등과 같은 세부 지리적 요소가 명확히 드러날 때는 이들 주제명의 계층에 따른 심도도 상대적으로 깊은 것으로 분석되었다.

#### 4.2 서지데이터에서의 주제명 부여 현황

서지데이터에서 하나의 레코드에는 한 개 이상의 주제명을 부여할 수 있다. 서지 레코드의 내용이 어떤 주제나 형식을 내포하느냐에 따라 그 주제나 형식을 표현하는 용어에 해당하는

하나 또는 그 이상의 주제명을 선정하게 된다.

국립중앙도서관의 서지데이터에서 개별 서지 레코드에 부여한 주제명 개수에 따른 서지 데이터 비율을 나타내면, 〈표 14〉와 같다. 하나의 서지 레코드에 하나의 주제명이 부여된 경우가 556,204건으로 전체 서지데이터의 45.6%에 해당하였다. 주제명이 2개 부여된 서지데이터가 39.3%, 3개 부여된 서지데이터는 12.8%로, 주제명이 3개 이내 부여된 서지데이터가 전체의 97.7%이었고, 이는 『국립중앙도서관 주제명표목 업무지침』(국립중앙도서관, 2021)에서 가급적 3개 이내로 주제명을 부여하도록 명시하고 있는바, 서지데이터에 부여된 주제명 개수는 적절하다고 판단된다.

이 밖에 주제명 수가 10개 이상 부여된 서지데이터의 사례는 〈표 15〉와 같이 확인할 수 있

〈표 14〉 주제명 부여 개수에 따른 서지데이터 비율

개수	1	2	3	4	5	6	7	8	9	10	11	12
서지(건)	556,204	478,901	155,606	24,670	2,798	506	105	40	22	3	10	2
비율	45.63	39.29	12.77	2.02	0.23	0.04	0.01	0.00	0.00	0.00	0.00	0.00

〈표 15〉 다수의 주제명이 부여된 서지데이터 사례(10개 이상)

서명	KDC	주제명 수
(2009 공인중개사 1차) 부동산학개론 1-3	321.32077	10
수도권의 변화	539.7	11
우리 소리의 맥脈을 찾아서	679.311	12

었는데, 표제가 『(2009 공인중개사 1차) 부동산학개론 1-3』인 도서의 실제 주제명은 “공인 중개사 시험[公認仲介士試驗]”, “부동산학 개론[不動產學概論]”, “부동산 금융[不動產金融]”, “부동산 투자[不動產投資]”, “부동산 경제[不動產經濟]”, “부동산 시장[不動產市場]”, “부동산 정책[不動產政策]”, “부동산 개발[不動產開發]”, “부동산 관리[不動產管理]”, “부동산 감정 평가[不動產鑑定評價]”이었으며, 표제가 『수도권의 변화』인 도서는 “수도권[首都圈]”, “주택[住宅]”, “주거 환경[住居環境]”, “공원 녹지[公園綠地]”, “녹지[綠地]”, “환경[環境]”, “교통[交通]”, “인구 이동[人口移動]”, “정보화[情報化]”, “교외화[郊外化]”, “산업[產業]”을 주제명으로 보유하고 있었다. 한편, 『우리 소리의 맥脈을 찾아서』가 표제인 도서는 주제명

이 모두 국내 지명이었다.

또한 이 연구는 주제의 특정성에 따라 부여된 주제명의 현황을 확인하기 위한 간접적인 방법으로 KDC의 분류 기호 길이에 해당하는 자릿수를 이용하였다. 일반적으로 KDC와 같은 십진분류법은 상위 주제를 승계하여 하위 주제로 세분하므로 분류 기호가 길어질수록 상위 개념에 종속되는 세부 주제를 표현하게 되어 있어 대상 자료의 주제가 특정성이 높아 복잡적이고 구체적일수록 긴 자릿수를 갖게 된다. 이러한 특성을 이용하여 주제의 특정성에 따른 주제명의 부여 현황을 그 개수와 심도로 분석하면 각각 〈표 16〉, 〈표 17〉과 같다. 〈표 16〉은 전체 서지 데이터 중 분류번호가 7자리 이상과 10자리 이상인 서지데이터들의 평균 주제명 부여 개수를 확인한 것으로, 전체 서지데이터에

〈표 16〉 특정성에 따른 주제명 부여 개수 현황

	전체 서지	KDC 7자리 이상	KDC 10자리 이상
개수	1,218,867	313,388	13,733
평균	1.72	1.94	1.97
편차	0.78	0.85	0.84
최대	12	11	7

서 주제명이 평균적으로 1.72개가 부여 되고 있는 것에 반하여, 7자리 이상은 1.94, 10자리 이상의 경우 1.97로 그 부여 개수가 증가하고 있는 것을 확인하였다.

〈표 17〉은 이에 따른 부여 주제명의 평균 심도의 현황으로 전체 서지데이터에 부여된 주제명의 평균 심도는 2.77이었으나, KDC 분류기호가 10자리 이상 서지에 부여된 주제명의 심도는 3.18로 나타나, KDC 분류 기호의 자릿수가 길수록 주제명 입력개수도 증가하는 경향에 맞춰, 심도 역시 깊어서 특정 영역의 주제를 분류한 서지일수록 주제명 부여에서도 연관이 있음을 보여주었다.

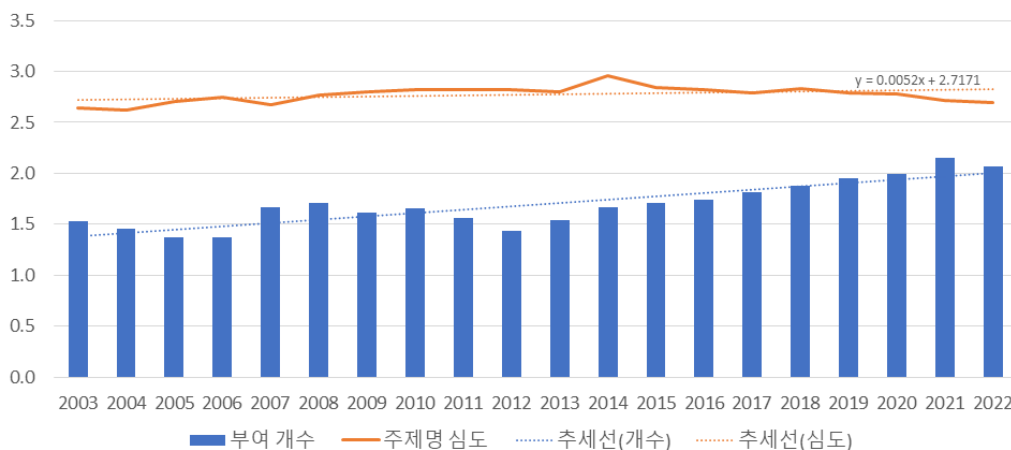
〈그림 3〉은 2003년 주제명표목표가 개발된

이래로 데이터가 입력된 연도별로 하나의 서지 레코드에 부여된 주제명표목의 평균 개수를 시계열 추이로 나타낸 것이다. 2003년 당시 주제명 부여 개수는 서지 레코드 당 1.53개였으나, 2022년에는 0.54개 증가한 2.07개로 나타나 최근으로 올수록 하나의 서지 레코드에 부여되는 주제명의 개수는 전반적으로 증가하는 추세를 보였다. 더불어 부여된 주제명의 평균 심도도 깊어지는 양상을 보이는데, 이는 추세선을 통하여 확인할 수 있다.

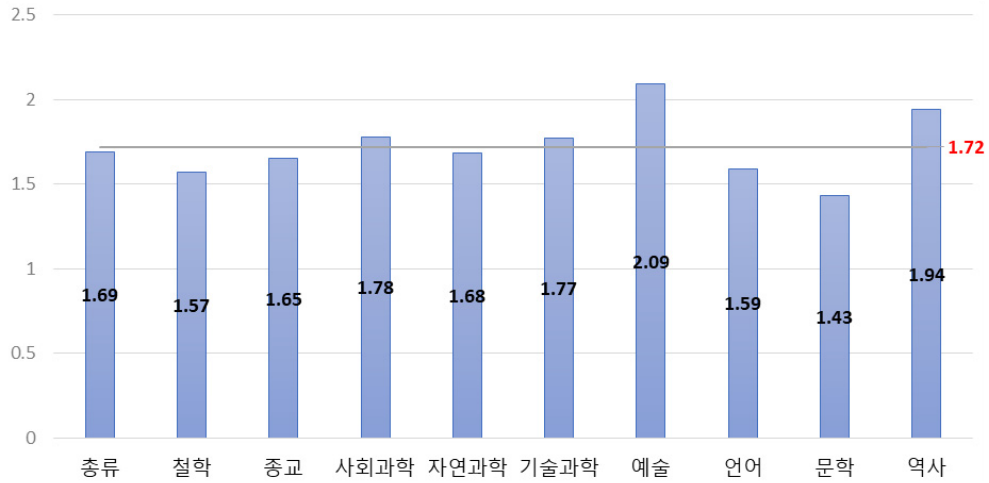
한편, KDC 주류 분야별 주제명 부여 개수의 현황을 살펴보면 〈그림 4〉와 같다. 〈표 16〉에서 전체 서지데이터 기준으로 주제명이 평균 1.72개 부여되는 것을 고려할 때, KDC 주류 분야인

〈표 17〉 특정성에 따른 주제명 부여 심도 현황

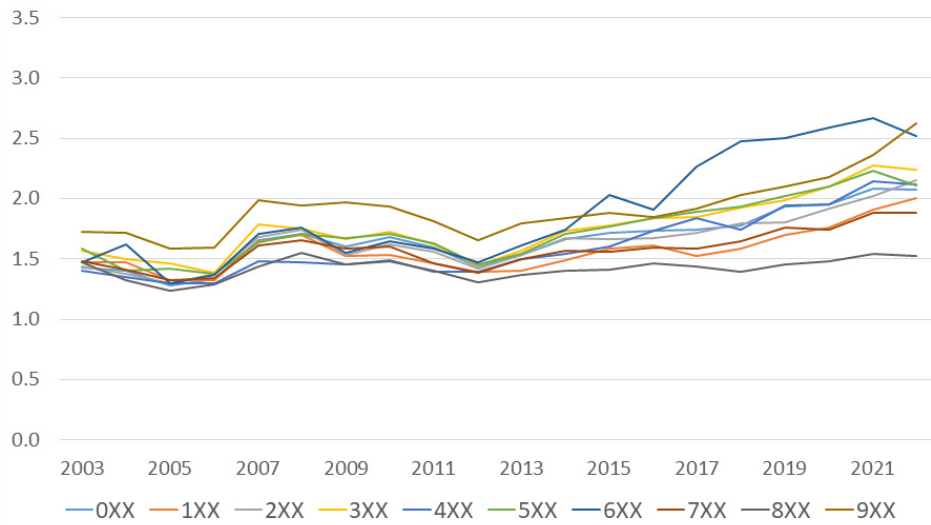
	전체 서지 부여 주제명	KDC 7자리 이상	KDC 10자리 이상
개수	2,097,682	609,191	27,021
평균	2.77	2.79	3.18
편차	2.56	2.47	2.90
최대	17	17	15



〈그림 3〉 서지 레코드 당 주제명 부여 개수의 시계열 추이(평균)



〈그림 4〉 KDC 주류별 주제명 부여 현황(평균)



〈그림 5〉 KDC 주류별 주제명 부여 개수의 시계열 추이(평균)

철학과 문학 등은 1.43과 1.57로 주제명 부여 개수가 상대적으로 적은 편이었다. 반면 예술 분야가 2.09개로 가장 높았고, 역사 분야가 다음으로 주제명을 많이 부여하고 있었다. 역사 분야의 경우 일반 주제명보다 상대적으로 지명 중심으로 주제명 부여가 많은 것으로 확인되었다.

KDC 주류별로 주제명을 부여한 개수의 평균을 연도별로 추이를 살펴보면 〈그림 5〉와 같으며, 대부분의 주류에서 평균 부여 개수가 증가하는 경향을 확인할 수 있었다. 그러나 문학의 경우 연도별 표준 편차가 0.1348로 다른 주류에 비하여 부여 개수가 연도별로 큰 편차 없

이 평이한 추이를 보이며, KDC 주류 분야 중에서 하나의 서지 레코드에 가장 적게 주제명을 부여하고 있는 경향을 나타냈으며, 최근 3년에서야 미세하게 증가하고 있었다. 한편, 가장 높은 증가세를 보이는 주류 분야는 예술로, 그 표준 편차는 0.4824였다. 이외 연도별 편차가 큰 주류 분야는 자연과학과 역사였으며, 이때 표준 편차는 각각 0.3124와 0.3014였다. 자연과학 분야의 경우, 평균 주제명 부여 개수는 1.68개로 <그림 4>에서 서지데이터의 전체 평균 부여 개수인 1.72개보다 0.04 낮은 수치였으나, 연도별 부여 개수의 증가세는 주류 분야 중에서 두 번째로 높은 것으로 나타났다.

## 5. 결론

이 연구는 국립중앙도서관에서 개발한 주제명표목표가 자체적으로 갖고 있는 다양한 구성 형식 및 체계에 대한 현황 분석과 이 주제명표목표를 사용하여 2003년부터 2022년까지 구축한 국가서지에 대하여 주제명의 부여 현황을 분석하였다. 주요 분석으로는 주제명표목표의 구성 현황으로 용어 통계, 관계 기호의 빈도, 계층 구조와 심도, 주제명의 범주 유형 및 활용 현황 등을 분석하였고, 국가서지에서 서지 레코드에 부여된 주제명표목의 평균 부여 개수를 분석하였다. 이때 부여 개수 분석은 서지 레코드가 가지는 주제의 특정성에 따른 분석과 연도별 시계열 추이에 따른 분석 그리고 KDC 주류 분야에 따른 분석을 포함하여 전체적으로 서지데이터에서 주제명표목의 부여 및 활용 현황을 파악하고자 하였다. 이를 위해 분석 대상

으로 용어 511,980개가 수록된 국립중앙도서관 주제명표목표와 함께 2003년부터 최근까지 국립중앙도서관이 주제명을 부여된 서지데이터 총 1,218,867건을 수집하고, 전처리를 수행하여 각 데이터 별로 다양한 측면에서 분석을 수행하였다. 우선 국립중앙도서관 주제명표목표의 주요 현황을 파악하기 위하여 수행한 분석의 결과는 다음과 같다.

첫째, 현재 국립중앙도서관 주제명표목표가 보유하고 있는 용어의 총수는 511,980개며, 이 중 우선어는 257,103개로 전체 주제명 개수의 50% 이상을 차지하고, 비우선어는 61,442개로 12%이며, 외국어가 169,466개로 전체 용어 중 33% 비율로 비우선어 보다 많은 것으로 나타났다.

둘째, 주제명표목표에서 서지레코드에 부여 가능한 우선어를 대상으로 활용 현황을 살펴보면 전체 중 실제 부여되어 활용된 주제명은 57,312종(22.3%)이며 나머지 77.7%에 해당하는 199,791종은 한 번도 부여 및 활용되지 않았다. 또한, 활용 횟수 100회 이상의 주제명은 3,505종에 불과하였고, 활용 주제명 중 54.7%인 31,351종의 활용횟수가 5회 미만으로 나타났다. 즉 주제명표목표에 구축된 전체 용어에 비하여 상대적으로 적은 주제명이 활용되고 있으며, 그나마 활용되는 주제명도 절반 이상이 적은 횟수로 사용되고 있는 현황을 파악할 수 있었다. 이를 방증하는 것으로 고빈도의 활용 주제명 5,312종(9.3%)의 활용 누적 횟수가 전체의 80% 이상으로 나타나 대체로 매우 적은 수의 주제명을 활용하여 서지데이터에 부여하고 있음을 알 수 있다.

셋째, 상위 활용 주제명 300종 중에서 갈래

없는 일반 주제어에 해당하는 주제명이 199종으로 다수를 차지했지만, LCSH의 세목 성격을 띠는 형식이나 지명을 나타내는 주제명이 111종으로 일반 주제명보다 88종이 적음에도 불구하고 이들 주제명의 활용 횟수는 300종의 총 활용횟수 중 51.0%를 차지하면서, 형식이나 지명과 같은 세목 성격의 주제명이 다소 많이 활용되고 있는 것을 파악하였다.

넷째, 범주 유형 측면에서 범주 갈래 있는 주제어보다 갈래가 없는 일반 주제어에 해당하는 주제명이 232,524종으로 90.4%의 비율을 보여 주었으며 그 다음으로 지명(8,836종, 3.4%), 동물<생물<주제어(6,411종, 2.5%), 식물<생물<주제어(4,173종, 1.6%) 순이었다.

다섯째, 관계지시기호 측면에서 RT 관계가 31.1% 이상의 최상위 빈도 기호로 나타났으며, 이어 계층 구조를 형성하는 BT와 NT 관계가 각각 14.6% 정도로 등장하였다. 그 다음으로 외국어와의 대응 관계에 대한 지시기호의 비율이 높았으며, 그 순서는 KEN, ENG, JPN 순이었다. 동등 관계를 나타내는 USE와 UF 기호는 각각 4.5% 정도의 비율을 보였다.

여섯째, 관계 기호를 중심으로 개별 주제명 간의 계층 관계를 확인했을 때, 최상위어에 해당하는 주제명은 12,602종이고 최하위어의 주제명은 143,704종이었다. 주제명표목표에서 구축된 최상위어 기준 최하위 계층까지의 최대 심도는 17단계이었다. 대체로 주제명표목표의 전체 우선어 257,103종의 평균 심도가 2.92이며 갈래가 없는 일반 주제어 범주에 해당하는 주제명의 평균 심도는 2.47로 주제명표목표의 전체 계층 구조에서 평균적인 주제명의 심도는 2-3단계 내외라고 볼 수 있다. 또한 1회 이

상 활용된 주제명의 평균 심도가 2.32이며 전체 서지데이터에 부여된 주제명의 평균 심도 역시 2.77 단계에 해당하므로 서지데이터에 부여된 측면에서 주제명의 심도는 2-3단계 내외에서 많이 형성되는 것을 확인하였다. 이는 다소 넓은 의미의 개념이나 주제에 해당하는 주제명을 서지에 부여하고 있음을 의미한다고 볼 수 있다. 이와 함께 주제명표목표의 계층 구조 측면에서 심도가 없는 단독 주제명이 우선어 전체의 28.5%에 이르고 1회 이상 부여된 주제명의 48.0%가 심도가 없거나 1단계에 해당하는 점을 비추어볼 때, 이용자의 탐색이나 브라우징을 지원하기 위해 빈약하거나 파편화되어 있는 주제명 간의 계층 관계를 재정립할 필요가 있으며 보다 다양한 주제명이 부여되어야 한다.

서지데이터를 중심으로 분석된 주제명의 부여 현황을 살펴보면 첫째, 주제명 부여 횟수는 평균 1.72회로 업무지침에서 제시된 3개 이내 부여라는 권고사항에 부합하였으며, 다만 대상 자료나 콘텐츠의 주제가 특정적일수록 그 자료에 부여된 주제명 개수가 증가하고, 심도 역시 깊어지는 경향을 보였다. 둘째, KDC 주류 분야에 따른 주제명 부여 개수를 보면, 예술이 하나의 서지 레코드 당 2.09개로 가장 많았고 문학이 1.43회로 가장 적었는데, 1.72회가 평균이라는 점을 감안 할 때 주제 분야에 따라 주제명 부여 개수의 편차가 다소 존재하는 것으로 나타났다. 셋째, 이에 따라 서지데이터 입력 연도에 따른 주제명 부여 현황을 시계열 측면에서 확인하면 주제명표목표 개발 이후 최근일수록 서지 레코드 당 주제명 부여 개수가 증가하고 있었으며, 부여 주제명의 심도도 깊어지는 양

상을 보인다. 즉 최근 작성된 서지데이터일수록 과거보다 상대적으로 자료의 주제를 잘 표현하도록 주제명을 부여하고 있음을 시사하였다. KDC 주류 영역별로는 주제명 부여 개수의 연도별 추이를 보면 예술이 가장 높은 증가세를 나타냈으며, 문학은 가장 낮고 평이한 모습을 보였다.

본 연구의 결과에 비춰봤을 때, 다음과 같이 국립중앙도서관 주제명표목표의 개선을 고려할 필요가 있다. 우선 국립중앙도서관의 주제명 부여 업무지침에 따라 올바르게 주제명이 부여되고 있기는 하나, 실제로 쓰인 주제명은 전체 우선어 주제명의 22.3%에 불과하고 고빈도 주제명 위주로 다소 한정적으로 활용하고 있다는 점에서, 국립중앙도서관 주제명표목표가 수록한 용어의 범위(coverage) 측면에서 그 규모가 적절한지 평가할 필요가 있다. 주제명 간의 계층 관계에 따른 심도가 없거나 매우 얇은 주제명이 우선어의 상당수에 해당하고 비교적 많이 부여되고 있기에, 과연 국립중앙도서관 주제명표목표에 수록된 주제명 사이의 계층 관계와 연관 관계가 현실 세계 또는 지식 세계를 적절히 표현하고 구조를 적합하게 형성하고 있는지 평가할 필요가 있다.

더불어 형식 관련 주제명 부여에 대한 언급은 이미 선행연구(노지현, 2015)에서도 언급되었던 바, 여전히 그 모습이 나타나고 있기에 LCSH의 형식세목 및 지리세목 또는 MeSH(Medical

Subject Headings)의 'Publication Type'처럼 패킷 개념(최윤경, 정연경, 2015) 등을 도입하여 자료의 실제 주제를 나타내는 주제명을 부여하고 추가로 세목 내지 패킷을 지정하여 접근점을 부여하는 방향으로 개선이 필요하다.

한편, 국립중앙도서관 주제명표목표의 개선을 통해 주제명의 계층 구조 및 심도를 재정비하는 것은 서지데이터의 접근점을 보다 특정성 있게 부여하게 되므로 정보 검색에서도 성능 향상을 위한 새로운 가능성을 제시할 수 있을 것으로 생각되며, 이는 현재 국가서지에 부여된 실제 주제명의 계층 관계를 세밀하게 분석하는 작업이 필요하므로 본 연구 연장선으로 후속 연구과제로 진행해 볼 수 있을 것이다.

2003년부터 2022년까지 구축된 실제 국립중앙도서관의 주제명표목표와 국가서지 데이터를 활용하였으며, 그 결과로 주제명표목표의 자체 현황을 파악할 수 있기에 보다 긍정적인 방향으로 개선하기 위한 통찰력을 제공할 수 있을 것으로 생각된다. 또한 이 연구가 목록 데이터에서의 주제 접근을 통한 검색이나 FRISAD에서 제시한 이용자 과업 중 네 번째인 탐험의 적용 등 전거를 기반한 다양한 후속 연구에서 포석의 역할로 활용될 수 있을 뿐만 아니라, 주제명표목표의 개선과 발전에 이바지하여 이용자에게도 도서관 검색이 더 효율적이고 만족할만한 결과가 도출할 수 있도록 기대한다.

## 참 고 문 헌

- 국립중앙도서관 (2021). 국립중앙도서관 주제명표목 업무지침. 서울: 국립중앙도서관.
- 김정현 (2010). 미국의회도서관 주제명표목표의 한·중·일 관련 주제명표목의 변천과정 비교 분석. 한국도서관·정보학회지, 41(2), 147-169. <http://doi.org/10.16981/kliss.41.2.201006.147>
- 김정현, 문지현 (2009). LCSH 한국관련 주제명표목의 변천과정에 관한 연구. 한국도서관·정보학회지, 40(3), 49-68. <http://doi.org/10.16981/kliss.40.3.200909.49>
- 김해인, 정연경 (2022). 구술 기록의 활용을 위한 패킷 기반 주제명표목 개발에 관한 연구: 해방촌 닷 사업 구술 기록을 대상으로. 한국기록관리학회지, 22(4), 67-85. <http://doi.org/10.14404/JKSARM.2022.22.4.067>
- 노지현 (2015). 주제명 데이터를 통해 본 현행 목록의 품질과 과제. 한국도서관·정보학회지, 46(4), 379-402. <http://doi.org/10.16981/kliss.46.4.201512.379>
- 문지현, 김정현 (2011). 『國立國會圖書館件名標目表』의 한·중 관련 주제명표목에 대한 비교 분석. 한국도서관·정보학회지, 42(1), 247-267. <http://doi.org/10.16981/kliss.42.1.201103.247>
- 박지영 (2011). 서지레코드의 주제전거를 위한 FRISAD 모형 연구. 정보관리연구, 42(1), 113-135.
- 백지원, 정연경 (2014). 국립중앙도서관 주제명표목표 검색 시스템 개선 방안에 관한 연구. 정보관리학회지, 31(1), 31-51. <http://doi.org/10.3743/KOSIM.2014.31.1.031>
- 여지숙, Kiduk Yang, 이토히로코, 이해경 (2022). 한인디아스포라 관련 주제명표목 개선 방안 연구: 국립중앙도서관 주제명표목표의 한인 관련 용어를 중심으로. 한국도서관·정보학회지, 53(1), 103-124. <http://doi.org/10.16981/kliss.53.1.202203.103>
- 정연경 (2010). 정보의 접근점으로서 미국의회도서관 주제명표목에 관한 연구. 한국문헌정보학회지, 44(2), 51-72. <http://doi.org/10.4275/KSLIS.2010.44.2.051>
- 정옥경 (2002). 표목통제를 위한 전거레코드 작성방안에 대한 연구. 한국문헌정보학회지, 36(1), 145-163.
- 최윤경, 정연경 (2014). 국립중앙도서관 주제명표목표의 고품질화 방안에 관한 연구. 한국문헌정보학회지, 48(1), 75-95. <http://doi.org/10.4275/KSLIS.2014.48.1.075>
- 최윤경, 정연경 (2015). 한국 주제명표목의 패킷 유형 개발에 관한 연구. 한국문헌정보학회지, 49(4), 179-201. <http://doi.org/10.4275/KSLIS.2015.49.4.179>
- 한국도서관협회 (2010). 문헌정보학용어사전. 서울: 한국도서관협회.
- Drabenstott, K. M. & Vizine-Goetz, D. (1994). Using subject headings for online retrieval: Theory, practice, and potential. San Diego: Academic Press.
- Frost, C. & Dede, B. (1988). Subject heading compatibility between LCSH and catalog files of a large research library: A suggested model for analysis. Information Technology and



- Libraries, 7(3), 288-299.
- IFLA Study Group on the Functional Requirements for Bibliographic Records (1998). Functional Requirements for Bibliographic Records: Final Report. München: K. G. Saur.
- Julien, C. A., Tirilly, P., Dinneen, J. D., & Guastavino, C. (2013). Reducing subject tree browsing complexity. *Journal of the American Society for Information Science and Technology*, 64(11), 2201-2223. <https://doi.org/10.1002/asi.22915>
- Julien, C. A., Tirilly, P., Leide, J. E., & Guastavino, C. (2012). Constructing a true LCSH tree of a science and engineering collection. *Journal of the American Society for Information Science and Technology*, 63(12), 2405-2418. <https://doi.org/10.1002/asi.22749>
- Library of Congress (2022). *Library of Congress Subject Headings*. Washington D.C.: Library of Congress.
- O'Neill, E. & Aluri, R. (1981). Library of Congress subject heading patterns in OCLC monographic records. *Library Resources and Technical Services*, 25, 63-80.
- Yi, K. & Chan, L. M. (2010). Revisiting the syntactical and structural analysis of Library of Congress Subject Headings for the digital environment. *Journal of the American Society for Information Science and Technology*, 61(4), 677-687. <https://doi.org/10.1002/asi.21295>

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

- Baek, Ji-Won & Chung, Yeon-Kyoung (2014). A study on improving access & retrieval system of the national library of Korea subject headings. *Journal of the Korean Society for Information Management*, 31(1), 31-51. <http://doi.org/10.3743/KOSIM.2014.31.1.031>
- Choi, Yoon Kyung & Chung, Yeon Kyoung (2015). Study on developing facets for subject headings in Korea. *Journal of the Korean Society for Library and Information Science*, 49(4), 179-201. <http://doi.org/10.4275/KSLIS.2015.49.4.179>
- Choi, Yoon Kyung & Chung, Yeon-Kyoung (2014). A study on improvements for high quality in national library of Korea subject headings list. *Journal of the Korean Society for Library and Information Science*, 48(1), 75-95. <http://doi.org/10.4275/KSLIS.2014.48.1.075>
- Chung Ok-Kyung (2002). A study on the guideline of authority record for heading control. *Journal of the Korean Society for Library and Information Science*, 36(1), 145-163.
- Chung, Yeon-Kyoung (2010). A study on the library of congress's subject headings as an information access point. *Journal of the Korean Society for Library and Information Science*, 44(2),

- 51-72. <http://doi.org/10.4275/KSLIS.2010.44.2.051>
- Kim, Haein & Chung, Yeon-Kyoung (2022). A study on developing facet-based subject headings of oral history records use: using oral history records of knit business in Haebangchon. *Journal of Korean Society of Archives and Records Management*, 22(4), 67-85. <http://doi.org/10.14404/JKSARM.2022.22.4.067>
- Kim, Jeong-Hyen & Mun, Ji-Hyun (2009). A study on the development of topical headings related to Korea in LCSH. *Journal of Korean Library and Information Science Society*, 40(3), 49-68. <http://doi.org/10.16981/kliss.40.3.200909.49>
- Kim, Jeong-Hyen (2010). A comparative study of subject headings related to Korea, China, and Japan in the LCSH. *Journal of Korean Library and Information Science Society*, 41(2), 147-169. <http://doi.org/10.16981/kliss.41.2.201006.147>
- Korean Library Association (2010). *The Glossary of Library and Information Science*. Seoul: Korean Library Association.
- Mun, Ji-Hyun & Kim, Jeong-Hyen (2011). A comparative analysis of subject headings related to Korea and China in the NDLSH. *Journal of Korean Library and Information Science Society*, 42(1), 247-267. <http://doi.org/10.16981/kliss.42.1.201103.247>
- National Library of Korea (2021). *National Library of Korea subject heading work guidelines*. Seoul: National Library of Korea.
- Park, Ziyong (2011). A study on the functional requirement of subject authority data for subject authority control. *Journal of Information Science Theory and Practice*, 42(1), 113-135.
- Rho, Jee-Hyun (2015). A study on the quality of subject data in library catalogs. *Journal of Korean Library and Information Science Society*, 46(4), 379-402. <http://doi.org/10.16981/kliss.46.4.201512.379>
- Yeo, Ji-Suk, Yang, Kiduk, ITO, HIROKO, & Lee, HyeKyung (2022). A study on the enhancement of Korean diaspora-related subject headings: focusing on Korean-related terminology in the national library of Korea subject headings. *Journal of Korean Library and Information Science Society*, 53(1), 103-124. <http://doi.org/10.16981/kliss.53.1.202203.103>