

주제명 데이터를 통해 본 현행 목록의 품질과 과제*

A Study on the Quality of Subject Data in Library Catalogs

노 지 현(Jee-Hyun Rho)**

〈목 차〉

I. 서론	1. 품질 분석의 방법
II. 주제의 개념과 속성	2. 품질 분석의 결과
1. 주제의 개념	IV. 현행 주제명 데이터의 한계와 과제
2. 주제의 속성	V. 결론
III. 주제명 데이터의 품질	

초 록

이 연구에서는 지금까지 국내 자료조직 연구에서 거의 다루어지지 않았던 주제명 데이터의 '품질'에 관해 논의하였다. 구체적으로, 우리나라 대표도서관으로서 주제명표목표의 개발과 이를 이용한 주제명 데이터의 구축을 주도해 온 국립중앙도서관의 실체를 있는 그대로 드러내 보이면서 과연 현재와 같은 품질을 유지하는 것이 바람직한지에 대해 폭넓게 논의해 보고자 하였다. 이러한 목적을 위해 이 연구에서는 먼저, 주제명 데이터에 반영되어야 할 '주제'의 의미와 속성에 대해 살펴보고, 이어 국립중앙도서관에서 구축한 주제명 데이터의 품질을 절대적 품질과 상대적 품질로 나누어 분석하였다. 절대적 품질 분석을 위해서는 '대중자료'와 '학술자료' 중 샘플자료를 임의로 추출하여 이들 자료에 반영된 주제명 데이터를 분석하였으며, 상대적 품질 분석을 위해서는 미의회도서관(이하 LC)과 국내의 A대학도서관에서 구축한 주제명 데이터와의 비교를 시도하였다.

키워드: 주제, 주제명, 주제명표목, 주제목록법, 도서관목록, OPAC, 편목, 목록의 품질

ABSTRACT

This study intends to make a comprehensive inquiry into the quality of subject data in library catalogs. On the basis of National Library Subject Headings (NLSH) developed in 2002, National Library of Korea (NLK) has input the subject data into cataloging records since 2003. However a serious question could arise regarding whether the data are appropriate and desirable as a subject of the work. Under these circumstances, this study examined as follows: First, the meanings and attributes of the 'subject' were examined with a comprehensive literature survey. Second, an experimental analysis was carried out to measure the quality of the subject data. Sample records were selected from NLK, LC and A university library in Korea, and absolute and relative quality were compared. Finally, the substantial and realistic strategies for improving the quality of subject data were discussed.

Keywords: Subject, Aboutness, Subject data, Subject headings, Subject cataloging, Library catalogs, Cataloging, OPAC, the Quality of catalogs

* 이 논문은 부산대학교 자유과제 학술연구비(2년)에 의하여 연구되었음.

** 부산대학교 문헌정보학과 교수(jhrho@pusan.ac.kr)

•논문접수: 2015년 11월 20일 •최초심사: 2015년 11월 26일 •게재확정: 2015년 12월 14일

•한국도서관·정보학회지 46(4), 379-402, 2015. [<http://dx.doi.org/10.16981/kliss.46.201512.379>]

I. 서론

자료검색의 방식은 크게 두 가지로 구분할 수 있다. 자료에 대한 구체적인 단서를 가지고 검색을 시도하는 경우(known-item searching)와 막연한 상태에서 주제에 대한 포괄적인 검색을 시도하는 경우(subject searching)가 그것이다. 도서관의 편목 업무를 기술편목(descriptive cataloging)과 주제편목(subject cataloging)으로 구분하는 이유도 바로 여기에 있다. 이러한 관점에서 볼 때 현재 우리 도서관에서 제공하는 OPAC은 커다란 한계를 안고 있다. 저자, 표제, 발행사, 심지어 ISBN과 같은 결정적인 단서 없이는 ‘주제’를 이용하여 원하는 자료에 접근하는 것이 거의 불가능하기 때문이다. 이러한 문제제기는 더 이상 새삼스럽지 않다. 우리의 OPAC에서 주제 검색의 한계는 지난 20여 년 간 끊임없이 지적되어 왔음에도 불구하고 여전히 해결하지 못하고 있는 고질적인 문제로 남아 있다.

그동안 자료조직 연구자들이 진단해 온 문제의 가장 큰 원인은 ‘주제명표목표의 부재’에 있었다. 주제 검색에 사용되는 데이터는 “통제된 용어로 표현될 때 효과적”인데, 우리의 경우 “용어 통제를 위한 주제명표목표가 정비되지 않아 주제명 데이터의 구축이 불가능하다”는 것이 주요 논점이었다. 그러다보니 해결책 또한 주제명표목표의 ‘개발’로 귀결되었다. “주제명표목표의 개발이 하루빨리 이루어져야 할 것”이라 채근하였고, 주제명표목표가 개발되면 주제 검색의 문제도 자연스럽게 해결되는 것처럼 문제의 본질을 호도하였다. 그러나 지난 2002년에 모든 주제를 포괄하는 시소러스형 주제명표목표를 개발하고, 이어 2013년부터 주제명표목표의 고품질화 작업을 추진하고 있는 국립중앙도서관에서도 주제 검색의 문제는 크게 개선되지 않고 있다.

우리가 이 문제에 대한 유일한 해결책으로 여겨온 주제명표목표의 개발이 근본적인 해결책이 될 수 없는 이유는 주제편목의 과정을 소개하는 개론서를 보아도 쉽게 알 수 있다. 개론서에서는 주제편목을 “첫째, 자료가 무엇에 관한 것인지를 파악하고, 둘째, 파악한 주제를 통제어휘나 분류기호로 변환하는 2단계로 이루어지며”(Taylor 2009, 305), “여기서 중요한 포인트는 그 출발점인 동시에 요체에 해당하는 첫 번째 단계”(윤희운 2015, 71)라고 강조하고 있다. 필자의 강의 경험에 비추어보아도, 주제편목을 학습하는 과정에서 학생들이 느끼는 실질적인 어려움은 주제명표목표의 활용이나 용어의 통제 과정이 아니라 자료의 주제를 파악하는 첫 번째 단계에 있음이 분명하게 드러난다. 이러한 상황은 현장의 사서들도 크게 다르지 않다. 그들이 자료의 핵심 주제라고 판단하여 제시한 주제명 데이터에는 주제에 대한 ‘그릇된 판단’이나 ‘부실한 선택’의 흔적이 적지 않게 포함되어 있기 때문이다.

상황이 이러한데도 불구하고, 자료의 주제를 파악하는 방법이나 그 과정에서 부딪히는 다양한 어려움을 해소하는데 도움이 되는 연구는 좀처럼 찾아보기 어렵다. 국내 선행연구의 대

부분은 주제명표목표의 개발이나 주제명표목표에 수록된 주제명에 대한 분석에 집중되어 있고, 자료조직 교육을 위한 대부분의 교재에서도 주제를 파악하는데 도움이 되는 ‘단서’만 순차적으로 나열하고 있을 뿐이다. 최근 국립중앙도서관에서 실무에 참조할 수 있는 매뉴얼을 발간하였지만(국립중앙도서관 2014), 이 또한 새로운 주제명의 생성이나 기존 주제명의 수정·변경 등에 초점이 맞추어져 있어 자료의 주제를 파악하고 그 결과를 주제명으로 표현하는데 참조하기에는 여러모로 부족한 상황이다. 주제명표목표의 개발과 개량이 실현되더라도 주제명 데이터의 품질이 획기적으로 개선되고, 그로 인해 주제 검색에 대한 이용자 만족도가 제고될 수 있는지에 대해 근본적인 의문이 드는 까닭이 바로 여기에 있다.

이러한 배경에서, 이 연구에서는 국내 자료조직 연구에서 거의 다루어지지 않았던 주제명 데이터의 ‘품질’에 대해 논의해 보고자 한다. 구체적으로, 우리나라 대표도서관으로서 주제명표목표의 개발과 이를 이용한 주제명 데이터의 구축을 주도해 온 국립중앙도서관의 실체를 있는 그대로 드러내 보이면서 과연 현재와 같은 품질을 유지하는 것이 바람직한지에 대해 폭넓게 논의해 보고자 한다. 이러한 목적을 위해 이 연구에서는 먼저, 주제명 데이터에 반영되어야 할 ‘주제’의 의미와 속성에 대해 살펴보았다. 이어, 국립중앙도서관에서 구축한 주제명 데이터의 품질을 절대적 품질과 상대적 품질로 나누어 분석하였다. 절대적 품질 분석을 위해서는 ‘대중자료’와 ‘학술자료’에 반영된 주제명 데이터를 분석하였으며, 상대적 품질 분석을 위해서는 미의회도서관(이하 LC)과 국내의 A대학도서관에서 구축한 주제명 데이터와의 비교를 시도하였다. 마지막으로, 이상의 연구결과를 토대로 주제명 데이터의 품질을 개선하기 위해 우리 도서관계가 고민해야 할 과제와 추진 전략에 대해 논의하였다.

Ⅱ. 주제의 개념과 속성

1. 주제의 개념

편목 대상인 자료가 “무엇에 관한 것”인지를 파악하는 작업이 주제편목의 핵심이다. 여기서 “무엇에 관한 것”은 결국 ‘주제’를 의미하는데, 이 용어가 가지는 의미는 상당히 추상적이고 모호하다. 따라서 자료의 핵심 주제를 파악하기 위해서는 주제가 가진 의미를 명료하게 정리해 볼 필요가 있다. 먼저, 사전적 의미를 살펴보면 주제는 ① 대화나 연구 따위에서 중심이 되는 문제, ② 예술 작품에서 지은이가 나타내고자 하는 기본적인 사상, ③ 주된 제목 등을 의미한다(국립국어원 2014). 여기서, ‘③ 주된 제목’은 ①이나 ②를 함축적으로 나타내거나 이와 완전히 무관한 것일 수 있기 때문에 논외로 하면, 주제는 ‘자료에서 다루는 중심 문제’

혹은 ‘저자가 전달하고자 하는 사상이나 주장’이라 할 수 있다. 주제편목에서 사용하는 ‘주제’의 개념도 크게 다르지 않다. 문제는 이 개념이 너무 포괄적이어서 다양한 해석이 존재한다는 데 있다.

가령, 분류기호의 생성을 위해 해당 자료의 주제를 파악해야 하는 DDC와 KDC에서는 “저자가 무엇을 어떤 관점에서 기술하였는지” 즉, ‘저자의 의도’가 곧 주제라고 밝히고 있다. 반면, 주제명목록에서는 전통적으로 ‘자료 전체를 포괄하는’ 또는 ‘가장 중요한’ 개념이나 대상, 이름 등을 주제로 간주하여 왔다. 자료의 주제를 결정하는 방법과 관련하여 폭넓은 이론적 접근을 시도했던 Patrick Wilson은 주제란 객관적으로 드러나는 대상이 아니라 개인이 주관적으로 인지하는 것이기 때문에 한마디로 정의하기 어려우며,¹⁾ ① 자료의 생산 목적, ② 자료에 두드러지게 나타나는 중심 사상이나 사건, ③ 자료에 빈번하게 등장하는 개념, ④ 자료 전체를 포괄하는 개념 중 하나 또는 이들을 종합한 것이 결과적으로 주제라고 정리하였다 (Taylor 2009, 315-316 재인용).

미국과 영국을 중심으로 하는 서구의 문헌정보학계에서는 1970년대부터 ‘자료의 주제를 무엇으로 볼 것인가’에 대해 폭넓은 논의를 전개한 것으로 보인다. 이러한 논의의 과정에서, 초기의 이론적이고 관념적인 논의는 점차 실무에의 적용을 위한 실용적 논의로 변모하였다 (Hjørland 1992, 2001). 실용적 논의에서의 핵심은 자료의 핵심 주제를 추출하여 논리적으로 표현하기 위한 객관적인 방법과 이러한 방법의 일관된 적용에 있었다. 이러한 이유로 “자료에 직접적 또는 간접적으로 드러나는 저자의 의도”나 “자료에 대한 독자들의 개인적이고 주관적 이해”라는 의미를 내포하고 있는 ‘subject’ 대신에, “무엇에 관한 것”을 직접적으로 지칭하는 ‘aboutness’를 보다 보편적으로 사용하였다 (Albrechtsen 1993, 220).²⁾

aboutness의 개념에 대한 문헌정보학 연구자들의 관심이 증가하면서 이 용어를 정리하고 개념화하고자 하는 다양한 노력도 나타났다. 먼저, 문헌정보학계에서 aboutness라는 용어를 처음 사용한 Fairthorn은 aboutness를 외연적 의미와 내포적 의미로 구분하고, 이를 각각 ‘자료에 객관적으로 드러나는 것’과 ‘자료에 대해 이용자들이 특히 관심을 가질 만한 것’이라 정리하였다 (Fairthorn 1969, 79). Lancaster도 ‘자료가 가진 고유한 주제’와 ‘이용자가 자

1) Wilson은 이처럼 소유(having)나 실체(being)의 개념이 아니라 개개인 이 자료에 대해 주관적으로 이해하는 것이기 때문에 주제에 대한 명확한 정의를 내리는 것이 불가능하다고 주장하였다. Hjørland는 Wilson의 이러한 입장을 ‘주관적 관념론’(subjective idealism)으로 간주하고, 이와 상반되는 입장으로 ‘객관적 관념론’(objective idealism) 혹은 ‘유물론’(materialism)을 소개한 바 있다 (Hjørland 1992, 177).

2) 이 용어는 1969년 Fairthorn이 처음 사용하였으며, 관련 연구나 편목업무에서 보편적으로 사용되기 시작한 것은 1970년대 후반부터였다. 『문헌정보학용어사전』에서는 aboutness를 ‘주제관련성’으로 번역하고, “문헌의 본문 내에서 암묵적으로 또는 명시적으로 언급된 의미나 주제를 총칭할 때 쓰는 용어”라고 풀이하고 있다 (한국도서관협회 2010, 311). 본문에 언급한 것처럼 subject을 대신하는 용어로 등장한 신조어인 aboutness를 단순히 ‘주제관련성’으로 번역하는 것은 한계가 있다고 판단하여, 이 연구에서는 번역어 대신 원어를 그대로 사용하였다.

료를 이용하고자 하는 목적'으로 구분하여 주제검색에 사용될 주제어(색인어)의 특성에 대해 논의하였다(Lancaster 1991, 36). Wellisch 역시 aboutness는 '자료가 무엇에 관한 것인지' 뿐만 아니라 '누가 이 자료를 이용할 것인지' 혹은 '이 자료는 누구를 대상으로 하는지'의 의미를 내포하는 개념이라고 강하게 주장하였다(Wellish 1988, 3).

한편, Maron은 정보검색에 직접적으로 활용할 목적으로 aboutness의 개념에 대한 보다 실증적 분석을 시도하였다. 분석결과, 자료에 객관적으로 드러나는 주제와 색인자나 목록자가 주관적으로 판단하는 주제, 그리고 이용자가 검색에 사용하는 주제 간에는 상당한 차이가 있으며, 이러한 차이를 고려할 때 주제는 이용자의 검색 측면에 맞추어져야 한다고 주장하였다(Rondeau 2014, 20 재인용). 주제편목의 관점에서 주제의 의미에 대한 냉철한 분석을 시도했던 Hjørland도 주제에 대한 저자, 이용자, 사서 간의 인식의 차이에 주목하였다(Hjørland 1992, 174-175). 이렇듯 aboutness의 개념을 둘러싼 그간의 논의를 살펴보면, 우리가 목록레코드에 반영해야 할 주제는 자료에 대한 이용자의 요구나 관심이라는 결론에 이르게 된다. 이러한 논리는 저작과 저작의 주제 간의 관계를 표현한 '서지레코드의 기능상의 요건'(FRBR)과 그 후속으로 발표된 '주제전거 데이터의 기능 요건'(FRSAD)에도 고스란히 반영되었다. 이들 개념모형에서는 "이용자의 관심사가 되는 핵심 대상"을 개체로 간주하고, 자료의 주제에 해당하는 개체를 '개념', '대상', '사건', '장소'로 정의하였다(IFLA 1997, IFLA 2011).

이상의 내용을 종합해 보면, 우리가 단순히 '주제'라고 부르는 용어에 상당히 복합적인 의미가 내포되어 있음을 알 수 있다. 특히, 서구의 도서관계에서 주제에 대한 논의가 주로 aboutness를 중심으로 이루어졌고, 이러한 aboutness의 개념에 자료에 대한 이용자의 관심이 강하게 내포되어 있다는 점을 고려하면, 주제란 자료에 표현된 저자의 의도라기보다는 자료에서 객관적으로 다루고 있는 무엇, 그 중에서도 '이용자가 특히 관심을 가질 만한 무엇'이라 결론내릴 수 있다. 그렇다면, '무엇'에 해당하는 구체적인 실체는 무엇일까? 관련 연구를 비롯하여 주제명 데이터의 구축이나 활용 사례 등을 살펴보면 '무엇'의 실체가 보다 분명하게 드러날 것으로 판단된다.

2. 주제의 속성

'자료에 취급된 내용 중 이용자가 특히 관심을 가질 만한 것'을 주제로 간주할 때, 주제에 해당하는 속성은 상당히 다양할 수 있다. 이 속성은 자료로부터 핵심 개념을 추출하는데 중요한 근거가 되므로 선행연구와 사례를 통해 보다 구체적으로 규명해 볼 필요가 있다.

Langridge(1989)는 ① 해당 자료가 무엇인지(what is it?), ② 누구를 위한 것인지(what

is it for?), ③ 무엇에 관한 것인지(what is it about?)와 같은 세 가지 질문에 대한 답이 결과적으로 자료의 주제라고 정리하였다. 그리고 이 질문에 대한 답을 유형화하면 ① 자료와 관련된 학문분야나 자료의 형식·장르·유형, ② 자료의 발간 목적이나 이용대상, ③ 자료에 취급된 주요 개념이나 대상 등이 되는데, 이들이 바로 자료의 주제적 속성이라 주장하였다. Langridge가 정리한 이들 속성은 20년이 지난 현재에도 중요하게 다루어지고 있다. 특히, 자료의 내용과 직접적으로 관련이 없는 ‘사전’, ‘전기’, ‘일기’, ‘단편소설’ 등과 같은 자료의 물리적 특성이나 장르·형식, 그리고 ‘초등학생’과 같은 이용 대상이나 지적 수준을 주제의 범주에 포함시켰다는 점에서 Langridge의 연구는 여전히 중요하게 평가되고 있다.³⁾

주제의 속성은 주제명표목표나 MARC 포맷에도 반영되어 있다. 전 세계적으로 가장 널리 사용되는 주제명표목표인 LCSH는 2015년 현재 357,090개의 주제명을 수록하고 있으며, 매년 약 5,000개의 용어를 추가 또는 갱신하고 있다(Library of Congress 2015). LCSH에 수록된 주제명은 단일개념으로 사용할 수도 있고, 주표목에 세목을 결합한 복합개념으로 사용할 수도 있다. 이 가운데, 주제명으로 바로 사용할 수 있는 주표목은 ① 자료에 취급된 개념이나 사물을 나타내는 ‘일반주제명’, ② 자료의 물리적 형태나 예술양식, 문학형식 등을 나타내는 ‘형식’, ③ 자료에 나타난 개인·단체·회의의 이름과 같은 ‘고유명사’ 세 가지이며, 여기에 ① 일반주제명, ② 형식, ③ 지리, ④ 시대와 같은 세목을 덧붙여 해당 자료의 주제를 보다 구체적으로 표현할 수 있다.

MARC에서는 이러한 속성을 중심으로 주제명 데이터(6XX필드)를 구조화하고 있다. <표 1>과 같이, 600-651필드는 자료의 내용과 직접적으로 관련된 주제명을, 그리고 655-658 필드는 자료의 범주를 나타내는 물리적 특성이나 장르, 형식, 자료와 관련된 직업, 기능, 교육 과정 등을 기술하도록 구조화하고, 주제명표목표나 시소러스에 의해 통제되지 않은 주제명의 기술을 위해 653필드를 마련하였다. MARC에서는 이처럼 자료의 ‘내용’과 물리적·내용적 ‘특성’에 해당하는 데이터를 명확하게 구분하고 있는데, 가령, 「판타지소설 쓰는 법」과 소설 「해리포터」가 있을 때, “판타지소설”이라는 동일한 주제명이라 하더라도 전자와 같이 자료의 내용에 해당하는 것은 ‘일반주제명’(650필드)에, 그리고 후자와 같이 자료의 물리적 특성이나 장르, 형식에 해당하는 것은 ‘장르/형식’(655필드)에 기술하도록 엄격히 구분하고 있다.

3) 자료의 물리적 특성이나 장르, 형식 등은 “무엇에 관한 것”이 아니라 “무엇”을 설명하는 용어에 해당한다. aboutness와의 구분을 위해 이를 ‘isness’(is와 -ness의 합성어)로 정의하기도 하는데, isness는 일반적으로 자료의 내용과 직접적인 관련이 없는 형식이나 장르, 관점, 연구방법, 대상 이용자 등을 포괄하는 의미로 사용되고 있다.

〈표 1〉 MARC21 및 KORMARC의 주제명 필드

주제명부출표목(통제)		비통제	색인어(통제)
600 개인명	648 시대	653 비통제 색인어	655 장르/형식
610 단체명	650 일반주제명		656 직업
611 회의명	651 지명		657 기능
630 통일표제			658 교육과정 목표

IFLA에서 발표한 FRBR과 FRASAD에서도 저작과 관련된 주제 개체를 명료하게 정의하고 있다. 특히, FRBR에서는 이용자가 서지레코드에서 어떤 정보를 찾고자 하는지를 충분히 검토한 후 주제에 해당하는 개체를 ① 개념(추상적인 관념이나 생각), ② 대상(사물), ③ 사건(행동이나 일어난 일), ④ 장소(위치)로 정의하였다. FRASAD 실무그룹이 구성된 후 FRBR에서 제시한 이들 개체가 과연 주제의 속성을 충분히 포함하는지에 대해 검토한 바 있지만(Delsey 2005), 2011년 최종 발표된 FRASAD에서 주제에 해당하는 개별 개체를 새롭게 정의하기보다 상위 개념인 ‘테마’(theme)의 관점에서 포괄적으로 논의하기로 결정함에 따라 FRBR에서 정의한 이들 4가지 개체를 그대로 유지하고 있다(IFLA 2011). FRBR과 FRASAD에서는 자료의 내용과 직접적으로 관련이 있는 속성만 ‘주제’로 취급하고, 내용과 관련 없는 자료의 형식이나 장르, 대상 이용자 등은 ‘저작’의 속성으로 간주하고 있다.

주제의 속성과 관련하여 살펴볼 또 다른 것은 FAST(Faceted Application of Subject Terminology)이다. OCLC에서 개발한 FAST는 LCSH가 가진 구조적 한계를 극복하고, 웹 환경에서 주제 접근성을 향상하기 위해 LCSH에 패킷 구조를 적용한 것이다. OCLC는 2001년에 WorldCat과 연계한 주제 검색 툴인 ‘searchFAST’를, 이어 자연어를 전거형 접근점으로 자동 변환할 수 있는 ‘assignFAST’를 개발하였다. FAST에서는 LCSH의 주제명을 8개의 패킷으로 범주화하고 있는데, 이 중 7개는 ‘주제’ 패킷이고, 나머지 하나는 ‘장르/형식’ 패킷이다. 주제 패킷에 해당하는 속성은 자료의 내용과 직접적으로 관련된 ① 일반주제명, ② 장소, ③ 시대, ④ 개인명, ⑤ 단체명, ⑥ 사건, ⑦ 표제 등 7개이며, 자료의 내용과 관련이 없는 장르·형식은 이들 주제 패킷과 분명하게 구분하고 있다(OCLC 2013).

이처럼 자료의 주제에 해당하는 속성은 자료에서 직접적으로 다루고 있는 내용, 즉 개념, 대상, 사건, 장소, 시대 등으로 한정할 수도 있고, 혹은 보다 포괄적 관점에서 장르나 형식 등 자료가 가진 고유한 특성이나 이용 대상 등을 주제의 범주에 포함시킬 수도 있다. 장르나 형식 등은 엄밀한 의미에서 주제의 속성에 해당하지 않는다고 볼 수 있지만, 특정 주제로 한정할 수 없는 자료나 다수의 주제를 취급하고 있는 자료, 또는 영화, 지도, 음악, 미술, 법률자료 등과 같은 특수한 유형의 자료에서 유용한 주제 접근점으로 활용될 수 있기 때문에 이를 주제의 속성에 포함시키고 있는 사례도 적지 않다. 이러한 점을 고려하여, 이 연구에서는 주

제의 속성을 다음 <표 2>와 같이 ① 자료의 내용과 직접적으로 관련이 있는 속성, ② 자료의 물리적·내용적 특징을 지칭하는 속성, ③ 자료의 이용과 관련된 속성으로 범주화하고, 이를 주제명 데이터의 품질 분석을 위한 근거로 활용하고자 한다.

<표 2> 주제의 속성 구분

구분	자료의 내용	자료의 물리적·내용적 특징	자료의 이용
주요 속성	개념(학문분야 포함) 대상(개인명, 단체명, 회의명, 사물(생물·무생물)의 이름) 사건 장소 시대	장르(스타일이나 기법) 형식(표현형식이나 서지적 형태) 물리적 특성(매체)	이용대상 직업 기능 교육과정

Ⅲ. 주제명 데이터의 품질

1. 품질 분석의 방법

주제의 개념과 실체에 대한 분석 결과를 토대로 하여, 지금부터는 국립중앙도서관의 주제명 데이터에 대한 품질 검증을 시도하고자 한다. 국립중앙도서관에서는 지난 2002년에 개발한 『국립중앙도서관 주제명표목표』를 토대로 2003년부터 학습서, 교과서, 만화, 학위논문을 제외하고 국내에서 발행된 모든 단행본에 대해 주제명을 부여하고 있다. 국가대표도서관으로서 서지정보의 표준화를 주도하고 있는 국립중앙도서관의 목록레코드는 국내 도서관계에서 실질적인 ‘표준’으로 인지되고 있다. 이에 이 연구에서는 국립중앙도서관의 주제명 데이터가 가진 품질을 냉철히 분석하고, 이에 내재된 근본적인 문제를 파악해 보고자 한다. 구체적으로, 주제명 데이터의 품질은 개별 자료에 부여된 주제명 데이터의 양, 주제명 데이터의 속성, 개별 용어의 수준(granularity) 등을 중심으로 분석하였으며, 분석은 ‘절대적 품질’과 ‘상대적 품질’로 나누어 실시하였다. 절대적 품질은 국립중앙도서관의 주제명 데이터가 가진 실재적 수준을, 상대적 품질은 비교 도서관과의 대조를 통해 상대적 수준을 가늠하는데 목적을 두었다.

상대적 품질 검증을 위한 비교 대상 기관으로는 LC와 국내의 A대학도서관으로 선정하였다. 주지하다시피 LC는 고품질의 데이터를 생산하기 위해 다양한 정책을 펴고 있으며, 국내는 물론이고 전 세계의 많은 도서관으로부터 그 품질을 널리 인정받고 있다. 주제명의 구조나 조합의 순서에 있어 국립중앙도서관과 차이가 나지만, 주제명 데이터의 생성과 관련하여서는 오랜 전통과 풍부한 경험을 가지고 있기 때문에 LC와의 비교를 통해 국립중앙도서관에서 구축한 주제명 데이터의 품질을 가늠하는 것이 의미가 있다고 판단되었다. 반면, A대학도서관

은 국내 도서관계에서 비통제 형식의 주제명 데이터를 제공하고 있는 대표적인 곳이다. 현행 OPAC에서 주제 검색이 가진 한계를 인지하고 있는 일부 도서관에서는 비통제 주제명의 구축을 통해서라도 자관 이용자의 주제 검색을 지원하고자 나름의 노력하고 있다. 동의어나 동형이의어의 처리 등 용어의 질에 있어서는 국립중앙도서관이나 LC와 분명한 차이가 있을 수 있지만, 주제명 데이터에 반영된 ‘속성’이나 ‘관점’은 충분히 비교해 볼 가치가 있다고 판단되어 A대학도서관을 비교 대상에 포함시켰다.

샘플 데이터는 국립중앙도서관, LC, A대학도서관에서 공통적으로 소장하고 있는 자료에서 추출하였다. 먼저, 자료의 이용목적이나 내용적 특성에 따른 차이를 고려하고자 ‘대중자료’와 ‘학술자료’를 나누고 두 부류에 해당하는 자료 중에서 ‘소설’과 ‘사회과학’을 임의로 선정하였다. 다음으로, 「대한민국 국가서지 2014」를 이용하여 세 도서관에서 공통적으로 소장하고 있는 소설 및 사회과학 분야의 번역서 각 30권을 무작위 선정하였다. 국립중앙도서관과 A대학도서관의 OPAC에서는 이들 번역서에 대한 주제명 데이터를, 그리고 LC의 OPAC에서는 번역서에 대한 원저를 탐색하여 주제명 데이터를 각각 확보하였다. 세 기관으로부터 확보한 데이터는 스프레드시트에 정리한 후 데이터의 양, 속성, 용어의 수준 등을 분석하였고, 직관적인 비교를 위해 이들 데이터를 워드 클라우드(word cloud)로 구현하였다.⁴⁾

2. 품질 분석의 결과

가. 절대적 품질

국립중앙도서관에서 개별 자료에 부여한 주제명 데이터의 양은 다음 <표 3>과 같이, 소설의 경우 평균 1.10개, 사회과학 자료의 경우 평균 1.50개로 나타났다(전체 평균 1.30개). 무작위로 선정한 30종의 소설 중 3종의 소설에만 2개의 주제명 데이터가, 나머지 27종에는 모두 1개의 주제명 데이터가 부여되어 있었다. 사회과학 자료에 부여된 주제명 데이터의 양은 소설 보다 약간 많은 편이었다. 30종 중 15종에는 2개의 주제명 데이터가, 나머지 15종에는 단 1개의 주제명 데이터가 부여되어 있었다.

<표 3> 주제명 데이터의 양

	소설	사회과학	계
주제명 데이터의 합계	33	45	78
저작 당 평균	1.10	1.50	1.30

4) 워드 클라우드는 오픈소스인 WordItOut (<http://worditout.com/word-cloud/make-a-new-one>)을 활용하여 생성하였다. 데이터를 변환하기에 앞서, 국립중앙도서관의 주제명 데이터는 가독성을 높이기 위해 각괄호([])에 병기된 한자를 삭제하였다.

다음으로, 주제명 데이터의 속성을 분석한 결과, 소설 30종에 포함된 주제명 데이터는 모두 자료의 물리적·내용적 특징에 해당하는 것으로 나타났다. 구체적으로, “미국 현대 소설[美國現代小說]”, “미국 소설[美國小說]”과 같이 ‘문학형식’을 나타내는 주제명 데이터가 33개 중 32개(97.0%)로 절대적인 비중을 차지하였으며, “추리 소설[推理小說]”과 같이 소설의 ‘장르’를 지칭하는 데이터가 1개 포함되어 있었다. 이에 비해, 소설의 내용과 직접적으로 관련된 개념, 대상, 사건, 장소, 시대 등에 해당하는 주제명은 전혀 존재하지 않는 것으로 나타났다.

소설과 달리, 사회과학 자료에 부여된 주제명 데이터는 대부분 자료의 ‘내용’과 관련된 것이었다(45건 중 44건(97.7%)). “사회 변동[社會變動]”, “정치 이론[政治理論]” 등 추상적인 관념이나 생각, 이론 등을 나타내는 개념이 42개(93.3%)로 가장 많은 비중을 차지하였으며, 개인의 이름에 해당하는 주제명이 2개(“마르크스[Marx, Karl Heinrich]”, “엔겔스(인명)[Engels, Friedrich]”), 자료의 형식에 해당하는 주제명(“자서전[自敘傳]”)이 1개로 나타났다(<표 4> 참조). 대상이용자, 직업, 교육과정 등 자료의 이용과 관련된 주제명은 소설과 사회과학 자료 모두에서 발견되지 않았다.

<표 4> 주제명 데이터의 속성

자료의 내용	소설		사회과학	
			개념	42 (93.3%)
		대상	2 (4.4%)	
자료의 물리적·내용적 특징	형식	32 (97.0%)	형식	1 (2.2%)
	장르	1 (3.0%)		
자료의 이용	-		-	
계	33 (100%)		45 (100%)	

주제명으로 사용된 용어의 특성을 보다 세밀하게 분석한 결과, 국립중앙도서관의 주제명 데이터는 대체로 자료 전체를 포괄하는 용어로 이루어져 있는 것으로 나타났다. 앞서 언급하였듯이, 소설 30종에 부여된 주제명은 대부분 ‘문학형식’을 나타내는 것이었는데, 그 중에서도 “미국 현대 소설[美國現代小說]”이라는 주제명이 가장 많은 비중을 차지하였다(<표 5> 참조). 문학 형식을 나타내는 이들 주제명 간에는 용어의 수준에 따른 차이가 분명하게 드러났다. 가령, 똑같이 ‘미국현대소설’에 해당하는 자료임에도 불구하고 자료에 따라 “영미 문학[英美文學]”, “미국 문학[美國文學]”, “미국 소설[美國小說]”, “미국 현대 소설[美國現代小說]” 등 상이한 계층이 적용되어 있었고, 하나의 자료에 2개의 주제명이 부여된 경우에도 상위어 “미국 문학[美國文學]”과 그 하위어인 “미국 소설[美國小說]”과 같이 동종의 주제명인 것으로 드러났다. 이밖에, 아프리카 소설에 “영미 소설[英美小說]”을 주제명으로 부여한 결정적인 오류도 있었다.

〈표 5〉 소설에 부여된 주제명

주제명	건수	주제명	건수
영미 문학[英美文學]	3	영국 소설[英國小說]	1
미국 문학[美國文學]	1	영국 현대 소설[英國現代小說]	1
미국 소설[美國小說]	6	캐나다 문학[--文學]	1
영미 소설[英美小說]	1	추리 소설[推理小說]	1
미국 현대 소설[美國現代小說]	18	계	33

사회과학 자료에 부여된 주제명도 주로 자료의 전체 내용을 포괄하는 용어이거나 해당 자료의 근원이 되는 학문분야의 명칭으로 판명되었다. 가령, 칸트, 헤겔, 마르크스, 프로이트 등의 이데올로기를 다룬 『광신: 어느 저주받은 개념의 계보학』(알베르토 토스카노 저)이라는 자료에 “정치 사회학[政治社會學]”이, 그리고 플루토크라트의 개념과 역사, 사회적 영향력 등에 대해 다룬 『플라토크라트』(크리스티아 프리랜드 저)에 “사회 계층[社會階層]”이라는 다소 포괄적인 주제명이 부여되어 있었다. 자료의 주제를 나타내는 용어가 이렇듯 포괄적이다보니 전혀 상이한 내용을 다룬 두 자료에 동일한 주제명이 부여되어 있기도 하였다. 인재양성에 초점을 두고 교육개혁을 주장한 『이노베이터의 탄생』(토니 와그너 저)과 창의적 아이디어의 개발에 있어 네트워크의 중요성을 강조한 『탁월한 아이디어는 어디서 오는가』(스티븐 존슨 저)라는 자료에 “혁신[革新]”이라는 주제명이 동일하게 부여되어 있었다.

표제의 특정 용어를 유일한 주제명으로 채택한 경우도 적지 않았다. 진화론적 관점에서 중년의 심리적·육체적·사회적 변화를 다룬 『중년의 발견』에 “중년[中年]”, 인종과 관련된 문화이론에 대해 다룬 『식민 욕망』에 “식민 주의[植民主義]”가 이들 자료에 부여된 유일한 주제명이었다. 이처럼 사회과학 자료 30종에 부여된 주제명 데이터는 전반적으로 다소 포괄적이고 모호하였으며, 자료의 세부 주제나 관점을 반영한 주제명을 거의 찾아볼 수 없었다.

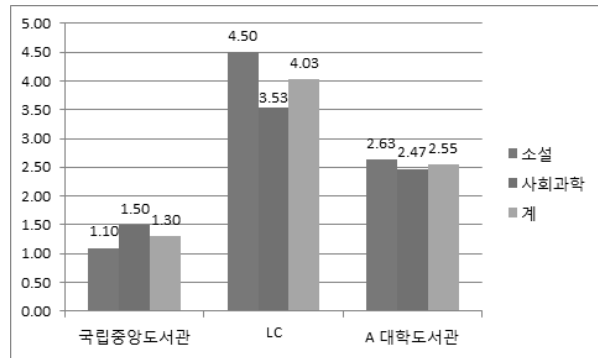
나. 상대적 품질

국립중앙도서관에서 구축한 주제명 데이터의 상대적 품질을 분석하기 위해 LC와 A대학도서관에서 구축한 주제명 데이터와 비교하였다. 비교에 앞서, LC 주제명 데이터에 대한 처리 방식을 설정하여야 했다. 주지하다시피 LC의 주제명은 국립중앙도서관이나 A대학도서관과 달리, “Police—California—Los Angeles—Fiction”처럼 대부분 주표목과 세목이 결합된 전 조합 구조로 이루어져 있기 때문이다. 이에 이 연구에서는 LC 주제명 데이터에 다음과 같은 기준을 적용하였다. 첫째, 결합된 구조 전체를 하나의 주제명으로 간주하였다. 세목은 주표목을 한정하거나 주표목의 하위주제를 표현한 것이므로 이를 분절할 경우 그 의미가 상쇄될 수 있다고 판단하였다. 둘째, 주표목과 세목으로 구성된 경우 주표목을 근거로 하여 속성을 분석

하였다. 가장 주된 주제는 주표목에 반영되고 있고, 더불어 LCSH의 논리적 결합 구조에 따라 주표목이 우선적으로 선정된 것이라는 점을 고려하였다.

이러한 기준을 적용하여 먼저, 샘플 레코드에 포함된 주제명 데이터의 양을 상호 비교하였다. 그 결과, 다음 <그림 1>과 같이

LC는 평균 4.03개, A대학도서관은 평균 2.55개의 주제명 데이터를 가지고 있는 것으로 나타나, 평균 1.30개에 불과한 국립중앙도서관에 비해 2~3배 이상 많은 주제명 데이터를 보유하고 있었다. 주제명 데이터의 누락은 A대학도서관에서만 발견되었는데 (60종 자료 중 10종 자료, 누락율 16.7%), 주제명 데이터가 아예 누락된 레코드를 제외하면 A대학도서관에서 개별 자료에 부여한 주제명 데이터의 수는 평균 3.06개 수준이었다.⁵⁾



<그림 1> 주제명 데이터의 수 비교

다음으로, 세 도서관에서 구축한 주제명 데이터를 속성에 따라 구분한 후 비교하였다(<표 6> 참조). 먼저, 소설에 대한 국립중앙도서관의 주제명 데이터는 앞서 살펴본 것처럼 대부분 ‘문학형식’(97%)에 해당하는 것이었지만, LC의 주제명 데이터는 이와 사뭇 다른 양상을 보였다. “미국 현대 소설”과 같이 문학형식을 직접적으로 표현한 주제명 데이터는 전혀 찾아볼 수 없었고, “Married people—fiction”, “Ants—fiction”과 같이 문학형식에 해당하는 “fiction”이 세목의 형태로 일관되게 나타났다. LC의 주제명 데이터(주표목)는 주로 소설의 ‘내용’을 드러내고 있었는데, “Brothers and sisters—Fiction”, “Physicians—Fiction”과 같이 소설의 등장인물과 관련된 ‘대상’이 가장 많은 비중을 차지했으며(33.3%), “Friendship—Fiction”, “Emotional problems—Fiction”과 같이 소설의 소재나 재료에 해당하는 ‘개념’(23.0%), 그리고 “Bronx (New York, N.Y.)—Fiction”과 같이 소설의 배경이 되는 ‘장소’(15.5%)가 그 뒤를 이었다. 국립중앙도서관에서 단 1개에 불과했던 ‘장르’ 주제명이 LC에서는 거의 모든 소설에 부여되어 있다는 점도 뚜렷한 차이였다.

한편, A대학도서관의 경우 ‘문학형식’에 해당하는 주제명의 비율이 절대적(66.7%)이라는 점에서는 국립중앙도서관과 유사하였지만, 국립중앙도서관에서 거의 볼 수 없었던 “스릴러소

5) A대학도서관에서는 통제된 용어를 사용하지 않기 때문에 ‘비통제 주제명’인 653필드에 자료의 주제와 관련된 용어를 입력하고 있다. 이들 데이터는 OPAC에서 <주제어 검색>을 위한 기초데이터로 활용되며, 상세 검색결과에 <키워드>라는 항목으로 제공되고 있다. 이처럼 A대학도서관에서는 이들 데이터를 ‘주제어’ 또는 ‘키워드’로 칭하고 있지만, 이 연구에서는 이를 엄격히 구분하지 않고 ‘주제명’이라는 용어로 통일하여 사용하였다.

설”, “가족소설”, “우정소설” 등 소설의 ‘장르’(21.8%)와 “시간”, “숙명”, “기다림” 등 소설의 소재나 재료에 해당하는 ‘개념’(11.5%)이 다수 포함되어 있었다. 자료의 이용과 관련된 데이터(이용대상, 직업 등)는 세 기관 모두에서 발견되지 않았다.

〈표 6〉 소설에 대한 주제명 데이터의 속성 비교

		국립중앙도서관		LC		A대학도서관	
		건수	비율	건수	비율	건수	비율
내용	개념	-	-	31	23.0%	9	11.5%
	대상	-	-	45	33.3%	-	-
	사건	-	-	4	3.0%	-	-
	장소	-	-	21	15.5%	-	-
	시대	-	-	-	-	-	-
특징	형식	32	97.0%	-	-	52	66.7%
	장르	1	3.0%	34	25.2%	17	21.8%
계		33	100.0%	135	100.0%	78	100.0%

동일 자료에 대해 세 기관에서 제공하는 주제명의 구체적인 사례를 비교해 보면 그 차이는 더욱 극명하게 드러난다. 〈표 7〉에서 보듯이, 국립중앙도서관에서는 소설의 주제명으로 문학 형식에 해당하는 주제명만을 제시하고 있지만, LC는 복수의 주제명을 통해 소설과 관련된 인물이나 장소, 소재 등을 비교적 상세하게 나타내고 있다. 뿐만 아니라 LC에서는 소설의 장르/형식을 자료의 내용을 나타내는 주제명과 분명하게 구분하고 있어 소설의 하위형식이나 서술양식에 따른 체계적인 접근도 지원하고 있음을 알 수 있다. A대학도서관의 경우에도, 비록 자료에 따라 용어의 수준이나 주제명의 속성에 차이가 있지만 문학형식 외에 장르나 소설의 세부 주제, 소재 등을 반영하고자 노력한 흔적을 엿볼 수 있다.

〈표 7〉 소설에 대한 주제명 데이터의 사례 비교

	국립중앙도서관	LC		A대학도서관
	주제명	주제명	장르/형식	키워드
나를 찾아줘 : 길리언 플린 장편소설 (길리언 플린 저)	미국 소설[美國小說]	Husbands—Fiction Married people—Fiction Wives—Crimes against—Fiction	Mystery fiction	현대소설 영미소설 장편소설 심리스릴러 스릴러
헛된 기다림 (나담 아슬람 저)	영국 소설[英國小說]	Interpersonal relations—Fiction Afghanistan—Social life and customs—Fiction	-	영미소설 기다림
가벼운 나날 (제임스 샬터 저)	미국 현대 소설[美國現代小說]	Married people—Fiction Divorced people—Fiction Hudson River Valley (N.Y. and N.J.)—Fiction	Domestic fiction	미국문학 미국현대소설

마지막으로, 소설에 대한 주제명 데이터 간의 관련성 혹은 분산의 정도를 비교하고자 샘플 자료에서 확보한 주제명 데이터를 워드 클라우드로 구현하였다. 다음 <그림 2>와 같이, 소설에 대한 국립중앙도서관의 주제명 데이터는 일부 용어(“미국 현대 소설”)에 대한 편중 현상이 두드러지게 나타나지만, LC나 A대학도서관은 동일 자료에 대해 훨씬 다양하고 상세한 주제명 데이터를 가지고 있음을 볼 수 있다. 주제명표목표를 토대로 용어를 엄격히 통제하고 있는 국립중앙도서관과 LC 두 곳의 데이터만 비교해 보더라도 용어의 수준이나 표현 방식에서의 차이가 뚜렷이 드러난다. “[주표목]-fiction”과 같은 표현의 일관성을 유지하고 있는 LC와 달리, 국립중앙도서관의 주제명 데이터는 문학형식을 나타내는 데이터 간에도 용어의 수준이 일관되지 않기 때문이다. 국립중앙도서관의 주제명 데이터가 가진 이러한 비일관성의 문제는 비통제어의 사용으로 이에 대한 근원적인 한계를 가질 수밖에 없는 A대학도서관과 오히려 유사한 양상을 보이고 있다. A대학도서관에서도 문학형식에 해당하는 주제명이 “미국문학”과 “영미문학”, “미국소설”과 “영미소설”, “미국현대소설” 등으로 상당히 흩어져 있다.



<그림 2> 소설의 주제명 클라우드 비교

한편, 사회과학 자료에 대해서는 세 도서관이 공통적으로 ‘개념’에 해당하는 용어를 주제명으로 선정한 비율이 높았다(<표 8> 참조). 개념에 해당하는 주제명 중에는 학문분야의 명칭도 포함되어 있었는데, 이에 대해서는 세 기관 간에 다소 차이가 있었다. 국립중앙도서관의 경우 42개 중 9개(21.4%)가 “사회학[社會學]”, “문화 사회학[文化社會學]”, “언어 사회학[言語社會學]” 등과 같은 학문분야의 명칭이었었는데, LC는 93개 중 7개(7.5%), 그리고 A대학도서관은 68개 중 4개(5.9%)에 불과해 국립중앙도서관에서 학문분야를 주제명으로 채택하는 비중이 가장 높게 나타났다. LC에서는 자료 자체가 특정 학문분야에 관한 것이거나 “Persuasion (Psychology)”, “Right and left (Political science)”와 같이 동일 주제명이 여러 학문분야에서 공통적으로 사용되는 경우 그 범주를 제한하기 위해 학문명을 한정어의 형태로 부기하고 있었지만, 국립중앙도서관에서는 자료의 직접적인 주제가 아닌 경우에도 이

를 해당 자료의 ‘유일한’ 주제명으로 채택하고 있다는 차이가 있었다.⁶⁾

자료와 관련된 ‘대상’을 주제명으로 채택하는 것과 관련하여서도, 국립중앙도서관과 LC 간에 뚜렷한 차이가 발견되었다. 국립중앙도서관에서는 표제에 관련 인물이 명확하게 드러나는 경우에 한해서만 이를 주제명으로 채택하고 있었지만(2건), LC에서는 자전적인 성격에 해당하는 자료의 저자나 자료에서 특히 자주 언급되는 인물도 주제명에 포함하고 있었다(11건). 특정 장소를 나타내는 주제명은 세 기관 중 LC에서만 유일하게 발견되었다. LC에서는 “Great Britain—Race relations”와 같이 자료와 관련된 구체적인 장소(국가명, 지리명 등)를 주표목으로 표현한 사례가 2건, “Political participation—United States”와 같이 세목에 포함된 사례가 13건 있었다. 이에 비해, 국립중앙도서관에서는 “미국 민주 주의[美國民主主義]”, A대학도서관에서는 “프랑스정치”가 장소(국가)를 포함하는 유일한 주제명이었다.

〈표 8〉 사회과학 자료에 대한 주제명 데이터의 속성 비교

		국립중앙도서관		LC		A대학도서관	
		건수	비율	건수	비율	건수	비율
내용	개념	42 (9)	93.3%	93 (7)	86.9%	68 (4)	91.9%
	대상	2	4.4%	11	10.3%	4	5.4%
	사건	—	—	—	—	—	—
	장소	—	—	2	1.9%	—	—
	시대	—	—	—	—	—	—
특징	형식	1	2.2%	1	0.9%	1	1.4%
	장르	—	—	—	—	1	1.4%
계		45	100.0%	107	100.0%	74	100.0%

‘개념’에 있는 ()안의 수는 학문분야의 명칭을 주제명으로 제시한 사례임

동일 자료에 대한 세 도서관의 주제명 데이터를 상호 비교해 보면, 주제를 나타내는 용어 간의 차이가 보다 뚜렷이 드러난다(〈표 9〉 참조). 가령, 17세기 영국의 철학자 토머스 홉스의 저서 『리바이어던』을 토대로 하여 도덕적 합리성에 대한 의문을 제시한 『리바이어던의 논리』(데이비드 고티에 저)에 대해, 국립중앙도서관에서는 “정치 이론[政治理論]” 1개의 주제명만 부여하고 있지만, LC와 A대학도서관에서는 자료와 관련된 구체적인 대상인 “홉스”, “리바이어던”과 이 자료에서 궁극적으로 논의하고 있는 “정치”, “윤리”, “도덕” 등을 주제명에 포함하고 있었다. 국립중앙도서관에서 “중년[中年]”으로만 표현한 자료에 대해서도, LC에서는

6) 국립중앙도서관 주제명표목표에서도 주제명의 의미가 불분명하거나 동음이의어인 경우, 혹은 여러 주제(학문) 분야에서 공통적으로 사용되는 경우 그 의미를 한정하기 위해 한정어를 원괄호 속에 기재하도록 하고 있다. 그러나 샘플 레코드에서 이러한 한정어가 사용된 사례는 찾아볼 수 없었다.

세부주제나 관점을 보다 구체적으로 나타내고 있는 점도 중요한 차이였다. 이러한 경향은 이 연구에서 선정한 샘플 자료 대부분에서 유사하게 나타나는데, LC와 A대학도서관의 주제명이 대체로 해당 자료의 주제를 구체적이고 상세하게 표현한 것이라면, 국립중앙도서관의 주제명 데이터는 다소 포괄적이고 일반적인 용어라 할 수 있다.

<표 9> 사회과학 자료에 대한 주제명 데이터의 사례 비교

	국립중앙도서관	LC	A대학도서관
리바이어던의 논리 (데이비드 고티에 저)	정치 이론[政治理論]	Hobbes, Thomas, 1588-1679 Leviathan Political science—Great Britain— History—17th century Ethics, Modern—17th century	리바이어던 홉스 정치이론 도덕이론 홉스사상 정치학
중년의 발견 (데이비드 베인브리지 저)	중년[中年]	Middle age Middle-aged persons—Life skills guides Middle age—Psychological aspects	인문교양 중년기
자기 절제 사회 (대니얼 액스트 저)	사회학[社會學]	Self-control Moderation Supply and demand	행동심리학 절제사회

사회과학 자료에 대해 세 기관에서 구축한 주제명 데이터를 워드 클라우드로 구현한 결과는 다음 <그림 3>과 같다. 소설과 달리, 사회과학 분야의 자료는 주제의 다양성으로 인해 특정 용어에 대한 편중 현상이 거의 나타나지 않았다. 동일 주제명이 2회 이상 사용된 경우는 국립 중앙도서관의 “사회학”(3회)과 “혁신”(2회), LC의 “Sociology”(2회), A대학도서관의 “아이디어”(2회) 뿐이었다. 소설과 마찬가지로, 사회과학 자료의 경우에도 국립중앙도서관의 주제명에서는 “사회 이론”, “서양 경제사”와 같이 다소 포괄적인 용어가 주제명으로 채택되었지만, LC에서는 “Middle-aged persons—Life skills guides”, “Cooperativeness—Psychological



<그림 3> 사회과학 분야의 주제명 클라우드 비교

aspects”와 같이 자료의 내용을 구체적으로 표현하는 용어로 구성되어 있음을 볼 수 있다. 어휘가 통제되지 않은 A대학도서관의 주제명에서 “발전”, “개발”, “development”와 같이 동일 개념의 상이한 표현이나 “공동체성”과 같이 의미가 모호한 주제명이 발견되었지만, 국립중앙도서관에 비해 대체로 용어의 수준이 상세하고 다양한 편임을 알 수 있다.

IV. 현행 주제명 데이터의 한계와 과제

앞서 2장에서 필자는 자료의 주제란 “자료가 무엇에 관한 것인지”를 의미하며, 그 중에서도 ‘이용자가 특히 관심을 가질 만한 무엇’이 구체적인 주제명으로 표현되어야 한다고 정리한 바 있다. 주제명 데이터는 그 자체로 이용자의 주제 접근을 지원하는 중요한 수단이 되기 때문이다. 비록 샘플 데이터의 수는 적었지만, 이러한 관점에 근거하여 국립중앙도서관에서 지난 10여 년 간 구축해 온 주제명 데이터의 실체를 파악한 결과, 주제검색의 기초데이터로 활용하기에 여러 한계를 가지고 있는 것으로 나타났다. 주제명표목표의 개발이나 국립중앙도서관 OPAC에서의 주제 검색 등 외형적으로는 커다란 변화가 있었지만, 주제명 데이터가 자료의 내용(주제)을 제대로 반영하고 있는지, 나아가 주제 탐색어로서의 기본 요건을 갖추고 있는지에 대해서는 결코 긍정적인 평가를 내리기 어려웠기 때문이다. 이에 지금부터는 앞서의 분석 결과를 바탕으로 현재 국립중앙도서관의 주제명 데이터가 가진 한계를 살펴보면서 향후 주제명 데이터의 생성과 관련하여 우리 도서관계가 고민해 보아야 할 추진 전략에 대해 논의해 보고자 한다.

앞서 살펴보았듯이, 소설에 대한 국립중앙도서관의 주제명 데이터는 대부분 ‘문학형식’에 해당하는 것이었다. 사회과학 자료의 주제명도 학문분야나 다소 포괄적인 용어로 구성되어 있었다. 개별 자료에 부여된 주제명 데이터의 양 또한 소설의 경우 평균 1.1개, 사회과학 자료의 경우 평균 1.5개에 불과해, 비교 대상 기관인 LC나 A대학도서관과 비교하여 그 차이가 두드러졌다. 주제명 데이터가 가진 이러한 용어의 특성과 양은 서로 밀접한 관계를 가진다고 볼 수 있다. 주제명을 1-2개의 용어로 표현하려다보니 자료의 전반적인 내용을 함축적으로 표현하는 용어를 선택할 수밖에 없기 때문이다. 주제명을 부여하는데 있어 이처럼 자료 전체를 아우르는 포괄적인 용어가 적절한지, 아니면 다수의 용어를 사용하여 자료의 다양한 측면을 깊이 있게 반영하는 용어가 적절한지에 대해서는 아직까지 이론적으로 검증된 바가 없다 (Taylor 2009, 310). 개별 도서관의 상황과 이용자의 요구에 따라 ‘정책적으로’ 판단하고 있을 뿐이다. 그러나 분명한 것은 이러한 판단이 주제 검색에 결정적인 영향을 미친다는 점이다. 국립중앙도서관의 사례와 같이, ‘one overall subject’ 만으로는 검색의 정확성은 물론이

고 내용 식별을 위한 기초 데이터로 사용하기에 커다란 한계가 있기 때문이다.

가치판단의 문제를 차치하고서라도, 현재 국립중앙도서관에서 제공하는 주제명 데이터는 그 자체로도 다음과 같은 한계를 가지고 있다. 먼저, 주제명 선정에 있어 가장 기본적인 원칙이라 할 수 있는 ‘특정성’(specificity)의 원칙에 전면 배치된다는 점이다. 특정성의 원칙은 자료의 내용을 직접적으로 나타내는 특징의 용어를 주제명으로 채택해야 한다는 원칙으로써 자료에서 다루는 내용보다 광의의 또는 상위의 용어가 적절하지 않다는 의미를 담고 있다(최달현, 이창수 2010, 246-247).⁷⁾ 그러나 이번 연구에서 살펴본 국립중앙도서관의 주제명 데이터는 대부분 이러한 원칙에서 크게 벗어나 있었다. 앞서 살펴본 『플루토크라트』라는 자료에 ‘특권 계층[特權階層]’이나 ‘상류 계층[上流階層]’ 대신에 이를 포괄하는 개념인 ‘사회 계층[社會階層]’이, 그리고 『자기 절제 사회』에 ‘사회 진화[社會進化]’나 ‘절제(자제) [節制]’ 대신에 ‘사회학’이라는 학문분야가 이 자료에 대한 유일한 주제명이었고, 주제명 가운데 구체적인 장소나 시대가 반영된 사례도 거의 찾아볼 수 없었다.

현행 주제명 데이터가 가진 이러한 특성을 보다 세밀히 추적해 보면, 주제명 데이터가 분류기호와 깊은 관련이 있음을 확인할 수 있다. 국립중앙도서관에서 개별 자료에 부여한 주제명을 KDC 분류기호와 대조해 보면, <표 10>과 같이 분류기호에 반영된 용어가 주제명과 정확하게 일치하는 경우가 적지 않다. 소설에 부여된 주제명도 분류기호의 조합 패킷인 「국가+문학형식+시대」를 그대로 반영한 것이라 볼 수 있다.

<표 10> 주제명과 분류기호의 관계

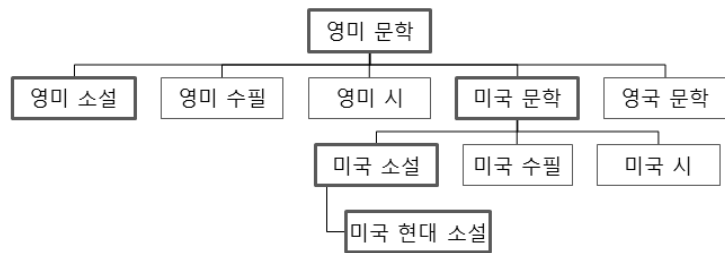
자료명	주제명	KDC 분류기호
플루토크라트	사회 계층[社會階層]	332.6 사회계층, 계급, 신분
중년의 발견	중년[中年]	331.235 중장년
다운사이징 데모크라시	민주 주의[民主主義] 미국 민주 주의[美國民主主義]	340.22 민주주의
훼손된 세상	인간 생태학[人間生態學]	331.4 인간생태학
갯 리얼	사회 변동[社會變動]	331.54 사회변동
눈물의 아이들	미국 소설[美國小說]	843.6 영미소설 - 21세기
아름다운 폐허	미국 현대 소설[美國小說]	843.6 영미소설 - 21세기

7) 특정성의 원칙은 국립중앙도서관에서 주제명표목의 생성과 관련하여 업무 참조용으로 발간한 『국립중앙도서관 주제명표목표 업무 지침』에도 분명하게 기술되어 있다(국립중앙도서관 2014, 4). 그 내용을 그대로 옮겨보면 다음과 같다.

- 자료의 내용에 대해 가장 특징이 되는 주제명을 부여함
- 가장 적절한 주제명이 존재하지 않고, 신설이 부적당한 경우는 가장 가까운 상위 개념에 해당하는 주제명을 부여함
- 지역, 시대를 한정된 주제가 다루어지고 있는 경우 주제명표목표에 등록된 지리명, 시대명을 확인하고 주제명을 부여함

분류기호와 주제명 모두 자료의 ‘주제’를 토대로 한 접근법이라는 점에서 분류기호를 주제명으로 대체하는데 크게 문제가 없어 보일 수도 있지만, 각각에 내재된 본질적 의미를 따져보면 그 결과는 판이하게 달라진다. 주지하다시피 분류기호의 궁극적인 목적은 “자료 전체 또는 핵심 내용을 토대로 유사한 것은 모으고 상이한 것은 구분함으로써 개별 자료의 서가상의 위치를 결정”하는데 있다. 반면, 주제명 데이터는 “개별 자료에 포함된 다양한 주제를 세밀하게 표현함으로써 해당 주제를 포함하는 자료에 대한 체계적인 접근을 지원”하는데 주된 목적을 두고 있다. 이러한 이유로, 분류기호는 세부 주제를 포괄하는 상위 주제나 저자가 특히 강조하는 주제, 혹은 복수 주제를 다룬 경우에 특정 주제를 반영할 수밖에 없는 구조적 한계를 지니지만, 주제명은 이러한 제한 없이 자료에 취급된 다양한 주제나 관점을 모두 반영하고 이를 접근점으로 활용할 수 있다는 점에서 커다란 차이가 있다. 이러한 점을 고려한다면 “분류기호 = 주제명”으로 간주하고, 심지어 이를 ‘유일한’ 주제명으로 활용하고 있는 국립중앙도서관의 업무관행에 대해서는 근본적인 재검토가 필요하다고 할 수 있다.⁸⁾

국립중앙도서관의 주제명 데이터에서 발견된 또 다른 문제는 표현의 일관성(consistency) 문제이다. 관련 자료에 대한 망라적인 검색을 지원하기 위해서는 주제를 일관되게 표현하는 것이 필수적이다. 그러나 이번 연구에서 살펴본 자료 중 소설에서 특히 이러한 일관성의 원칙은 거의 지켜지지 않고 있었다. 30개의 소설 중에서 소설의 장르가 주제명에 포함된 사례는 단 1개에 불과하였으며, 용어의 수준에 따른 비일관성의 문제도 매우 심각한 것으로 드러났다. <그림 4>와 같은 계층구조를 지닌 국립중앙도서관 주제명표목표에서, 동일 부류에 속하는 자료임에도 불구하고 개별 자료에 부여된 주제명은 “영미 문학”, “영미 소설”, “미국 문학”, “미국 소설”, “미국 현대 소설” 등 적용계층이 모두 상이하게 나타났다. 사회과학 자료에서도 “민주주의”와 “미국 민주주의”와 같이 동일 계층에 해당하는 상·하위 개념이 하나의 자료에 동시에 부여되어 있었다.⁹⁾ 문제는 이러한 데이터를 주제 검색에 그대로 사용할 경우 이용자가 선택한



<그림 4> 국립중앙도서관 주제명표목표의 계층 구조

8) 이처럼 분류기호를 그대로 주제명으로 옮겨놓는 수준이라면, 현재와 같이 사서가 직접 이 데이터를 입력하는데 시간과 노력을 들일 필요가 있는지에 대해 의문이 제기된다. 가뜰이나 부족한 입력구조에서, 분류기호와 주제명 상호 간의 변환프로그램을 활용하는 것이 비용 효과적이기 때문이다.

9) 『국립중앙도서관 주제명표목표 업무 지침』에도 상위 개념을 다루면서 특정 하위 개념에 중점을 두거나 일반적인 경우를 다루면서 특정 사례를 드는 경우를 제외하고는, 상위 개념의 주제명과 하위 개념의 주제명을 동시에

검색어에 따라 그 결과가 확연히 달라진다는 데 있다.

이에 더해, 소설의 유일한 주제명으로 이러한 문학형식이 과연 적절한지의 문제도 여전히 남아 있다. 소설에 대한 국립중앙도서관의 주제명은 97%가 ‘문학형식’인데 반해, LC의 주제명은 75%가 소설의 ‘내용’, 나머지 25%가 소설의 ‘세부장르’에 해당하는 것으로 나타나 극명한 차이를 보였다. 주지하다시피 문학형식은 “무엇에 관한”이 아닌 “무엇”을 의미하는 것으로서 자료의 ‘내용’을 직접적으로 나타내는 데이터라고 보기 어렵다. LC에서 문학형식이나 장르를 자료의 내용과 관련된 주제명과 엄격히 구분하고, 기존의 주제 검색에서 장르/형식 검색을 분리하고 있는 까닭도 여기에 있다. 이러한 점에서 볼 때, 문학형식을 소설의 유일한 주제로 취급하고 있는 이러한 관행을 앞으로 우리 도서관계가 ‘표준’으로 삼기에 적절한지에 대해서도 면밀한 검토가 필요해 보인다.

이상의 결과를 종합하면, 현재와 같은 수준의 품질로는 주제명 데이터가 수행해야 할 기본적인 기능¹⁰⁾은 물론이고, 최근 우리 도서관계가 야심차게 추진하고 있는 ‘개체-관계를 이용한 서지적 관계의 효과적 구현’도 불가능하다는 결론에 이르게 된다. 그렇다면 지금 이 시점에서 주제명 데이터의 품질을 제고하기 위한 우리 나름의 해법은 무엇일까? 장기적이고 거시적인 관점에서는 체계적인 직무교육과 훈련을 통해 이와 관련한 사서들의 자질과 역량을 점진적으로 향상시키고, 더불어 사서들이 자료에 대한 장악력과 주제 진단 능력을 겸비하게 된다면 이 문제는 자연스레 해결될 것이다. 그러나 인력 부족이나 잦은 부서이동 등 우리 도서관 업무구조의 낙후성을 고려할 때 사서 개인의 자질과 역량에 전적으로 의존하는 것 또한 머지않아 한계에 봉착할 가능성이 농후하다. 이러한 상황에서 최선의 방법은 우리의 실정에 보다 쉽게 적용할 수 있는 현실적인 방안을 찾는 것이라 할 수 있다.

주지하다시피 주제명 데이터의 생성 과정에서 가장 어려운 점은 해당 자료가 무엇에 관한 것인지를 해석하고 판단하는데 있다. 국립중앙도서관의 업무지침에 나와 있는 것처럼 “자료의 전체 내용을 가장 잘 표현하는 주제명” 혹은 “자료의 가장 중요한 주제에 대한 주제명”(국립중앙도서관 2014, 3)을 선택하는 일은 말처럼 그리 간단한 작업이 아니기 때문이다. 이러한 측면에서 볼 때, 사서의 업무 부담을 덜면서 동시에 주제 검색을 획기적으로 개선할 수 있는 일차적 방안은 주제가 가진 이러한 ‘애매모호한’ 개념을 구체화시켜주는 것이다. 가령, 이번 연구에서 살펴본 것처럼 자료가 가진 다양한 주제를 ‘개념(소재)’, ‘인물’, ‘사건’, ‘장소’,

부여하지 않는 것을 원칙으로 한다고 분명하게 기술되어 있다(국립중앙도서관 2014, 4). “민주 주의”와 “미국 민주 주의” 등 2개의 주제명이 부여된 이 자료는 자료의 내용 자체가 미국의 민주주의에 대한 자료이므로 상위 개념인 “민주 주의”는 불필요한 주제명이라 볼 수 있다.

10) FRSAD에서는 주제명 데이터의 기능을 다음과 같이 4가지로 정의하고 있다(IFLA 2012, 13) : ① 이용자가 제시한 기준에 적합한 주제를 검색, ② 유사한 특성을 가진 둘 이상의 주제를 구분하고 적합한 주제를 확인, ③ 이용자의 요구에 적합한 주제를 선정, ④ 주제 간의 관계를 탐험.

‘시대’ 등으로 구분하고, 자료의 유형이나 특성에 따라 문학작품에 대해서는 ‘문학형식’이나 ‘장르’를, 학술자료에 대해서는 ‘관련 학문’과 같은 추가적 속성을 세부적으로 마련하여 제시해 준다면, 최소한, 막연한 상태에서 무엇을 자료의 주제로 간주할 것인가에 대한 업무 부담을 대폭 경감시키면서 동시에 데이터의 질에 대한 효율적인 관리도 가능할 것으로 보인다.

이와 관련하여서는 LC를 비롯한 북미 도서관계에서 실질적인 업무 지침서로 활용하는 SHM(Subject Headings Manual)을 참고해 볼 만하다.¹¹⁾ 300개 이상의 항목으로 세분화된 SHM는 업무의 일관성과 표준화를 위해 주제명 데이터의 수와 용어의 특정성, 깊이 등에 대한 기본 원칙과 함께 자료의 유형이나 매체, 언어 등에 따라 각 자료에 부여해야 할 주제명의 속성을 구체적으로 제시하고 있다. 가령, SHM에서는 문학작품을 일반 문학작품, 희곡, 소설, 우화나 전설, 시, 다양한 언어로 된 문학작품 등으로 구분하고 이들 각각에 대한 세부 지침을 제시하고 있는데, 이 가운데 소설에 대해서는 ① 자료의 형식이나 장르, ② 등장인물, ③ 배경(장소, 사건, 시간), ④ 세부 주제(소재)에 해당하는 각각의 용어를 주제명으로 표현하도록 명시하고 있다. 뿐만 아니라 개인 저자와 복수 저자가 쓴 소설, 역사소설이나 전기소설 등과 같이 자료가 가진 개별적 특성에 따른 사례도 풍부하게 제시하고 있다. 현재 LC에서는 SHM에서 제시하는 이러한 속성을 편목과정에만 활용할 뿐 주제 검색에 직접 활용하지는 않고 있지만, 주제검색에 이들 속성을 이용하여 독립적인 접근점을 제공하는 방안을 적용한다면 검색의 정확성은 물론이고 현재 우리가 가진 주제검색의 문제를 개선하는데 크게 기여할 수 있을 것이라 판단된다.

주제명 데이터 자체가 부재하거나 부실한 우리 도서관의 실정에서 이러한 세부 지침의 개발은 그 자체로 주제명 데이터의 품질 제고를 위한 ‘갯대’가 될 수 있다. 물론 이러한 지침의 개발을 위해서는 우리 이용자들이 자료의 주제나 유형별로 어떻게 접근하는지를 폭넓고 철저히 파악하는 작업이 선행되어야 할 것이며, 이러한 이용자 요구에 근거하여 우리의 실정에 적합한 주제명의 속성을 보다 세밀하게 개발해야 할 것이다. 국가대표도서관인 국립중앙도서관에서 ‘표준’ 주제명표목표의 개발에만 머물지 않고 이러한 도구의 활용을 위한 세부 지침의 마련에 앞장선다면, 현재의 OPAC이 안고 있는 주제검색의 문제를 단숨에 개선시키는 중요한 계기가 될 수도 있을 것이다. 이렇게 될 때 주제명 데이터는 비로소 ‘원하는 자료를 포괄적으로 탐색할 수 있는 장치’로서 그리고 ‘검색된 결과 중에서 적합한 자료인지를 식별하는데 도움이 되는 장치’로서의 기본적인 기능을 수행할 수 있게 될 것이며, 나아가 FRBR을 비롯하여 시맨틱 웹이나 링크드 데이터 등을 효과적으로 구현하기 위한 기초데이터로 활용할 수 있게 될 것이다.

11) 2008년부터 LC에서 발간하고 있는 『SHM』(Subject Headings Manual)는 1984년부터 2008년까지 『SCM : SH』(Subject Cataloging Manual : Subject Headings)이라는 명칭으로 발간되었다. SCM : SH는 1996년까지 5권이 발간되었으며, 이후 2008년까지 연2회 개정되었다(LC 2014).

V. 결론

지금까지 우리의 관심은 주제명표목표의 개발이나 이에 포함된 주제명의 개선에 집중되어 있었다고 해도 과언이 아니다. 이에 반해, 많은 비용과 노력을 들여 개발한 주제명표목표를 이용하여 주제명을 어떻게 부여하고 있는지에 대해서는 거의 관심을 가지지 않았다고 하여도 과언이 아니다. 이에 이 연구에서는 우리 목록레코드에 포함된 주제명 데이터의 품질에 대해 논의해 보고자 하였다. 자료에 대한 주제가 어떠한 데이터로 어떻게 구성되어 있는지, 그리고 이러한 데이터는 이용자의 주제 검색에 활용하기에 적절한지를 국립중앙도서관의 사례를 통해 미시적으로 접근해 보고자 한 것이다. 샘플 레코드의 수가 적긴 하였으나 우리나라 ‘대표’ 도서관에서 구축하고 있는 주제명 데이터를 분석한 결과, 현재의 주제명 데이터는 여러 면에서 문제를 지니고 있는 것으로 나타났다. 외형적으로 볼 때 주제명 데이터가 포함되지 않은 목록레코드는 거의 없었지만, 품질 면에서 이들 데이터가 자료가 가진 주제를 ‘제대로’ 반영한 것이라 보기에 여러모로 부족했기 때문이다.

그렇다면 이러한 결과를 초래한 원인은 어디에 있을까? 단언컨대 어느 하나가 아니라 여러 가지 원인이 복합적으로 작용한 결과일 것이다. 도구(주제명표목표)의 개발과 활용에 정도되어 본연의 목적을 망각했던 편목정책의 근시성, 주제명 데이터의 기능도 제대로 이해하지 못한 채 자료의 외형이나 표제에 포함된 일부 키워드를 기계적으로 변환하는데 급급했던 업무 관행, 자료조직 교육과정이나 도서관 현장에서의 체계적인 교육과 훈련의 부재, 그리고 자료의 주제를 분석하고 판단하는 자질과 능력의 부족 등등. 자료조직 업무 전반에 걸쳐 있는 이러한 문제는 당장 주제명 데이터의 품질 문제를 개선하기 위한 해결책으로 삼기에는 너무나도 커다란 한계가 존재한다. 따라서 이번 연구에서는 보다 현실적이면서 실무에 직접 적용할 수 있는 방안으로써 자료로부터 파악해야 할 주제의 속성을 체계적으로, 세밀하게 개발한 후 이를 실무에 적용하는 방안을 제안하였다.

이번 연구에서 이렇듯 국립중앙도서관의 주제명 데이터를 적나라하게 드러내 보인 궁극적인 목적은 이 문제의 심각성에 대해 우리 학계와 현장이 깊이 자각하고, 이를 개선하기 위한 구체적인 전략을 시급히 마련할 것을 촉구하기 위함이었다. 더불어, 지금 이 시점에서 이 문제에 대한 심층적 논의를 이끌어내지 못한다면, 주제명 데이터가 아예 누락된 현재의 목록레코드를 개선하기가 더욱 어려워질 것이라는 확신 때문이었다. 부족하지만 이 연구가 주제명 데이터의 품질에 대한 우리 도서관계의 관심을 촉발하고, 우리의 수준에 맞는 다양한 개선책을 검토하여 이를 단계별로 그리고 지속적으로 실행해 나가는 계기가 되기를 희망한다.

참고문헌

- 국립국어원. 2014. 『표준국어대사전』. <http://stdweb2.korean.go.kr/main.jsp> [cited 2015. 8. 7].
- 국립중앙도서관. 2002. 『국립중앙도서관 주제명표목표 개발』. 서울 : 국립중앙도서관.
- 국립중앙도서관. 2014. 『국립중앙도서관 주제명표목표 업무지침』. 서울 : 국립중앙도서관.
- 김태수. 2008. 『목록의 이해』. 서울 : 한국도서관협회.
- 백지원, 정연경. 2014. 국립중앙도서관 주제명표목표 검색 시스템 개선 방안에 관한 연. 『정보관리학회지』, 31(1): 31-51.
- 윤희윤. 2015. 『정보자료분류론』. 개정증보 제5판. 대구 : 태일사
- 최달현, 이창수. 2010. 『정보자료의 분류와 주제명』. 개정판. 서울 : 한국도서관협회.
- 최윤경. 2015. 『주제명 표목의 패킷 구조화 모형 개발에 관한 연구』. 박사학위논문, 이화여대 대학원 문헌정보학과.
- 한국도서관협회. 2010. 『문헌정보학 용어사전』. 서울 : 한국도서관협회.
- Albrechtsen, Hanne. 1993. "Subject Analysis and Indexing: from Automated Indexing to Domain Analysis." *The Indexer*, 18(4): 219-224.
- Delsey, T. 2005. "Modeling Subject Access: Extending the FRBR and FRANAR Conceptual Models." *Cataloging & Classification Quarterly*, 39(3/4): 49-61.
- Fairthorn 1969, 79.
- Hjørland, Birger. 1992. "The Concept of 'subject' in Information Science." *Journal of Documentation*, 48(2): 172-200.
- Hjørland, Birger. 2001. "Towards a Theory of Aboutness, Subject, Topicality, Theme, Domain, Field, Content ... and Relevance." *Journal of the American Society for Information Science and Technology*, 52(9): 774-778.
- Hoover, Lona. 2005. "A Beginners' Guide for Subject Analysis of Thesis and Dissertations in the Hard Science." *Cataloging & Classification Quarterly*, 41(1): 133-161.
- IFLA. 2002. 서지레코드의 기능상의 요건(FRBR) 개념 모형. 김태수 역. 서울 : 국립중앙도서관.
- IFLA. 2012. 주제전거 데이터의 기능 요건(FRSAD) 개념 모형. 박지영 역. 서울 : 국립중앙도서관.
- Lancaster, F. W. 저. 1999. 색인 초록의 이론과 실제. 장혜란 역. 서울 : 구미무역.
- Langridge, D. W. 1989. *Subject Analysis: Principles and Procedures*. London : Bowker-Saur.

- Library of Congress. 2015. Introduction to Library of Congress Subject Headings. <<http://www.loc.gov/aba/publications/FreeLCSH/lcshintro.pdf>> [cited 2015. 8. 9].
- Library of Congress. 2014. List of the Subject Heading Manual PDF Files. <<http://www.loc.gov/aba/publications/FreeSHM/freeshm.html>> [cited 2015. 8. 9].
- OCLC. 2013. FAST(Faceted Application of Subject Terminology). <<http://www.oclc.org/research/themes/data-science/fast.html>> [cited 2015. 9. 10].
- Rondeau, Sophie. 2014. "The Life and Times of Aboutness: a Review of the Library and Information Science Literature." *Evidence Based Library & Information Practice*, 9(1): 14-35.
- Taylor, Arlene G. and Daniel N. Joudrey. 2009. *The organization of information*. 3rd ed. Westport, Conn.: Libraries Unlimited.
- Wellisch, Bella Hass. 1988. "Why Indexing Fails the Researcher," *The Indexer*, 16(1): 3-7.

국한문 참고문헌의 영문 표기

(English translation / Romanization of reference originally written in Korean)

- Baek, Ji-Won and Yeon-Kyoung Chung. 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.
- Choe, Dal-Hyun and Lee Chang-Soo. 2010. *Classification and Subject Cataloging. Revised ed.* Seoul: Korean Library Association.
- Choi, Yoon-Kyung. 2015. Study of Developing a Facet Application Model of Subject Headings in Korea. Dissertation, Ewha Women's University.
- Kim, Tae-Soo. 2008. Understanding Cataloging. Seoul: Korean Library Association.
- National Library of Korea. 2002. *A Study on Development of National Library Subject Headings*. Seoul : National Library of Korea.
- National Library of Korea. 2014. *A Study on Improvements for High Quality in National Library of Korea Subject Headings List*. Seoul : National Library of Korea.
- Yoon, Hee-Yoon. 2015. *Library Classification : Theories and Practices*. Daegu: Taeilsa.