

국가장서 포괄성 측정 연구: 판매도서를 대상으로

Measuring the Coverage of Book Collection in the National Library of Korea

장혜란(Hye-Rhan Chang)*

홍현진(Hyun-Jin Hong)**

노영희(Younghee Noh)***

오의경(Eui-Kyung Oh)****

초 록

국가도서관은 문화유산의 일환으로 국가의 출판물을 망라적으로 수집하고 보존한다. 본 연구는 서지대조법을 활용하여 국립중앙도서관의 판매도서 포괄성을 측정하였다. 2007년과 2008년 출판 도서에 대해 교보문고 판매용목록을 기준서지로 설정하고 10%를 체계적으로 표집한 후, 국립중앙도서관 온라인목록을 탐색함으로써 수록 비율을 측정하였다. 탐색 시점에서, 2007년 출판 도서의 포괄성은 92.25%, 2008년 출판 도서의 포괄성은 78.10%, 누적 포괄성은 85.16%로 나타났다. 주제별 분석에 따라 포괄성이 낮은 분야를 알 수 있었고, 미소장도서의 출판사 규모를 분석하였으며, 국내외 선행연구 결과와 포괄성을 비교하였다. 납본지연 등 불완전한 수집의 원인이 식별되었으며, 포괄성 향상 및 후속 연구를 위하여 제언하였다.

ABSTRACT

The purpose of the national library is to collect all publications as part of the national heritage and preserve them. This study is undertaken to determine the coverage of the book collection of the National Library of Korea. Bibliographies for the year 2007 and 2008 publications are obtained from the Kyobo Book Centre and sampled 10 % systematically to be used as checklists. Coverage was determined by searching the citations against the OPAC of the National Library of Korea. Results indicate 92.25% of coverage for the books published in 2008, 78.10% coverage for the books published in 2007, and the cumulative coverage is 85.16%. The analysis shows time lag and subject fields with low coverage. Results are compared with previous studies. Reasons for the unpossessed books, such as delay of deposit and publisher characteristics, are identified. Recommendations for the coverage improvements and future research are suggested.

키워드: 포괄성, 국가장서, 국립중앙도서관, 서지대조법, 도서
coverage, national collection, National Library of Korea, bibliography checking
method, books

-
- * 상명대학교 문헌정보학과 교수(chrhan@smu.ac.kr) (제1저자)
** 전남대학교 문헌정보학과 교수(hjhong@chonnam.ac.kr) (공동저자)
*** 건국대학교 문헌정보학과 부교수(irs4u@kku.ac.kr) (공동저자)
**** 건국대학교 문헌정보학과 강의교수(ohspace@kku.ac.kr) (공동저자)
- 논문접수일자 : 2009년 11월 18일 ■ 최초심사일자 : 2009년 11월 23일 ■ 게재확정일자 : 2009년 11월 30일
■ 정보관리학회지, 26(4): 187-200, 2009. [DOI:10.3743/KOSIM.2009.26.4.187]

1. 서론

도서관장서는 도서관의 존재이유이며, 국가 도서관에서는 정보자료의 수집이 가장 중요하고 기본적인 기능이다. 일반적으로 도서관장서의 수집 범위는 도서관의 목적, 이용자 요구, 사용가능한 예산 등 여러 가지 요인에 의해 결정되지만, 국가도서관은 기본적으로 포괄적인 수집을 전제로 한다.

국가도서관에 대한 여러 가지 정의가 있으나, 비교적 최근에 출판된 ISO 2789에서는 다음과 같이 서술하였다(ISO 2006):

국가도서관은 자국의 적합한 문헌을 모두 입수하고 보존할 책임이 있는 도서관으로 납본도서관의 역할을 수행할 수 있다. 또한 국가도서관은 일반적으로 국가서지 생산, 자국에 대한 문헌을 포함한 외국문헌의 소장과 최신성 유지, 국가서지정보센터의 역할, 종합목록 편찬, 다른 도서관행정의 관리 및 협력 촉진, 연구개발 서비스의 조정 등의 기능 중 일부 또는 전체를 수행한다.

우리나라의 국가도서관은 국립중앙도서관이며, 국립중앙도서관은 도서관법에 의해 국가대표도서관으로서의 법적 지위를 보장받고 있다. 따라서 국립중앙도서관은 국가장서를 구축하고 보존하여 정보를 활용하게 함으로써 지식사회를 선도하고 나아가 미래 세대에 문화유산의 전승할 의무와 책임을 가지고 있다. 이러한 사명을 완수하기 위해 국립중앙도서관은 일차적으로 우리나라에서 생산되는 모든 출판된 자료를 망라적으로 수집해야 한다.

여러 해 전부터, 세계의 국가도서관들이 도

서관 성과의 품질을 평가하기 위한 지표에 대한 일반적 동의의 필요성을 인식하여 왔으며, 관련 기관에서는(예, IFLA Section of National Libraries, Conference of European National Libraries 등) 노력을 경주하였다.

Poll과 Jonsson-Adrial(2006)은 국가도서관에 대하여, 국가장서 구축, 서비스 접근, 참고봉사 제공, 장서보존, 경영효율성, 발전을 위한 잠재력 등 6개 부문에 대한 지표를 제안하였는데, 여기에는 국가장서 구축에 대한 세부 지표 중 국가출판물에 대한 포괄성 측정이 의무사항(proposal status, "obligatory")으로 표시되어 있다.

2009년 4월, 세계표준기구는 ISO/TR 28118: Information and documentation - performance indicators for national libraries(ISO 2009)를 출판하였다. 이것은 기존의 도서관 성과지표 기준, ISO 11620 으로부터 국가도서관에 적합한 성과지표를 선택하였고, 또한 국가도서관에서 사용되고 검증되었으나 ISO 11620에서 고려되지 않았던 새로운 지표를 포함하고 있다. 국가도서관 성과지표는 A1부터 A9까지 9가지 활동 부문에 대한 30종의 지표로 구성되어 있다. 그 중 A1은 국가장서 구축에 관한 부문으로 다음과 같은 두 가지 지표로 구성되어 있다: 국가의 출판물 중 국가도서관이 수집한 자료의 비율(A.1.1), 이용자가 요청한 자료 중 국가도서관에 소장된 자료의 비율(A.1.2).

국가도서관에 대한 가장 기본적인 성과 지표는, 해당 국가에서 생산되는 출판물에 대한 장서의 포괄성이다. 국가장서의 포괄성은 국가에서 생산된 출판물(표본)수와 국가도서관 소장 자료의 수를 근거로 하며, 서지대조법을 사용

하여 측정한다. 측정결과는 도서관이 국가 지식을 적절히 소장하고 있는지 여부를 나타내며, 또한 납본도서관의 경우에는 납본업무의 이행 절차가 효과적인가를 나타낸다.

본 연구의 목적은 판매도서에 대한 국립중앙도서관 장서의 포괄성을 측정하기 위한 것이다. 국립중앙도서관에 의해 수집된 자료는 목록데이터베이스에 수록되며, 국립중앙도서관은 연보를 통해 매년 자료 유형 및 수집 방법에 따른 장서 수에 대한 통계를 상세하게 보고하고 있으나, 우리나라에서 출판되어 판매되는 자료 중 어느 정도가 실제로 수집되었지, 즉 국가장서의 판매도서 포괄성에 관한 사항은 찾아보기 어렵다.

이와 같은 포괄성 측정 연구는 국가장서 데이터베이스의 품질을 평가하는 것이며, 연구결과는 국립중앙도서관의 도서 수집 및 제공 능력을 나타내고 포괄성이 낮은 주제 분야를 식별함으로써, 장차 수집업무 개선에 기여할 수 있다. 포괄성 측정은 또한 다른 나라 국가도서관의 성과와 비교할 수 있는 정보를 제공하며, 측정대상 기간에 출판된 문헌에 대한 국가장서 데이터베이스 최신성에 대한 정보를 부가적으로 제공할 수 있을 것이다.

2. 연구 방법과 내용

장서의 포괄성에 대한 올바른 평가를 위해서는, 객관적인 기준과 방법이 강구되어야 한다. 우리나라 국가장서 포괄성을 분석하기 위해서, 장서평가에서 보편적으로 사용되며, ISO TR/28118에서 국가도서관 포괄성 측정 방법으로

제시하고 있는 서지대조법을 사용하였다. 서지대조법은 평가의 기준을 신뢰할만한 서지류에 두고, 해당 도서관이 서지에 포함되어 있는 문헌을 어느 정도 소장하고 있는가를 조사한다. 측정 결과는 비율(%)로 나타낸다.

2.1 기준서지의 설정

서지대조법에서 중요한 측면은 기준이 될 적합한 서지를 선정하는 것이다. 서지가 발전하지 못한 환경에서는 연구 목적에 맞는 서지를 발견하기가 어렵다. 또한 기준서지는, 편찬될 때, 평가될 데이터베이스를 참고로 하지 않았다는 전제조건이 충족되어야 한다. 만약 적합한 서지가 없다면, 연구에 사용할 목적으로 서지를 편성하여 기준서지로 사용할 수도 있다.

포괄성 측정을 위한 기준서지를 설정하기 위하여, 대한출판문화협회에서 발행하는 『한국출판연감』의 사용 가능성에 대하여 검토하였다.

『한국출판연감』은 매년 출판되는 국내 유일의 출판 종합 자료집으로서 6편 〈목록편〉에서 한 해 동안의 도서 목록을 한국십진분류법(KDC)에 의해 분야별로 정리하여 수록하고 있다. 그러나 수록된 목록은 납본된 데이터를 근거로 집계하였으므로, 국내 유통 출판물 전체에 대한 데이터로 볼 수 없으며(http://www.kpa21.or.kr/bbs/board.php?bo_table=d_total), 또한 대한출판문화협회는 현재 국립중앙도서관의 납본대행 기관으로서 『한국출판연감』 수록 자료는 거의 모두 국립중앙도서관에 수집될 것으로 추정되었다.

망라적 상업서지가 존재하지 않으므로, 국내 대형서점 목록의 사용 가능성을 검토하였다. ISO

TR 28118에는 포괄성 측정의 구체적 방법으로 상업적 출판물만을 대상으로 하는 방법, 상업적 출판물과 비상업적 출판물을 포함시키는 방법, 상업적 출판물과 비상업적 출판물을 분리하여 장서 포괄성을 측정하는 방법을 제시하였다. 우리나라의 주요 대형서점으로는, 교보문고, 영풍문고, 리브로 등이 있다. 1981년 대한교육보험(주)의 자회사로 설립된 교보문고(<http://www.kyobobook.co.kr>)는 각종 도서정보를 발간하고 산업정책연구원 슈퍼브랜드를 4년 연속 수상하였으며, 총 자산 1,838억 원, 2007년도 매출액 3,976억에 달하는 대기업 수준의 국내 최대 서점이다. 영풍문고(<http://www.ypbooks.co.kr>)는 1992년 설립되어, 2007년 현재 총 자산 853억 원, 매출액 1,340억 원에 이르며, 2003년 1억불 수출의 탑, 2005년 2억불 수출의 탑, 2008년 5억불 수출의 탑을 각각 수상하기도 하였으며, 교보문고에 이어 국내 2위 수준의 서점이다. 또한, 리브로(www.libro.co.kr)는 2000년 을지서적을 인수한 이후, 2007년 현재 총 자산은 264억, 매출액 805억의 인터넷 서점으로도 유명한 후발 서점이다(매일경제신문사 2009, 두산백과사전 엔사이버 <http://www.encyber.com>). 주요 대형서점들의 판매용 목록데이터베이스는 현재 국내에서 유통되는 도서를 망라적으로 수록하여, 기준서지로 사용하기에 적합한 것으로 판단하였다. 모든 서점들의 목록데이터베이스를 통합하여 기준서지를 편찬하는 것이 이상적인 방법이지만, 주어진 연구 환경에서 현실적인 어려움이 있어서, 가장 유통 규모가 큰 교보문고의 목록데이터베이스를 기준서지의 근거로 사용하였다.

기준서지로 사용하기 위해서는 교보문고에

서 유통되는 모든 국내 도서의 서지데이터가 필요하였으나, 이러한 데이터를 수록한 목록은 제작 배포되지 않고 있다. 적절한 서지를 얻기 위해서, 교보문고 홈페이지 상에서 발행일을 통제하여 탐색함으로써, 기준서지의 기초 데이터를 산출하고자 시도하였다. 그러나 교보문고의 온라인 자료탐색시스템은, 외부에서 이용할 경우, 탐색어를 반드시 한 가지 이상 입력해야만 탐색결과를 얻을 수 있으므로 망라적 서지를 산출하는 것이 불가능하였다. 따라서 교보문고 홍보팀에 직접 협조를 구하여, 교보문고 내부시스템에서 출판연도로 서지데이터를 제한함으로써 연구에 적합한 서지를 엑셀 파일로 제공받았다. 이 파일에는 교보문고의 자체 분류 기준에 따른 22개 주제별로 레코드가 수록되어 있고, 각 주제 내에서는 발행일자순으로 배열되어 있다. 만화, 아동, 어린이영어, 유아, 외국어 등 5개 주제는 대부분이 국립어린이청소년도서관 소장 자료인 것으로 조사되어, 측정에서 제외하였다. 국립어린이청소년도서관이 국립중앙도서관의 분관이고, 국립중앙도서관 목록데이터베이스 탐색 시 미소장의 경우, 국립어린이청소년도서관으로의 추가 탐색을 공지하는 메시지가 제공되고 있으나, 본 연구는 국립중앙도서관 홈페이지에서 탐색되는 것을 기준으로 삼았고, 이러한 주제 영역의 서지들이 대조과정에서 미소장으로 처리되어 포괄성 측정의 신뢰도와 타당도를 저하시키는 요인이 될 수 있으므로, 기준서지에서 제외하였다.

〈표 1〉은 대한출판문화협회(출협)의 2007년도와 2008년도 『출판통계』(http://kpa21.or.kr/bbs/board.php?bo_table=d_total)에서 학습참고, 아동, 만화 부문을 제외한 신간 발행 중

〈표 1〉 출협 출판통계와 기준서지의 수록 종수 비교

구 분	수 록 종 수	
	2007년	2008년
출협, 출판통계	24,748	26,354
기준서지	38,481	38,570

수와 기준서지로 사용하기 위해 편성된 2007년과 2008년 출판 대상 교보문고 판매도서 파일의 종수를 비교한 것이다.

장서나 데이터베이스의 포괄성을 평가하기 위하여 완벽한 서지를 필요로 하는 것은 아니다. 필요한 것은 대표할 만한 문헌의 표집이다 (Lancaster 1999, 181). 따라서 교보문고의 서지가 우리나라에서 출판된 도서 전체를 포함하지 않는다는 점은 문제가 되지 않으며, 다만 우리나라에서 출판된 판매도서를 대표할 수 있는 서지라는 점에서 기준서지로 설정하기에 무리가 없다.

2.2 표집 방법

장서 포괄성 측정을 위해서는, 확보된 서지 정보를 접근점으로 하여 국립중앙도서관의 소장 여부를 확인하고 소장 비율을 산출해야 한다. 그러나 전수조사는 보통 시간적·경제적 이유로 불가능하므로, 모집단으로부터 추출된 표본을 대상으로 조사한다. 이 경우 표본의 대표성이 중요하다.

일반적으로 표본의 추출은 확률추출법과 비확률추출법으로 구분되는데, 전자가 좀 더 과학적이고 유용하며, 모집단을 정확히 대표할 수 있다. 확률추출방법에는 무선표집, 체계적표집, 층화무선표집, 군집표집 등이 있다. 본 연구

에서는 조사를 위해 체계적표집 방법을 채택하였다. 체계적표집은 모집단의 리스트가 긴 경우, 무선표집보다 간단하고 편리하다는 장점을 갖는다. 자모순 리스트와 같이 리스트에 어떤 패턴이 존재하지 않는 경우, 체계적표집은 모집단을 대표한다는 점에서 대체로 만족스럽다고 볼 수 있다.

체계적표집에서는 전체 리스트 중에서 매 n 번째 나타나는 요소들을 차례로 선택해 나가게 된다. 첫 번째 추출을 위한 출발점은 임의로 결정되고 마지막 추출은 첫 번째 추출의 바로 전 구간에서 이루어지도록 되어 있기 때문에 리스트가 고리모양을 이루게 된다. 본 연구에서는 서지대조작업을 위해 기준서지의 10%에 대해 체계적 표본을 추출하여 조사를 수행하였다.

2.3 데이터수집 방법과 기간

서지대조 작업은 온라인 탐색을 통하여 수행되었고, 1차로 2008년 출판 도서에 대하여 조사하였다. 기준서지에 수록된 도서의 10%를 체계적 표집한 3,857건에 대하여 2009년 5월 1일부터 5월 10일까지 국립중앙도서관 온라인열람목록을 탐색하여 소장 여부를 조사하였다.

탐색에는 서명, 저자명, 출판사명, ISBN 등 가능한 모든 탐색 접근점을 철저히 사용하였고, 미소장 자료인 경우에는 정확성을 위해서 재차

탐색하였다.

조사 시점을 기준으로 볼 때(1차 조사: 5월 1일), 2008년 출판도서에 대한 탐색은, 출판 후 4개월~16개월의 시차를 허용하고 있다. 국립중앙도서관 자료수집방법 중 가장 큰 비율(약 70% 이상)을 차지하는 납본에 의한 자료를 기준으로 했을 때, 법정 납본기한이 발행 후 30일 이내임을 감안하면, 본 조사가 허용하는 시차는 국립중앙도서관에 입수되기에 충분한 시간으로 간주되었다.

1차 조사의 결과 포괄성 수치가 다소 낮은 것으로 나타나, 출판 년도를 2007년으로 소급하여 2차 조사를 수행하였다. 2차 조사는, 2009년 10월 12일부터 10월 29일까지, 기준서지 10%를 표집한 3,848건에 대하여 탐색과 대조 작업을

하였다. 2차 조사의 시점을 기준으로 볼 때, 2007년 출판 도서에 대하여 출판 후 22개월~35개월의 시차를 허용하고 있다.

3. 데이터 분석 결과

3.1 2008년 출판 도서에 대한 포괄성

2008년 출판 도서에 대하여 측정된 국립중앙도서관의 포괄성은 <표 2>와 같다. 체계적 표집에 의하여 추출된 3,857건의 표본 중 중 62건은 다이어리, 사은품, 비도서, 부정확한 서지사항 등의 이유로, 탐색과 대조과정에서 추가로 제외시키고, 최종적으로 3,795건에 대하여 탐

<표 2> 2008년 출판 도서에 대한 포괄성

주제 구분 (교보문고 자체분류)	기준서지 표본 (건 수)	국립중앙도서관 목록 수록 (건 수)	포괄성 (%)
가정/생활	47	36	76.60
건강	58	47	81.03
경영/경제	353	274	77.62
과학	91	82	90.11
기술/공학	345	267	77.39
소설	652	580	88.95
시,에세이	304	245	80.59
여행	69	54	78.26
역사	60	46	76.66
예술	207	153	73.91
요리	34	23	67.64
인문	483	378	78.26
자기계발	116	106	91.38
정치/사회	482	358	74.27
종교	409	261	63.81
취미/스포츠	69	40	57.97
컴퓨터	16	14	87.50
전 체	3,795	2,964	78.10

<기준서지: 교보문고 작성 2008년 출판 도서목록, 탐색일: 2009년 5월 1일 ~ 5월 10일>

색하고 대조한 결과, 2,964건이 국립중앙도서관 온라인목록에 포함되어 있는 것으로 나타났다. 전체적인 포괄성은 78.10%이다.

기준서지는 총 17개 분야로 구분되어 있다. 주제분야별로는 '자기계발'이 91.38%로 포괄성이 가장 높았으며, 그 다음으로 '과학(90.11%)', '소설(88.95%)', '컴퓨터(87.50%)', '건강(81.03%)', '시·에세이(80.59%)', '인문(78.26%)', '여행(78.26%)'의 순이며, 전주제의 평균 포괄성 78.10%에 도달하지 못한 주제는 '경영/경제(77.62%)', '기술/공학(77.39%)', '역사(76.66%)', '가정/생활(76.60%)', '정치사회(74.27%)', '예술(73.91%)', '요리(67.64%)', '종교(63.81%)', '취미/스포츠(57.97%)' 등이다.

3.2 2007년 출판 도서에 대한 포괄성

자료의 발행일로부터 탐색일까지의 시차를 더 길게 허용하였을 때, 수집이 보다 완전해지는지 알아내기 위하여, 2007년 출판 도서로 소급하여 포괄성을 조사하였다. 1차 조사와 마찬가지로, 10%의 표본 중에서 적합하지 않는 서지엔트리 67건을 발견하여 제외함으로써, 대조작업은 3,781건에 대하여 수행되었다.

다음의 <표 3>은 2007년 출판도서에 대한 국립중앙도서관 온라인목록 포괄성 측정 결과이다.

<표 3>에서 나타난 바와 같이, 전체적인 포괄성은 92.25%이다. 2008년 출판 도서에 대한 포괄성 78.10% 보다 14.15% 증가하였다. 주제별로 모든 부문에서 포괄성의 증가가 있었으며,

<표 3> 2007년 출판 도서에 대한 포괄성

주제 구분 (교보문고 자체분류)	기준서지 표본 (건 수)	국립중앙도서관목록 수록 (건 수)	포괄성 (%)
가정/생활	45	39	86.67
건강	61	56	91.80
경영/경제	308	290	94.16
과학	82	81	98.78
기술/공학	313	294	93.93
소설	645	610	94.57
시,에세이	328	298	90.85
여행	57	51	89.47
역사	124	120	96.77
예술	199	172	86.43
요리	32	29	90.63
인문	405	370	91.36
자기계발	141	133	94.35
정치/사회	418	398	95.22
종교	399	340	85.21
취미/스포츠	61	52	85.24
컴퓨터	163	155	95.09
전 체	3,781	3,488	92.25

<기준서지: 교보문고 작성 2007년 출판 도서목록, 탐색일: 2009년 10월 12일 ~ 10월 19일>

증가의 범위는 최소 2.97%(자기계발)에서 최대 27.27%(취미/스포츠)까지 다양하다.

주제분야별로는 '과학(98.78%)', '역사(96.77%)', '정치/사회(95.22%)', '컴퓨터(95.09%)' 등의 포괄성이 95% 이상으로 높고, 그 다음으로 '소설(94.57%)', '자기계발(94.25%)', '경영/경제(94.16%)', '기술/공학(93.93%)', '건강(91.80%)', '인문(91.36%)', '시·에세이(90.85%)', '요리(90.63%)' 등이 90% 이상이며, 상대적으로 포괄성이 낮은 주제는 '가정/생활(86.67%)', '여행(89.47%)', '예술(86.43%)', '취미/스포츠(85.24%)', '종교(85.21%)' 등이다. 2008년 출판 도서에 대한 측정결과와 비교하면, 주제별 포괄성의 순위는 포괄성이 높은 부문에서는 차이가 있으나, 포괄성이 낮은 예술, 종교, 취미/스포츠 부문에서는 유사하다고 볼 수 있다.

소급조사를 통하여, 시간이 경과됨에 따라 국립중앙도서관의 수집률이 보다 충실해짐을 알 수 있다. 이것은 도서가 출판된 후 국립중앙도서관에 수집되어 목록에 탑재될 때까지, 상당한 시간적 지연이 있다는 것, 즉 국가장서 데이터베이스의 최신성에 문제가 있음을 나타내고 있다. 국립중앙도서관 장서에서 납본이 차지하는 비율을 감안할 때, 상업적 출판물의 납본 과정에 비효율성이 존재함을 유추할 수 있다.

3.3 2007년~2008년 출판 도서에 대한 누적포괄성

두 차례의 조사 결과, 탐색 시점에서 과거로 소급할수록 포괄성이 크게 증가하는 것을 알 수 있다. 부문별 문제점을 파악하기 위해서는 1년 단위의 분석보다는 누적 데이터가 신뢰할

수 있을 것으로 판단하여, 두 차례의 조사 결과를 합산하여, <표 4>와 같이 누적포괄성을 산출하였다.

<표 4>에서 보는 바와 같이 누적포괄성은 85.16%로 측정되었다. 주제별로는 '컴퓨터'가 94.41%로 가장 높고, 다음으로 '과학(94.21%)', '자기계발(93.00%)', '소설(91.75%)', '역사(90.22%)', '건강(86.55%)', '시·에세이(85.92%)', '경영/경제(85.33%)', '기술/공학(85.26%)'의 순이며, '인문(84.23%)', '정치/사회(84.00%)', '여행(83.33%)', '가정/생활(81.52%)', '예술(80.04%)', '요리(78.79%)', '종교(74.38%)', '취미/스포츠(70.77%)'는 평균에 미치지 못한 것으로 분석되었다. 가장 포괄성이 높은 '컴퓨터'와 가장 낮은 '취미/스포츠'간의 차이는 23.44%에 달한다.

최소 4개월부터 최대 35개월의 시차가 허용된 점을 감안할 때, 85.16%의 포괄성은 불완전한 수집 양상을 보이고 있으며, 주제별 포괄성에 대한 편차와 포괄성이 일관성있게 낮은 주제의 식별은 장서 수집상의 구조적인 문제가 내재되어 있을 가능성을 나타내고 있다.

미소장 도서에 대해 납본 대상 기관인 출판사별 특성을 파악하기 위하여, 미소장 도서의 리스트를 출판사명으로 정렬하여 분석하였다. 결과는 <표 5>와 같다.

<표 5>에서 나타난 것을 보면, 2007-2008년 출판 미소장도서 대상 출판사는 모두 858개 사에 달하며, 그 중 82.40%에 해당하는 707개 출판사가 미소장도서 1건을, 10.72%에 해당하는 92개 출판사가 미소장도서 2건을 가지고 있다. 미소장도서 1~2건을 가지고 있는 출판사의 미소장도서 수는 전체의 79.27%를 차지하고 있다.

〈표 4〉 2007-2008년 출판 도서에 대한 누적포괄성

주제 구분 (교보문고 자체분류)	기준서지 표본 (건 수)			국립중앙도서관 목록 수록(건 수)			누적 포괄성 (%)
	2007년	2008년	소계	2007	2008	소계	
가정/생활	45	47	92	39	36	75	81.52
건강	61	58	119	56	47	103	86.55
경영/경제	308	353	661	290	274	564	85.33
과학	82	91	173	81	82	163	94.21
기술/공학	313	345	658	294	267	561	85.26
소설	645	652	1,297	610	580	1,190	91.75
시.에세이	328	304	632	298	245	543	85.92
여행	57	69	126	51	54	105	83.33
역사	124	60	184	120	46	166	90.22
예술	199	207	406	172	153	325	80.04
요리	32	34	66	29	23	52	78.79
인문	405	483	888	370	378	748	84.23
자기개발	141	116	257	133	106	239	93.00
정치/사회	418	482	900	398	358	756	84.00
종교	399	409	808	340	261	601	74.38
취미/스포츠	61	69	130	52	40	92	70.77
컴퓨터	163	16	179	155	14	169	94.41
계	3,781	3,795	7,576	3,488	2,964	6,452	85.16

〈표 5〉 미소장 도서 건별 출판사 수와 비율

미소장 도서 (건 수)	출판사 수(기관)				미소장 도서	
	2008년 출판	2007년 출판	전체	비율 (%)	계 (건 수)	비율 (%)
1	511	196	707	82.40	707	62.90
2	64	28	92	10.72	184	16.37
3	25	7	32	3.73	96	8.54
4	12	3	15	1.75	60	5.34
5	6	0	6	0.70	30	2.67
6	2	0	2	0.23	12	1.07
7	1	0	1	0.12	7	0.62
8	0	1	1	0.12	8	0.71
9	0	0	0	0.00	0	0.00
10	2	0	2	0.23	20	1.78
계	623	235	858	100.00	1,124	100.00

대한출판문화협회 출판통계 『2007년 출판사 간행물 실적 현황』을 보면, 전체 2,333개 출판

사 중에서, 1종을 발행한 출판사는 18.7%에 해당하는 437개 사이며, 2종을 발행한 출판사는

12.0%에 해당하는 281개 사로 나타나 있다(http://kpa21.or.kr/bbs/board.php?bo_table=d_total).

불완전한 수집이 출판사의 영세성에 기인하는지를 조사하기 위하여, 출판유통진흥원의 통계정보서비스를 이용하였다(www.booktrade.or.kr/statistics). 출판유통진흥원은 출판유통현대화를 위한 출판유통 표준 연구, 서지·유통 DB 및 정보시스템 개발 구축, 정보 교류 및 교육 등 공공 성격의 사업을 추진하고 있으며, 실시간 출판통계서비스 등의 부가서비스도 제공하고 있는 사단법인이다(<http://www.booktrade.or.kr/overview/ceo.jsp>). 출판통계서비스는 출판사별, 분야별, 내용분류별로 제공하고 있으며, 출판사별 통계의 경우, 기간을 설정하고, 출판사명을 입력하면, 해당기간 동안의 출판실적을 알 수 있다(www.booktrade.or.kr/statistics_by_publisher.jsp).

미소장도서가 1~2건으로 나타난 출판사들에 대한 출판실적을 조사하여 분석한 결과가 <표 6>에 나타나 있다. 탐색 결과, 2007년 해당 미소장 출판사 중 224개사(74.11%)와 2008년 해당 미소장 출판사 중 361개사(62.78%)에 대한 도서출판 실적을 찾을 수 있었다. 출판유통진흥원 통계에 의하면 우리나라 출판사의 평균

출판실적은 2007년에는 13.44종, 2008년에는 13.12종이다. 2007년 미소장도서 출판사들의 실적을 조사한 결과, 최소 1종의 도서를 출판한 회사로부터 최대 833종의 도서를 출판한 회사까지 실적이 상당히 분산되어 있는 것으로 나타났다. 따라서 출판실적 100건이 넘는 이상값을 가진 출판사 12개사를 제외하고 평균과 사분편차를 산출하였다. 2008년 미소장도서 출판사들의 실적도 최소 1종의 도서를 출판한 회사로부터 최대 636종의 도서를 출판한 회사까지 분산되어 있어, 100건이 넘는 이상값을 가진 9개사를 제외하고 평균과 사분편차를 산출하였다. 미소장도서 출판사들은 2007년에는 평균 15.34종, 2008년에는 평균 15.33종의 도서를 출판한 것으로 나타났다. 이 수치는 출판사 전체에 대한 평균치를 상회한다. 남분이나 구입에서 누락되었던 미소장도서가 실적이 낮은 소규모 출판사의 도서가 아니라는 사실을 알 수 있다.

3.4 국내외 선행연구 결과와 비교분석

일반적으로 포괄성이 높을수록 우수한 장서로 간주된다. 본 연구에서 측정한 국가장서 포괄성의 세 가지 수치, 즉 2008년 출판 판매도서에 대한 포괄성 78.10%, 2007년 출판 판매도서

<표 6> 미소장 도서 출판사의 출판 실적 분석

실적 년도	미소장 도서 출판사				출판유통진흥원 수록 출판사	
	출판사 수	도서 수(종 수)	평균치(종 수)	분포		평균치(종 수)
2007	154*	2,362	15.34	범위	91	13.44
				사분편차	10.60	
2008	352**	5,397	15.33	범위	95	13.12
				사분편차	12.02	

* 이상값 12개 제거 후의 값임. ** 이상값 9개 제거 후의 값임.

에 대한 포괄성 92.25%, 2007-2008 누적 포괄성 85.16%는 선행 연구 및 외국의 포괄성 측정 결과와 비교해 볼 수 있다.

한국출판협동조합 “판매도서목록”과 대한출판문화협회 “한국출판연감”을 기준서지로 하여 국립중앙도서관목록의 포괄성을 측정했던 장혜란(1994)의 연구에서는, 유통도서 전반에 대한 포괄성은 69.83%, 1993년 출판도서에 대한 포괄성은 79.86%로 나타났다. 비록 데이터 수집 방법 및 내용이 동일하지는 않지만, 15년이 경과한 시점에서, 포괄성이 크게 향상되었음을 알 수 있다. 주제별 분석을 보면, 선행 연구에서는 문학이 81.02%로 가장 높은 수치를 보였는데, 이것은 본 연구에서 ‘소설’의 높은 포괄성(91.75%)과 상응하는 결과로 볼 수 있다. 한편 선행 연구에서는 순수과학과 기술과학에 대한 포괄성이 평균 수준이었으나, 본 연구에서는 일관성 있게 평균치 보다 높이가 나타나, 과학 기술 부문 도서의 수집률이 크게 향상된 것을 알 수 있다. 또한, 선행연구에서 1994년 출판 도서 포괄성을 별도로 측정하여(40.00%) 목록데이터베이스 최신성을 추정해 보았는데, 비록 본 연구의 허용시차와는 차이가 있으나, 국립중앙도서관 목록데이터베이스의 최신성도 과거보다 크게 향상된 것을 알 수 있다.

국가도서관은 역사와 전통이 각각 독특하므로 포괄성의 단순 비교는 가능하지 않다. 국가도서관은 각 나라 고유의 기관으로, 역사와 발전, 규모, 자료수집 범위, 기능, 서비스 등에서 다양하며 도서관의 성과는 쉽게 비교할 수 없다. 1909년 이후 100년에 걸쳐 저작권을 부여하며 납본 업무를 수행해 온 미국 국회도서관이나 또는 전 세계 지적 자원의 망라적 수집을

목표로 하는 영국 국립도서관과 기타 다른 나라의 도서관들의 성과를 단순비교하여 평가할 수는 없을 것이다. 그러나 선진국 대열에 있는 우리나라 국립중앙도서관과 외국 국가도서관의 포괄성의 비교는 측정 결과의 해석과 향후 개선을 위해 유의할 것으로 보인다.

영국 국립도서관의 홈페이지에서 제공하는 연간보고서(Annual Report 2005/06)에는 당해 년도의 “납본율”(percentage of UK legal deposit material acquired)이 97%로 명시되어 있다(British Library 2006).

네덜란드 국가도서관은 2003년 이 “납본 수행 출판사 비율”(percentage of publishing companies covered)라는 지표를 사용한다. 이것은 Royal Dutch Book Trade Association 회원 출판사 중 해당 연도에 국가도서관에 납본한 출판사의 비율이다. 측정 결과는 2004년에는 98%, 2005년에는 97%를 달성하였고(Koninklijke Bibliotheek 2005), 2008년에는 96%를 달성하였다(Koninklijke Bibliotheek 2008). 한편 2005년 전반기에 수행된 한 연구에서는 2003년 출판물에 대하여 ISBN을 갖고 있는 도서와 ISBN이 없는 도서/정기간행물을 분리하여 조사하였다(Voorbij and Lemmen 2006). ISBN을 갖고 있는 도서에 대한 포괄성은 95%, ISBN이 없는 도서에 대한 포괄성은 70%로 나타났다.

독일의 국가도서관은 독일국가서지에 수록된 출판물의 비율을 유형별로 측정하였다. 2005년과 2006년 2년에 대한 수록 비율은 상업적 출판물에 대해서 91.2%, 비상업적 출판물에 대해서는 78.2%로 나타났다(ISO 2009, 20).

일본의 國立國會圖書館은 2007년 10월부터 12월까지 2달 동안에, 2005년 출판 자료에

대한 수집비율을 정부간행물과 민간출판물, 판매자료와 비매자료 등으로 나누어서 측정하였다(國立國會圖書館 2008). 그 결과를 보면 시판 자료 전반에 대한 포괄성은 89.8%, 민간출판(상업적) 도서에 대한 포괄성은 88.0%로 나타나 있다.

위에서 열거한 외국 국가도서관의 포괄성 측정치를 비교할 때는, 지표의 정의와 계산 방식을 고려해야 한다. 측정 내용과 조사 시점은 상이하지만, 영국, 네덜란드, 독일 등은 상업적 출판물에 대해 90% 이상의 포괄성 내지는 납본율을 달성하고 있으며, 조사 국가 중 포괄성이 비교적 낮은 일본의 경우도 민간출판 판매도서에 대하여 88.0%의 포괄성을 달성하고 있다.

4. 결론 및 제언

본 연구는 서지대조법을 적용하여 국가도서관 장서 중 판매도서의 포괄성을 측정함으로써 국가도서관의 수집 현황을 평가하였다. 서지대조 작업을 위한 기준서지로 2007년과 2008년 출판 도서에 대한 교보문고의 목록파일을 구하여 10%를 체계적으로 표집하였고, 표본 레코드를 직접 국립중앙도서관 온라인목록에서 탐색하고 소장여부를 조사하여 포괄성을 측정하였다.

주요 결과와 시사점은 다음과 같다.

첫째, 2008년 출판 판매도서에 대한 포괄성은 78.10%, 2007년 출판 판매도서에 대한 포괄성은 92.25%이며, 2007-2008 2년간의 누적포괄성은 85.16%이다. 이 수치는 과거에 비하여 상당히 개선된 양상을 나타내고 있으나, 선진

국에 비하여 여전히 낮은 수준이며, 향후 포괄성 향상을 위한 다각적인 노력이 강구되어야 할 것이다.

둘째, 도서 출판년도와 탐색일자 간의 시차가 커짐에 따라 포괄성이 크게 증가되었으며, 이것은 모든 주제 분야에 공통적인 현상이다. 이것은 국가장서 데이터베이스의 최신성에 대한 문제를 지시하고 있으며, 탐색 시점까지의 충분한 허용시차에도 불구하고 비교적 낮은 포괄성은, 국립중앙도서관 장서 수집의 가장 큰 비중을 차지하는 납본이 지연되고 있음을 나타내고 있다. 납본, 미납본 도서에 대한 청구, 구입 등 자료수집 속도의 효율성을 제고시킬 필요가 있다. 최신성의 문제는 국가서지의 생산과 배포 및 활용에도 지대한 영향을 미친다.

셋째, 주제별 분석을 통하여 포괄성이 일관성 있게 낮은 수집 취약 분야가 식별되었다. 취미/스포츠, 종교, 요리, 예술, 가정/생활, 여행, 등의 주제 분야 누적포괄성의 범위는 70.77% - 83.33%에 있다. 이 분야의 도서는 내용의 본질과 수준 그리고 형태면에서 다양하다. 국립중앙도서관은 이 분야 도서의 내용별 형태별 상세 분석을 통해 수집정책을 정교화 하는 한편 수집대상 자료와 수집제의 자료를 명백히 함으로써 포괄성을 향상시킬 수 있을 것이다.

넷째, 미소장도서의 출판사 조사 결과를 보면, 미소장도서가 1~2건인 출판사는 전체의 93.12%인 799처에 달한다. 미소장도서 출판사의 도서출판 실적 분석 결과를 보면(2007년 평균 15.34종, 2008년 평균 15.33종), 우리나라 출판사 평균 실적(2007년 13.44종, 2008년 13.12)을 약간 상회하는 수준으로 나타났다. 납본이나 구입에서 누락된 미소장도서가 소규모 출판

사의 도서에 국한된 것이 아님을 알 수 있다. 국립중앙도서관은 출판계 전체를 대상으로 출판물 모니터링 업무를 강화할 필요가 있을 것으로 생각된다.

후속 연구를 위하여 다음과 같은 제언을 하고자 한다.

장서의 포괄성은 국가도서관에 있어 가장 중요한 평가 지표이다. 이러한 포괄성 측정 연구는 일회적인 것이 아니라 정례화된 업무로서 지속적으로 수행되어야 할 것이다. 선진국의 사례를 보면, 국가도서관에서 국가장서의 완전성을 나타내는 포괄성을 측정하여 공식적으로

연간보고서나 홈페이지 상에 제공하고 있으며, 보다 발전된 경우에는, 매년 예상 목표치에 대한 실적을 공표하고 있다. 최근 가까운 일본의 경우에도 국립국회도서관 수집기획위원회에서 당관의 직원으로 구성된 위원회를 만들어 국내 출판물 납본율을 조사하여 연보에 공표한 바 있다(國立國會圖書館 58-59). 본 연구는 판매도서에 대한 포괄성을 측정한 것이다. 도서뿐만 아니라 정보자료의 유형별로 국내 생산 자료의 포괄성을 측정함으로써 수집 현황을 평가하고 개선한다면, 보다 완전한 국가장서 구축에 기여할 수 있을 것이다.

참 고 문 헌

- 국립중앙도서관. 2009. 『국립중앙도서관연보 2008』. 서울: 국립중앙도서관 [cited 2009.4.30]. <http://www.nl.go.kr/pds/research_data/yunbo.php>.
- 대한출판문화협회. 2009. 『한국출판연감 2008』. 서울: 대한출판문화협회.
- 매일경제신문사. 2009. 『회사연감 2009』. 서울: 매일경제신문사.
- 장혜란. 1994. 상업서지를 이용한 국립중앙도서관 목록데이터베이스 평가. 『정보관리학회 전국논문대회 논문집』, 1: 147-150.
- 한국. 문화체육관광부. 2009. 도서관법 [일부개정 2009.3.25 법률 제9528호]. <http://likms.assembly.go.kr/law/jsp/Law.jsp?WORK_TYPE=LAW_BO N&LAW_ID=A0742&PROM_NO=09528&PROM_DT=20090325&HanChk=Y>.
- 國立國會圖書館. 2008. 『國立國會圖書館年報平成19年度』. 東京: 國立國會圖書館. [cited 2009.10.1]. <<http://ndl.go.jp/jp/publication/annual/h19/nen19.dpf>>.
- British Library. 2006. *Thirty-Third Annual Report and Accounts 2005/2006. Key performance indicators*. <<http://www.bl.uk/about/annual/2005to2006/pdf/kpis.pdf>>.
- Comer, C. H. 1981. "List checking as a method for evaluating library collections." *Collection Building*, 3(3): 26-34.
- Evans, G. Edward., Saponaro, Margret Zar-

- nosky. 2005. *Developing Library and Information Center Collections*. 5th ed. Westport, Conn.: Libraries Unlimited.
- International Standardization Organization. 2006. *ISO 2789, Information and Documentation International Library Statistics*. Geneva: ISO.
- _____. 2008. *ISO 11620, Information and Documentation-Library Performance Indicators*. Geneva: ISO.
- _____. 2009. *ISO/TR 28118, Information and Documentation-Performance Indicators for National Libraries*. Geneva: ISO.
- Koninklijke Bibliotheek, 2005. *Annual Report 2005, Performance Indicators*. [cited 2009.10.1].
<<http://www.kb.nl/bst/jaar/kb2005/bedrijfsvoering5-en.html>>.
- Koninklijke Bibliotheek, 2008. *Annual Report 2008, Performance Indicators*. [cited 2009.10.1].
<<http://www.kb.nl/bst/jaar/kb2008/eng/kbannualreport2008.pdf>>.
- Lancaster, F. W. 장혜란 역. 1999. 『색인 초록의 이론과 실제, 제9장 색인 초록의 평가』. 서울: 구미무역.
- Poll, Roswitha. 2008. "The cat's pyjamas? Performance Indicators for National Libraries." *Performance Measurement and Metrics*, 9(2): 110-117.
- Poll, Roswitha and Christina Jonsson-Adrial. 2006. *Performance Indicators for National Libraries: A list of possible indicators*, taken from the new draft of the standard ISO 11620 and form practical examples tested by national or regional libraries. [cited 2009.10.1].
<<http://www.ifla.org/VII/s1-PerformanceIndicators2006.pdf>>.
- Steenbakkers, Johan. 2000. The NEDLIB Guidelines. [cited 2009.10.1].
<<http://nedlib.kb.nl/results/NEDLIBguidelines.pdf>>.
- Tenopir, C. 1982. "Evaluation of Database Coverage: A Comparison of Two Methodologies." *Online Review*, (6): 423-441.
- Tokuhara, N. 2008. "ISO/TR 28118: Establishment of Performance Indicators for National Libraries." *Current Awareness*, 295(march 20). [cited 2009.10.1].
<<http://current.ndl.go.jp/ca1653>>. (Japanese text only)
- Voorbij, H. and Lemmen, A. 2006. "Examining the Coverage of a National Deposit Library: a Case Study in the Netherlands." *Libri*, 56(3): 191-199.