

단과대학별 도서관 장서 활용 현황 분석을 위한 대출데이터 기반 대출지수 비교

Analysis of Circulation Measures based on Circulation Data of Colleges in a University

최상희 (Sanghee Choi)*

이재윤 (Jae Yun Lee)**

초 록

대출데이터는 대학도서관에 축적된 중요한 데이터로서 도서관 장서개발이나 서비스 개선에 활용될 수 있는 중요한 데이터이다. 이 연구는 대출빈도를 기반으로 한 다양한 대출관련지수를 비교분석하여 지수별 특성을 파악한 후 도서관 운영에 적용할 수 있는 타당성을 평가하고자 하였다. A 대학도서관의 10개 단과대학별 대출데이터를 대상으로 비교분석한 지수는 대출빈도, 대출엔트로피, 대출 h-지수, 대출주제차별지수 등 총 4개의 지수이다. 이 지수들을 적용하여 단과대학별 대출현황을 분석하였고 단과대학별로 나타나는 대출주제의 특성을 표하는 각 지수의 특성을 비교 분석하였다. 분석 결과 대출 엔트로피는 여러 대학이 공통으로 선호하는 주제를 표현하는 성향이 있는 것으로 나타났다. 반면 대출주제 차별지수는 특정대학에서만 특화되어 대출되는 주제를 표현하는 성향이 있는 것으로 나타났다.

ABSTRACT

Circulation data is a key data set of academic libraries in terms of collection development and service improvement. This study aims to identify the characteristics of circulation measures and their feasibility. This study collected the circulation data of 10 colleges in a university and analyzed 4 measures based on the circulation data: circulation frequency, circulation entropy, circulation h-index, and circulation divergence. These measures are to present the circulation topics of each college. This study identified that circulation entropy tends to present general topics which are popular for many colleges, but circulation divergence tends to present specific topics which are preferred by a specific college.

키워드: 대학도서관, 대출데이터, 대출 엔트로피, 대출 h-지수, 대출주제차별지수
university library, circulation entropy, circulation h-index, circulation divergence,
circulation data, data analysis

* 대구가톨릭대학교 도서관학과(shchoi@cu.ac.kr) (제1저자)

** 명지대학교 문헌정보학과(memexlee@kgu.ac.kr) (교신저자)

■ 논문접수일자: 2018년 11월 15일 ■ 최초심사일자: 2018년 12월 26일 ■ 게재확정일자: 2018년 12월 28일
■ 정보관리학회지, 35(4), 125-140, 2018. [http://dx.doi.org/10.3743/KOSIM.2018.35.4.125]

1. 서론

최근 빅데이터 분석이 사회전반에서 주목을 받게 되면서 도서관계에서 데이터 분석에 대한 중요성이 더욱 강조되고 있다. 도서관에 축적되고 있는 데이터 중에 대출데이터는 가장 대량으로 생산되고 축적되고 있는 데이터이다. 대출데이터는 월별, 연간별 이용자별, 주제별로 산출될 수 있으며 누적되는 시기별로 기초 통계로 산출되어 활용되고 있다.

대출데이터는 이용 현황을 파악하기 위해서 많이 활용되고 있지만 한편으로는 도서관이 소장하고 있는 장서의 특성을 분석하기 위해서도 활용되고 있다. 특히 다양한 주제 분야의 자료를 소장하고 있고 이를 기반으로 주제 서비스를 하고 있는 대학도서관에서는 대출데이터를 분석하여 장서를 개발하고 주제서비스를 운영하고자 하는 시도를 하고 있다. 최근에는 대학도서관의 예산이 크게 줄고 전자장서에 대한 비용이 증가하면서 효과적으로 장서를 개발하기 위하여 각각의 도서관에 최적화된 주제 분야를 파악해야하는 필요성이 증대하고 있다. 그러나 도서관에서 대출 데이터를 활용하는 수준은 대부분 대출빈도를 기반으로 한 단순 통계 수준에서 활용되는 수준인 경우가 많다. 대출데이터를 수치화하여 활용하고자 하는 의도는 대학도서관이 장서를 개발하거나 서비스를 개발하는 근거로 사용하고자 하는 것이지만 단순 대출빈도는 이러한 근거로 활용되기에는 부족한 측면이 있다. 따라서 대출통계를 장서관련 실적보고용으로 활용하거나 장서이용통계 수준으로 단순하게 해석하는 것보다 다양한 방식으로 대출데이터를 활용하는 시도를 수행하

여 도서관내에 축적되는 데이터 분석 기법을 개발해야할 필요성이 있다.

이 연구에서는 대출데이터를 기반으로 한 다양한 지수를 산출하여 대출데이터를 다양하게 활용하고자 하는 시도를 하고자 총 4종류의 대출관련 지수를 비교분석하였다. 비교 분석한 대출관련 지수는 대출빈도, 대출 엔트로피, 대출 h-지수, 대출주제차별지수인데 대출지수의 특성을 살펴보기 위하여 대학도서관의 대출데이터에 적용하여 각 지수가 대학도서관의 자료 이용 특성을 표현하는 성향을 조사하였다. 실제 데이터 분석에 적용하기 위하여 A 대학의 1년치 대출데이터를 수집하여 단과대학별로 대출빈도를 구분하여 대출관련지수를 산출하였고 4개의 대출관련 지수가 단과대학별로 대출되는 자료의 주제를 표현하는 성향에 차별성이 있는지를 파악하였으며 각 특성을 비교분석하였다.

2. 선행연구

대출통계분석은 보편적으로 장서개발을 목적으로 수행되어 왔는데 도서관의 예산이 감축되는 환경에서 효율적인 장서개발을 하는 도구로 활용되어 왔다. Knievel의 연구에서는 장서대출 통계를 분석하는 것이 도서관 장서개발을 위한 주요 평가 방식이 되어야 한다면서 대출통계 분석의 중요성을 주장하였다(Knievel, 2006). 이와 유사한 연구로서 도서관 대출 통계가 장서개발 정책평가도구가 될 수 있는지를 분석한 연구도 있는데(Adams & Noel, 2008) 이 연구에서는 평균대출빈도에 대한 분석을 발행자별

로 하고 대출빈도수가 높게 나타나는 자료들을 파악하여 장서개발정책에 적용할 수 있는지 여부를 분석하였다. 최근 학술단행본의 대출 빈도와 인용 빈도를 분석하여 폐기에 적용할 수 있는 척도가 되는지를 분석한 연구가 있는데 이 연구에서 두 요소 모두 폐기결정에 적용해야만 한다고 기술하였으며 과거 대출 기록을 기반으로 미래 대출행태를 예측할 수 있다고 하였다(White, 2017).

대학도서관의 대출통계를 기반으로 분석을 한 연구로서 특정 대학의 1년치 대출데이터를 기반으로 이용행태를 분석한 국내 연구가 있다(김선애, 2013). 이 연구에서는 자료의 주제별, 이용자 유형별로 구분하여 분석하였는데 장서의 주제별 이용 현황에서는 경제학, 예술, 한국문학, 철학 순으로 나타났고 주제별 이용 요인분석에서는 철학과 동양문학이 높게 나타났다. 이러한 주제별 자료 이용 현황은 이용자의 유형에 따라서 다르게 나타났는데 학부생은 경제학, 대학원생은 사회/교육학, 교원은 예술 그리고 직원은 동양문학에서 이용률에 차별성을 보이고 있는 것으로 밝혔다. 반면에 철학은 모든 이용자가 이용을 많이 하는 분야로 도출되어 이용자별로 대출주제간 차별성과 보편성을 분석 제시하였다.

양지안(2017)의 연구에서는 대학도서관 대출데이터를 활용하여 장서의 대출수명과 주제 영역, 특정 이용자층과의 연관성 등의 요소로 대출행태를 분석하였다. 이 연구에서는 특정 대학도서관의 10년간 대출데이터를 분석하여 매해 20%의 대출도서가 전체 대출의 50%를 차지한다는 패턴을 분석하였고 각 대출도서의 도서연령과 대출횟수의 연령별 누적분포를 구

하여 도서 반감기를 산출하였다. 또한 대출되는 주제 분야를 종단적 관점과 횡단적 관점으로 분석하였는데 종단적 관점 분석에서 10년 동안 핵심장서로 분류되는 도서들의 역할이 도출되었다. 이는 장기적인 시점에서 대출이 지속되는 핵심장서의 특성이 명확하다는 의미라고 해석하였고 누적된 대출통계 분석을 통해 효율적인 장서구성정책을 개발할 수 있는 장서 대출데이터 분석기법을 제시하였다.

전자자료의 활용이 활성화되면서 전자책의 대출 데이터에 초점을 맞추어 분석한 연구도 수행되었는데 이 연구에서는 대학도서관 전자책의 이용현황을 인쇄 단행본 대출데이터와 비교 분석하여 전자책의 이용현황을 파악하고자 하였다(정진한, 2009). 이 연구에서 수행한 전자책 대출분석 결과, 전자책 시장과 서비스가 전자책 대출을 활성화할 수 있는 수준으로 지원하고 있지 못하다는 문제점이 도출되어 대학도서관의 장서개발 차원이 아닌 보다 거시적인 문제점이 분석되었다.

상호대차 데이터를 분석한 사례도 다수 있었는데 박영애의 연구에서는(박영애, 이재운, 2010) 작은 도서관을 대상으로 대출통계를 기반으로 분석을 한 사례이다. 이 연구에서는 한 지역에서 상호대차를 하는 작은 도서관 14관을 대상으로 장서량과 이용량을 측정하여 도서관의 특성을 파악하였는데 이용량으로 대출빈도와 상호대차 빈도를 측정하여 분석하였다. 특히 이 연구에서는 상호대차 의존도와 주제 분야를 교차분석하였는데 주제별 도서이용 수준이 높은 주제 분야가 도서관별 상호대차 의존도에서도 높게 나타났다는 결과를 제시하였고 이와 같은 결과를 바탕으로 작은 도서관은 개별도서관의 장서

개발시 상호대차와 대출통계를 바탕으로 주제 분야에 대한 고찰이 필요하다고 지적하였다.

3. 단과대학별 대출데이터 분석

3.1 분석데이터 및 연구방법

각 단과대학별로 대출하는 도서 주제영역 및 단과대학별 특성을 살펴보기 위하여 A 대학의 학부생을 대상으로 1년간 단행본 대출기록을 수집하여 통계분석하였다. 수집한 대학에 소속된 단과대학은 총 12개인데 이 중 캠퍼스가 분리되어 있어 개별적으로 도서관을 운영하는 신학대학과 의과대학을 제외한 단과대학 총 10개를 대상으로 분석을 실시하였다.

- 간호대학
- 공과대학
- 글로벌비즈니스대학
- 디자인대학
- 바이오메디대학
- 사범대학
- 사회과학대학
- 약학대학
- 음악대학
- 자율전공학부

A대학 경우 단과대학의 명칭이 변경되고 일부 대학은 통합이 된 시기이어서 이에 따라 단과 대학명을 전거하여 하였다.

1년간 A대학도서관에서 학부생이 대출한 횟수는 단행본 기준으로 82,125이며 대출통계 분석은 전체 대출 데이터에서 직원 대출은 제외하고 휴학생도 제외하였다. 또한 복본 구분이 되어 있는 데이터는 데이터가 분산되는 것을 방지하기 위하여 복본도 전거하였다. 또한 딸림자료 대출은 단행본자료 대출에서 제외하였다.

대출빈도를 기초 통계로 산출하였고 대출빈

도와 대출 엔트로피, 대출 h-지수, 다이버전스 지수를 산출하여 각 대출빈도와 차별성을 비교하여 단과대학별로 대출로 나타나는 특화 주제를 파악할 수 있는지를 조사하였다.

대출 엔트로피는 개별 학문분야의 학제성을 조사한 이재윤(2008)의 연구에서 활용한 엔트로피 개념을 적용하였다. 엔트로피는 확률값이 고르게 분포되어 있을수록 높게 나타나는데 이 연구에서는 특정 학문분야가 다른 학문분야와 가지는 학술지 공유도를 확률값으로 변환한 후 학술지 공유도 엔트로피를 산출하였다. 이 개념을 적용하여 이 연구에서는 이 개념을 단과대학과 대출 분산도를 산출하는데 적용하여 대출 분산도에 따른 단과대학별 대출 주제 특성을 살펴보고자 하였다. 대출 엔트로피는 해당 주제 장서 중에서 1책만 대출된 경우에는 아무리 대출빈도가 높아도 0이 되고, 여러 책이 고르게 많이 대출될수록 높게 산출된다.

대출 h-지수는 연구자의 영향력을 평가하는 지수인 h-지수를 대출의 주제에 적용하여 산출한 것으로서, h-지수의 개념은 특정 과학자의 논문을 인용빈도 순으로 나열하고 일련부호를 부여하였을때 상위 h개의 문헌이 최소 h 인용되고 h+1 문헌의 인용빈도가 h+1 미만일때 특정 연구자의 h-지수가 라는 개념을 주제와 대출 빈도에 적용한 것이다. 즉, 이 연구에서는 특정 과학자를 주제 분야로 대체하고 인용빈도를 대출빈도로 하여 특정 주제의 영향력을 대출 h-지수로 산출한 것이다. 대출 h-지수는 해당 주제 장서 중 대출횟수가 높은 책이 많을수록 높게 산출되며, 한 두 책이 집중적으로 이용되는 것만으로는 대출 h-지수가 높아지지 않는다.

단과대학별 대출수에 기반하여 단과대학별

로 특화된 주제를 파악해보기 위하여 김하수, 손현정, 이재운, 강범일(2013)의 연구에서 적용한 다이버전스 개념을 이용하여 주제 차별성을 측정하였다. 김하수 외는 선거 후보가 사용하는 어휘 차이를 다이버전스로 측정하여 선거 후보간 차별성을 파악하려고 시도하였다. 이들은 Kullback-Leibler 다이버전스(Kullback, 1968; Kullback & Leibler, 1951) 공식을 사용하여 각 후보의 어휘들을 대상으로 각각의 어휘가 '각 후보의 자료에서 출현할 확률들의 평균값'과 '특정 후보의 자료에서 출현할 확률값'을 산출하여 이 두 값의 차이를 나타내는 다이버전스값을 구했다. 이와 같은 방식으로 이 연구에서는 각 단과대학별로 대출되는 단행본의 주제를 대상으로 각각의 주제가 '전체 단과대학 대출에서 출현할 확률들의 평균 값'과 '특정 대학 대출에서 출현할 확률 값'을 산출하여 이 두 값의 차이를 나타내는 다이버전스 값을 구하여 특정 대학에서 특화되어 나타나는 주제 분야가 있는지 조사하였다. 주제 W가 특정 대학에서 출현할 확률 $q_A(W)$ 와 전체 대출에서 출현할 확률의 평균 $r(W)$ 의 다이버전스값인 $DA(q_A||r)$ 는 선행 연구에서 Kullback-Leibler 다이버전스를 기반으로 제시한 다이버전스 공식에 의해 산출되었다(이재운, 2007). 이 연구에서는 산출된 다이버전스를 대출주제차별지수라고 지칭하였다. 대출주제차별지수는 해당 주제 장서를 특정 단과대학에서 다른 단과대학보다 두드러지게 많이 이용할수록 높게 산출된다. 대출주제차별지수 공식은 다음과 같다.

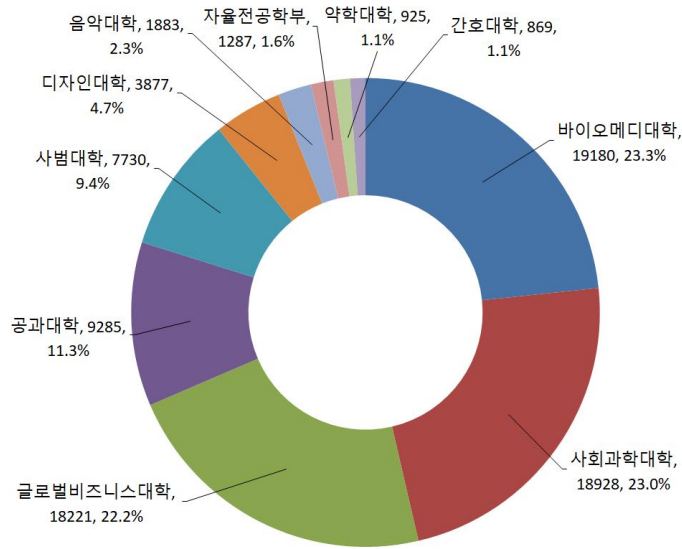
$$D_A(q_A||r) = q_A(W) \times \log \frac{q_A(W)}{r(W)}$$

3.2 대출현황 및 대출주제

단과대학별 대출현황을 조사한 결과 글로벌비즈니스대학, 바이오메디대학, 사회과학대학 순으로 대출빈도가 높은 단과대학인 것으로 나타났다. 이는 단과대학의 규모에 따른 것으로 3개 단과대학이 소속 학과 수도 많고 학생 수도 많아 대출빈도에 직접적인 영향을 미친 것이다 (<그림 1> 참조).

주제 분야별 대출현황을 조사해보면 대출비율이 가장 높은 분야는 경제학(10.2%)이고 한국문학(9.8%), 사회학(5.6%), 의학(5.5%), 심리학(4.7%)의 차례로 많이 대출되었다(<표 1> 참조). 그런데 이 수치는 대출 비중이 높은 특정 단과대학의 영향을 많이 받는 경향이 있다. 예를 들어 의학의 경우는 바이오메디대학(16.2%), 약학대학(47.2%), 간호대학(15.9%)만 대출비율이 높을 뿐이고, 다른 단과대학에서는 단과대학별 대출 중에서 의학이 2% 이상 차지하는 경우가 없다. 또한 사회학의 경우도 사회과학대학(14.6%)과 전공자율학부(5.6%)를 제외하면 전체 대출비율인 5.6%를 넘게 대출하는 단과대학이 없다.

따라서 주제 분야별 대출비율이 비중이 높은 특정 단과대학에 좌우되지 않도록, 10개 단과대학별로 각 주제의 대출비율을 산출한 후 10개 값의 평균을 구해서 <표 1>에 함께 제시하였다. 단과대학별 주제 대출비율의 평균이 높은 주제는 한국문학(8.9%), 의학(8.9%), 경제학(7.3%), 음악(5.4%), 심리학(4.7%)의 순이었다. 사회과학대학의 대출이 크게 반영되어 전체 대출비율에서는 3위였던 사회학 분야가, 단과대학별 비율의 평균 순위에서는 6위로 하락하였다. 반면에 전체



〈그림 1〉 단과대학별 1년간 단행본 대출 비율

〈표 1〉 전체 대출 중 주제별 도서의 대출비율과 10개 단과대별 대출 중 주제별 도서의 대출비율의 평균(전체 대출비율 상위 20개 주제만 제시함)

주제	전체 대출비율(순위)	10개 단과대별 대출비율의 평균(순위)
경제학	10.2% (1)	7.3% (3)
한국문학	9.8% (2)	8.9% (1)
사회학	5.6% (3)	4.1% (6)
의학	5.5% (4)	8.9% (2)
심리학	4.7% (5)	4.3% (5)
교육학	3.8% (6)	3.7% (7)
영미문학	3.3% (7)	3.3% (9)
지리	2.5% (8)	3.3% (8)
총류	2.4% (9)	2.1% (14)
영어	2.3% (10)	1.9% (19)
수학	2.3% (11)	2.0% (17)
윤리학	2.2% (12)	2.2% (13)
음악	2.2% (13)	5.4% (4)
일본문학	2.1% (14)	2.3% (12)
가정학	2.1% (15)	1.8% (20)
아시아 역사	2.0% (16)	2.0% (16)
문헌정보학	1.9% (17)	1.4% (25)
생물과학	1.9% (18)	2.0% (15)
법학	1.7% (19)	1.5% (24)
기독교	1.6% (20)	2.4% (11)

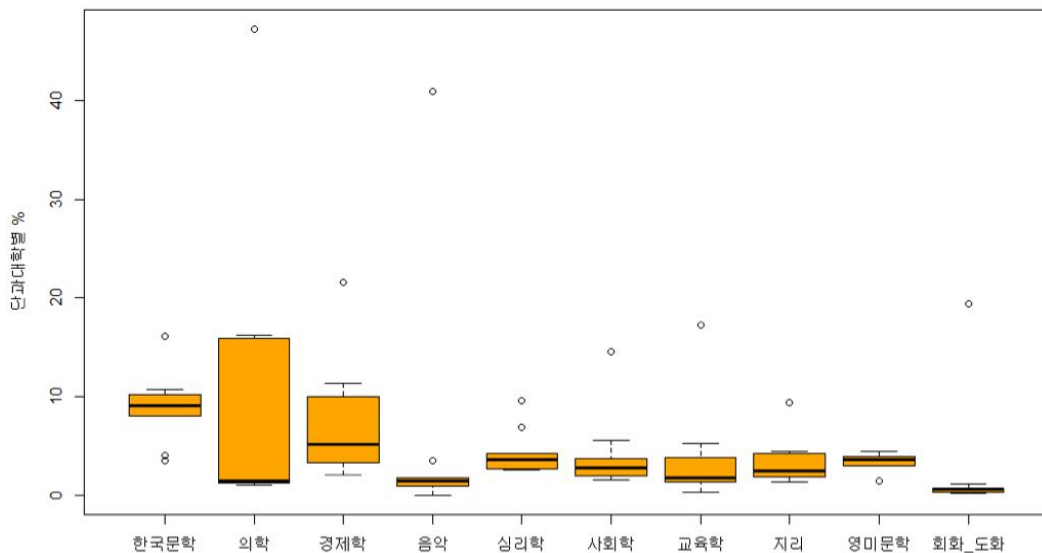
대출비율이 2.2%로 13위였던 음악 분야가 4위로 평균 순위가 크게 상승하였다. 이는 음악대학에서 음악 주제의 대출비율이 40.9%로 예외적으로 높았기에 평균이 높게 산출된 경우로서, 나머지 대학에서는 대부분 1% 내외에 불과하였다. 따라서 단과대학별 주제 대출비율의 평균을 산출한 결과에서도 여러 단과대학에서 크게 대출되는 주제 분야를 파악하기는 쉽지 않다.

여러 단과대학에서 공통적으로 많이 대출되

는 주제 분야를 파악하기 위해서 각 주제 분야가 단과대학별 대출 중에서 차지하고 있는 비율의 분포를 파악해보았다. 전체 대출비율 상위 10개 주제에 대해서만 단과대학별 대출비율의 분포를 제시하면 <표 2> 및 <그림 2>와 같다. <그림 2>를 보면 의학이나 음악 분야는 단과대학별 대출비율 평균은 높으나 일부 단과대학에서만 대출이 많이 되고 있어서 중앙값(상자 안의 굵은 가로선)과 1사분위수(상자의 하단)는 매우 낮게 나타났다.

<표 2> 10개 단과대학별 대출비율 평균이 높은 상위 10개 주제 분야의 대출비율 분포 통계량 (단위: %)

	한국문학	의학	경제학	음악	심리학	사회학	교육학	지리	영미문학	회화/도화
최댓값	16.2	47.2	21.6	41.0	9.7	14.6	17.3	9.4	4.4	19.5
3사분위수	10.1	12.4	9.4	1.7	4.1	3.6	3.4	4.0	3.9	0.7
평균	8.9	8.9	7.3	5.4	4.3	4.1	3.7	3.3	3.3	2.5
중앙값	9.1	1.5	5.2	1.5	3.7	2.8	1.8	2.5	3.6	0.7
1사분위수	8.1	1.3	3.5	1.0	2.9	2.1	1.4	1.9	3.1	0.4
최솟값	3.6	1.1	2.1	0.0	2.6	1.6	0.3	1.3	1.5	0.3



<그림 2> 10개 단과대학별 대출비율 평균이 높은 상위 10개 주제 분야의 대출비율 분포

1사분위수 기준으로 상위 주제 분야는 한국문학(8.1%), 경제학(3.5%), 영미문학(3.1%), 심리학(2.9%), 사회학(2.1%)의 순이었다. 이들은 3/4의 단과대학에서 2%이상의 대출비율을 차지하는 분야들이다. 최솟값 기준으로 상위 주제 분야는 한국문학(3.6%), 심리학(2.6%), 경제학(2.1%), 사회학(1.6%), 영미문학(1.5%)의 순이었다. 이들은 모든 단과대학에서 1.5% 이상의 대출비율을 차지하는 분야들이다. 결국 여러 단과대학에서 고르게 많이 대출하고 있는 주제 분야는 1사분위수와 최솟값 기준으로 상위 5위에 공통적으로 포함되는 한국문학, 경제학, 심리학, 사회학, 영미문학 분야인 것으로 나타났다. 이 분야들은 대학교육의 교양영역에 해당하는 것으로 대학도서관 장서의 교양 단행본 주제로 해석할 수 있다.

4. 대출 지수 특성 분석

대출빈도, 대출 엔트로피, 대출 h-지수, 대출 주제차별지수 등 4가지 대출 데이터를 기반으로 산출된 지수의 특성을 조사하기 위하여 가장 기본적인 통계지수인 대출빈도를 기준으로 하여 각각의 지수를 비교하고, 마지막에 가장 상반된 지수인 대출 엔트로피와 대출주제차별지수를 비교하였다.

4.1 대출빈도와 대출 엔트로피

각 단과대별로 대출 엔트로피 값을 산출한 후 높은 순으로 상위에 나타나는 주제를 추출하여 단과대학별 대출빈도 상위에 나타난 주

제와 비교하였다. 총 10개의 대학에서 최상위 주제가 변경된 경우는 <표 3> ~ <표 7>에서 나타났듯이 자율전공학과와 공과대학, 바이오대학, 사범대학, 사회과학대학 등 5개 대학이었다.

<표 3> 대출빈도 vs. 대출 엔트로피 - 자율전공학부

대출빈도	대출엔트로피
한국문학 / 132	경제학 / 6.16
경제학 / 129	한국문학 / 6.09
법학 / 92	법학 / 5.79
정치학 / 74	영미문학 / 5.40
사회학, 사회문제 / 72	일본문학 / 4.94

자율전공학부에서는 단순빈도에서는 한국문학이 최상위 주제였으나 대출엔트로피에서는 경제학이 최상위 주제로 나타났으며 대출빈도 상위 5개 주제로 나타나지 않았던 영미문학과 일본문학이 주제로 나타났다.

<표 4> 대출빈도 vs. 대출 엔트로피 - 공과대학

대출빈도	대출엔트로피
총류 / 899	한국문학 / 8.11
한국문학 / 769	총류 / 7.81
경제학 / 710	문헌정보학 / 7.14
문헌정보학 / 667	일본문학 / 7.12
전기공학, 전자공학 / 586	경제학 / 7.03

공과대학에서는 단순빈도에서는 총류가 최상위 주제였으나 대출엔트로피에서는 한국문학이 최상위 주제로 나타났으며 대출빈도 상위 5개 주제로 나타나지 않았던 일본문학이 주제로 나타났다.

〈표 5〉 대출빈도 vs. 대출 엔트로피 - 바이오메디 대학

대출빈도	대출엔트로피
의학 / 3116	한국문학 / 9.15
한국문학 / 1695	의학 / 8.48
생물과학 / 1061	경제학 / 8.16
경제학 / 951	지리 / 7.53
수학 / 783	가정학 및 가정생활 / 7.35

바이오메디 대학에서는 단순빈도에서는 의학이 최상위 주제였으나 대출엔트로피에서는 한국문학이 최상위 주제로 나타났으며 지리, 가정학 및 가정생활이 대출빈도 상위 5개 주제로 나타났다.

〈표 6〉 대출빈도 vs. 대출 엔트로피 - 사범대학

대출빈도	대출엔트로피
교육학 / 1338	한국문학 / 8.89
한국문학 / 1250	교육학 / 8.15
아시아(아세아) / 862	아시아(아세아) / 7.74
수학 / 339	심리학 / 6.75
지리 / 333	지리 / 6.62

사범대학에서는 단순빈도에서는 교육학이 최상위 주제였으나 대출엔트로피에서는 한국문학이 최상위 주제로 나타났으며 심리학이 대출빈도 상위 5개 주제로 나타났다.

〈표 7〉 대출빈도 vs. 대출 엔트로피 - 사회과학대학

대출빈도	대출엔트로피
사회학, 사회문제 / 2765	한국문학 / 9.56
경제학 / 2161	사회학, 사회문제 / 9.16
한국문학 / 1846	경제학 / 8.86
심리학 / 1827	심리학 / 8.14
법학 / 1078	법학 / 7.97

사회과학대학에서는 대출빈도에서는 사회학, 사회문제가 최상위 주제였으나 대출엔트로피에서는 한국문학이 최상위 주제로 나타났지만 대출빈도 상위 5개 주제에는 변화가 없었다.

최상위 순위변동이 일어나지 않은 대학은 간호대학, 약학대학, 디자인대학, 음악대학, 글로벌비즈니스 대학 등 비교적 단과대학의 전공 주제가 단일 성향이 있는 대학이었고 순위변동이 있는 대학들은 자율전공학부나 공과대학, 바이오메디, 사범대학, 사회과학대학 등 전공 주제가 좀 더 학제적이거나 여러 학문 분야로 구성된 복합적인 성격이 있는 대학이었다.

대출엔트로피로 인해 순위가 상승된 주제 분야를 살펴보면 한국문학, 영미문학, 일본문학 등 전공 주제보다는 보편적인 교양에 해당하는 주제 분야였다. 즉, 대출 엔트로피가 특정 대학에서만 보는 세부 전공분야보다는 여러 단과에서 공통적으로 보는 주제를 도출해내는 역할을 할 수 있는 것으로 해석된다.

4.2 대출빈도와 대출 h-지수

각 단과대별로 대출h-지수값을 산출한 후 지수값이 높은 순으로 상위에 나타나는 주제를 추출하여 단과대학별 대출빈도 상위에 나타난 주제와 비교하였다. 총 10개의 대학에서 최상위 주제가 변경된 경우는 〈표 8〉, 〈표 9〉, 〈표 10〉에서 나타났듯이 자율전공학부와 공과대학, 사회과학대학 등 3개 대학이었다.

자율전공학부의 경우 대출빈도로 나타난 최상위 주제는 한국문학, 경제학, 법학 순이었고 엔 대출 h-지수로 나타난 주제는 경제학, 정치학, 한국문학 순이었다. 그러나 상위 5개 주제 분야에

대출 h-지수로 새로 나타난 주제 분야는 없었다.

〈표 8〉 대출빈도 vs. 대출 h-지수 - 자율전공학부

대출빈도	대출 h-지수
한국문학 / 132	경제학 / 4
경제학 / 129	정치학 / 4
법학 / 92	한국문학 / 3
정치학 / 74	법학 / 3
사회학, 사회문제 / 72	사회학, 사회문제 / 3

공과대학은 대출빈도 순으로는 총류, 한국문학, 경제학 순이었으나 대출 h-지수로는 전기, 전자공학, 공학, 공업일반, 총류 순으로 나타났다. 대출 h-지수로 공과대학의 주제성이 부각되는 주제 분야가 상위권으로 이동한 것으로 조사되었다.

〈표 9〉 대출빈도 vs. 대출 h-지수 - 공과대학

대출빈도	대출 h-지수
총류 / 899	전기공학, 전자공학 / 13
한국문학 / 769	공학, 공업일반 / 12
경제학 / 710	총류 / 11
문헌정보학 / 667	문헌정보학 / 11
전기공학, 전자공학 / 586	경제학 / 10

사회과학대학은 대출빈도 순으로는 사회학, 사회문제, 경제학, 한국문학 순이었으나 대출 h-지수로는 심리학, 사회학, 사회문제, 경제학 순으로 나타났다. 그러나 상위 5개 주제 분야의 변동은 없는 것으로 조사되었다.

3개 단과대학을 제외한 나머지 7개 대학에서는 최상위 주제가 변동되지 않았고 상위 5개 안의 주제 변동도 거의 없었다. 대출엔트로피와

비교하면 대출 h-지수는 대출빈도와 거의 유사한 결과를 나타나는 것으로 조사되었다.

〈표 10〉 대출빈도 vs. 대출 h-지수 - 사회과학대학

대출빈도	대출 h-지수
사회학, 사회문제 / 2765	심리학 / 18
경제학 / 2161	사회학, 사회문제 / 16
한국문학 / 1846	경제학 / 16
심리학 / 1827	법학 / 12
법학 / 1078	한국문학 / 11

4.3 대출빈도와 대출주제차별지수

대출주제차별지수 값을 산출한 후 지수값이 높은 순으로 상위에 나타나는 주제를 추출하여 단과대학별 대출빈도 상위에 나타난 주제와 비교하였다. 총 10개의 대학에서 최상위 주제가 변경된 경우는 자율전공학부 1개 대학으로 최상위 주제가 변경되는 변화는 크게 나타나지 않았다(〈표 11〉 참조).

〈표 11〉 대출빈도 vs. 대출주제차별지수 - 자율전공학부

대출빈도	대출주제차별지수
한국문학 / 132	법학 / 0.10
경제학 / 129	정치학 / 0.07
법학 / 92	지리 / 0.02
정치학 / 74	사회과학 / 0.02
사회학, 사회문제 / 72	행정학 / 0.02

자율전공학부의 대출주제는 한국문학, 경제학, 법학 순이었고 대출주제차별지수로 나타난 주제는 법학, 정치학, 지리 순이었다. 상위 5개 주제 분야에서 나타난 변화는 다른 지수에서

항상 상위로 나타난 한국문학이 상위 5개 주제 분야에서 나타나지 않은 것이다.

최상위 주제가 변화하는 특성은 보이지 않았으나 대출주제차별지수는 위 5개 주제내의 변화는 대출빈도 주제와 비교하여 위 두 지수보다 크게 나타났다. 이러한 주제 변화는 특히 2-5위내에서 많이 나타났고 새로 등장한 주제 분야는 해당 단과대학과 관련된 주제성이 높은 것으로 조사되었다.

〈표 12〉 공과대학의 사례와 〈표 13〉 바이오메디대학 사례를 살펴보면 공과대학의 경우 최상위 주제는 총류로 동일하지만 대출빈도 2위에 해당하였던 한국문학이 대출주제차별지수의 상위주제에서 나타나지 않았다. 대출빈도 상위 주제였던 경제학도 대출주제차별지수로는 상위에 랭크되지 못하였지만 공학, 공업일반과 건축공학과 같은 공과대학의 전공주제와 연관성이 높고 다른 대학의 전공과는 연관성이 낮은 주제 분야가 상위 주제 분야로 도출되었다.

〈표 12〉 대출빈도 vs. 대출주제차별지수 - 공과대학

대출빈도	대출주제차별지수
총류 / 899	총류 / 0.13
한국문학 / 769	전기공학, 전자공학 / 0.12
경제학 / 710	문헌정보학 / 0.09
문헌정보학 / 667	공학, 공업일반 / 0.07
전기공학, 전자공학 / 586	건축공학 / 0.06

바이오메디 대학 사례에서도(〈표 13〉 참조) 최상위 주제는 의학으로 동일하지만 대출빈도 2위에 해당하였던 한국문학이 대출주제차별지수의 상위주제에서 나타나지 않았고 경제학이나 수학처럼 다른 대학에서도 교육을 하는 기

초나 교양과목 주제에 해당하는 주제 분야가 상위주제에서 누락되었다. 반면에 화학, 화학공학, 농업, 농학과 같은 바이오메디 대학의 전공 주제와 연관성이 높은 주제 분야가 상위에 랭크되는 현상이 나타났다.

〈표 13〉 대출빈도 vs. 대출주제차별지수 - 바이오메디대학

대출빈도	대출주제차별지수
의학 / 3116	의학 / 0.17
한국문학 / 1695	생물과학 / 0.05
생물과학 / 1061	화학 / 0.03
경제학 / 951	화학공학 / 0.03
수학 / 783	농업, 농학 / 0.02

공과대학과 바이오메디 대학의 사례에서 나타났듯이 대출주제차별지수 분석에서 해당대학의 전공과 연관이 높은 주제가 상위주제로 랭크되는 현상은 글로벌 비즈니스 대학이나 디자인, 음악, 사회과학대학 등에서 나타났다. 또한 대출빈도분석에서 상위 주제로 도출되었던 한국문학, 영국문학, 경제학 등과 같은 교양과 관련된 주제 분야가 상위주제에서 하위로 이동되는 현상이 나타나 대출주제차별지수가 단과대학의 전공 주제와 연관성이 높은 주제 분야를 파악하는 역할을 하는 것으로 분석되었다.

4.4 대출엔트로피와 대출주제차별지수

비교한 4개 지수 중에서 대출빈도와 대출 h-지수는 유사한 결과를 보이고 있다. 대출량에 크게 좌우되는 대출 h-지수와 달리 대출엔트로피는 대출의 다양성을 대표하고 대출주제차별지수는 대출의 특이성을 대표하는 것으로 간주

할 수 있다. 대출엔트로피와 대출주제차별지수의 상반된 특성은 <표 14>의 공과대학 사례에서 명확히 대비된다. 대출엔트로피로 상위 주제로 도출된 주제 분야를 보면 한국문학, 일본문학, 경제학 등 교양주제가 주를 이루었지만 대출주제차별지수로 도출된 주제 분야는 전기공학, 전자공학, 공학, 공학일반, 건축공학 등 공과대학의 전공과 연관성이 높은 주제이다.

<표 14> 대출 엔트로피 vs. 대출주제차별지수 - 공과대학

대출엔트로피	대출주제차별지수
한국문학 / 8.11	총류 / 0.13
총류 / 7.81	전기공학, 전자공학 / 0.12
문헌정보학 / 7.14	문헌정보학 / 0.09
일본문학 / 7.12	공학, 공업일반 / 0.07
경제학 / 7.03	건축공학 / 0.06

이와 같은 대출주제차별지수의 특성은 특히 대학도서관이 주제성이 모호한 단과대학의 학생들이 관심 있어 하는 전공주제를 파악하는데 효과적일 수 있다. 예를 들어 자율전공학부와 같이 명칭만으로는 전공 파악이 어려운 단과대학을 위해 장서를 개발하거나 서비스를 운영할 때 도움이 될 수 있다.

<표 15>에서 자율전공학부의 사례를 보면 대출엔트로피에서는 경제학, 한국문학, 법학, 영미문학, 일본문학 등이 상위 주요 주제로 도출되었는데 대출주제차별지수에 의하면 법학, 정치학, 지리, 사회과학, 행정학 등이 주요 주제로 나타났다. 이는 A대학의 자율전공학부 학생들이 다른 단과대학과 비교하여 법학, 정치학, 지리, 사회과학, 행정학 등의 주제 단행본을 상대적으로 많이 대출하는 것을 의미하며 이러한

주제가 자율전공학부의 주제적 성격을 나타낸다고 해석할 수 있다.

<표 15> 대출 엔트로피 vs. 대출주제차별지수 - 자율전공학부

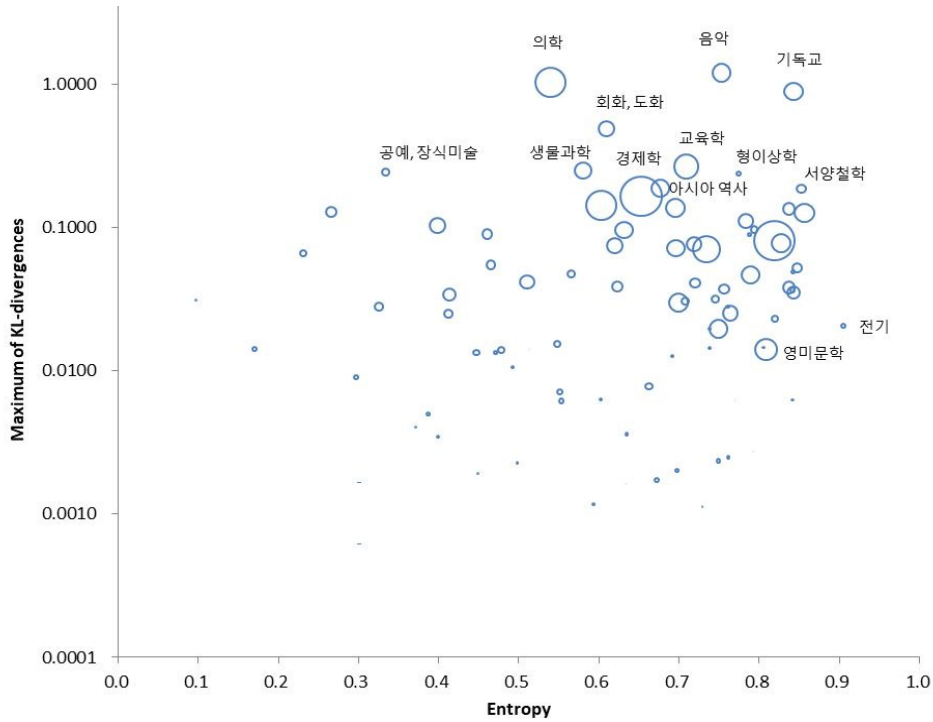
대출엔트로피	대출주제차별지수
경제학 / 6.16	법학 / 0.10
한국문학 / 6.09	정치학 / 0.07
법학 / 5.79	지리 / 0.02
영미문학 / 5.40	사회과학 / 0.02
일본문학 / 4.94	행정학 / 0.02

이외에도 대출엔트로피와 대출주제차별지수의 특성을 살려 대학도서관의 장서개발을 위해 소속대학의 주요 주제 분야 특성을 파악하는 목적으로도 활용할 수 있다.

<그림 3>을 살펴보면 우측 상단에 분포되어 있는 주제들은 일반적이면서도 단과대학의 주제성이 강한 주제이며 좌측으로 올수록 일반적인 주제성보다는 특정 단과대학의 주제성이 강한 주제이다. 반면에 우측 하단으로 내려갈수록 단과대학의 주제성은 낮아지고 교양과 같은 일반적인 주제성이 높아지는 것을 알 수 있다. 각 동그라미의 크기는 대출빈도에 비례하는 것이어서 동그라미가 커질수록 대출빈도가 높은 것이다.

가장 우측 상단에 있는 기독교는 특정 단과대학에서 대출하는 주제성도 강한 반면에 모든 대학에서도 고르게 많이 대출하는 분야이다. 이는 A대학이 기독교계열의 학교이어서 기독교관련 교양과목이 많은 것과도 연관성이 있다.

우측 하단에 있는 영미문학은 단과대학별 주제성은 약하지만 여러 단과대학에서 공통적으로 대출하는 주제 분야로 해석되며 서양철학도와 유사한 주제 분야로 나타났다.



〈그림 3〉 대출엔트로피와 대출주제차별지수로 도출된 주제 분야(원의 크기는 대출 빈도에 비례)

의학은 대출빈도도 높지만 모든 대학에서의 대출보다는 특정대학에 특화되어 대출되고 있는 것으로 나타나 단과대학의 전공과 연관이 깊은 전문 주제 분야이며 공예, 장식미술과 같은 분야는 특정 대학에 집중되어 대출되는 전문주제성이 매우 높은 분야인 것으로 판명되었다.

이처럼 상반된 특성인 대출엔트로피와 대출주제차별지수를 통해 각 주제별 장서의 활용양태를 입체적으로 파악할 수 있다.

5. 결론

대출데이터를 이용하여 단과대학별로 대출

주제 특성을 파악할 수 있는 대출관련 지수를 대출빈도, 대출엔트로피, 대출 h-지수, 대출주제차별지수 총 4개를 선정하여 특성을 비교한 결과 아래와 같은 사항이 분석되었다.

첫째, 대출 엔트로피는 확률값이 고르게 분포되어 있을수록 높게 나타나는 성향을 가지고 있어 각 단과대학에서 공통적으로 선호하는 주제를 파악하는데 효과적인 것으로 나타났다. 이러한 특성은 대출빈도와 비교에서 나타났는데 대출 엔트로피가 대출빈도로 산출된 최상위 주제가 가장 많이 변동된 대출주제지수였으며 대출 엔트로피가 상위 주제로 대출빈도와 다르게 도출한 주제 분야는 한국문학, 영미문학, 일본문학 등 단과대학의 전공주제보다는 교양에 가까

은 주제 분야였다.

둘째, 대출빈도와 대출 h-지수는 주제를 도출하는 성향이 유사하여 각 지수로 산출한 상위 주제 분야가 순위에 변동은 미비하게 있었으나 상위 주제 분야가 변동되는 비율이 적어 대출 주제 분야를 파악하는 성향이 거의 동일한 것으로 나타났다.

셋째, 대출주제차별지수는 '전체 단과대학 대출에서 출현할 확률들의 평균 값'과 '특정 대학 대출에서 출현할 확률 값'을 산출하여 특정 대학에서 특화되어 나타나는 주제 분야를 추적할 수 있는 지표로 제안되었다. 이 지수는 대출빈도와 비교하여 최상위주제를 변동시키는 역할은 하지 못하였지만 상위주제를 도출하는데 있어서 문학 분야나, 경제 분야 등 단과대학의 전공주제와 연관성이 낮은 주제 분야를 하위로 이동시키고 상대적으로 단과대학의 전공주제와 연관성이 높은 주제를 상위로 이동시키는

역할을 하는 것으로 나타났다.

넷째, 대출의 규모를 반영하는 대출빈도(또는 대출 h-지수), 대출 단과대의 다양성을 반영하는 대출엔트로피, 대출의 특정 단과대 집중도를 반영하는 대출주제차별지수를 통해서 주제별 장서의 대출 양태를 입체적으로 파악할 수 있다.

이 연구에서는 대학도서관에서 장서를 평가하거나 개발할 때 대출 데이터를 다양하게 활용할 수 있도록 대출데이터를 기반으로 4개의 지수를 비교분석하여 각 특성을 도출하였다. 향후에는 규모나 특성이 다른 대학의 대학도서관, 그리고 공공도서관에도 적용하여 대출지수 활용방안을 연구할 필요도 있다. 이용자 소속 구분이 있는 대학도서관과 다른 상황인 공공도서관의 경우에는 이용자 집단을 연령대 및 성별로 구분하여 이 연구에서 제안한 지수를 측정할 수 있을 것이다.

참 고 문 헌

- 김선애 (2013). 대학도서관의 자료 이용에 관한 연구. 한국비블리아학회지, 24(4), 115-131.
<http://dx.doi.org/10.14699/kbiblia.2013.24.4.115>
- 김하수, 손현정, 이재윤, 강범일 (2013). 정치와 언어의 관계에 대한 양적 분석 시론. 담화와인지, 20(1), 79-111.
- 박영애, 이재윤 (2010). 대출 및 상호대차 통계를 활용한 작은도서관 장서 평가에 대한 실험적 연구. 한국문헌정보학회지, 44(2), 333-356. <http://doi.org/10.4275/KSLIS.2010.44.2.333>
- 양지안 (2017). 대학도서관 대출데이터분석을 통한 장서 이용행태 및 특성에 관한 연구. 한국도서관·정보학회지, 48(2), 263-293. <http://dx.doi.org/10.16981/kliss.48.201706.263>
- 유종덕 (2013). 경기도 공공도서관 상호대차 네트워크 분석. 정보관리학회지, 30(2), 83-99.
<http://dx.doi.org/10.3743/KOSIM.2013.30.2.083>

- 이재운 (2007). 분포 유사도를 이용한 문헌클러스터링의 성능향상에 대한 연구. 정보관리학회지, 24(4), 267-283. <http://doi.org/10.3743/KOSIM.2007.24.4.267>
- 이재운 (2008). 연구자의 투고 학술지 현황에 근거한 국내 학문분야 네트워크 분석. 정보관리학회지, 25(4), 327-345. <http://doi.org/10.3743/KOSIM.2008.25.4.327>
- 정진한 (2009). 대학도서관 전자책서비스 활성화 방안에 관한 연구. 한국문헌정보학회지, 43(1), 113-133. <http://doi.org/10.4275/KSLIS.2009.43.1.113>
- Adams, B., & Noel, B. (2008). Circulation statistics in the evaluation of collection development, *Collection Building*, 29(2), 71-73.
- Kniewel, J. E., Wicht, H., & Connaway, L. S. (2006). Use of circulation statistics and interlibrary loan data in collection management. *College and Research Libraries*, 67(1), 35-49.
- Kullback, S. (1968). *Information Theory and Statistics*. 2nd Edition. New York: Dover Books.
- Kullback, S., & Richard A. L. (1951). On information and sufficiency, *The Annals of Mathematical Statistics*, 22(1), 79-86.
- White, B. (2017). Citations and circulation counts: Data sources for monograph deselection in research library collections. *College & Research Libraries*, 78(1), 53-65.

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

- Jung, Jin-han (2009). A study on the scheme to activate an e-book service in academic libraries. *Journal of the Korean Society for Library and Information Science*, 43(1), 113-133. <http://doi.org/10.4275/KSLIS.2009.43.1.113>
- Kim, Ha-Soo, Son, Hyunjung, Lee, Jae Yun, & Kang, Beomil (2013). A quantitative approach to the relation between politics and language. *Discourse and Cognition*, 20(1), 79-111.
- Kim, Sun-Ae (2013). A study on the collection use of an academic library: Focused on the analysis of circulation statistics of the A-academic library users. *Journal of the Korean Biblia Society for Library and Information Science*, 24(4), 115-131. <http://dx.doi.org/10.14699/kbiblia.2013.24.4.115>
- Lee, Jae Yun (2007). Improving the performance of document clustering with distributional similarities. *Journal of the Korean Society for Information Management*, 24(4), 267-283. <http://doi.org/10.3743/KOSIM.2007.24.4.267>
- Lee, Jae Yun (2008). Analyzing the network of academic disciplines with journal contributions of Korean researchers. *Journal of the Korean Society for Information Management*, 25(4),

327-345. <http://doi.org/10.3743/KOSIM.2008.25.4.327>

Park, Young-Ae, & Lee, Jae Yun (2010). An experimental study on small library collection evaluation utilizing circulation statistics and interlibrary loan data. *Journal of the Korean Society for Library and Information Science*, 44(2), 333-356.

<http://doi.org/10.4275/KSLIS.2010.44.2.333>

Ryoo, Jong-duk (2013). An analysis on interlibrary loan network of public libraries in Gyeonggi Province. *Journal of the Korean Society for Information Management*, 30(2), 83-99.

<http://dx.doi.org/10.3743/KOSIM.2013.30.2.083>

Yang, Ji-Ann (2017). A study on use behaviors of books through circulation data analysis in an academic library. *Journal of Korean Library and Information Science Society*, 48(2), 263-293.

<http://dx.doi.org/10.16981/kliss.48.201706.263>