

基于引文分析的交叉学科 资源建设研究*

□ 沈丹 周锐 王晓丽 / 中国科学技术信息研究所 北京 100038
李俊莉 / 中国科学技术信息研究所 北京 100038
/ 河南警察学院 郑州 450000

摘要:当前,图书馆学科资源建设主要按文献类型以及按学科两种方式进行组织,较少关注交叉学科的资源建设。期刊论文是用户研究成果的体现,对期刊引文数据的分析有助于发现用户对不同学科资源的需求。文章以2010-2011年两年的中文科技期刊的引文数据为研究对象,将期刊间的引用转化为学科间的引用,从整体学科层面发现用户对交叉学科的资源需求,结果表明用户对交叉学科资源存在不同程度的需求,资源建设需要考虑交叉学科。

关键词:引文分析,交叉学科,资源建设,用户需求
DOI: 10.3772/j.issn.1673—2286.2013.05.004

1 引言

信息资源建设是图书馆开展信息服务工作的基础,长期以来一直是图书馆实践分析的重要内容。在传统模式下,图书馆资源建设工作主要按文献类型进行组织,由采访馆员对所有学科进行统一采选。近二十年来,随着学科不断细分,采访馆员无法对所有学科进行跟踪了解。为实现对信息资源的知识化、专业化管理,学科化资源建设组织模式受到越来越多的关注^[1],学科馆员制度被国内外图书馆普遍采用^[2],目的是利用学科馆员熟悉对口学科特点、了解专业信息资源分布的优势来提高资源建设质量^[3]。

信息时代图书馆不仅要起到资源保障的作用,更需要随时跟踪学

科发展的前沿,主动为用户提供学科化的信息服务。学科馆员制度在实践工作中取得了一定的成效,但仍存在以下不足:首先,对国内期刊的总体学科交叉研究不多。科学技术的发展越来越呈现出多学科交叉渗透、高度融合的趋势,学科通过引入、吸收、整合其他学科理论、方法来促进学科自身的发展是学科创新最常见的途径^[4]。国外在研究跨学科方面已经取得了不少成果, Van Leeuwen基于ISI数据的期刊分类利用期刊互引进行了跨学科研究^[5]。Leydesdorff则通过ISI期刊分类互引网络构建了学科研究图谱^[6]。目前国内跨学科研究主要基于ISI数据或集中在某一或某几个学科范围内。国内期刊的总体跨学科引用情况如何,以引用为度量

方式的中文期刊相似性怎样等中文资源建设问题,还没有很好的结论。其次,在了解用户需求方面,学科馆员一般通过直接推荐的形式或从访谈中获取用户需求,受访对象知识面以及访谈对象数量的限制,推荐的范围有限,无法全面了解到用户对交叉学科的资源需求。因此,要进一步推进信息资源学科化建设工作,除依靠学科馆员之外,有必要寻求其他方法来加以决策支撑。

引文分析是研究学科知识交流的有效方法。文章以2010-2011年两年的大规模中文科技期刊的引文数据为研究对象,分析当前我国学科交叉总体状况并提出基于引文分析的资源建设方法,为国内交叉资源学科资源建设、期刊学科标识等提供思路和决策参考。

* 本文系中国科学技术信息研究所科研预研项目“数字环境下科技查新引业务创新模式研究”(编号:YY201203)的研究成果之一。

2 研究思路与数据

2.1 研究思路

引文分析是研究学科知识交流的有效方法,大体可分为论文引用、期刊引用、学科引用三个层次^[7]。当前基于引文分析的交叉学科研究主要从论文引用以及期刊引用的角度进行,且多为对个别学科领域的研究^[8,9],基于引文分析的资源建设研究多为从文献保障程度的角度对某一个机构的资源建设情况进行分析^[10,11]。本文利用期刊引文数据开展交叉学科资源建设研究,主要从学科引用的角度进行分析,其思路主要包括以下几点:(1)用户引用行为是用户信息需求的外在表现形式,通过分析用户引用行为可以直接地了解用户对文献资源的需求;(2)引文数据中包含了用户对非本领域学科的资源需求,将论文间的引用统一到论文所在期刊间的引用,进而转化为期刊所属学科间的引用,有助于从整个学科范围更为宏观地了解用户需求,为交叉学科的资源建设提供参考。

本文提出的基于引文分析的交叉学科资源建设研究,主要通过以下几个步骤来进行:(1)从引文数据中抽取能反映用户信息需求和资源引用状况的相关信息,包括引用

期刊、引用期刊所属学科、被引期刊、被引期刊所属学科、引用次数等,并对期刊所属学科为空的记录补充学科分类号。期刊的学科分类参考2011年版《中国期刊引证报告(扩刊版)》^[12]中对期刊的学科分类,具有多个分类号的期刊,取第一主要学科分类号。(2)对期刊所属学科分类号进行规范化处理,统一到中图分类法的一级类目级别(工程技术类规范到二级类目),并与类目名称一一对应。(3)将期刊间的引用转化为学科间的引用,得到学科间引用矩阵比例表,分析学科之间是否明显存在交叉引用的情况。(4)若存在,进一步分析并找出具有相同引用特征的学科群,为典型学科在交叉学科的资源建设上提出针对性的建议。

2.2 数据描述

本文利用2010-2011年两年的中文期刊引文数据进行分析,共有2,009,320条期刊引用记录,涉及引用次数8,156,700次,数据清洗结果如表1所示。经清洗,约15%的数据为无效数据,这部分主要为大学学报类的综合性期刊,因涉及学科范围广而无法界定其所属学科类别。最终获取到有学科分类的引用记录1,498,884条,两年引用次

数共6,951,893次。其中,人文与社会科学类学科涉及的引用次数为1,110,878,仅占总次数的13.6%,其余均为理、工、农、医类学科。本文仅抽取理、工、农、医类26个学科的引用情况加以分析,共涉及引用次数5,841,015次。

3 分析过程及结果

3.1 学科自引与他引对比分析

学科自引与他引对比分析的目的在于验证学科之间是否明显存在交叉引用的情况,资源建设是否需要考虑交叉学科。将引用次数按学科进行分类求和,得到26个学科之间的引用次数矩阵。考虑到不同学科的引用总次数相差较大,对引用次数进行百分比处理,得出学科之间的引用比例矩阵。在矩阵中,斜对角线上的百分数为学科自引比例,除此之外,比例最高的学科即为他引最多的学科,是最值得研究的交叉学科对象。将自引百分比及他引最多学科的百分比从矩阵中抽取出来,进行对比分析,结果如图1所示。

图1中,蓝色为自引比例,红色为除自引外引用最多的第一他引学科。从中可以看出,理、工、农、医类学科引用最多的均为本学科,但自引比例有高有低,具体可分为以下几种情况:

(1)学科引用行为主要发生在学科内部,对其他学科的引用较少且分散。包括电工技术、天文学地球科学、石油天然气工业等,尤其医药卫生,90%以上的引用行为均发生在学科内部。

(2)学科引用行为主要发生在

表1 引文数据清洗结果

	引用记录数	学科类别	引用次数总和	所占比例
有效数据	1498884	人文与社会科学类 理工农医类	1110878 5841015	13.6% 71.6%
无效数据	510436	无	1204807	14.8%
总计	2009320	—	8156700	100%

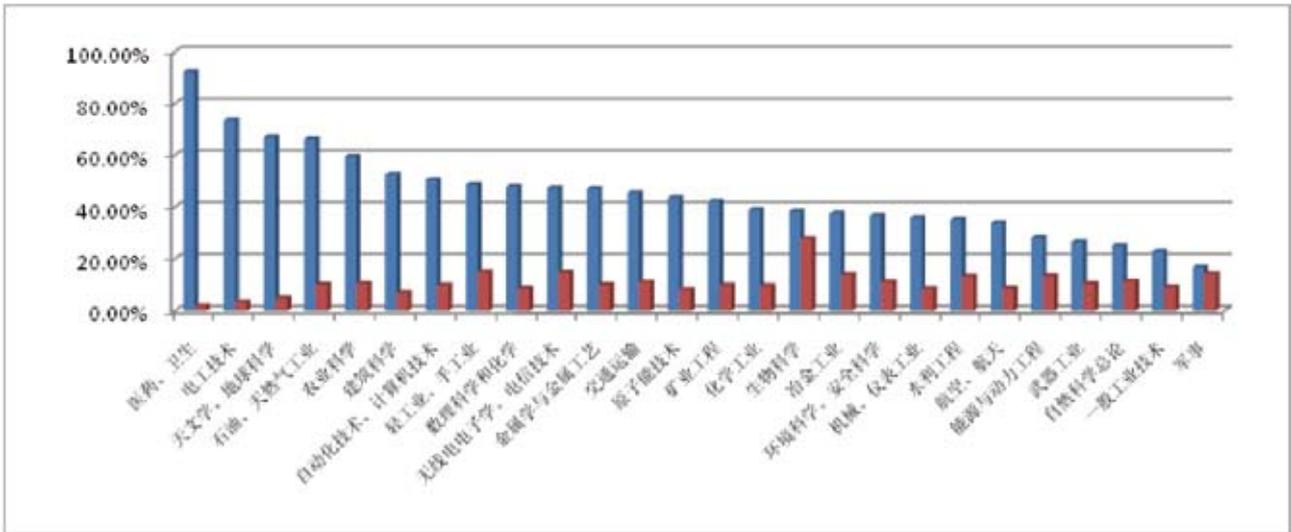


图1 理、工、农、医类学科自引及第一他引对比分析图

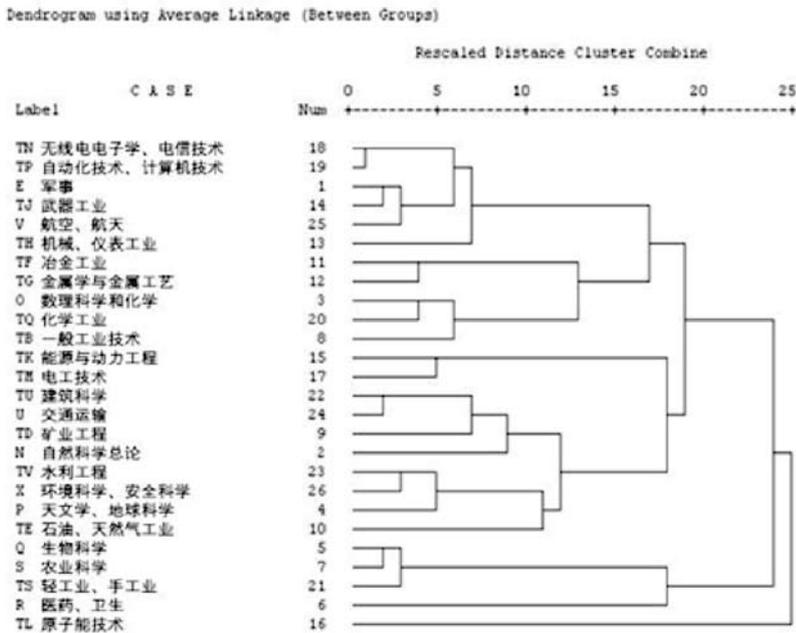


图2 学科相似度系统聚类图

存在不同程度的需求。

3.2 学科聚类

学科分析的结果仅对学科间引用进行了初步的了解，在学科引用比例矩阵的转化过程中发现，不同的学科会同时引用相同的某一个或几个学科，即学科共引。通过学科共引分析，能揭示出不同学科类别之间的交叉关系及相似性，发现联系较为紧密的学科群，为交叉需求学科的确立提供参考。对学科分析过程形成的引用矩阵比例表进行学科共被引分析，对引用比例矩阵进行聚类分析，对对角线数值统一取最大值，在距离测度的计算方式上，采用夹角余弦距离。聚类分析的结果如图2所示。

从图2中可以看出，同一学科群的学科在引用行为上存在相似性，学科之间存在亲疏关系。下面分别抽取典型学科群的引用比例矩阵表，对聚类结果进行讨论：

(1) 军事类相关学科群

学科内部，但对其他个别学科资源也有一定的需求。包括轻工业手工业、无线电电子与电信技术、冶金工业、水利工程等。尤其生物科学，对农业科学资源的需求已经接近自身学科。

(3) 学科引用行为较为分散，

自引比例低，对其他学科资源的需求明显。包括军事、一般工业技术、自然科学总论、武器工业、能源与动力工程等学科。

由此可见，在理工农医学科范围里，学科间的交叉引用情况普遍存在，用户对非本领域学科的资源

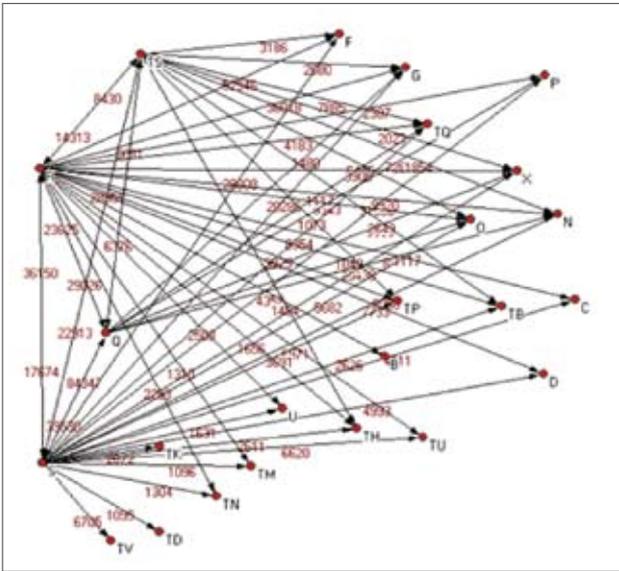


图4 生物类学科群高引用关系网络 (引用频次高于1000)

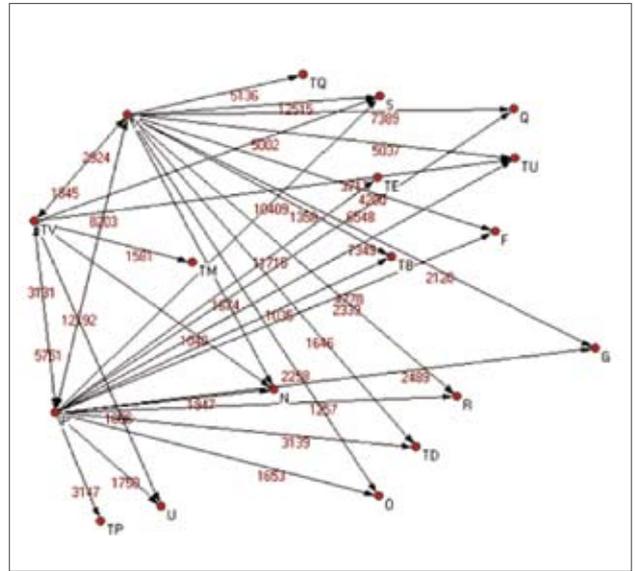


图5 水利工程类学科群高引用关系网络 (引用频次高于1000)

表3 生物科学等三个学科的部分引用数据

学科类别	自引比例	第二引用学科	比例	第三引用学科	比例
生物科学	38.43%	农业科学	27.90%	医药卫生	16.68%
农业科学	59.57%	生物科学	10.63%	医药卫生	4.56%
轻工业、手工业	48.83%	农业科学	15.07%	医药卫生	9.41%

表4 水利工程等三个学科的部分引用数据

学科类别	学科分类号	自引比例	第二引用学科	比例	第三引用学科	比例	第四引用学科	比例	第五引用学科	比例
水利工程	TV	35.25%	P	13.32%	S	11.55%	TU	8.48%	X	7.02%
环境科学、安全科学	X	36.78%	S	11.24%	P	11.05%	Q	7.11%	TQ	4.65%
天文学、地球科学	P	67.12%	TE	5.00%	S	4.45%	X	3.53%	TU	3.14%

(3) 水利工程类相关学科群
水利工程类相关学科主要包括水利工程、环境科学两个学科。在3.1的分析中,天文学、地球科学为

自引比例较高的学科,但其作为水利工程及环境科学的高被引学科,从而在聚类过程中与二者被划分到一起,形成水利工程类相关学科

群。从图5和表4可以发现天文学、地球科学引用其他学科不少但总量比例较低;农业科学(S)为水利工程和环境科学的高被引学科;水利

工程学科引用的其他学科要明显低于环境科学和天文学、地球科学学科。

表5分别给出了农业科学被水利工程和环境科学安全科学两个学科引用最多的十种期刊。

表5 农业科学被引前10期刊

被引次数 排序	期刊名称 (农业科学)
1	农业环境科学学报
2	水土保持学报
3	农业工程学报
4	中国农村水利水电
5	土壤学报
6	安徽农业科学
7	土壤
8	水土保持研究
9	土壤通报
10	灌溉排水学报

此外,建筑科学和交通运输之间、冶金工业与金属学金属工艺之间、数理科学和化学与化学工业之间的学科联系同样密切。而自引比例最高的医药卫生自成一类,相同情况的有原子能技术,虽然自引比例不太高,但由于除自引外引用的学科极为分散,且不被其他学科引用,因此也自成一类。总体来说,聚类结果与实际数据相符。

4 结论与对策

前文分析结果表明,许多学科之间有着非常紧密的联系,因而在信息资源建设工作中必须考虑学科交叉情况,以实现资源建设的全面性。从分析中可以发现,学科

之间的交叉引用大致可分为三种情况:(1)相互引用,例如无线电电子技术和自动化计算机技术之间、生物科学和农业科学之间;(2)链式引用,例如环境科学安全科学引用农业科学最多,而农业科学引用生物科学最多;(3)单向引用,例如医药卫生和天文学地球科学,被相关学科引用较多,但极少引用其他学科。就上文分析结果,对交叉学科资源建设从以下几个方面进行初步探讨。

对交叉资源需求明显的学科。例如军事、武器工业以及航空航天三个学科,除考虑互相补充外,还应多加吸收无线电电子技术与计算机技术的相关资源;生物科学、农业科学除互相补充外,还应多加吸收医药卫生类学科的相关资源,轻工工业对农业科学类特定资源的需求同样值得考虑。水利工程和环境科学除应互相补充外,还可着重考虑天文学地球科学以及农业科学类的相关资源。金属学与冶金工业、建筑科学与交通运输之间应考虑互相吸收等等。实际工作中,有必要进一步找出这些学科交叉引用次数较高的期刊,为资源建设提供直接参考,同时也可作为推荐资源向用户提供推送服务。

对交叉学科资源需求不明显的单向引用学科。例如医药卫生、电工技术、天文学地球科学等学科,主要引用行为都发生在本学科内部,这部分学科的引用行为有待在学科分类上进一步细化,发现其在各自学科内部的交叉引用和资源需求,为用户提供更为细化和深入的服务。

在图书馆在交叉学科资源建设

上的具体操作方式上,对于大学图书馆或大型的专业图书馆,可将关联学科划分出来后,对其资源建设统筹考虑。负责该学科群的学科馆员联合起来,组成学科服务资源建设小组。学科服务小组的成员作为相关学院的联系人,在了解所负责领域研究人员的需求后,可互相探讨及推介资源,共同建设相关领域的资源。对于小型的专业性研究机构,由于不可能将相关学科的资源补充齐全,有必要将机构用户对交叉学科资源的需求进一步明确化,借助大型图书馆的资源,为用户提供服务。

5 结语

网络环境下,馆藏资源建设逐渐由传统按文献类型遴选转变为以用户需求为中心的学科化资源建设,是图书馆以服务为主导的理念在新时代的具体体现。引文数据能客观准确地反映用户的需求,通过引文分析发现交叉需求学科,是学科化资源建设实施中的一项新思路。本文基于用户引文数据的分析,能更全面、客观地了解用户需求,弥补当前学科化资源建设的不足,提高资源建设的水平,更好地促进图书馆学科化服务。

本文的研究还存在几点局限性:首先,仅对两年的引文数据进行了分析,不能全面反映用户的需求;其次,期刊之间的学科引用不足以准确反映文章之间的学科引用;此外,数据的分析还需要与定性分析结合起来,其实用性及实际应用方式还有待深入研究。

参考文献

- [1] 韩丽风,钟建法.图书馆资源建设学科化的组织模式与保障机制研究[J].图书馆工作与研究,2011(8):72-76.
- [2] HOFFERT B. Book Report, Part 2: What Academic Libraries Buy and How Much They Spend [J]. Library Journal, 1998, 123(14): 144-146.
- [3] 韩丽风,张秋.图书馆资源建设学科化的实践与思考[J].图书情报工作,2011,55(7):63-67.
- [4] 魏建香,孙越泓,苏新宁.基于聚类分析的学科交叉研究[J].情报学报,2010,29(6):1066-1073.
- [5] VAN LEEUWEN T, TIJSSEN R. Interdisciplinary Dynamics of Modern Science: Analysis of Cross-disciplinary Citation Flows [J]. Research Evaluation, 2000, 9(3): 183-187.
- [6] LEYDESDORFF L. "Betweenness Centrality" as an Indicator of the "Interdisciplinarity" of Scientific Journals [J]. Journal of the American Society for Information Science and Technology, 2007, 58(9): 1303-1309.
- [7] 彭继东,谭宗颖.纳米科技学科领域的知识交流——基于期刊引文网络的分析[J].图书情报工作,2011(4):15-18.
- [8] 郑惠伶,付小爱,李荣明,等.基于引文的学科文献需求特征分析[J].情报理论与实践,2012(8):48-51.
- [9] 魏海燕,尹怀琼,刘莉.基于引文分析的情报学与相关学科的研究[J].情报杂志,2010(2):38-43.
- [10] 张家生,吴卫华,黄转红.基于引文分析的专业学科信息资源建设研究[J].图书馆论丛,2012(2):19-22.
- [11] 李海霞,刘贵占.引文分析法在图书馆期刊资源建设中的应用研究[J].情报理论与实践,2012(1):99-102.
- [12] 曾建勋.中国期刊引证报告(扩刊版)2011年版[M].北京:科学技术文献出版社,2011:364-459.

作者简介

沈丹(1986-),女,硕士,研究方向:外文书科技资源建设。E-mail: shend@istic.ac.cn

Interdisciplinary Resources Construction Research Based on Periodical Citation Analysis

Shen Dan, Zhou Rui, Wang Xiaoli / Institute of Scientific and Technical Information of China, Beijing, 100038
Li Junli / Institute of Scientific and Technical Information of China, Beijing, 100038
/ Henan Police College, Zhengzhou, 450000

Abstract: Currently the construction of disciplinary resources is mainly organized by document types or disciplines in the libraries. The construction of interdisciplinary resources is less concerned. Periodical papers reflect users' research, and users' resources demand can be found by citation analysis. Taking the citations of Chinese periodicals between 2010 and 2011, this paper transforms the citations between periodicals to citation between subjects, in order to find out the resource demand on interdiscipline. The results show that users demand exists in different degree of interdiscipline and interdisciplinary resource construction is needed.

Keywords: Citations analysis, Interdiscipline, Resource construction, User demand

(收稿日期: 2013-03-25)