

高校学生用户开放存取知识需求分析*

晁亚男 毕强 滕广青

摘要 从国内高等院校学生用户知识需求的角度出发,通过问卷调查获得数据,采用常规统计与形式概念分析相结合的方法,对高校学生用户开放存取知识需求进行综合分析。结果显示,开放存取的检索效率是高校学生用户共同关注的核心焦点;关注文献的传统形式仍然是部分高校学生用户获取知识过程中的思维定势;经济成本在高校学生用户中受关注程度并不突出;学历层次和研究层次是影响高校学生用户间知识需求差异的主要因素。通过对高校学生用户开放存取知识需求特征的揭示,可从用户知识需求的角度为国内高校开放存取的相关理论研究和实践应用提供借鉴。图3。表4。参考文献11。

关键词 开放存取 知识需求 高校学生 常规统计 形式概念分析

分类号 G252

Analysis on Knowledge Needs of University Student Users about Open Access

Chao Yanan, Bi Qiang & Teng Guangqing

ABSTRACT Starting from the perspective of the student users' knowledge needs in the domestic universities, based on questionnaire survey data, combining conventional statistics and formal concept analysis method, a comprehensive analysis on knowledge needs of university student users about open access is executed. The results show that the retrieval efficiency of open access is the core focus of the common concern of university student users; the traditional formal characters of literature are still mindset in the process of part of the university student users to obtain knowledge; the degree of concern of economic cost is not prominent in the university student users; the academic level and research level are primary causes of the differences in knowledge needs in university student users. By disclosing characters of knowledge needs of university student users about open access, it provides reference for theoretical study and practical application of open access in domestic universities with the view of users' knowledge needs. 3 figs. 4 tabs. 11 refs.

KEY WORDS Open access. Knowledge needs. University students. Conventional statistics. Formal concept analysis.

1 研究背景与相关工作

开放存取(Open Access, OA)^[1]是网络环境下信息交流与知识共享的一种理想模式,具有的方便

性、快捷性、低成本等优势,已为国际学术界所认可,并被迅速地引入到我国数字图书馆与知识工程的建设中。自2005年国内50余所高校联合签署《中国大学图书馆合作与资源共享武汉宣言》以来,国内高校陆续开始建立各具特色的“学术机构

* 本文系教育部人文社会科学研究青年基金项目“基于概念格的高校图书馆开放存取(OA)资源知识服务体系构建研究”(编号:10YJC870035),教育部高等学校博士学科点专项科研基金项目“基于领域本体的开放存取资源在线集成服务门户研究”(编号:20100061110085)和吉林大学“985工程”项目的研究成果之一。

通讯作者:毕强, Email: biqiang12345@163.com

仓储”,以促进学术资源的开放存取。其后的数年间,国内学术界对开放存取相关问题的研究热情持续高涨,部分高等院校和知识机构的开放存取资源建设也初见成效。然而另一方面,我们不得不面对一个事实,即经过数年的曲折发展,许多高校从早期如火如荼的“各自为战”正在逐渐蜕变为今天的“无心恋战”。一些曾经有声有色的高校开放存取网络服务平台(如浙江大学开放存取资源平台等)已经无法访问,更有许多高校(同济大学、北京交通大学等)仅仅提供了部分公共开放存取资源的链接,总体形势不容乐观。造成上述情况的主要原因之一是学生作为高校的主要用户群对于开放存取资源,特别是高校自建的开放存取资源少有问题,这直接反映出高校学生用户的知识需求与高校开放存取资源建设间存在着断层。

用户知识需求及其特征规律的相关研究一直是图书情报学研究的一个热点领域。早在1962年,Robert S. Taylor就提出了用户知识需求的“四层次”理论^[2]。Taylor从用户个体的角度出发,认为知识需求首先是内在需求,这一层次上的需求是真实的但是未能表达出来的需求;处于其外层的是意识需求,能够在用户的大脑意识中予以描述却难以言表;再外层是形式化的需求,这一层次的知识需求可以实现形式化的表述;最外面一层则是提交在检索系统中的具体提问。其后,进一步研究发现用户的知识需求与用户期望的目标是相联系的^[3],不同的用户由于其期望的目标不同,在具体的知识需求上会体现出很大的差异。国内学者从虚拟社区的角度出发,认为用户知识需求具有社会影响性、时间迁移性和需求层次性的特征^[4]。近年来随着开放存取运动的发展,用户知识需求的相关研究又面临着新的挑战。一方面,国外相关研究表明,尽管开放存取文献的下载量明显优于传统期刊,但是开放存取文献的被引率却处于较低的水平,且低于传统期刊^[5]。而另一方面,国内学者的研究则表明,开放存取资源在下载总频次、总被引频次、影响因子、即年指数等方面均优于非开放存取期刊^[6]。显然,不同条件下不同用户群对于开放存取的知识需求存在较大的差异。其中高等院校作为特殊的

用户群体,经常被作为开放存取相关研究的调研对象,但是专门针对高校学生用户的开放存取知识需求问题却鲜有研究。

本文从国内高等院校学生用户知识需求的角度出发,通过对不同学校类别、不同学历阶段、不同学科背景的学生在开放存取方面的知识需求进行问卷调查,获取基础数据,采用常规统计方法对其中具有代表性的用户知识需求特征进行了分析。在此基础上,借助国际数据分析领域新兴的形式概念分析(Formal Concept Analysis, FCA)^[7]的理论与方法,构建高校学生用户知识需求概念格(Concept Lattice),借助形式概念分析对人类认知、思维、争论、交流等活动的支持,在呈现高校学生用户开放存取资源知识需求特征的同时,揭示其背后的关联规则,以期从用户知识需求的角度为国内高校开放存取的理论研究和实践应用提供借鉴。

2 基于常规统计学的用户知识需求分析

2.1 数据采集与预处理

研究中,首先针对开放存取资源的方便性、快捷性、低成本、权威性、创新性等特征在国内高等院校学生用户群中进行需求特征问卷调查。被调查的对象按所在院校类别划分为一般院校、211高校以及985高校的学生;按照用户身份划分为博士研究生、硕士研究生、本科生;按照用户所处学科划分为文史类和理工类。各类院校的差异主要体现在经费支持和培养目标上,985高校一般具有较好的经费支持(馆藏资源丰富、购买数据库多)和较高的研究型人才培养目标(研究生比例大、文献需求层次高),211高校和一般院校的相关指标依次递减。但调研中也包括一些东南沿海经济发达地区的一般院校,其文献资源建设经费由于获得地方财政支持也较为宽裕。调研中共发放问卷583份,收回有效问卷531份,回收率为91.1%。问卷中,被调查者可对知识诉求点进行多项选择,即被调查的学生可以在“低成本”、“被引率”等多个诉求点上同时选择“关注”。由于问卷数据量过于庞杂,需要根据用户对知识需求特征的支持程度设置一定

的数据筛选阈值,对问卷中的知识需求特征进行预处理。预处理过程中,如果阈值选择过大,则仅能囊括极少量的知识需求特征(如阈值大于 60% 则仅有 4 项知识需求特征),且一些知识需求特征仅有某一类用户数据支持。如果阈值设置过小,则一

些原本显著的知识需求特征会湮没在其他特征中难以得到彰显。研究中设置数据筛选阈值为 25% (至少有 25% 的某类用户提出该项诉求),得到国内高校学生用户开放存取知识需求调查汇总表(见表 1)。

表 1 开放存取知识需求调查汇总表(节选)

序号	用户知识需求诉求点	院校类别			用户身份			所处学科	
		一般/198	211/231	985/102	博士/97	硕士/253	本科/181	文史/327	理工/204
T1	低成本或免费	85	51	21	26	25	106	109	48
T2	发文时间(新)	7	48	30	50	29	6	39	46
T3	一站式集成	122	192	87	88	191	122	235	166
T4	传统期刊刊载	79	119	89	65	180	42	213	74
T5	权威期刊刊载	25	61	32	42	69	7	77	41
T6	文章创新性	43	126	96	89	131	45	127	138
T7	文章被引率	10	80	77	22	93	52	88	79

由于篇幅所限,表 1 中没有列示出调查中涉及的全部用户知识需求,仅将其中满足汇总阈值(25%)要求的 7 项诉求点及其相关数据集中进行了列示。将表 1 中的数据视为各个用户子群在相关问题上的得分,计算获得克隆巴哈(Cronbach) Alpha 信度系数估计值为 0.823,基于标准化项的克隆巴哈(Cronbach) Alpha 信度系数为 0.836,由于信度系数在 0.80—0.90 之间,说明表 1 中经过预处理的用户知识需求诉求点具有较强的内部一致性。同时,采用因子分析法对结构效度进行检验,结果显示,当提取 3 个公共因子时,所提取的 3 个公共因子特征值的累计贡献率达到 92.419%,且经过阈值筛选的用户知识需求诉求点的结构与理论预设的经济维度、使用维度、学术维度的基本结构一致,结构效度达到了要求。

2.2 用户知识需求的常规统计分析

对上述调研数据进一步分析发现,表 1 中,7 项开放存取用户知识需求特征(诉求点)在调查对象中被诉求的程度不尽相同,其总体情况如图 1 所示。

图 1 中,横轴为 7 项知识需求诉求点,纵轴为

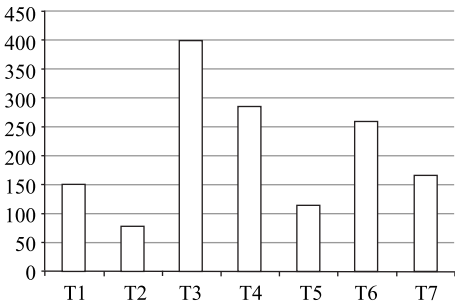


图 1 高校学生用户开放存取知识需求特征总体情况

表达诉求的人数。可以看出,大多数被调查用户对开放存取资源的“一站式集成(T3)”表达了强烈的诉求,占被调查对象的 75.6% (401/531),其后依次为“传统期刊刊载(T4)”占 54.0% (287/531)、“文章创新性(T6)”占 49.9% (265/531)、“文章被引率(T7)”占 31.5% (167/531)等。以“一站式集成(T3)”这一最具显著性的用户知识需求特征为例,依据常规的分类汇总方法,从用户所在“院校类别”、“用户身份”、“所处学科”三个方面做进一步分析,不同的“用户身份”对开放存取提出“一站式集成(T3)”诉求的分布情况如图 2 所示。

从图 2 中可以看出,提出“一站式集成(T3)”

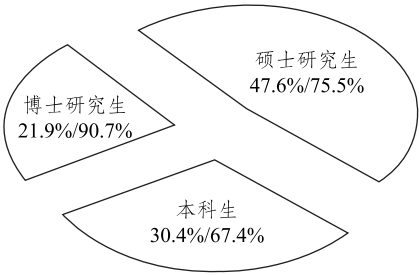


图2 “一站式集成”诉求点的用户身份分布

诉求的用户群依次是:“硕士研究生”占全部用户的47.6%(191/401),“本科生”占30.4%(122/401),“博士研究生”占21.9%(88/401)。表面上看,处于学历中间阶段的“硕士研究生”是提出“一站式集成(T3)”诉求的主要用户群。然而,结合表1中各用户子群参与调查的用户数量做进一步分析可以发现,在所有被调查的“硕士研究生”中,提出“一站式集成(T3)”诉求的占75.5%(191/253)。同时,这一指标在“本科生”中为67.4%(122/181),在“博士研究生”中为90.7%(88/97)。这一方面说明有超过一半的用户提出该项诉求,解释了图1中“一站式集成(T3)”诉求点高居榜首的原因。另一方面,分析结果显示随着被调查用户群学历层次的提高(本科生→硕士研究生→博士研究生),对开放存取“一站式集成(T3)”的诉求表达越发强烈。从这个意义上讲,开放存取的“一站式集成(T3)”知识诉求点主要是被“博士研究生”(90.7%)用户子群拉高了比率。或者可以这样说,用户的学历层次越高越会追求文献检索的效率。

采用同样的分类汇总思想(基于比率),可以对表1中的汇总数据进行更全面的分析。结果显示:大多数用户并不关心“低成本或免费”,只有“一般院校”和“本科生”两个用户子群略显在意;除部分“博士研究生”外,大多数用户并不在意“发文时间(新)”;很多用户注重“传统期刊刊载”,且在“985院校”和“文史类”用户中表现比较突出,但是否为“权威期刊刊载”则少有关注;“985院校”和“博士研究生”两个用户子群都关注“文章创新性”,且“理工类”用户要超过“文史类”用户;对于

“文章被引率”的关注不是很明显,但“985院校”用户子群中的诉求比率要高于其他用户群。

然而,常规分类汇总方法尽管能在各口径指标内部的分析中较好地发挥作用,但在揭示各指标间的关联关系时却显得相对薄弱。要对开放存取的各类用户知识需求之间以及不同用户子群之间做更深入的关联分析,就必须引入能够更有力地支持关联分析的数据分析方法。

3 基于形式概念分析的用户知识需求关联分析

3.1 用户知识需求概念格的构建

形式概念分析的理论与方法自诞生以来,其在关联规则挖掘领域中的优势已被学术界所共识^[8]。采用形式概念分析方法进行数据分析,首先需将调查汇总数据进行形式化处理,构建用户知识需求的形式背景。研究中仍然采用最初设定的数据筛选阈值(25%),表1中某一知识需求诉求点如在某一用户子群中被超过25%的用户提出,则视为该诉求点满足阈值要求。以用户子群为对象,用户知识诉求点为属性,根据表1基础数据构建用户知识需求单值形式背景(见表2)。

表2 高校学生用户开放存取知识需求形式背景

	T1	T2	T3	T4	T5	T6	T7
U1	×		×	×			
U2			×	×	×	×	×
U3		×	×	×	×	×	×
U4	×	×	×	×	×	×	
U5			×	×	×	×	×
U6	×		×				×
U7	×		×	×		×	×
U8			×	×		×	×

注:U1:一般院校,U2:211院校,U3:985院校,U4:博士研究生,U5:硕士研究生,U6:本科生,U7:文史类,U8:理工类

表2中,一个形式概念是二元组(U,T),其中

用上文中设定的规则筛选阈值,获得高校学生用户子群间关联规则(见表4)。

表4 高校学生用户子群间关联规则

规则	条件	结论	支持度(%)	置信度(%)
RU01	U2	U3 U5	71	100
RU02	U5	U2 U3	71	100
RU03	U4	U3	86	83
RU04	U3	U2 U5	86	83
RU05	{ }	U4	100	86
RU06	{ }	U3	100	86
RU07	{ }	U7	100	71

表4所列示的规则表达式中有些可以揭示出不同高校学生用户子群间关联关系的规则。以关联规则“RU02”为例,该规则表达式为“〈U5〉〈71%〉→〈U2 U3〉〈100%〉”,说明在所有具有显著性特征的知识诉求点中,有71%的开放存取知识诉求点被“U5(硕士研究生)”这一用户子群所关注,而这部分被“U5”用户子群所关注的知识诉求点同时还全部(100%)被“U2(211院校)”和“U3(985院校)”这一用户子群所关注。再以关联规则“RU03”为例,该规则表达式为“〈U4〉〈86%〉→〈U3〉〈83%〉”,说明在所有具有显著性特征的知识诉求点中,有86%的开放存取知识诉求点被“U4(博士研究生)”这一用户子群所关注,其中有83%的知识诉求点还同时被“U3(985院校)”这一用户子群所关注。所以,用户子群间的关联规则能够反映出不同用户子群间开放存取知识需求的相似性与差异性,进一步揭示出不同用户子群对开放存取知识需求的诉求差异与联系,为针对高校学生用户群提高服务水平提供了依据。

4 关于高校学生用户开放存取知识需求的讨论

将表3和表4中的两组关联规则以及前述常规统计的汇总结果进行综合,可以初步得出国内高校学生用户对于开放存取的知识需求具有以下

特征:

首先,开放存取的检索效率是高校学生用户共同关注的核心焦点。调查中,“一站式集成”是检索效率的直接体现,图1中“一站式集成”是大多数高校学生用户对于开放存取最具显著性的诉求,占全部被调查用户的75.6%。并且在各个用户子群中均有超过一半以上的用户提出该项诉求(见表1),分布比较均匀。在进一步获取的关联规则(见表3)中表明,“一站式集成”涉及所有满足阈值要求的关联规则。因此,开放存取的检索效率是被高校学生用户所共同关注的核心焦点。这一结果进一步反映了开放存取实践中,用户对资源规模的偏好。尽管一些高校或组织自建的开放存取资源很有特色,但如果用户为了获取特定的知识信息不得不同时在几个不同的资源平台之间转换,就会在一定程度上消减用户对于开放存取的热情。

其次,关注开放存取文献的传统形式仍然是部分高校学生用户知识获取过程中的思维定势。开放存取原本以预出版或网络出版见长,但是调查中有54.0%的用户仍然会关注文献是否被“传统期刊刊载”(见图1)。这一结果在表面上看似有悖常理,其本质则反映了高校学生用户对于文献质量的诉求。表3中除“RT03”和“RT04”两条规则外其余5条关联规则均显示出“传统期刊刊载”与“文章被引率”、“文章创新性”存在不同程度的关联。这进一步反映了以完成学业和学术研究为目的的高校学生用户在知识获取过程中对文献质量有着较高的要求,并且在没有更好、更权威的开放存取评价指标的情况下,许多用户以“传统期刊刊载”作为反映文献质量的参考因素。在针对这一问题的补充调查中,也有部分用户表示,提出这一知识需求的原因是为了便于在研究成果利用时能够按照传统方式对文献进行标引。事实上,其本质还是出于用户对开放存取文献质量的忧虑。

再次,开放存取的经济成本在高校学生用户中的受关注程度并不突出。调查中,作为开放存取主要特征的“低成本或免费”的受关注程度远低于其他需求,并没有成为高校学生用户知识需求的关注焦点。数据显示仅有部分“一般院校”、“本科生”、

“博士研究生”、“文史类”用户提出该项诉求。因为对于大多数高校学生用户而言,学校已经购得的数据库,在免费使用的情况下,开放存取的经济成本优势从个体用户的角度被淡化了。当然,在一些资金投入有限的院校中,“低成本或免费”还是受到了部分用户的关注(见表1),但是其关联性很弱,在表3和表4两组具有一定阈值筛选要求的关联规则中均没有体现。补充调查显示,如果没有学校统一付费购买的数据库,会有更多的用户关注“低成本或免费”。同时,对于数据中“博士研究生”关注这一诉求的现象,补充调查显示是由于部分博士研究生不在校园内生活,而一些院校没有对学生开通VPN,学生无法在校园以外利用已付费的数据库。

最后,学历层次和研究层次是影响高校学生用户间知识需求差异的主要因素。针对某一特定的知识需求,不同用户会依据其学历层次和研究层次表现出不同的偏好。图2中从本科生到硕士研究生,再到博士研究生对“一站式集成”表现出的关注程度递增的趋势就很好地说明了这一点。表4中的关联规则(“RU03”)显示,“博士研究生”用户子群与“985院校”用户子群之间具有十分紧密的关联关系。由于被调查对象中的博士研究生主要来自985院校,因此两个用户子群在学历层次和

研究层次上相对接近,其表现出的知识需求也比较相似。此外,“硕士研究生”和“本科生”两个用户子群与“博士研究生”用户子群的关联关系依次减少,表明高校学生用户中各个用户子群开放存取的知识需求差异在很大程度上受到学历层次和研究层次的差异性影响。

5 结语

通过以上对国内高校学生用户开放存取知识需求的综合分析可以看出,对于高校学生用户这一特殊的用户群体,由于各高校在不同程度上采用统一付费的方式购买数据库供其用户使用,使得原本作为开放存取最主要特征优势之一的“低成本或免费”在高校学生用户群中被淡化,而在以往研究中被忽略的检索效率、文献质量等需求特征则凸显出来。尽管调研中的数据尚不足以覆盖国内高校学生用户开放存取知识需求的全貌,但这却正是我国当前高校信息资源建设中传统期刊(含数据库)与开放存取并行的现实反映。如何在这一具体环境下,以最大程度满足用户知识需求为导向,调整和推进我国开放存取建设实践,是今后研究的热点方向。

参考文献:

- [1] Suber Peter. Open access overview[EB/OL]. [2012-5-25]. <http://www.earlham.edu/~peters/fos/overview.htm>.
- [2] Taylor Robert S. The process of asking questions[J]. American Documentation, 1962, 13(4): 391-396.
- [3] Green A. What do we mean by user needs?[J]. British Journal of Academic Librarianship, 1990, 5(2): 65-78.
- [4] 滕广青,董立丽,田依林,等. 基于概念格的社区用户知识需求模型研究[J]. 情报科学, 2011, 29(1): 108-112. (Teng Guangqing, Dong Lili, Tian Yilin, et al. Research on demand model of knowledge about community users based on concept lattice[J]. Information Science, 2011, 29(1): 108-112.)
- [5] Sotudeh H, Horri A. The citation performance of open access journals: A disciplinary investigation of citation distribution models[J]. Journal of the American Society for Information Science and Technology, 2007, 58(13): 2145-2156.
- [6] 程维红,任胜利,刘旭. 我国农学期刊网上学术影响力分析[J]. 中国科技期刊研究, 2006, 17(4): 555-558. (Cheng Weihong, Ren Shengli, Liu Xu. Analysis on online academic influence of journal of agricultural science in China [J]. Chinese Journal of Scientific and Technical Periodicals, 2006, 17(4): 555-558.)
- [7] Wille R. Restructuring lattice theory: An approach based on hierarchies of concept[C] // Rival I. Ordered Sets. Boston: D Reidel Publishing Company, 1982: 445-470.

- [8] 毕强,滕广青. 国外形式概念分析与概念格理论应用研究的前沿进展及热点分析[J]. 现代图书情报技术, 2010, 26(11): 17-23. (Bi Qiang,Teng Guangqing. Analysis of the progress and hotspots in applied research of FCA and concept lattice theory abroad[J]. New Technology of Library and Information Service, 2010, 26(11): 17-23.)
- [9] Stumme Gerd. Efficient data mining based on formal concept analysis[C] //Hameurlain A,Cicchetti R,Traunmüller R. Database and Expert Systems Applications. London: Springer-Verlag, 2002: 534-546.
- [10] 滕广青,毕强. 概念格构建工具 ConExp 与 Lattice Miner 的比较研究[J]. 现代图书情报技术, 2010, 26(10): 17-22. (Teng Guangqing,Bi Qiang. Comparative study on ConExp and Lattice Miner[J]. New Technology of Library and Information Service, 2010, 26(10): 17-22.)
- [11] 甘特尔 B,威尔 R. 形式概念分析[M]. 马垣,张学东,迟呈英,等,译. 北京: 科学出版社, 2007. (Ganter B,Wille R. Formal concept analysis [M]. Ma Yuan, Zhang Xuedong, Chi Chengying, et al., Trans. Beijing: Science Press, 2007.)

晁亚男 吉林大学管理学院博士研究生。通讯地址:长春市人民大街 5988 号。邮编:130022。

毕 强 吉林大学管理学院信息管理系教授,博士生导师。通讯地址同上。

滕广青 吉林大学管理学院博士研究生。通讯地址同上。

(收稿日期:2012-06-01)