

基于 OPAC 统计数据的借阅率提升策略探讨*

□ 孟德泉 董颖 沙娅弘 孟桂荣

摘要 目前图书馆馆藏资源的借阅率普遍偏低,利用布拉德福定律对外借图书进行分区,可以发现导致低借阅率的影响因素,提升馆藏图书的利用率。该文借助问卷调查,利用主成分分析方法得出影响用户借阅行为的因素:作者、出版社、出版日期等。选取“TP313 汇编程序”这一领域的外借图书数据,首先利用定量分析方法和专家咨询方法验证前述因素如何影响图书的外借率,然后根据研究结果提出具体建议。

关键词 使用挖掘 借阅行为 布拉德福定律

1 概况

由于网络化、电子化资源的发展和人们阅读方式的转变等多种因素的影响,许多图书馆的借阅率呈现下滑趋势。虽然各高校提出众多提高图书利用率的对策,但仍然改变不了图书借阅量逐年下滑的现状^[1](参见图 1 和图 2)。另一方面,根据《2012 年高校图书馆发展报告》,543 所高校图书馆提交了纸质文献资源购置费的统计数据,平均值约为 250 万元(2011 年约为 253 万元),虽然延续了 2010 年以来的持续走低趋势,但是下降幅度并不是很大,因此可以看出图书馆对于纸质资源的购置量依然稳定(如图 3 所示)。

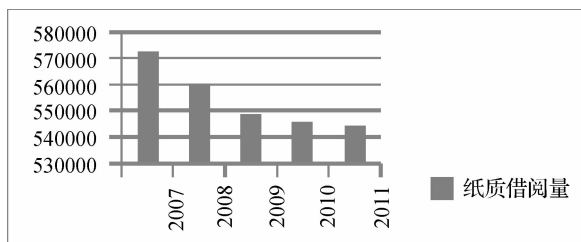


图 1 吉林大学图书馆纸质资源借阅统计数据^[2]

原因没有被借阅。因此在用户借阅图书的时候,可通过技术手段以及其他手段,充分推荐用户借阅那些借阅次数较少但是有价值的图书,从而提高馆藏的利用率。

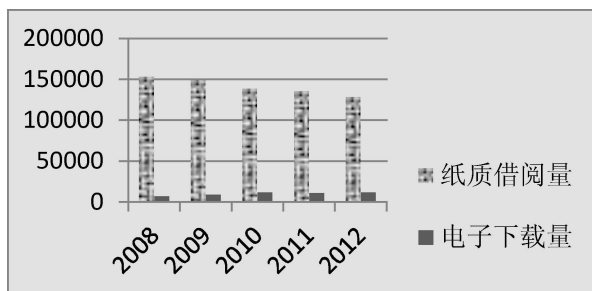


图 2 山东大学图书馆纸质资源和电子资源借阅统计数据^[3]

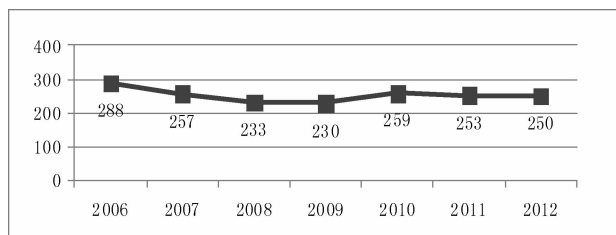


图 3 2006—2012 年高校图书馆纸质文献资源的平均购置费(万元)

然而,图书馆采购的大部分图书被借阅的次数偏少,一些内容质量不错的、有价值的图书由于某些

* 本文为国家社会科学基金项目“数字图书馆动态组合学科服务模式研究”(项目编号:14BTQ017)、教育部人文社会科学研究规划基金项目“团体导向的数字图书馆评价研究机制”(项目批准号:12YJA870003)以及 2012 年哈尔滨师范大学学术团组项目“信息资源管理网络化研究”(项目编号 SYG2012-04;SYB2012-02;SYS2012-03;SYB2012-03)的研究成果之一。

2 布拉德福定律的指导意义

2.1 定律简述

布拉德福定律由英国著名文献学家布拉德福于 1934 年首先提出^[4]：某一学科领域的核心区与相继各区(各区载文量大致相等)的期刊数量成 $1:n:n^2$ ，定量地揭示了科学论文在期刊中的集中与离散分布规律。布拉德福定律是文献计量学中的二八定律，即图书馆 80% 的流通量由 20% 的馆藏来满足^[5]。在核心期刊界定方面，早期的学者从布拉德福定律出发，认为对于某一主题而言，平均每 30 种学术期刊中有 1 种核心期刊^[6]。后来有学者认为，可根据二八定律，选取各学科前 20% 的期刊作为核心期刊^{[7][8]}。对于图书馆的图书资源而言，可以根据布拉德福定律以及二八定律挑选出借阅量排在前 20% 的图书，根据这些图书的特征对后 80% 的图书进行聚类，再次筛选出有较高借阅价值的书，进而提高这部分图书的借阅率。

2.2 基于布拉德福定律的图书借阅分区

根据布拉德福定律和二八定律，80% 来图书馆的读者只是查阅了馆藏图书的 20%。可以从用户借阅行为上考虑，探索影响用户借阅行为的各种因素，找出低借阅量图书与高借阅量图书的共同点，以前 20% 的图书带动后 80% 的图书的借阅量。本文主要根据借阅次数来划分图书的类别：排序在前 20% 的图书借阅次数高，作为对比的参照对象；而后 80% 的图书是本文研究的重点，二次开发和利用也是针对这 80% 的图书来进行。实践上，通过具体案例，借助定量和定性的分析方法来进行实证分析。

3 基于用户借阅行为的 OPAC 资源借阅情况调查与分析

3.1 调查问卷的设计

采用七级等距离结构来设计调查问卷的选项，“1”代表强烈不影响，“7”代表强烈影响，中间值“4”代表基本不受到影响，“>4”代表用户受到影响，“<4”代表用户没有受到影响。问卷包括三部分内容：用户基本情况调查、影响用户借阅行为的因素调查和用户借阅后的评价调查，其中第二部分影响因素的自变量包括 4 类共 27 个。通过文献调研，结合高校图书馆学生和教师科研人员的特殊性，把影响用户借阅行为的因素归纳成四部分：外部因素、推荐因素、书本身的因素、个人偏好因素，并基于以上影响

因素设计问卷的变量。影响借阅的外部因素有：是否有检索系统、检索系统中是否有被查阅的书籍、书架上是否可以找到、书籍是否被借出、借阅次数、是否全部在馆、摆放位置、馆员服务、课堂必备、书籍分类准确性；影响借阅的推荐因素有：导师推荐、专家推荐、网络评分；影响借阅的书本身因素有：题目、封皮设计、封面评论、摘要导言、目录、内容、知名度、实用性；影响借阅的个人偏好因素：作者、出版社、版本、新旧、纸张手感、中英版。

3.2 问卷分析

此次问卷调查以北京理工大学的学生作为主要对象，后面的实证分析也以北京理工大学的馆藏资源作为分析对象。调查问卷采用随机抽样和网络途径发放方式，在图书馆、学生宿舍以及网络上共发放 230 份，调查对象包括本科生、硕士研究生和博士研究生，问卷回收了 226 份，224 份有效。

从均值统计结果来看，检索系统、作者、出版社、推荐、版本等观测变量的度量均值都在 4 分以上，问卷中主要变量的均值分布分别为：检索系统影响 5.82、作者影响 6.11、出版社影响 5.80、导师朋友推荐影响 5.55、权威专家推荐影响 5.59、课堂必备影响 5.92、版本影响 5.71。其他因素，如图书摆放区域、馆员服务态度、图书封皮设计、图书封面评论、图书新旧以及图书纸张等变量的均值较低。具体来说：

(1) 用户受检索系统、作者、出版社、版本、导师朋友推荐、权威专家推荐以及课堂必备图书的影响最为强烈；

(2) 借阅次数、图书分类的准确性、图书的网络购买量、网络评分、图书内容、知名度以及图书中英文版本的影响力强烈；

(3) 图书全部借出或全部在馆、检索系统中是否有该书、题目、摘要和导言、目录、图书新旧以及实用性的影响程度中等；而书架上没有、想借阅的图书已被借出、图书摆放区域、馆员服务态度、图书封皮设计、封皮上的评论以及纸张手感等因素的影响力不大，均值都在 4 分之下。

3.3 影响用户借阅行为的主成分因素分析

影响用户借阅一本图书的原始因素有 27 个，但是真正具有继续研究价值的、方差累计值超过 80% 的影响因子是 15 个，分别是：作者、出版社、出版日期、权威专家推荐、有无检索系统、课堂必备、中英文

版不同、目录、借阅次数、馆员服务态度、图书摆放区域、图书内容、图书分类的准确性、实用性、题目。由旋转后的正交因子载荷矩阵和得分矩阵可以得出：

“作者”、“出版社”和“版本”在旋转成分矩阵中相应的因子载荷量分别为 0.849、0.885 和 0.884，在成分得分矩阵中的载荷量分别为 0.683、0.623 和 0.515，这三者的影响强度都很大，并且相关性最强。基于此，论文拟定这三个因素作为定性和定量分析的变量。

“权威专家推荐”在旋转成分矩阵中相应的因子载荷量为 0.873，在成分得分矩阵中的载荷量为 0.530。因为与导师朋友推荐的相关性很高(0.640)，因此在成分矩阵中导师朋友推荐的载荷量也有 0.506，两者差距不大。可见权威专家和导师朋友推荐共同影响力还是比较强的。这两个因素只能作为定性分析，可以在专家访谈和鉴定的实证分析中引入专家推荐环节。

“有无检索系统”在旋转成分矩阵中相应的因子载荷量为 0.845，在成分得分矩阵中的载荷量为 0.559，载荷量一般，影响程度一般。考虑到图书馆检索系统出现崩溃的状况不会频繁发生，因此对检索系统的影响不做太多考虑。“图书分类的准确性”和“图书的摆放区域”在旋转成分矩阵中相应的因子载荷量分别为 0.810 和 0.805，在成分得分矩阵中的载荷量分别为 0.667 和 0.853，载荷量相对较高，所以影响强度较高。但是对此因素做具体分析的价值不大，只需馆员准确分类和摆放即可。“目录”、“题目(图书内容)”在旋转成分矩阵中相应的因子载荷量分别为 0.883、0.806，在成分得分矩阵中的载荷量分别为 0.658、0.749，影响程度都相对较高。因此在专家鉴定图书的时候，可以采用题目、目录与图书内容考察相结合的方式进行鉴定。“借阅次数”在旋转成分矩阵中的因子载荷量为 0.823，在成分得分矩阵中的载荷量为 0.670，因此影响程度相对较高，并且借阅次数在第四章定量分析以及提取数据的时候是一个重要的参考指标。“馆员服务态度”在旋转成分矩阵中的因子载荷量为 0.805，在成分得分矩阵中的载荷量为 0.691，与初期作者假设的影响程度较低的观点相比，馆员服务态度的影响程度相对较高。因此，馆员如何更好地服务用户是一个值得深思的问题。“实用性”在旋转成分矩阵中的因子载荷量为 0.899，在成分得分矩阵中的载荷量

为 0.871，影响程度很高。关于图书的“实用性”，作者随机询问了 50 个同学，46 人认为“实用性”在阅读完图书后才能判断，图书对自己的帮助越大，实用性越高。而且在采访中，不同的人对于同一本书的实用性判断不尽相同，因此实用性的影响不能进行定量评估，只能根据用户的个人偏好来决定。所以对于实用性的影响不再进行探讨。“课堂必备”在旋转成分矩阵中的因子载荷量为 0.951，在成分得分矩阵中的载荷量为 0.921，几乎所有受访者都认为老师要求课上必须用到的图书，他们肯定会到图书馆借阅，因此影响程度非常高，对于本科生来说更是如此。“中英文版不同”在旋转成分矩阵中相应的因子载荷量为 0.915，在成分得分矩阵中的载荷量为 0.924，说明图书语种对于用户的影响非常高，被调查的用户主要是研究生，大部分同学倾向于阅读中文图书；而本科生用到的英文图书较少，因此这个影响因素主要是针对研究生的。

由以上分析可以发现，“作者”、“出版社”和“版本”是主要的影响因素，将作为下一步定量和定性实证分析的重点。

4 基于 OPAC 统计数据的借阅率提升策略实证分析

选取“TP313 汇编程序”这一类图书作为实证分析的案例。通过定量分析初步筛选图书，并通过定性分析对结果进行验证和加强。

4.1 定量分析

通过 CNMARC 格式著录细则，提取出所需图书的 MARC 码以及所对应的关键字：题目(200)、作者(200)、版本(205)、出版社(210)、出版日期(210)、摘要(330)、学科名称主题(606)等进行对比和分析。按照借阅率从高到低的顺序，划分出前 20% 和后 80% 的图书。

“TP313 汇编程序”类图书截止到 2013 年 11 月 18 日总共有 284 本，借阅量最高的是温冬婵等编著的《IBM PC 汇编语言程序设计例题习题集》，达 422 次。前 20% 的图书 56 本，借阅率最低的有 142 次；后 80% 的图书共有 228 本。

4.1.1 作者因素分析

抽取借阅量在前 20% 的图书，其中 66.7% 的作者所编著的图书也出现在了后 80% 的图书中。如钱晓捷编著的《汇编语言程序设计》的 2000 版、2003 版以及 2006 版的借阅次数高达 184、159 和 161 次；

但是此书 2012 版的借阅次数仅为 8 次。她编著的《32 位汇编语言程序设计》2007 版和 2011 版的借阅次数分别仅为 24 次和 5 次,《基于 Windows 环境的汇编语言程序设计》的借阅次数为 0 次。

这些数据带来一个问题:在同一门类,由同一个作者编著的图书的借阅次数为何会有如此大的差距?也许那些没有被借阅或者借阅次数很低的图书,它们的内容其实非常有价值,只是我们没有通过很好的渠道让用户去发现它们。

4.1.2 出版社和版本因素分析

清华大学出版社的图书最多,总共 53 本;其次是电子工业出版社,共 36 本;另外还有机械工业出版社 25 本、高等教育出版社 14 本、科学出版社 11 本、人民邮电出版社 9 本、中国水利水电出版社 6 本、中国电力出版社 4 本、华中科技大学出版社 5 本、北京理工大学出版社 3 本、冶金工业出版社 3 本、南开大学出版社 2 本、西北工业大学出版社 2 本、天津大学出版社 2 本,以及航空工业出版社 1 本。

尽管清华大学出版社的图书最多,但是借阅次数在前 20% 的图书只有 25%,低于电子工业出版社的 31% 以及机械工业出版社的 28%。而高等教育出版社的借阅次数在前 20% 的仅有 21%。电子科技大学出版社共有 9 本书,最高的借阅次数为 133 次,最低的借阅次数为 7 次;武汉大学出版社的图书最高借阅次数为 65 次,最低为 24 次;中国电力出版社的图书最高借阅次数为 90 次,除去一本 2011 年出版的书的借阅次数为 0,其余三本的借阅次数均在 40 次以上。其他类似的例子不在此一一列举了,由这些数据可以得出:尽管有些出版社的图书没有进入到前 20%,但是其借阅次数也是非常可观的。

从版本因素看,新版本的借阅量大部分都明显低于老版本,原因之一可能是新版本的借阅次数累积量不够多。

4.1.3 复合因素分析

从作者和出版社的复合因素来看,前 20% 的图书中,将近三成图书的作者所编著的其他著作均在后 80%,75% 的作者会选择同一家出版社出版著作。而以作者、出版社和版本作为复合因素考虑时,在图书的题目、作者和出版社相同的情况下,用户借阅次数主要受版本不同的影响。比如郑初华编写的《汇编语言、微机原理及接口技术》(2003 年出版)的

借阅次数为 165,2006 年和 2010 年的版本的借阅次数分别为 102 和 30。造成借阅次数严重下降的原因可能由以下几点:

(1)出版周期短。郑初华的《汇编语言、微机原理及接口技术》2003 版的借阅次数是 165,虽然 2006 年版和 2010 年版的借阅次数没有进入前 20%,但是它们的借阅次数也分别达到 102 次和 30 次。尤其是 2006 版,只比 2003 版的借阅次数少 63 次,由此可以看出,此书并不是因为老化引起的借阅次数低,而是出版的时间周期短引起的。

(2)图书不同版本的馆藏数量不同。笔者任意挑选作者“钱晓捷”和“郑初华”作为关键词检索,发现两位作者的新版书的馆藏复本小于旧版。一旦新版书复本都被借出,用户必然会选择借阅旧版。这也是造成新版书籍借阅次数暂时偏低的原因之一。因此,对于旧版借阅次数很高的图书,可以适当加大新版图书的数量。

(3)借阅系统能否推荐新版本图书。钱晓捷的《汇编语言设计》由于图书馆已有从 2000 到 2006 年的三个版本,所以 2012 年版本的借阅次数非常少,只有 8 次。前三版的借阅次数可以反映出此书使用价值极高,第四版之所以现在借阅次数不高,一方面是因为新书的借阅次数积累不够,另一方面是新版图书没有一个完善的推荐系统为用户推送。

4.2 定性分析

通过征询专家意见,与上文中借阅量在后 80% 的图书中有较高借阅价值的书进行实际验证。

4.2.1 图书的选取

根据作者因素,抽取借阅量排在前 20% 的图书的作者,把他们所编著的借阅量在后 80% 的图书挑选出来,共计 17 本。将这些作者的所有图书(总共 42 本)放在一起,请专家评定。再用同样的步骤,根据出版社因素选定 118 本。经过去重等,最终需要专家鉴定的共计 131 本。

4.2.2 专家咨询

选择在汇编语言领域具有 5 年以上教学经验的老师来为图书的内容把关。经过两天共计 8 小时的审查和翻阅,131 本书全部鉴定完毕。经鉴定,图书被划分为 7 类:

(1)内容比较新颖,编著质量不错:这类书总共 19 本,占鉴定总数的 14.5%。

(2)内容比较新颖,编著质量一般:这类书共有

7 本。

(3) 内容比较新颖, 编著质量差: 仅有 1 本。

(4) 内容相对陈旧, 质量不错: 这类书共有 35 本, 占鉴定总数的 26.7%。

(5) 内容相对陈旧, 质量一般: 这类书共有 41 本, 占据的比例最大。

(6) 内容相对陈旧, 质量差: 共有 10 本。

(7) 工具书一类的边缘书籍: 这类书共 11 本, 把它们划分到汇编语言这个领域不太合适, 是边缘书籍。

4.2.3 作者因素结果分析

根据专家鉴定结果, 笔者发现: 很多作者编著的借阅次数高的图书的质量反而不如他们编著的借阅次数低的图书。比如张雪兰和谭毓安编著的《Windows 汇编语言程序教程》, 虽然这本书是在 2005 年出版的, 但是内容新颖, 编著质量超过了很多 2005 年以后出版的相关图书。尽管如此, 它从 2005 年到 2013 年的累计借阅次数只有 122 次, 并没有进入到前 20%。另一本图书《Windows 汇编语言程序设计实验指导》的借阅次数仅有 78 次。然而, 两位作者编著的《Intel 80X86/Pentium 汇编语言程序设计》(2001 版和 2004 版) 以及《汇编语言程序设计》, 其质量虽然不错, 但是内容已经陈旧, 可借阅次数高达 346、259 和 215 次, 居前 20%。葛洪伟的《Intel 汇编语言程序设计》和《Intel 汇编语言程序设计习题集》虽然内容上有些陈旧, 但是整体质量都很不错, 借阅次数只有 41 和 77 次; 而他编著的《汇编语言程序设计习题与解析》不仅内容陈旧, 而且质量相对一般, 较前两本差距很大, 但是借阅次数却高达 215 次。戴水贵编著的《Windows 汇编语言程序设计教程》及其配套的《Windows 汇编语言程序案例解析》的编著质量非常出众, 但是借阅次数仅有 39 和 23 次。其中《Windows 汇编语言程序设计教程》更是被专家鉴定为是一本非常值得借鉴的书。

基于作者的定性分析结果与定量分析结果是一致的, 说明图书馆确实有很多内容新颖且质量不错的图书没有被用户发现和充分利用。

4.2.4 出版社和版本因素结果分析

对比其他出版社的图书资源, 清华大学出版社以及电子工业出版社的图书内容质量更好。但是这两个出版社在后 80% 的图书数量占他们图书总数的比例约有 73%, 证明大部分的图书没有被用户充

分开发和利用。鉴定结果显示: 后 80% 中质量不错并出于这两个出版社的图书共有 12 本, 而前 20% 的图书为 24 本, 说明没被充分利用的图书的比例是很大的, 因此可以推荐的潜在空间也是巨大的。

新版图书的内容新颖、质量好、封面优良、纸张也很新, 各方面均优于老版图书, 但是借阅次数远远低于老版。造成这种反差的原因主要是因为新出版的图书借阅次数的积累量少, 图书馆推荐力度不够大, 大部分用户暂时还不知这本书已经重新出版。吴向军编著的《汇编语言程序设计》2009 版的借阅次数只有 48 次, 远远低于 2002 版的 208 次。但是鉴定的结果显示: 2009 版的图书质量非常好, 而且内容的新颖程度远远高于 2002 版。钱晓捷编著的《汇编语言程序设计》的 2000、2003 以及 2006 版的内容质量均相对一般, 但是借阅次数进入前 20%; 而此书的 2012 最新版本, 内容质量都非常优秀, 但是借阅次数仅有 8 次。

5 建议

根据之前的研究结果, 对比北京大学图书馆、北京航空航天大学图书馆、哈尔滨工业大学图书馆、哈尔滨师范大学图书馆等 10 所高校的数字图书馆平台, 针对 OPAC 信息查询系统, 可以就以下几个方面进行改进:

(1) 在检索结果页上显示图书的封面。虽然用户可以根据题目、作者以及出版社来确定图书, 但封面设计对于用户借阅行为的影响强度也是相对较强的, 封面图片可以使用户更加快捷地找到所需图书, 提高图书馆服务的质量。

(2) 增加二维码的应用。像哈尔滨工业大学图书馆和北京航空航天大学图书馆的用户可以通过扫描二维码将书目信息转移到手机移动端, 减少因查阅不同分类号下的多本图书而花费的记录图书物理方位的时间。

(3) 完善书目信息。书目信息中缺少必要的文字说明, 这使得没有目的的用户很难快速判断这本图书是否有用。因此很多用户采取翻看图书的方式来确定其是否和自己的需求匹配, 这样浪费了很多时间和精力。为了改善用户的借阅体验, 可以增加书目信息中的摘要部分, 方便用户判断。

(4) 检索结果的排序需要完善。除了常见的按照题名、责任者、索书号、出版社、出版日期以及入藏

日期的各种排序选择外,借阅次数强烈影响了用户的借阅行为,因此可以将图书按照借阅次数进行排序。第二是名家作品往往多次重版,如果检索结果的显示中能够体现作者同一作品的不同版本,那么有助于用户更快找到需要的图书。

(5)OPAC 管理系统的统计评价功能需体现。很多数字图书馆平台没有热门借阅排行系统,或者借阅榜并没有按照图书分类号来分类,不便于用户按类查看。这一点北京航空航天大学图书馆做得很好,热门借阅、热门收藏、热门评价等的分类功能都有所体现。

(6)以作者、出版社或者版本为主线来完善推荐系统。作者、出版社以及图书版本在影响用户借阅行为中充当了很重要的角色,确实有很多图书由于作者、出版社和版本的原因没有被用户充分利用。基于借阅次数,可对借阅量在前 20% 和后 80% 的图书根据作者、出版社、版本因素进行聚类,缩小推荐查找的范围。

(7)学科馆员参与流通工作。让拥有相关背景知识的馆员负责相应学科的图书区域,在适当的时候向用户提供帮助。只有了解图书借阅,才能有效

展开学科服务工作。

参考文献

- 1 李美珠. 高校图书馆馆藏图书借阅率调查分析——关于福建农林大学新书借阅调查. 农业图书情报学刊, 2006, 18(3): 158-161
- 2 邹梅. 图书借阅量下降原因及对策. 吉林工程技术学院学报, 2013, 29(7): 54-56
- 3 候蕾. 基于借阅数据分析的读者阅读倾向及服务对策. 图书馆刊, 2012, 12: 90-92
- 4 邱均平. 文献信息离散分布规律——布拉德福定律. 情报学理论与实践, 2000, 4(23): 315
- 5 Garfield, E. Citation Analysis as a Tool in Journal Evaluation. Science, 1972, (178): 471-479
- 6 邱均平, 李爱群, 周明华, 舒明全, 周春雷, 马瑞敏. 中国学术期刊评价的特色、做法与结果分析. 重庆大学学报(社会科学版), 2008, 14(4): 64-69
- 7 邱均平, 燕今伟, 周明华等. 中国学术期刊评价研究报告. 北京: 科学出版社, 2009: 21
- 8 张其苏, 蔡蓉华. 核心期刊数量的界定. 大学图书馆学报, 1999, (3): 1-6

作者单位: 哈尔滨师范大学, 哈尔滨, 150025

收稿日期: 2014 年 5 月 18 日

The Promotion Strategy of Lending Rate Based on the Statistical Data of OPAC

Meng Dequan Dong Ying Sha Yahong Meng Guirong

Abstract: The library resource usage ratio reflects the quality and lever of library service quality. Through user survey on library collection usage, it is found out that the low utilization of library collection is brought about by user lending behavior. Therefore, the paper finds out the factors affecting user borrowing behavior and analyzes these factors based on the principal component analysis method. Then it summarizes the main factors influencing the user lending behavior as author, publisher, publication date and etc.

Keywords: Usage Data Mining; Borrowing Behavior; Bradford's Law