

# 基于用户驱动的数字信息配置模式研究

刘 微

**摘 要** 图书馆在馆藏建设与读者需求之间始终存在“藏非所需”的矛盾。这一矛盾在造成图书馆资金使用效益低下的同时,也降低了检索效率,加大了读者时间成本,背离了图书馆以读者为中心的初衷。本文仅就当下数字信息所面临的问题入手,借助案例分析的方法,试图构建一种基于用户驱动的数字信息配置模式,旨在降低图书馆数字资源购置成本,提高用户检索数字信息效率。

**关键词** 图书馆;馆藏建设;数字信息;用户驱动;信息配置

**分类号** G253.1

**Abstract** In the library, this is “collection not required” contradiction between library construction and readers demands. This contradiction causes the low benefit of using the library, reduces the efficiency of retrieval, increases the time cost of readers, all of them deviate from the original intention of library reader centered. Starting from the problems in digital information, and using case analysis, this paper tries to construct an information procurement mode based on users' needs, in order to reduce acquisition cost of library, and improve the retrieval efficiency of users.

**Keywords** Library; Collection construction; Digital information; User driven; Information configuration

**Class Number** G253.1

随着互联网时代的来临,数字信息在图书馆已经占据了重要位置,无论馆藏资源数量、经费投入还是读者利用已经有超越传统印刷型图书资料的趋势。然而,各类数据库的信息内容、阅读方式、下载格式、结算价格等,都是由数据供应商单方决定,图书馆、信息中心等大量信息消费者处于被动地位,完全没有话语权。图书馆花费巨额经费购买的数字信息中充斥着大量的无用信息甚至垃圾信息,既浪费了经费和存贮空间,增加了信息检索的难度,降低了查准率,又加大了用户的时间成本。本文仅就当下数字信息所面临的问题入手,借助案例分析的方法,试图构建一种基于用户驱动的数字信息配置模式,旨在降低图书馆数字资源购置成本,提高用户检索数字信息效率。

## 1 图书馆数字信息配置问题

### 1.1 从计量分析看数字资源配置的现状

笔者以全球最大的中文数据库清华同方的中国知网(CNKI)为例,运用文献计量的方法分析和探讨图书馆数字信息藏与用的结构比例。中国知网用户数量4000多万,中心网站及镜像站点年文献下载量突破30亿篇次,而且还会逐年增加。然而,约5500万篇庞大的文献总量其参考价值究

竟多大则很难考证。

众所周知,布拉德福文献分散定律、洛特卡倒数平方定律和齐夫词频定律是文献计量学研究领域最著名的三大定律。笔者尝试运用这三大定律分析探讨CNKI数据库5500万篇文献中核心文献的数量分布。(1)按照布拉德福定律,学术期刊分为核心区、相关区和非相关区。处于核心区的学术期刊数量是期刊总量的 $\frac{1}{1+n+n^2}$ ,其中 $n$ 为布拉德福常数,约等于5。<sup>[1]</sup>据此公式计算,CNKI收录的5500万篇文献中核心区期刊文献约为总量的 $\frac{1}{31}$ ,即177万篇,相关区期刊文献约为总量的 $\frac{5}{31}$ ,即887万篇。诚然,学术相关性分为近相关(密切相关)和远相关(疏离相关),二者文献各占约 $\frac{1}{2}$ 。如果取全部核心区和相关区中近相关的50%文献,则共有620万篇核心和密切相关的文献,占文献总量的11.28%。(2)按照洛特卡定律,写1篇论文的作者占全部作者的60%,写 $N$ 篇论文的作者数量是写1篇论文作者数量的 $\frac{1}{n^2}$ 。如果将写5篇及5篇以上论文的作者视为核心作者,其数量是写1篇论文的作者数量的 $\frac{1}{25}$ <sup>[2]</sup>,即5500万篇文献中写1篇论文的作者为3300万,那么写5篇及以上论文的核心作者是132万人,其撰写的文献数量 $\geq 132 \times 5 = 660$ 万篇以上,占文献总数的12%。(3)按照齐夫词频定律,包含最常用的出现

频率很高的关键词的文献与词频排名成反比。<sup>[3]</sup> 2012年出版的《现代汉语词典(第六版)》中收录汉语词汇约7万个,国家汉办出版的《汉语水平词汇与汉字等级大纲》中甲、乙、丙、丁四级词汇数量分别是1033个、2018个、2202个、3569个,共8822个。很显然,这8822个汉语词汇是使用频率较高的常用词汇,占全部汉语词汇量的12.6%。那么,据此计算,5500万篇文献中包含常用高频词汇的文献约有693万篇。

按照三大定律的推演均已证明,图书馆在经费并不十分充足的条件下,其数字信息配置充斥着大量重复无用的信息资源,导致当下图书馆数字资源配置已然形成的现状。

## 1.2 数字资源配置给图书馆带来的困扰

无论从信息供应方——图书馆,还是从信息消费方——读者来看,“大而全”的巨无霸数字信息建设模式都是不科学不明智的选择。

首先,增加了资源配置成本,资金效益较低。通过上面的计算我们发现,文献计量学的三大核心定律分别从不同的视角揭示文献的特征,布拉德福定律是从学科的角度研究核心期刊分布规律,洛特卡定律是从作者的角度推算核心作者群分布,齐夫定律是从关键词的角度揭示高频关键词出现的文献分布规律。<sup>[4]</sup> 尽管三个定律出发点不同,计算方法不同,但计算结果却比较接近,最大差距只有1.32%。为了提高计算的相对科学性,克服某一种计算方法的缺陷和偏差,综合平衡学科、作者、关键词对文献的影响,取三个定律计算结果的平均数为标准值,那么CNKI有价值的比较重要的核心文献约658万篇,约占全部文献数量的12%。这就意味着图书馆要用5500万篇文献的经费购买了658万篇比较有参考价值的文献。如果是本地镜像存储,还需要增加存储设备,5500万篇全文文献的存储与运行设备至少在百万元以上,而且设备的维护保养需要不断的资金投入。上述计量分析数据表明,在图书馆馆藏与利用之间存在着巨大的数据落差,而且这种落差不仅体现在同一个数据库中,不同的数据库之间这种落差更为明显。如维普、万方、同方三个数据库供应商都提供学术期刊全文数据,其重复率超过60%。同一个数据库中的有效资源比例本来就很低,不同数据库之间的重复率又如此之高,二者效果叠加,整体资源配置的效益完全可以用惨不忍睹来形容。

其次,降低了信息检索效率,加剧了时间成本

的提高。在信息检索中,查全率与查准率历来就是一对矛盾的存在。庞大的数字信息资源配置,在相对提高查全率的同时,必然大大降低文献的查准率。以上述CNKI数据为例,如果精确检索,查找某一具体的文献,在全部资源库中,其查准率为1/5500,这无异于大海捞针。如果模糊检索,以篇名、作者、机构、关键词等检索,一次检索后,检索结果数量比较大,仍然需要进行二次甚至三次、四次检索,大大增加了用户的检索难度,同时也浪费了大量时间。很多时候,信息检索者都不是精确检索,也就意味着绝大部分用户其检索效果的满意度、认可度非常低。

## 2 用户驱动采购模式的选择

### 2.1 用户驱动采购模式

用户驱动采购模式是一种数字信息许可使用的订购模式,意在让图书馆采购部门从花费了可观经费却购买了大量鲜有人问津的图书的窘境中走出来。一份关于学术图书馆未来的报告,将这种基于需求的服务模式视为那些面临压力的图书馆一个不可逆转的发展趋势,图书馆通过这种方式来证明其采购经费的支出是与其本身的价值相匹配的。一家设在华盛顿的咨询公司发布的一项报告表明,用户驱动采购方式在根本上改进了图书馆低效的书目传递模式,在这个传统的图书传递模式中,由图书馆员来主观臆断读者所需的阅读书目,而不是允许用户来引导馆藏配备和采购流程。该报告预言,图书馆向其用户提供图书方式的转变反映了在数字媒体领域的一个更明朗的发展趋势,图书馆将会逐步抛弃在流通率较低的开架借阅区配备大量实体图书的采购模式,与电子书供应商达成许可使用合作协议,让图书馆纸本藏书数量和所呈现出的需求量具备较为合理的匹配度,以此来减少副本数。<sup>[5]</sup>

### 2.2 用户驱动采购的必然性

2010年美国大学和图书馆协会发布了未来大学图书馆的十大发展趋势,排名第一的是“馆藏的增长取决于用户”,这个调查结果印证了用户驱动、以用为主已经是图书馆发展的必然。我国现行的财管理政策是预算管理和项目管理,公共图书馆、高校图书馆、科研院所图书馆等采购经费基本都是项目管理、国库支付,需要进行项目绩效

2014年10月 October, 2014

评价,即对资金使用的长期绩效目标和年度绩效目标进行考核。在政府划拨财政预算项目申报文本中,对项目实施绩效考核的一级指标包括产出指标、效益指标、服务对象等,二级指标包括数量指标、质量指标、时效指标、成本指标、社会效益、经济效益、环境效益、可持续影响等。文本还要求对项目实施绩效的上年实际值、预期实现值进行考核。由此可见,国家财经政策、项目管理制度都要求我们降低投入成本,提高资金利用效益,图书馆已经没有选择的余地,其信息配置必须转变理念,藏用并举,以用为主。

### 2.3 用户驱动采购的可行性

现代网络技术的发展及其在图书馆信息服务领域的广泛应用,为图书馆数字资源的用户驱动采购提供了现实可能性和技术操作性,在数字信息配置中,已经有大量的用户试图通过驱动采购模式进行学术研究和实践尝试。用户驱动采购将图书馆信息配置权授予全体读者而非某一个或几个图书馆员、学术专家,通过读者对信息利用的时间和频率来决定付费方式和数量:利用率很低的文献完全免费试读;利用率适中的文献按照实际使用情况计时收费;利用率较高的文献采用买断使用权的方式长期提供阅读和下载服务。这样一来,一方面促使数字信息供应商优化产品结构,加大优质信息宣传,努力提供适销对路的信息产品;另一方面提高了信息的针对性和适用性,降低了文献采购成本。这无疑是信息生产和消费双方风险分担、互利共存的信息配置模式。

## 3 数字信息配置的用户驱动采购模式

### 3.1 用户驱动采购模式构建

用户驱动的数字信息采购模式首先要求数字信息生产和经销商将其所有数字产品的目录和试读内容公开呈现在用户页面,包括社会大众用户和特定集团消费用户,供其浏览、阅读。产消双方签订协议,根据数据库文本的大小,设置相应的计费触发点,达到触发点后开始计费,否则视为体验性阅读,完全免费。比如,可以根据数据库文本的容量大小,分别为期刊论文、硕士学位论文、博士学位论文、电子图书等设置2分钟、5分钟、10分钟、20分钟的计费触发时间,该时间点以内的浏览是读者

的考察、犹豫和选择期,可以完全免费;该时间点以上的阅读视为有效阅读行为,开始计费。作为一个产消双方双向选择的过程,这样可以促使数字产品供应商组织专业人员进行产品选题,加强产品的优化,注重文献的内容选择和形式表达,通过更认真、更专业的筛选,优胜劣汰,努力为用户提供数字精品。科学地讲,这种计费触发点应该是一系列的,触发时间越长费用越高,最后一个触发点应该是全文下载。如果用户最终决定下载全文数据,费用应该是最高的。图书馆最后根据触发时间和触发人数的综合计量来支付费用。这种计费方式充分体现了互利互惠、公平竞争、各取所需的市场原则。当然,可以根据用户单位的服务对象、学科设置、专业建设、科研方向等特点,对不同板块的信息资源设置不同的权重系数,如普及性、公共性的数据信息,其用户面广读者众多,但往往是消遣性阅读,下载量不是很大,这种信息产品的权重系数相对较低,适用于薄利多销的销售原则,以量取胜。专业性强、偏重学术性、用户面较小的专业性数据产品权重系数可以设置大一些,通过较大的权重设置保证专业数据供应商的基本利润,引导其产品供应的方向和比例。

### 3.2 用户驱动采购模式案例分析

2009,美国大河谷州立大学与电子书图书馆(Ebook Library,EBL)签订了合作协议。该大学图书馆购买的电子图书定价由一系列触发点来设定:如果一个学生在某本书上停留的时间少于5分钟,那么图书馆则无需支付任何费用。如果超过5分钟,触发点会自动启动一日租赁付费模式,此时EBL则会向图书馆收取该书定价的10%—20%作为电子书使用费。超过5天的电子书租赁行为则会自动触发购买整本电子书,到那时,学生们可以随时访问这些内容,不会给图书馆带来额外的费用。只有当电子书每年的综合使用率超过320小时,则自动触发图书馆购买该书的第二副本。如果购买EBL提供的目录中所有5万种电子书,图书馆需支付300万美元采购费用。购买2009年图书馆用户所使用过的6239种电子书需要花费55万美元。但实际上仅有343种电子书是满足触发购买行为条件的。基于此,图书馆仅需支付69,000美元。2010年,可供浏览的EBL书目数量翻了三番,但图书馆采购(下转第101页)

## 2.3 要开拓创新,着力提高农家书屋后续管理的科学化水平

推进农家书屋后续管理工作,为广大群众提供更多的实惠,既是责任,也是使命。全面推进这项工作,既要脚踏实地更要开拓创新,只有开拓才有出路,唯有创新才有活力。一是机制制度创新。要对现有的机制制度重新进行审视,推优去劣,激浊扬清。要建立好分级负责、分类指导、分层管理的良性运行机制,适时健全驻村干部联系制度、乡镇村骨干定期蹲点制度等,并在此基础上健全考核评价和奖惩激励机制,将考评结果纳入乡镇村或乡镇村干部绩效考核范畴,逐步提高层级权量,以完善的制度机制来推动后续管理工作的管理运行。二是活动载体创新。要紧密结合社情民意,开展丰富多彩的读书活动,引导更多的群众参与其中,并通过深切的感同身受来激发灵感,提出更多有创意重实效的活动载体。当前,要认真组织两部分人走进农家书屋,一部分是中小學生。这些学生处于奇思妙想的喷发期,引导他们参与管理、自主管理,说不定一不小心就会冒出创新的火花,让整个书屋的读书活动焕发生机活力。另一部分是回乡打工人员。这部分人群见多识广,经验丰富。要充分利用这一资源优势推进后续管理工作,让“肥水不流外人田”。做好以上“两个

引进”工作后,就要积极打造“活动品牌”。品牌的树立不能好高骛远,不能贪大求全,要立足本土文化沃土,大胆创新,真正实现“根植沃土发芳华”。像享有“湖南书画第一村”美誉的宁远县湾井镇朵山村就是一个很好的例子。该村农家书屋根据全村三分之一村民习书作画优长,积极争取资金办起了书画馆,其书画品牌更加耀眼。三是理念观念创新。观念一变天地宽。为此,要进一步解放思想,转变观念,把农家书屋后续管理工作办出亮点特色。要彻底摒弃等人上門的思想与行为,主动走出去,沉下去,广泛收集意见建议,全面掌握第一手材料,为工作决策提供有力参考。同时,加强战略性和宏观性研究,增强农家书屋后续管理工作的前瞻性和预见力。要加强资金投放的科学化管理,把好钢用在刀刃上。要切实加强对整个工作的信息化管理,提高现代化水平。农家书屋管理员也要主动上门服务,全面掌握群众所需所求,以便及时调整状态与思路,为群众提供优质服务。要强化农家书屋的动态化管理,因人而异,因地制宜,真正把农家书屋办成广大农村群众的“精神家园”。

作者简介:

李晓文(1976—),男,馆员,湖南宁远县图书馆,湖南,宁远,425600。

(上接第72页)

电子书的数量却大致维持在同一水平。<sup>[5]</sup>基于使用率来决定是否购买,大河谷州立大学每年约需购进350种电子书。而如果按传统采购方式,满足同样的需求则需配备约10,000种实体书,且仅有一半数量的图书会被学生们外借使用。

大河谷州立大学图书馆馆藏和学术信息交流部的负责人道格·卫表示,从这一次试验中收集来的数据是颇具启发性的。他认为,实体图书的采购往往是基于一些高远的目标,如传承人类文明、收藏民族文化、服务社会等。但是,图书馆并不具备这样的能力。如果在10,000种图书上花费600万美元,那么其中就有300万美元的开销完全是浪费的。<sup>[6]</sup>而且,这也并非是一次性的成本,纸本图书还会带来很高的利用和贮存成本。

综上,现代网络技术的发展及其在图书馆服务中的广泛应用,成为图书馆数字信息用户驱动采购的契机。数字信息生产和消费双方通过协商,根据数据库文献单元的内容和数量,设置不同的用户阅读时间计费触发点,并根据用户单位的

学术相关性综合设定数据库的权重系数,最终实现“多用多付、少用少付、不用不付”的数字信息采购机制,提高投入产出比,进而提高图书馆资金的利用效益和资源利用效益。因此基于用户驱动的数字资源配置模式无疑为图书馆数字资源建设起到了引领示范作用,预期此举在业界将有很广泛的发展空间。

参考文献:

- [1] 苏文. 基于B/S结构的体育统计文献检索系统设计与实现[D]. 武汉:武汉体育学院,2010.
- [2] 朱娴淑. 结构化对等网中基于复制的访问热点消除方法研究[D]. 长沙:湖南大学,2007.
- [3] 尹丽春. 科学学引文网络的结构研究[D]. 大连:大连理工大学,2006.
- [4] 邱红. 关于无标度网络的理论探讨及其在信息分析中的运用研究[D]. 上海:上海大学,2009.
- [5] 吴建中. 转型与超越:无所不在的图书馆[M]. 上海:上海大学出版社,2012:119-122.
- [6] PDA:用户驱动的图书馆采购新模式[EB/OL]. [2014-06-10]. [http://blog.sina.com.cn/s/blog\\_5deec7c30100wryf.html/](http://blog.sina.com.cn/s/blog_5deec7c30100wryf.html/).

作者简介:

刘微(1974—),女,馆员,中南大学图书馆,湖南,长沙,410083。

2014年10月 October, 2014