

论图书馆数字化信息资源建设中的信息精选

陈 雅, 陆宝益, 郑建明

(南京大学 信息管理系, 江苏 南京 210093)

摘 要: 文章简要阐述了信息精选的含义和理论基础, 分析了信息精选的意义, 并在此基础上进一步讨论了图书馆数字化信息资源建设中精选信息的必要性, 进而提出了一套信息精选测评指标体系, 初步设计了信息精选流程, 并对可能影响信息精选的不利因素进行了分析。

关键词: 信息精选; 数字图书馆; 信息资源建设; 数字化信息资源

中图分类号: G250.76 **文献标识码:** A **文章编号:** 1007-7634 (2006) 11-1662-07

On the Information Selecting in the Library Digital Information Resources Construction

CHEN Ya, LU Bao - yi, ZHENG Jian - ming

(Department of Information Management, Nanjing University, Nanjing 210093)

Abstract: This article briefly describes the meanings and theoretical foundation of the Information Selecting, analyzes the significance of the Information Selecting, and based on it, discusses the necessity of the Information Selecting in library digital information resources construction. Moreover, the author puts forward a suit of information selecting evaluating indicator system, designs the procedure of the Information Selecting, and analyzes the probably negative influence factors on the Information Selecting.

Key words: information selecting; digital library; information resources construction; digital information resources

信息的无限量增长, 给人们信息的检索、获取、利用, 乃至文献信息收藏与服务机构的信息资源建设等都带来了一定的困难, 对信息的评价、过滤、筛选等成为信息管理学界近些年研究的重要问题之一, 人们提出了多种多样的理论或方法。近来, 有的学者进而提出了“信息精选理论”的概念。鉴于此, 我们结合图书馆多年的工作实践, 对当前正在进行中的数字图书馆建设以及图书馆数字化信息资源建设中的信息选择问题有了进一步的认识, 产生了一些新的想法, 在此付诸文字, 以便与同行商榷。

1 信息精选

1.1 信息精选的含义

所谓“信息精选”, 就是根据特定的目的或要求, 按照一定的标准与程序, 对一定范围内的有关信息进行精心的评价与筛选, 以确定核心信息区域的过程。

长期以来, 信息精选一直存在于我们的实际工作之中。对于图书馆来说, 日常的图书选购、期刊

收稿日期: 2005-05-29

基金项目: 江苏省高等教育教改立项研究课题 ([2005] 27号-68)

作者简介: 陈 雅 (1965-), 女, 上海人, 副教授, 从事图书馆学、社会信息化、出版管理、图书馆法规等研究;
陆宝益 (1969-), 男, 淮阴人, 博士生, 从事社会信息化、目录学、信息咨询等研究; 郑建明 (1960-), 男, 福建人, 教授, 从事目录学、信息咨询等研究。

选订、数据库购买前的大量试用与比较,以至于新书报道与推荐等,实质上都是一个对有关的文献信息进行精选的过程,以期把最好的文献信息资源纳入馆藏,把最有针对性、最好的文献信息传递给读者。进一步地说,信息精选存在于信息生命周期的每一环节,从文献信息的生产、加工、传递,到收集、保存、利用、销毁等过程,都在实施着实质性的信息精选。如书刊出版时对录用稿件的评审、营销商对经销范围确定及具体文献信息产品的选择、传统文献检索工具及各种学术数据库收录来源文献的选择、文献收藏单位对于过时(期)文献或冗余信息的剔除与销毁时的甄别,等等,无不是一个信息筛选的过程。在当今的信息时代,图书馆数字化信息资源建设,同样应该是一个信息精选的过程。可以说,“信息精选理论”虽然没有被明确提出和定义,但是一直为图书情报界所实践着。

1.2 信息精选的理论基础

自文献存在以来,人类就在对其进行有效的管理。随着文献生产技术的进步,文献产量的不断积累与增加,人们一直探索对文献进行有序化管理和利用的方法。到19世纪后期,随着机械化印刷技术的日臻成熟,文献生产进入了一个飞速发展的时期,各种形式的文献大量涌现,促进了有关学科领域的专家对文献生产、存在、管理、利用等内在规律的研究。20世纪上半叶,利用文献计量学的方法,人们先后发现或总结了多个文献管理的定律或原则。如1926年洛特卡(A.J.Lotka)提出了科学文献生产中作者与论文之间关系的“洛特卡定律”;1934年布拉德福(S.C.Bradford)提出了反映文献分布规律的“文献分散定律”;1948年,齐夫(G.K.Zipf)通过对自然语言中的词频分析,得出了“最小努力原则”;1960年代,加菲尔德(E.Garfield)统计出引文分布定律^[1];以及反映文献增长的普赖斯定律、文献老化逐渐过时律,等等。所有这些定律或原则,成为整个20世纪中后期人类管理、选择与利用文献(包括分析、研究与选择、利用情报)的重要的基础理论。

人类进入21世纪以后,随着现代信息技术的发展及其在文献信息生产与管理领域中的广泛应用,以及人们对文献信息需求的不断专业化、个性化、多样化,国内外信息管理领域、特别是图书情报界的专家,进行了大量的研究,提出了多种视角、各有侧重的,现代信息社会、网络环境下的信

息资源管理理论,推动了人类对于信息资源的管理从“传统文献的管理”转向“对知识单元的管理”。这些理论体系中融入了更多的技术因素,但是对于有关信息资源的分析与评价、选择,仍然离不开传统的文献情报规律的指导。如孟广均教授等人合著的《信息资源管理导论》中在论及“信息源分析”时,仍然运用了传统的文献计量学三大定律,即布拉德福定律、洛特卡定律、齐夫定律^[2]。可以说,现在谈论的信息精选、或者说以后可能建立的信息精选理论,也将是一个建立在传统文献情报规律与管理理论基础之上的新型理论^[1],将更加适用于信息时代对于各种载体信息资源的生产、管理与利用。它与传统文献情报规律与管理理论的联系表现在以下几个方面。

(1)“信息精选”建立在一系列传统文献情报规律与管理理论的综合基础之上,但不等同于任何一种传统的文献情报规律。信息精选过程中对于待选信息的评价,既要借鉴传统的文献情报规律与管理理论来分析信息资源的学科性质、存在状况、读者对象等信息资源自身的内外特征,又要认真分析新条件下用户的专业化、个性化、多样化的信息需求特点与变化规律,以保证信息资源的提供者与使用者都能达到“最小努力原则”的效果。

同时,“信息精选”将在处理信息的对象、适用范围、信息处理方法等多方面突破传统文献情报规律及管理理论的界限,它们的关系如图1所示。

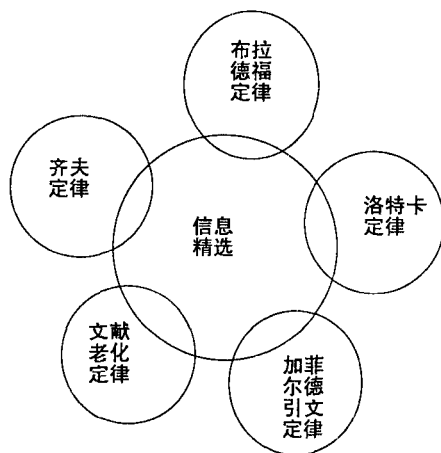


图1 信息精选理论基础构造图

(2)信息精选比传统文献情报规律具有更广泛的分析研究对象。传统的文献情报规律理论主要以文献的数量、作者、单词、时间等为分析单元,研究的对象是以文献信息的外部特征为主;而信息精

选则要分析信息资源所涉及的所有因素,不仅要研究其外部特征,更要深入地研究信息资源的内部特征,即信息资源的内容属性及其质量水平。

(3) 信息精选比传统文献情报规律具有更宽泛的应用范围。后者主要适用于对文献情报的选择、保存与利用等部分的文献情报管理环节;而前者则适用于信息生命周期的全过程,从信息的生产/创建、流通/传递、采集、组织/整合、存储,到检索/利用、归档/销毁等每一个信息运动与管理环节,都应该对有关的信息进行必要的评估与选择。

1.3 信息精选的意义

当今时代,信息技术高度发达,世界范围的信息产量与信息流量呈指数级爆炸性增长,有人估算,现在每最近四年的信息产量就等于之前历年信息产量的总和,任何一个专家学者都不可能用一生的时间来阅读完其所在学科领域的一年之中产生的有关的信息与文献资料。这样环境下,实施信息精选具有非常重要的意义。

(1) 可以在一定程度上舒缓信息恐慌。美国著名社会学家约翰·纳斯比特曾说过:“我们正被信息淹没却乏于知识”。

在信息海量增长与累积的情况下,人们面对汹涌而来的信息,一方面怀有对信息的强烈渴求,另一方面却不知如何从大量的信息中挑选信息,自己真正所需的有用信息几乎都为大量不相关的信息所覆盖和淹没,想到要从信息的“海中捞针”,非常害怕和不安。而通过信息精选,可以相对地控制信息用户所能接触到的信息数量,特别是对精选过的信息进行有序化的组织以后,可以让信息用户需要选择的信息范围大大缩小,从而把对所需信息的选择与利用过程变得更加轻松、快捷和方便。

(2) 利于塑造信息精品。在信息的生产与传递更加自由的现代信息技术条件下,大量的伪劣信息、重复信息鱼龙混杂来到世人的面前,致使信息的整体质量下降。美国曾有人形象地形容:“Internet上的信息如果有一英里宽,则只有一英寸深”。要改变这样的状况,只有提倡在信息的生产、传递、保存等信息生命周期的各环节上对所处理的信息进行精选,逐步淘汰、销毁劣质信息、重复信息,不断净化信息的内在纯度,力争信息生产、管理与利用的每一过程中的信息都能以优质、精品的信息产品形式存在。如出版社、杂志社通过信息精选,能够更多地载发优秀论文、出版精品图书;文

献收藏机构通过精选,努力把优质的文献及其他各种载体的信息资源纳入馆藏,以保证为用户提供优良的精品信息资料,进而创建一个高可信度的“精品信息通道”——创作/生产(高水平作者)→出版/发布(权威出版社、杂志社、网站)→流通/营销(有信誉的经销商)→存储/馆藏(有能力的机构)→利用/再生→归档/永久保存。

(3) 促进信息流通的良性循环。信息环境是由信息本身、信息技术与设备、信息人员(包括信息生产者、传播者、用户等)、信息政策与信息伦理等多种因素构成,它们相互影响、共同制约着信息环境及其信息流通质量的优劣状况。而其中信息本身是其他各因素研究和处理的共同对象和存在的意义所在,信息本身质量的优劣是信息环境好坏的重要标志。通过信息精选,可以促进信息环境的不断优化与信息流通的良性循环。这不仅表现在信息的精选可以尽可能地保证优良的信息产品在信息生命周期各环节上流动,减少信息传递过程中的信息失真,控制信息流量、提高信息传输速度,还反映在它可以刺激人们的信息质量意识、信息法律意识、信息伦理意识等信息观念的培育,以及信息评估机制、标准的制订,等等。

(4) 还可以引导学术规范。在竞争日趋激烈的当前社会,由于种种原因,我国的学术研究失范现象呈不断加重的趋势,学术规范问题已经成为我国政府有关部门与学术界关注的热点问题。促进学术规范的手段和途径可以是政策引导、法律制裁,也可以是舆论监督、道德制衡,等等,而信息精选同样不失为一种有效的刺激性因素。在信息的生产、传输、存储、利用等各环节,通过精选工作,把有违学术道德以及不符合学术规范的信息(产品)排除出信息流通领域,使其永无见天之日,让符合学术规范的信息独占天下。比如图书出版、期刊编辑机构在审稿时,对于那些存在结构不完整、关键词标引错误、参考文献著录不合规范、作者信息欠缺、内容抄袭等问题的作品坚决拒用;检索工具编制、学术数据库建设选择来源书刊时,对于不合学术规范的文献或作品不予收录;信息与文献收藏机构在选购馆藏时,对于存在学术规范问题的信息资源不予收藏;研究人员在选择参考文献时,对于存在引文来源不明、著录不准的作品,以及存在学术道德问题的人的所有作品坚决不予引用。如此,将在很大程度上刺激学术研究人员、出版发行机构等信息生产者自觉遵守有关的学术研究规则与道德准

则,引导学术研究的规范不断走向完美。

2 图书馆数字化信息资源建设中信息精选的原因

图书馆数字化信息资源建设的内容,主要包括传统馆藏文献信息的数字化转换(如书目回溯建库、特色数据库、期刊篇目库、书刊全文库等)、商业性学术数据库购买(如中外文全文库、文摘题录库、引文库等)、网络信息资源开发与利用(如网络导航、网上资源下载与保存等)。在网络环境下,图书馆可以进行数字化处理的信息量非常庞大,但是任何一个图书馆的数字化信息资源建设都不可能涵盖所有的信息资源,而必须对其进行适度的控制,这不仅仅是基于信息精选的重要意义的考虑,更是由于数字图书馆建设过程中各种内在因素协调发展的需要。读者的信息需求特点、建设成本、信息产权、信息质量、建设时效等是图书馆数字化信息资源建设过程中需要对相关信息进行精选的重要影响因素。

(1) 读者需求问题。现代网络条件下读者对信息的获取变得越来越容易,图书馆对于馆藏建设的指导思想也基本上从“重藏轻用”转变为“藏用并举、以用为主”,满足读者的文献信息需求是图书馆信息资源建设的主要目的,数字化信息资源建设也是为了更快、更好地实现这一目的。所以,对于一个图书馆来说,数字化信息资源建设在不可能全面铺开的情况下,首先应在结合学校的学科建设发展方向,及地方社会、经济、教育与文化建设的发展战略的情况下,认真分析不同学科专业背景、不同年龄层次、不同职称档次、不同研究领域的读者,对信息的越来越多样化、专业化、个性化的需求特点与变化规律,以及对信息的理解与接受能力,对数字化建设的信息资源进行精选,以保证建成的数字化信息资源首先可以满足重点学科、重点读者、重点领域的信息需要,从而发挥最大的学术价值、社会效益与经济效益。其实,这也是图书馆数字化信息资源建设是否具有针对性的问题。

(2) 建设绩效问题。数字化信息资源建设与传统文献信息建设的最大不同在于其需要投入更多的建设成本,如传统文献的数字化转换需要一般性的计算机、描述仪等录入设备,研发或购买数据库软件系统,还需要添置较大容量的存储设备;购买学术数据库除了价格比较昂贵外,如果要作本地镜像

或数据保存,也需要大容量的存储设备;下载保存网上有用的信息需要占用很大的存储空间。此外,为了实现数字化信息资源的有效利用,还要完善校园或图书馆内部的网络设施以及可供读者检索、利用信息资源的机器设备;同时要对读者进行必要的培训。所有这些都需要投入大量的人力、资金与时间,而目前我国绝大多数图书馆都存在着经费紧张和专业人才不足的问题。因此,图书馆在进行数字化信息资源建设时必须对有关的信息进行评估,预测其将来的实用价值,看其能发挥多大的社会效益或经济效益,是否值得列入本馆的数字化信息资源建设范围,从而精选出最有价值前途、性价比最高、最好的信息资源。这也是图书馆数字化信息资源建设是否具有经济性的问题。

(3) 信息产权问题,包括数字化信息资源的长久使用权问题。在知识经济时代,人们的知识产权保护意识不断增强。作为公益性为主体的图书馆、特别是高校图书馆,对信息的收集、保存、提供、利用等,主要是为了满足法定范围内读者的教学、科研、学习、消遣等需要,而不以营利为目的。但是即便如此,也不能完全排除意外地涉及法律纠纷,或造成对他人合法权益的侵害,或自身的合法权益受到他人的损害,此类事件在媒体上时有报道。故此,图书馆在建设数字化信息资源时必须对相关信息的产权状况进行认真研究。传统文献中哪些是产权保护之外可随意进行不同媒体形式转换的,哪些是禁止的;网上信息中哪些是公共、免费可以任意下载、保存、传递与利用的,哪些是只能个人在线阅读的;商业性学术数据库中哪些是完全解决知识产权问题的,哪些是存在产权纠纷的,哪些是既授予使用权又提供原数据的,哪些是只提供使用权的,哪些是解决已购数据长久使用权问题的,等等,都是数字化信息资源建设时要进行信息精选的重要影响因素。这是一个合法性的问题。

(4) 信息资源质量问题。数字化信息资源建设需要很大投入,保证建成的数字化信息资源具有较高的质量水平是非常必要的,也是应该的。但是图书馆数字化建设过程中所接触的信息资源的质量是参差不齐的,比如说馆藏图书有权威出版社出版、学术水平很高的;有一般出版社出版、学术水平一般甚至是低劣的。期刊有权威期刊、核心期刊、普通期刊,学术性期刊、非学术性期刊之分。数据库有全文数据库、文摘数据库、引文数据库,综合性数据库、专科学性数据库之别,而且不同数据库的收

录容量、功能配置、检索途径、输出效果等也各不相同。故此,对来源信息进行精选是图书馆数字化信息资源建设过程中必不可少的步骤,这是保证数字图书馆建设整体优越性的问题。

(5) 建设时限问题。包括两方面的含义:一是信息资源本身的出版发行时间。网络条件下信息更新速度很快,人们总是希望获得最新的信息资料,那么作为以先进的现代信息技术手段为依托的数字化信息资源建设必然也要将最新的信息资源作为处理和向读者提供服务的对象,无论是传统馆藏还是网上信息。二是数字化信息建设是一项耗财、耗力、耗时的系统工程,然而读者对其却是寄予真诚希望的,他们总希望能够早日从中受益,特别是读者的参与将更加强化他们的这种期望,图书馆任何一项数字化信息资源建设项目都不能无视读者的期望及其容忍度。在此考量下,对有关的信息资源进行精选,以尽可能新的和相对较小数量的信息资源进入分期或分阶段的数字化建设项目,来加快馆藏信息资源数字化建设的整体进程,同样是必要和可行的。这样可以很好地解决图书馆数字化信息资源建设的时效性问题。

3 图书馆数字化信息资源建设中信息精选的测评指标体系

3.1 信息精选测评指标体系

信息精选的过程,也就是对有关信息资源进行评价、淘汰和挑选、保留的过程。制定一套系统、完整的信息评估指标体系对于信息精选工作来说具有非常重要的意义。信息精选对于信息的测评不同于单纯的信息评价,单纯的信息评价只考虑信息本身的质量因素,而信息精选不仅要考虑信息自身的质量,还要考虑其读者对象及将来的效用情况等因素。当然,单纯信息资源评价中的有关指标与方法如田建平的图书评价八大值素^[3]、邱均平等的期刊评价指标与方法^[4]、刘颖的电子文献选购原则与评价指标^[5]、罗春荣的因特网信息资源评价指标与方法^[6]、应峻的全文数据库评价方法^[7]等,对于信息精选仍然有一定参考价值。结合以前的相关研究成果,笔者提出如表1所示的信息精选综合性测评指标体系。

在这个测评指标体系中,共有25项测评指标。定性指标16项,定量指标9项;适用于传统馆藏

文献14项,商业性学术数据库24项,网络信息资源16项。

表1 图书馆数字化信息资源建设中信息精选的测评指标

| 序号 | 指标名称 | 传统馆藏 | | 学术数据库 | | 网络信息资源 | |
|----|-------------|-----------|-----|-----------|-----|-----------|-----|
| | | 选用 (√) | 权重 | 选用 (√) | 权重 | 选用 (√) | 权重 |
| 1 | 学科针对性 | √ | | √ | | √ | |
| 2 | 学术性 | √ | | √ | | √ | |
| 3 | 出版商 | √ | | √ | | √ | |
| 4 | 发行(代理)商 | √ | | √ | | | |
| 5 | 第一责任者 | √ | | √ | | √ | |
| 6 | 功能配置 | | | √ | | | |
| 7 | 数据提供及访问方式 | | | √ | | | |
| 8 | 技术要求 | | | √ | | √ | |
| 9 | 文字表达效果 | √ | | √ | | √ | |
| 10 | 学术规范程度 | √ | | √ | | √ | |
| 11 | 知识产权解决方案 | √ | | √ | | √ | |
| 12 | 社会效益 | √ | | √ | | √ | |
| 13 | 安全性 | | | √ | | √ | |
| 14 | 稳定性 | | | √ | | √ | |
| 15 | 互补性 | √ | | √ | | √ | |
| 16 | 界面友好性 | | | √ | | √ | |
| 17 | 出版时间 | √ | | | | √ | |
| 18 | 收录时间跨度 | | | √ | | | |
| 19 | 更新周期 | | | √ | | √ | |
| 20 | 利用率 | √ | | √ | | | |
| 21 | 收录书刊数量 | | | √ | | | |
| 22 | 专业书刊比例 | | | √ | | | |
| 23 | 与著名评价工具的耦合度 | | | √ | | | |
| 24 | 价格 | √ | | √ | | | |
| 25 | 经济效益 | √ | | √ | | √ | |
| 26 | 合计 | | 14项 | | 24项 | | 16项 |

(1) 学科针对性,是图书馆数字化信息资源建设精选信息时首先要考虑的因素,因为建成以后的数字化信息资源是否具有较高的实用价值,是评价数字图书馆建设是否成功的重要标志。评价信息资源的学科针对性主要是看其所含内容涉及的学科范围与本校、本系统的专业设置、重点建设学科、研究课题、博硕士点设置,以及本地区的经济发展特点、社会文化建设工程等,是否相符合、吻合的程度有多大。

(2) 功能配置,主要是相对于数据库而言的,不同的数据库提供的功能与服务不尽相同。如《中国学术期刊全文数据库》(CNKI)和《中文科技期刊全文数据库》(维普)同样是全文数据库,但是二者除了在收录的期刊品种和数量上存在较大差别外,在检索途径及其他配套功能也有不同。CNKI有15个检索入口,而维普只有10个;CNKI除了提供原文全文阅读、下载服务外,还提供引文检索与题录下载服务,而维普如要查询引文,还得要另

购其引文数据库。另外,有些数据库除了提供学术论文全文或摘要、索引外,还提供详细的投稿信息以及该数据库使用培训材料等。这些都是精选数字化信息资源时需要认真评估的内容。

(3) 数据提供及访问方式,亦即是使用方式的问题,有些数据库提供网上包库访问,有些数据库则提供镜像使用,有些数据库则二者均提供,有的除此以外还提供数据光盘。不同的数据提供与访问方式对于其后续服务的影响很大,比如镜像使用的数据库,内容更新不如包库访问的来得快,而且购买单位需要添置一定容量的存储设备或服务器并要安排专人予以维护,但是一旦互联网发生故障或停止购买,原有的数据还可以继续使用。否则,可能变得一无所有。因此,这也是选购数据库时需要考虑的重要因素。

(4) 学术规范程度,一定程度上反映作者治学的严谨性、引用数据或事实的核查性。需考查的项目包括作者署名方式、关键词标引的规范性、正文结构的完整性、行文描述的客观性、注释标示的准确性、参考文献著录的标准性等。

(5) 知识产权解决方案,是目前商业性学术数据库业发展中比较引人关注的问题,有些数据库开发者与源信息出版商或作者个人达成版权使用协议,有的则没有。一旦出现产权纠纷,不仅数据库开发商和代理者的销售行为受到影响,购买者的使用同样也会受影响。因此,图书馆应尽可能地选购那些知识产权问题解决的比较好的数字化信息资源。同时,在数字化馆藏传统文献时,亦要注意其版权状况。

(6) 社会效益,主要指有关信息资源的社会影响,包括相关的报道、评论、转载、被引情况等,过去有多少,当前有多少,以后可能还会产生什么样的社会影响等。

(7) 互补性,指拟建设的数字化信息资源与传统馆藏文献资源之间、数字化信息资源相互之间,在学科内容、功用等方面存在的相互补充的程度。相互之间的互补性越大说明这样的数字化信息资源建设的作用和意义越大,建成后的使用价值可能也就越高。

(8) 利用率,包括馆藏传统文献资源的借阅率与复制率、网上信息资源的点击率、商业学术数据库试用期间的访问与下载率等,一般说来,信息资源的利用率越高,读者需求越是强烈,越是值得列入数字化建设范围中来。

(9) 与著名评价工具的耦合度,即有关的学术数据库中所收录的期刊与著名评价工具中所收录的期刊的吻合程度,如 CNKI 中的期刊有多少种被列入 CSSCI、CSCI 的来源期刊,Elsevier SDOS 中的期刊有多少种是 SCI 收录的,等等。一般来说,与著名评价工具的耦合度越高,表明这个数据库的学术性越强、学术价值就越高。

其他指标的含义比较容易理解,不作一一解释。

3.2 权重问题

给每项测评指标赋予适当的“权重”,是建立一套科学合理的评价指标体系的一项重要工作。笔者在“网络信息资源的评价”^[8]一文中曾经利用“调查求重”方法,为网络信息资源的评价指标体系中的每项指标求得一个权重,但是在数字化信息资源建设信息精选的测评指标体系中,笔者仅是列出“权重”项目,并没有为每项指标求得权重,而是留待具体的测评者去求取。笔者以为,不同的图书馆由于其所在学校的学科性质、专业及其课程设置、学科建设侧重点与发展方向、服务对象不同,或所在地区的社会经济、文化等的发展特点、以及国民素质状况不同,而对于信息资源的价值认识与要求存在一定的差异,即使是同样的信息对于不同图书馆的重要性也不相同,不同的图书馆完全有可能为相同信息的同一指标赋予不同的权重。因此,为信息精选测评指标体系中的每项指标分配固定的权重是不妥的,而应由各图书馆根据本馆及其读者的实际信息需求情况,通过一定的方法来求得相关指标的权重。

4 图书馆数字化信息资源建设中信息精选流程

信息精选是一项复杂而细致的工作,实际操作时应制定详细的精选策略和实施步骤。为保证精选质量,至少要经过精选信息范围划定、信息测评指标体系制定、对具体的信息资源进行评价与筛选、专家(读者)对精选初步结果进行评议、精选结果调整等流程(如图2所示)。

(1) 划定精选信息范围,主要是根据图书馆数字化信息资源建设的目标及数字图书馆建设的整体战略,对精选来源信息的数量予以适当控制,既避免大海捞针、做大量无效劳动,又可减少评价信息资源总量、提高评测速度,保证精选质量效果。

(2) 确定信息精选数量(比例),就是根据本单位(地区)的学科建设、经济与文化建设、以及读者的实际需要、本馆的具体馆情等因素,来确定数字化信息资源的建设规模。如要从传统馆藏文献资源中选出多少类别多少种或者说以什么样的比例来进行数字化转换,要购买多少种学术数据库,等等,在正式实施信息精选前都应该有一个相对明确的概念。

(3) 拟定信息评测指标及权重,其主要工作是求取各项指标在整个评测体系中的权重。要注意的是,即使是同一项指标,在评估不同类型信息资源时其权重值也是不同的。权重的求取可以采用特尔非法,即专家多次调查的方法,也可以采用一般性读者调查求重的方法,由他们为每项指标的重要性程度打分,然后进行统计分析计算出各项指标的权重值。这样求得的权重比较客观,较少带有评价者个人因素。

(4) 对精选范围内所有信息进行综合测评与筛选,就是根据拟定的评测指标体系对有关信息资源进行全面的评价与打分,再根据其综合得分情况,按照预定的精选数量或比例,选择其中最好的部分。测评时,无论定量指标还是定性指标,最终都将其评价结果予以量化,以便同一测评体系内的所有指标最终可以按照综合得分进行统一排序。至于打分方法可以采用百分制、十分制、或五分制,但是无论采用什么计分方法,都应该制订一个比较详细的赋分标准,而且要保持一定时期内的相对稳定性。

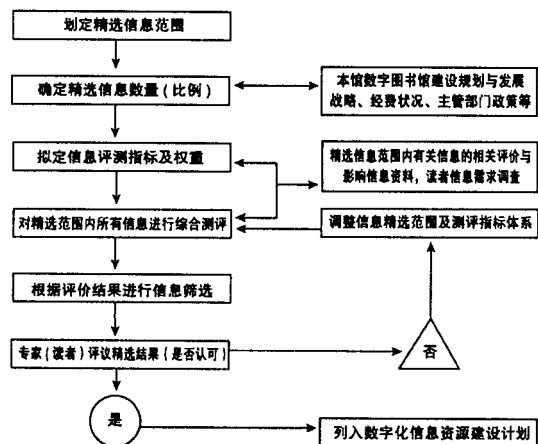


图2 信息精选流程示意图

(5) 专家(读者)评议精选结果,这也是再次征求读者反馈意见的过程,将精选出来的初步结果提交专家或广大读者讨论或试用,检验其是否真正符合用户现实的信息需求,是否可以激发或满足他

们潜在的信息需求,同时进一步征求他们的意见或建议。对于他们认可的则将其列入本馆的数字化信息资源建设计划;否则,进行再一次的精选。

(6) 精选结果调整,也就是对有关信息进行再次评价与选择。当图书馆精选出来的信息资源得不到专家(读者)认可时,则图书馆非常有必要进行再次甚至多次的精选工作。主要是结合专家(读者)的评议意见,对精选来源信息范围、评价指标体系(包括指标权重)进行适当的调整,使其更加科学、合理。

5 结 语

图书馆作为集保存人类文化遗产、提供社会信息服务、辅助国民终身教育等多种功能于一体的,文化性、公益性为主体的服务机构,在市场经济全球化与知识经济逐渐成熟的信息社会环境中,无论从建设成效、传承优秀文化,还是从读者服务的角度考虑,其数字化信息资源建设都要进行信息精选。但是由于信息精选的概念始终未被明确提出,更无成熟的理论产生并对实践予以指导,所以,虽然其一直为图书情报界、编辑出版界等所实践,但也无法提供成功的案例以供分析或借鉴。笔者在此仅以自己浅显的理解,试图通过对图书馆数字化信息资源建设中的信息精选问题的分析来阐述“信息精选”的某些方面,其中的有些提法不一定准确,观点不一定正确,乃笔者水平所限,真诚欢迎同行批评指正。

参考文献

- 游红.论布拉德福德定律齐普夫定律洛特卡定律的关系[J].情报杂志,1995,(6):68-69.
- 孟广均,等.信息资源管理导论(第2版)[M].北京:科学出版社,2003.7.
- 田建平.编辑评价图书的八大值素[J].编辑之友,2003,(5):24-27.
- 邱均平,等.期刊评价指标体系及定量方法研究[J].现代图书情报技术,2004,(7):23-26.
- 刘颖,刘霞.网络环境下电子文献的评价与选购[J].图书馆建设,2001,(5):39-41.
- 罗春荣,曹树金.因特网的信息资源评价[J].中国图书馆学报,2001,(3):45-47,52.
- 应峻,徐一新.网络全文数据库资源评价[J].现代图书情报技术,2005,(3):57-59,15.
- 陆宝益.网络信息资源的评价[J].情报学报,2002,(1):71-76.

(责任编辑:徐波)