



基于机构知识库成果引文分析的数字资源保障策略研究*

□陈雅迪 史淑英 张雪蕾 邵晶

摘要 在新形势下,大学图书馆普遍面临着“文献购置经费的增长速度低于数字资源的价格涨幅”的严峻问题,如何利用有限的经费最大限度地满足学科建设与发展的需要,成为图书馆数字资源建设面对的核心问题。要把有限的经费用在“刀刃”上,必须聚焦用户的核心信息需求。为了全面评估图书馆的文献保障水平,文章构建了基于机构知识库成果引文分析的自动化分析流程,考察全校及各学院发文成果参考文献的“收藏”和“缺藏”情况,建立可推广的数字资源保障方案。

关键词 引文分析 机构知识库 保障策略 数字资源

分类号 G252

DOI 10.16603/j.issn1002-1027.2022.02.005

1 引言

经过 20 余年的发展,多数图书馆在数字资源建设方面基本形成了自己的保障体系。在数字资源高速发展的过程中,一般会存在重采购、轻评估,重资源价值、轻使用价值等问题。如何对保障体系进行综合评价,对现有体系进行微调,在经费受限的前提下尽可能满足读者的需求,是现阶段数字资源建设的重点和难点。在评估数字资源的用户需求方面,较为传统的方法是通过数据库商提供的统计报表、基于网络流量、反向代理等方式以分析读者的使用情况,尽管这类方法能够评估用户对已购数字资源的使用偏好,但却不能发现未购置数字资源中是否有用户的核心需求,难以判断图书馆所采购的数字资源是否保障到位。

为了全面评估用户需求,学界提出通过引文分析来检验图书馆的文献信息保障水平,聚焦用户的研究成果,考察高频引用数字资源的保障情况。引文模式揭示了科学知识的产生、验证和交流发展的一般性规律,因此,通过引文分析,图书馆可确定用户在科研过程必不可少的资料范围。引文模式能够更全面、客观、有针对性地为馆藏建设提供决策支持,其优势在于:(1)图书馆可针对机构、院系、学科的特定用户群体,开展学术成果引文分析,实现精准

文献保障评估;(2)引文来源渠道多元化^[1],并不局限于图书馆已购馆藏数字资源。因此,引文模式既可以用于衡量用户对已购数字资源的真实需求,为未来数字资源续订提供决策依据;也可以用于查漏补缺,发现读者高频引用但未被购买数字资源,进而调整馆藏数字资源配置。

2 引文分析相关研究

引文分析的发展可划分为四个阶段:萌芽期(20 世纪 20—50 年代)、奠定期(20 世纪 50—80 年代)、发展期(20 世纪 80 年代以后)和繁荣期(至今)。在引文分析的萌芽期,虽未能形成系统性的理论,但相关探索为后期研究奠定了知识基础^[2],如格罗斯(Gross)首次通过引文分析法确定了化学专业的常用期刊范围,开创了引文分析的先河^[3]。在奠定期,加菲尔德(Garfield)提出了引文索引检索方法^[4],《科学引文索引》发布,更加完善的引文分析开始被应用于馆藏建设与评估^[5]。但是在本阶段,统计过程只能通过手工查阅,限制了引文分析的效率。在发展期阶段,随着计算机技术的发展,引文索引数据库在线开放,参考文献获取效率显著提升,同时大量引文分析研究成果涌现^[6]。到繁荣期阶段,引文数据库不再被 Web of Science 垄断^[7],Scopus、谷歌学

* 高校图书馆数字资源采购联盟(DRAA)课题“基于引文分析的数字资源保障策略研究”(编号:2021DRAA18)的研究成果之一。

通讯作者:陈雅迪,ORCID: 0000-0003-3874-1958,邮箱:chenmodidi@xjtu.edu.cn。



术(Google Scholar)等多个平台相继出现,对数字资源的覆盖范围也更加广泛^[8],用户可自由检索来自不同来源的引文数据来评估各类数字资源的引用。

虽然引文繁荣于商业领域,但是作为学术界科学对话的一种方法,也激发了图书馆在数字资源保障评估和数字资源建设配置等方面的研究。早在2005年,刘庆亮等就针对学位论文引文进行了数字资源保障分析^[9]。之后也有不少学者采用类似方法对某一学科或某一类型文献的引文进行分析,评估数字资源保障情况^[10-14]。随着引文分析方法逐渐被图书馆所重视,形成了以CASHL为代表的基于引文分析的数字资源评估实践^[15]。上海交通大学在通过引文分析作者引用行为的基础上,探索引文分析系统的需求,联合南京昆虫软件公司合作开发了数字资源利用绩效分析平台(ERS),ERS系统被广泛应用于数字资源保障评估实践^[16],这标志着将引文分析用于数字资源保障评估形成了产业化。虽然基于引文分析的利用涉及到方方面面,但在自动化收割特定对象的数字资源保障情况方面尚未有系统解决方案,受制原因主要包括以下几个方面:

(1)缺少精准分类的数据基础。在评估特定科研群体保障情况时,由于不同科研人员提交的学术成果存在机构地址不规范、全称缩写不一致等问题,图书馆通常需要投入大量人力和时间,进行繁琐而耗时的成果归属匹配工作,制约了引文分析开展的范围和频率。

(2)缺少多源数据集成机制,已有研究大多只针对特定机构的某一个学科分析,其调查范围局限、样本量较少^[17],不能反映用户信息需求的全貌;已有研究大多针对期刊层级,缺少数据库级统计,在包库订购模式已经成为主流采购方式的今天,仅对期刊进行引用分析,难以为采访决策提供直接性的依据。

(3)缺少对读者常用数字资源的自适应界定方法。常见的经验性方法,如采用对引用频次前99种期刊计算保障率^[18],或剔除被引频次数小于1或小于5的数字资源^[19]等方式受主观因素影响较大,不同的研究者的标准也存在差异,不利于实施推广。也有研究采用“二八定律”或“布拉德福定律”界定常用数字资源,但实际的数据分布可能并不符合合理规律。如希尔维亚(Sylvia)在研究中发现,排名最高的前23%的期刊仅产生了66%的引用^[20];扎夫伦尼沙(Zafrunnisha)在研究中发现,引文的过度集中

导致数据分布不符合布拉德福定律^[21]。因此,如果直接沿用理论公式,则可能导致常用数字资源的估计过于狭窄或宽泛。

引文分析可准确评估机构科研群体对于数字资源的使用情况,通过调整采购策略满足各科研群体的核心需求。本文针对科研群体的引文分析研究,直接复用西安交通大学机构知识库中精准分类到学院的研究成果数据,从期刊、数据库维度综合评估,全面把握校级、院级用户的常用数字资源;在常用数字资源的界定上,采用引文曲线“肘点”作为判别依据,自适应地确定用户常用数字资源范围,从而保证数字资源评估的全面性、客观性和准确性。

3 基于机构知识库成果引文分析的数字资源保障评估关键技术

3.1 运用引文分析实现数字资源统计利用的技术框架

机构知识库已完成了西安交通大学成果院系的归属分配,开展引文分析时可直接抽取院系成果数据;建立应用程序编程接口(Application Programming Interface, API)数据通道,打通分散的异构数据平台,整合机构知识库、引文数据库和已购数字资源目录,形成多源数据集成分析的有效机制,并运用自适应算法界定读者的常用数字资源,获得客观性的评估结果。技术框架如图1所示。

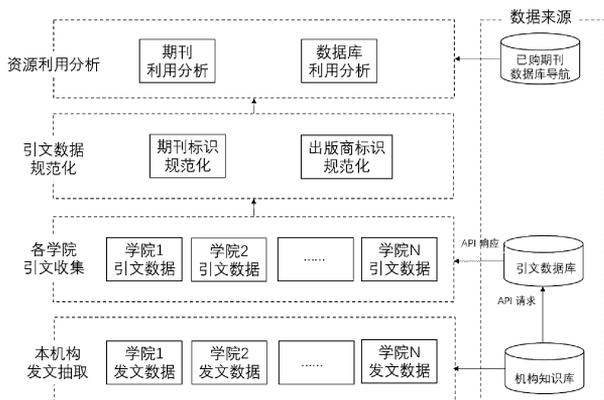


图1 基于机构知识库的引文分析技术框架

3.2 机构知识库中各学院发文的引文收集过程

之所以使用机构知识库的成果数据进行引文分析,是因为机构知识库已实现对研究成果按署名单位归属到学院,能为学院级的引文分析提供精准分类的数据基础。借助引文数据库的API,可实现机构知识库中各学院发文的引文收集。其具体流程如下:

(1)获得引文数据库的API访问权限,根据引

文数据库的 API 调用规范,对机构知识库的研究成果建立数据访问策略,包括对引文信息的请求方式、请求间隔控制、响应格式处理等。

(2)通过引文数据库的 API 发送数据请求,请求参数包括研究成果的识别号、请求参数;服务器经验证后返回对应参考文献。根据参考文献入藏号,进一步利用 API 数据视图,查询参考文献对应的出版来源。

(3)依据机构知识库对研究成果的精准分类,统计全校、各学院研究成果所引用的参考文献的来源期刊、出版商等信息,并根据其引文曲线识别其中的常用数字资源,进一步评估已购数字资源对常用期刊、数据库的覆盖情况。

3.3 常用数字资源的界定方法

高需求和低需求数字资源可根据引文信息的离散分布来区分。大量研究表明,在特定领域中,学者对文献的引用呈现出均匀分布规律;其“期刊—被引量”的降序排列曲线呈现“长尾”形态,即位于头部的少数期刊集中了大量被引量。在经费限制条件下,头部高频使用的数字资源,应当被甄别为常用数字资源,对其优先保障;频次较低、处于引文分布尾部位置的期刊应酌情考虑。在界定常用数字资源范围时,应当从引文分布实际情况出发,对引文分布中集中了高频引用的“头部”数字资源实现自适应划分。

引文分布曲线呈现为“长尾形态”,需要明确“头部”与“尾部”的最佳分割点。如图 2,由引文曲线的形态可知,随着排序增加,引用量先是陡峭下降,随后平缓;其下降幅度逐渐减小,并最终趋于稳定。因此,“头部”与“尾部”数字资源的最佳分割点即为引文曲线降幅明显放缓的点,即曲线的“肘点”。引文曲线的肘点可通过计算曲线相邻两点间夹角的最小值进行判别,该方法被称为“肘形判据”。采用“肘形判据”界定常用数字资源范围的步骤如下:

将某用户群体所用数字资源按引用量降序排列,其数字资源集合为 $X = \{x_1, x_2, \dots, x_n\}$,数字资源引用量序列为 $Y = \{y_1, y_2, \dots, y_n\}$,数字资源排名序列为 $R = \{1, 2, \dots, n\}$,假设序列 Y 严格单调递减,由图 2 可知,某点 (i, y_i) 与相邻两点连线的夹角 θ_i 为:

$$\theta_i = \frac{\pi}{2} + \theta_a + \theta_b \quad (1)$$

其中,

$$\theta_a = \arctan \frac{1}{y_{i-1} - y_i} \quad (2)$$

$$\theta_b = \arctan(y_i - y_{i+1}) \quad (3)$$

采用“肘形判据”,则 Y 的肘点 (z, y_z) 应满足:

$$\theta = \min_{1 < i < n} \left[\frac{\pi}{2} + \arctan \frac{1}{y_{i-1} - y_i} + \arctan(y_i - y_{i+1}) \right] \quad (4)$$

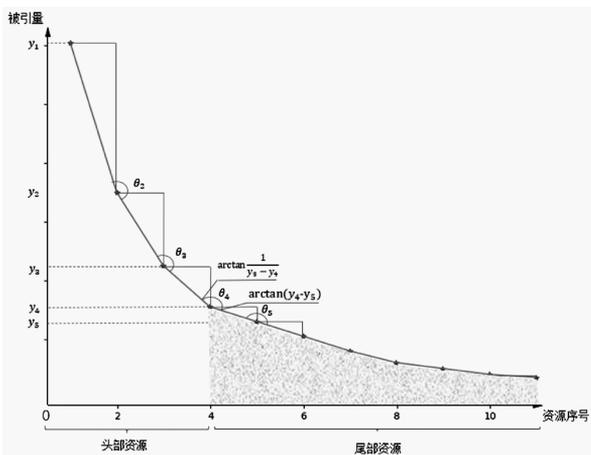


图 2 引文曲线肘点计算示意图^①

需要注意的是,“肘形判据”适用于严格单调递减的曲线。但在实际中,存在多个数字资源具有相同引用量的情况,导致引文曲线呈现为如图 3 所示的阶梯形态。如果直接应用肘形判据,则肘点计算可能会出现偏差。

为了避免对计算造成干扰,应当对引文曲线的构造方法进行改进:将数字资源按照引用量降序排列,然后将相同引用量中排名最后一位的数字资源抽取出来,建立对应的排名序列 $\bar{R} = \{r_1, r_2, \dots, r_m\}$ 和引用量序列 $\bar{Y} = \{y_1, y_2, \dots, y_m\}$,该方法可确保引文曲线严格单调递减,满足“肘形判据”的应用条件。改进后,引文曲线肘点的计算公式为:

$$\theta = \min_{1 < i < n} \left[\frac{\pi}{2} + \arctan \frac{r_i - r_{i-1}}{y_{i-1} - y_i} + \arctan \frac{y_i - y_{i+1}}{r_{i+1} - r_{i-1}} \right] \quad (5)$$

求取 θ 最小值处的肘点 (r_z, y_z) ,则用户所引用的前 r_z 个数字资源就是常用数字资源 $C(X)$:

$$C(X) = \{x_i, 1 \leq i \leq r_z\} \quad (6)$$

^① 在本例中引文曲线在 $(4, y_4)$ 处取得 θ 的最小值,故可确定肘点为 $(4, y_4)$ 。

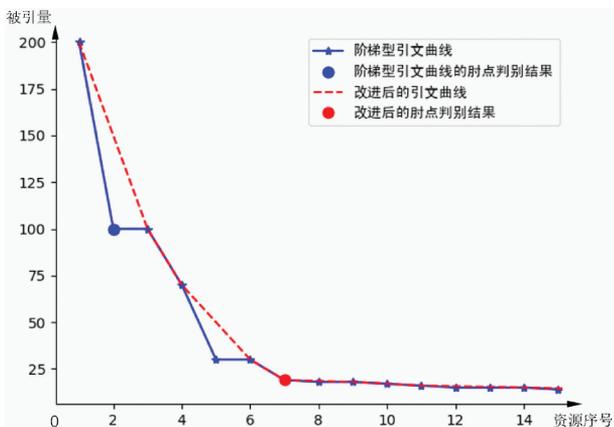


图3 阶梯型引文曲线、改进后的引文曲线肘点判别结果对比^①

3.4 校级重要核心数字资源的构建方法

为了最大限度保障学校重要核心数字资源,应兼顾校级用户整体的常用数字资源评估和院级常用数字资源的评估,以便将学院高频利用的专业性数字资源纳入到校级常用数字资源列表中,最终构建出校级重要核心数字资源列表。其构建策略如下:

假设学校有 N 个学院,在机构知识库中抽取各学院的研究成果集,经过引文分析后,第 i 个学院所引用的数字资源集合为 X_i ,其常用数字资源集合是 $C(X_i)$,则各学院的常用数字资源集合的并集可表示为 $\bigcup_{1 \leq i \leq N} C(X_i)$ 。将全校用户视为整体,其引用的数字资源集合为 $\bigcup_{1 \leq i \leq N} X_i$,其常用数字资源集合为 $C(\bigcup_{1 \leq i \leq N} X_i)$,构建的校级重要核心数字资源 C_H 为:

$$C_H = \bigcup_{1 \leq i \leq N} C(X_i) \cup C(\bigcup_{1 \leq i \leq N} X_i) \quad (7)$$

4 基于机构知识库成果引文分析的数字资源保障评估实践——以西安交通大学为例

4.1 校级科研成果发文的引文分析的数据模型构建过程

以西安交通大学机构知识库作为学校各学院科研成果发文的信息来源,选择 2019 年第一单位或第一作者单位为“西安交通大学”的学术成果。数字资源类型为期刊、会议等连续出版物(在后文统称为数字资源),共计 11064 条。以 Scopus 数据库作为参考文献信息的主要来源,通过 API 接口获取研究成果的参考文献信息,构建如图 4 所示的数据结构 ER 模型,其过程如下:

(1)数据导入。连接机构知识库数据库,创建机构发文数据表,其属性包括:机构知识库成果 ID、标题、作者、发表时间、所属学院、成果的 SCOPUS ID。

(2)收集参考文献数据。在 Elsevier Developers 中申请数据请求权限、身份标识秘钥 API KEY。对每一条发文数据,通过 GET 方法向 API 中传入 API KEY 和 SCOPUS ID,选取检索分面 VIEW=REF,获取参考文献的 SCOPUS ID。通过 API 接口进一步查询参考文献的详细信息,获得题名、作者、ISSN 号、出版商名称等字段并创建参考文献信表。

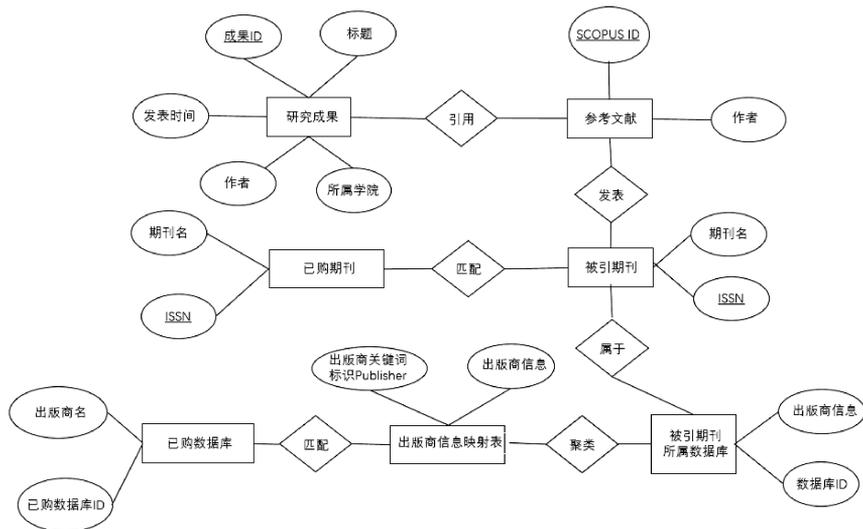


图4 西安交通大学机构知识库成果引文分析的 ER 模型

^①如阶梯型引文曲线直接应用“肘型判别”,由于受到曲线阶梯形态的干扰,肘点判别结果并非头部数字资源和尾部数字资源的最佳分界点;改进后的引文曲线为严格单调递减,采用“肘型判别”计算可获得正确的结果。



(3)规范引文数据。由于引文数据中,同一种期刊、数据库存在多种写法,如 ISSN 号为 0003021X 的期刊就有三种写法:JAOCs, *Journal of the American Oil Chemists' Society*、*Journal of the American Oil Chemists Society*、*Journal of the American Oil Chemists' Society*; Elsevier 出版商写法多达 17 种,如“Elsevier Ltd”“Elsevier Science Ltd”“Elsevier Sci LtdExeter, United Kingdom”等。不规范的写法显著影响后续统计结果,应进行数据清洗和规范化处理:采用 ISSN 作为期刊唯一标识;通过关键词聚类+人工甄别,将出版商信息统一标识为出版商关键词 Publisher,并建立出版商信息关键词映射表。

(4)统计期刊引文频次和数据库利用频次。对机构发文数据表按照学院分组,共获得 17 个学院发文数据表。将机构发文数据表、学院发文数据表与参考文献信息表通过 SCOPUS ID 连接,对参考文献信息表中的 ISSN 和 Publisher 应用 COUNT 方法进行统计,获得学校、学院对期刊的引文频次及数据库的利用频次,按照引用量降序排列,建立校级、

院级的期刊引用频次统计表和数据库利用频次统计表。

(5)期刊与数据库的利用分析。以学校、学院为统计单位,分别获得校级和院级期刊引文曲线、数据库利用曲线,得到全校、各学院的常用期刊和常用数据库,然后对常用期刊、常用数据库与已购数字资源建立关联查询,评估已购数字资源对常用数字资源的覆盖情况。

4.2 常用期刊的保障分析

根据建立的校级、院级期刊引用统计表,分别绘制全校用户及各学院的期刊引文曲线。如图 5 所示,引文分布均呈长尾形态,但其集中程度有所不同。从实际引文分布来看,校级用户和大部分学院的引文分布均不符合“二八分布”。当研究范围越广泛时,所引用的文献分布的集中程度越弱;当研究范围越集中时,所引用的文献分布的集中程度也越强。因此,出现了实际的引文分布不符合理论规律的情况。因此,采用“肘型判据”区分引文曲线中的头部数字资源、确定校级及院级的常用期刊范围更为合理。

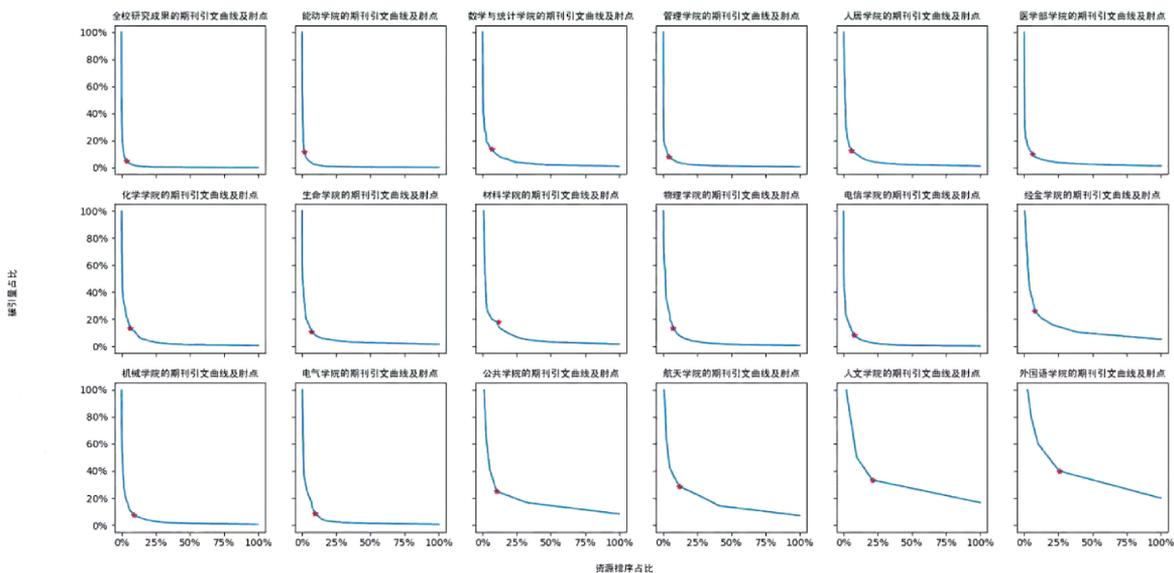


图 5 西安交通大学全校、各学院对期刊的引文曲线及肘点分布

通过计算肘点,将排名在肘点位置之前的期刊划分为“全校常用期刊列表”“各学院常用期刊列表”。从常用期刊的数量来看,各学院存在明显差异。对于电气工程学院、材料科学与工程学院、外国语学院等学院,其引用的期刊种类相对较少,因此常用期刊的数量较少;对于医学部、生命科学与技术学

院、机械工程学院等学院,其引用的期刊种类较多,因此其常用数字资源的数量也较多。该现象可能与学科分支、交叉程度有关,交叉和分支越多的学院,其引用的来源期刊也越广泛。虽然不同学院的常用数字资源数量存在差异,但其累积引用率大体相近,均位于 71.82% 附近。



由“全校常用期刊列表”“各学院常用期刊列表”可知,如图6所示,有624种期刊既存在于“全校常用期刊列表”中,也存在于“各学院常用期刊列表”中。有1026种期刊存在于“各学院常用期刊列表”中,但并未被“全校常用期刊列表”包含,表明这些期刊尽管在校级用户的期刊引用列表中排名靠后,但对特定学院科研群体不可或缺。有121种期刊存在于“全校常用期刊列表”中,但未被“各学院常用期刊列表”包含,表明这些期刊的主要用户分散于不同学院,尽管在每个学院的期刊引用列表中其排位都不够突出,但综合来看,却反映了不同学院用户群体的共性需求。因此,在构建“校级重要核心期刊集合”时求取并集,即1771种期刊。

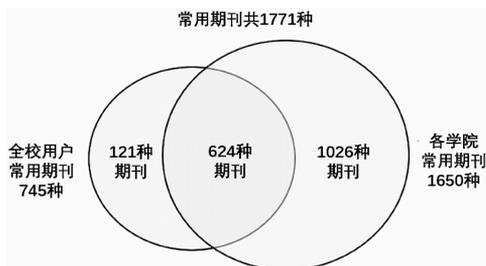


图6 西安交通大学全校常用期刊和各学院常用期刊关系

将“校级重要核心期刊集合”在图书馆已购期刊导航中关联查询,分析其保障率,计算结果如下:已购期刊对“校级重要核心期刊”的总体保障率为98.31%,各学院的保障情况如图7所示,其中8个学院的常用期刊达到了100%保障,9个学院的常用期刊保障率超过了95%;未保障期刊的数量仅占1.69%,整体表明图书馆校级、院级的常用期刊保障充分。

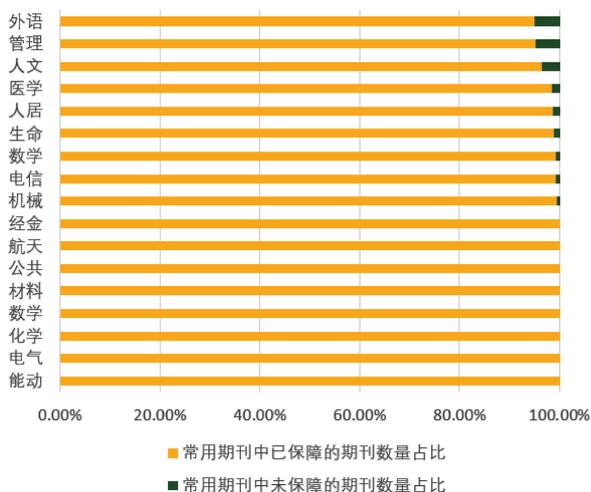


图7 西安交通大学各学院常用期刊保障情况

在未保障的期刊中,有一些是小型学协会或小语种期刊。例如APA出版的 *Journal of Personality and Social Psychology*、*Psychological Bulletin* 和 *Psychological Methods*,均为管理学院的常用期刊。尽管其引用量在校级期刊引用统计表的排位相对靠后,但在管理学院的期刊引用统计表中排在前列。日本 JAPAN Institute of Metals(JIM)出版的 *Materials Transaction*,尽管不是各学院头部引用的期刊,但在校级期刊引用统计表中排在前十,体现了多个学院的共同需求。对于这类常用期刊,可以根据需要予以保障,例如购买纸质期刊,或以文献传递的方式进行保障。

4.3 常用数据库的保障分析

为了确定全校、各学院的常用数据库,首先将参考文献归属到对应的数据库中,获得各个数据库的被利用频次,按照利用频次降序排列绘制全校、各学院对数据库的利用频次曲线。如图8所示,全校数据库利用频次曲线呈高度集中态势。根据引文曲线的肘点计算,发现其肘点位于第53个数据库,可确定校级常用数据库的范围为排名前5.4%的数据库,其提供的引用量占总量的90.54%。其中,Elsevier、IEEE、ACS、Wiley、AIP等数据库是利用频次最多的数据库。通过与数据库导航比对,校级常用数据库的保障率达到了100%。

各学院的数据库利用频次曲线如图8所示,经计算,其肘点位置分布于数据库的前6%至33%之间。相较于全校数据库利用频次曲线,学院的引文分布相对分散;表明校级常用数据库的优势在学院细分后出现下降,一些在整体排位相对靠后的数据库显现出对特定学科的支撑作用。

经计算,学院常用的数据库共101个。其中有30个数据库位于校级常用数据库的列表中。有71个数据库虽不在全校数据库利用频次曲线的头部,但却是学院常用的数据库。如,医学部排名第8位的BioMed Central数据库,在校级数据库利用频次排行中位于第69位;人居环境与建筑工程学院引用排名第3位的American Society of Civil Engineers(ASCE),在校级数据库利用频次排行中位于第105位;管理学院引用排名第6位的American Psychological Association数据库,在校级数据库利用频次排行中位于第119位。因此,仅以全校常用数据库作为参考会忽视不同学院科研所需的特色型数据库。

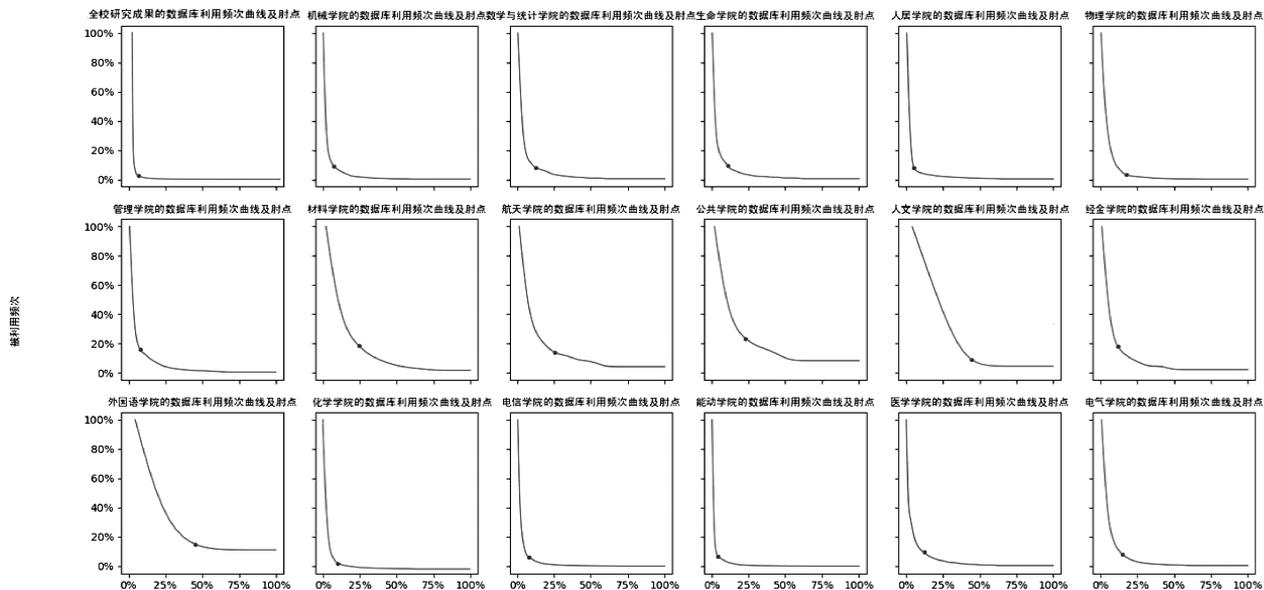


图8 西安交通大学全校、各学院对数据库的利用频次曲线及肘点分布

因此,基于统筹兼顾原则,通过对校级常用数据库和各学院的常用数据库求取并集,构建“校级重要核心数据库集合”。通过与已购数据库目录进行比较,分析其保障情况。经过计算,“校级重要核心数据库集合”的保障率为 93.71%,各学院的常用数据库保障情况如图 9 所示。

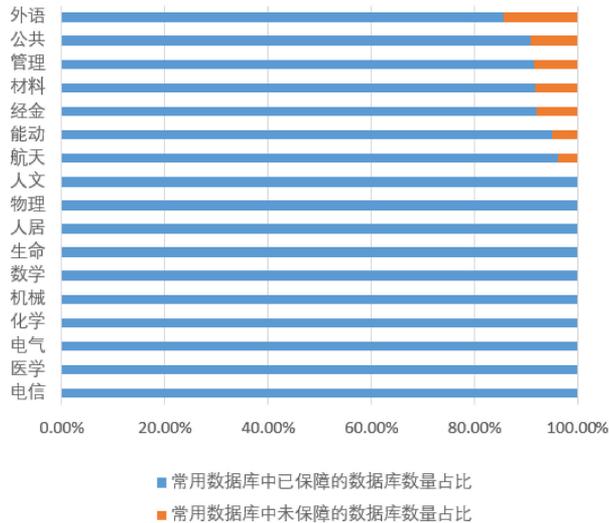


图9 西安交通大学各学院常用数据库的保障情况

其中,10 个学院的常用数据库的保障率达到 100%,7 个学院的常用数据库保障率超过 85%。总体来看,对全校、各学院常用数据库保障都处于较高水平。

对于未保障的数据库集合进行分析,发现数据库中 43% 的数字资源可以在 Elsevier、Wiley、

Springer、Taylor & Francis 等数据库中找到全文;没有替代性全文数字资源的数据库仅有一家,即 American Psychological Association (APA) 数据库。由于该数据库同时位于管理学院、公共政策与管理学院、外国语学院等学院的常用数据库中,因此可考虑需求针对性采购。

5 总结与思考

5.1 利用机构知识库开展数字资源保障评估是一种周期短、见效快的方法

目前,数字资源保障评估的难点和痛点在于,难以掌握学院对数字资源的利用情况、发现未购买的潜在的专业性数字资源需求。在开展数字资源评估时,往往是对全校用户整体发文情况分析,忽视了各学院对数字资源利用情况的不同。根据研究结果,如果仅对全校用户整体发文进行引用分析,就会导致学院特有的、专业性的常用数字资源需求难以被发现,不能判断学院的常用需求是否保障到位。学院发文的引文分析之所以难以开展,其根本原因在于研究成果的学院归属确认需要耗费巨大工作量。研究结果表明,利用机构知识库开展数字资源评估是一种周期短、见效快的方法,能够实现在校级、院级常用数字资源需求的全面评估。因为机构知识库已经实现了机构发文可精确归属到所属学院,省去了引文分析中学院归属的工作,因此能够显著提高数字资源评估分析的效率。



5.2 以“肘点”法界定常用数字资源相较“二八分布”更为客观合理

研究表明,大多数学院发文对期刊的引文曲线、对数据库的利用频次曲线均不符合“二八分布”。因此,在确定常用数字资源范围时,不可照搬理论公式,否则可能导致常用数字资源的估计范围过大或过小。在确定常用数字资源的范围时,应当采用自适应的方法。通过“肘形判别”确定曲线“肘点”、界定常用数字资源范围是一种行之有效的方法,对于不同文献集中程度的曲线均具有很好的适应性,能够有效识别常用数字资源的范围。

5.3 基于期刊的引文分析和数据库的利用分析有利于重建图书馆话语权

数字资源的采购往往是以数据库为单位而进行的一种大宗货物采购,在购置之初确实极大地丰富了图书馆数字资源数量,使得用户能够获取的信息内容成倍增长。但是,随着数字资源提供商之间的合并、重组、代理权的变更等,甚至于以盈利为目的各种资本的注入,以“数据库”为单位的“一揽子”交易,给图书馆的经费使用带来了巨大的压力,而这些“一揽子”交易中的数字资源内容往往由数字资源提供商确定,图书馆在这个过程中话语权缺失。本研究有助于图书馆掌握每个数据库的成本效益;有助于图书馆掌握每种期刊的使用情况,确定其中被高频引用的常用数字资源。从而有助于重拾图书馆在采购方案制定过程中的话语权,促进图书馆支付的成本产生与之相匹配的科研产出效益。

6 结语

文章提出了基于机构知识库成果引文分析的实施思路,并以西安交通大学为例,评估图书馆对全校用户及17个学院的公开发表成果所引用参考文献的保障水平。相较于传统方法,基于机构知识库成果引文分析方法有效提升了处理效率,有利于开展期刊级别和数据库级别的全面分析,有助于评估机构整体核心需求和学院特色需求。

本研究主要针对校级、院系公开发表文献的引文数据进行分析,未对学位论文、课题申报书等其他未公开类型资料进行统计。未来可进一步拓宽研究对象的范围,进行更为全面的文献保障评估研究。此外,由于引文信息不能反映文摘型、数值型等其他非全文数据库的被利用情况,因此,对于这类数据库的保障评估方案还应进行进一步的探索。

参考文献

- 1 沈艳红,吴信岚.基于中国引文数据库的核心书目测定研究[J].现代情报,2011,31(9):136-139.
- 2 侯剑华.基于引文出版年光谱的引文分析理论历史根源探源[J].情报学报,2017(2):26-34.
- 3 Gross P L K, Gross E M. College libraries and chemical education [J]. Science, 1927(66): 385-389.
- 4 Garfield E. "Science Citation Index"—a new dimension in indexing; this unique approach underlies versatile bibliographic systems for communicating and evaluating information [J]. Science, 1964, 144(3619): 649-654.
- 5 Kriz H M. Citation counting and the future of engineering libraries [J]. Engineering Education, 1977(67): 707-710.
- 6 Devin R B, Kellogg M. The serial/monograph ratio in research libraries: budgeting in light of citation studies [J]. College & Research Libraries, 1990, 51(1): 46-54.
- 7 Meho L I. The rise and rise of citation analysis [J]. Physics World, 2007, 20(1): 32-36.
- 8 Moed H F. New developments in the use of citation analysis in research evaluation [J]. Archivum immunologiae et therapeuticae experimentalis, 2009, 57(1): 13-18.
- 9 刘庆亮. 博士论文引文分析与信息数字资源建设——南开大学经济、管理学科部分重点学科博士论文引文分析 [J]. 图书情报工作, 2005(8): 131-135.
- 10 吴卫华. 从近5年引文分析看高校图书馆藏书建设——以河北理工大学冶金与能源学院教师所发论文为例 [J]. 河北理工大学学报(社会科学版), 2010, 10(5): 74-77.
- 11 高仕健, 段晓玲. 博士学位论文引文分析——面向多学科馆藏建设提供信息支持的新思路 [J]. 图书馆理论与实践, 2010(5): 9-11, 28.
- 12 刘良璧. 从硕士学位论文引文分析透视重点学科文献保障情况 [J]. 图书馆建设, 2010(4): 58-61.
- 13 郭璇. 我国图书馆学专业博士学位论文的引文分析 [J]. 图书馆工作与研究, 2011(2): 65-68.
- 14 雷顺利. 高校人文学科馆藏资源的保障与结构优化——基于CSSCI的引文分析 [J]. 图书馆, 2012(1): 118-121.
- 15 CASHL. 2020年CASHL前瞻性课题报告 [EB/OL]. [2022-03-02]. http://www.cashl.edu.cn/qzktbg?field_nianxian_value=&field_xuexiao_value=&field_zuoze_value=&page=0.
- 16 张赞玥, 秦鸿, 杜丽. 应用ERS评估外文电子书绩效的实践探索——以电子科技大学为例 [J]. 图书情报工作, 2020, 64(21): 50-57.
- 17 路一, 陈茹, 孟惠玲. 理工院校文科用户的文献信息需求及其保障 [J]. 情报资料工作, 2002(1): 51-53.
- 18 侯利娟, 郝群, 张立彬. 高校图书馆外文文献资源保障研究——以复旦大学数学学科为例 [J]. 图书馆, 2018(1): 92-100.
- 19 徐志玮, 郑建瑜. 高校化学学科用户对纸本/电子期刊需求研究——以中山大学化学学科用户为例 [J]. 图书情报知识, 2010(4): 44-50.
- 20 Sylvia M J. Citation analysis as an unobtrusive method for journal collection evaluation using psychology student research bibliographies [J]. Collection Building, 1998, 17(1): 20-28.
- 21 Zafrunnisha N. Bradford's zones and productivity of journals in psychology doctoral theses [J]. Annals of Library and Information Studies, 2012, 59(1): 39-52.

作者单位:西安交通大学图书馆,陕西西安,710049

收稿日期:2021年4月9日

修回日期:2021年10月15日

(责任编辑:关志英)

(转第26页)