

《中文图书引文索引·人文社会科学》 示范数据库研制过程、意义及其启示

□叶继元

摘要 学术图书是学者研究成果的集中体现。作者在写作过程中,参考、引证了大量文献,将这些被引文献析出即形成索引。因此,将精选学术图书作为来源文献,创建图书引文数据库是现实可行的。基于这一设想,并付之于实践,即形成了《中文图书引文索引》(CBkCI)。该数据库以精选学术图书作为来源文献(统计源),统计、分析图书作者引用图书、期刊论文、报告等所有文献资料的情况。CBkCI 示范数据库的研制成功,不仅填补了国内在图书引文领域的空白,促进学术图书出版质量的提升,而且有助于图书馆进行图书采访,精选馆藏,并为学术评价提供坚实的基础。

关键词 中文图书 引文数据库 引文索引 人文社会科学

1 引言

50 多年前,美国著名学者加菲尔德博士首创《科学引文索引》(SCI)、《社会科学引文索引》(SS-CI)等各种数据库,形成了“引文分析”新的研究领域,极大促进了图书情报学实践与理论的发展,为文献计量学、信息计量学、网络计量学,以及科学学、人才学等的发展做出了重要贡献。10 多年前,在参考国外研究成果的基础上,我国建立了《中国科学引文数据库》(CSCD)、《中文社会科学引文索引》(CSS-CI)等数据库,同样促进了中国图书情报实践与理论的发展。然而我们在引进、消化国外研究成果时,能否有“再创新”?这涉及到我国的图书情报学是否能自主创新、立于世界学科之林的问题。早在 8 年前,笔者及所在团队在深入开展中文引文数据库实践和研究的基础上,首次设想并提出研制《中文图书引文索引·人文社会科学》(CBkCI·H&SS,以下简称 CBkCI),当时国外还没有人提出此种设想,更没有图书引文数据库的出现,直到 2012 年 5 月美国 Thomson 才在中国发布以英文为主的《图书引文索引》(BkCI)的有关信息。因此,CBkCI 在字段设计、功能实现等多个方面具有独特性和创新性。回顾研制 CBkCI 数据库的过程、意义,不仅有助于回答上述是否能创新的问题,而且能加深对研制中文图书

引文索引及其数据库体系重要性的认识,增强图书情报学的学术创新意识和知识产权的保护意识。

2 提出研制 CBkCI 的最早设想

2004 年 4 月 21—22 日南京大学中国社会科学评价中心在四川大学组织“CSSCI 专家咨询委员会暨全国人文社会科学研究定性定量评价研讨会”,会上有的学者提出应将“以书代刊”的出版物纳入“来源期刊文献”中的建议,后又发文论述^[1],但由于这些出版物具有书号而无 CN 号,尽管其内容与“期刊”并无二致,但毕竟不是正式的“期刊”,不符合 CSSCI 来源期刊的基本条件,因此会议未采纳此建议。尽管未采纳,但笔者从此建议中敏锐地联想到一个关键问题,即图书能否作为来源文献(统计源)建立引文索引。此问题甫提出,立即引起团队中同仁的注意和探讨。

经过讨论,我们认为,从理论上讲,将图书作为来源文献是完全可以的,因为图书与期刊一样,都是重要的研究成果的体现者,作者在写作时,参考、引证了大量文献,完全可以将这些被引文献分析出来,形成索引。何况广义的图书,除了单本以外,还包括丛书,即有相对统一的题名、序号、连续出版,丛书与“连续出版物”最根本的区别是,前者的连续是“有限

的”,而后者则是意图“无限的”,正是由于这个意图“无限”,才使得从属于连续出版物的期刊作为来源文献具有了优越性——每年都可连续进行引文统计。而图书,尤其是单本图书,一次统计后,就需要再寻“统计源”,这使得筛选统计源的工作量很大,且图书数量庞大,筛选难度更大。因此,从实践上看,当时世界上还没有学者提出过这一问题,更没有一种图书引文索引数据库,尤其是久负盛名的美国ISI,自首创SCI等以来的几十年间,都没有研制图书引文索引,可见难度很大。“以书代刊”是我国一大“特色”,是一种非书非刊、介于书和刊之间的一种出版物,对这种出版物进行研究和探讨,进而筛选出“来源文献”,建立引文索引,不失为新的尝试。对“以书代刊”出版物的研究与思考,催生出了要研制图书引文索引的最早思想。

3 来源学术集刊的筛选及其引文示范数据库的建立

“以书代刊”是对无法申请到CN号(连续出版物号),只得在出版社出版(具有书号)的期刊的俗称,如果要从学理上研究这类出版物,应该进行命名并给予较严格的定义。经过对能收集到的数百种这类出版物的实物的阅读、分析和对已有术语的提炼、概括,课题组将这类出版物命名为“学术集刊”,并为其下了明确的定义:具有相对稳定、统一的题名,以分册形式,以年、半年、季度、双月、月、半月、双周、周等间隔定期出版,并有年、卷、期等标识序号,计划无限期出版,具有正式书号,能反映当前我国哲学社会科学各个学科、领域最新研究成果的连续出版物^[2]。该定义经过CSSCI咨询专家委员会的审议,遂成为较正式的术语。

2005年我们对学术集刊进行了摸底问卷调查。一共发出调查问卷323份,截止到2006年3月共收回有效问卷225份,占调查样本总数的70%。通过统计、分析,获得了225种集刊的创刊时间、出版周期、开本、页码及平均每期的载文量、发行价格、发行范围及每期印量、集刊所属学科情况、地区分布、主办单位、编辑部人员、经费来源等基本数据和内容概况。从调查结果来看,我国学术集刊都有自己明确的办刊宗旨,其栏目各不相同,但都有自己一些特色和固定的读者群。它们为研究人员成果的推出提供了一个强有力的平台,作者群多是相关专业的权威人士。此次调查为筛选来源集刊做好了较充分的

准备。

根据教育部社科司、学术批评网提供的资料,我们进行了补充、考证,当时有目录控制的集刊有539种。来源集刊的筛选需要集刊编辑部申报,这不同于CSSCI来源期刊不需申报的规定。截至2006年1月5日,共收到申报表和集刊样本243份(种),涉及20个学科,集刊的主办单位涉及全国19个省市,既有北大、清华、人大、北师大、复旦等数十所重点高校,又有相关领域的专业协会、学会。在243种集刊中,2000年或以前创办的、符合遴选条件的集刊为98种,涉及19个类。我们将98种集刊的定量评价数据(总被引次数、2004年的影响因子和加权总值)发给CSSCI指导委员会各位委员和部分教育部社会科学委员会委员进行定性评价,最终评选出来源集刊47种,其中33种有2005年版本,另14种无2005年版本。

自2006年12月25日开始建立学术集刊示范引文数据库,首先组织人员对32种集刊进行标引,到2007年3月20日全部数据初步总校完毕,先后历时3个月。32种集刊共有来源文献1178篇,引文22194条(其中外文引文2954条)。平均每种集刊约有论文36.8篇、平均每篇论文有引文18.8条。2005年集刊已标引、输入、初校完成。

学术集刊毕竟不是正式的期刊,由出版社出版,编辑、出版的规范性较期刊为差,使标引工作十分困难,如序号不全。有10种集刊,在版权页或封面页上同时标出了年度和卷、期、辑等序号;另有23种集刊只有卷、期、辑序号而无年代。“出版地”、出版者多头。论文数目信息不全,如文章大都没有关键词、中图分类号、中英文文摘;参考文献大都是以脚注的形式出现,且重复率很高。不少集刊繁体汉字印刷,给录入、校对工作增加了困难。有的集刊很难获得,到刊率很低,使数据加工工作无法进行。集刊每期刊载文量不一。必须对二年内出版的各期集刊的载文量逐篇统计,而后相加,计算其影响因子较为困难,等等。从总体看,学术集刊在编辑出版方面不大规范,但在质量方面确有不少精品,国际化程度较高,几万字的长篇小说不在少数,较符合学术发表要求和规律,但有些集刊未必如此,质量参差不齐。学术集刊来源文献的筛选和示范引文索引数据库的研制为图书引文索引数据库的研制奠定了基础。

4 CBkCI 示范数据库的研制

2004年,笔者作为首席专家主持国家社科基金重大项目《建立与完善哲学社会科学评价体系研究》,至2008年笔者及所在团队已发表有关评价的论著数十篇(部),在学术界、管理界产生较大反响。2008年在撰写该项目研究总报告时,有专章论述图书评价及其图书引文索引研制问题。2009年4月,笔者组织了《首届人文社会科学评价暨第五届科研绩效评价学术研讨会》,来自国内外学者、评价专家70余人参加了会议。作为组织者,笔者曾约请美国ISI中国区代表在会上谈论有关图书引文方面的问题,该代表曾答应。但由于时间限制及其他原因,该代表在会上未及谈论图书引文问题。2010年笔者指导一名博士研究生专门研究图书评价及其图书引文索引问题,撰写成博士论文,获得通过。2010年8月3日、2011年11月笔者在《光明日报》多次发文,明确提出要“建立评论数据库、图书引文数据库等,为创新和质量精细评价提供重要保障和基础条件”^[3]。2011年5月在南京召开、由爱思维尔组织的有关引文计量和评价的国际学术研讨会上,笔者作主旨报告《基于创新测度的科研及期刊“全评价”体系》,其中专门谈到图书引文索引问题,引起听众极大兴趣,在提问阶段,反响热烈。同年5月在北京参加国际知识共享与资源建设学术研讨会上,11月在杭州参加信息资源共享国际会议上,以及在苏州参加教育部人文社科高层论坛上,笔者应邀作主旨报告中都谈到这一问题。2011年1月,国家社科基金重大项目《建立与完善哲学社会科学评价体系研究》结项,《中文图书引文索引·人文社会科学》(CBkCI·H&SS)示范数据库作为该重大项目的副产品,提交到国家社科规划办,其有关信息被头版头条发表在国家社科规划办官网上^[4]。

4.1 CBkCI 数据库简介

CBkCI数据库来源图书管理模块主要设置的字段有:来源题名、副题名、英文题名、来源责任者等数十项。数据库文献检索模块包括来源图书检索和被引文献检索,可以提供地区、机构、学科、学者等多种类型的统计分析数据,从而为制定科学研究发展规划、科研政策提供决策参考;可以提供多种定量数据,如被引频次(可以细分为正面引用、负面引用与中性引用)、图书影响广度、地域分布等。学科类别分别依据《中华人民共和国学科分类与代码国家标

准GBT 13745—2009(人文社科类)》和《中国教育部学科分类表》。

来源图书检索入口主要有:来源责任者、责任者姓名拼音、第一责任者、英文题名、学位分类、学科分类、出版者或期刊社、出版时间或年代卷期、基金类别、基金、机构名称、第一机构、地区、主题词等。

被引文献检索入口主要有:被引文献责任者、被引文献题名、被引期刊或被引图书出版者、被引文献年代或出版时间、被引文献类型、被引文献细节、被引频次(可以区分正面引用、负面应用与中性引用)等。

4.2 CBkCI 数据库统计源的确定

课题组和我的博士生王兰敬将笔者提出的“全评价”体系应用于人文社会科学学术图书评价领域,拓展了学术图书质量评价框架体系,建立了中文人文社会科学学术图书“全评价”模型。根据形式评价、内容评价和效用评价新概念组合,采用引文等文献计量评价与同行专家评价相结合、以同行专家评价为主导的评价方法,基于人文社会科学学术图书“全评价”模型筛选出高质量的学术图书“代表作”,再通过学科同行专家小组审核方能确定为精品图书,即作为统计源(来源图书)构建中文人文社会科学学术图书引文索引数据库。

4.3 CBkCI 系统的设计

社会科学研究尤其是人文科学研究往往注重图书出版物,侧重于文本的解读,因此建立人文社会科学图书引文评价数据库,以图书引文方式来评价图书学术创新与学术质量是一种更好的途径,图书引文评价软件的设计显得至关重要。2011年11月初,征求有关评价专家意见后,课题组起草了图书引文评价软件即CBkCI数据库系统用户需求报告。之后,多次邀请有关评价专家、计算机工程师根据用户需求报告共同论证设计图书引文评价软件应该注意的一些问题。2011年11月底,有关评价专家对软件设计提出了一些修改意见和建议,到12月17日,软件设计基本完成。CBkCI是以评价中文人文社会科学学术图书质量为设计目的、兼及检索中文人文社会科学首版图书的数据库,该数据库以“全评价”体系为指导,定性与定量结合筛选来源图书。系统分为文献管理与文献检索两大模块。文献管理新增文献可以进行来源图书的添加、删除、修改和查询操作,来源图书的记录包括图书唯一标识、图书名

称、作者名称、出版社、出版时间、学科分类、ISBN、中图分类号以及作者维护与被引文献维护两个添加与修改模块。文献检索模块包括来源文献检索和引文检索两个模块，来源文献检索可以提供图书唯一标识、文献名称、英文名称、副标题、ISBN、文献版本、文献出版社、文献出版地、丛书信息、关键词、基金细节等字段按照特定的年份进行检索；引文检索可以提供被引文献责任者、被引文献题名、被引文献出版社、被引文献年代、被引文献类型等字段检索。具体检索界面见图1—图6(注：此界面系示范数据库最初设计的界面，目前界面、系统题名等均有所修改)。



图3 《中文图书引文索引·人文社会科学》作者维护界面



图4 《中文图书引文索引·人文社会科学》引文信息列表界面



图1 《中文图书引文索引·人文社会科学》用户登录界面



图5 《中文图书引文索引·人文社会科学》来源文献检索界面



图2 《中文图书引文索引·人文社会科学》来源文献录入界面

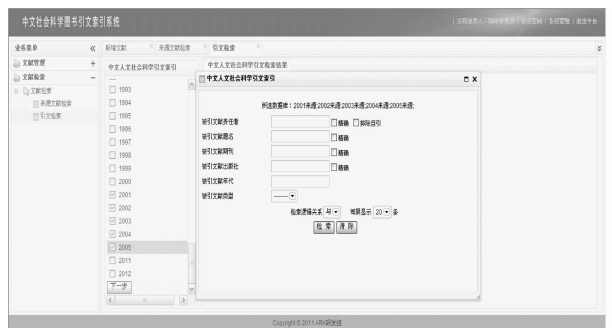


图6 《中文图书引文索引·人文社会科学》被引信息检索界面

4.4 CBkCI 系统的维护

CBkCI 系统软件安装试用后,课题组首先对图书引文评价数据库软件的数据字典进行维护,录入各种必需的字典数据,如中国地区代码表、中华人民共和国学科分类与代码国家标准、中国教育部学科分类表等。接着,根据中文人文社会科学学术图书“全评价”结果,将筛选出的来源图书逐项录入数据库,建立 CBkCI 示范数据库。在数据录入的过程中,课题组发现软件设计存在一些问题需要进行修改,如关于出版社和引文出版地的添加引文信息对话框里提供省份和城市下拉单以供选择,国外出版社就无法录入,不能保存记录,课题组在输入《现代性命运》(漆思著)第 8 页第③注释项就遇到这种情况,应该增加手工输入窗口;引文年月不应设置为必填项,为便于著录,应该增加手工录入方式。随着时间的推移,数据库的图书记录将会不断增加,数据越多,越能起到图书引文索引的检索、评价数据库的作用,充分显示其各项功能。

5 CBkCI 示范数据库研制的意义

CBkCI 是以精选的学术图书作为来源文献(统计源),统计、分析图书作者引用图书、期刊论文、报告等所有文献资料的情况,是以检索和评价中文人文社会科学学术图书质量为目的的数据库。现有的《中文社会科学引文索引》(CSSCI)实际上是“期刊引文数据库”,因为其来源文献是精选的期刊。这两个引文索引的来源文献不同,主要功能亦不同,二者各有侧重,形成互补。对于文科许多学科而言,如文史哲,学者更重视图书的作用。CBkCI 可以整合中文人文社会科学诸多学科领域图书和图书上的全部引用文献,为人文社会科学研究人员提供一个崭新的中文人文社会科学图书检索、分析与评价资源,帮助研究人员针对人文社会科学图书等文献进行定量、系统、客观、科学的统计与分析,具有检索功能。CBkCI 可以进行引文排序、高被引图书、引文分析方面的检索,是一种综合性人文社会科学文献检索工具,此外它还具有引文分析与评价功能,其评价功能是建立在引文分析基础之上的。CBkCI 可以提供被引频次(可以细分为正面引用、负面引用与中性引用)、图书影响广度、地域分布等多种定量数据。引文分析与评价功能主要体现在以下几个方面:首先

是及映各学科研究成果交流、借鉴等情况,其次是科研成果的评价。引文指标是比较有效的衡量科研成果质量的测度指标,一般情况下图书被引频次越高,反映了图书受学术共同体关注的程度越高,其著者的影响力越高,一定程度上反映了图书应用的价值越大。再次是人才评价,CBkCI 通过科研人员图书被引用或进入高被引,显示出科研人员的学术影响与研究水平,反映出科研群体对其依赖程度。使用该指标时,应结合其他指标特别是同行专家精细评审来合理地评价人才,应科学地、有分析地使用。还有科研机构评价,引文索引在机构评价中彰显了它独特的作用。该数据库将产出高被引图书的机构关联在一起,按共引核心图书进行排序,能使科研人员了解影响力突出的机构。此外,还有学科领域评价、出版社评价、期刊评价、作者评价等。

总之,CBkCI 示范数据库的研制成功,不仅填补了国内该领域的空白,而且可促进学术图书出版质量和论文等各种文献的质量的提升。它有助于图书馆推荐好书,精选馆藏图书,提高馆藏质量,有助于图书出版规范化建设,提高图书出版质量,有助于研究人员及研究生、大学生迅速获取精品图书及其他文献,为公正、公开、公平进行学术评价提供坚实的基础。

6 几点启示

CBkCI 的研制,从问题的提出、设想到示范数据库的编制,历时 8 年。回顾其研制过程,可以得到以下启示。

6.1 要有学术自信心

图书馆学起源于欧美各国,多年来,我国专业研究者一直处于引进、吸收发达国家的研究成果的阶段,这对于相对落后的学科的发展是必须的。但近 10 余年来,随着我国文献信息的激增,尤其是信息化、网络化、数字化的迅速发展,我国图书情报工作实践丰富多彩,所面临的问题,在发达国家也同样遇到,他们也在探讨这些问题,并没有提供现成的答案。因此,在这种情况下,我们应有足够的学术自信,在引进、吸收、消化的基础上,大胆再创新,往往能出奇制胜。CBkCI 的研制过程,充分说明了这一点。

6.2 要有知识产权尊重心

保护知识产权,尊重某种思想、观点提出的首发权,普及学术引文、注释规范,对于鼓励学术创新至关重要。应该说,我国在此方面还有很长的路要走。在我国图书情报界,不尊重知识产权的事情时有发生,比如,从学术论文、著作,或从学术会议资料和演讲中获得了有关学术信息、观点、思想,不少同行在自己成果中不加以明确说明,为了某种近期目的,盲目或重复研究,既劳民伤财,又牵涉侵权,严重影响他人研究的积极性。对此一定要高度重视和防范。

6.3 要有质量责任心

CBkCI 是一项提高我国文科研究质量的基础性工程,“基础不牢,地动山摇”,学术界、科研管理界必须有强烈的责任心,加强数据库数据质量控制。从大的方面看,每位图书的作者、每个出版社都必须严格执行写作规范和引文规范,不仅重视引文格式规范,更要重视内容引用规范。从索引编辑者方面看,必须严格数据标引、校对、反馈、核验等各个环节和程序。只有大家一起努力,数据库的质量才能得到保证,才能发挥引文数据库应有的作用。

6.4 要有学术耐心

CBkCI 是大型信息数据库,又是一项复杂、细致的系统工程,工作量大,尤其要对引文深度标引和加

工,更需要投入大量的人力物力,因此不论是索引研制者,还是广大学者、专家,一定要有足够的耐心,对不断出现的新问题,应积极面对,正确探讨解决之道,如此,该数据库才能在大家共同支持、帮助下不断完善。

致谢

课题组袁培国研究员、吴向东副研究员、郑德峻教授、袁曦临研究员和王兰敬博士对本文提出了许多很好的修改意见,特此致谢!

参考文献

- 1 杨玉圣. 值得关注的学术集刊现象. 云梦学刊, 2004(4): 29-30
- 2 叶继元等. 中文人文社会科学学术集刊现象的调查分析. 中国出版, 2006(10): 17-21
- 3 叶继元. 基于质量与创新测度的人文社会科学“全评价”体系及其应用. 光明日报, 2011-11-25(14)
- 4 叶继元. 建立和完善以质量和创新为主的哲学社会科学评价体系. [2012-02-10]. <http://www.npopss-cn.gov.cn/GB/219468/17082632.html>

作者单位: 南京大学信息管理学院, 南京, 210093

中国人文社会科学评价国家创新基地, 南京, 210093

收稿日期: 2012年9月10日

The Innovation, Significance and Enlightenment of the Database of Chinese Book Citation Index (CBkCI) Research & Development

Ye Jiyuan

Abstract: Academic book is a core expression of the academic research. In the process of writing, the author cited many literatures such as journals, reports and articles. Analyzing and fetching these citations of selected academic books, is the way to create citation indexes. Based on this concept and design, the Chinese Book Citation Index (CBkCI) is produced. The value of CBkCI is not only to fill the domestic blank in books citation, but also to promote the quality of academic book publishing, to contribute to library collection, and more importantly, to help to build up a solid foundation in the area of academic evaluation.

Keywords: Chinese Academic Books; Citation Database; Citation Index; Humanities & Social Sciences