

北京地方文献及其报刊索引数据库的开发与建设

[作者] 常林

[单位] 首都图书馆

[摘要] 介绍了首都图书馆北京地方文献及其报刊索引的发展历史。论述了北京地方文献及其报刊索引数据库的开发与建设。

[关键词] 数据库建设, 信息资源, 北京地方文献

1 导言

1.1 建设具有民族特色的数据库

在推进信息化建设中,保护和发扬民族优秀的文化传统,已引起世界各国普遍重视。我国有丰富的信息资源,中华民族悠久的历史 and 灿烂的文化,社会主义建设的经验和成果,以及经济建设、文化教育、科学研究等方面的重大进展,都是极其宝贵的信息资源。因此在建设具有民族特色的数据库领域,时代赋予我们的责任重大。

1.2 开发地方文献报刊资源

地方文献是指记录着某个地方自然、历史、社会现象的一切文献。这些文献包括历史地理、政策经济、语言文字、民族宗教、矿产资源、水文气象、农林牧副等,且表现于各种记载形式,如:图书、杂志、报纸、图片、照片、影片、画片、唱片、拓片、表格、传单、影片、画片、唱片、拓片、表格、传单、票据、文告、手稿、印模、簿籍等等。〔1〕。

其中,地方文献报刊是地方文献产生以来最大的信息源,主要反映最新的社会生活动态,而且还对地方历史进行着回顾和总结研究。因而由于其内容丰富、涉及面广的特点,使其系统的汇集起来,就会成为一部非常有价值的地方百科全书。利用先进的现代化信息处理技术,对作为各省市公共图书馆的重要文献信息资源的地方文献报刊进行深层次的开发,把分散于各报刊中的有价值的各种文献信息进行系统的选择、收集、整理、加工,使其实现数字化、网络化,生产与利用以充分满足用户需要已成为当务之急。

2 北京地方文献及其报刊 索引建设源流

2.1 北京地方文献资源现状

首都图书馆是一个综合性的省市级公共图书馆,处在北京这个图书馆、文献情报机构林立的环境之中,在“夹缝中生存”是其主要特点。环境迫使首都图书馆较早的走上了一条以特色馆建设求发展的道路。由于北京地方文献工作在北京市的两个文明建设中的重要地位,北京地方文献资源建设和开发利用一直是首图工作的重点。〔2〕

从 1958 年起，首图建立了专门进行北京地方文献收集、整理与研究的北京地方文献部，该部目前已收集、典藏了北京地方文献 2.9 万种，4.5 万册(件)。其中包括中外文图书、报刊、学术论文、舆图、会议记录、文件汇编、统计图标、谱录、商标、文契、戏剧说明书、乐谱、拓片、形象资料，以及地方文献专用工具书等多种文献类型。这些文献的入藏标准是以内容上具有北京地域特征，同时具有学术价值与信息利用价值为标准，而不考虑出版地、出版时间、文种、文献类型和载体形式等因素，这些文献从自然地理、人文环境、社会生活、文化事业、经济建设等不同角度记录着北京的历史、现状及未来发展趋势。其主要文献分布状况如表 1〔3〕：

主要类型	馆 藏	
	种	册
总 合 计	29132	45055
古籍	330	2306
图书	7193	15733
报刊	443	3245
论文	416	603
照片	3875	3875
拓片	2090	3676
“文革”小报	324	1012
其它	14461	14603

2.2 北京地方文献的社会需求

北京具有悠久的历史 and 丰厚的文化沉积，自古以来就是华北地区的重镇，在政治、经济、军事诸方面举足轻重，早在辽、金时期，就已是北方少数民族国家的都城，在元朝以来，更成为全中国统一的首都。新中国成立以后，作为全国政治、文化中心首都，其在政治、文化诸方面的飞速发展，令世人瞩目。随着北京向国际大都市的进一步发展，无论是北京的党政机关、科研人员、教育机构，还是与北京建设相关的各类企事业单位，以及国内外学术界都将大量产生并获得全面准确的地方信息，这就使得北京地方文献报刊的开发利用有着广泛的社会源泉和需求。

因此，首都图书馆在 1958 年建成的北京地方文献部，其主要任务就是广泛收集、科学整理北京地方文献，并在其基础上加工生产各类文献情报产品，为首都党政部门和科研单位提供服务。在北京地区现有的图书馆系统之中，它是唯一的一家已建立了研究级的藏书体系，采用独具特色服务手段的北京地方文献机构。在全国范围内已成为与广东中山图书馆特藏部、甘肃省馆地方文献部并列的三大地方文献工作机构。为了提高北京地方文献的馆藏质量和文献保障网络，该部从建立初起，便对首都重点院校、科研单位的文献信息机构所收藏的有关北京方面的文献进行了系统的普查，并一一制成卡片，逐步建立起协作网。经过近 40 年的建设，首图北京地方文献部已形成一个独立、完整的采访、专藏、检索与服务体系。

2.3 卡片式北京地方文献报刊索引系统建设

由于在北京地方文献部的专藏中，报刊部分共有 443 种、3245 册。其中大部分为本世纪初至今的各种京版报刊，早期藏品有清代道光二十九年(1849)至宣统元年(1909)的木活字版《京报》(后期改为铅印)，其它的，从各时期政府所办的大型综合性报纸到“文革”时期群众

组织的小报，从几十年一贯制的日报，到专门为亚运会、七运会而编辑出版的临时报刊，从历届政权的政府公报，到本地区的行业报刊、企业报刊和院校报刊，共同构成了北京地方文献报刊系统。

由于北京市社会科学研究对这部分文献的需求较为突出，对其进行深层次的揭示，成为北京地方文献部的基础性工作。1963年起，首都图书馆从首图、北京图书馆、北京大学图书馆、清华大学图书馆等十几个大中型图书馆所藏的清末至1963年的2800余种报刊中，检索出有关北京的索引条目7万余条，编制成大型题录式卡片索引

《北京地方文献报刊索引》，此索引大大方便了科研机构及读者的需要，在北京研究界引起巨大反响。《北京地方文献报刊索引》是一种篇目索引，或称题录索引，主要是记录报纸和期刊中所包含的地方文献篇目，按分类和专题相结合的办法进行编排。每个条目均注明篇名、著者、报刊名称、出版日期和卷期号、内容提要、附注等项。举例如下：

北平研究院与北平图书馆合组拓片展览会之经过/徐炳昶//国立北平研究院院务汇报。—1936, 7(6)。—196“文革”结束后，首图恢复了中断多年的北京地方文献报刊索引工作，每年新生产索引条目1—3千条，并将1964—1977年的报刊回溯完成。到1995年，手工索引系统一直正常发展，年索引量也已达1.5万条以上，总累计量达20余万条〔4〕。这一索引系统的建成，为北京地方文献索引数据库的建设奠定了资源基础。

2.4 北京地方文献报刊索引的利用

除利用这套索引及北京地方文献专藏库服务外，还针对重点需求，编制了多种检索工具，主要有：《北京地方文献报刊索引·地理与名胜古迹部分(清末—1949年)》、《北京地方文献报刊索引·历史部分(清末—1949年)》、《北京地区党史资料报刊索引(1919—1949)》、《北京青年英烈报刊资料索引》等。随着业务工作规模的扩大，首图开展了较大规模的专题文献服务工作，如编制《北京青运史资料索引》、《北京印刷工业志料》、《北京园林绿化志料》、《北海公园历史资料汇编》、《北京出版志料》等大中型课题。自1984年到1994年底，已为研究和编写北京各行业史志、城市规划与建设，文物古迹修复、文化商业街建设、园林建设、影视剧拍摄等方面的用户提供了大量文献资料〔5〕。共编制专题文献汇编36种、435册，图片1646年，3000余万字。由于这种服务方式为有偿服务，工作人员根据合同要求，利用《北京地方文献报刊索引》等检索工具进行检索，而后将索出文献进行编排，最后装订成册。这种不同于常规的服务方式，由于针对用户需要，编排上又利于使用，提供的文献精而全，很受用户欢迎，社会效益和经济效益显著，仅1993年咨询收入就达13万元。〔6〕

2.5 传统模式下北京地方文献报刊索引系统的局限性

随着北京地方文献信息源的增加、信息量的成倍增长，处理难度和劳动强度不断增加。经过数十年的基础建设，北京地方文献报刊索引现已积累起来20万多张卡片，并且每年仍在以2万余张的速度增长，占据了首图紧张的馆舍。特别是改革开放以来，随着北京城市建设与精神文明建设的发展，对北京地方文献的需求日益迫切，要求提供咨询课题服务的大中型课题呈上升趋势。在这些课题的服务过程中，用户对查全率、查准率、时效性的要求使得卡片式报刊索引不便于检索的矛盾显得更为突出，就近几年来完成的《北京文化志料》、《北京房地产志料》、《北京饮食服务业志料》、《北京出版志料》等20余项大中型课题来看，对北京地方文献索引中所揭示文献的需求已达40%，其它的包含于北京地方文献内容的图书或档案资料，课题工作人员近70%以上的工作时间花费在对报刊文献的检索上。可见建立一个高性

能的数据库系统已是迫在眉睫。

这样，以已建成的卡片式《北京地方文献报刊索引》为基础，研制开发相应的计算机数据库系统从 1996 年起成为首图业务自动化建设的首要工程。

3 数据库系统的开发方案及其技术实现

由于北京地方文献报刊索引数据库所收录的文献以反映社会发展状况的报纸和期刊为主，基本上属于一种专题性的题录库，其主要的功能要求是提供文献信息检索服务，数据收录范围比较有限，数据内容比较规范，数据结构也较为稳定。但在标引方面，无论是分类标引还是主题标引方面，都比馆藏书目数据库的标引深度要高，难度也大〔7〕。这些实际情况就决定了该数据库系统的研制目标、软硬件条件及性能指标。

3.1 研制目标

满足北京市各级决策与科研机构对北京地方文献的需要，同时替代原有的手工生产系统，实现每天加工处理录入 100 条以上的，并完成全部数据的回溯工作。定期编制和发行书本式检索刊物，并可根据普遍的社会需求和特定用户的需要，编制各种专题索引，开展计算机检索服务，满足参考咨询工作的需要。该系统要求实用、操作简便，同时便于数据积累和对外交换。

3.2 硬件环境

适合采用 IBM 奔腾系列以上微型计算机，内存推荐 16 兆以上，外存 850 兆以上，显示器为 VGA 彩显，外接针打或激光打印机。可采用单机或网络两种运行方式。

3.3 软件环境

操作系统及汉字环境采用中文版 Windows95，应用软件采用 QuickIMSV4.0 微机通用信息管理系统作为开发平台，进行二次开发。由于该系统是一个基于非关系型数据库设计的具有中国自主知识产权的商品化软件，适合于中国人的使用习惯，能够方便的处理中文：价格合理，便于维护：具有符合国际标准的文献数据交换功能：特别是有遍布于国内政府机关和企事业单位的 5 千余家用户的开发使用经验可供借鉴的优势，使得这个软件成为建立中文文献检索系统的较佳选择。

该软件具有高性能的技术特点，如：面向最终用户，具有多媒体信息管理功能、实现全文检索、运行速度快，可方便的进行数据库定义、维护、任意方式索引、组合条件检索、排序输出等功能。单双字节的完全兼容，使得显示和输出结果更美观，可用于直接印刷排版处理。这些功能是建立和进一步发展北京地方文献报刊索引数据库系统所必须的。

3.4 采用的主要技术标准

3.4.1 文献著录等标准

标准化程度高的系统，将会为资源共享打下良好的基础。因此数据库的著录、编辑排版建库等工作严格遵守以下国家标准、部颁标准：

《GB3468—83 检索期刊编辑总则》
《GB3793—83 检索期刊条目规则》
《GB3860—83 文献主题标引规则》
《GB3469—83 文献类型与文献载体代码》
《GB3792—85 文献著录总则》
《GB3792.3—85 连续出版物著录规则》
《GB3792.2—85 普通图书著录规则》
《中国机读目录格式》

3.4.2 分类与主题标引

该数据库收入的文献暂时使用《北京地方文献非控主题词表》进行标引，用《北京地方文献报刊索引分类表》分类。

(1) 分类表

由于北京地方文献报刊中有关北京的信息涉及社会生活的各个方面，如果直接采用《中国图书分类法》，由于类目的概念泛指，造成大量文章过于集中：类目的数量和详尽程度与北京地方文献的实际状况具有较大差距：为此，在《中国图书分类法》的基础上，根据北京地方的自然与人文环境和各项事业的特点，以及研究者多年使用习惯，编制了该分类表，将类目设置为“自然环境”、“人文环境”、“地方经济”、“地方事业”四大部类、17 大类。

(2) 非控主题词表

由于北京地方文献报刊索引读者范围及其使用文献的广泛性，决定了文献主题标引的专指度不易过高或过低。标引深度过低，会埋没许多有价值的文献和误检率增大：标引深度过高，会使检出文献过于分散，检出率下降。只有对地方文献进行全面考察，科学分析，权衡利弊，并在实践中不断总结经验，才能编制出满足适度标引、质量较高的词表。1996 年开始数据库系统研制，制定了“主题标引规则”。该规则规定，由于北京地方文献的特殊性，决定了人名、地名、团体机构名称、年代范围等专有名词在检索中的特殊作用，而这些词按照一定的规则规范化即可作为主题词使用，成为人名、地名、团体机构名称、年代范围主题词。而现阶段的主题词标引则采取“非控主题词”标引的方式，大部分“非控主题词”与《北京地方文献报刊索引分类表》类目相对应，最后利用 QuickI MSV4.0 系统得出检索词在全数据库中出现频率，以及各检索词的“用”、“代”、“属”、“参”关系进行主题词表的编制。

3.5 主要性能指标的实现

3.5.1 数据库的基本操作

(1) 定义数据库

利用 QuickI MSV4.0 的数据库定义功能，可实现对北京地方文献索引数据库结构的定义修改。依据 3.4 小结中的技术标准，将北京地方文献数据库的字段定义如下表，并相应建立录入工作单、可检索字段的倒排用字段选择表、排版与卡片输出格式文件等。

(2) 管理功能

- 进入系统时，可根据预先设计的用户名和口令进入主菜单，用户分超级用户和普通用户两种，从而实现数据库的保密管理。普通用户不能对该数据库的数据进行增、删、改操作，只能进行检索、查询和输出操作。
- 对数据库进行备份、恢复和重新组织，初始化主文件。
- 业务工作统计，如工作人员工作量统计，显示不同期刊的索出率，各个时期的索引量，

各类文献的比率等。

· 收费管理功能，可对使用该数据库开机费、机时费、检索费和打印费等进行收费统计。

(3)数据生产

录入员可根据预先定义好的数据录入格式及自行设定缺省字段后进行数据的生产工作，对已有的数据可在授权下进行删除或修改，还可对指定条件的数据进行批式修改。如对同一本期刊的论文，某些字段内容是相同的，采取定义隐含值方式，可减少录入工作量。

3.5.2 检索功能

系统可提供文献类型，分类号，题名，个人责任者，团体责任者，出版日期，专栏，个人主题，团体主题，地名主题，年代范围主题，非控主题，母体文献名称等检索。还可进行布尔逻辑检索，并根据需要显示或保存检索结果。还提供中文题名及文摘的单汉字、单词的全文检索。

3.5.3 显示功能

可以以记录号为显示条件，通过表格方式或预定义显示格式来显示。

采用的卡片式和书本式两种格式均可以国家标准局颁发的《检索期刊条目著录规则》所制定的标准输出。举例如下：文献顺序号文献题名/著者：译者姓名//报刊名.一版本.一年，卷(期).一起页～止页，转页.一文摘或附注。图片*表格*照片*参考文献*010240 中国图书馆界早期参与国际图联重要活动的追溯/丘东江//图书馆杂志.—1995，14(5).—54～56.一本文介绍了1925年至1939年中国图书馆界人士参加国际图联大会的情况，其中特别介绍了清华大学图书馆馆长戴志骞、北平图书馆馆长袁同礼等人代表北京的图书馆参与各届会议的情况。

个人主题：戴志骞：袁同礼

团体主题：中华图书馆协会：北京图书馆：

清华大学图书馆：非控主题：图书馆工作：国际图联大会

对于多媒体数据，在显示时，点击“多媒体显示”按钮既可显示。

3.5.4 打印功能

按需要的格式与版面打印输出检索结果、指定范围的记录、数据库内容校对清样，卡片文本文件或直接打印，生成可供排版的文本文件，甚至直接生成激光印刷清样。该功能是通过定义不同的打印格式来实现的。

3.5.5 交换数据

可提供转出、读入符合 ISO-2709 格式标准的数据文件功能，可实现与其它系统的数据交换。

4 数据生产与质量控制

4.1 数据生产工作流程数据

生产的主要工作程序是在限定摘录报刊对象范围的基础上，依据既定采摘标准进行文章选择、内容分析、主题分类标引、填写工作单和录入数据库系统。如图 1：

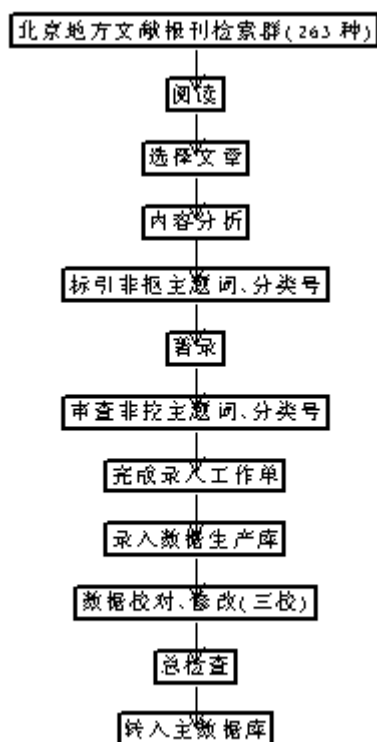


图1 数据生产工作流程

4.2 数据质量控制

质量控制是北京地方文献报刊索引数据库建设的首要问题。具体方法包括：

- (1)根据实际工作要求，制订北京地方文献报刊索引的标引、著录规则，工作单录入规则，以及有关注意事项，并注意解决发展中的问题。
- (2)注意提高所有索引工作人员的专业技术水平，经过编目规则、分类、主题标引规则以及计算机操作技术基础知识的培训并考核合格以后，经试用期满才能正式从事工作。
- (3)严格实行三校工作：标引人员填制工作单 专业录入人员录入 标引人员进行校对 修改 再校对 修改 三校对 总检查。
- (4)设立专门的质量检查员，对工作人员的标引结果进行抽查，在数据库中设置标引、录入人员代码字段，实行岗位责任制。
- (5)利用数据库的检索程序对无题名、母体文献、出版年与卷期等错误记录自动进行检测。
- (6)制定合理的工作量计量标准。数据产量与质量控制是一对矛盾，过分强调产量则会造成质量的下降，经过多次实践最后确定合理的日产量。

5 使用效果

利用该系统实现了定期编制和发行书本式检索刊物，产可根据普遍的社会需求和特定用户的需要，编制各种专题索引，满足参考咨询工作的需要。系统可输出：《北京市情文献索引》月刊(正文、主题、著者、文献号索引)和年度索引(主题、著者、文献号索引)，检索卡片，以及各种专题索引。

该系统从正式运行起，经过1年半的时间，已将1996年以前的22万张卡片全部回溯进

入数据库,现在每月已实现对 3000 余条以上最新地方文献索引的加工处理,大大超过原设计目标。1997 年新出版的北京地方文献索引数据正在及时的输入该系统,可供时效性强的用户使用。利用北京地方文献索引数据库的检索输出功能,已编制《北京市情文献索引(文化事业部分)》(1995-1996)第 1 期,《北京市情文献索引(城市规划部分)》第 1 期,分送市级和局级领导使用,得到好评。现正在积极筹备,拟在条件成熟时,将该数据库通过因特网播送,使更多的用户可通过网络使用,更好的发挥其效益。首图现正在利用该系统的技术成果,积极筹建北京地方文献图书内篇目索引数据库、北京地方文献照片数据库、北京地方文献拓片数据库、北京地方文献全文数据库的建设工作。

参考文献

- 1 杜定友.地方文献的搜集整理与使用.南京图书馆,19572
- 2 常林.建设有特色的公共图书馆——公共图书馆繁荣与发展的新契机.北京图书馆馆刊,1995,(3/4):107-112,1323
- 3 北京市文化局、首都图书馆编.首都图书馆馆史.19944
- 4 吴开华.用计算机开发利用图书馆报章信息资源——谈香港报章资料库的建设.中国图书馆学报.1997,(1):38-465
- 5 北京市文化局编.图书馆与经济建设成果汇编,1996