

● 何建新

我国综合性文献光盘数据库的比较研究

ABSTRACT A comparative study is conducted on the two kinds of domestic comprehensive optical disc data bases: the "data base for Chinese scientific and technical periodicals" and the "data base for Chinese scientific and technical documents". Suggestions about the way to improve the structural quality of optical disc data bases and to raise the functions of its retrieval software are put forward 2 tabs 3 refs

KEY WORDS Document data bases Optical disc data bases Comparative study

CLASS NUMBER G356 1

我国目前发行量较多、应用较广的综合性文献光盘数据库,主要有中国科技信息研究所重庆分所开发研制的《中文科技期刊数据库》(以下简称《中刊库》)和中国科技信息所等单位共同开发研制的《中国科技文献数据库》(以下简称《文献库》)两种。其中《中刊库》为我国第一个中文光盘数据库,是目前国内最大的综合性中文文献光盘数据库,于1992年6月建成^[1];《文献库》则是我国有史以来学科覆盖面最广、文摘量最高、时间跨度最长的光盘文献数据库,是由我国第一家专业数据库公司万方数据公司联合全国近20家专业信息单位,于1995年建成的^[2]。几年

来,笔者作为用户在使用《中刊库》和《文献库》过程中积累了一些经验,也发现了不少问题。现对《中刊库》和《文献库》进行分析和比较研究。

1 基本情况的比较

在国家科委1995年进行的普查中列入《中国数据库大全》的有1038个数据库^[3],其中光盘数据库有几十种,能够形成规模,走向市场,在一定范围提供服务的有10来种,而综合型文献库只有《中刊库》和《文献库》。两库基本情况见表1。

表1 《中刊库》和《文献库》基本情况比较

	《中刊库》	《文献库》
建成时间	1992年6月	1995年
库容量 (截止1995年底)	460兆 170万条记录	750兆 120万条记录
年报道量	26万条记录	27万条记录
光盘数	1张	2张
记录年限	1989~1995	1980~1995

	《中刊库》	《文献库》
文摘率	1994 年起部分记录有文摘	50% 以上记录有文摘
系统软件	用 C 语言自行编写专业检索软件	用 CD S/IS IS 情报检索软件包开发
用户手册	无, 只有 4 页说明书	有, 但与万方其他数据库的手册合在一起, 其中该库内容也只有 4 页
版本	只有单机 DOS 版	单机、网络、DOS、WINDOW S 版
专业子库版	有	1996 年版取消
系统安装与启动	方便容易	繁琐, 每次启动要用系统盘
收录范围	一切自然领域及部分社会科学(经济、文化、教育、图书情报等)领域, 并含港、台核心期刊	国内公开发表的科技论文(不含社会科学)包括期刊、会议录、汇编、专著、科技报告、专利文献、学位论文、译文等
更新周期	一年二次	一年一次
价格	订购费 9000 元/换版费 3500 元	订购费 6800 元/换版费 3800 元

由表 1 可以看出,《中刊库》和《文献库》在规模和容量上旗鼓相当,两库的年增长数又都接近 30 万篇,都是目前国内少有的大型文献数据库。从收录范围来看,《中刊库》侧重于国内出版的全部期刊,是以“全”为主,除自然科学外,还收录了部分社会科学的内容,并包含了港、台期刊,截止 1995 年底共收录 5218 种期刊,体现了其“中文期刊”的内涵。而《文献库》则主要收录科学技术论文,除了期刊以外,还收录专著、报告、专利、学位论文和译文等其它类型的文献,但这些类型的文献含量极少,仅占 0.6%。《文献库》是纯自然科学文献,它强调“精”,同时拓展了收录文献的类型。

从收录文献的年代跨度看,《文献库》年限表面上长于《中刊库》,但《文献库》是分为 19 个子库,并不是每一个库收录的年限都一致,其中有长有短。如光学子库为 1984~1995 年,计算机科学子库为 1985~1995 年,天文学子库为 1986~1995 年,而农业科学子库为 1969~1995 年。同时,《文献库》的大量文献也集中在 1989~1995 年当中,如农业科学子库 1988~1995 年的文献占 96.7%,而前面 19 年的文献仅为 3.3%。

从记录的文摘率看,《中刊库》在 1994 年以前为纯粹的题录式,1994 年开始有 40% 左右的记录有文摘;而《文献库》文摘率在 50% 以上,但各个专业库的文摘含量又不尽相同,有多有少。目前国外同类数据库的文摘率都在 90~95% 以上。

从操作和掌握系统软件的难易程度方面看,《中刊库》和《文献库》都存在不足。《中刊库》的出发点是能使用户在无指导的情况下,独自安装,独立运行,简便、易学易掌握,用户可无师自通,因此其随盘附件只有一份 4 页的说明。但事情往往不是这么简单,因为用户的层次参差不齐,所使用的硬件设备和软件环境不同,以及用户的各种心理因素等等,会出现各种各样的问题和困难,短短 4 页说明书是难以解决的。在这方面,《文献库》虽然随盘附有一本 60 页的用户手册,然而其出发点似乎并不是为新用户而准备的,或者说不是为各个数据库专门准备的。它是把万方数据公司现有的 8 个光盘数据库产品合在一起编写,其中专门介绍《文献库》的也只有寥寥 4 页,而且没有一个实例,用户使用起来要前后来回翻阅查找,如同捉迷藏一样,很不方便。

另外,从光盘数据库更新周期上看,《文

献库》和《中刊库》都与国外同类产品存在差距。目前《文献库》的更新周期为一年;《中刊库》为半年。周期太长的原因可能和制作成本及发行量相差太多有关。国外光盘数据库产品大都在一季度甚至一月更新一次。

2 实际使用情况的比较

光盘数据库研制开发的目的是为用户服务的,因此其系统的功能如何,检索的速度快慢,操作是否灵活方便,用户界面是否友好,数据库结构是否合理等等,只有通过用户的实际使用以后才能给出科学的、公正的评价。下面把《中刊库》和《文献库》的系统功能及库结构加以比较。见表 2。

表 2 《中刊库》和《文献库》系统功能及库结构比较

	《中刊库》	《文献库》
用户界面	屏幕底有选择检索项目提示,最多 2 步即可进入检索式输入菜单,检出后按回车可进入下一轮检索	有下拉式菜单,但操作较繁琐;数据子库用拼音字头,难以记忆;至少要 8 步才能进入检索式输入,进行下一轮检索要退出菜单,重新进入
检索速度 (386 兼容机 4M 内存)	输入主题词“水稻”,6 秒检出 8572 条;输入主题词“水稻?”,6 秒检出 8927 条	输入关键词“水稻”,58 秒检出 7319 条;输入关键词“水稻 \$”,5 分钟(300)秒仍未检出
自由检索	无	有
逻辑检索	可进行复合检索,功能符与“*”、“+”、“非”、“-”	可通过菜单进行检索项目选择组合,也可直接写出布尔逻辑检索式,功能符与“*”、“+”、“非”、“-”
限制检索	有,可将检索词限制在屏幕底 4 个检索项(字段)其中之一,年代限定有窗口提示	有,可通过专项检索限制检索目(字段),还可在自由检索中直接写出检索式,功能符为“/(字段标识)”
截词检索	只有右截断(前方一致),功能符“?”	只有右截断(前方一致),功能符“\$”
相邻检索	无	无
精确检索	无词表	有检索词典库,但不规范
全文检索	篇名字段可进行全文检索	不可以
检索结果显示	较方便,按 F1 键为全屏显示,按 F2 键为浏览(只显示分类号和篇名),但要经过 2 次选择检索式,并且不可以修改显示输出格式	较复杂,要退回菜单再进入显示菜单。自由与专项检索显示屏不一致。有三种可选显示格式,还可修改显示格式,但非经专门培训难以掌握
检索结果的打印	可以全屏打印,浏览打印,亦可按标记选择打印,不能进行字段和打印输出格式的选择	可以全屏打印,不能进行标记选择打印,浏览打印只能打印当前记录。打印格式有三种可选:1 全记录;2 篇名和作者;3 文献题录。还可自定义打印格式,但同样难以掌握

	《中刊库》	《文献库》
检索结果的套录	可以根据显示/打印菜单提示将检索结果全部(前 500 条记录)或按标记套录到指定的磁盘上	不能套录, 可以将检索结果保存在磁盘上的 *. SAV 文件中(* 为自定义), 但这只是记录在数据库的指针, 是一个不可读的文件, 只能备以后在本系统打印输出用
检索策略的保留	不能存储到磁盘上, 而且屏幕上只容纳 20 条检索式, 超出 20 条将不予接受	可执行以前的检索式, 不能存储到磁盘上。可修改检索式, 同一检索式可访问其它子库, 但手册未注明
数据库字段	8 个	21 个
可检索字段	5 个	11 个
联机帮助	只是 4 页说明书的文本文件	无, 只有屏幕底提示

从表 2 可以看出, 无论从系统功能还是数据库结构方面, 《文献库》和《中刊库》都存在一定的缺憾和不足。总的看来, 《文献库》的库结构似乎较为合理一些, 其字段的设置多达 21 个, 接近国外数据库产品, 但可检索字段只有 11 个, 这又显得大为逊色。主题(关键词)标引的不规范是两库共同存在的问题, 限于篇幅, 不予赘述。

从系统检索的功能看, 《中刊库》表现为出色, 其用户界面较为友好, 操作方便快捷, 检索速度较快, 在 170 万条记录中, 检中 200 条的记录几乎按键即可检出。而《文献库》进入子库在 26 万条记录的情况下, 同样 200 多条检中记录约 10 秒才能完成。另外, 笔者曾试用截词检索功能比较两库的性能, 《中刊库》表现出较高的检索速度, 一般 6~8 秒即可检出, 而《文献库》则需时几分钟至几十分钟, 甚至经常出现不能检出(死机)情况, 如利用检索式“S5\$(/610)”、“水稻\$”等都会造成死机。《文献库》除检索速度缓慢外, 还表现出检索过程机构复杂, 不够灵活, 用户只能被动地按系统提示操作。

从检索结果的显示与打印以及套录功能来看, 也是《中刊库》略胜一筹, 显示结果方便, 直接按键显示, 无须退回菜单。在浏览显示结果时可随时进行标记, 打印结果可按标记打印, 并可将检出结果套录在指定磁盘上,

这些功能对于使用《文献库》的用户是无法做到的。目前的数据库不具备套录功能可以说是过于保守或者说是落后的。

另外, 两库共同存在的问题是, 不能进行检索策略的保留。检索策略的保留非常重要, 可以再利用, 特别是较为复杂的检索式, 不但省时省力, 还可保证检索策略的一致性, 从而提高检索效率。

联机帮助的详尽和可读性强也是不可或缺的。国外同类产品对联机帮助相当注重, 用户(特别是新用户)往往通过联机帮助即可解决检索中遇到的所有困难和问题, 而无须翻看厚厚的用户手册。这一点无论是《中刊库》还是《文献库》都未能达到。

3 结论和建议

《中刊库》与《文献库》都是我国大型综合性文献数据库, 基本上代表了我国数据库, 特别是光盘数据库的发展水平。提高和发展我国光盘数据库业是每个用户与研制开发者的共同愿望。

综合两库的基本情况和实际使用情况, 笔者认为, 《文献库》的数据库结构是发展方向, 但还应适当增加可检索字段与限定字段。ISBN 和 ISSN 字段等应设置为可检索字段, 以增加检索入口和扩大数据库使用范围。分

类标识和出版日期应设置为限定字段,用户在检索大于某一类号或出版年代时可使用功能符“>”,小于可使用“<”,在某段之间使用“-”等等。《文献库》结构的缺陷是其采用各专业子库“各自为政”的形式,这自然增加了用户选库的负担,而且其库名又用汉语拼音首字母,很不直观,也不便记忆。其实这完全可能通过分类号限制进行检索。《文献库》毕竟是由近 20 家单位联合建立的,因此存在库结构不尽相同的问题,如分类标识就多达 4 种分类号(法),造成了全库的不统一不规范,给用户带来麻烦。

再好的库结构如果没有优良的软件支持也是收不到好的使用效果的。由于软件跟不上,许多好的数据库半途而废,前功尽弃的已有前车之鉴。《文献库》可以明显的反映出 CDS/ISIS 情报检索软件的一些弊端,如格式语言的复杂,有些命令(操作)只能执行一次,重复进行则必须退回菜单再从头一步一步来,缺乏容错功能和安全保护功能,如对于误操作 ESC 键时,常常会造成输入的东西全部丢失,甚至会退出系统或造成死机,而在检索和打印等过程中又无法中止运行。因此,研制开发具有中国特色的汉字检索软件势在必行。在此,笔者认为《中刊库》的软件值得提倡和发扬,特别是其检索速度及篇名全文检索功能尤其受到用户的青睐。

文献数据库的发展方向是光盘化、网络化、全文型、实用型和多媒体。优胜劣汰是必然的,竞争也是激烈的。因此,两库都要发挥各自的优势和专长,向特色方向发展。

《中刊库》应在尽量收全整个世界范围内的中文期刊文献上作出努力,《文献库》则应

在收全各种类型的文献上再下苦功,不能满足于只是象征性的收录。

《中刊库》的库结构应作适当调整和增加,起码应增加作者单位(地址)字段才不失为一个好的数据库结构。《文献库》应使检索速度和复杂的检索过程彻底改观。

光盘的制作也可采取多种形式的版本。可按专业学科分,如《中刊库》可分为科技版和社科版;也可按年代分,如《文献库》可把 19 个子库合并为一个库,以 5 年为 1 卷等等,以适应不同用户的需求。

两库都应力求缩短更新周期,增强文献信息的时效性。这也是衡量和评价光盘数据库质量的重要标准之一。

要及时搜集用户的反馈意见和要求。用户对数据库的意见和要求往往是最切合实际的,甚至有时是研制者所想象不到的。把用户的建议和要求进行综合研究,有的放矢,不断改进,不断完善提高,才能赢得用户,占领市场,才能有旺盛的生命力。

参考文献

- 1 陈源曙等 我国第一个中文 CD-ROM 光盘库的研制与开发 情报科学, 1992(5): 1~ 6
- 2 温红彦 “万方”为信息高速公路备“快车” 人民日报, 1996-05-03(5)
- 3 陈宝通 我国数据库资源现状及“九五”期间对策建议 中国信息导报, 1996-02-22~ 2

何建新 南京农业大学图书馆 通讯地址: 南京市, 邮编 210014。

(来稿时间: 1996 8 19。编发者: 李万健)