

几种光盘数据库的比较研究

吴丹 邹瑾 黄睿 魏凤萍

(华中师范大学信息管理系 武昌 430079)

[摘要] 从基本情况、收录范围、检索功能、输出形式、用户界面、网上服务等比较分析我国三大光盘数据库,并对其存在的缺陷提出了改进建议。

[关键词] 光盘数据库 CD-ROM 比较研究

[分类号] G255

A Comparative Study of Three CD-ROM Databases

Wu Dan Zou Jin Huang Rui Wei Fengping

(Department of Information Management, Central China Normal University, Wuchang)

[Abstract] Analyzes the comparison and research of three CD-ROM databases in respects of content including range, search function, output form, system requirement and online service. Meanwhile, the main problems and objective evaluation are given. Suggestions are put forward for users and manufacturers as well.

[Keywords] CD-ROM databases CD-ROM comparative study

光盘数据库检索以其信息获取量大、提取方便快捷、检索可靠性好,投资成本低廉等特点深受用户欢迎,广泛地运用于图书情报部门。

从1983年第一张光盘问世到现在短短十几年内,光盘数据库在品种、数量、质量以及检索性能等方面都发展得极为迅速,而且内容正向多学科延伸,涉及自然科学、社会科学、工程技术、生命科学、计算机、法律、医学等各个领域。据Gale Directory of Databases统计,1996年全世界共生产商用CD-ROM数据库2971种。我国目前生产的各种CD-ROM数据库也达1000多种,数量呈直线上升趋势。

随着越来越多的光盘数据库投放市场,用户的选择余地也相应地越来越大,因此,普遍面临着一个对光盘数据库的比较和科学选择的问题。要在众多良莠不齐的CD-ROM数据库中找到最适合自己的一种,用户除了考虑其内容范围、数据来源、更新频率、标引质量以及出版商的权威性等因素外,更为关注的是文献标引深度、检索途径、输出形式、检全

率、检准率以及用户界面等更深层次的因素。本文将就上述问题,对目前国内三大光盘数据库《中国学术期刊(光盘版)》、《中文科技期刊数据库》、《人大复印报刊资料系列光盘》进行分析和比较研究,以供用户在选择、利用CD-ROM数据库时参考。

1 三大光盘数据库的基本情况

《中国学术期刊(光盘版)》(以下简称《CAJ-CD》),全称为《Chinese Academic Journal(CD)》,是由国家新闻出版署批准、清华大学研制的我国第一个大规模全文现刊集成化学术电子期刊,是一套具有文献导读、评价、检索、咨询等功能的实用化电子文献系统。《CAJ-CD》自问世以来,已在全国各地的图书馆和科研机构中建立了700多个检索站,是目前我国图书馆电子阅览室常备的检索光盘之一。在使用过程中,《CAJ-CD》以其完善的功能和突出的优点,受到了广大用户的

收稿日期:2000-03-16

欢迎,并在全国首届“国家电子出版物奖”评选活动中荣获“国家电子出版物荣誉奖”第一名。

《中文科技期刊数据库》(以下简称《中刊库》)是由中国科技信息所重庆分所重庆维普资讯公司于1989年开发的一种光盘数据库检索系统。该库是我国图书情报界自建的一个公开发行的光盘数据库。也是国家科委计算机文献检索系统“八·五”重点项目。自1989年以来,《中刊库》相继获得国家科委科技进步二等奖和国家科技进步三等奖,是目前国内容量最大的综合性科技文献数据库之一。

《人大复印报刊资料系列光盘》(以下简称《系列光盘》)是由中国人民大学书报资料中心和北京新北成实业开发公司联合研制推出的信息含量最大、检索功能最全、最具权威性的大型索引型电子出版物。该《系列光盘》分为“复印报刊资料索引”、“复印报刊资料专题目录索引”和“人大复印报刊资料全文”三种,为科研人员从相关报刊中获取专业资料提供了方便,同时也是图书情报部门编制题录、文摘、索引的重要工具。

上述三种光盘数据库都是我国CD-ROM数据库中的精品,也是国内各图书情报部门常备常用的检索工具。为了提高用户文献信息检索的效率,本文试对上述三种光盘数据库进行比较分析,以就教于同行。

2 三种光盘数据库的比较

2.1 数据库的类型、内容范围、数据来源及更新频率比较

近10年来,数据库发展最引人注目的变化就是全文数据库的迅速崛起。全文数据库具有其它书目数据库所无法比拟的优点,其检索结果直接、详尽,使用简便、快捷,特别适合最终用户使用。从这个角度来看,《CAJ-CD》较《中刊库》和《系列光盘》更能满足用户准确快速获取文献的需要。尽管《系列光盘》目前也推出了全文型,但毕竟还不够成熟,有待进一步完善。从另一方面来看,全文数据库和书目数据库对数据质量的影响也不同。《中刊库》和《系列光盘》是二次文献数据库,因而其中的关键词、分类号一般不依赖于原文提供,而需经专业人员重新标引,所以数据质量主要受专业人员的业务水平和工作态度影响。《CAJ-CD》是一次文献数据库,数据完全依赖于原文,因此数据质量受原文学术水平的影响较大,该数据库的主要不足是录入文章的著录项目不

全。

从收录范围和数据来源方面看,《CAJ-CD》是综合性数据库,而《中刊库》则侧重于自然科学,《系列光盘》侧重于社会科学。正是由于《中刊库》和《系列光盘》各有侧重,而《CAJ-CD》虽具综合性,但主要收录核心期刊和有专业特色的期刊,因而若单从自然科学期刊或社会科学期刊的数量上比较,《CAJ-CD》则远不及《中刊库》和《系列光盘》。由此可见,《CAJ-CD》注重的是“精”,而《中刊库》和《系列光盘》注重的是“全”。另外,《中刊库》从1989年问世至今已有10年,《系列光盘》收录文献的时间跨度为1978年至今,而《CAJ-CD》正式出版不过3年时间。因此,《CAJ-CD》和《系列光盘》在收录期刊种类和收录年限上占有绝对优势。《CAJ-CD》虽可即检即得,但需分8个专辑分别检索,如理工A、B、C也分辑检索,为用户大范围了解信息带来了不便。

数据库的数据更新速度是衡量一个数据库好坏的重要标志。《CAJ-CD》与印刷版期刊几乎同步出版,即一个月更新一次,《中刊库》和《系列光盘》目前是季度出版,较之前者时效性显然要差一些。

2.2 数据库检索功能比较

表1 数据库检索功能比较

光盘名称	《CAJ-CD》	《中刊库》	《系列光盘》
检索途径	关键词、分类、作者、机构、篇名、中英文文摘、引文、基金、自由词、蕴含、关联、连续、整刊13个检索点	主题词、分类号、著者、篇名、刊名、复合检索6个检索点	“索引盘”:标题、作者、分类号、分类名、年份、原文出处、出版地、期号、任意词、复合检索10个检索点 “全文盘”:原文出处、原刊地名、分类号、分类名、复印期号、作者、任意词、复合检索8个检索点
检索方式设置	先输入检索词,再确定检索途径,该检索词可一次输入多途径	先选中检索途径,再确定检索词,一次只能进行一项检索	先选中检索途径,再确定检索词,一次只能进行一项检索
检索速度(同一计算机上)	输入关键词“情报”2秒检出	输入主题词“情报”10秒检出	输入任意词“情报”5秒检出
逻辑检索	可运用“布尔逻辑运算符”进行组配与检索	可运用“布尔逻辑运算符”进行组配与检索	可运用“布尔逻辑运算符”进行组配与检索
截词检索	前方一致截词,后方一致截词	前方一致截词,后方一致截词	前方一致截词
限制检索	在选盘时已限制内容和年代、期数	检索年代限制	检索内容、年代限制
联库检索	有	不需要	无

数据库检索系统功能的强弱,在很大程度上取决于作为文献检索途径的信息检索入口的多少。从表1可看出,《CAJ-CD》在三种数据库中检索途径最多。众所周知,检索点不同其作用也不同,检索点越多,检索越方便,速度也越快。但是,尽管《CAJ-CD》的检索点多达13个,其中对检索效果影响最大的著录项目,如关键词、分类号、摘要等的遗漏或缺少情况也屡见不鲜。另外值得一提的是,《系列光盘》“全文盘”

竟然没有篇名这一重要检索途径,实在令人遗憾。在检索方式的设置上,《CAJ-CD》明显比《中刊库》和《系列光盘》要灵活得多,其一词多途径的检索方式省去了用户重复输入检索词的不必要操作。检索速度也是衡量一个光盘数据库质量的重要因素,上述三种数据库的检索速度都较快,几乎按键即可检出,差别不算太大,相比而言,《CAJ-CD》最快,《中刊库》最慢。对光盘数据库而言,检索方式的灵活多样性应该是影响用户检全率与检准率的关键因素,也是决定数据库性能的深层次因素。目前的数据库都比较注重提供完整的逻辑检索功能,如布尔运算符、截词运算符等,此外还有位置运算符、限制检索等。利用这些运算符进行组配检索与模糊检索,将大大提高文献的检索效率。在这方面,三种光盘数据库都具备上述功能,只是《系列光盘》的截词功能不够完整,完整的截词功能应是既能进行词首截,又能进行词尾截。另外,提供联库检索功能也非常重要,因为随着文献时间的跨度增大,盘片也就随之增多,如果没有此项功能,当用户要查找时间跨度很大的文献时,必须一片一片更换盘片,既麻烦又浪费时间。有了此功能就可对光盘库或光盘库中的光盘进行一次检索,不需退出重新进入,从而可大大节省检索时间。《CAJ-CD》具有较好的联库检索功能,可在一张光盘中选择路径进入另一张,而《系列光盘》则必须关闭一张再进入另一张,当然,《中刊库》因其本身的光盘数就是一张,自然不存在此问题。

2.3 数据库检索结果输出功能比较

输出格式是否提供多种方式选择、输出形式是否灵活多样、是否保留检索策略等问题,也是光盘数据库性能的重要方面。从表2可以看出,在输出格式方面,三种数据库均提供有多种选择方案,可供用户对所检结果按自己的要求有选择性地打印拷贝或显示。但能提供套录功能的却只有《中刊

表2 三种数据库检索结果输出功能比较

光盘名称	《CAJ-CD》	《中刊库》	《系列光盘》
检索策略处理	保留部分检索策略	不可保留检索策略	不可保留检索策略
文献条目输出	摘要显示,原文显示	既提供浏览显示,也提供全屏显示	屏幕浏览和摘录
检索结果输出	打印,拷贝记录,但不可以拷贝全文	拷贝、打印、套录、建库	“索引库”不可拷贝、打印,“索引汇总库”和“全文库”可拷贝、打印、选择记录条目输出
输出格式	与期刊登载的格式完全相同	分类号、著者、篇名、刊名、主题词、信息出处、文摘	“索引库”为文摘号、原文出处、分类号、标题、作者、原文出处、年份、期号、页码 “全文库”为文献号、原文出处、原刊期号、原刊页号、分类号、分类名、复刊期号、标题、作者、作者简介、内容提要、关键词、正文、责任编辑

库》。从这一点来说,《中刊库》要略胜一筹。并且《中刊库》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。三种数据库存在的共同问题是不能保留检索策略或者只能保留部分检索策略。事实上,检索策略的保留是非常重要的,特别是对于较为复杂的检索式,若能保留下来再利用,不但节省时间,还可以保证检索策略前后的一致性。检索策略的保留问题不仅仅是上述三种光盘的缺陷,这也是国产光盘的普遍不足之处,而发达国家,如美国的《BA》(《生物学文摘》)、《ERI》(《教育文摘》)、《PSYCLIT》(《心理学文摘》)等光盘既可将检索策略进行屏幕保留,也可进行磁盘保留,有助于用户随时修改或开展定题服务。另外,《CAJ-CD》的输出形式与原文完全相同,缺乏自身的规范性和统一性。

2.4 数据库用户界面与管理服务功能比较

用户界面和系统管理、服务功能虽然对检索结果不会造成直接影响,但会在某种程度上影响光盘数据库的使用频率及在用户心目中的印象。用户的满意程度是对光盘数据库质量的最大认可,因此,生产厂家应把用户摆在第一位,以人为本,为人服务。由表3可见,上述三种数据库均以 Windows

表3 数据库用户界面与管理服务功能比较

光盘名称	《CAJ-CD》	《中刊库》	《系列光盘》
操作平台	Windows	Windows	Windows
系统管理与服务	用户帐户维护、费用查询、数据备份、密码修改、期刊管理与订购、入编期刊简介、光盘记到、英汉双向字典等	修改馆藏号、主题分类表、同义词表、期刊目录等	无
情报服务	定题服务、项目背景分析、引文检索、科研产出能力评价、学术期刊评价	无	无

为操作平台,使用起来灵活简便,尤其是《中刊库》启动后直接进入系统,无繁琐的操作程序,而《CAJ-CD》则不然,其路径设置和盘片的选择须经专业人员指点,且容错功能不强,稍有不慎便容易死机,无法正常使用。但从另一个角度来看,《CAJ-CD》具有强大的系统管理与服务功能,特别是还有五大情报服务功能,这些都是该系统设计的独到之处,是《中刊库》与《系列光盘》所无法比拟的。

2.5 网上服务情况比较

Internet的诞生使全球进入了网络化信息时代,上网查询信息不仅方便、迅速,而且实现了全方位的信息资源共享。我国在建设信息高速公路的过程中,将数据库也推上了因特网。数据库上网不仅为用户进行信息检索开辟了新的领域,同时也使各种信息得到了更广范围的传播。《CAJ-CD》于1998年在中国教育科研网(CERNET)上开通(<http://www.cajcd.edu.cn>),用户可通过关键词、著者、篇名、全文、中文摘要等途径免费查询,返回结果为篇名、光盘号、文献号。另外,

用户还可在中国学术期刊网(<http://www.chinajournal.net.cn>)上检索,检索方式和结果与上述相同。可见,《CAJ-CD》在网上的检索范围和方式都比原来扩大了,用户不仅可根据自己课题的需要对其8个专辑有选择性地检索,一次可选一个或多个,还可以通过不同的网站进行检索。《中刊库》上网时间较《CAJ-CD》早一年,且上的是中国信息网(Chinainfo),其网址为:<http://www.chinainfo.cn.net>,用户上网经过注册后即可使用,检索方式基本上与光盘数据库相同。《人大复印报刊资料》目前只有印刷版、电子版、缩微版三种,电子版的《系列光盘》只支持在局域网(NOVELL、NT)上运行(10个终端),还不能在因特网上大范围传播。

3 三种光盘数据库的改进建议

3.1 存在的问题及改进建议

3.1.1 《CAJ-CD》《CAJ-CD》作为我国连续出版物的大规模集成化、多功能全文检索数据库系统,它将学术期刊原文和计算机检索技术结合起来,从而达到“即检即得”、快速查找文献的目的。该数据库除分类号由清华大学光盘公司统一标引外,其它各项检索入口的数据均直接按期刊论文本身提供的著录项目和内容编排,这样就容易造成著录数据项有遗漏的情况,同时也影响了数据标引的质量。为改变这种情况,笔者建议光盘制作者应本着严谨务实的态度对每篇文章的著录项目即使不必详尽地一一标引,至少也要增加审核校对环节。另外,作为全文数据库,要保证屏幕输出及打印格式与期刊原文一致,但入编期刊编排格式不统一、标识项目不完全,导致《CAJ-CD》的输出缺乏规范性。笔者认为屏幕输出大可不必强求与原文完全一致,只要与原文有关的信息齐全便可,但作为数据库,其输出应该讲究自身的规范性。除此之外,《CAJ-CD》还应增强容错功能,简化入库的复杂程序,这方面可借鉴《中刊库》和《系列光盘》。

3.1.2 《中刊库》《中刊库》虽然建库较早,但在制作上却不及《CAJ-CD》和《系列光盘》精致。该数据库期刊收集不够完整,论文数据遗漏较多,且输入信息也有遗漏或出错现象。另外,《中刊库》的分类号单一,只按一篇文献的主要内容所在学科性质给一个类号。但是随着现代科学的发展,目前许多新兴学科层出不穷,交叉学科、边缘学科也大量涌现,许多文献具有多重属性,《中刊库》这种单一分类的做法势必影响

检索的检全率和检准率。笔者认为,强化附加分类标引功能,将类目进一步细分是《中刊库》亟待解决的问题。由于《中刊库》为题录型数据库,比起全文型数据库(如《CAJ-CD》)并没有绝对优势,若能在制作上下功夫,形成自身特点,进一步提高检全率,便会在某一方面占有相对优势。

3.1.3 《系列光盘》《系列光盘》是《人大复印报刊资料》的派生产品,该刊物的权威性使得《系列光盘》成为目前文科类的权威性检索工具,同时它也为《人大复印报刊资料》注入了信息技术的活力。但可能由于该光盘分的系列太多,以至全而不精。仅就“索引盘”而言,就存在不少问题:首先是索引号不规范问题。索引号的意义未在帮助信息中予以说明,这就给非专业人士的检索带来了不便。另外索引号仅能指示期号和位置,不能指示年度,且打印不出来。其次,“索引盘”未将复印文章和非复印文章加以区分,用户必须找到《复印报刊资料》才知是否为复印文章,建议增设这样一个字段以作区分。再次,“索引盘”的帮助信息不够明确。帮助信息是为帮助用户正确使用光盘数据库而设置的提示信息,若帮助信息不明,就会影响用户的使用。建议明确给出在线帮助信息,以方便非专业用户使用。

(本文在写作中得到了华中师范大学信息管理系统张帆教授的指导,在此仅表感谢)。

参考文献:

- 1 韩改样.我国两大期刊文献数据库比较研究.图书情报工作,1996(6):45~46,8
- 2 丁中桃,雷菊霞.中美光盘数据库检索软件功能比较.大学图书馆学报,1999(4):31~35
- 3 郭依群.文献型CD-ROM数据库的评价与选择.现代图书情报技术,1997(6):17~21
- 4 柳丹枫.对《中国学术期刊(光盘版)》的认识和利用.河北科技图苑,1999(2):56~57
- 5 韩改样.学术期刊出版规范化与其光盘版的检索效率.情报学报,1998(2):121~125
- 6 杜伟,王君.“维普”中文科技期刊数据库的检索.云南图书馆季刊,1999(1):47~49
- 7 陈玉凤.浅谈《中文科技期刊数据库》存在的问题及对策.文献信息学刊,1996(3):42~44
- 8 袁毅,郭秋英.《报刊资料索引》光盘数据库的特色及使用.图书馆杂志,1998(1):10~11

【作者简介】吴丹,女,1978年生,发表论文7篇。邹瑾,女,1979年生,发表论文2篇。黄睿,女,1978年生,发表论文数篇。魏凤萍,女,1979年生,发表论文1篇。