

●卜书庆 宋安莉

《中图法》电子版的原理、结构及应用

摘要 《中图法》电子版的研制具有重要意义。其原理是将《中图法》第四版印刷本数据机读化,包括《中图法》MARC格式研制及其数据库的建立,以及功能软件的研制。《中图法》电子版具有完善的体系结构,功能完备。参考文献3。

关键词 《中图法》电子版 研制原理 体系结构 应用功能 使用环境

分类号 G254.1

ABSTRACT In this paper, the authors introduce the characteristics, MARC format design, database development and software development of the electronic version of the fourth edition of *Chinese Library Classification*. 3 refs.

KEY WORDS *Chinese Library Classification* electronic version. Development principles. Structure. Application functions. Environments.

CLASS NUMBER G254.1

《中图法》电子版是由《中图法》编委会和北京丹诚软件有限责任公司共同开发研制而成的一部利用计算机技术实现多功能检索和多维显示的分类法,已由北京图书馆出版社出版发行。

1 《中图法》电子版研制的意义

《中图法》的电子化是历史发展的必然。

首先,从网络环境看,网络资源的膨胀和增殖,直接影响到《中图法》在网络环境下的潜在使用功能,网络环境下资源组织的实践给《中图法》的理论发展和技术进步带来了巨大的挑战和机遇。国内图书馆自动化、网络化得到快速发展,作为机读化的分类数据应和书目数据,规范数据一样具有标准化的交换格式,以便数据的利用和开发。

其次,从《中图法》更新版本、维护管理来看,由于缺乏计算机管理,无法控制类目概念重列、类号类级排版错误,不能建立和控制健全完善的类目参照关系等,分类法的编制质量受到直接影响。从《中图法》各系列版本同步更新发展来看,由于缺乏行之有效的电子化产品和机读化的管理手段,很难达到同步一致的修订。

第三,从计算机编目工作来看,用计算机检索分类法可以帮助分类人员自动标引和辅助标引,提高计算机编目效率和编目质量。另外,由于缺乏《中图法》机读数据及其与书目数据的联机管理控制,无论是《中图法》数据还是书目数据,都不能实现实时修

改,每一版更新都对书目数据造成负面影响,最终导致影响读者分类检索。

第四,从计算机检索环境来看,由于缺乏机读化的分类数据,所以读者很难实现分类检索。

第五,从世界分类法发展看,在国外,走在世界分类法计算机化前列的美国《杜威十进分类法》(DDC)在20世纪80年代已开始研制计算机管理系统,1992年推出DOS环境下的DDC电子版的第一版,1996年推出Windows环境下的DDC视窗版,并通过与《美国国会图书馆标题表》(LCSH)的连接,已具备了分类主题一体化的检索功能,是比较完善的电子版。

《中图法》电子版研制出版有其深远意义和重大影响。它标志着《中图法》已步入文献信息分类电子化时代,对《中图法》发展起到里程碑的作用。而且它也是我国第一部综合性电子分类法,将推动文献信息组织和应用的电子化。它的应用范围非常广泛,可适用于任何图书馆、档案馆、情报所、书店、电子网站的各种类型、各种载体形式的文献、信息的分类。它既适用于传统文献的分类标引及快速查找类目,又适用于电子信息的有序组织,为实现机助标引和智能自动标引奠定了基础。它既适用于文献信息脱机标引,又适用于文献信息联机网络标引。

2 《中图法》电子版研制原理

《中图法》电子版的研制原理,是将《中图法》第

四版印刷本的全部数据机读化,即将其印刷排版数据转换为分类法 MARC 格式数据存储在数据库中,然后通过光盘介质由软件提供检索、显示等功能。软件功能的基础是建立在数据库中的 MARC 数据记录上的。因此,电子版的研制原理主要由三部分构成:一是《中图法》MARC 格式的研制;二是《中图法》MARC 格式数据库的建立;三是功能软件的研制。电子版的研制过程是数据编制和软件研制相互磨合的过程。它经历了制作数据—编写程序—校验数据—修改程序—修改数据—再修改程序这样一个相互影响相互完善的过程。

2.1 《中图法》MARC 格式的研制

20 世纪 90 年代中期“USMARC Classification Format”在美国诞生,并依据该格式生产了《杜威十进分类法(DDC)》的机读数据。1995 年 IFLA 成立了 UNIMARC Classification Format 研究小组,开始进行研制,至 2000 年 1 月 25 日才发布了“Concise UNIMARC Classification Format”。

《中图法》编委会 1995 年开始对分类法机读数据格式进行调研,1999 年根据“USMARC Classification Format”的设计思想和 UNIMARC 规范格式(UNIMARC Authorities Format)的结构特点,结合《中图法》类目体系和类号结构特点草拟出“《中图法》机读数据格式”即“CNMARC Classification Format”,因为它先于“UNIMARC Classification Format”的发布,所以它虽然与“UNIMARC Classification Format”设计思想和原则吻合,但还有待完善。

首先,《中图法》的 CNMARC 格式的设计须与 CNMARC 的书目记录格式、规范记录格式、馆藏记录格式相匹配。因此,它也分成记录头标区、目次区、数据区三部分,同样遵循“文献目录信息交换用磁带格式”标准(CB29-1-82,ISO2709)。

其次,分析类目结构和类目间的关系,用机读符号对它们详细描述。类目是分类法机读数据的款目记录,它包括类号、类名、类级、注释、参照根查等项目。类号、类名是类目的主体,是被记录和描述的对象,也是必备项目,需通过字段和子字段来描述。类号类型如主类号、通用复分号、专类复分号,以及类目的性质如使用类目、交替类目、停用类目、指示性类目等分别构成记录类型,通过头标区定义记录类型代码。类号的编辑制度如层累制、八分法、借号法、顺序制等是通过类目字段的指示符来描述的;除了指示类目以外,类级是类目概念逻辑限定的重要

手段,在印刷版中只需通过类名排版缩位则可表达,但在机读格式中除了用指示符描述类号编辑制度外,需要通过子字段来补充描述。注释是类目的必要补充信息,在《中图法》中主要有三种类型:含义注释、使用注释、沿革注释。不同注释类型在类目使用中起着不同的作用,因此,需要用不同的字段表述。与其他类发生关系的注释信息,如仿分复分注释、使用例举注释、参见注释等又采用了子字段和控制字段进一步描述。参照根查是用来建立类目之间交替关系、属分关系和相关关系的,分别通过不同的字段来描述。为了表达清楚类目之间不同关系,还采用了子字段控制。类目与主题词之间的关系、分类法之间的关系也是类目的补充信息,也需要不同的字段来描述。一条类目的记录由以下几部分构成:

(1) 记录头标区,包括记录长度、记录状态、记录类型代码、指示符长度、子字段标识符长度、数据地址、字段长度等 23 位固定长度;

(2) 目次区,包括 001 记录控制号字段、005 记录处理时间字段、152 分类表与版本字段等;

(3) 1--编码信息块,包括 100 一般处理数据字段,23 位固定长度;

(4) 2--类目标识块,包括 260 类目字段,有两位指示符,类号用 \$a \ \$b \ \$c,类名用 \$m、类级用 \$h 来描述, \$z 为 8 个通用复分表和 60 个专类复分表的控制标识符;

(5) 3--类目注释块,包括 330 涵义注释、331 使用注释、332 沿革注释,通过各字段的指示符表达了注释类型,如宜入注释、仿分复分注释、参见注释等,也表达了类号变更情况,如新增类号、停用类号、占用类号等;

(6) 4--无效类号根查块,460 无效类号根查,主要是交替类目、停用类目、指示性类目的根查;

(7) 5--有效类号根查块,560 有效类号根查,主要是属、分、参类目的根查;

(8) 其他块,包括 6--其他分类号块、7--主题词块、8--来源信息块、9--国内使用块。

这套格式基本能把《中图法》类目体系描述清楚,为《中图法》建立机读数据库,生成电子版奠定了坚实的基础。

2.2 《中图法》MARC 数据库的建立

依据《中图法》机读格式把所有排版数据转换成 MARC 格式数据,共 53817 条,装入数据库中。在数

据机读化过程中,人工重点解决了两方面的问题,这些问题解决与否,直接关系到电子版软件功能的实现。

2.2.1 类号非规范化问题的处理

在《中图法》中有 20 % 的类目为特殊类号,即非层累制,类目加细一级,类号不是延长一位,但 18 % 的数据有规律可循,需要人工一一编辑。有如下几种:

(1) 类目加细一级,类号长两位,类号为八分法或双位制或借下级,如:

C91 社会学 其上位类号为 C,机读数据为 2601 \$h02 \$aC91 \$m 社会学

H211 少数民族古语言 其上位类号为:H2,机读数据为 2602 \$h03 \$aH211 \$m 少数民族古语言

K825.19 法律 其上位类号为 K825,机读数据为 2604 \$h05 \$aK825.19 \$m 法律

(2) 类目加细一级,类号长度不变,类号尾数递增,递增数值为 1,2,3...9,则类号为借上级,如:

K812 人物总传 其上位类号为 K811,机读数据为 2603 \$h04 \$aK812 \$m 人物总传:按时代分

(3) 类目加细一级,类号长一位,但是类号倒数第 2 位号码以其上位类类号尾数递增的顺序发生变化,一般递增数值为 1,2,3...9,则类号为借同级,如:

Q959.471 颞针鱼目 其上位类号为 Q959.46,机读数据为 2605 \$h07 \$aQ959.471 \$m 颞针鱼目

(4) 类目加细一级,类号长两位,其中一位为 0 或为 -,类号是借 0 或借 -,如:

K231.04 古代史籍 其上位类号为 K231,机读数据为 2006 \$h05 \$aK231.04 \$m 古代史籍

J292.2-3 碑帖源流考证、题跋、译文 其上位类号为 J292.2,机读数据为 2607 \$h05 \$aJ292.2-3 \$m 碑帖源流考证、题跋、译文

(5) 在《中图法》中用单字母或双字母作类号的类目的编辑制度称为顺序制,如:A、B、C、D...Z 或 DF、TB、TD、TE...TV,机读数据为:

2608 \$h01 \$aA \$m 马克思主义、列宁主义、毛泽东思想、邓小平理论(基本大类,无上位类)

2608 \$h02 \$aDF \$m 法律 其上位类为 D

以上数据有规律可循,但有 2 % 的类目类号极不规范,无规律可循,缺乏逻辑性。其中有些是类号编辑不规范;有些是违背借号规律,造成借号泛滥;有些是几种借号方法应用到一个类号的编辑制度中。

为使电子版能正确显示类目关系,我们对违背逻辑的类号进行了修改,主要集中在注释描述中,对已既成事实的类号不宜修改,因为影响类分文献,为此,我们定义类目字段第一指示符为 n,来进行特别标识,在生成电子版时需要特殊编辑,人工协助处理。如:

(1) 类目主体类号不规范

260n \$h07 \$aQ969.431.1 \$m 斑蛾总科 其上位类号为 Q969.42+3

260n \$h04 \$aJ60-02 \$m 音乐的哲学基础 其上位类号为 J60

260n \$h05 \$aD922.291.91 \$m 企业法、公司法 其上位类号为 D922.29

(2) 起止类类号书写不规范

在印刷版中,起止类类号的省略书写法很多,规则不统一,影响到电子版的类目正确显示。电子版的数据库对其类似的数据作了统一处理:起类类号的阿拉伯数字小于等于三位时,/后省略大类字母;如果大于三位时,/后则省略第一位小数点或-前的所有相同部分。例如:

S851.34+7.1/.7 等同于 S851.34+7.1/S851.7,该号为正确省略方式。

P1-093/097 改为 P1-093/-097

S945.4+1/945.4+9 改为 S945.4+1/.4+9

D5/7 改为 D5/77,因为 D7 不存在,计算机不识别该类号。

(3) 注释字段类号描述的不规范

注释表述自然语言化,注释中出现的类号随意性太强,缺乏逻辑性和规律性,其中有许多是超文本显示信息,必须进行规范,否则影响电子版的超文本正确显示。例如:

D52 世界政治制度与国家机构

专论一国的有关著作入 D6/7。应改为...入 D6 与 D73/77。

TK423、TK433、TK441+.3、TK473 等类下注释仿 TK413/414 分,应改为仿 TK413 分(TK414 不存在,TK414.1 是 TK413 的下位类)。

2.2.2 复分表和指示性类目的处理

专类复分表和通用复分表分别是供某些类目和全表类目使用的,都是供集合类目复分使用,但是每个复分表又是相对独立集中、自成体系的类目表。因此,为了电子版正确显示复分表,其机读数据的编制是把每个复分表连接在一个可概括的上位类下或

增设可概括的指示性类目下,同时将复分表中每个类目又与相关联的上位类或指示性类目链接,这样可正确显示复分表。如:D676.58/.59是为处理其下的专类复分表专门增设的指示性类目。指示性类目在类表中与使用类目的功能和显示位置不同,需要特殊描述,单独处理。

2.3 功能软件的研制

电子版是在《中图法》(第四版)数据机读化基础上,由软件来实现类目检索和类目显示的。因此,软件研制主要包括以下几方面内容。

2.3.1 类目详表、简表界面软件的研制

把每个类目作为类目“树”的结点,通过对结点的点击,生成显示详简不同的等级表。该软件的实现主要是通过机读数据补充必要的类目信息。

(1)每个类目进行上下位类的链接,通过560字段表示出类目间等级关系。

(2)产生结点999字段。

(3)辅助人工建立特殊类号的类目、指示性类目、专类复分表等结点。

2.3.2 文本显示、超文本显示的软件研制

该软件主要是通过类目结点、浏览窗口、类目注释等实现文本显示和超文本显示,包括多窗口显示、单条类目的显示、单级类目的显示(类目范围显示)、上下位类的显示,对不同的数据类型做不同的显示。

2.3.3 检索窗口及其功能软件的研制

类目要素检索、全文检索、逻辑检索等多窗口检索功能软件的研制。

2.3.4 应用功能软件的研制

(1)编目接口软件的研制,在编目同时可自动启动电子版,并对编目类号起到复制校验等功能。

(2)电子记事本功能软件的研制,对使用过程中发现的问题和解决问题的方案随时记录在案,对网络版用户起到规范文档的作用。

3 《中图法》电子版的体系结构及其功能

它主要包括两大部分:类目检索部分和类目显示部分。类目检索部分包括类号、类名、类级、复分号、复分表、注释等检索点,包括前方一致、精确一致、任意一致、全文检索、逻辑检索等检索方式。显示部分包括主表和附表类目等级关系显示(也称

类目树浏览显示)、某一类目的详细内容显示和邻近类目等级关系的显示等显示范围,包括浏览显示、多窗口多维显示、任意类目的超文本显示、印刷排版格式的显示和类目MARC格式的显示等显示方式。它除了提供和印刷本非常相似的“类目树”层次显示界面外,还表达了类目的体系结构,便于使用者浏览检索。在一个类目的详细信息显示中,除包含了和印刷本兼容的信息以外,与印刷本相比还有以下功能:

(1)从多途径快速精确检索类目。它不仅可从类号、类名、类目的注释文字等途径跨类快速查找所需类目,也可通过类号、类名、注释等概念符号组合检索所需类目,也可实现全文检索,即类表中出现过的任意概念、符号、类号的检索。总之可实现从主题概念、类目短语或类号快速检索到专指类目和相关类目,也可实现从学科角度进行任意等级类目的检索,随时扩大和缩小检索范围,实时灵活地改变检索策略,便于用户选择最专指的类目,提高了文献标引效率。

(2)实现多窗口多维检索,打破了印刷本的线性检索体系,便于用户对类号类目的选择判断、甄别和组配,起到了辅助标引的作用,为文献选择分类标引方式提供了必要条件,提高了文献标引质量。

(3)实现类目任意概念、类号的模糊检索和全文检索,包括类号、类名的后方一致的检索,满足了新用户和对《中图法》类目体系、类目概念的学科属性不熟悉的用户的需求。

(4)实现类目的多层面、多窗口、超文本的显示,特别是参照类号的超文本显示。类目的多维显示打破了印刷版的线性体系,大大加快了文献分类标引和检索的速度,起到了智能辅助标引的作用。

(5)根据类目的属性,通过不同的符号、不同的颜色对类目进行多层面的等级显示,对类目概念及其使用起到注释的作用,降低了标引难度,用户界面更友好,通俗易懂。

(6)起到分类目录检索系统和分类目录主题索引的作用。作为读者可根据主题概念和主题因素查找所需文献的类号,可根据类目树状显示了解学科范围,鸟瞰学科体系,选择合适的类号检索文献,是读者的必要检索工具。

(7)起到了分类规范文档的作用。在任意类目的固定窗口可增加标引员的注释,起到使用注释文档的作用,便于分类查重,也便于分类法修订。

4 《中图法》电子版的应用环境

《中图法》电子版软件对运行环境要求不高,只为大陆简体 Windows95/ Windows98/ Windows NT 4.0 以上版本/ Windows 2000,建议计算机内存在 16M 以上。如果是 Windows95 环境,须先安装 IE 4.01 以上版本后则可安装。安装方式分为两种:一种是最小空间的安装。这种方式只拷贝程序文件和配置文件到计算硬盘上,而将占用空间很大的数据库文件保留在光盘上。这种方式只占用 20M 左右的硬盘空间。另一种是最大空间的安装。这种方式将所有程序文件、配置文件和数据库文件都拷贝到计算机硬盘上,因而将占据约 200M 左右的硬盘空间。

《中图法》电子版使用操作较为方便简单,易掌握。只要具备一般的计算机使用常识和分类法常识则可操作使用。

《中图法》电子版的载体是一张光盘,安装极为简单,用户只需将这张光盘插入到计算机的光盘驱动器中,并运行光盘根目录中的 setup. exe 程序,安装后即可使用。

参考文献

- 1 戴维民等主编. 面向 21 世纪的情报语言学. 北京:北京图书馆出版社,2000
- 2 卜书庆,汪东波主编.《中图法》电子版. 北京:北京图书馆出版社,2001
- 3 王忠红. DDC 电子化历程概述. 大学图书馆学报,1999 (1)

卜书庆 国家图书馆副研究馆员。通讯地址:北京中关村南大街 33 号。邮编 100081。

宋安莉 北京图书馆出版社图书馆学编辑室主任,副编审。通讯地址:北京文津街 7 号。邮编 100034。

(来稿时间:2001-03-19)

(上接 24 页)

续表 6

教育学			体育学			心理学			统计学		
名次	地区	篇数	名次	地区	篇数	名次	地区	篇数	名次	地区	篇数
18	山西	62	18	湖南	8	18	山西	6	16	河南	1
19	广西	44	19	吉林	5	19	云南	5	16	甘肃	1
20	河北	41	20	广西	4	20	贵州	4	16	广西	1
21	天津	29	20	江西	4	20	福建	4	16	云南	1
22	江西	28	20	新疆	4	22	江西	3	16	山西	1
23	云南	24	20	内蒙古	4	23	广西	2			
24	青海	15	24	黑龙江	3	23	新疆	2			
25	宁夏	12	24	云南	3	25	内蒙古	1			
26	贵州	11	26	青海	2	25	黑龙江	1			
27	新疆	8	27	贵州	1	25	青海	1			
28	海南	5	27	宁夏	1						
29	西藏	2									
29	内蒙古	2									
	总计	2822		总计	850		总计	483		总计	98

参考文献

- 1 《中文社会科学引文索引》课题组. 中文社会科学引文索引(1998). 南京:南京大学出版社,2000
- 2 邹志仁. 中文社会科学引文索引的研制、意义与功能. 南京大学学报(社科版),2000(4)
- 3 苏新宁. 引文索引的数据质量控制研究. 中国图书馆学报,2001(2)

注 本文为教育部重大课题成果之一。课题号: 99JBZD870001。

韩新民 南京大学社会科学处副处长,副研究员。通讯地址:江苏南京。邮编 210093。

苏新宁 南京大学信息管理系教授,博士生导师。

(来稿时间:2001-01-10)