

●文榕生

机读目录的检索原理与功能

ABSTRACT With huge storage and fast computing speed of computers, MARC has powerful search functions that card catalogue does not have. There is still a lack of standardisation in computer cataloguing, which is affecting MARC quality. 9 refs. 1 tab.

KEY WORDS Cataloguing. MARC. Retrieval.

CLASS NUMBER G254.36

检索是书目的主要功能之一,著录信息中相当大部分与检索点密切相关。当今,书目形式已由传统的书本式、卡片式向机读式发展,检索过程也由人工查检改进为计算机操作。然而,从一些文献编目规则的制定和编制的机读目录数据来看,存在与计算机特点脱节的情况,影响了机读目录功能的发挥。

1 计算机的检索原理与特点

机读目录是在书本式与卡片式目录基础上发展起来的现代目录形式,其检索方式可以从前二者中觅到某些相似之处,但它又具有自己的特点(计算机具有逻辑判断功能,对“是”与“非”迅速做出判断与选择),便于处理有规律性问题,其检索有如下特点:

(1) 前方一致匹配。运用存同除异排除法,计算机可以自动在已存储的大量信息中将具有相同字符的信息集中。当首字符相同时,再判断比较次字符……如此类推,经过逐级筛选,达到精选目的。这也是卡片目录的基本排检原理,但计算机是机械性操作,严格的一一对应;而卡片目录采用人工操作,有一定的机动性,易出现偏差,甚至排错位置。

(2) 限定字长。从理论上说,只要存在相似的情况,区分就可以继续下去,但机读目录往往根据一般情况限定检索字符的长度,超出部分需

要人工进行判断、选择或改用其他检索途径。卡片目录的排检无所谓限定长度,可以一直比较下去;当常规比较难以区分时,可以再规定简便易行的方式,直至区分开^[1];当前方相同字符达到一定数量时,可设置导片分流。

(3) 模糊检索。卡片目录可以根据文字特点将题名、责任者、主题等按字顺混合排列成字典式目录(混排式目录)。机读目录也具有如此模糊检索功能,计算机可以根据单纯的表意文字、表音文字、数字等特点或它们混合形式进行自动排检。故在同一机读数据中没有必要设置相同的检索点(尽管其内涵不同)。

(4) 指令检索。卡片目录可以将款目(以标目为特征)按分类、题名、责任者、主题等不同途径分门别类排检(单排式目录),这样检索更有的放矢。机读目录也可以通过指令从预设的不同检索字段的信息中检索,提高检索精度。输入完整信息,在定长字符内计算机可以迅速准确地检索到书目数据;而卡片目录不仅需要人工不断干预,若遇上排列误差的款目,还要费尽周折。

卡片目录中可以通过设置参照款目说明检索点的简化、省略、批量规范等内容;各款目中的不少著录信息往往需要重复著录。机读目录数据相当卡片目录的通用款目,但前者检索点不需重复;一般不做说明。而是通过完整客观和规范两种著录信息增加检索自由度;用划分层次方式,既保持多层次著录信息的客观逻辑性,又缓

解字符定长区分困难的矛盾。

2 机读目录实现了检索体系理论

根据人们认识客观事物的规律和文献特点,通过科学编制方法和组织形式,充分揭示文献的特点以及它们之间的相互关系,能够形成有机检索整体。尽管在按文献类型、文种、各种著录信息特点等检索体系的理论研究方面取得一定进展,但囿于以往的手工操作条件,未能在实践中得以全部实现。

卡片目录除了编制主要款目,还要设立各种标目,但要对检索点加以限制,通常只设2~3种检索途径,辅助款目与参照款目设置也十分有限,一般情况一种文献的卡片目录需10~30张款目,故难以满足多途径检索要求,实现检索体系。

机读目录兼容不同类型、文种文献,每一条数据中记录的字段、子字段可以完善检索体系所赋功能。当然,机读目录不等同于卡片目录:前者难以仿造后者某些功能,如单纯参照,某些功能采用不同方式实现。现将两种目录实现检索(目录)体系作比较,见表1。

机读目录的检索字段除了一般检索功能,还有其他功用。

(1) 交替题名。200字段仅第一交替书名有检索意义,其余的应将纯书名部分分别著录在517字段,实现附加功能。

(2) 统一题名。名称不同,内容相似的书,宜设统一题名规范,实现相互参照功能。如《金陵十二钗》、《石头记》、《绣像红楼梦》、《脂砚斋重评石头记》、《程甲本红楼梦》、《四十八本红楼梦》、《金玉缘》、《乾隆抄本百廿回红楼梦手稿》等宜分设500字段:

200 1k \$a 金陵十二钗

500 1k \$a 红楼梦

机读目录200和500字段皆有检索意义,各数据皆设“500 1k \$a 红楼梦”,检索时若键入上述任一书名,皆出现统一的《红楼梦》;若键入《红楼梦》,便会显示所有预设上述各种书。

(3) 规范题名。500字段又有规范作用,可

实现一般参照功能,用户不必学习过多检索规则便可从客观或规范任一方检索文献。如:

200 1k \$a 一九八九上海文化年鉴

500 1k \$a 上海文化年鉴

200 1k \$a 钦定全唐诗

500 1k \$a 全唐诗

(4) 补充检索点。著录信息源有团体责任说明,又有责任说明页记载个人责任说明,宜选适当的作为检索点,实现附加功能。如:

200 1k \$a 中国文献编目规则 \$f《中国文献编目规则》编撰小组编

701 k0 \$a 黄俊贵 \$4 主编

711 02 \$a《中国文献编目规则》编撰小组 \$4 编

(5) 规范团体责任者。同时采用全称及简称,保持客观(非规范)与规范两种检索点。如:

200 1k \$a 材料新星 \$e 纳米材料科学 \$f 严东生,冯瑞主编

225 2k \$a 攀登计划普及丛书 \$f 国家科委基础研究高技术司组织

701 k0 \$a 严东生 \$4 主编

701 k0 \$a 冯瑞 \$4 主编

711 02 \$a 国家科委 \$b 基础研究高技术司 \$4 组织

711 02 \$a 国家科学技术委员会 \$b 基础研究高技术司 \$4 组织

(6) 规范个人责任者,补充个人责任说明,集中同责任者著作。如:

200 1k \$a 总理全集 \$f 孙文 著

300 kk \$a 孙中山(1866~1925)即孙文

701 k0 \$a 孙文 \$f 1866~1925 \$4 著

701 k0 \$a 孙中山 \$f 1866~1925 \$4 著

200 1k \$a 国父全集 \$f 中国国民党中央委员会党史委员会编订

300 kk \$a 本书由孙中山著

701 k0 \$a 孙中山 \$f 1866~1925 \$4 著

711 02 \$a 中国国民党中央委员会 \$b 党史委员会 \$4 编订

表 1 卡片目录与机读目录实现目录体系的不同措施

款目类型与著录方式		款目种类	机读目录相关字段或子字段说明 (数字表示机读目录字段)
主要款目(基本著录方式)	按著录方式	1. 分别著录(分析著录)	主要著录方式
		2. 综合著录(整体著录)	不另做数据,通过“连接款目块-4”的有关字段达到此种效果
	按检索途径	1. 分类	按确定的分类法采用相应字段:675(UDC)、676(DDC)、680(LC)、686(其它分类法)、690(CLC)、692(CASLC)
		2. 题名	200(题名与责任者项)
		3. 责任者	根据 200 具体责任说明采用 701(人名-等同责任者)、702(人名-次要责任者)、711(团体名称-等同责任者)、712(团体名称-次要责任者)
		4. 主题	按确定的主题方式采用相应字段:606(普遍主题,即叙词)、610(非控制主题词,即关键词)
		5. 出版者	根据文献上实际记载,采用 010(ISBN)、011(ISSN)、091(统一书刊号)等
		6. 出版时间	100(一般处理数据)
		7. 登录号	905(馆藏信息)
	辅助款目(附加方式)	1. 分类	适用于多主题内容的文献,重复已确定分类法的相应字段
		2. 题名	适用于与 200 有区别的题名,采用相应字段:510(并列题名)、512(封面题名)、516(书脊题名)、517(其它题名,交替题名入此)等
		3. 责任者	适用于与 200 相同责任方式的其它责任者,根据具体责任说明采用或重复 701, 702, 711, 712
		4. 主题	适用于多主题内容的文献,重复已确定主题方式的相应字段或子字段:606, 610
		5. 出版者	适用于其他合作出版者,重复或选择相应字段:010, 011, 091 等
参照款目(参照方式)	单纯参照	1. 分类	机读目录不采用单纯参照(说明参照)功能
		2. 题名	
		3. 责任者	
		4. 主题	
	相互参照	1. 分类	a. 分类附加具有相互参照功能;b. 非确定的其它分类法,采用相应字段,或重复相应字段(多主题内容)
		2. 题名	a. 题名附加具有相互参照功能;b. 规范题名作用的 500(统一题名);c. 各种需连接的题名(连接款目-4)
		3. 责任者	a. 责任者附加具有相互参照功能;b. 多层次著录信息文献的不同层次责任说明,根据具体情况采用或重复 701, 702, 711, 712;c. 规范责任说明,根据具体情况采用或重复 701, 702, 711, 712;d. 宜补充的责任说明(多用于个人责任者)选择或重复 701, 702
		4. 主题	a. 主题附加具有相互参照功能;b. 非确定的另一种主题方式,采用相应字段,或重复相应字段、子字段(多主题内容)
		5. 文种	453(译为)、454(译自)
		6. 载体	451(同一载体的其它版本)、452(不同载体的其它版本)
	一般参照	1. 分类 2. 题名 3. 责任者 4. 主题	机读目录不另设数据,卡片目录的此种作用是通过部分字段功能实现

3 编目规范文件的欠缺

文献编目的规范指导性文件近年都经过修订,如卡片目录的GB3792系列修订后形成一系列国家标准报批稿^[2]和《中国文献编目规则》^[3],机读目录的修订本有《中国机读目录格式》^[4]和《〈中国机读目录格式〉手册》^[5]等都较以往版本有较大改进,但它们仍存在不尽人之处。

《机读目录格式》逐一按字段进行规定、说明,反映了机读目录兼容性(对目录形式、著录项目、项目标识、文献类型、不同文种文献等)强的特点;但它没有对规定著录信息源的选取进行规定,没有制定规范原则对不规范检索点实施控制,对交替书名、卷标识的著录闪烁其辞,实例少,某些实例与规定、或规定之间互相矛盾^[6~8],检索字段设定过于繁琐(既然其主要功能是检索,在终端又不完全显示,设定过细反不便计算机软件设置,影响功能发挥)等等,往往造成编目员操作困难,影响数据质量。

《中国文献编目规则》详细规定了著录信息源选取,各项说明均有实例;但也存在规定之间、或实例与规定不一致等问题。尤其是标目法部分,编目员难以操作,也不便于计算机运作。主要表现在:

(1) 多重标准,掌握困难。例如“中国共产党”就分成“中国共产党中央”(全称)、“中共中央”(半简称)、“中共×××委”(大部分简称)三种类型标目,而各民主党派则只有“以其全称为标目”一种规定,各级科学、教育、文化机构又一般直接以其名称为标目;青少年、妇女、工人、学生等各种人民团体、群众组织再以惯用名称为标目。又如“多级机构”按“直接选择独立机构”、“以其最高及最低层次的机构名称”、“以全称”三种形式“规范”。且不论用户,就是编目员操作时对如此多种规定的掌握都有困难。

(2) 以简称代全称,本末倒置。简称是对较

复杂名称的简化形式,其根源是完整的名称(全称),“采用全称”^[9]进行规范是避免滥用简称或简化不当引起偏差甚至造成不必要麻烦的简便易行规定。如“中国科学院地理研究所”可以任意简称为“中科院地理研究所”、“中国科学院地理所”、“中科院地理所”、“中地所”、“地理所”等,“计量科学研究院”可任意简称“计量院”、“计科院”、“计院”等,它们在一定范围内语义相通。标目法中采取了相当数量的简称、俗称取代全称,颠倒了规范与非规范关系。

参考文献

- 1 文榕生. 刍议现行中文书名目录的排检法. 图书馆建设, 1993(3): 46~47
- 2 朱南主编. 科学技术信息系统标准与使用指南. 第五卷, 情报文献工作标准. 北京: 中国标准出版社, 1996
- 3 《中国文献编目规则》编辑小组编. 中国文献编目规则. 广州: 广东人民出版社, 1996
- 4 朱岩等起草. 中华人民共和国文化行业标准. WH/T0503-96: 中国机读目录格式. 北京: 文化部印刷厂印刷, 1996
- 5 北京图书馆《中国机读目录格式》编委会. 《中国机读目录格式》手册. 北京: 华艺出版社, 1995
- 6 文榕生, 赵春旻. 多层次著录信息图书标识之管见. 河北科技图苑, 1997(4): 26~28, 36
- 7 文榕生. 论书名著录的规范. 图书馆建设, 1997(1): 35~37
- 8 文榕生. 略论机读目录格式中的卷标识. 江苏图书馆学报, 1997(1): 24~26
- 9 图书在版编目国家标准起草小组起草. 中华人民共和国国家标准. GB12450-90: 图书书名页. 见: 作者编辑出版常用国家标准. 北京: 中国标准出版社, 1993

文榕生 中科院文献情报中心, 馆员. 通讯地址: 北京中关村. 邮编 100080.

(来稿时间: 1998-05-05. 编发者: 赵薇)