

陈树年

在机编和机检中对“组配编号”和“多重列类”的再认识

ABSTRACT Since we are using computers for cataloguing and retrieval, we should restructure hierarchical classification systems and enhance their multiple-approach search functions by using the method of multiple division as well as the method of pre-coordinate class number and post-coordinate searching.

KEY WORDS Hierarchical classification. Multiple division. Facet analysis. Indexing. Retrieval.

CLASS NUMBER G254

1 体系分类法增加组配功能,并不能显著提高检索效率

组配,是将两个或多个简单的要素,按一定的规则组合成一个复杂的要素。在叙词语言中是指将两个或多个简单的概念合成一个新的复杂的概念;在分类语言中是一种类号合成技术,即按一定的规则或使用特定的组配符号,利用已有的类号(包括复分表的子目号),合成一个新的更专指的类号。

体系分类法(以下均以《中图法》为例)运用组配的方法大体有三种类型:

(1) 主类号使用通用复分表或专用复分表的子目进行细分;

(2) 主类号仿照已列出的有关类目细分;

(3) 主类号之间直接进行组配。

前两种组配形式的主要作用是提高类目的细分能力,而又不增加类表的篇幅。例如《公路桥冬季施工技术》可标引为U 448.145.43,其中“543”(冬季施工技术)是仿U 44分展开的。当使用特定的组配符号与有关子目组配时,还能在现有类目体系的基

础上提高对文献主题的描述(标引)能力。例如《美国对晚期肺癌治疗研究的最新进展》可标引为R 734.205(712)=5 147,这个类号中包括“肺肿瘤”、“治疗”、“美国”、“现代”、“晚期”等主题因素,后三个主题因素是原类目体系所不能提供的。

由于提高了类号的专指度,从理论上讲可提高检准率。但实际上体系分类法结构的复杂性和组号的复杂性,使一般用户很难掌握到标引人员的程度,检准率的提高就会大打折扣。上例类号虽是组配而成的,但是先组性的,一旦合成后就成为一个凝固的整体,其中所包含的各个主题因素便不能离析出来(特别是第二种方式合成的类号完全不能离析),因此不能作为独立的检索入口使用,多途径检索也就无从谈起。

第三种组配方式是比较灵活的,除可用于类目细分之外,还可合成新的主题概念,例如“生物社会学”可组配标引为C919:Q(分科社会学:生物学)。但这种组配在手工编目、手工检索和现有类目结构条件下,不论从标引还是从检索角度来说其作用都是十分有限的,原因是:

第一,手工编目条件下类号的长度受限

制,不可能用很多的主类号组配成长长的字符串(如“用于化纤织物染整自动控制的模式识别装置”可组配为 TS190.645: TP273: TP394.1),这样对组织文献和组织分类款目都不方便。

第二,体系分类法编制受逻辑划分规则和划分层次的制约,文献主题中所有具有检索意义的主题因素(属性),不会都作为分类标准使用,也就是说分类法没有列出描述某些主题因素的类目,那么使用主类号组配某些新主题概念就成了无米之炊。此外,为保证现有类目体系不受组配编号的“冲击”,主类号直接组配法从标引规则上也严格地进行了控制。

第三,在手工检索条件下,无法实现后组配式的逻辑检索,用主类号逻辑组配起来的类号只能作为先组类号使用,即使进行轮排,也只相当于附加分类号的作用,而在手工编目条件下也难以广泛应用。

2 体系分类法中“多重列类”的功能不能充分发挥

等级体系分类法是我国各种分类法编制的基本模式,这种分类法用建立等级结构的方式组织和显示人类的全部知识。为揭示各知识领域及各学科门类的从属关系、并列关系和它们之间亲疏远近,首先确定一个知识分类大纲即基本大类,然后以基本大类为起点,逐级进行概念划分。在对一个较大的概念进行划分时,遵循逻辑划分的规则,即每一层次的划分只能选择一个分类标准(事物的某种属性),使划分出来的子目(子概念)增加相同的内涵,从而限制了其外延。通过逻辑划分形成的子目之间没有外延交叉,这对类分文献是十分实用的。然而事物的属性是多方面的,当某事物的几种属性都具有检索意义时,应首先使用哪种属性进行划分呢?于是就产生了“分类标准的引用次序”,即人为地确定类目划分时所使用分类标准的先后。虽然某事物所具有的几种属

性都作为分类标准被使用,但文献只能按最先使用的分类标准集中,第二划分层次的标准仅能在第一层次划分形成的若干区间分别集中有关文献,越后使用的分类标准按该属性集中文献的能力越弱,使用该属性作为检索入口检全文献也越难。

为了使事物所具有的多种属性在一个相对较高的划分层次(通常选择某类事物的类称)下同时作为分类标准使用,避免在多层次划分中被先使用的分类标准“层层切割”,降低聚类的能力,于是等级体系分类法采用了变通的方法,允许部分类目在同一划分层次同时使用几个分类标准进行划分,这就是所谓的“多重列类”。这样在一个上位类下就形成几组并列的、外延有交叉的同位类,例如:

- U 488 各种桥梁
- 448.1 各种用途桥梁
- 448.11 人行桥
- 448.2 各种结构桥梁
- 448.21 梁式桥
- 448.3 各种材料桥梁
- 448.31 木桥、竹桥
- 448.4 各种形式桥梁
- 448.41 斜桥
-

“桥梁”在这里是事物的类称,以此为多重划分的起点,就可以把各种属性或特征的桥梁同时平行地列举出来。其作用在于:如果分别从不同的属性论述桥梁,都可以选择相应的类号进行标引,用户也能以不同的属性作为检索入口检索文献。但如果文献主题同时涉及事物的几种属性,由于同位类存在外延交叉,就有几个类号可用于标引,为保证主题文献标引的一致性,就人为规定采用“最前编号法”或“最后编号法”。这时,多重列类的优越性就无法体现。

多重列类法是编制分面分类法的基本技术——分面分析在体系分类法中的有限应用。分面分析是对主题概念进行范畴化处理的技

术,即通过对每个大类(知识领域)中大大小小的概念进行分析,将其按不同的特征分成性质纯粹、相互排斥的组面,在一个组面之下可以继续分出亚面或直接列举类目。例如:

- FC 成人教育
- FC1 继续教育 (按对象分)
- FC2 农民教育
(按教育形式分)
- FC6 函授教育
- FC9 培训
- FC91 基础培训
(按培训地点的)
- FC97 国外培训
(摘自《教育主题词表》)
- LL 民法
(按法律关系分)
- LL 1 债 (按生产根据分)
- LL 12 行政命令之债
(按行为方式分)
- LL 16 债的转移
(按法律规范分)
- LL 2 合同法
- LL 3 继承法
(摘自《社会科学检索词表》)

采用分面分析产生的一组组性质纯粹、内涵相对较浅的类目(与叙词接近),主要作用是用于对文献进行组配标引,并为机检的后组检索提供条件。因此,在体系分类法中多重列类法不与组配技术相结合时,其作用是十分有限的,主要是在一定程度上提高对文献主题的描述能力,而对检索效率影响不大。

3 使用多重列类法同时使用组配标引法,才能充分发挥多重列类的功能,这也是提高体系分类法多途径检索功能的基本出路

当我们使用后组式检索语言(如《汉语主

题词表》)编制传统的手工检索工具时,虽然也采用组配技术,但标引的结果是先组式的主题词串(相当于组配式的标题),不能用于后组检索,那么后组式检索语言的优越性也就难以发挥。同样,体系分类法引进分面分析技术,采用多重列类法,如仅用于编制传统的手工检索工具的话,其优越性也不能有效地发挥。

图书情报单位普遍实现计算机编目和计算机检索,为通过采用多重列类提高体系分类法多途径检索功能提供了技术条件。实现的思路是:在修订或编制分类法时,尽可能地广泛采用多重列类法编制类目,为不影响体系分类法的系统检索功能,主要是在较低的级位上应用。这样,在不增加类目划分层次的基础上,增加描述文献主题某一属性(特征)的类目,从而提高独立从特定角度描述文献主题的能力——这是体系分类法逐步进行分面改造的过程;其次,在文献标引时放弃惯用的先组配的思路(指主类号间组配),因为这种组配是无多大意义的,转而采用多个列举的类号后组标引。例如《德国钢筋混凝土大跨度公路桥冬季施工技术》一文,按最前标号法只能标引为 U 448.14-115.65,不能提供从“钢筋混凝土桥”、“长跨桥”、“桥梁冬季施工”、“德国”检索的途径,而我们现在如果标引为:

- 200 \$ a 钢筋混凝土大跨度公路桥冬季施工技术 \$ f 李鸿志编译
- 690 \$ aU 448 34
- 690 \$ aU 448 43
- 690 \$ aU 448 14
- 690 \$ aU 445 43
- 905 \$ b0078432 \$ dU 448.14 \$ e4034

不但分别提供了 U 448.14 (公路桥)、U 448.34 (钢筋混凝土桥)、U 448.43 (长跨桥)以及 U 445.43 (桥梁冬季施工技术)的检索途径,还可以运用“U 448.14 AND U 448.34”、“U 448.34 AND U 448.43”、“U 448.43 AND U 448.34 AND U 445.43”等检索式进行后组配检索。这样一来,多重列

类的基本优越性才得以充分发挥, 由此对检索效率的影响显而易见。

多重列类法还不能作为体系分类法的主要列类技术, 所提供的多途径检索功能还是有限的。为此, 我们还可以从改进用特定组配符号连接通用复分表子目的组配标引方法入手, 挖掘潜力。基本思路是: 首先大大扩充现有各通用复分表的子目, 特别是总论复分表、地区表和时代表, 使之能提供更多的描述文献主题特征的子目; 其次, 在 CNMARC 格式的 690 字段标引时, 将用特定组配符号连接的复分表子目作为一个独立的分类号和检索标识来使用, 这在手工标引和编制手工检索工具中是不允许也是无意义的, 但在计算机编目和检索的条件下, 不但可以实现, 而且赋予了新的功能。例如《英国十七世纪造船史》一文, 手工标引为 U 66-095. 16, 使用这个分类号只能从造船工业入手进行检索, 而“英国”、“17 世纪”、“历史”均不能作为检索入口。现在我们标引为(这里假定以“= 4117”作为 17 世纪的时代标识):

200 \$ a 英国十七世纪造船史 \$ f(英) 史密斯著
690 \$ aU 66
690 \$ a-09
690 \$ a(561)
690 \$ a= 4117
690 \$ b0023415 \$ dU 66 \$ e5504

由于复分子目号前均有特定的组配符号做区分标识, 因此在计算机检索中均可作为独立的检索标识参与组配检索, 如用“U 66 AND = 4117”(检索式中至少有一个主类号)组配可以检索出“十七世纪造船工业”的文献。这样一来, 原来凝固在先组类号中的主题因素便被离析出来, 成为可独立使用的检索点。当然, 采用这种新的标引方法必须有相应的规则加以限定, 否则会出现很大的随意性。

通过上述分析可以看出, 原先在手工标引和检索条件下多重列类和组配技术只能发

挥极有限的作用, 而在计算机编目和检索的条件下, 只要我们稍作变革, 它们自身的优越功能将发挥得淋漓尽致。

计算机技术的普遍应用, 使文献编目和文献检索发生了重大变革, 我们的观念也必须随之改变。捷克学者托曼(Toman, I)曾深刻地指出:“对字顺主题和分类系统的研究表明, 任何一种系统只不过是对同一组织原则进行另一种组合而已”;“当这些系统用于电子计算机时, 所有的差别都消失了。”在机读目录中, 一条记录著录有文献的各种数据和多种检索标识, 分类目录、主题目录、著者目录的界线已经消失。而分类检索语言和主题检索语言也由原来的泾渭分明、相互对立, 转而相互渗透、相互结合、相互包含, 它们之间的界线也变得越来越模糊。现在计算机检索系统已可以在分类号和主题词之间进行逻辑和、逻辑与、逻辑非的组配检索, 这使体系分类法的分面改造降低了难度, 因为当分类法不易或不能通过多重列类提供更多的能独立描述文献主题因素的类目时, 可以通过用主题词与分类号在检索阶段的组配作补充。这时分类检索语言和主题检索语言的相辅相成的作用, 在更深的层次上体现出来。

总之, 计算机技术的应用为体系分类法的改造和功能的发挥, 为分类检索语言和主题检索语言的结合, 注入了无穷的活力。不论是检索语言的编制和修订, 也不论是标引技术和检索技术的改进与改革, 都必须彻底摆脱“手工阶段”所固有的思路和观念, 只有这样, 才能在新的信息时代满足联机检索、联网检索对情报语言日益提高的要求。

陈树年 华东理工大学石油化工学院图书馆研究馆员, 《中国图书馆图书分类法》副主编。通讯地址: 上海市金山区。邮编 201512。

(来稿时间: 1997. 12. 30。编发者: 赵薇)