

20世纪中国图书馆学文库·72

主题法导论

侯汉清 马张华 主编

国家图书馆出版社

20 世纪中国图书馆学文库·72

主题法导论

侯汉清 马张华 主编
张涵 戴维民 王荣授 陈树年 编著

圖 國家圖書館出版社

本书据北京大学出版社 1991 年 9 月第 1 版排印

前 言

近年来随着科技情报检索系统(包括科技检索刊物体系)的建立和办公室自动化的发展,国内出现了“词表热”。政府部门和专门机构纷纷组织力量编制、出版叙词表,各地情报机构和专业图书馆利用叙词表进行主题标引,建立起一批文献数据库,并为几十种检索刊物编制主题索引。每年数百家高等院校的数以万计的学生在文献检索课堂上学习使用叙词表。这种形势有力地推动了主题法、特别是叙词法理论与实践的发展。我们作为图书馆学和情报学专业的教师,愿趁此机会向图书情报界的同行献上这本小书。

本书是一本全面论述主题法的著作,其中以主题检索语言为主线,对叙词语言、标题语言和关键词语言及其标引进行了详细的阐述,同时设置专章介绍了主题检索工具的编制。为了反映80年代以来国内外主题法研究和应用的最新成果,我们在书中介绍了在主题法领域出现的标准化、自动化、兼容化、自然语言化、分类主题一体化等趋向。当然这本书的写作也概括了我们多年来从事主题法教学、参加叙词表编制的实际经验。书中还包含我们对一些主题法理论和技术问题的观点和看法,敬请同行们指正。

本书由侯汉清(第二章、第三章)、马张华(第二章、第六章)、张涵(第一章、第七章)、戴维民(第四章、第八章、第九章)、王荣授(第四章)、陈树年(第五章)共同撰写。本书第七章介绍保留上下文索引系统一节特请黄水清同志撰写。全书由侯汉清、马张华统

稿。

本书在编写过程中参考了国内外一些文献(见各章的引用书目和书末的参考文献目录)。本书还得到北京大学图书馆学情报学系和北京大学出版社的支持和帮助,顺此一并表示深切的谢意。

本书是《图书分类》(书目文献出版社出版)的姐妹篇。《图书分类》印数已逾十余万册,修订本已出版。愿本书像《图书分类》一书一样受到读者的欢迎。

编著者

1991年5月

目 次

第一章 导言	(1)
一、什么是主题法	(1)
二、主题法的发展及类型	(6)
三、主题法与分类法	(8)
第二章 主题法的原理	(12)
第一节 情报检索中的词汇控制	(12)
一、为什么要在情报检索中实施词汇控制	(12)
二、词汇控制的原则	(15)
三、词汇控制的范围	(16)
四、词汇控制的利弊	(22)
第二节 句法控制	(25)
一、句法控制的意义	(25)
二、引用次序	(27)
三、控制符号	(37)
四、句式变换——轮排	(43)
第三章 标题语言	(45)
第一节 标题法的发展及原理	(45)
一、标题法的由来及发展	(45)
二、标题法的词汇控制	(47)
三、标题的形式	(56)
第二节 标题表	(61)
一、标题表的性质和结构	(61)

二、《美国国会图书馆标题表》结构剖析	(63)
三、《美国国会图书馆标题表》简评	(67)
第三节 标题表的使用	(75)
一、标题的选择	(75)
二、标题的复分	(80)
三、标引深度和专指度的确定	(87)
四、标题变动的处理与主题标准档的建立	(89)
第四章 叙词语言	(93)
第一节 叙词语言的发展及原理	(93)
一、叙词语言的前身——元词语言	(93)
二、叙词语言的产生及发展	(96)
三、叙词语言的基本原理	(100)
第二节 叙词语言的词汇控制	(104)
一、词形控制	(105)
二、词义控制	(108)
三、词间关系控制	(110)
四、先组度控制	(117)
第三节 叙词表的结构	(119)
一、叙词表的微观结构	(120)
二、叙词表的宏观结构	(125)
第四节 叙词表的编制和管理	(138)
一、叙词表的编制方式	(138)
二、叙词表的编制程序	(139)
三、叙词表的管理	(144)
第五节 叙词表的评价	(147)
一、叙词表评价的内容和方法	(147)
二、《汉语主题词表》及其评价	(151)
第五章 分类主题一体化语言	(159)
第一节 分类主题一体化语言的发展及原理	(159)
一、什么是分类主题一体化语言	(159)

二、分类主题一体化语言研究进展	(161)
三、分类主题一体化语言的原理	(165)
第二节 分类主题一体化词表的结构	(167)
一、一体化词表的宏观结构	(168)
二、一体化词表的微观结构	(171)
第三节 分类主题一体化词表的编制	(179)
一、一体化词表编制方式的选择	(179)
二、一体化词表的编制程序	(181)
三、一体化词表的编制技术	(184)
第六章 主题标引工作	(195)
第一节 主题标引工作概述	(195)
一、主题标引的方式	(195)
二、主题标引工作程序	(199)
三、主题标引的质量控制	(202)
第二节 主题标引方法	(206)
一、主题分析	(206)
二、查表选词	(217)
三、标引词的处理	(229)
第三节 主题标引规则	(241)
一、主题标引的基本规则	(241)
二、各种主题、文献类型的标引	(248)
第七章 主题检索工具	(263)
第一节 主题目录	(263)
一、主题款目编制法	(264)
二、主题目录参照编制法	(265)
三、主题目录指导片编制法	(271)
四、主题目录的组织	(273)
第二节 分类目录字顺主题索引	(276)
一、分类目录字顺主题索引的功能	(276)
二、分类目录字顺主题索引的类型	(278)

三、分类目录字顺主题索引的编制	(280)
第三节 检索刊物主题索引	(286)
一、检索刊物及其索引	(286)
二、检索刊物主题索引的类型	(287)
三、检索刊物主题索引的结构	(289)
四、检索刊物主题索引的编制	(291)
第四节 机编索引系统	(293)
一、PRECIS 的发展	(293)
二、PRECIS 的款目结构	(295)
三、PRECIS 的标引	(298)
四、PRECIS 的开放词表	(303)
第八章 自然语言标引与检索	(306)
第一节 关键词语言	(306)
一、关键词索引法的历史发展	(306)
二、关键词语言的构成原理	(307)
三、关键词法的类型	(308)
第二节 自动标引	(311)
一、自动标引的发展	(311)
二、自动标引的主要方法	(314)
三、中文文献自动标引	(315)
第三节 自然语言检索	(317)
一、自然语言检索技术	(317)
二、全文检索系统	(319)
第四节 情报检索语言的自然语言化	(321)
一、情报检索语言的自然语言化趋势	(321)
二、自然语言与受控语言的结合	(323)
第九章 主题法发展趋势	(326)
第一节 标准化	(327)
一、叙词表编制标准化	(327)
二、主题标引标准化	(330)

第二节 计算机化	(333)
一、计算机辅助编制和管理词表的发展	(334)
二、叙词表的计算机辅助编制技术	(335)
三、叙词表的计算机管理	(338)
第三节 兼容与互换	(339)
一、情报检索系统的兼容性	(339)
二、兼容与互换的模式与方法	(341)

第一章 导 言

一、什么是主题法

美国未来学家奈斯比特(Naisbitt, J.)在《大趋势》一书中指出:“我们淹没在信息中,但是却渴求知识。”为什么当各类信息像洪水一样向我们涌来时,我们仍然缺乏所需要的信息呢?这是因为在信息社会之中,“没有控制和没有组织的信息不再是一种资源。它倒反而成为信息工作者的敌人。”*

我们图书馆和情报工作者在信息社会中的重要任务之一,就是对众多的文献信息加以控制和组织。文献控制的基本方法是:①文献的描述,即客观地对文献的内容和形式特征进行分析、选择和描述;②文献的浓缩,即把文献的内容加以压缩、提炼或概括。通过描述和浓缩,把文献加工成便于人们查找的替代品(或称替代文献);③文献的整序,即把众多的杂乱无章的替代文献按照文献特征井然有序地排列起来,建成文献的检索系统。

一篇文献,我们可以把它加工成摘录、文摘或提要,还可以进一步浓缩为扩充篇名或篇名。浓缩程度最高、涵义最明确的替代文献形式是主题词(或称主题词集合)(参看图1-1)。当然还可以进一步压缩为文献号,但此时已经成为一种代码,失去了原来的

* 见奈斯比特《大趋势——改变我们生活的十个新趋向》(新华出版社,1984年)。

涵义。我们通过对文献内容的分析,可以把文献所论述的对象(或事物)——主题概括出来,再使用规定的词汇,并按照一定的规则加以描述,就可以得到文献之精髓——主题词。

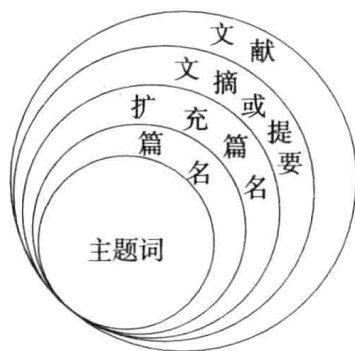


图1-1 文献浓缩示意图

把文献加工成替代品——主题词,尚未完成文献控制的全过程,还必须把代表众多文献的成千上万个主题词排列成为一种有序系统,使每一种具有特定主题内容的文献都固定在这个有序系统的某一特定位置上,这样,用户就可以准确地获取所需文献。在这个由主题词组成的有序系统中,主题词是按其相互之间的关系予以存放的,因而用户不仅可以查到关于某一主题的文献,而且可以查到相关主题的文献。这就是主题整序法。

我们再从情报交流的过程考察一下文献控制。维克利(Vickery, B. C.)曾经这样描绘情报交流的过程(参看图1-2)*。人类环境中的一个对象(事物、现象、过程等),经过研究,在信息发源者——科技工作者的头脑中上升为一个概念。这个概念综合到科技工作者的个人知识结构之中,然后他又以语词或符号把它表达出来。为了传递有关此概念的情报,这些语言符号又作为一条信息或一篇文献发出。后来这篇文献或信息又被收入一个经过组织的信息集合体(如图书馆或情报机构)之中,构成了公共知识,在那里,这篇文献被标引人员赋予一个或多个标识,这些标识又被组织在一个或几个标识集合体(如索引)之中。潜在的信息接收者(即用户)则根据其知识结构,也以语言符号的形式表达其情报需

* 见维克利等《情报科学的理论与实践》(科学技术文献出版社,1990年)。

求。这些提问被赋予一个或多个标识。此后这些标识又与组织好的标识集合体进行相符性比较,从而完成文献的检索。

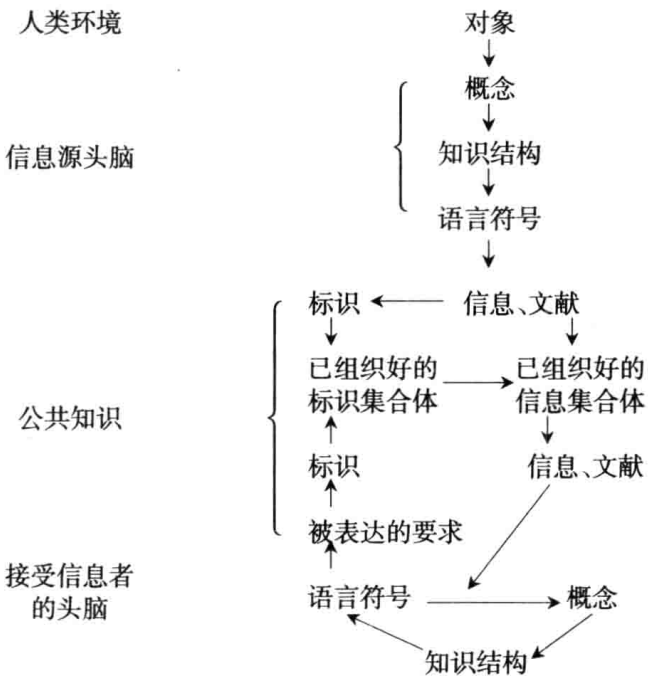


图1-2 情报交流过程

在这一情报交流过程中,从文献到标识集合体的过程就是一个对文献进行浓缩、描述和整序的过程,即情报存储的过程,而从用户的情报需求到标识集合体的过程,则恰好是文献控制或情报存储的逆过程。如果我们把这幅图中的标识理解成主题标识(主题词),把标识集合体理解成主题检索工具或检索系统,我们完全可以把它看成是一幅主题法的清晰图解。

从上述介绍可以看出,主题法实际上有两重涵义:

其一是指上文所述的主题整序法,是一种用语词标识处理文

献、组织主题检索工具或检索系统的方法。即把众多的文献根据其内容主题,强制性地用一个语词标识系统予以组织。(如图表1-2所示)。主题整序开始时,必须把文献的主题概念转换为经过规范的语词标识(即主题词),这一过程被称为主题标引。同样,用户进行检索时,必须把其提问的概念转换为语词标识,并据此进行查找,这一过程被称为主题检索。

其二是指主题检索语言。我们知道情报检索语言有两大语系,一个是以码号为文献主题的标识的分类检索语言,另一个是以语词为标识的主题检索语言。前者简称为分类法,后者则相应地被人们称之为主题法。

主题语言与自然语言一样,也是由词汇和语法(主要是句法)两部分组成。奥斯汀(Austin, D.)曾用下图来说明这两部分之间的关系。

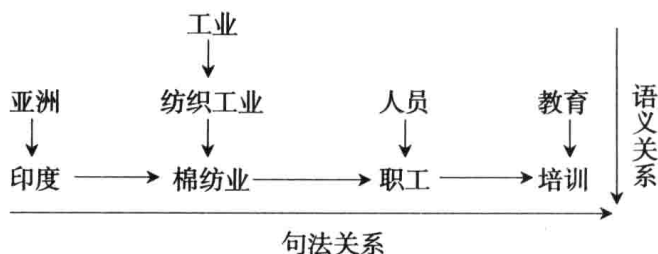


图1-3 主题语言的语义关系和句法关系

词汇是主题语言中的主体,其具体体现是各种主题词表,如标题表、叙词表等。词汇之间的语义关系是先天的,与具体的文献无关,这种关系通常在编表时予以控制和显示。主题语言的句法是指将主题词组合成检索语句以表达文献主题的造句规则。句法关系是后天的,像图1-3中印度、棉纺业等四个词之间的关系,是依具体的文献而存在的,因而标引时要加以控制。关于词汇控制和句法控制,详见第二章。

无论是主题检索语言,还是主题整序法,它们的基本特点是:

- ①按文献所论述的事物(即主题)集中文献;
- ②用人们日常使用的名词术语经过规范化后直接表达文献所论述的事物;
- ③用参照系统等间接显示事物概念之间(即主题之间)的关系;
- ④用字顺序列直接提供按事物名称检索文献的途径(即主题检索途径)。

综上所述,主题法研究的内容包括:

- (1)对词汇控制的原理和方法的研究;
- (2)对主题语言历史发展、结构原理及其性能、特点的研究;
- (3)对各类词表(包括分类主题一体化词表)编制及管理的技术方法的研究;
- (4)对文献主题的结构、特点及检索课题的结构、特点的研究;
- (5)对叙词表、标题表的使用及标引技术的研究,以及对各种类型、各个学科文献标引规则的研究;
- (6)对各文种文献主题描述语句及其表达方式的研究;
- (7)对主题检索工具与主题检索系统的研究;
- (8)对自然语言在情报检索中应用的研究;
- (9)计算机、人工智能技术在主题法领域的应用,包括自动标引、计算机辅助标引、计算机自动编制叙词表等的研究;
- (10)对索引法、汉字排检法及其他相关问题的研究。

* 见张琪玉《情报语言学基础问题选讲》附录五(武汉大学出版社,1987年)。

二、主题法的发展及类型

最早的主题法直接采用自然语言。这可以从我国古代的书、12 世纪流传至今的手抄本《圣经索引》及 17、18 世纪图书馆采用的书名主词款目中找到佐证。类书和圣经索引均是以书中出现的语词为标目,后者是以书名中的主词(Catchword),即相当于现今的关键词为标目。这些不加控制或少加控制的自然语言,成了最早的主题检索语言。从早期的主题法论著——英国克里斯塔多罗(Crestadoro, A.)的《图书馆编目技术》(1856 年)和英国索引之父惠特利(Wheatly, H.)的《什么是索引》(1878 年)中,可以发现有关的论述。

由于主题法直接以事物的名称作为标识,又采用了世界上最重要的精确工具之一的字顺次序,因而在文献管理、索引工作和图书馆工作中得到广泛的应用,反过来又推动了主题法理论和技术的发展。19 世纪在主词款目、书后索引以及字顺分类目录的基础上,出现了标题法。美国图书馆学家克特(Cutter, C. A.)编写的《字典式目录条例》,就是标题法实践的全面总结。所谓标题,即通用语文中定型的事物的名。从这一简短的定义可以看出,标题语言仍然是以自然语言为基础,但是第一要挑选定型的名称,第二对不定型、半定型的名称要进行筛选和加工。苏联学者切尔内(Черный, А. И)等人把标题法命名为“字顺主题分类法”,正好反映出标题法作为传统的主题法与传统的体系分类法一脉相承的关系。

随着阮冈纳赞分析兼综合分类法——《冒号分类法》的问世和 50 年代英国分类法研究小组对阮氏分类理论的研究和普及,主题法领域也受到这种分面分类理论的冲击。在标题表中出现了越来越多的自由浮动复分标题,英国柯茨(Coates, E. J.)等人还把组配技术引入到标题法之中,在检索刊物的主题索引中允许标引员

自由地使用若干主题词组配标引复杂概念。50年代初,美国陶伯(Taube, M.)在40年代英国巴顿(Batten, W. E.)研制的比孔式组配索引法基础上创制了元词法系统。元词法是以最小的字面单元——元词来标引文献,通过元词的组配来检索文献。这可以说是最早的分析兼综合的主题法。元词法首次运用机械(穿孔卡片机、光学比孔卡等)进行标引和检索,可以说开机器检索的先河,为以后计算机运用于主题检索奠定了基础。

60年代初,计算机技术在图书馆及情报工作中的应用,使采众家之长而发展起来的叙词法成为“后起之秀”。以概念组配为基本原理的叙词法淘汰了元词法,在很多图书馆和情报机构中代替了传统的标题法和分类法,成为当代情报检索语言的主流。60年代成为叙词表时代。数以百计的叙词表的出版,推动了计算机检索的发展,主题法的研究也进入一个新的阶段。

早期的叙词表几乎完全是字顺形式的,后来字顺形式的名词一览表的不足之处日益显露出来,分类词表与字顺词表的结合形式被广泛采用了。到了60年代末,英国叙词表专家艾奇逊(Aitchison, J.)终于研制成功世界上第一部分类主题一体化词表——《分面叙词表》,首次建立了一个真正实现了分类法和主题法二者合一的系统。美国兰开斯特(Lancaster, F. W.)曾经预言,这种“具有全面优点”的词表将可能成为受控词表的楷模。在它的影响下,果然出现了一批这样的词表,并有力地推动了主题法兼容化和计算机化的发展。

计算机的广泛应用既推动了叙词检索语言及分类主题一体化检索语言的发展,同时也促进了自然语言的回归。早在1958年美国卢恩(Luhn, H. P.)等人就发表了用计算机编制关键词索引的论文。随着机编关键词索引的发展,自动标引、全文检索、自然语言与受控语言结合的混合系统等,都迅速地发展起来。目前,这种不用受控词表、不用主题分析、不用人工标引的关键词法正在与叙词