



普通高等教育“十一五”国家级规划教材

文献分类法 主题法导论

(修订版)

马张华 侯汉清 薛春香 编著

国家图书馆出版社



国家重点图书出版项目 “十一五”国家重点图书出版规划项目

文献分类法 主题法导论

（第四版）

冯晓波 杨红英 张新科 主编

图书情报学类专业教材

文献分类法主题法导论

(修订版)

马张华 侯汉清 薛春香 编著

 國家圖書館出版社

图书在版编目(CIP)数据

文献分类法主题法导论(修订版)/马张华,侯汉清,薛春香
编著. —北京:国家图书馆出版社,2009. 12

ISBN 978 -7 -5013 -4082 -8

I. 文… II. ①马…②侯…③薛… III. ①分类法—概论
②主题法—概论 IV. G254. 11 G254. 21

中国版本图书馆 CIP 数据核字(2009)第 150259 号

书名 文献分类法主题法导论(修订版)

著者 马张华 侯汉清 薛春香 编著

出版 国家图书馆出版社(原北京图书馆出版社)

(100034 北京市西城区文津街 7 号)

发行 010 -66139745 66151313 66175620 66126153

66174391(传真) 66126156(门市部)

E-mail btsfxb@nlc.gov.cn(邮购)

Website www.nlcpress.com→投稿中心

经销 新华书店

印刷 北京联兴盛业印刷股份有限公司

开本 880 × 1230(毫米) 1/32

印张 13.75

版次 2009 年 12 月第 1 版 2009 年 12 月第 1 次印刷

字数 280 千字

书号 ISBN 978 -7 -5013 -4082 -8

定价 38.00 元

目 录

第一章 导 言	(1)
一、信息资源的组织和内容检索途径	(1)
二、信息组织与检索语言	(5)
三、信息组织的发展	(8)
四、检索语言的类型	(13)
第二章 检索语言原理	(16)
第一节 情报检索中的词汇控制必要性	(16)
一、词汇控制的要求	(17)
二、词汇控制的范围	(19)
三、分类法、主题法词汇控制的不同特点	(24)
第二节 词汇控制和检索效率	(26)
一、检索效率是词汇控制的出发点	(26)
二、词汇控制对检索效率的影响	(28)
第三章 文献分类法	(33)
第一节 分类法概述	(33)
一、文献分类法的意义	(33)
二、文献分类法的类型	(36)
三、学科分类与文献分类	(39)
第二节 分类法结构剖析	(44)
一、类目体系	(44)
二、标记符号	(69)
三、说明、注释和类目索引	(78)
第三节 《中国图书馆分类法》评介	(83)
一、由来与发展	(83)
二、类目体系	(85)

三、辅助表	(89)
四、标记符号	(90)
五、管理	(92)
六、简评	(92)
第四章 主题法	(95)
第一节 主题法概述	(95)
一、主题法的意义	(95)
二、主题法的类型	(96)
三、主题法的特性	(102)
第二节 叙词语言的词汇控制	(104)
一、词类控制	(105)
二、词形控制	(106)
三、词义控制	(109)
四、词间关系控制	(111)
五、先组度控制	(117)
六、词汇性能的测评	(119)
第三节 叙词表的结构	(122)
一、字顺显示	(122)
二、系统显示	(126)
三、图形显示	(132)
四、叙词表的组成模式	(134)
第四节 叙词表的编制与维护	(136)
一、叙词表的编制	(136)
二、叙词表的维护	(140)
三、词表管理软件及其使用	(141)
第五节 《汉语主题词表》评介	(144)
一、由来及发展	(144)
二、主表	(145)
三、附表	(147)

四、辅助索引	(148)
五、概要评价	(151)
第五章 分类主题一体化词表	(156)
第一节 分类主题一体化词表的发展及原理	(156)
一、分类主题一体化词表的由来及发展	(156)
二、分类主题一体化词表的类型和性能	(158)
三、分类主题一体化词表的原理	(161)
第二节 分面叙词表	(163)
一、分面叙词表的由来及发展	(163)
二、分面叙词表的宏观结构	(164)
三、分面叙词表微观结构的改造——词汇控制	(165)
四、分面叙词表的编制	(167)
第三节 分类主题词表	(170)
一、分类主题词表的由来及发展	(170)
二、分类主题词表的原理和结构	(171)
三、分类主题词表的编制	(174)
第六章 文献标引概述	(178)
第一节 文献标引的意义和要求	(178)
一、文献标引的意义	(178)
二、分类标引和主题标引的特点	(179)
三、文献标引质量的测评	(181)
第二节 文献标引的规范和控制	(184)
一、确定标引工具和标引规则	(184)
二、标引方式	(187)
三、标引工作程序	(189)
四、标引人员和标引质量管理	(192)
第三节 主题分析	(195)
一、分析主题类型	(195)
二、分析主题结构	(198)

三、主题分析方法	(201)
四、主题分析误差及克服方法	(206)
第七章 分类标引与检索工具	(208)
第一节 分类标引方法	(208)
一、分类法使用本的确定	(208)
二、号码配置方法	(211)
三、类目辨析	(215)
四、分类法的改用和修订时的改编	(219)
第二节 分类标引规则	(223)
一、基本分类标引规则	(223)
二、一般分类标引规则	(225)
第三节 分类检索工具的组织	(234)
一、分类索书号	(234)
二、分类检索工具的编制	(244)
第八章 主题标引和主题检索工具	(251)
第一节 主题标引方法	(251)
一、查词途径	(251)
二、主题概念的分解转换	(254)
三、标识的确定	(256)
四、《中分表》在主题标引中的作用	(261)
第二节 主题标引规则	(264)
一、主题标引的基本规则	(264)
二、各种主题、文献类型的标引	(270)
第三节 主题检索工具	(284)
一、主题检索系统	(285)
二、主题数据的输入	(293)
第九章 自然语言标引与检索	(305)
第一节 自然语言与受控语言	(305)
一、自然语言与受控语言的比较	(305)

二、自然语言在信息检索中的应用	(310)
第二节 关键词语言	(313)
一、关键词法的由来及发展	(313)
二、关键词语言的构成原理	(315)
三、关键词语言在索引工作中的应用	(316)
四、关键词语言的性能	(320)
第三节 自由标引	(322)
一、自由标引的性质和特点	(322)
二、自由标引的种类	(323)
三、自由标引的基本要求和方法	(324)
四、自由标引的规则	(325)
五、自由标引的缺点及其对策	(327)
第四节 后控制词表	(328)
一、编制后控制词表的必要性	(328)
二、后控制词表的特点和结构	(329)
三、后控制词表的编制	(330)
四、后控制词表轮排索引的编制	(334)
五、后控制词表的编制流程	(335)
第五节 汉语自动切词	(337)
一、词典匹配切词法	(337)
二、设立切词标志法	(338)
三、理解式切词法	(339)
四、歧义切分字段的处理	(340)
第十章 知识组织系统及其在网络环境中的应用	(341)
第一节 从情报检索语言到知识组织系统	(341)
一、知识组织系统的概念及由来	(341)
二、知识组织系统的类型	(344)
三、知识组织系统的特征和作用	(354)

第二节 知识组织系统在网络环境中的应用	(357)
一、分类法在网络信息组织中的应用	(357)
二、叙词表在网络信息组织中的应用	(371)
三、本体在网络信息组织中的应用	(375)
第三节 知识组织系统的互操作及其实现	(388)
一、知识组织系统互操作的概念	(388)
二、知识组织系统互操作的模式	(389)
三、知识组织系统互操作的实例	(394)
第四节 结语	(405)
主要参考文献	(406)
再版修订说明	(409)
一版后记	(410)
中英文简称缩略语表	(412)
名称索引	(415)
主题索引	(421)

第一章 导 言

一、信息资源的组织和内容检索途径

当今的社会是一个信息社会。其特点是:一方面,随着现代科学技术的迅速发展,各种媒体的文献信息急剧增加;另一方面,计算机技术与通信技术的结合使用,进一步加快了信息处理和传输的速度;与此同时,信息的生产以及信息资源的有效开发利用在社会发展中占有越来越重要的位置,成为社会发展的关键因素之一。越来越多的人认识到,在信息爆炸的时代,“原始信息本身并不能产生价值”,只有将其加以有效的组织,按特定的需要集中和揭示,才能产生价值。^①而要使数量巨大、种类多样的文献信息资源得到有效使用,就必须采用相应的文献控制方法进行处理和组织。知识组织正是信息资源有效开发利用的关键之一。

在现代文献中,知识组织一词,具有多种含义。广义的知识组织,能涵盖人类对知识、信息组织化、系统化的各种不同类型的活动,甚至可以包括文章的写作。某一专门领域中的知识组织,则往往有其特定的含义,如人工智能中的知识组织,是指建立一种能反映人类知识结构的推理机制。文献信息资源管理中的知识组织,主要是指以文本及各种类型的信息源为对象,根据使用的需要,通过对其内容特征等的分析、选择、标引、处理,使其成为有序化集合的活动,^②因此亦可称为信息组织或信息资源组织。这一活动目的是为了对信息资源进行有效的开发利用。在当前,一般是在文献处理的基础上,通

① Best, D. The Fourth Resource: In formations and its management. Hampshire. Gower House, 1995:84

② http://www.db.dk/bh/lifeboat_ko/CONCEPTS/knowledge_organization.htm

过编制或建立各种类型检索系统,多层次、多角度地对信息资源进行有效的组织和揭示。

文献处理是对现代文献进行多层次、多角度组织和揭示的基础。其基本做法是:①对文献进行描述。包括:文献著录或编写题录,对文献外部特征进行描述,用于确认文献;②对文献内容进行浓缩。包括编写内容提要或文摘,以使用户能迅速了解文献内容,选择使用;③提供检索点。主要是确定检索内容及可检信息,以便据此对文献信息进行有序化组织。文献著录中描述的文献特征,如:著者名、文献名、出版机构名等,均可作检索点。同时也可以在对主题内容进行分析的基础上确定相应检索标识,包括分类号和主题词。在计算机检索系统中,文摘和提要中的语词亦可以作为检索依据。在完成文献处理的同时,一方面,将文献本身纳入文献库,按一定的方式加以组织;另一方面,将文献处理的结果,即文献记录,组织成相应的检索工具。

文献组织是信息资源组织的重要形式,但它与文献记录亦即检索工具的组织是不同的。以图书馆和文献中心为例,将文献按一定方式进行组织,可以提供一种直接使用文献的检索途径,但同时也会受到文献实体的限制。例:一图书馆如选择按内容之间关系组织文献,就不能同时按照作者或文献名等对文献进行组织;在按内容之间关系组织时,对于同时具有两种以上内容的文献,通常只能按其中一种排列,无法同时按其它内容加以组织;此外,为了方便读者使用,某些文献如工具书、特定对象的文献如以儿童读物、盲人读物等往往需要单独组织;为了合理利用空间,不同媒体的文献,如光盘、录音带、缩微品等一般是分别组织的;这些情况都会造成同一内容文献的分散。至于期刊、报纸中文献的内容,就更无法在文献组织中揭示了。而文献记录作为文献的替代物,则不必考虑文献的载体形式,可以方便地按统一的方式,从多个角度、多种层次组织检索工具和系统,因此检索工具对信息资源的揭示作用更加充分,对信息资源的使用更加重要。

根据对文献信息资源中不同对象和层次揭示的需要,目前文献检索体系是一个由目录、索引以及机读数据库等多种形式的检索工具组成的完备的系统。

传统的目录、索引是经过长期发展建立起来的,从不同层次、不同角度对文献揭示的系统。其中,文献目录是以各种媒体的文献单元为对象,对其进行揭示、报道的工具,包括文献收藏目录和书目。前者以一个或多个文献单位收藏的文献为对象加以揭示;后者则以某一领域、地域或时限内出版的文献为对象进行组织,为用户提供较为全面的文献信息。索引则是以文献或文献集合中包括的信息内容为其揭示对象的工具,分为期刊论文索引、书后索引等。它一般可以给用户提供某一内容、特征的查找线索,其揭示深入到文献所包含的信息单元,将其与目录结合,就构成了对不同层次文献信息的有效揭示体系。

机读数据库是一种依托现代计算机技术,以机读形式建立的检索系统。数据库通常由某种同类记录按照一定的方式组织而成。按照数据的性质,传统的机读数据库通常可以分为文献数据库和非文献数据库两类。文献数据库包括书目数据库和全文库,书目数据库通常收录文献目录或索引数据,是手工文摘索引刊物和文献目录的机读形式;全文库则收入文献全文或部分原文数据,用户可以从多种形式入手查阅文献原文。非文献库则包括事实数据、数值数据、图像数据等多种类型的数据。机读数据库一般数据完备,检索手段先进,并可以通过联机方式查找使用,受到广泛的欢迎。上世纪90年代来出现的网络搜索引擎,则可以看作是机读数据库的一种特殊形式。网络搜索引擎以网络资源为对象,它不仅包括各种类型的网络文本资源,而且还包括聊天室、BBS、网上即时播报等动态资源以及音频、视频资源等,并可以通过超文本链接的方式直接访问网络信息资源本身,是目前发展最为迅速、使用面广的检索工具。

为了满足用户的不同需要,上述各种检索工具在对文献信息进行多层次揭示的同时,一般均提供多种检索途径,方便用户从不同角

度检索查找。

根据所依据的文献检索特征,信息资源组织和检索的途径大致可以分为两种类型:

一类检索途径直接来自文献的外部特征,可以据以查找一种或数种文献。常用的包括:责任者名、文献名、机构名、出版地、出版社、标准书号、专利号、档案号、文献登录号等。这类途径通常可以通过已知的文献特征,得到明确的查找结果。有的特征,如标准书号、专利号,可以据以查出相应的特定文献;其他特征,如责任者名、文献名等,一般可将文献范围缩小到便于查检的数量。与此同时,还可以通过数种特征的组配进行查找,如通过出版者与出版时间、文献名与出版者等的组配,查找特定文献。此外,还有一种根据文献的引证关系建立的文献检索途径。这种方法将文献写作过程中引用的文献作为组织检索系统的依据,编制引文索引,供用户从引证关系入手对文献进行检索。这一途径也可以通过文献提供的数据直接进行编制。

另一类检索途径则来自文献的主题内容,包括分类法和主题法两种形式。分类法以表达文献内容的标记符号为标识,主题法直接以表达文献内容的语词为标识,分别依据主题之间关系或主题字顺的方式提供检索途径。使用此类途径进行检索,不必预先了解具体的文献特征,可以直接从所需要的主题内容入手进行查检。例如,在主题检索系统中,可以直接使用“信息管理”一词进行检索,获取论述该主题内容的有关文献。同时还可以根据“信息管理”这一概念与其他主题词的联系,进行相关内容文献的查找。

从上面的介绍可以看出,上述两类检索途径的特点是不同的,前一种途径可供用户从已知文献数据的角度进行查找,后一类途径则可以根据用户的需要,从需要的主题内容出发对文献进行检索。比较而言,前一种途径可以迅速、准确地查找特定文献,但需要预先掌握确切的数据;后一种途径则不必事先了解相应的数据,用户不仅可以根据使用需要直接检索特定主题内容的文献,而且还可以通过系统中主题内容之间的联系,扩大或缩小检索范围,进行相关文献的查

找,有利于信息资源的开发利用,是更为重要的检索途径,因而受到广泛的重视。本书将主要对后一类检索途径的理论方法进行介绍。

二、信息组织与检索语言

如前所述,主题内容角度的检索途径是信息组织中一种依据其内容特征加以组织与揭示的方法。由于这一方法依据的对象为文献主题内容,因此国外有人将其概括地称为主题途径(subject approach);又因为这种方法一般需要对主题内容进行归纳、分析,在概括为类的基础上进行的,因此也将其统称为广义的分类法;此外,还有人因这一方法的操作过程是与主题分析密切联系的,因而把它称为主题分析(subject analysis)。

与其他检索途径相比,内容检索途径对文献处理的不同之处在于:首先,其揭示的对象不是外在的形式,而是文献内容,因此不能简单从文献外部特征中选取,必须根据检索需要,通过对文献内容的分析、选择,概括出文献论述的事物对象,再以一定的标识,如主题词或分类号集中加以表达;其次,在赋予标识的同时,还必须通过对标识的组织,将文献记录组织成有序的信息检索系统,在这一系统中,每一个标识亦即文献集合有其固定的位置,能够以各种方式显示出不同主题内容之间相互联系,从而使用户不仅可以通过这一系统对特定主题内容的文献集合进行查找,并可以根据这一系统扩大或缩小检索范围,进行相关内容的查找。要达到这一目的,通常必须根据一定的知识结构预先建立起一个标识系统或组织标识系统的规范。因此,在实际操作中,从内容角度进行组织往往是按照预先确定的概念标识系统或规则,对文献内容进行分析和处理的,通过对文献赋予系统规定的相应标识,将文献纳入这一标识系统。用户使用,一般也需按照该系统的要求,将其信息需求转换成一个或多个标识,再通过检索系统中标识的相符性比较,实施文献检索(见图1-1)。

从上面的分析可以看出,从内容角度进行组织实际上具有两重含义:

其一,信息组织方法。即是一种通过依据文献主题内容或特征,将文献组织为相应检索系统的方法。这种方法一般依据一定的信息标识系统处理文献、组织检索工具,包括存储与检索两个方面。在存储阶段,通常需要通过主题标引操作,在文献主题内容提炼的基础上,以一定的词表或分类表为工具,将主题分析的结果转换成相应标识,纳入文献库或检索系统。用户检索时,则依据该检索系统的特点,首先将检索提问转换成相应标识,再根据系统提供的途径进行匹配查找。

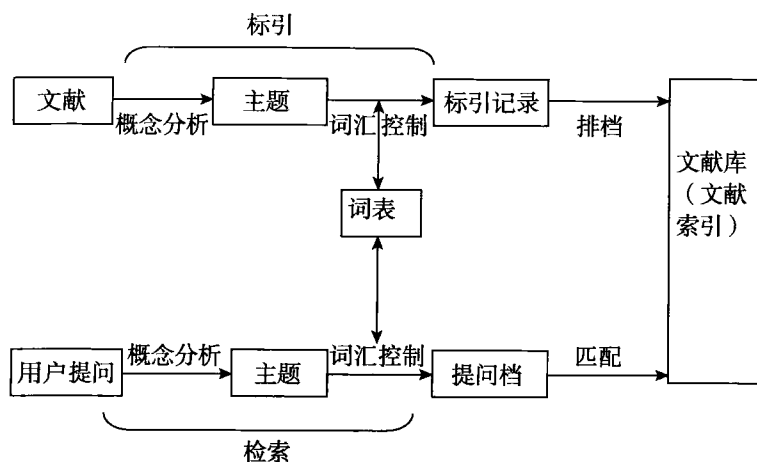


图1-1 信息组织与信息检索语言

其二,信息组织语言。通常指标引语言(indexing language)或情报检索语言(information retrieval language)。这是一种由表达文献主题概念及其相互关系的词汇及规则组成的人工语言系统。这一语言系统是根据检索系统的特点,按照使用需要和对知识关系的理解预先确定的,它是信息组织的体现,也是文献标引和检索的依据。从国内对这两个术语的限定看,除了它们的侧重点有所区别外,其基本内容并没有本质不同。从本书讨论的内容重点而言,使用标引语言这一术语可能是更适合的。但考虑到国内对检索语言一词接受的普遍程度,本书在下面对相应内容统一采用情报检索语言或检索语言一

词加以表达。

如同自然语言一样,检索语言由词汇和语法两部分组成。其中,词汇集是检索语言的主体,一般用于对所使用的词汇及其相关性进行规定,是文献存储和检索的语言保障。其表现形式是各种词表,包括分类表或标题表、叙词表等。所谓语法,则是分类标引和主题标引的规则系统,它们规定词汇集的使用方法,是文献标引必须遵守的规范。自由标引属于自然语言应用,不属于受控语言标引。

一般来说,无论是作为信息组织方法还是检索语言,一般都具有以下特点:

- ①采用概念标识对文献内容进行类聚和描述;
- ②从一定的角度出发组织和集中文献;
- ③根据系统的需要,揭示主题概念之间的相关性;
- ④对复合主题的标引和检索的方法加以规定;
- ⑤以一定方式排序,提供检索途径。

内容检索途径作为从内容角度对知识进行组织和揭示的方法,历来是文献信息资源组织化、结构化的基本手段,对它的研究是与社会需要和技术改进密切联系的,也是与知识的发展以及人们对知识之间关系的了解相一致的。随着现代信息技术发展的日益推进,其研究受到社会的广泛重视,研究领域也在不断扩大。在目前情况下,其研究领域一般包括下述内容:

- ①信息组织基本理论方法的研究,包括词汇控制理论、分类理论、概念逻辑理论、文献主题结构层次理论等的研究;
- ②信息组织的发展及其类型的研究;
- ③类表、词表的基本理论及方法研究,包括结构原理、性能、特点、编制及管理的技术方法等的研究;
- ④各种类型类表、词表研究,包括各种综合性类表、词表研究,专业性类表词表研究等;
- ⑤标引理论与技术的研究,包括各种标引方式、类型研究以及各种文献、主题类型标引方法的研究等;