

情报检索语言

张琪玉 编著

国家图书馆出版社

20 世纪中国图书馆学文库·36

情报检索语言

张琪玉 编著

圖 國家圖書館出版社

本书据武汉大学出版社 1983 年 6 月第 1 版排印

前 言

情报检索语言是专门用于各种手工的和计算机化的文献情报存贮检索系统,表达文献主题概念和检索课题概念的人工语言,包括体系分类法、组配分类法、标题法、单元词法、叙词法、关键词法以及代码语言等多种类型。这类语言之对于图书情报工作,如同自然语言之对于人类社会生活一样重要。检索系统效率的高低,在很大程度上取决于所采用的情报检索语言的质量以及对它的使用是否正确。

从图书情报工作现代化的需要看,作为受过高等学校图书馆学或情报学教育的专业人才,除了必须掌握体系分类法这种传统的情报检索语言的知识外,还有必要掌握更多的情报检索语言基本知识,以便在工作中能正确地使用这类语言,以及向读者普及这类语言的知识,乃至进一步有所研究;结合实际需要改进现用的语言和创制新的语言。

由于情报检索语言知识的用途比较广泛,武汉大学图书馆学系从1977级起,为图书馆学专业和科技情报学专业的本科生、函授生、走读生以及部分研究生、进修生普遍开设了《情报检索语言》课程。从1980年上半年以来,已对800多名学生讲授了该课程知识。本书即是为该课程编写的教材,现被列入教育部主持制订的《高等学校文科教材编选规划》出版。

本书以检索效率问题为中心,讲述和讨论了情报检索语言的

基本理论、各种类型情报检索语言的原理、编制法、使用法、性能、在方法上互相吸取的情况,以及文献分析与标引的一般问题,力求使学习者获得有关情报检索语言的比较完整、系统的基本知识,并对各种类型语言所存在问题的解决有所思考。

本书初稿写成于1980年3月,这是第二次修订本。但是,将各种类型情报检索语言知识统一于一门课程中,连贯起来作分析比较式的讲授,这还是一种探索。编者深知,本书作为一种教材还不够完善,虽经两次修改,一定还有很多缺点。现借出版的机会,恳请专家和读者批评指正,使它能得到改进。

本书在编写和修改过程中参考了一些文献(见参考文献目录),在这里我谨向作者致以谢忱!本书在出版过程中曾得到图书馆学专业教材编审组关懿娴、彭斐章、来新夏等同志,不辞辛苦前来参加审稿会议的曾世荣、韩承铎、范世伟、黄万新、史永元同志,提出书面意见的许培基同志,本系孙冰炎、周继良、杨元生同志,武汉大学出版社刘文华同志等的支持、指正和帮助,对此我表示深切的谢意!

张琪玉

1982. 12. 24 于武汉大学

目 录

第一章 绪论	(1)
第一节 情报检索语言在情报检索中的作用	(1)
第二节 对情报检索语言的共同要求	(4)
第三节 概念逻辑和知识分类是情报检索语言的基础	(11)
第四节 情报检索语言的分类	(22)
第二章 等级体系分类语言	(25)
第一节 体系分类法的原理	(25)
一 体系分类法的构成原理	(25)
二 类目的划分	(27)
三 类目的排列	(31)
四 类目名称及涵义	(35)
五 学科之间相互关系的处理	(38)
六 某些图书资料分类的特殊性	(41)
第二节 体系分类表	(42)
一 分类表的功用和结构体系	(42)
二 分类表的编制法	(44)
三 分类表的设计和编制程序	(59)
四 分类表的使用法	(61)
五 分类表的管理	(63)
第三节 体系分类法的性能	(65)
一 体系分类法的检索效率	(65)
二 体系分类法的某些问题	(66)

第三章 分析 - 综合分类语言	(75)
第一节 组配分类法的原理	(75)
一 组配分类法的构成原理	(75)
二 “面”的划分和细分	(80)
三 类号的组配和排列	(82)
四 组配分类法的类型	(84)
第二节 组配分类表	(86)
一 组配分类表的结构体系	(86)
二 组配分类表的编制法	(87)
三 组配分类表的使用法	(88)
第三节 组配分类法的性能	(90)
一 组配分类法与体系分类法的比较	(90)
二 组配分类法的某些问题	(91)
第四章 标题词标引语言	(94)
第一节 标题法的原理	(94)
一 标题法的构成原理	(94)
二 标题及其类型	(96)
三 标题词的范围及选词原则	(99)
四 标题词的规范	(100)
五 标题词相互关系的显示	(106)
第二节 标题表	(112)
一 标题表的功用和结构体系	(112)
二 标题表的编制法	(113)
三 标题表的使用法	(114)
四 标题表的管理	(122)
第三节 标题法的性能	(124)
一 标题法与体系分类法的比较	(124)
二 标题法的某些问题	(126)
第五章 单元词描述语言	(132)
第一节 单元词法的原理	(132)

一	单元词法的构成原理	(132)
二	单元词的范围及选词原则	(134)
三	单元词的规范	(135)
第二节	单元词表	(138)
一	单元词表的功用、结构体系和编制法	(138)
二	单元词表的使用法	(139)
第三节	单元词法的性能	(141)
一	单元词法与标题法、组配分类法的比较	(141)
二	单元词法的某些问题	(143)
第六章	叙词描述语言	(147)
第一节	叙词法的原理	(147)
一	叙词法的构成原理	(147)
二	叙词的范围及选词原则	(151)
三	叙词的组配	(154)
四	叙词的规范	(163)
五	叙词相互关系的显示	(172)
第二节	叙词表	(183)
一	叙词表的功用和结构体系	(183)
二	主表的编制法	(187)
三	辅表的编制法	(194)
四	叙词表的设计和编制程序	(203)
五	叙词表的使用法	(207)
六	叙词表的管理	(212)
第三节	叙词法的性能	(213)
一	叙词法与单元词法、标题法、组配分类法、体系分类法的 比较	(213)
二	叙词法的某些问题	(215)
第七章	关键词描述语言	(224)
第一节	关键词法的原理	(224)
一	关键词法的构成原理	(224)

二	关键词法的表现形式	(225)
三	关键词的范围	(232)
四	关键词的规范	(233)
第二节	关键词法的性能	(234)
第八章	文献分析与标引问题	(237)
第一节	文献的分析与标引是影响情报处理质量的关键性工序	(237)
一	保证分析标引质量的重要性	(237)
二	评价分析标引质量的标准	(238)
第二节	文献分析与标引的一般方法和规则	(239)
一	分析水平与标引深度	(239)
二	分析角度	(240)
三	分析方法	(241)
四	标引方式方法	(241)
五	导致检索效率降低的分析标引误差	(243)
第三节	文献分析与标引的某些问题	(244)
一	适度标引问题	(244)
二	集中标引问题	(246)
三	利用已有标引成果问题	(246)
	参考文献目录	(248)

第一章 绪 论

第一节 情报检索语言在情报检索中的作用

情报检索语言是根据情报检索的需要而创制的人工语言。又称情报语言、检索语言、情报存贮与检索语言、文献语言、文献工作语言、索引语言、标引语言、标引符号、标识系统,等等。

目前世界上有成千种情报检索语言。例如,《中国图书馆图书分类法》、《汉语主题词表》、《国际十进分类法》、《杜威十进分类法》、《NASA 叙词表》等,都是情报检索语言的一个语种;对图书文献资料所标的分类号或检索词,即是情报检索语言的语词。有数以千万计的图书馆、情报机构、档案部门以及检索刊物的工作人员和读者,正在使用着这类语言。

情报检索是情报传递中的重要环节。在当代科学技术文献数量庞大,内容专深且互相渗透,而科研、生产、教学、行政部门对情报的需要十分迫切,检索课题又很专指的情况下,情报检索问题尤其重要,并变得相当复杂;为要达到较高的检索效率,就必须作出很大的努力。

情报检索语言与情报检索效率有最密切的关系。它在情报检索过程中所起的作用是极为重要的。

情报检索的全过程包括情报的存贮和检索两个方面。情报存贮是指编制检索工具和建立检索系统;情报检索即是利用这些检

索工具和检索系统来查找所需的情报。所以,情报的存贮和检索是两个有着密切联系的过程,其关系如图1。

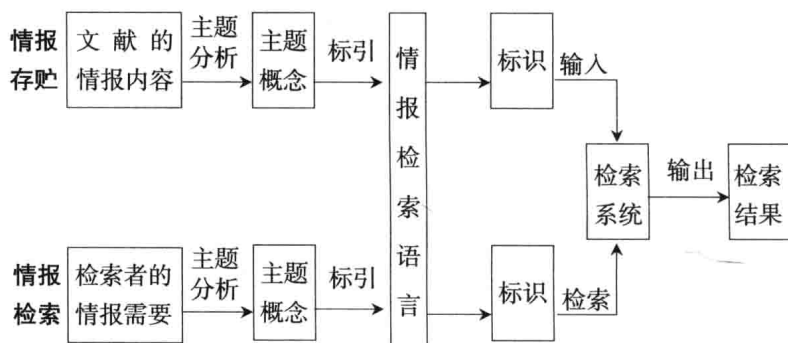


图1 情报存贮与检索过程及情报检索语言在其中的作用

当存贮情报时,文献标引人员首先要对各种文献进行主题分析,即把它所包含的情报内容分析出来,使之形成若干能代表文献主题的概念,并用情报检索语言的语词(标识)把这些概念标示出来,然后纳入检索工具或检索系统。

当检索情报时,情报检索人员也首先要对检索课题进行主题分析,即把它所涉及的检索范围明确起来,使之形成若干能代表情报需要的概念,并把这些概念转换成情报检索语言的语词(标识),然后从检索工具或检索系统中查找正是用该语词标引的文献,从而找到包含有所需情报的文献。

由此可见,情报检索语言是情报检索系统的重要组成部分,在情报检索系统中起着语言保障的作用,是沟通情报的存贮和检索两个过程,标引人员和检索人员双方思想的桥梁。如果没有情报检索语言作为标引人员和检索人员的共同语言,就很难使标引人员对文献情报内容的表达(标引用语)和检索人员对相同内容的情报需要的表达(检索用语)取得一致,情报检索也就不可能顺利

实现,甚至根本不能实现。

情报检索语言作为情报检索专用语言的特点,是能简单明白又比较专指地表达文献及检索课题的主题概念,容易将概念进行系统排列,在检索时便于将标引用语和检索用语进行相符性比较;语词与概念一一对应,排除了多词一义、一词多义和词义含糊的现象,并且还能显示出概念之间的相互关系,从而能帮助情报检索人员又全、又准、又快地检索到含有所需情报的文献。

情报检索语言就其实质来说是表达一系列概括文献情报内容的概念及其相互关系的概念标识系统。它可以是从自然语言中精选出来并加以规范化的一套词汇,可以是代表某种分类体系的一套分类号码,也可以是代表某一类事物的某一方面特征的一套代码(如化合物的各种代码),用以对文献内容和情报需要进行主题标引、特征描述或逻辑分类。

情报检索语言由词汇和语法组成。词汇是指登录在类表、词表中的全部标识,一个标识(分类号、检索词、代码)就是它的一个语词,而分类表、词表则是它的词典;语法是指如何创造和运用那些标识(单个标识或几个标识的组合)来正确表达文献内容和情报需要,以有效地实现情报检索的一整套规则。

情报检索语言的主要功能大致可归纳为如下四点:

- (1)对文献的情报内容(及某些外表特征)加以标引;
- (2)对内容相同及相关的情报加以集中或揭示其相关性;
- (3)对大量情报加以系统化或组织化;
- (4)便于将标引用语和检索用语进行相符性比较。

任何一种情报检索语言,都要采取一定的方法和手段,将它的各种要素按照一定的结构结合成一个有机的统一体,以便在图书馆工作、情报工作、档案工作中发挥这四项功能;并不断改进它的结构,以发挥最佳功能。

情报检索语言的质量高低及其使用正确与否,对情报检索效

率有重大影响。所以,如何提高检索效率,特别是提高检全率和检准率以及情报检索计算机化程度,是现今情报检索语言研究的主要目标。

第二节 对情报检索语言的共同要求

检索效率是由许多因素构成的,所以对情报检索语言也必须提出多方面的要求:

1. 保证较高的检全率和检准率。或者说,漏检率和误检率要能控制在允许的范围内。

最高的检全率是指对各种文献用某种情报检索语言标引并纳入检索工具或检索系统之后,在需要时能将其中载有有关情报的那些文献全部检出。事实上,很难达到全部检出的要求,而只能达到某个百分比。既然检全率一般只能达到某个百分比,就会有一部分有关的文献未能被检出,这一部分文献的数量可用漏检率来表示。与某项情报需要有关并且被检出的文献量和虽有关但未能被检出的文献量之和,等于有关该项情报的全部文献量。由此可见,检全率和漏检率是可以说明同一问题的两个相对的概念。检全率高则漏检率就低,而漏检率高则检全率就低;如果检全率为60%,则漏检率就是40%。检全率是情报检索质量的主要标志。因为漏检率高表明重要情报被遗漏的危险性大,所以在情报检索中,总是力求提高检全率,或者说力求降低漏检率。

最高的检准率是指检出的文献全部与情报需要有关。事实上,在情报检索中,当检出与某项情报需要有关的文献的同时,常常会带出一些无关的文献。要完全避免带出无关文献也是很难达到的,所以检准率也只能达到某个百分比。换句话说,会存在一定的误检率。与某项情报需要有关并且被检出的文献量和虽无关但

也被带出的文献量之和,等于被检出的全部文献量。由此可见,检准率和误检率也是可以说明同一问题的两个相对的概念。检准率高则误检率低,而误检率高则检准率低。因为误检率高表明用于甄别所检出文献是否真正有用的劳动的增加和时间的浪费,虽不如遗漏重要情报那样严重,但也会影响检索效率,所以在情报检索中,总是力求提高检准率,或者说力求降低误检率。

关于检全率、检准率、漏检率、误检率的概念见下面的图示及公式(公式有助于明确概念,虽然可以应用于评价试验,但实际很难进行十分精确的计算):

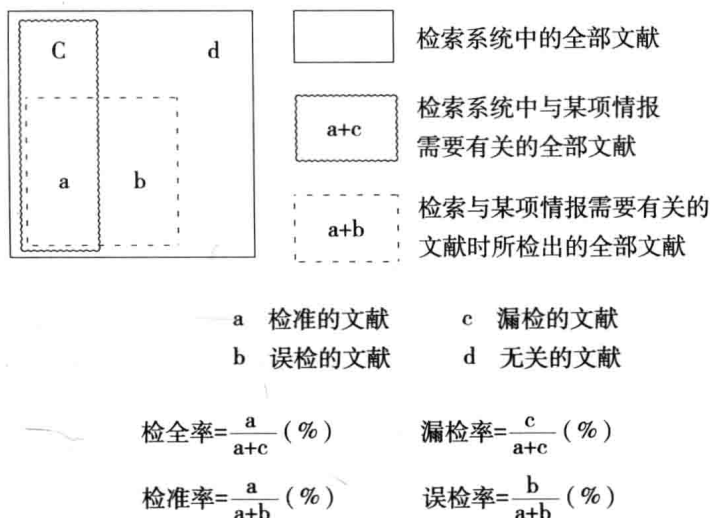


图2 检全率、检准率、漏检率、误检率概念的图示及公式

保证较高的检全率和检准率,是对情报检索语言的主要要求。

检全率主要与一种情报检索语言是否能比较全面地显示概念之间的等级关系和相关关系,以及在它的词汇中是否能排除同义

现象有关;检准率主要与它的标识(单个标识或若干标识的组合)是否能达到较高的专指度,在它的词汇中是否能排除多义现象和同形异义现象,以及它是否能正确地、恰如其分地显示概念之间的关系有关。情报检索语言的选词和列类是否能符合文献主题的实际情况和情报检索的实际需要,则与检全率和检准率都有关。

实践证明,在检全率和检准率之间存在着互相制约的现象,即提高检全率会使检准率下降,提高检准率会使检全率下降。这主要与标引过程和检索过程有关,但也与情报检索语言本身有关。例如,提高标识的专指度,可提高检准率,但会影响检全率;若降低标识的专指度,则可提高检全率,但又会影响检准率。不过应当指出,并不是情报检索语言所采取的一切方法和手段都会出现这种现象。例如,在它的词汇中排除同义现象只会提高检全率,不会影响检准率;排除多义现象和同形异义现象只会提高检准率,不会影响检全率。

检全率和检准率除与以上诸因素有关外,还与下列许多因素有关。

2. 能满足多种检索要求。社会的情报需要是多种多样、千变万化的,即使是同一个人,也会提出不同类型的检索要求。区分检索要求的类型,有助于了解情报检索语言的性能。典型的检索要求大体可分下列五种:(1)检索某一事物的某一方面文献;(2)检索某一事物的全部有关文献;(3)检索许多事物的同一方面文献;(4)检索一类事物的全部有关文献;(5)浏览一个学科或专业范围的文献。各种情报检索语言在满足这五种典型的检索要求上是有很大差异的。往往是一种语言对这一种检索要求的适应性比较好,而对那一种检索要求的适应性却比较差,另一种语言的情况则又不同,但总不能很好地同时满足这些检索要求。从检索系统完善性、使用方便性和经济性的观点来看,一种语言能同时满足多种检索要求是非常必要的。

情报检索语言对各种检索要求的适应性,主要与它们的构成原理有关,即与它们用什么方式来构成概念标识,以及按哪些原则、方式来排列、组织概念标识和显示概念之间的关系有关。

一种情报检索语言是否能满足多种检索要求,主要看它的下列各种性能如何:(1)是否既能从学科、专业出发进行检索,又能从事物出发进行检索;(2)是否既能较好地进行族性检索,又能较好地进行特性检索;(3)是否能视检索过程中出现的具体情况而自由地扩大、缩小或改变检索范围;(4)是否能进行多途径检索;(5)是否能进行多因素检索,即按任何特征组配检索,或者用布尔代数表达检索课题和进行逻辑运算,实现精确检索。

族性检索是范围较广泛的检索。前述第四种检索要求和第五种检索要求可以认为是族性检索要求。特性检索是范围较狭窄的检索。前述第一种检索要求可以认为是特性检索要求。至于第二种检索要求和第三种检索要求,相对于第一种检索要求来说,可以认为是族性检索要求;相对于第四种检索要求来说,可以认为是特性检索要求。总之,所谓族性检索和特性检索是两个相对的概念,在它们之间没有明确的界限。一般认为,从学科、专业出发的检索要求是族性检索要求,从事物出发的检索要求是特性检索要求;认为分类法系统各种语言的职能主要是满足族性检索的要求,主题法系统各种语言的职能主要是满足特性检索的要求。其实,无论是从学科、专业出发的检索(可以称学科检索或分类检索),还是从事物出发的检索(可以称事物检索或主题检索),都有族性检索,即范围较广泛的检索,和特性检索,即范围较狭窄的检索。任何情报检索语言,都应采取一定的方法和手段,力求既适合于族性检索的要求,也适合于特性检索的要求。

3. 易于标引,易于检索。是否易标易检,不仅与标引、检索的速度有关,而且与标引、检索的质量有关。因为一种情报检索语言如果不是易于为标引人员和检索人员所掌握,必将导致标引误差

和检索误差的增加。保证易标易检的条件是多方面的,包括:(1)语词或符号含义的明确性。例如,在体系分类法中通过类名措词、等级隶属结构(形成语言环境)、类目注释、类目划分规则等来明确分类号的含义;(2)语法的严密性。即凡是会出现分歧的地方都要有明确规定,规则不能模棱两可和自相矛盾。例如,在叙词法中明确规定“凡词表中已有专指词者不得用泛指词组配标引”,“凡可以用组配标引又可以用上位词标引者优先用组配标引”;(3)标识的直观性及其排列次序易于理解。在这方面,语词标识比符号标识直观性好,先组式的标题词比组配式的叙词直观性好,组配分类法的分类号比体系分类法的分类号直观性好,层累制的分类号比顺序制的分类号直观性好;号码或语词的排列,应尽量避免特殊的、使人不易理解的规定;(4)查词查号手段的多样性。例如,体系分类表编有类目索引,叙词表编有分类索引、等级索引、轮排索引、双语种对照索引等,使标引人员和检索人员从各种方便的角度都能查到所需的标识;(5)整个语言包含概念的丰富和完备性。所谓包含概念丰富,如体系分类表中类目多,以类目注释方式列出的概念多;叙词表中叙词多,作为检索“入口”的非叙词多等。所谓包含概念完备,是指不论任何主题的文献都能有类可归,有适当的语词可以用来标引,即使是用泛指类目和泛指词也可以。

4. 具有对先进的检索方式和检索设备的适应性,以及对多种检索方式和检索设备的适应性。

检索方式即文献情报存贮与检索方式,可以分为文献单元方式和标识单元方式。文献单元方式也称顺检方式、顺排档,其特点是文献标识直接加在文献款目上作为标目,文献款目即按文献标识排列,查到文献标识即可见到文献著录事项。我国图书馆和情报机构目前普遍使用的卡片式目录都是属于文献单元方式。标识单元方式也称逆检方式、倒排档,其特点是检索工具分两部分,一部分是文献卡(文献款目),另一部分是标识卡,标识卡上只记有