



高等学校图书馆学核心课教材

情报检索语言实用教程

张琪玉 主编 侯汉清 副主编

0
1

武汉大学出版社

高等学校图书馆学核心课教材

情报检索语言实用教程

张琪玉 主 编

侯汉清 副主编

武汉大学出版社

图书在版编目(CIP)数据

情报检索语言实用教程/张琪玉主编;侯汉清副主编. —武汉:武汉大学出版社,2004.7

高等学校图书馆学核心课教材

ISBN 7-307-04219-3

I. 情… II. ①张… ②侯… III. 检索语言—高等学校—教材
IV. G254.0

中国版本图书馆 CIP 数据核字(2004)第 044104 号

责任编辑:严红金鑫等 责任校对:黄添生 版式设计:支笛

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件:wdp4@whu.edu.cn 网址:www.wdp.whu.edu.cn)

印刷:武汉市新华印刷有限责任公司

开本:880×1230 1/32 印张:11.625 字数:318千字 插页:2

版次:2004年7月第1版 2004年7月第1次印刷

ISBN 7-307-04219-3/G·671 定价:18.00元

版权所有,不得翻印;凡购买我社的图书,如有缺页、倒页、脱页等质量问题,请与当地图书销售部门联系调换。

内 容 提 要

为了适应近年来高等学校图书馆学、信息管理学等专业教学改革的新趋势,减少情报语言学领域诸课程之间的重复,精简讲授内容、突出实用性,本教程从原有诸课程中选取最必要讲授的知识作重点讲授,减少了实际工作中不常用的知识,内容包括情报检索语言及情报语言学概论、等级体系型分类检索语言、叙事型主题检索语言(包括分类主题词表)、情报检索语言构造基本方法、自然语言在情报检索中的应用、文献数据库及计算机检索技术、互联网上资源的检索等。关于情报检索语言应用,以讲解其在计算机检索系统(数据库)方面的应用为主。最后设“进一步学习指导”一章,也颇有特点。

本教程供高等学校图书馆学、信息管理专业本科教学使用,也可供图书馆、情报机构、档案部分工作人员参考。

前 言

本教程系受武汉大学出版社委托编写,供高等学校图书馆学专业、信息管理专业等本科教学使用。编写目的主要是减少情报语言学领域诸课程之间的重复,精简讲授内容,突出实用性,以适应近年来教学改革发展的新趋势。

本教程从原“情报语言学基础”、“图书分类学”、“主题法导论”及其他同类课程中选取比较实用的、最必要讲授的知识作重点讲授,减少在实际工作中不常用的情报检索语言类型和语种、历史及其他知识的讲授,其基本内容分为如下七章:第一章 情报检索语言及情报语言学基本问题;第二章 等级体系型分类检索语言及其应用;第三章 叙词型主题检索语言(叙词法)及其应用;第四章 情报检索语言构造基本方法;第五章 自然语言在情报检索中的应用;第六章 文献数据库与计算机检索技术;第七章 因特网信息资源的检索。有些非重点但仍有一定需要的知识则附在有关章节中作简略讲授。

关于情报检索语言的应用,以讲解其在计算机检索系统(数据库)方面的应用为主,在手工检索工具(卡片式和书本式目录索引)方面的应用比原有的教材简略,以适应文献检索领域的发展趋势。

本教程在最后设有“进一步学习指导”一章,分相

应各章编写，内容包括：进一步学习和研究的课题、课题内容简介、学习和研究方法、参考文献等，为希望在情报检索语言及相关领域深造的学习者提供通过自学进一步扩展知识和深入学习的途径以及可读的主要文献。这一章不作为必讲和考试的内容。

本教程作如上安排，是企图达到：保障必要的基本理论（第一章和第四章），突出最常用的技术和方法（第二章和第三章），扩大相关的知识（第五章、第六章和第七章），对所精简的内容作适当的处理（第八章），既精简、突出实用性，又不致使学习者陷于只知技术不理解原理以及缺乏相关知识的偏漏，使各个方面能妥善安排。但这只是初步尝试，还有待教学实践来不断检验。希望使用本教材的单位和教师，能将意见和建议寄给我们，以便修改。

本教程主要作者情况如下：

张琪玉：南京政治学院上海分院信息管理系

侯汉清：南京农业大学信息管理系

陈树年：华东理工大学图书馆

曹树金：中山大学信息管理系

本教程编写计划和大纲由张琪玉草拟，第一、第四、第五章由张琪玉编写，第二章由陈树年编写，第三章由侯汉清编写，第六章由曹树金、胡岷编写，第七章由曹树金、杨涛编写，第八章由各人编写相应部分，最后由侯汉清担任全书统稿。

张琪玉

2003年1月6日于上海

目 录

第一章 情报检索语言及情报语言学基本问题	1
第一节 情报检索语言及其功能	1
一、建立情报检索系统是开发利用知识信息资源的 基本措施	1
二、情报检索系统与情报检索语言	2
三、情报检索语言的基本功能	8
第二节 情报检索语言对检索效率的影响	9
一、检索效率概念	9
二、情报检索语言编制过程对检索效率的影响	12
三、情报检索语言使用过程(文献标引过程) 对检索效率的影响	15
四、情报检索语言使用过程(情报检索过程) 对检索效率的影响	15
第三节 情报检索语言的基础——术语学、 概念逻辑和知识分类	15
一、术语学与情报检索语言	15
二、概念逻辑与情报检索语言	16
三、知识分类与情报检索语言	21
第四节 情报检索语言的表达对象——文献主题	22
一、文献主题的构成因素及层次	22
二、文献主题的类型	26
第五节 情报检索语言的类型	26
一、按其结构原理分	27

二、按所包括的学科、专业范围分	28
三、按类目或检索词所用自然语言分	28
四、按其标识的组合使用方法分	28
第六节 情报语言学的一般问题	28
一、情报语言学的研究对象和研究目的	28
二、情报语言学的研究范围	29
三、情报语言学的研究材料	29
四、情报语言学的研究方法	30
第二章 等级体系型分类检索语言 (体系分类法) 及其应用 ..	32
第一节 体系分类法的构成原理与检索性能	32
一、体系分类法的构成原理	32
二、体系分类法的检索性能	46
第二节 体系分类法的结构功能及《中国图书馆分类法》 ..	48
一、体系分类法的宏观结构与功能	49
二、体系分类法的微观结构与功能	50
三、《中国图书馆分类法》概述	52
第三节 体系分类法的使用	63
一、针对体系分类法的文献主题分析	63
二、体系分类法的标引规则	65
三、体系分类法的组号技术	75
四、体系分类法的标引技巧	84
五、用于组织目录的分类和用于组织排架的分类	94
六、用于机读文档的标引方法	96
第四节 分类标引工作的管理	101
一、文献分类标引的程序	101
二、制订分类法使用本与分类细则	103
三、分类标引的质量管理	105
四、文献重新分类工作的组织	107
第五节 分类目录和分类排架	111
一、分类目录及同类书区分	111
二、分类排架及索书号	119

第三章 叙词型主题检索语言（叙词法）及其应用	124
第一节 叙词法的原理及性能	124
一、叙词法的构成原理	124
二、叙词法的词汇控制	126
三、叙词法的性能	136
第二节 叙词表的结构及其功能	138
一、叙词表的性质和功能	138
二、叙词表的微观结构	138
三、叙词表的宏观结构	141
第三节 叙词法的应用	148
一、《汉语主题词表》概述	148
二、叙词法的文献标引方法	150
三、主题目录和检索刊物的主题索引	178
第四节 分类法主题法一体化词表	190
一、分类法主题法一体化词表概述	190
二、分面叙词表	195
三、分类主题词表	204
四、《中国分类主题词表》及其应用	209
第四章 情报检索语言构造基本方法	218
第一节 情报检索语言结构与功能的关系	218
一、情报检索语言的宏观结构	218
二、情报检索语言的微观结构	219
第二节 情报检索语言为实现四项基本功能在其构造中	
所采用的基本方法	220
一、情报检索语言为实现“对文献的情报内容（及某些外表特征）	
加以标引”这项功能所采用的方法	220
二、情报检索语言为实现“对内容相同及相关的情报加以集中或	
揭示其相关性”这项功能所采用的方法	225
三、情报检索语言为实现“对大量情报加以系统化或组织化”	
这项功能所采用的方法	228
四、情报检索语言为实现“便于将标引用语和检索用语进行	

相符性比较”这项功能所采用的方法	229
第五章 自然语言在情报检索中的应用	232
第一节 自然语言的应用是当今情报检索领域的一种 重要发展趋势	232
第二节 自然语言在情报检索中应用的各种方式	233
一、不标引(无标引)方式	233
二、自动抽词标引方式	234
三、关键词索引方式	239
四、人工自由标引方式	243
五、以自然语言作为自由词,与情报检索语言并用的方式	247
六、自然语言接口方式	247
七、用后控制词表弥补各种自然语言应用方式的不足	248
第三节 自然语言检索中各种因素对检索效率的影响	254
一、作为检索依据的文本的类型对检索效率的影响	255
二、检索用语的专指度对检索效率的影响	255
三、在句、段、节、篇不同范围内进行组配检索 对检索效率的影响	256
四、文本用词的不规范性对检索效率的影响	257
五、不同标引方式对检索效率的影响	257
六、对自然语言进行词表控制的程度对检索效率的影响	258
七、使用自然语言应注意之点	259
第四节 自然语言或情报检索语言的未来	260
一、自然语言和情报检索语言发展的大趋势	260
二、自然语言词表在自然语言与情报检索语言结合中的 关键作用	261
第六章 文献数据库与计算机检索技术	263
第一节 文献数据库相对于手工检索工具的优异检索性能	263
一、内容更及时,报道时差缩小	264
二、信息存储量大,检索方便	264
三、检索速度快	264

四、文献特征描述更详细, 检索途径多, 可灵活调整	
检索策略	265
五、检索结果输出方式灵活、实用	267
第二节 文献数据库是当代主流文献检索工具	267
一、文献数据库的发展	268
二、文献数据库的类型	281
第三节 情报检索语言与文献数据库的配合	288
一、情报检索语言与文献数据库互相配合的重要性	288
二、情报检索语言应充分发挥数据库可组配检索的重要特点	288
三、情报检索语言应充分发挥数据库便于进行扩检、缩检 的特点	289
四、自然语言文本的语词匹配检索和字段检索只能在数据库中 实现	290
五、后控制词表只能用于数据库	291
第四节 计算机检索技术	292
一、计算机检索技术的发展阶段	292
二、计算机检索的基本原理	293
三、计算机检索方法	294
四、国外主要的联机情报检索系统介绍	298
五、我国计算机检索技术的现状	300
六、计算机检索技术发展的展望	302
第七章 因特网信息资源的检索	309
第一节 因特网上信息资源极其丰富但检索困难	309
一、网上信息资源的种类	309
二、网上信息资源检索的困难性	313
第二节 网上信息资源的检索工具	314
一、网上信息资源检索工具的类型及其构造原理	314
二、中外主要搜索引擎简介	319
第三节 网页、HTML 语言与元数据	327
一、网页的基本概念和基本结构	327
二、HTML 的概念和功能	330

三、元数据在提高网页质量中的作用	333
------------------------	-----

第八章 进一步学习指导	341
一 第一章的进一步学习、研究课题	341
二 第二章的进一步学习、研究课题	345
三 第三章的进一步学习、研究课题	346
四 第四章的进一步学习、研究课题	349
五 第五章的进一步学习、研究课题	352
六 第六章的进一步学习、研究课题	355
七 第七章的进一步学习、研究课题	356

第一章

情报检索语言及情报 语言学基本问题

第一节 情报检索语言及其功能

一、建立情报检索系统是开发利用知识信息资源的基本措施

知识是人类创造的一切财富中最宝贵的财富。这里所说的知识财富，是指古今中外的人类全部知识积累，尤其是指世界各国科学技术（包括社会科学）的最新成果。人类知识财富也称为社会信息资源。

知识和信息具有可分享性。知识在使用过程中不但不会消耗，而且还会增殖，不断丰富、完整。人类社会也只有在继承、利用现有知识财富的基础上才能进一步向前发展。小至个人，大至国家，对于充分利用人类知识财富或者说利用信息资源的能力和程度，对本身发展的意义，可以用“知识就是力量”和“科学技术是第一生产力”两句名言来概括说明。一个国家要快速发展，就要靠充分掌握和利用知识财富或者说信息资源，要以世界先进的科学技术成果作为自己发展的起点。当今世界已进入知识经济时代，科学技术成果的利用在国民经济的发展中起着主要的作用，故充分掌握和利用知识财富或者说信息资源尤为重要。

文献是存储和传递知识、信息的主要载体，所以，

利用文献就成为获取知识、信息的主要途径。要利用文献,首先必须找到所需要的最有参考价值的文献。然而,当今科学技术高速发展,文献数以亿计,纷繁无序,可以说是文献的海洋。人们利用文献是有高度针对性的。而要从这文献海洋中“全、准、快、便、省”地找出所需要的最有参考价值的文献,却是一个难题。建立高效率的情报检索系统,是解决这个难题的惟一有效的方法,是使文献资源得到充分开发、利用的基本措施。

“情报检索”一词,其广义是指通过建立情报检索系统来解决文献查找问题(也就是知识和信息的查找问题)的整个实践,其狭义仅指根据某种需要从情报检索系统中查找出相关文献线索(也称文献检索)或文献中相关内容(也称事实检索、数据检索或全文检索)的操作过程。

抛开“情报”与“信息”两个概念异同的争论,本教材把“情报检索”与“信息检索”以及“文献检索”作为同一概念,把“情报检索系统”与“信息检索系统”以及“文献检索系统”作为同一概念。

二、情报检索系统与情报检索语言

1. 情报检索系统的原理与功能

情报检索系统是为了情报检索的目的而建立的,可以根据某种(或某些种)文献特征信息从其中查检出含有特定情报(知识和信息)的文献的文献特征信息逻辑系统,是对巨大的不断增长的文獻流进行控制的手段。

所谓文献特征信息,是指文献所固有的,可藉以确认某一或某些文献并将其从其他文献中识别出来的那些特征的信息。如文献的题名、著者、出版者和出版年份、发表文献的刊物名称和期号、文献原始编号(如专利号)、文献主题内容等等。

将某一文献的各种特征信息加以记载,成为一条文献记录(也叫文献款目),再将属于一定范围的文献的全部文献记录按一定规则组织起来,就成为一个情报检索系统。所以,一个情报检索系统

是一定范围文献的全部文献记录的有序集合。情报检索系统之所以能使人们顺利地达到检索目的,是由于:

(1) 它是一定范围的文献记录的集合,对属于该范围的文献起着登记、控制的作用。同时,其中的文献记录对每一文献的外部特征和内容特征进行了描述和记载,因而能使检索者易于识别和确认某一文献是否符合检索要求。

(2) 它能通过各种检索标识提供检索途径。文献要有检索标识才能对它进行检索,犹如某一事物要有名称或代号才能使人易于与其他事物区别开来并且易于查找。在前述的各种文献特征信息中,有一部分具有较大的检索意义,如著者、题名、文献原始编号等,可以直接作为检索标识来提供检索途径。文献主题内容具有最重要的检索意义,但要以经过标引工序所赋予的分类号和检索词(包括关键词)等作为检索标识来提供检索途径。不同的检索途径可以回答从不同角度提出的检索提问。

(3) 它能通过对文献检索标识的排列,使全部文献记录(它们代表着各种不同的文献)有序化。这样,在进行检索时,就能大大缩小查寻的范围,加快检索的速度。同时,如果在固定的位置未发现相应的检索标识,就可以肯定在该检索系统中没有所需文献,而无需再到处搜索。并且,由于检索标识具有简明、惟一等特点,使表达检索提问的检索标识与表达文献特征的检索标识的相符性比较过程变得极为方便。

(4) 书本式、缩微式、机读式检索系统具有可以大量印刷、复制和远距离检索等特点,可对一定范围出版或收藏的文献进行宣传报导和信息交换,为知识财富或者说信息资源的共享提供条件。

2. 情报检索系统的类型与检索途径

情报检索系统种类繁多,功能各异,编制方法也不同。划分情报检索系统的类型,是深入、全面、系统地情报检索系统进行研究,掌握各种系统的特点、功能、适用范围、编制和使用方法要点,便于将各种类型系统的性能和优缺点进行分析比较以及进行改进、创新的重要方法。

情报检索系统大体可以按其所输出的结果、所提供的检索途径、所收录的文献、所采用的载体形式四个角度划分类型。

(1) 按输出结果(即检索结果)划分类型。可分为文献检索系统(提供文献线索)和文本检索系统(提供文献中的相关内容,如全文检索系统、事实检索系统、数据检索系统、图像检索系统、多媒体检索系统等)。文献检索系统在情报检索系统中占主要地位。下述三种分类角度,都是指文献检索系统的分类。

(2) 按检索途径划分类型。可分为单一检索途径的检索系统和非单一检索途径的检索系统。单一检索途径的检索系统又可分为内容检索途径(如分类、主题、化合物代码等检索途径)的检索系统和非内容检索途径(如文献著者、原始编号、书名和刊名等检索途径)的检索系统两大类,其中提供内容检索途径的检索系统是基本的,在情报检索中占主要地位;提供非内容检索途径的检索系统是辅助性的,在情报检索中只占次要地位。单一检索途径的检索系统是单功能的,也可称为简单检索系统。若干单一检索途径的检索系统的有机结合,构成一个体系,通常称为目录体系。附有索引的检索工具书刊相当于一个目录体系,属于非单一检索途径的检索系统。绝大多数的电子计算机检索系统是非单一检索途径的检索系统,具有多种检索功能。

(3) 按收录文献范围划分类型。情报检索系统可按其所收录的文献从多方面划分类型。例如,按收录文献的学科分,按收录文献的出版类型分,按收录文献的文种分,按收录文献的出版时间分,按收录文献的密级分等情报检索系统类型。

(4) 按载体形式划分类型。作为情报检索系统基本构成因素的文献特征信息必须记录在某种载体上,载体的种类和形式多样,从而形成按载体形式分的情报检索系统类型,有卡片式检索系统、书本式检索系统、电子计算机检索系统以及其他许多现已很少采用的检索系统形式。

任何一种具体的情报检索系统,都可按上述四个角度划分类型。各种类型情报检索系统的检索功能或用途,是与各种不同的检

索要求或文献情报服务的侧重点相适应的,所以每种类型都有一定的客观重要性。

任何一种具体的情报检索系统,都提供一种或多种检索途径。不提供检索途径的系统,就不能称为情报检索系统。提供检索途径,使检索者能循着所提供的与文献特征相吻合的检索途径查找到载有所需知识或信息的文献,是情报检索系统的主要特征。

每种检索途径由两个因素构成:一是检索标识,二是可检序列。

3. 内容检索途径与情报检索语言

可以这样说,在情报检索中,不管是从哪种检索途径入手进行检索,其最终目的都是一样的,即都是为了要查找到具有特定内容的文献。虽然,有时从非内容检索途径入手进行检索更为简便,但那些检索途径满足情报检索要求的能力都毕竟有限,不可能很充分地满足大多数检索要求。所以,内容检索途径是不可缺少的基本的检索途径,非内容检索途径只是辅助的检索途径。

但是,任何提供内容检索途径的检索系统,都有一个用某种语言来作为检索标识以表达文献内容的问题。这是一个十分复杂的问题。

任何文献都有题名,大多数文献还有章节题名。文献的题名是文献著者经过深思熟虑后拟定的,用以表达文献主题内容的名称。一般而言,一个完整的文献题名,能够较好地表达文献的主题内容。但实际上,文献题名虽能较好地表达文献主题内容,却不能直接作为一种内容检索途径(它只能作为一种文献外表特征——已知题名的检索途径)。只有当文献题名转化为关键词时,才能作为一种内容检索途径——题名关键词检索途径。

本来就是用以表达文献主题内容的文献题名之所以不能直接作为一种文献内容检索途径,其原因是多方面的。例如,它在字顺序列中不能很好地提供准确、有效的文献内容检索入口,它缺乏良好的聚类功能(充分地、正确地、明确地显示文献主题概念之间关系的功能),部分文献的题名揭示文献主题内容不充分或不确切等等,使它不可能达到较满意的检索效果。