

Internet

网络信息资源检索

主编 叶青方倪 郭璐 吴桂金



东北林业大学出版社

Internet 网络信息资源检索

主编 叶青 方倪 郭璐 吴桂金

东北林业大学出版社

· 哈尔滨 ·

图书在版编目 (CIP) 数据

Internet 网络信息资源检索 / 叶青, 方倪, 郭璐等主编.

--2 版. --哈尔滨: 东北林业大学出版社, 2016. 7

ISBN 978 - 7 - 5674 - 0814 - 2

I. ①I… II. ①叶… ②方… ③郭… III. ①网络检索 IV. ①G354. 4

中国版本图书馆 CIP 数据核字 (2016) 第 149699 号

责任编辑: 卢 伟 张红梅

封面设计: 彭 宇

出版发行: 东北林业大学出版社 (哈尔滨市香坊区哈平六道街 6 号 邮编: 150040)

印 装: 三河市佳星印装有限公司

开 本: 787mm × 1092mm 1/16

印 张: 14

字 数: 310 千字

版 次: 2016 年 8 月第 2 版

印 次: 2016 年 8 月第 1 次印刷

定 价: 65.00 元

目 录

1 科技信息检索基础知识	(1)
1.1 中国图书馆图书分类法	(1)
1.2 主题词的选择	(3)
1.3 数据库及文档结构	(9)
2 网络信息资源的类型及应用	(13)
2.1 信息资源的类型	(13)
2.2 Internet 网络信息资源	(16)
2.3 网络搜索引擎的利用	(18)
3 网络中文报刊论文数据资源检索	(27)
3.1 中国期刊网	(27)
3.2 维普中文科技期刊全文数据库	(31)
3.3 万方系统数字化期刊数据库	(33)
3.4 人大复印资料全文数据库	(33)
4 网络图书数据资源检索	(36)
4.1 书生之家图书检索	(36)
4.2 超星数字图书馆	(37)
4.3 电子图书	(38)
4.4 Apabi 高校教参系统	(39)
4.5 馆藏信息资源的检索	(41)
4.6 美国 NetLibrary 电子图书	(43)
5 特种文献数据资源检索	(45)
5.1 专利数据资源检索	(45)
5.2 标准数据资源检索	(58)
5.3 会议信息检索	(72)
5.4 学位论文数据库	(76)
5.5 国务院发展研究中心信息网	(79)
5.6 中国高等教育文献保障系统	(84)
5.7 万方数据资源系统	(86)
6 外文数据库	(90)
6.1 Kluwer 电子期刊数据库检索	(90)
6.2 EBSCO 公司数据系统	(93)
6.3 OCLC FirstSearch 基本组数据库	(97)
6.4 Elsevier Science Direct Onsite 外文电子期刊全文	(104)

6.5	SpringerLink 科技期刊及丛书	(109)
6.6	Biosis Previews 数据库	(113)
6.7	CAB 国际生物和农业中心数据库	(120)
6.8	AGRIS 数据库	(122)
6.9	ProQuest 生物、农业全文、Agricola 数据库	(124)
6.10	John Wiley InterScience 电子期刊	(128)
6.11	Blackwell 数据库	(131)
7	参考工具书	(134)
7.1	字词(辞)典	(134)
7.2	百科全书	(136)
7.3	年鉴	(138)
7.4	手册	(139)
7.5	名录	(140)
7.6	表谱	(141)
7.7	工具书指南	(142)
8	著名四大参考数据库的概况	(144)
8.1	SCI 科学引文索引	(144)
8.2	EI 美国工程索引网络数据库	(146)
8.3	ISTP 国际科技会议录索引和 ISSHP 社会科学 与人文科学会议录索引	(148)
8.4	SSCI 社会科学引文索引和 A & HCI 艺术与人文科学引文索引	(149)
8.5	国外期刊收录对稿件内容和学术水平的要求	(150)
9	DIALOG 国际联机检索系统	(154)
9.1	DIALOG 系统资源	(154)
9.2	DIALOG 系统检索基本知识	(155)
9.3	DIALOG 系统检索指令	(160)
9.4	检索运算符	(168)
9.5	检索策略和实例	(175)
10	STN International 系统	(187)
10.1	STN 联机系统简介	(187)
10.2	STN 系统检索技术	(189)
10.3	检索实例	(196)
10.4	专利检索方法	(197)
10.5	化学文摘 CA 检索	(198)
10.6	STN 系统中国代理网站的使用	(204)
11	创新能力的培养	(205)
11.1	知识经济时代的特征	(205)
11.2	创新人才的培养	(207)

11.3 信息能力	(208)
11.4 科学研究的步骤	(211)
11.5 科技查新的概念	(212)
参考文献	(215)

1 科技信息检索基础知识

信息获取和利用的能力是知识经济时期人们必须具备的基本素质之一。信息检索的能力是评价一个人信息能力的主要标志。一个科研工作者如果了解数据库的组成和结构,掌握查找所需信息的方法和技巧,可以达到事半功倍的效果。不但可以节省很多的时间和精力,也可以提高工作效率,及时了解本学科前沿的研究动态和发展趋势,不断地进行创造性的研究。目前在网络上运行的中外文数据库不断增加,我们只要掌握其基本检索方法,就可以熟练应用。下面介绍信息检索中需要掌握的基础知识。

1.1 中国图书馆图书分类法

1.1.1 分类检索语言

检索语言是编制检索工具书和建立数据库使用的一种特定语言。它是根据信息检索的需要而创建的人工语言,专门用于各种手工的和计算机的文献信息存储系统,表达文献主题概念和检索课题概念。目前在数据库检索中常用的检索语言有分类语言和主题语言。

分类检索语言是按学科的知识集中文献,并将每个学科赋予一个分类代号,用分类号作为检索标识。常用的分类语言属于等级体系分类语言。等级体系分类语言是按学科范畴概念而构成的一种语言体系,它集中体现学科的系统性,反映事物从属、派生的关系,从上至下、从总体到局部划分、展开,是一种等级体系分类结构。

它的结构原理是以学科分类为基础,从学科的大类开始,依次根据学科所属的分支层层划分下一级类目,从大类到小类反映了一个学科的全貌。例如“建筑设计”这一概念,又可根据各种建筑设计不同特性,进一步划分为民用建筑、农业建筑、工业建筑、军用建筑等。分类语言的类目及文字表达体现在各种分类工具书中。

1.1.2 分类工具书

目前国内主要采用的图书分类工具书是《中国图书馆分类法》(简称《中图法》);国外采用的分类工具书大多是《美国国会图书馆分类法》(Library of Congress Classification)(简称 LC 分类法),《杜威十进制分类法》(Decimal Classification)(简称杜威法),《国际十进制分类法》(Universal Decimal Classification)(简称 UDC 分类法),《国际专利分类表》等,这些都是典型的分类语言。下面以《中图法》为例介绍它的编排及使用。

《中图法》由北京图书馆组织全国力量编写,该分类法广为全国的图书馆和信息研究部门所采用。主要为图书、期刊等文献,以及课题、论文给出分类号。我们在用国内

检索工具或数据库查找信息线索时,也要用此分类法寻找类目。目前,全国统一使用“第四版”。

《中图法》依据发展原则和客观原则,遵循从总到分、从一般到具体、从低级到高级等逻辑规则,层层隶属,采用英文字母与阿拉伯数字相结合的混合制号码。字母的顺序反映大类的序列,字母后面用数字反映大类下类目的划分。每一个类目都有一个类号和类目名称。采用分类表中的类目等级而编制的索引就是我们常用的分类索引。情报人员以分类类目或号码作标目,读者则以其为检索的入口点。

《中图法》将人类知识分为五大基本部类:马列主义毛泽东思想、哲学、社会科学、自然科学、综合性图书,下面再分为 22 个基本大类。简表具体见图 1.1.1,大类下再分小类,层层展开,形成一个树形等级结构。例如:我们要检索有关“森林保护”方面的图书,在用分类检索时,首先找出它的大类号 S 农业科学为一级分类, S7 林业为二级类目,继续查找 S76 为“森林保护”。S76 的下面集中了所有有关“森林保护”的资料。在 S76 下还可找到 S761 气象灾害及其防治、S762 林火、S763 森林病虫害及其防治等下一级分类号。选择合适的分类号作检索词找出所需文献。

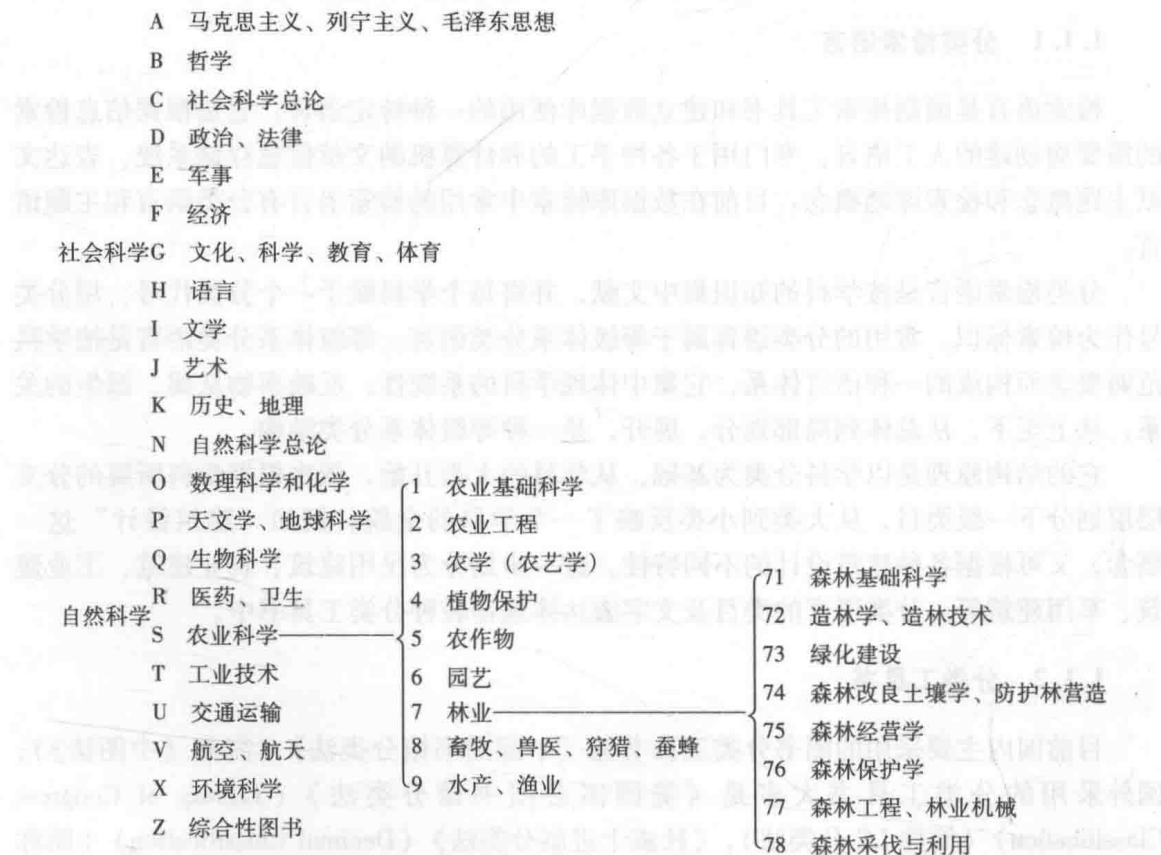


图 1.1.1

可以看出,分类号的变化体现了各学科体系的从属关系,分类号位数每增加一位,则级别更低一级。确定分类号可从两种途径入手:

直接查找:如果课题有一个主题,则直接根据主题入手,按学科一级级查找即可;

如果课题有两个主题,可在两个主题领域给出两个分类号或根据从属关系给出一个主分类号。

间接查找:可利用现成的已有分类号的对口文献标出的分类号,参考后再进行查找。

我们在使用《中图法》时,要注意分类号的表达方式有以下几个问题:

分类号的类别字母要大写,例如:

F28 基本建设经济

Q981 古人类学

为适应“工业技术”图书资料分类的需要,对其下一级类目的复分,也采用字母标志,即工业技术所属的二级类,采用双字母。如:铸造工业,分类号为TG24;陶瓷工业,分类号为TQ174。

一个课题跨两个学科时,要把两个分类号都找出来,主要的分类号放在前面。例如:棉纺织机械TH6 TS112。

《中图法》中几个符号的应用:

例如:S718.55*1 森林生态系统类型。

上标有“+”号,表示资料法的分类。在用《中图法》分类时“+”号不用写上。

例如:S717.2/7 森林地区分布。

2/7 表示地区分布,按世界地区表复分。

通常为图书、期刊论文、学位论文、课题查找分类号时要用《中图法》。在一些数据库检索中凡是提供了分类号字段,都可以用分类号进行检索。用分类号检索的特点是只要输入一个学科的分类号,即可以把全学科的文献都检索出来,不会出现漏检的问题。查找中图分类号可参考上海交通大学图书馆 <http://www.lib.sjtu.edu.cn>, 点击网上参考咨询台的中图法简表,也可以使用维普科技期刊数据库中的分类检索。

1.2 主题词的选择

检索词是用户在数据库检索中输入的检索项,它的选择与课题的相关性和准确性大小直接影响检索结果的查全率和查准率。所以在直接进入检索前,一定要先分析课题,提取准确的检索词。上面介绍的分类号是检索词的一种形式,利用一个分类号可以把整个学科的文献查出来,同样道理如果利用一个主题词,就可以把同一个主题的文献检索出来。了解主题检索语言的特点会帮助我们选择正确的主题词。

1.2.1 主题检索语言

主题检索语言的特点是把同一个主题的文献进行集中。以主题词汇构成索引词表,并对文献进行标引。主题语言表达的概念比较准确,具有较好的灵活性和专指性,不同的检索系统、不同的专业领域可以有各自的主题词表。目前国内出版的叙词表有综合性和专科性两种,详见表1.2.1。凡是国内外公开出版的叙词表,都可以用于对文献的主

题标引和建立主题检索系统。主题检索语言常用的有两种,即叙词语言和关键词语言。

表 1.2.1

名 称	编 者	出版 时间	专 业 范 围	词 量	
				叙 词	非叙词
航空科技主题表	航空工业部情报系统	1971	宇宙航行航空	4 330	199
电子技术汉语主题词表	第四机械工业部第一研究所	1977	电子工业	75 900	1 000
常规武器专业主题词表	常规武器系统 20 个单位共同 编制	1977	常规武器	4 398	1 625
国防科学技术主题词典	国防科工委情报所	1978	国防科技	17 173	3 719
原子能科技资料主题词表	二机部情报所	1978	原子能科技	15 179	3 363
机械工程主题词表	一机部情报所	1979	机械工程、电机 工程、仪器仪表	9 667	1 533
汉语主题词表	中国科学技术情报所、北京 图书馆	1991	哲学、社会 科学、自然科学	93 945	17 410
国防科学技术叙词表	国防科工委科技情报所	1985	科学技术	29 774	4 742
化工汉语主题词表	化学工业部科学技术情报所	1983	化工	16 602	3 075
中国分类主题词表	《中图法》编委会	1994	综合	21 万	
铁路汉语主题词表	铁道部科学技术情报所	1987	铁道等	4 447	82
新闻叙词表	新华通讯社新闻资料检索研 究室	1988	社会科学	8 603	1 203

1) 叙词语言

叙词是指从自然语言优选出来,经过规范化处理的名词术语。叙词法除具备主题法的特征外,还有两个特征:a. 规范性。叙词法采用的检索语言均需要事先进行规范化处理,列于叙词表中,标引和检索都从叙词表中选词,因而保证了标引和检索的一致性。b. 概念性。叙词法从单元概念出发去揭示文献内容,而不是单元词,在检索时可以用一个单元概念,也可以用多个单元概念组合,因而保证了用多个概念组配检索复合概念的课题,具有灵活性。概念组配的原理决定了叙词语言的基本特征和性能。按照概念之间的逻辑关系,叙词的组配有以下三种。

交叉组配。交叉组配指使用两个或两个以上具有交叉关系的叙词的组配。组配结果可以形成一个新概念,例如:“林业”与“机械”,将两词组配后的概念是“林业机械”。

限定组配。这是一个表示事物的叙词与另一个表示事物某一方面的叙词之间进行某一概念限定的组配,即某一方面的组配。例如:将“计算机”和“显示器”这两个分

别表示事物及其部分的叙词进行组配,表示将“显示器”限定在计算机范围内。

并列组配。表示两个或两个以上主题概念之间有一定联系的一种组配,组配的结果不形成一种新的概念,只揭示参加组配概念之间的某种联系。例如:“病害”与“虫害”组配成“病虫害”。

采用叙词组配的方式表达文献主题,可以提高标引的深度,有助于充分表达文献的主题,满足从不同主题查找文献的需求。

2) 《汉语主题词表》的体系结构

叙词表是概括某学科领域并由语义相关、族性相关术语组成的一种规范化的动态词典。可以通过词间关系的揭示来表达其语义关系,如等同关系、属分关系、相关关系以及等级关系,同时运用字顺、分类或范畴、词族等方法予以全面显示;保证标引人员和用户用语一致,便于叙词的规范化管理,便于按其组织检索系统。

《汉语主题词表》(以下简称《汉表》),是我国最大的一部综合性的叙词表。全书分为自然科学和社会科学两大部分。社会科学部分共两册,自然科学部分共五册(于1991年修订)。共收词条93 945条。全书由主表、附表和辅助索引组成。写文章标引主题词时可以到此词表中选词。

主表(叙词字顺表):包括全部叙词和非叙词,全部按汉语拼音字母顺序编排,并标有注释及参照项。主表是进行主题概念转换,确定主题词的主要途径,只要分析出的主题概念比较稳定,能够使用明确的语词形式表达,就可以通过字顺系统进行探索性查找,还可以通过参见系统进一步查找。每个款目除了叙词外,还著录有用词参照项等。主表的著录格式如图1.2.1所示。

Xian xiang guan			
款目主题词→	显像管	[56E]	← 范畴号
英文译名→	picture tubes		
代项符号→	D 电视显像管		← 非正式主题词
	监视管		
分项符号→	F 彩色显像管		
	固体显像管		← 下位主题词
	黑白显像管		
属项符号→	S 电子显像管		← 上位主题词
族项符号→	Z 电子管*		← 族首词
			(* 族首词符号)
参项符号→	C 显示管		
	指示管		← 相关词

图 1.2.1

例如:奥林匹克委员会

Y 奥委会

《汉表》中的参见符号及几种常用外语叙词表使用的参见符号见表1.2.2。

表 1.2.2

词间关系	参照符号含义	汉语拼音符号	英文索引	法文符号	德文符号	国际通用符号
等同关系	用	Y	Use	EM	BS	→
	代	D	UF	BF	=	
	组代		UFC			
等级关系	分	F	NT	TS	VB	>
	属	S	BT	TG	OB	<
	族	Z	TT			
	(属种)		BTG			
	(属种) 分		NTG			
	(整部) 整		BTP			< p
相关关系	(整部) 分		NTP			> p
	参	C	RT	VA	VB	—

附表：包括四个部分：“世界各国政区名称”“组织机构名称”“自然地理区划名称”和“人名”。它也是主表的一部分，可以用做标引和检索的直接依据。它主要是为了控制主表的词量而将一些专有名词抽出来，分别按字顺编排而成的。

辅助索引中有范畴索引、词族索引、英汉对照索引。

范畴索引：是根据叙词的学科和词义范畴，划分若干类，每类下再按字顺组织的分类系统。此索引有利于从学科分类角度查找叙词，主要用于两种情况：一种是某些文献涉及的主题概念集中于某一学科时，为了提高标引效果，可直接查范畴索引；另一种是在有些语词形式一时无法确定，但知其学科归属，或没有相对应的叙词，需选用近义词加以标引时，也可查范畴索引。范畴索引著录格式见表 1.2.3。

词族索引：词族，是指一组具有层层隶属关系即词族关系的叙词。该索引按族首词（即一组具有族性关系的叙词中概念最广泛的词，该词的右上角标有“*”号）的字顺编排，从族首词就可查得其层层下属的全部叙词。用此表时一般要通过其他途径查出族首词后，才能使用。该索引的著录格式见表 1.2.4。

英汉对照索引：此索引是用来从英文查检汉语叙词，也可以再以对应的中文主题词回查主表，利用主表的叙词再进一步确定词义，从而选定标引的主题词。也是参考英文叙词的一种辅助工具。

表 1.2.3

表 1.2.4

范畴索引	词族索引
49JH 木工机床	动物 *
拌胶机	D 动物学
刨光机	· 保护动物
Y 刨木机	· 哺乳动物
刨机	D 兽类
Y 刨木机	· 观赏动物
刨木机	· 野生动物
刨片机	· · 野禽
刨削机	· · · 猛禽
拨料机	· · · 珍禽

3) 关键词语言

关键词语言是主题检索语言之一，以关键词作为检索标识的标引和检索文献的方法。关键词指从文献题名或文摘以及正文中抽取的，能够表达文献主题并具有检索意义的词。但与叙词语言不同，它是未经规范的纯自然语言语词。关键词语言是检索语言中自然语言化的重要标志之一。这种语言不受任何词表的控制，标引较方便，并且可以从多个词作为入口点。

因为关键词语言未经规范处理，有时标引人员用词和检索人员用词不一致，会造成漏检，查全率一般较低，所以在数据库检索中一般采用两种语言并用，互为补充，以保证较高的检索效率。关键词与叙词的区别在于前者要用词表，后者不用词表。

1.2.2 提取检索词的方法

检索词是表达文献信息需求的基本元素，也是计算机检索系统中进行匹配的基本单元。检索词选择正确与否，直接影响着检索结果。在全面了解检索课题的相关问题后，提炼主要概念与隐含概念，排除次要概念，以便确定检索词。

1) 确定检索词的原则

(1) 利用数据库提供的词表进行选择。当所选的数据库具有规范化的词表时，应优先选用该数据库词表中与检索课题相关的规范化主题词，从而可以获得最佳的检索效果。一些外文数据库都提供了较丰富的各类词表。

(2) 选用数据库规定的代码。许多数据库的文档中使用各种代码来表示各种主题范畴，有很高的匹配性。例如，世界专利文摘数据库中的分类代码，化学文摘数据库中的化学物质登记号。

(3) 选用常用的专业术语。在数据库没有专用的词表或词表中没有可选的词时，可以从一些已有的相关专业文献中选择常用的专业术语作为检索词。

(4) 选用同义词与相关词。同义词、近义词、相关词、缩写词、词形变化等应尽量选全，提高查全率，避免漏检。

(5) 使用各类文献的分类表, 利用分类号进行检索。如《中国图书馆分类法》《标准文献分类法》《国际专利分类表》等分类的工具书。

(6) 使用机构名、专家学者名、邮政编码等作为检索词, 同样可以达到理想的检索效果。

2) 确定检索词的注意事项

(1) 如果课题较新且面较窄, 希望查准, 可用分类号进行检索; 同时结合关键词进行检索。也可用作者、机构、出版者等提高查准率。

(2) 对课题名称语句进行切分, 找到核心概念, 进行核心词的提取。

(3) 对隐含概念的发掘。隐含的概念指课题中没有明确指出的但又与课题密切相关的概念, 包括相关的概念和上下位的概念。这些词需要从课题所属专业性角度进行深入分析, 才能提取出确切反映课题内容的检索概念。如病虫害这样外延比较宽的词可转换成具体的病害和虫害的名词表示。转换时要注意选择专指概念的主题词。

(4) 选择检索词时, 要尽量使用专指性强的词, 避免使用抽象的、泛指的概念词, 例如: 系统; 无关紧要的词, 例如: 研究、作用、发展; 一些限定过窄的词, 例如: 可控硅电池。对课题中一些比较模糊、狭窄或不可行的词可用明确的可行的相关词或同义词进行替换。

(5) 选择专业数据库检索时, 相关概念就没有必要提取。如: 用医学数据库时就不用提 Medicine 一词。

(6) 检索必须选择有实质意义的词。尽量避免一句话或者短语。可以使用单词、词组或者不可分割的短语。一个课题检索词很多时, 可以利用逻辑符进行逻辑组配。

1.2.3 逻辑符的使用

布尔逻辑符是国内外数据库中通用的符号, 目前使用的逻辑符有逻辑“与”或“and”、逻辑“或”或“or”、逻辑“非”或“not”, 用布尔逻辑符可以进行检索词的逻辑组配, 以实现检索课题的要求。

除了使用布尔逻辑算符外, 一些数据库还利用位置算符、截词符等限定符来进行检索提问式的编写, 提问式的编辑是否合理, 直接影响查全率和查准率。

检索提问式。在检索中经常会遇到一些课题的概念很复杂, 需要几个检索词来表达课题的概念, 这时在检索中就要对课题编写检索提问式。例如课题“冰箱节能阀的研究”, 检索提问式可以写成:

中文检索式: (冰箱 or 制冷) and 节能 and 阀

英文检索式: (Refrigerat * or freezer *) and valve and (save or saving)

上面两个检索提问式用到了逻辑符、截词符。在一些数据库检索中, 必须要利用检索提问式才能进行检索。关于检索提问式的编写技术在各类数据库中将详细介绍。

1.3 数据库及文档结构

1.3.1 数据库的定义

数据库 (Database) 是计算机技术与信息检索技术相结合的产物, 是计算机检索的基础, 要建立计算机检索系统, 首先必须对所收集到的大量文献资料按一定的体系和规则加以处理, 使之成为机读形式的数据。这种在计算机存储设备上按一定方式存储的相互关联的数据集合, 就是数据库。根据 ISO/DIS5127 (文献与情报工作术语) 标准, 数据库 (Database) 定义为: “至少由一种文档 (File) 组成、能满足某一特定目的或某一特定数据处理系统需要的一种数据集合。”

目前国际上的检索系统以 DIALOG 系统、STN 系统、ORBIT - BRS 比较著名, 包含的数据库也较多。国内的检索系统规模比较小, 检索的文献类型比较单一。除了万方检索系统是综合的数据库外, 如中国期刊网、书生之家、超星系统、重庆维普检索系统、中国专利网、中国标准网等检索系统, 包含的数据库均较少, 但是发展得很快。

1.3.2 数据库的类型

数据库是情报检索系统的核心部分之一。数据库的质量和建设是非常重要的。用户在进行检索时, 选择和了解数据库是非常重要的, 它直接影响到检索的质量。我们目前用的数据库有以下几种类型:

1) 题录文摘型数据库 (Abstract)

主要是科技、专利类信息。这类数据库每条记录的内容和格式与书本式的检索工具相同, 主要通报各条信息的题目、作者、原文出处及论文摘要, 不直接显示原文, 如果需要原文要进一步获取。

2) 目录手册型数据库 (Directory)

主要汇集了某一或若干学科或专业领域的基本知识、参考资料或数据。如各种工商企业名录、商业、公司、研究机构、学校、世界名人、医学专家、产品手册、药典、百科全书、年鉴、指南、各种学会、协会、基金会等叙述性记录。

3) 全文型数据库 (Full text)

多数以专题形式或提供某一种出版类型的文献为特征, 用户通过检索可以获得原始文献全文, 主要是市场行业报告、分析报告和工业报告、新闻报道、期刊报纸全文数据库等。商情信息的检索多数用此类数据库。

4) 数值类信息数据库 (Numeric)

主要给用户提供各种统计数据, 诸如价格、进出口数据、生产、销售数据等信息。

5) 图像数据库 (Image)

主要提供图像, 可以形象、直观地为用户提供所需图片信息。这在专利、商标、化学化工等方面有十分重要的检索价值。

1.3.3 数据库的组成和结构

一个检索系统包含诸多的数据库，一个数据库由若干的文档（File）组成，一个文档由诸多的记录（Record）组成，一条记录由若干个字段组成，结构如图 1.3.1 所示。

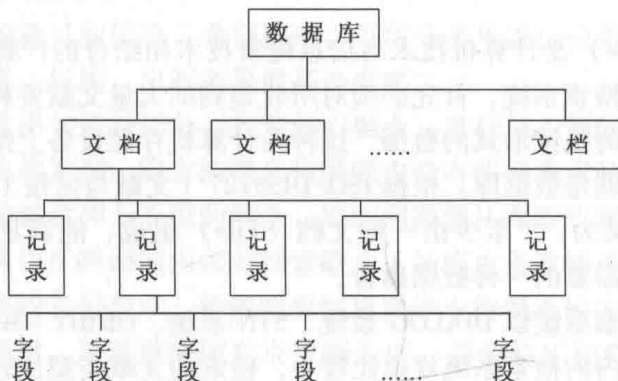


图 1.3.1

文档是一种数据和信息的容器，是可以包含文本、图像、视频、音频、动画等复杂非结构化数据的集合。一般来说，一个数据库至少包括一个顺排文档和多个倒排文档。顺排文档是将数据库的全部记录按照记录号的大小排列形成的文献集合，但内容庞大且各种文献特征处于无序状态，直接用于检索时会大大影响检索速度。而倒排文档是将记录的某一特征作为排列依据，其后列出含有此特征的记录号，检索时可以大大提高计算机运算速度，提高检索效率。

一条记录对应一篇文献，是对某一实体的全部属性进行描述的结果，是多个字段的集合。若干条记录构成一个文档。在全文数据库中检索一条记录相当于一篇完整的文献；在书目数据库中检索一条记录相当于一篇文摘或题录。

字段是记录的基本组成单元，是有关一篇文献或一条记录的基本数据单元，每一个字段都反映该篇文献的一个方面的信息，组合在一起形成对一篇文献的内容特征和外表特征的完整描述。在检索中其含义是指检索某一记录时的检索入口，如书名字段、著者字段、文摘字段、出版项字段等。各字段的组合就成为记录。

可以从不同角度选出多个检索入口词。检索词可以是单词、词组、字母、数字和符号，输入的检索词必须指定出现在某一个字段，一般的数据库文档都设有默认的检索字段。不同字段的检索对提高文献的查找效率有很大的作用。字段通常又可分为基本索引字段和辅助索引字段。

基本索引字段：与主题内容直接相关的、描述文献内容特征的检索字段称为基本索引字段，联机检索中一般用后缀表示，如标题（TI）、规范词（DE）、文摘（AB）等。

辅助索引字段：除基本索引外，反映文献外部特征的字段，如机构名称（CS =）、公司名称（CO =）、文献类型（DT =）、作者姓名（AU =）、语种（LA =）、出处（SO =）等。联机检索中一般用前缀表示。若不做任何限定，系统将默认在基本索引字

段中进行检索。

下面是两条记录的格式：

(1) 中文数据库记录格式，例如：万方资源系统的学位论文数据库的记录格式。它由 12 个字段组成。

【论文题名】垂直管外降膜吸收传热传质过程强化的研究

【论文作者】陈沛

【作者专业】化学工程

【授予学位】硕士

【导师姓名】陈嘉宾

【授予学位单位】大连理工大学

【馆藏号】Y0331995

【分类号】TQ051

【论文页数】55 页

【出版时间】2000. 3

【关键词】数值计算；强化热—传质传递；溴化锂；垂直管外降膜吸收

【文摘】对 LiBr 溶液在垂直的 $\phi 19 \times 1$ 铜管和四种换热强化管外降膜吸收的传热传质情况进行了实验研究，通过分析、比较，得到了其中的最佳管型。通过对 LiBr 溶液在垂直管外降膜吸收过程中热—质传递现象的研究，分析了非绝热吸收过程中传热和传质相互作用、相互影响的关系，建立了吸收过程热—质传递的数学模型，并对垂直管外 LiBr 溶液降膜吸收过程进行数值计算。

(2) 外文数据库记录格式。

下面是 DIALOG 系统一条公司名录数据库的部分记录样例，它是由不同的字段组成的。

0190126	DIALOG 系统存取号
FAMIRIMATO KK	公司名称
FAMILYMART CO LTD	
26 - 10, HIGASHIKEBUKURO 4 - CHOME	联系地址
TOSHIMA - KU, TOKYO 170	
Telephone: 03 - 3989 - 6600	联系电话
Primary Industry:	
87910 (BUSINESS CONSULTANTS)	主要业务范围
Date of Incorporation: 08/1948	公司成立年月
Paid - in Capital (Y000): 13, 238, 000	
Paid - in Capital (\$ 000): 65 - 660	资产额
Credit Scoring: B (82)	资信等级
Number of Employees: 1, 132	雇员人数
Number of Shareholders: 6, 276	销售情况
.....	

以上只是一条记录的片段，有公司的股票价格，与公司往来的银行名单，公司性质，公司负责人姓名、联系电话、出生年月日、出生地址、教育背景等事项。每一条记录有一个唯一的存取号。存取号是每一条记录在数据库中的位置。DIALOG 系统首先把每个存取号按年代或卷期的先后顺序排成一个目录文档，也叫主文档。检索结果就是从主文档中按存取号调出来的。其次，检索系统把每条记录的其他可检字段中的词抽出来，按字顺排序。从每个字段抽出来的词，按顺序排成不同的索引词典。

美国 DIALOG 系统数据库的一条记录格式如表 1.3.2 所示。