

信息检索

何晓萍 胡德华 主编

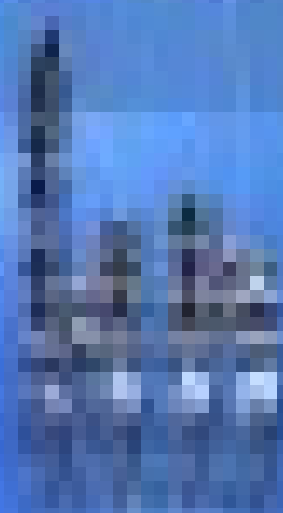


海洋出版社

信息检索

第2版 2006年12月 120页

ISBN 7-309-05411-2



中国海洋大学出版社

信 息 检 索

主 编 何晓萍 胡德华

副主编 罗时民 韩亚兰

海 军 出 版 社

2002·北京

内 容 简 介

本书由信息检索概论、基础信息检索和数字信息检索三大部分组成。第一部分主要介绍信息检索概论、信息检索方法和技术,内容包括信息与文献的基本概念、信息检索与检索系统、检索语言、检索方式、检索程序和检索技术;第二部分主要介绍检索工具的检索和利用,内容包括相关基础检索工具的著录格式、检索方法与检索步骤,相关数字数据库的使用指南和检索技巧;第三部分主要介绍网上信息资源检索和利用、网上信息资源与信息检索,包括中国高等教育文献保障系统(CALIS)引进数据库的检索应用、方法及技巧,因特网上信息检索技巧。书中将基础信息检索和数字信息检索两部分内容合为一体,便于读者系统地了解本学科及相关学科文献检索工具和高质量地综合利用多种检索工具。

本书可作为高等院校学生文献检索课教材,亦可作为科技人员检索信息的参考书。

图书再版编目(CIP)数据

信息检索/何晓萍等编. —北京:海洋出版社,2002.8

ISBN7-5027-5440-7

I. 信... II. 何... III. 情报检索 IV. G252.7

中国版本图书馆 CIP 数据核字(2002)第 048204 号

责任编辑:高朝君

特约编辑:王海峰

海洋出版社 出版发行

<http://www.oceanpress.com.cn>

(100081 北京市海淀区大慧寺路 8 号)

北京昌平前进印刷厂印刷

新华书店经销

2002 年 9 月第 1 版 2002 年 9 月北京第 1 次印刷

开本:787mm×1092mm 1/16 印张:20.25

字数:491 千字 印数:1~8 000 册

定价:29.80 元

海洋版图书印、装错误可随时退换

前言

21 世纪,是信息技术飞跃发展的时代,也是网络信息时代,整个世界在激烈的竞争中不断前进,信息化的程度已经被当作衡量一个国家综合国力的重要尺度,也是衡量一个国家竞争能力的核心要素。在人类文明和科技高速发展的今天,谁掌握了信息,谁就掌握了无穷的财富。

信息产业部信息化推进司司长宋玲指出,信息是 21 世纪人类发展的核心,推进信息化建设的首要任务是加强信息基础设施建设,关键和核心是要推动信息资源的深度开发利用。现代社会数字化信息浩如烟海,面对各种不同类型的数据库如何去检索所需信息资料;面对网上无序的信息又如何去搜索、分析、筛选、下载储存有用信息。这些正是文献检索课需要解决的问题。

教育部曾于 1984 年就做出决定,要求各高等学校以图书馆为基地开设“文献检索与利用”课程,以适应科学技术发展,增强大学生、研究生的信息意识,培养和提高他们检索和利用信息的技能。随着科学技术的发展和创新,新的科技文献不断涌现,特别是计算机技术和网络技术的广泛应用,使检索手段日趋现代化。为适应数字化、网络化的时代特征,编者在总结多年教学实践的基础上,仔细研究分析了文献检索课程相关教材,针对传统教材侧重讲述手工检索工具的特点和数字信息检索的时效性,以手工检索为基础,以计算机检索为重点,突出讲述数字信息检索内容。

本书作为高校学生文献检索课程教材和科技人员检索信息的参考用书。它可以帮助读者掌握信息检索的方法与技能,使读者学会获取各类信息和新知识,并引导读者到更广泛的知识范围中去进行检索,建立一个不断演进的知识体系。

本书由信息检索概论、基础信息检索和数字信息检索三大部分组成。第一部分主要介绍信息检索概论、信息检索方法和技术(第 1、2 章)。内容包括信息与文献的基本概念、信息检索与检索系统、检索语言、检索方式、检索程序和检索技术。第二部分主要介绍检索工具的检索和利用(第 3、4、5、6 章)。内容包括相关基础检索工具的著录格式、检索方法与检索步骤;相关数字数据库的使用指南和检索技巧。第三部分主要介绍网上信息资源检索和利用(第 7 章),网上信息资源与信息检索,包括中国高等教育文献保障系统(CALIS)引进数据库的检索应用、方法及技巧、因特网上信息检索技巧。书中将基础信息检索和数字信息检索两部分内容合为一体,便于读者系统了解本学科及相关学科文献检索工具和质量地综合利用多种检索工具。

参加本书的编写人员有:何晓萍(第 1 章、第 2 章、第 4 章第 1 节);罗时民(第 6 章);胡德华(第 4 章第 2、3、4 节,第 7 章);韩亚兰(第 3 章、第 5 章)。本书由何晓萍、胡德华同志担任主编,负责编写大纲和全书的统稿工作。袁芳、梁红妮同志担任校核工作。

网络信息资源与现代信息技术发展迅速,限于编者的学识水平,书中难免有疏漏和不足之处,敬请学界同行和读者批评指正。

编者

2002 年 8 月

目次

第1章 信息检索概论.....	1
1.1 基本概念.....	1
1.1.1 信息、文献、知识.....	1
1.1.2 文献类型.....	1
1.2 信息检索与检索系统.....	4
1.2.1 信息检索.....	4
1.2.2 检索系统.....	4
1.2.3 检索系统的组织.....	6
1.2.4 检索系统的构成.....	7
1.3 检索语言.....	8
1.3.1 检索语言的类型.....	8
1.3.2 体系分类法.....	9
1.3.3 人工语言和自然语言.....	11
第2章 信息检索方法和技术.....	14
2.1 信息检索方法.....	14
2.1.1 检索途径.....	14
2.1.2 检索方法.....	15
2.1.3 检索程序.....	16
2.2 数字信息检索技术.....	19
2.2.1 检索方式.....	19
2.2.2 检索功能.....	20
2.2.3 检索步骤.....	21
2.3 数字图书馆.....	22
2.3.1 数字图书馆的主要特征.....	22
2.3.2 数字图书馆的信息资源及其检索.....	23
2.4 网络环境下信息检索.....	24
2.4.1 信息检索模式.....	24
2.4.2 网络信息资源组织与揭示.....	24
2.4.3 网络信息检索策略.....	25
第3章 中文检索工具.....	27
3.1 《全国报刊索引》与检索.....	28
3.1.1 简介.....	28
3.1.2 印刷版.....	28
3.1.3 《全国报刊索引数据库》光盘版.....	30
3.2 《中国学术期刊》与检索.....	33
3.2.1 简介.....	33

3.2.2 检索指南.....	34
3.2.3 检索实例.....	38
3.2.4 《中国期刊题录数据库》检索.....	44
3.3 《中文期刊数据库》与检索.....	45
3.3.1 简介.....	45
3.3.2 检索指南.....	46
3.3.3 检索实例.....	50
3.4 《人大复印报刊资料数据库》与检索.....	53
3.4.1 简介.....	53
3.4.2 检索指南.....	53
3.4.3 检索实例.....	58
3.5 万方数据资源系统与检索.....	60
3.5.1 系统简介.....	60
3.5.2 科技信息子系统.....	61
3.5.3 数字化期刊子系统.....	67
3.5.4 商务信息子系统.....	70
3.5.5 万方数据公司单一类型数据库简介.....	72
3.5.6 检索实例.....	73
第4章 综合性检索工具.....	78
4.1 美国《工程索引》与检索.....	78
4.1.1 概况.....	78
4.1.2 《工程索引》印刷版.....	79
4.1.3 《工程索引》光盘版.....	89
4.1.4 《工程索引》网络版.....	101
4.2 美国《化学文摘》与检索.....	112
4.2.1 概况.....	112
4.2.2 印刷版.....	113
4.2.3 光盘数据库.....	128
4.3 英国《科学文摘》与检索.....	134
4.3.1 概况.....	134
4.3.2 《科学文摘》印刷版.....	135
4.3.3 《科学文摘》光盘版.....	143
4.3.4 《科学文摘》网络版.....	145
4.4 美国《科学引文索引》与检索.....	149
4.4.1 概况.....	149
4.4.2 《科学引文索引》印刷版.....	150
4.4.3 《科学引文索引》电子版.....	163
第5章 专业性检索工具.....	173
5.1 英国《食品科技文摘》与检索.....	173

5.1.1 概况.....	173
5.1.2 印刷版.....	173
5.1.3 FSTA 数据库及其使用.....	177
5.2 美国《数学评论》与检索.....	186
5.2.1 概况.....	186
5.2.2 印刷版.....	188
5.2.3 数学文现库及其使用.....	193
5.3 美英联合《金属文摘》与检索.....	198
5.3.1 概况.....	198
5.3.2 印刷版.....	199
5.3.3 《金属文摘》光盘数据库.....	204
5.4 英国《建筑期刊索引》与检索.....	208
5.4.1 概况.....	208
5.4.2 编排与结构.....	208
5.4.3 检索方法与实例.....	209
5.5 美国《应用力学评论》与检索.....	210
5.5.1 概况.....	210
5.5.2 编排结构.....	210
5.5.3 检索方法与实例.....	212
5.6 美国《生物学文献》与检索.....	213
5.6.1 概况.....	213
5.6.2 印刷版.....	214
5.6.3 BIOSIS Previews 生物学文献数据库.....	221
第6章 单一类型文献及其检索	226
6.1 专利文献及其检索.....	226
6.1.1 专利基本知识.....	226
6.1.2 国际专利分类表.....	229
6.1.3 中国专利文献及其检索.....	233
6.1.4 因特网上专利信息检索.....	242
6.1.5 德温特专利索引及其检索.....	244
6.2 会议文献及其检索.....	260
6.2.1 概述.....	260
6.2.2 世界会议.....	261
6.2.3 科技会议录索引.....	263
6.3 学位论文及其检索.....	271
6.3.1 概述.....	271
6.3.2 国际学位论文检索.....	271
第7章 因特网信息资源及其检索	274
7.1 因特网信息检索.....	274

7.1.1 WWW 服务.....	274
7.1.2 搜索引擎.....	276
7.1.3 搜索引擎介绍.....	278
7.2 中国高等教育文献保障系统 (CALIS)	284
7.2.1 中国学术资源网.....	284
7.2.2 其他数据库.....	287
7.3 Dailog 联机检索系统.....	291
7.3.1 Dialog 联机检索系统简介.....	291
7.3.2 Dialog 数据库简介.....	292
7.3.3 WWW 环境下的 Dialog 系统信息检索.....	294
7.4 OCLC New FirstSearch.....	297
7.4.1 简介.....	297
7.4.2 检索指南.....	298
7.4.3 检索实例.....	303
7.5 联机公共书目查询系统 (OPAC)	307
7.5.1 南昌大学图书馆 OPAC 系统.....	307
7.5.2 中国国家图书馆公共检索服务系统.....	309
7.5.3 清华大学图书馆联机公共检索目录查询.....	311
7.5.4 中国科学院图书馆书目数据查询系统.....	314
参考文献.....	315

第 1 章 信息检索概论

1.1 基本概念

1.1.1 信息、文献、知识

1) 信息: 信息一词来自拉丁词源, 意思是通知、报道或消息。在中国历史资料中, 信息一词最早出自唐诗, 是音信、消息的意思。一直沿用到 20 世纪中叶, 其科学含义才被逐渐揭示出来。事实上, 信息是事物的存在方式和运动状态的表征。信息对人类的最重要意义是它在人类认识和实践中的特殊作用。人类认识世界的过程, 实际上就是不断地从外界取得信息和加工信息的过程; 而人类改造世界的过程, 则是把加工外部信息所得到的“主观”信息(表现为方针、政策、计划等等)反作用于外部世界的过程。因此, 没有信息, 人类就不能认识世界; 没有信息, 人类也不能有效地改造世界。

2) 数字信息: 通信系统传输和处理的对象, 泛指消息和信号的具体内容和意义。通常须通过处理和分析来提取。数字信息就是以数字化的形式存储和检索的信息。

3) 文献: 文献是记录有知识和信息的一切载体。由 4 个要素组成: ①所记录的知识和信息, 即文献的内容。②记录知识和信息的符号, 文献中的知识和信息是借助于文字、图表、声音、图像等记录下来并为人们所感知的。③用于记录知识和信息的物质载体, 如竹简、纸张、胶卷、胶片等, 它是文献的外在形式。④记录的方式或手段, 如铸刻、书写、印刷、复制、录音、录像等, 它们是知识、信息与载体的联系方式。

4) 信息管理: 是以信息组织为中心, 注重静态对象; 以文献信息为核心, 注重加工与保存、外部形态整合; 是对文献信息进行分类与检索、整序与整合、过滤与选择、应用与共享、分析与评价的过程。

5) 知识与知识管理: 知识是人类认识的成果或结晶。知识的特征有隐性的——存在头脑中的经验、直觉和社会网络; 镶嵌的——在程序、过程、系统和产品之中; 显性的——在数据库、文件和表格中(专利权、程序指南、科学作品和图表)。知识管理是指将各种信息转化为知识, 并将知识与人的需求和决策联系起来的过程。它以知识创新为中心, 注重动态过程。以人和用户为核心, 注重应用与共享、知识资源增值。知识管理的目标是确保知识积累、知识获得、知识导航、知识开发、知识共享、知识创新和知识保护。

1.1.2 文献类型

1. 文献按载体形式分

可分为印刷型、缩微型、电子型和声像型 4 种。

(1) 印刷型文献 以纸张为载体, 以手写、印刷为记录手段而产生出来的一种传统的文献形式。其优点是便于阅读和流传, 不受时间、地点和条件的限制; 其缺点是存储密度太低, 体积笨重, 所占储藏空间过多, 不易管理和保存, 难以实现自动输入和自动检索。

(2) 缩微型文献 以感光材料为存储介质,以缩微照相为记录手段而产生出来的一种文献形式。它包括缩微胶卷、缩微胶片和缩微卡片,随着激光和全息照相技术的应用,又出现了超级缩微胶片和特级缩微胶片,其最高缩小倍率已达 $1/22500$ 。一张全息胶片可存储 20 万页文献。其优点是体积小,存储密度高,传递方便,可以大大节省储藏空间;其缺点是不能直接阅读,需要借助缩微阅读机阅读。

(3) 电子型文献 是指经过数字化技术处理后存储在一定的特殊介质上并被计算机阅读和检索的文献资料。电子文献的基本特征是:文献的文本是以计算机可读数据的形式加以表示的;电子文献的信息载体是磁性或光学信息存储介质;电子文献经过格式化处理因而具有数据库结构;电子文献的使用要有相应的检索软件和其他编辑软件的支持并借助于计算机进行。其优点是存储密度高,存储速度快,可对所记录的信息进行各种处理,如转存、检索、传送、提取、检测和输出等;其缺点是使用费用较高。

(4) 声像型文献 又称为视听资料。这是一种以磁性材料和光学材料为存储介质,以电磁转换或光电转换技术为记录手段,直接记录声音、图像的一种文献形式。它包括:唱片、录音带、幻灯片、电影片、电视片、录像带、激光唱盘、激光电视录像盘、多媒体学习工具、程序设计工具等等。声像型文献的优点是存储密度高,内容直观,表现力强,易理解接受,传播效果好。这类文献在帮助人们观察罕见的自然现象、探索物质结构等方面,具有独特的作用。

在上述几种载体中,印刷型文献具有基本的和首要的意义,电子型文献是发展方向。本书所涉及的内容,以印刷型文献和电子型文献为主。

2. 按文献内容的加工与传递层次分

可分为一次文献、二次文献和三次文献。

(1) 一次文献 是指作者以本人的研究成果为基本素材而创作(或撰写)的文献。一般一次文献记录的信息比较具体、详尽和系统化,大部分期刊论文、科技报告、专利说明书、会议论文等都是—次文献。习惯上一次文献也称为原始文献。

(2) 二次文献 是指文献工作者对一次文献进行加工整理之后所得到的产物,是为了便于管理和利用一次文献而编辑、出版和累积起来的工具性文献。它包括书目、索引、文摘等。二次文献的重要性在于它可以帮助人们查找一次文献内容,提供一次文献的线索。

(3) 三次文献 是指利用二次文献,选用一次文献内容,而编纂出的成果。如百科全书、年鉴、手册、专著、教科书、述评、文献指南等等。

从一次文献到二、三次文献是一个从分散的原始文献加工整理系统化的过程。一般说,一次文献是基础,是检索的对象;二次文献是检索一次文献的工具;三次文献是一次文献的浓缩。

3. 按出版形式分

可分为图书、期刊(连续出版物)、会议文献、科技报告、技术标准、产品样本、专利文献、学位论文、档案文献、政府出版物等 10 大类型。也可将非书非刊的其他文献统称为特种文献,即图书、期刊和特种文献三大类。

(1) 图书 包括专著、教科书、参考工具书、丛书等。其特点是:内容比较系统、全

面、成熟、可靠，有一定的新颖性，但编辑出版时间过多，传递信息的速度太慢，时间上难以满足获取最新科技信息的要求。

(2) 期刊 是一种周期性的连续出版物。它具有品种多、数量大、出版周期短、报道速度快、发行流通面广泛、连续性强、内容新颖、能及时反映当前科技水平等特点。一直是科技人员最重视的信息源之一。

(3) 科技报告 是关于某项科研成果的正式报告，或者是研究过程中每一个阶段进展情况的实际记录。它的特点是每份报告自成一册，有连续编号，内容专深、详尽，由于它是研究的记录和成果，因而代表着一个国家和该专业的科研水平，对科研工作能起直接借鉴作用。

(4) 会议文献 在科技会议上科技工作者发表的论文，以及讨论记录等材料，多数都由主办学会以会议记录的形式全文发表。此类文献传递信息比较及时，传递的信息针对性较强，能反映科技的最新成就和发展趋势。会议文献一直受到科技界的高度重视，成为科技信息的又一重要来源。

(5) 专利文献 一切与专利制度有关的各种专利文件统称为专利文献。专利文献包括专利说明书、专利公报、专利文摘、专利分类表、与专利有关的法律文件及诉讼资料等。狭义的专利文献是指专利说明书及其所派生的各种二次文献（专利公报、文摘），主要是指专利说明书。其特点是内容广泛、完整实用、描述详细、新颖及时。对同一项技术来说，专利文献报道的速度比一般科技文献快得多。专利文献是一种兼具法律、经济、技术三方面效能的文献资料，其中蕴藏着丰富的法律信息、经济信息和技术信息。无论是科研、设计、管理和对外贸易人员都能从利用专利文献中受益。

(6) 学位论文 是高等学校博士和硕士研究生或本科生撰写的评定学位的论文，主要指博士和硕士论文。论文质量参差不齐，有的论文水平一般，有的论文在某些方面有其独到的见解，有一定的学术参考价值。

(7) 技术标准 主要指对工农业产品和工程建设的质量、规格及其检验方法等方面所作的技术规定。是从事生产、建设所必须遵守的技术法规，有一定的法律约束力。随着经济条件和技术水平的提高，标准常常要进行修订，或者以新标准代替旧标准，阅读时要选取最新标准。

(8) 产品样本 是制造厂商为了推销产品而印发的以介绍产品性能为主的出版物。它图文并茂，形象直观，出版发行迅速，有些产品资料反映的技术较为成熟可靠，便于科技人员选型、仿制及设计新产品时参考。

(9) 档案文献 它是在生产建设中和科技部门的技术活动中形成的，有一定工程对象的技术文件。其内容包括：任务书、协议书、技术经济指标和审批文件、研究计划、方案、大纲和技术措施；有关的技术调查材料、设计计算、试验项目、记录、数据和报告；技术鉴定书等其他技术资料。技术档案对以后从事相近或相似的专业研究活动，有直接的参考价值。但技术档案一般具有保密性，只能在内部借阅。

(10) 政府出版物 是各国政府部门及其所属机构所发表的文件。它的内容广泛，大致可分为行政性文件和科技文献两大类。后者约占 30%~40%。科技方面的政府出版物中，有许多在未列入政府出版物系统之前，往往已被所在单位出版过，因此跟其他文献类型（如科技报告）有重复。政府出版物对了解一个国家的科学技术和经济政策及其演变情况，有

一定的参考价值。

此外，还有报纸、科技译文、新闻稿件、手稿等。

4. 按电子文献载体形式划分

其类型有：磁性型文献、光盘型文献和网络型文献。

(1) 磁性型文献 也称计算机阅读型文献，它是通过编码和程序设计把文献变成计算机可识别的数字语言，即机器语言，储存在磁带、磁盘等载体上。阅读时，再由计算机将存入的内容转换成人能直接读的自然语言，即通过数字技术将存储在物理介质上的信息转化为数字信息。它的特点是存储容量大，并能按一定的程序设计快速输出文献单元及知识单元。

(2) 光盘型文献 光盘是一种利用激光技术在特制圆盘上记录（写成）和再现（读出）信息的载体。具体来说，利用激光作为光源照射在光盘表面，使表面物质产生变化而记录信息，在读出信息时，也是利用激光在光盘上产生出不同强度的反射光而判断信息内容所代表的是“1”或“0”。光盘可分为记录模拟信息的视频式光盘和记录数字信息的数字式光盘两种，分别大致相当于磁记录中的调频式记录与数字式记录。视频盘适合于记录彩色或活动图像，而数字式光盘则适合于记录分辨率要求较高的图像。数字式光盘采用了自动纠错技术，带信息的光盘可以无损地复制，其信息寿命接近于无限。

(3) 网络型文献 网络信息具有信息的时效性、内容的广泛性、访问的快速性、资源的动态性、搜索的相关性、组织的相对无序性。而网络型文献是指通过网络传播，并经过组织、制作成有序的数字信息文档。

1.2 信息检索与检索系统

1.2.1 信息检索

信息检索（Information Retrieval）是指将信息按一定的方式组织和存储起来，并根据信息用户的需要找出有关的信息的过程和技术。它的全称又叫“信息存储与检索”，这是广义的信息检索。狭义的信息检索则仅指该过程的后半部分，即从信息集合（库）中找出所需要的信息的过程，相当于人们通常所说的信息查询。在这里，存是指一种面向来自各种渠道的大量信息而进行的高度组织化的存储。而所谓取，就是面向随机出现的信息需求而进行的高度选择性的检索，要求能快速便利地查找与需求有关的信息。

信息检索的基本原理，我们可以用一句话把它最本质的部分概括为：对信息集合与需求集合的匹配与选择。即从用户需求出发，对一定的信息集合采用一定的技术手段，根据一定的线索与准则找出相关的信息。

1.2.2 检索系统

信息检索系统（Information Retrieval System）：利用一定设备和方法从某种载体（书本、胶卷、磁带或光盘、网络）上的文献、事实或数值记录集合中查找所需信息的系统。信息检索系统具有存储和检索两种基本功能。存储，即将搜集到的文献、数据进行分析、加工，给出检索标识和地址，组成卡片目录或记录在磁带或光学的载体上组成数据库。检索，即人工地或自动地从书本式的检索工具或数据库中查找所需的信息。因此，信息的存

储和检索是检索系统的两个不可分割的环节。

检索系统按其功能划分,可分为3种类型:目录检索、文献检索和事实检索。

1. 目录检索系统

目录性检索系统是对出版物进行报道和对图书资料进行科学管理的工具。有出版发行目录、馆藏目录、资料来源目录等之分。根据文献出版类型,又可分为图书目录、报刊目录、标准目录等。

馆藏目录反映图书资料的收藏情况,是用户经常使用的一种目录,以此来确定原(全)文的收藏地点。

馆藏目录可以是单个图书信息中心的,也可以是许多图书信息中心的馆藏的集合。后者称作联合目录,它可以反映一个地区、一个集团以至一个国家的图书资料收藏情况。目录系统的载体也是多样的,手工检索的目录有卡片目录和书本式目录。计算机检索的电子版目录包括机读目录以及在网络上运行的联机公共检索目录(OPAC)。联机公共检索目录(OPAC)可提供联机目录查询,网络把许许多多图书信息中心连接起来,提供地区性的乃至全球性的目录检索服务。

2. 文献检索系统

传统的文献检索系统是采用对自然语言事先规范而形成的受控语言(如分类表、主题词表),来描述文献信息特征、生成概念及其概念标识系统,人们通过分类表中的分类符号或主题词表中的主题词(或叙词)作为控制检索的入口格式进行检索。

文献检索系统提供对参考文献的检索,检索结果往往是一些可提供研究课题使用的参考文献的线索,即来源信息。它是一种间接的相关性检索,即只能查出与用户提问相关的文献,供用户参考,而不能直接解答用户所提出的问题。

从文献检索到原文获取,往往需经过目录检索这一环节,从已检出的文献出处(如刊名及卷期号)等来源线索去查找馆藏单位,索取原文。

索取原文的方法有:①由馆藏目录获得馆藏信息;②从联机公共检索目录(OPAC)所提供的联机目录查询、目录检索服务,从而获得馆藏信息。

目录检索与文献检索系统总称书目检索,其数据库称为书目数据库,也称二次文献数据库,它们提供的信息是文献的线索,而不是文献全文。

3. 事实检索系统

事实检索系统包括对事实、数据和全文的检索。从广义上讲,事实和数据也是一种全文,只是内容比较特殊、比较简短的全文。事实检索提供事实、数据等信息的原始资料,是一种直接的确定性检索。

事实检索的手检工具有:词典、百科全书、年鉴、手册、指南、名录等。与之相对应的计算机检索系统称源数据库。它主要有以下4种类型:

(1) 数值数据库 专门提供数值方式表示的数据,直接提供解决问题所需的数据,是进行各种分析、统计、决策的重要工具。如统计数据库、财务数据库等。

(2) 文本-数值数据库 能同时提供文本信息和数值信息。如产品数据库、商情数据库

等。

(3) 属性数据库 提供物理、化学、生物等科学数据。

(4) 全文数据库 提供文献的全文或其中的某些部分。如报刊全文库、法律法规全文库等。

1.2.3 检索系统的组织

检索系统是有序的文献信息集合。每篇文献都需要经过加工,把文献的特征描述下来,成为一条著录款目,亦称一条记录。将一个个记录按一定序列编排起来便组织成一个可供检索的系统。

文献检索系统按其检索手段划分为手工检索系统和计算机检索系统;按其报道范围划分为多学科的综合性检索系统和专业性检索系统;按其文献来源划分为多种出版物类型检索系统和单一出版物类型检索系统。

1. 手工检索系统

根据著录方式不同,手工检索系统可分为4种:目录、题录、文摘、索引。

(1) 目录 是图书、期刊等单位出版物外表特征的揭示和报道,是按某种顺序编列的文献清单。目录通常是以一个完整的出版单位或收藏单位为著录的基本单位,即以文献的“本”“种”或“件”(如一种期刊或一件专利)为报道单位。它对文献的描述比较简单,每条款目的著录项目有:著者/编者、书名/文献名和出版事项等。

(2) 题录 是单篇文献外表特征的揭示和报道。题录通常是以一个内容上独立的文献单元(如一篇文章或书中的某一部分,甚至整本出版物)为著录的基本单位。每条款目的著录项目通常包括:篇名、著者和来源出处。

(3) 文摘 是带有内容简介的、扩展了的题录。其著录项目比题录至少多一个文摘项。文摘也是以一个内容上独立的文献单元为著录的基本单位,它不仅描述文献外表特征,而且还包括文献的内容特征。文摘按揭示文献内容的详细程度可分为报道性文摘、指示性文摘和评论性文摘。

(4) 索引 是对一组信息集合的有系统的指引,一般只起指引特定信息内容及其存储地址的作用。索引款目(条目)一般有2个著录项目:检索标识(标目)和存储地址。检索标识是索引款目所指示的文献某方面的特征,其属性值有著者姓名、主题词等。存储地址是指文献属性值所对应的特定信息内容在信息集合中的地址,多数是流水号(如文摘号)。不同的标识系统构成不同的索引,有主题索引、分类索引、著者索引、关键词索引、图书索引等等。比如主题索引就是以主题词作标识(目)。索引不能作为一个单独的检索系统存在,它是目录、题录、文摘检索系统中的一个重要组成部分,在检索系统中占有重要位置,它提供了从不同角度(检索点)对特定信息集合检索的途经。

2. 计算机检索系统

计算机检索系统由硬件、软件和数据库组成。就检索而言,检索系统是指数据库。其数字检索机制主要包括4个部分:①检索界面模块:接受用户检索要求,有一般检索界面和高级检索界面。②检索策略模块:将用户输入的检索要求编制成计算机可执行的规范化检索式。③检索执行模块:利用检索式检索索引数据库,并保证检索的速度和准确性。④

检索结果组织模块：对检中记录的整理组织。

计算机检索系统有题录型、文摘型、指南数值型、全文型 4 种数据库。

(1) 题录型数据库 数据库存储某个领域文献的题录，包括文献的题名、作者、出处、主题词等。如：维普公司题录、文摘版数据库；德温特世界专利索引数据库等。

(2) 文摘型数据库 存储文献内容中主要部分的数据库，包括文献的题名、作者、出处、关键词、文摘等，基本上每一种书本式的检索工具都同时出版电子版，供用户在计算机系统中建立数据库。如：美国工程索引数据库、科学文摘数据库、化学文摘数据库等。

(3) 指南、数值型数据库 数据库的记录主要是指关于一些机构、公司、企业、名人、结构、主要产品及其产量、价格、规格型号等信息的简单描述，通过这些数据库可以查到公司、机构的地址、产品目录、研究项目或名人简历等信息。如：万方数据公司的中国企业、公司及产品数据库 (CECDB)、维普公司的中文期刊 (引文索引版) 数据库等。

(4) 全文型数据库 数据库存储文献内容全文或其主要部分的数据库。如：中国学术期刊 (全文版) 数据库、维普报讯 (全文版) 数据库等。

1.2.4 检索系统的构成

检索系统的构成一般是指手工检索系统和计算机检索系统的内容构成。

1. 手工检索系统

手工检索系统通常称之为书本式检索工具，一般由正文部分、辅助索引、使用说明和附录 4 个部分组成。

(1) 正文部分 作为手检工具的主体，正文部分是文献内、外部特征著录款目 (条目) 排成的有序集合。每一篇入存文献，只作一条款目，每条款目都有一个号码以与其他款目相区别。由于这一号码同时起排检顺序的作用，故称为顺序号或文摘号。不论什么检索工具，著录款目至少要表述文献的篇名、著者和文献来源 (出处) 等几个重要著录项目。大多数检索工具的正文部分按照学科分类体系的序列编排组织，因此提供了分类检索途径 (检索点)。

(2) 辅助索引 辅助索引是针对正文单一线索检索方式的补充。索引的种类越多，检索的途径就越多。常见的有著者索引和主题索引。按出版时间划分，索引有：期索引、卷索引、年度和多年度累积索引。

(3) 使用说明 使用说明是用户使用某种检索工具的入门向导与指南。一般包括编制目的、学科范围、收录年限、著录方式和使用方法。

(4) 附录部分 其内容包括：引用期刊一览表、刊物名称的简称与全表、机构名称缩写表、分类表和主题词表等。分类表和主题词表是用户用于浏览、确定分类类目、类号和主题词的工具，分类号 (类目名)、主题词作为检索点，用于正文或索引的准确检索。

2. 计算机检索系统

计算机检索系统由硬件、软件和数据库组成。数据库由字段、记录、文档组成。

(1) 字段 (Field) 书目数据库中基本的信息单位著录项目在数据库中称作字段。它是文献著录加工的最基本单元，描述的是文献外部特征和内容特征。字段包括：文摘字段、题名字段、著者字段等等。每一字段均有其标识符。

(2) 记录 (Record) 在数据库中的一篇文献称为一条记录,它是由若干个不同的著录单元(即字段)组成的。在数据库中每一条记录都有一个记录号,惟一地标识这条记录。

(3) 文档 (File) 由某一类型的若干条记录组成的信息集合称之为文档。它是数据库的基本形式,包括顺排文档和倒排文档。

顺排文档指文档中的记录按序存放,记录的存取按物理顺序进行。书目数据库中的主文档通常是顺排文档,主文档与手工检索工具中的正文部分相对应,是数据库的主体。主文档像手工检索工具一样,需要辅助索引的配合即倒排文档的配合。

倒排文档是将每篇文献记录中的标识(如文献的主题词、题名、著者等)抽取出来,同时记录下该检索标识出现过的文献记录号,然后再按序组织起来成为可以用作索引的文档,这种倒排文档亦称索引文档。如对著者字段,则将其所有记录字段中的著者名抽取出来,按序组织成著者倒排档(即著者索引),以指引与特定著者有关的记录在主文档中的地址。倒排文档又可以分为主索引文档和辅助索引文档。

数据库是计算机信息检索系统的组成部分。它是在计算机的存储设备上合理存放的相互关联的数据的集合。通常,由存储信息记录及其索引的若干文档组成。不同的数据库,存储不同的主题内容、不同的时间范围和信息类型以及不同的标引方式,提供不同的检索范围和检索途径。因此,掌握了数据库的特点,就可以自由灵活地运用检索技术,确保检索效果。

1.3 检索语言

检索语言 (Retrieval Language) 是用于描述检索系统中文献主题内容特征及外表特征和表达用户信息提问的一种专门语言。在文献检索的存储和检索过程中,检索语言起着重要的语言保障作用,是沟通文献存储和检索两个过程、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内容的表达(标引用语)和检索人员对同一内容文献需要的表达(检索用语)取得一致,文献检索也就无法进行。所以检索语言实质上是标引和检索之间的约定语言。使用检索工具必须掌握检索语言,如同使用计算机必须掌握计算机语言一样。掌握检索语言是掌握检索技能的基础。

检索语言作为文献存储检索系统的一个要素,在其中起着语言保证作用。其基本功能是:

- 1) 对文献的情报内容(及某些外部特征)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。

1.3.1 检索语言的类型

检索语言按其组配方式来划分,可分为先组式检索语言和后组式检索语言;按其表达文献特征来划分,可分为描述文献外表特征的检索语言和描述文献内容特征的检索语言。描述文献外表特征的检索语言,最常用的有引文语言、题名(书名、刊名、文献篇名等)、著者、文献号码(入藏登记号、ISSN号、专利号、报告号等)。描述文献内容特征的检索