



21世纪普通高等学校信息素质教育系列规划教材

数字信息资源检索 方法与实践 (理工版)

◎ 主 编 戚 敏 梁晓天

◎ 副主编 储冬红 廖 球 何海钊



华中科技大学出版社

<http://www.hustp.com>



21世纪普通高等学校信息素质教育系列规划教材

数字信息资源检索 方法与实践 (理工版)

◎ 主 编 戚 敏 梁晓天
◎ 副主编 储冬红 廖 球 何海钊



华中科技大学出版社

<http://www.hustp.com>

中国·武汉

内 容 提 要

本书是面向理工科高等院校的信息检索与利用课程的通用教材,立足于务实、易懂与新颖,全面揭示各种文献类型的理工类数字信息资源,介绍了外文图书、中外文期刊、学位论文、会议论文、专利、标准、科技报告的检索方法及综合学科、综合文献类型的数据库的检索方法,对一些常用的、重要的检索工具做详细介绍,展示检索工具最新的内容及功能,加强了开放获取信息资源、专业免费资源的介绍,全面深入介绍了信息鉴别及信息利用的内容。全书加入大量案例,将检索方法实践有机结合,为读者提供了全面、实用的信息检索与利用知识。

本书可作为理工科高等院校高年级本科生、研究生的教材,又可作为教学、科研工作者、各界信息用户的参考书。

图书在版编目(CIP)数据

数字信息资源检索方法与实践(理工版)/戚 敏 梁晓天 主编. —武汉:华中科技大学出版社,2011.4

ISBN 978-7-5609-6904-6

I. 数… II. ①戚… ②梁… III. 计算机应用-情报检索 IV. G252.7

中国版本图书馆 CIP 数据核字(2011)第 016810 号

数字信息资源检索方法与实践(理工版)

戚 敏 梁晓天 主编

策划编辑:袁 冲

责任编辑:张 琼

封面设计:刘 卉

责任校对:张 琳

责任监印:张正林

出版发行:华中科技大学出版社(中国·武汉)

武昌喻家山 邮编:430074 电话:(027)87557437

录 排:武汉兴明图文信息有限公司

印 刷:华中科技大学印刷厂

开 本:710mm×1000mm 1/16

印 张:22.25

字 数:450 千字

版 次:2011 年 4 月第 1 版第 1 次印刷

定 价:35.00 元



本书若有印装质量问题,请向出版社营销中心调换
全国免费服务热线:400-6679-118 竭诚为您服务
版权所有 侵权必究

总 序

自 1984 年国家教委(现更名为教育部)发布《关于在高等学校开设文献检索与利用课的意见》文件以来,文献信息检索教育在高校陆续展开。经过 20 多年的发展,文献信息检索课程内容逐步丰富。尽管各校开设该课程的部门可能不同,课程名称各异,发展状况也有很大不同,但是毫无疑问,信息检索教育已成为大学生信息素质培养和提高的一个重要途径。

信息素质教育不仅能培养大学生的文献信息检索技能和计算机应用技术,更重要的是能培养大学生对现代信息环境的理解能力、应变能力及运用信息的自觉性、预见性和独立性。从信息素质教育的本质内容来看,它是以培养创新能力为目标,培养学生猎取信息、加工信息和处理信息能力的教育活动,其最终目标是培养学生利用信息解决实际问题,并在此过程中实现创新的能力。

为了满足不同地区、不同层次、不同专业的信息检索的教学需要,推动信息素质教育的发展,我们成立了 21 世纪普通高等学校信息素质教育系列规划教材编委会,在专家的指导下,分批组织编写适合新形势、体现新成果、反映新思路的信息检索课程教材。

首期推出的有《大学生信息素养教程》、《信息检索基础教程》、《文理信息检索与利用》、《数字信息资源检索原理与实践(理工版)》、《数字信息资源检索原理与实践(社科版)》、《网络信息检索》、《现代信息资源检索案例化教程》、《经济信息检索》、《医学信息检索》等九本教材。

我们力求这套教材能突出如下特点。

(1) 选材新。作者选取的工具书的版本是常见的、容易获得的;网络信息资源是最新的,能够体现电子信息资源最新的建设成果。

(2) 分层次编写。专科层次要求内容组织上以够用为原则,适当兼顾新生入馆教育。本科层次强调系统性,兼顾课堂教学和课外拓展。

(3) 内容实用。不同教材的内容侧重有所不同,根据面向专业对象的不同而选材,强调实用。有的着重介绍文理信息检索基础知识,有的侧重文科信息检索,有的则重点介绍科技文献检索的内容,等等。

(4) 突出实践内容。尽量减少纯理论知识的介绍,合理分配检索理论和实践内容的比例,重点向读者介绍检索技巧、具体数据库检索技能,适当增加案例内容,由检索任务引领,以工作过程为导向,以活动为载体,按照检索流程组织教材内容,使读者在实践中掌握相关知识,培养、提高读者的检索能力。

(5) 强调直观性,便于自学。充分考虑检索课教学课时较少的现状,在教材中



编写,7.1.3由罗小芬编写,7.2.1至7.2.4由何海钊编写,7.2.5由储冬红编写,7.2.6的第1点至第3点由何海钊编写,及7.2.6的第4点至第10点由储冬红编写;第8章由廖球编写。戚敏、梁晓天进行全书的组织策划、体系拟定、统稿审定工作。

在编写过程中,参考了许多教材、课件、论文和网页,由于篇幅所限,参考文献中的一些网页、课件不能一一列出。在这里我们向所有参考文献的作者表示诚挚的谢意。

由于编者水平所限,书中难免有错误、疏漏之处,敬请同行、专家及读者批评指正。

编 者

2010年11月

前 言

随着网络技术、计算机技术和信息技术的飞速发展,E-learning、E-search 等以 Internet 为基础的学习方式渐渐成为自主学习的主流方式,如何获取和利用信息资源成为信息社会人们学习的前提和基础,同时,随着信息的多媒体化,对如何评价和理解信息提出了新的要求。信息素养作为终生学习的基础和促进因素,是所有学科、所有学习环境和所有教育水平都应共有的基本素养。拥有信息素养不仅能更好地掌握学习内容、扩展研究,而且还能对自己的学习进行自我指导和自我控制。就国家而言,信息素养已成为评价国民综合素质的一项重要指标。可以说,信息素养已成为个人和国家在信息时代的生存技能。为全面提高大学生的综合素质,国内外高校纷纷将信息素养教育作为培养人才的重要内容。信息检索课是实施信息素养教育的必修课程,目标是培养学生良好的信息道德修养和敏锐的信息意识,教授学生掌握基本的信息知识,增强学生各种获取信息的能力,尤其是学生信息利用和知识创新的能力。

为了适应当前高等学校课程建设和人才培养的需要,满足对大学生、研究生进行信息素养教育的需要,结合我国高校信息检索课程教学改革现状,我们在对现有的信息检索教材做了大量调查,并参考了大量相关文献的基础上,编写了此书。针对现有教材存在的实用性、可读性、新颖性不够强,介绍检索工具没有注意把检索技巧、检索实例结合到其中,缺乏对实际问题解决的实例,信息鉴别及信息综合利用部分很少提及或探讨不深入等问题,教材编写立足于务实、易懂与新颖,突出实践性。在检索理论部分只做概论论述,重点在尽可能全面揭示各种文献类型的理工类数字信息资源,对一些常用的、重要的检索工具做详细介绍,展示检索工具最新的内容及功能,书中配有大量的各种检索界面和检索结果的图像示例及标注说明,直观易懂;针对当代大学生的信息利用行为,加强了 Internet 信息资源和专业免费资源的介绍;书中加入多个案例,将检索理论与检索实践有机结合,增强可读性、趣味性,在让学生了解资源的基础上学会把对课题的分析、检索策略的选择、检索工具的运用融合在一起,使学生在学习和实践中提升信息素养,培养学生解决学习和科研中遇到的问题的能力,并转化为研究工作所应具备的基本技能。教材在信息鉴别及信息利用内容的广度和深度方面较现有教材有较大的突破,有助于提高学生综合利用信息的能力。

本书第 1 章由戚敏编写;第 2 章由贺梅萍编写;第 3 章除 3.2.3 的第 6 点由陈岚编写外,其余由戚敏编写;第 4 章 4.1 由罗小芬编写,4.2 由储冬红编写,第 5 章由廖球编写;第 6 章由储冬红编写;第 7 章 7.1.1 由石德万编写,7.1.2 由王凤翠



尽量采用图片、表格等形式展示知识要点和检索方法等,让读者能够按图索骥,现学现用。

随着计算机技术、网络技术、通信技术的快速发展和变化,电子文献迅速崛起,电子文献数据库和网络学术资源必将成为信息检索课程教学的主要内容。而这种变化又受制于学校的软件和硬件等实际条件,为了满足不同学校的教学需要,我们将适时对已经出版的教材进行修订改版,条件成熟的还将配备课件,提供案例、检索过程演示、题库等相关素材,并逐步创造条件建立网络资源平台,欢迎志同道合的专家和老师积极参与项目,我们的专用邮箱是 abcbook@126.com,检索课程教学探讨 QQ 群群号是 118459295。

这套教材的编写,得到了图书情报界同仁们的关心和支持,尤其是广大主编院校领导的帮助,在此,我们向大家表示由衷的感谢。相信随着更多专家和老师的参与,我们的教材会越来越完善。

编委会

2010 年 3 月

目 录

第 1 章 数字信息资源及其检索概述	(1)
1.1 数字信息资源概述	(1)
1.1.1 数字信息资源基本概念	(1)
1.1.2 数字信息资源类型	(1)
1.1.3 数据库类型	(3)
1.2 数字信息资源检索原理	(4)
1.3 数字信息资源检索技术	(6)
1.3.1 布尔逻辑检索	(6)
1.3.2 截词检索	(8)
1.3.3 字段限定检索	(9)
1.3.4 词位置检索	(10)
1.4 数字信息资源检索策略	(11)
1.4.1 制订检索策略的作用	(11)
1.4.2 制订检索策略的步骤	(12)
1.4.3 几种联机检索策略	(15)
1.4.4 检索策略的评价及调整	(17)
思考题	(19)
第 2 章 外文图书信息检索	(20)
2.1 外文图书书目信息检索	(20)
2.1.1 单个馆外文图书书目信息检索	(20)
2.1.2 主要外文图书联合书目信息检索系统	(27)
2.1.3 外文图书出版信息检索	(32)
2.2 外文电子图书信息检索	(36)
2.2.1 OCLC NetLibrary 电子图书检索系统	(36)
2.2.2 国际工程联合会电子图书数据库	(40)
2.2.3 Ebrary 电子图书数据库	(42)
2.2.4 Safari 电子图书服务系统	(44)
2.2.5 SpringerLink 线上丛书数据库	(47)
2.2.6 主要的电子工具书数据库	(51)
2.2.7 其他外文电子图书数据库	(62)



思考题	(64)
第3章 科技期刊信息检索	(65)
3.1 中外文科技期刊信息检索概述	(65)
3.1.1 科技期刊的特点及作用	(65)
3.1.2 科技期刊论文信息检索的基本思路与方法	(66)
3.2 中外文常用期刊二次文献库的特点及检索方法	(67)
3.2.1 国内常用期刊二次文献库的特点及检索方法	(67)
3.2.2 外文常用期刊二次文献库的特点及检索方法	(72)
3.3 中外文常用全文电子期刊数据库的特点及检索方法	(97)
3.3.1 中文常用全文期刊数据库的特点及检索方法	(97)
3.3.2 外文常用全文期刊数据库的特点及检索方法	(106)
思考题	(131)
第4章 学位论文、会议论文信息检索	(132)
4.1 学位论文信息检索	(132)
4.1.1 学位论文概述	(132)
4.1.2 中文学位论文数据库的特点与检索方法	(132)
4.1.3 外文学位论文数据库的特点与检索方法	(137)
4.2 会议文献信息检索	(142)
4.2.1 会议文献概述	(142)
4.2.2 中文会议论文数据库的特点与检索方法	(143)
4.2.3 外文会议论文数据库的特点与检索方法	(145)
思考题	(151)
第5章 开放获取文献及网上其他免费资源的检索	(152)
5.1 开放获取概述	(152)
5.1.1 开放获取及其特征	(152)
5.1.2 开放获取的发展历程	(153)
5.1.3 开放获取资源的实现途径	(155)
5.2 主要的开放获取资源的特点与检索方法	(156)
5.2.1 中文主要开放获取资源的特点与检索方法	(156)
5.2.2 国外主要开放获取资源的特点与检索方法	(160)
5.2.3 检索开放获取资源的工具	(164)
5.3 学科信息门户资源的特点与利用	(167)
5.3.1 学科信息门户资源及特点	(167)

5.3.2 国内学科信息门户资源导航系统的特点及利用	(169)
5.3.3 国外学科信息门户资源导航系统的特点及利用	(174)
5.4 网络免费学术信息的检索与利用	(177)
5.4.1 主要的搜索引擎的特点及使用	(177)
5.4.2 主要的联合参考咨询系统的特点及使用	(184)
5.4.3 主要的信息共享工程系统的特点及使用	(189)
思考题	(191)
第6章 特种文献资源的检索	(192)
6.1 专利文献的检索	(192)
6.1.1 专利文献基础知识	(192)
6.1.2 中国专利及检索	(198)
6.1.3 外文专利文献的检索与利用	(204)
6.2 标准文献的检索	(212)
6.2.1 标准文献基础知识	(212)
6.2.2 中文标准文献网站	(218)
6.2.3 外文标准文献资源的网站	(221)
6.3 科技报告的检索	(227)
6.3.1 科技报告基础知识	(227)
6.3.2 科技报告的检索	(229)
6.3.3 科技报告原文的获取途径	(233)
思考题	(234)
第7章 综合性数据库的检索与利用	(235)
7.1 中文综合性数据库的检索与利用	(235)
7.1.1 超星系列数据库	(235)
7.1.2 爱迪科森网上报告厅	(241)
7.1.3 综合考试系统	(243)
7.2 外文综合性数据库的检索与利用	(249)
7.2.1 多学科专业的综合性数据库	(249)
7.2.2 计算机科学及相关专业领域数据库	(254)
7.2.3 交通、航空航天及相关领域数据库	(259)
7.2.4 光学及相关领域数据库	(261)
7.2.5 地质和地球科学文献库	(265)
7.2.6 其他学科专业数据库	(268)



思考题.....	(295)
第8章 信息的鉴别与利用	(296)
8.1 信息的鉴别	(296)
8.1.1 信息鉴别的方法	(296)
8.1.2 网络信息资源的鉴别与评价	(299)
8.2 个人文献信息管理工具	(300)
8.2.1 国内个人文献信息管理工具的特点及利用	(301)
8.2.2 国外个人文献信息管理工具的特点及利用	(306)
8.2.3 其他个人文献信息管理工具的特点及利用	(308)
8.3 信息的分析利用	(310)
8.3.1 信息分析的概念	(310)
8.3.2 信息分析的方法	(311)
8.3.3 信息利用中的道德与法律规范	(313)
8.3.4 学术论文的撰写	(317)
8.3.5 科学研究的一般流程及信息利用案例	(330)
思考题.....	(338)
参考文献.....	(340)

第1章 数字信息资源及其检索概述

1.1 数字信息资源概述

1.1.1 数字信息资源基本概念

数字信息资源从狭义讲,亦可称为电子资源,指一切以数字形式生产和发行的信息资源。所谓数字形式,是以能被计算机识别的、不同序列的“0”和“1”构成的信息形式。数字资源中的信息,包括文字、图形、声音、动态图像等多种形式的信息,都是以数字代码方式存储在磁带、磁盘、光盘等非纸介质上,并通过网络通信、计算机或终端等方式再现出来的资源。目前,把通过网络传递的信息资源和仅在本地计算机上使用,没有通过网络传递的信息资源都归为数字信息资源。

1.1.2 数字信息资源类型

随着信息处理技术、计算机技术和网络技术的飞速发展,数字信息资源越来越丰富,种类越来越繁多,划分的依据不同会得到不同的分类结果。一般来说,数字信息资源类型可根据传统文献类型、信息载体、信息的交流表现形式、媒体信息类型等方式来划分。

1. 按传统文献类型分类

传统信息检索书上将文献分为一次文献、二次文献、三次文献等。根据这种分类方式,数字信息资源也有相应的分类。

1) 一次文献型

一次文献即原始文献,指未经他人加工、重组、压缩的原始信息。一次文献按出版的形式,可以分为图书、期刊、会议文献、学位论文、标准文献、专利文献、科技报告、产品资料、档案文献、政府出版物。一次文献型数字资源指各种数据库(单机、联机或网络的全文数据库)、电子期刊、电子图书、电子报纸,以及各种国际组织、政府机构、学术团体、教育机构、企业部门等单位在网上发布的信息等。



2) 二次文献型

二次文献指对一次文献进行加工、整理、简化,将可供检索的信息特征按一定的规律组织而形成的便于利用一次文献的信息资源,二次文献型数字资源包括书目、索引、文摘型的二次文献数据库、网络资源学科导航、搜索引擎、分类指南等。二次文献可以把大量分散的一次文献按学科或主题集中起来,组织成无数相关信息的集合,向公众报道原始信息产生和存在的信息;同时也是一种有效的检索工具,供用户查找信息线索之用。

3) 三次文献型

三次文献是对一次、二次文献进行综合分析、加工、整理后,按知识门类或专题综合加工而形成的信息资源,三次文献型数字资源有各种类型的网上电子参考工具书等。

4) 零次文献型

零次文献,如电子新闻、电子邮件、电子布告和电子论坛等。

2. 按所存储信息的载体分类

磁盘、光盘、硬盘是目前数字化信息的主要载体,因此,依所存储信息的载体,数字化信息可分为磁盘型信息资源、光盘型信息资源、网络型信息资源等三种。

1) 磁盘型信息资源

在电子化初始阶段,磁盘是数字化信息最原始、最主要的载体。

2) 光盘型信息资源

光盘是一种用激光来记录和再现信息的高密度存储介质。1985年,世界上第一个商品化的CD-ROM数据库——BiblioFile(美国国会图书馆的MARC,即机读目录)推出。光盘技术迅速发展,凭借其容量大、体积小、质量小、价格低并可同时存储声音、图像、文字,且携带、邮寄方便等优势,使信息处理、传播、检索和利用的模式发生了革命性的变化。

3) 网络型信息资源

20世纪60—70年代,世界发达国家和地区相继建立起计算机联机信息服务系统,如美国的Dialog、欧共体的ESA和德国的STN系统等,都为全世界联机用户提供了丰富的数字信息资源,但长期以来其费用昂贵,许多用户不敢问津。20世纪90年代以来,Internet迅猛发展,网络资源极大丰富,价格相对低廉,且有许多免费资源,这使得人们越来越多地从网络上检索、获取信息,通过网络来共享全球信息。

3. 按信息的交流表现形式分类

人类信息常以正式出版或非正式出版两种形式来表现,按人类信息交流的方式将电子信息资源划分为正式出版信息、半正式出版信息(即灰色信息)和非正式出版信息。



1) 正式出版信息

正式出版信息,是由正式出版机构、出版商、数据库商出版发行的,在数字学术信息资源中所占比例最大,包括各类数据库和电子期刊、电子图书等。其特点是:学术信息含量高;具备检索系统,便于检索利用;出版成本高,必须购买使用权才可以使用,并不是面向社会公众免费开放的。

正式出版的数字资源包括:①一次文献,主要有全文数据库、电子期刊、电子图书、电子报纸、电子名录、电子地图、数字音像和视听制品等;②二次文献,主要有书目数据库、文摘、题录数据库和引文数据库等;③三次文献,主要有事实数据库、数值数据库。

2) 半正式出版信息

半正式出版信息,又称为“灰色信息”,指受到一定产权保护但没有纳入正式出版信息系统中的信息,如各种学术团体和教育机构、企业和商业部门、国际组织和政府机构、行业协会等单位等在网上正式发布的网页及其信息,亦属于一次文献类型;这部分信息主要从搜索引擎/分类指南、网络学术资源学科导航、图书馆印刷型馆藏的联机公共目录(OPAC)等二次文献资源中获得。

3) 非正式出版信息

非正式出版信息指流动性、随意性较大的电子信息,信息的质量难以保证和控制的动态信息,如FTP资源、电子新闻、电子邮件、电子布告和电子论坛等。

4. 按媒体信息类型分类

媒体信息主要有文本信息、图形信息、图像信息、动画信息、视频信息等。根据媒体信息的种类,数字信息资源又可以分为文本信息资源、超文本信息资源、多媒体信息资源和超媒体信息等五种。

1.1.3 数据库类型

1. 全文数据库

全文数据库即收录有原始文献全文的数据库,是在二次文献数据库的基础上发展起来的数据库,所以一般的全文库都有文摘提供,此外还有全文可以下载。这样大大简化了信息寻找的过程和时间,因此受到了普遍欢迎,成为当今数据库的主流产品。全文数据库以收录期刊论文、会议论文、学位论文、政府出版物、研究报告、法律条文和案例、商业信息等的全文为主。

2. 二次文献数据库

二次文献数据库中的记录是通过对数据、信息或知识的再加工和过滤后形成的,如编目、索引、摘要、分类等。它只提供原文的相关信息,比如题名、作者、出处、摘要等,便于读者快速了解原文的内容,并能方便地找到它。

二次文献数据库主要包括:书目数据库、文摘数据库和索引数据库。书目数据



库主要用于揭示图书的基本信息,例如图书馆的馆藏联机目录数据库。文摘数据库则是对期刊论文、会议论文、专利文献、学位论文等类型文献的内容和相关信息的报道,往往具有强大的分析和挖掘功能,如新出现的 Scopus 数据库。索引数据库中的引文数据库则是从文献之间相互引用的独特角度出发,力图反映文献的质量及某个科学问题的发展历史的数据库,例如科学引文索引(SCI)、社会科学引文索引(SSCI)和我国的中文社会科学引文索引(CSSCI)都是索引数据库。

3. 事实和数值型数据库

事实和数值型数据库指包含大量事实、数据,直接提供原始资料的数据库,它的作用相当于印刷文献中的字典、词典、手册、指南等工具书。常见的数值数据库有:统计数据库、财务数据库、科技数据库、化学反应数据库等。常见的事实数据库有:公司信息库、产品市场报告、物性数据库等。百科全书电子版和网络词典也属于此类数据库。

4. 声音数据库

声音数据库主要以收集声音信息为主。

5. 图像数据库

图像数据库主要收集各种类型的图形、图像,如一些素材数据库等。

6. 视频数据库

视频数据库主要收集视频信息。

7. 多媒体数据库

多媒体数据库提供文本、图像(图表、图画、照片、动画或活动影视)、声音(语言、音乐或其他音响)及它们的结合等信息。

8. 超媒体数据库

超媒体数据库通过外部树形的链接将多种类型的媒体连成一个集合。

1.2 数字信息资源检索原理

计算机系统内的数字信息检索就是指人们在计算机和计算机检索网络或终端上,使用特定的检索指令、检索词和检索策略,从计算机检索系统的数据库中检索出所需要的信息,然后再由终端设备显示和打印的过程。计算机检索是以计算机为基础的信息的存储与检索,是在人和计算机的共同作用下去完成信息存取操作,从而迅速、准确地给用户提供符合特定需要的信息的一种方法与过程。

为实现这种信息检索,必须事先将大量的原始信息加工处理、存储在各种信息载体上待用,所以计算机信息检索广义上讲包括信息的存储和检索两个方面。

信息存储的过程就是将所选中的一次文献进行主题分析、标引和著录,按一定格式输入计算机,构成机读数据库记录及文献特征标识,为检索提供经过整序(即形成检索途径)的信息集合的过程。在存储文献时,文献信息标引人员要进行信息的标引和著录。信息的标引就是对文献信息的内容进行主题分析,使之形成若干能反映文献信息主题的概念,然后借助于检索语言(分类法、主题词表等)把这些概念转换成检索词汇,形成检索标识,如给出分类号或主题词的过程。所谓信息的著录是按照一定的规则对信息的外部特征和内容特征加以简单明确的表述(信息外部特征包括信息的著者、来源、卷期、页次、年月、号码、文种等,信息内容特征包括题名、主题词和文摘)。然后把这些经过“前处理”的数据按一定格式输入计算机存储起来,计算机在程序指令的控制下对数据进行处理,形成机读数据库,存储在存储介质(如磁带、磁盘或光盘)上,最后形成称为检索工具的检索体系(印刷型/电子型),完成信息的加工存储的过程。

信息检索则是存储的逆过程。检索人员对检索课题加以分析,确定所需信息的实质内容,形成主题概念,然后借助各种检索语言,把主题概念转换成检索词汇,并按实际需求把这些词汇之间的逻辑关系表达出来,形成检索表达式。从用户提问进行主题分析到检索表达式形成的这一过程,就是编制检索策略的过程。如同把存储阶段形成的标引记录看成是信息著录一样,也可以把检索阶段形成的检索表达式看成是用户提问著录。两种著录的区别在于,前者的著录结果是对信息内容和形式特征的逐一记录,而后的著录结果所表达的各主题概念之间通常含有逻辑性。检索表达式形成后,即可输入到计算机进行查找。这一查找的过程实际上是计算机自动比较、匹配的过程,当检索标识、检索策略与数据库中的信息的特征标志及其逻辑组配关系相一致时,则属“检索命中”,即找到了符合要求的信息。在手工检索过程中,相符性的比较是由人脑进行的,而在计算机检索过程中,则由计算机担负两者间的匹配工作。至此,一个具体的信息检索过程结束。计算机检索的基本原理如图 1-1 所示。

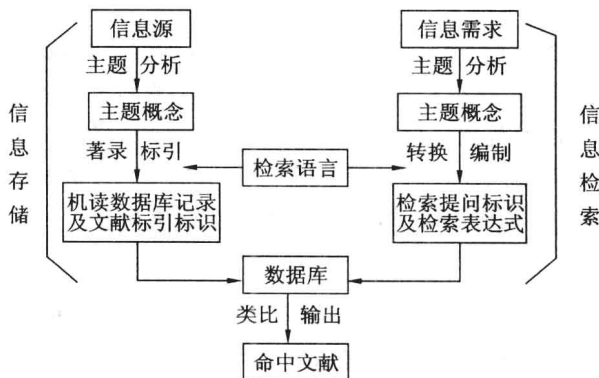


图 1-1 计算机检索的基本原理示意图



1.3 数字信息资源检索技术

1.3.1 布尔逻辑检索

在实际检索中,检索提问涉及的概念往往不止一个,而同一个概念又往往涉及多个同义词或相关词。为了正确地表达检索提问,系统中采用布尔逻辑运算符将不同的检索词组配起来,使一些具有简单概念的检索单元通过组配成为一个具有复杂概念的检索式,用于表达用户的信息检索要求。布尔逻辑检索是一种开发较早、比较成熟、最为流行的经典检索技术,它具有简单、易理解、容易在计算机上实现且检索速度快等优点,在许多逻辑检索系统中得到应用,例如 Yahoo、Google 等诸多网络检索站点均采用布尔逻辑检索技术。布尔逻辑检索是指通过布尔逻辑检索符号与检索词之间的逻辑关系而进行检索的方法,即采用布尔逻辑表达式来表示用户提问,通过对文本标识与用户给出的检索式进行逻辑比较来检索文档。常用的逻辑算符主要有布尔代数的“与”、“或”、“非”。

1. 逻辑或

逻辑或也称为逻辑加,用“OR”或“+”表示,是用来组配检索词的并列关系的。两个检索词以“OR”(或“+”)相连,表示要检索含有两词之一或同时包含两词的文献。比如说,要求查找计算机或机器人方面的文献,可向计算机检索系统提交下列提问逻辑式:“计算机+机器人”或者“计算机 OR 机器人”,运算的结果是含有“计算机”,或者含有“机器人”的文献均被检索出来。可见,用逻辑或扩大了检索范围。

2. 逻辑与

逻辑与也称为逻辑乘,用“AND”或“*”表示,是用来组配不同的检索词限定关系的。两个检索词以“AND”(或“*”)相连,表示要检索同时包含两词的文献,因而用逻辑与缩小检索范围。比如说,要求查找计算机软件方面的文献,可向计算机检索系统提交下列提问逻辑式:“计算机*软件”或者“计算机 AND 软件”,运算的结果是同时含有“计算机”和“软件”的文献才被检索出来。

3. 逻辑“非”

逻辑“非”用“NOT”或“-”表示,是用于从某一检索范围中排除不需要的概念。这种组配可以缩小检索范围。例如,在含有检索词 A 的文献集合中,排除同时含有检索词 B 的文献,可表示为:“A NOT B”或“A-B”。由于会丢失部分信