

医药学文献计算机检索

光盘 Internet 联机检索方法与技巧



主 编

于 泓
魏 良
张红萍

- 国内首部医学文献计算机检索实用工具书
- 循序渐进 图文并茂 实例丰富
- 全新 WINDOWS 版 MEDLINE CBM CMCC 光盘检索
- 全新 Internet 网上医学检索引擎及信息检索
- 首推 WINDOWS 版 CA 光盘检索

吉林科学技术出版社

医药学文献计算机检索

——光盘、Internet 信息资源和联机检索方法及技巧

主 编

于 泓 魏 良 张红萍

副 主 编

周晓政 高大林 李欣欣
蔡忆宁 栾晓艳 张宝珍

长 春

吉林科学技术出版社

内 容 提 要

本书为我国第一部专门论述计算机检索医学文献的实际操作、指导性用书。重点阐述了国内外 12 种生物医学及相关科学光盘数据库检索方法,并详尽介绍了 Internet 网上的各种一般检索引擎、医学检索引擎,还特别介绍了三个站点中免费 MEDLINE 的检索方法,以及最新国际联机检索方法。其中,中国生物医学光盘数据库(CBMdisc)、中文生物期刊数据库(CMCC)、MEDLINE 光盘数据库等 Windows 版本的操作方法介绍,以及美国化学文摘光盘数据库操作方法介绍,这在我国尚属首次。MEDLINE 数据库的资料介绍截止 1999 年 9 月。

全书共四篇三十章,其特点是:注重创新、循序渐进、图文并茂、实例丰富,科学性、新颖性、针对性、实用性较强。

本书可作为高等医药院校本科生、研究生的文献检索课教材,及教学、科研、医疗工作者继续教育的教材,还可作为图书情报信息专业教师、工作人员参考或自学用书,具有较强的可读性。

医药学文献计算机检索

于 泓 魏 良 张红萍 主编

责任编辑:齐 郁

封面设计:陈 勇

吉林科学技术出版社出版	787×1092 毫米	16 开本	444,000 字	18 印张
吉林科学技术出版社发行	1999 年 9 月第 1 版	1999 年 9 月第 1 次印刷		
六〇六所印刷厂印刷	ISBN 7-5384-2154-8/Z·76	定价:26.00 元		

地址 长春市人民大街 124 号 邮编 130021 电话 5635177 传真 5635185
电子信箱 JLKJCS @ public.cc.jl.cn

《医药学文献计算机检索》编委会

顾 问：吴观国 南京医科大学
董大钧 沈阳市医学信息中心
朱雪奇 江苏省医学情报研究所
郭继军 中国医科大学

主 编：于 泓 魏 良 张红萍

副主编：周晓政 高大林 李欣欣 蔡忆宁 栾晓艳 张宝珍

编 委：（按章节先后次序排列）

魏 良 中国医科大学
于 泓 沈阳医学院
张云秋 白求恩医科大学
侯跃芳 中国医科大学
周晓政 南京医科大学
黄亚明 中国医科大学
李欣欣 白求恩医科大学
张红萍 江苏省医学情报研究所
陈守鹏 南京中医药大学
高大林 江苏省医学情报研究所
栾晓艳 中国医科大学
蔡忆宁 江苏省科技情报研究所
于丹辉 长春图书馆
陈万福 江苏省医学情报研究所
徐 莉 大连大学
张宝珍 沈阳市医学信息中心
史春杨 沈阳医学院
邸 明 沈阳医学院
杨秀梅 沈阳医学院
曹伟平 沈阳市医学信息中心

序 言

教育部（国家教委）1984年起为在高校开设《文献检索与利用》课程而连续下达了三个文件，因此十几年来这个课程发展很快，取得了较大成绩。由于这是一门新的课程，又是实用性很强的方法课，要使之巩固，逐步提高教学质量，除了要充实设备及工具书刊的准备工作外，主要是要有一支高水平的教师队伍及能达到教学目的的教材。这些年来，医学文献检索教材已有十几种版本，能反映出文献课蓬勃发展的势头，但优秀成熟的并不多见。突出的问题是低水平的重复。现行教学模式多用讲授与实习相结合的方式，偏重介绍传统的文献检索基础理论知识和手工检索方法，局限于培养学生利用书本型检索工具查阅文献技能。教材也侧重描述介绍检索工具的编制体系结构，使用方法。在课程创始阶段，这样既是必要的，又是必然的。当时对学生学会基本检索知识和方法起了很大作用。但九十年代中期以来，随着现代科技的发展，计算机及光盘技术的应用，因特网的开通与普及，各种中外文检索数据库相继问世，利用计算机进行文献检索已逐步代替传统的手工检索。在这种形式下，广大的大学生、教师及医务工作者都感到在课堂上学到的检索理论和手工检索的传统方法已不能适应需要。《医药学文献计算机检索》适时出版了。这是一部有创新意义的书，专门论述计算机检索医学文献的实际操作的指导性用书，作为补充其他教材的不足而出版是非常必要的。本书实用性强，对那些想通过计算机自己动手操作来检索文献的医务工作者来说是十分有用的。

根据我国国情，文献检索课发展改革的方向，是由手工检索与向手检与机检相结合过度，就是以机检为主，手检与机检并存。因其各有利弊及特点，就要互相补充，取长补短，相辅相成以满足各层次读者要求。在相当长的时期内，机检不会完全代替手检，当前急需做的就是加大机检在课程中的比重及主导地位，而这就首先应从教材中体现。《医药学文献计算机检索》的出版，正是弥补了已出版的此类教材的不足。

尤其难能可贵的是，这本书的编写人员均为年轻的情报工作者和文献课的教师，都有十年以上的工作经历，具有较丰富的实践经验，绝大多数是医学情报或医学专业的毕业生，学有专长又是第一线工作人员，这就决定了本书是实用性、针对性及专业性强的著作，必将受到广大医务工作人员的欢迎。作为一个终身从事医学图书馆工作的老战士，又是我国正式出版的第一种医学文献检索教材的主编，为有这些能著书立说，热爱专业的接班人倍感欣慰。教材建设是一项长期而艰巨的任务，衷心祝愿他们百尺竿头更上一步，在不断实践的基础上，为今后文献检索课教材改革起到更大的作用。

吴观国

一九九九年八月于南京

前言

奋战了 10 个月的书稿终于杀青了,这是我们献给全国医学、生物学工作者的一份礼物,亦是作为工作在第一线的医学情报专业人员献给新中国诞辰五十周年的一束鲜花。

随着现代科技的发展,各种中外文检索数据库相继问世,计算机、光盘及因特网技术已经普遍应用于文献检索中,利用计算机进行文献检索已逐步代替传统的手工检索。在这种形式下,课堂上学到的检索理论和手工检索的传统方法已不能适应时代的需要。因此,在现实生活中常听到医科大学的研究生及本科生、医务人员的抱怨,在课堂上学到的检索理论,在实践中不会操作,迫切地需要一本能一步一步地教我们操作如何检索医学信息的书。本着重实用性、指导性的原则,我们全体编写人员在专家的指导下,共同努力,适时地为广大读者推出了《医药学文献计算机检索》一书。

本书为我国第一部专门论述计算机检索医学文献的实际操作、指导性用书。本书既可作为医学情报专业计算机检索的教科书,也可作为在职人员继续教育的教材。对于那些想通过计算机自己动手操作来检索生物医学文献的大学生、研究生、医务工作者来说,本书是一位极好的家庭教师。

本书系统、全面地介绍了国内外 12 种生物医学及相关科学的光盘数据库的检索操作方法。其中图例 225 幅,以图文并茂的形式阐述数据库的检索方法,重点突出实际操作方法。

本书内容新颖,首次介绍了中国生物医学光盘数据库(CBMdisc)、中文生物期刊数据库(CMCC)、美国化学文摘光盘数据库(CA)等 Windows 版本的操作方法;首次系统介绍了 MEDLINE 光盘数据库 Windows 版、Internet 上医学搜索引擎及免费 MEDLINE 数据库检索的操作方法;书中许多数据资料介绍均截止 1999 年 9 月。

本书实用性极强,除了 12 种光盘数据库的操作方法介绍外,因考虑到光盘数据库一般均很昂贵,从数千元到数十万元一套不等,只有单位才能购买。对于广大医学院师生和医疗卫生单位的科研人员来说,通过 Internet 国际互联网来检索,可免费获得医学

文献，这是一条捷径。本书详尽地介绍了 Internet 上的各种一般检索引擎、医学检索引擎，还特别介绍了三个站点中免费 MEDLINE 的检索方法。这样，使您在家中就免费拥有一座医学图书馆，且查询极为方便。

参加本书编写的同志均为工作在医学情报第一线，其中大多数的同志是我国第一代医学情报专业本科生，并运用计算机检索医学文献已有 10 年以上经验，可以说，本书是共同经验的综合，集体智慧的结晶。

人们常说，外语和电脑是科技人员通向 21 世纪的一对翅膀。通过计算机检索，我们可以在几分钟内搜索到过去需要花几天甚至几周才能获得的信息，通过 Internet 网我们可以轻而易举地获取万里之外国际上最新的医学信息。获取和跟踪信息的效率数十成百倍地提高，大大加快了科研工作的进程，加快了更新知识的速度，增强了教学、临床工作的能力。

我们的共同心愿就是为全国广大医务工作者和医学院校师生提供一本能够帮助大家自己动手利用计算机检索医学文献的书，愿我们的这本书能助您一臂之力。

本书在编写过程中参考了大量论著，对提高本书的质量起了重要的作用，谨此向这些论著作者深表感谢！

编 者

1999 年 9 月

目 录

第一篇 医学文献计算机检索概论

第一章 计算机信息检索的含义、原理及优点.....	(1)
第二章 计算机信息检索的简史及现状.....	(2)
第三章 计算机信息检索系统的构成.....	(5)
第四章 数据库类型及医学数据库.....	(6)
第五章 计算机信息检索的服务方式.....	(8)
第六章 光盘检索系统及其特点.....	(9)

第二篇 医学文献主题分析

第一章 主题分析的概念.....	(11)
第二章 主题分析的方法和步骤.....	(11)
第一节 文献或检索课题的审读.....	(11)
第二节 提炼主题概念.....	(11)
第三节 转换主题概念.....	(12)
第四节 合理组配.....	(13)
第五节 查找特征词.....	(18)
第三章 主题分析举例.....	(19)

第三篇 光盘检索数据库检索方法和技巧

第一章 MEDLINE 光盘数据库.....	(20)
第一节 数据库概述.....	(20)
第二节 数据库检索系统.....	(23)
第三节 检索途径及操作方法.....	(32)
第四节 检索策略.....	(45)
第二章 荷兰医学文摘光盘数据库 (EM)	(46)
第一节 数据库概述.....	(46)
第二节 数据库检索系统.....	(49)
第三节 检索途径及操作方法.....	(51)
第四节 实例分析.....	(57)
第三章 美国生物学文摘光盘数据库 (BA)	(63)
第一节 数据库概述.....	(63)
第二节 数据库检索系统.....	(65)
第三节 检索途径及举例.....	(67)

第四节 检索策略.....	(74)
第四章 美国化学文摘光盘数据库 (CA)	(75)
第一节 数据库概述.....	(75)
第二节 数据库检索系统.....	(76)
第三节 检索途径和方法.....	(79)
第五章 科学引文索引光盘数据库 (SCI)	(91)
第一节 数据库概述.....	(91)
第二节 数据库检索系统.....	(92)
第三节 检索途径和方法.....	(93)
第四节 检索限定及检索结果输出.....	(96)
第五节 实例分析.....	(99)
第六章 国际药文学文摘光盘数据库 (IPA)	(103)
第一节 数据库概述.....	(103)
第二节 数据库检索系统.....	(105)
第三节 指令检索介绍.....	(111)
第四节 实例分析.....	(115)
第七章 中国生物医学文献光盘数据库 (CBMdisc)	(118)
第一节 数据库概述.....	(118)
第二节 数据库检索系统.....	(118)
第三节 检索途径及操作方法.....	(125)
第四节 检索策略及检索举例.....	(134)
第八章 中文生物医学期刊数据库 (CMCC)	(137)
第一节 数据库概述.....	(137)
第二节 数据库检索系统.....	(137)
第三节 检索途径及操作方法.....	(147)
第九章 中国科学引文索引光盘数据库 (CSCI)	(152)
第一节 检索简介.....	(152)
第二节 检索途径.....	(152)
第三节 检索方法.....	(153)
第四节 显示 / 输出.....	(158)
第五节 实例分析.....	(163)
第十章 中国学术期刊 (光盘版) 全文数据库.....	(168)
第一节 数据库概述.....	(168)
第二节 数据库检索系统.....	(168)
第三节 检索途径及操作方法.....	(170)
第十一章 中国中医药文献光盘数据库.....	(176)
第一节 数据库概述.....	(176)
第二节 数据库检索系统.....	(176)
第三节 检索途径及操作方法.....	(180)
第十二章 中国生物学文献光盘数据库 (CBA)	(186)

第一节	数据库概况.....	(186)
第二节	数据库检索系统.....	(188)
第三节	检索途径及操作方法.....	(193)
第四节	实例分析.....	(196)

第四篇 Internet 网络生物医学信息检索方法及技巧

第一章	Internet 简介.....	(201)
第一节	Internet 概况.....	(201)
第二节	如何联入 Internet.....	(201)
第三节	Internet 基本功能介绍.....	(202)
第二章	WWW 浏览器介绍.....	(205)
第一节	Netscape Communicator.....	(205)
第二节	Internet Explorer.....	(207)
第三章	WWW 检索工具介绍.....	(209)
第一节	一般主题指南.....	(209)
第二节	一般检索引擎.....	(211)
第三节	中文检索引擎.....	(215)
第四节	医学检索引擎及主题指南.....	(218)
第四章	网上 Free Medline 检索.....	(224)
第一节	PubMed.....	(224)
第二节	HEALTHGATE.....	(231)
第三节	Evaluated Medline.....	(234)
第五章	Internet 上的生物医学电子期刊查找及利用.....	(237)
第一节	查找生物医学电子期刊的几种途径.....	(237)
第二节	如何利用网上提供的电子期刊全文.....	(238)
第三节	网上部分全文生物医学期刊网址.....	(238)

第五篇 医学文献联机检索

第一章	概论.....	(240)
第二章	国际联机情报检索系统的构成及主要功能.....	(240)
第三章	DIALOG 系统中若干重要医药卫生数据库简介.....	(242)
第四章	BRS 系统的主要医药卫生数据库.....	(247)
第五章	医学文献分析与检索系统 (MEDLINE) 简介.....	(248)
第六章	ELHILL 检索软件的使用.....	(254)
参考文献		(275)

第一篇 医学文献计算机检索概论

第一章 计算机信息检索的含义、原理及优点

计算机信息检索是本世纪50年代出现的一门新兴学科。它开辟了人类获取信息的新纪元,创立了信息检索的新篇章。

一、计算机信息检索的含义和原理

所谓计算机信息检索(下称机检)就是指人们根据特定的信息需求,利用计算机从相关的机读数据库中识别并获取所需的信息。

关于机检的原理,与手工检索的原理在本质上相同,但又有所不同。机检的基本原理是计算机将输入机检系统的用户提问标识(检索词)与已存贮在系统中数据库内的文献特征标识(标引词)进行机械性匹配比较,凡符合给定的比较原则和逻辑运算条件者即为命中信息。手工检索时,检索策略是由人脑记忆的,匹配比较是通过人对检索工具的手翻、眼看、大脑不停地思考和判断而完成的。这种匹配比较具有概念思维性、随机应变性,可随时修改检索策略。而机检是计算机按照人们给定的检索策略在机读数据库中进行高速、机械匹配比较而完成的。例如,人们让计算机用“AIDS”一词在有关的医学数据库中检索,其本意是想检索有关爱滋病方面的文献,但是,计算机会把含有“Hearing aids”(助听器)或者“Financial aids”(财政资助)的文献检出来。同样,人们如果让计算机用“Cancer”一词在数据库中进行自由词检索,希望获取有关癌症方面文献,那么,计算机在文献记录中出现的Carcinoma、Neoplasm、Tumor以及Cancers等词则一律视而不理。因为现有的计算机还不具备人的那种概念思维能力,也没有人脑那种随机应变能力。它是按照人们给定的字符串去进行机械的匹配比较。虽说人们已通过种种手段和技巧让计算机减少这种机械匹配的缺点,但在自由词检索、跨文档多数据库日益发展的今天,这种缺点还依然存在。

二、计算机信息检索的优点

计算机信息检索,为人类获取信息增添了无比的活力。尽管匹配比较的机械性还不能完全克服,但是与手工检索相比,机检却具有许多无可比拟的突出优点。归纳起来主要有以下几个方面。

1. 检索速度快、效益高:机检查找一个课题,少则3~5分钟,多则10分钟,即可完成手工检索几天甚至几个月的工作量。
2. 使用方便,实现了资源共享:用户不受地理位置的限制,可以通过设在家中、办公室或附近某一单位的联机终端,实现联机检索,查找几百里、几千里、几万里之外的主机系统所提供的数据库,从而实现地区性、全国性、乃至全球性的资源共享。
3. 检索内容新、范围广、数量大:机检可以使用户尽早获得最新的信息资料。许多机检系统在收到最新文献资料后,首先用于更新数据库,供用户使用,然后才利用机读磁带出版书本式的检索工具,而这两者之间往往存在一两个月的时差。

数据库可以编制得极大。用户一次可以检索收录几年、十几年、几十年的文献。例如,著名的医学文献数据库MEDLINE收录了自1966年以来全世界70多个国家出版的3900多种生物

医学期刊中的900多万条文献记录。另外,大型机检系统往往收录有几十个、乃至几百个各种类型的数据库。其专业范围十分广泛。用户可以根据需要,实行跨文档检索,即一个课题可一次检索许多个相关内容的数据库,这些都是手工检索无法比拟的。

4. 检索手段灵活,检索途径繁多:机检时,可以采用系统提供的布尔逻辑检索、位置逻辑检索、限定检索、截词检索、加权检索等手段,根据用户对查全率、查准率、文种、文献类型、时间跨度、经费支付能力等方面的要求,设计各种各样的检索式。

机检文献的标引深度要比手检的高得多。例如,一篇文献在美国“医学索引”(IM)中,一般标1~3个主题词。而在相应的机检数据库中,一般要标8~12个主题词,最多可达到25个主题词。再加上采用自由词标引,一篇文献在数据库中的检索点可达上百个。这样,大大增加了文献的检索途径、提高了文献的查全率。

5. 计算机不会像人那样因产生疲劳而致错:巨大的文献量若用手工查找,要花费大量时间和人力,而且还会因人眼和大脑的疲劳而错断。而计算机去完成此项工作,则不会有上述现象。

6. 以满意的方式为用户提供检索结果:计算机检索可将检索结果按用户要求的打印格式和排序方式加以整理,整理后的结果可以采用显示、联机打印或转存(软盘拷贝)等方式提供给用户。必要时,还可以为用户联机订购原文,这一点在逐步走向资源共享的高度信息化社会中也十分重要。

第二章 计算机信息检索的简史及现状

一、国外机检简史及现状

计算机信息检索由美国麻省理工学院Bagley等人于1961年率先开始研究。就计算机应用于信息检索而言,先后大体上经历了三个阶段。

1. 脱机检索阶段(1954~1964)

1954年,美国海军兵器中心使用IBM-701型计算机建立了世界上第一个计算机检索系统,开创了计算机信息检索的先河。1964年,美国国立医学图书馆(NLM)建成了当今世界上最具权威性的医学文献分析与检索系统MEDLARS(Medical Literature Analysis and Retrieval System)。用于编制印刷本的医学索引(IM)。同时开始了医学文献的脱机批式检索服务。

2. 联机检索阶段(1965~1970)

由于大容量计算机分时系统以及强功能检索软件的研制成功,机检的发展很快就进入了联机检索阶段。由中大型计算机做主机,并通过公共电话线路和许多终端机相连接,用户通过终端可以和主机实现联机对话,检索主机系统所拥有的数据库。1965年,美国麻省理工学院(MIT)以及系统发展公司(SDC)相继建成了他们的联机检索系统。随后,美国的DIALOG、欧洲空间研究组织(ESRO)的ESA/IRS等联机检索系统都相继问世。美国国立医学图书馆也于60年代末建成了世界上第一个医学文献联机分析与检索系统MEDLINE(MEDLARS on-line,能同时允许45个终端进行联机检索,并很快发展为世界上最有影响的联机检索系统之一。

3. 联机检索网络化阶段(1970~)

进入70年代以后,由于网状数据库管理系统的诞生,以及数据通讯网和分组交换通讯网等通讯技术的发展,加之高性能分时计算机、带终端的远程处理系统、廉价的大容量随机存取器等技术的发展,使得联机检索由原先的内部使用发展为面向社会公众的商业性服务。随着TELENET、TYMNET等数据通讯网的诞生,在美国首先形成了区域性、全国性的机检系统的联机网络。随之,欧洲、日本、加拿大等地也相继建成了全国性联机网络。在这些网络内,原先的每一个机检系统的主机成了网络上的一个个节点,每个节点都可以连接许多终端。各节点之间依靠通讯线路彼此相连,从而构成纵横交错、互相利用的机检网络。用户可以通过网络内的任何一台终端,检索网络内任何一台主机系统所拥有的数据库。信息检索在一个地区或一个国家内实现“资源共享”。

随着卫星通讯、公共数据通讯网等技术的进一步发展,计算机远程终端走出了国门,实现了跨国性的远距离联网检索,即国际联机检索。用户可以通过设在附近的国际联机终端,与远隔重洋、相距万里的主机系统直接进行联机对话,查找其所拥有的各种数据库,几分钟之内,即可找到用户所需的信息。这对用户来说,真可谓远在天边,近在眼前。

随着光缆通讯技术的发展,随着信息高速公路事业在全世界迅猛发展,计算机信息检索走向了全球大联网。美国的国际互联网络(INTERNET)的出现,使世界变成了一个地球村。世界各国的各类联机检索系统以及其它各种计算机信息系统纷纷加入互联网络。至1995年底,全球已有155个国家和地区的4万多个不同规模的计算机网络进入INTERNET。其上网户头已达到4000多万个。世界著名的大型国际联机检索系统均已加入INTERNET。世界各地人们可以利用就近的终端通过互联网络进入世界各大机检系统,检索其所需的数据库。除此之外,人们还可以利用互联网络传递电子邮件、进行文件传输、发布新闻消息、商讨疑难问题、进行信息咨询等。乡村医生利用终端,进入互联网络,与全世界医学专家对疑难病人进行会诊已成为现实。在美国等发达国家,INTERNET已成为人们社会生活中重要组成部分。

据资料表明,目前世界上已有联机检索系统600多个。其中,大型的联机系统计有100多个,可供联机检索的公用数据库已超过5000个,可检数据库的专业范围几乎无所不包。检索网络覆盖范围几乎遍及世界各国。其中,特别著名的国际联机检索系统有美国的DIALOG系统、ORBIT系统、BRS系统以及MEDLARS系统,还有欧洲的ESA/IRS系统、英国的BLAISE系统、德美日三国联合经营的STN系统,以及日本的JOIS系统等。国际互联网络的产生与发展,使得国际联机检索更是如虎添翼。今后,随着计算机技术的进一步智能化、数据库载体的进一步高密度化和多媒体化、通讯技术的进一步网络化,计算机信息检索将走向办公室化、家庭化。

另外,进入80年代中期以后,国际上还出现了一种光盘——微机检索系统。这种系统除了不需要大型计算机以外,更主要的是它使用了一种称之为CD-ROM光盘(Compact Disk-Read Only Memory)的新型存储器。这是一种高密度、大容量的存储介质。一张12cm光盘的存储容量为550~600兆字节,相当于1600多张普通软磁盘的总存储量。人们只要有一台配有光盘驱动器的微机,加上光盘数据库,便可实现光盘信息检索。目前,光盘检索也可在局域范围内实现多工作站的网络化检索。一个局域光盘网络可支持255个用户和多至240个光盘。网络内用户可在任意时间查访塔式驱动器内多达21个光盘。光盘检索系统因其具有设备简单、可随地安装、使用方便、检索费用低等优点,已被世界各地广泛使用。据1992年统

计,全世界共发行光盘数据库已达2200多种。多数医学数据库都有了相应的光盘,累计有200多种。这种分散的小型光盘—微机检索系统加速了人类信息化社会的进程。

二、国内机检简史及现状

我国计算机信息检索起步较晚,于1975年才开始这项工作。主要可以概括为以下五个方面。

1. 从国外引进数据库开展机检服务

1975年,我国开始从国外引进磁带式数据库,开展机检服务。先后有几十家信息单位和高等院校引进了60多种磁带式数据库。其中有近20种与生物医学有关。例如,中国医科院信息所引进的MEDLINE磁带、化工部信息所引进的CA磁带、南京大学引进的BA磁带、北京文献服务处引进的GRA磁带以及WPI磁带等。这些单位利用引进的磁带数据库建立自己的机检系统,开展机检服务。有些现已发展为跨地区的联机检索系统。中国医科院信息所通过引进MEDLINE磁带,于1987年发展成为美国MEDLARS在中国的分中心。从而逐步建成了中国的MEDLARS,面向全国开展机检服务。

2. 设立国际联机终端开展机检服务

1980年中国建筑科学院情报所等15个单位在香港联合租用一台国际联机终端,开展与DIALOG、ORBIT等系统的联机检索。检索结果由香港邮寄至北京。自那以后,先后在北京、南京、上海、西安、长沙、广州等50多个城市设立了130多台国际联机检索终端。仅北京一地就有20多台。这些终端开通了中国与DIALOG、ORBIT、ESA/IRS、STN等12个世界上著名的大型国际联机检索系统的检索,极大地促进了中国机检事业的发展。

3. 自建数据库

我国从80年代初就开始建设自己的中文数据库。1983年,《中国药学文摘》实现了计算机化生产,并同时开始了中文医学文献机检的试验。80年代中后期,自建数据库事业有了迅猛的发展,截止1991年底,全国登记注册的各类自建数据库就有806个,其中大部分为中文数据库。近几年发展更快,数据库质量也明显提高,数据库规模日益扩大。例如,中国医科院信息所研制的《中国生物医学文献数据库》、中国医药管理局研制的《中国药学文摘数据库》、中国中医研究院情报所研制的《中医药文献数据库》、中国专利局专利信息中心研制的《中国专利数据库》等等,都已投入使用。

4. 发展光盘—微机检索

鉴于大型计算机价格昂贵、国际联机检索费用过高,光盘检索深受国内广大用户欢迎。1986年4月,中国海洋信息研究所首次引进水产科学和渔业文摘光盘(ASFA)。此后,全国各大科研机构 and 高等院校的信息部门纷纷从国外引进相关学科的光盘数据库品种达上百种之多,其中,尤以MEDLINE光盘引进的单位最多。医学方面的EMBASE系列光盘、BA光盘、SCI光盘、CC/Life Science光盘等,均有引进。

此外国内也开始研制自己的光盘数据库。我国第一张光盘数据库《中文科技期刊篇名库光盘》于1992年问世。由中国医科院信息所研制的《中国生物医学文献光盘数据库CBMDisc》(1981~)已在医学界得到广泛地应用,并已走出国门。

5. 加入Internet

1994年4月,中国通过登记注册,正式与Internet联网。以国家教委、中国科学院、邮电部等系统发展和Internet联网。医学系统以中国医学科学院信息所为中心和六大地区医学院校联网,发展中国医学信息网络(CMINET)。

第三章 计算机信息检索系统的构成

由计算机信息检索系统，我们可以从它的逻辑构成和物理构成两个方面来认识它。所谓逻辑构成主要是从系统功能上看，它通常包括以下几个子系统：1. 信息源选择与采集子系统；2. 标引子系统；3. 建库子系统；4. 词表管理子系统；5. 用户接口子系统；6. 提问处理子系统。其中后三者与用户关系密切。从物理构成角度看至少包括三个基本部分，计算机硬件、软件及数据库三个部分。对于一个完整的联机检索系统，还必须包括通讯线路以及检索终端两个部分。

下面分别对这五个部分予以简介。

一、硬件

计算机硬件通常指计算机主机及其外围设备。主要包括以下五个部分。

1. 输入装置：主要指键盘、鼠标器、扫描识别仪、光笔等及用于输入程序和数据等的设备。
2. 运算器：在主机内，用于进行算术和逻辑运算以及其他有关的运算操作。
3. 控制器：在主机内，用于控制、协调计算机各部件的工作。例如何时输入数据、怎样进行运算、何时输出结果、中间结果存放何处等等，都由控制器指挥。
4. 存贮器：包括内存贮器和外存贮器，用于存放原始数据和各种操作程序以及运算过程中得出的中间结果和最终结果。
5. 输出装置：主要指显示器和打印机等，用于显示和打印运算结果和其它信息。

二、软件

计算机软件是计算机系统使用的各种程序的总称。按功能划分主要有以下几种：

1. 系统操作程序：用于对计算机系统进行有效的整体控制和管理的程序，以确保计算机系统正确、高效地运行。
2. 数据库管理程序：是对数据库整体进行控制和管理的程序，是信息检索系统软件的最核心部分。它包括对数据库的编制、检索、更新和排它控制等。
3. 联机控制程序：用于对机检系统的联机系统部分进行有效地控制和管理的程序，包括控制和管理检索终端、通讯线路和信息排队等。
4. 应用程序：指为完成用户不同需求而设计的各种实用程序如输入输出程序、词表管理程序、定题服务（SDI）程序、回溯检索程序、记帐程序等。

三、通讯线路

主要指电话通讯网、数据通讯网、卫星通讯网等。用于建立网内各主机之间、特别是检索终端和主机之间的通讯联系，以确保联机检索过程中能高速、有效地传递信息。

四、检索终端

检索终端是信息用户与检索系统主机进行人机对话，实现联机检索的设备。一个有效的联机检索终端还必须附有上网设备、打印机和调制解调器。上网设备负责与通讯网联系。如果是电话通讯网，那么，一台电话机便是合适的上网设备。打印机配合终端机负责信息的发送和接收。调制解调器负责把终端的数据信号调制成适合于通讯网上传输的模式。

信号,同时,把来自通讯网上经过调制的模拟信号解调并恢复成调制以前的数字信号。从而实现信息的有效传递。

五、数据库

数据库是计算机信息检索系统的核心。没有数据库,信息检索就成了无米之炊。根据国际标准化组织的文献与信息术语标准(ISO/DIS 5127号标准)的定义,数据库是指至少由一个文档(File)构成,并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。根据美国著名信息检索专家Williams ME所下的定义,数据库是指包含书目及与文献有关数据的机读记录(Record)的有序集合。我们还可以形象地说,书本式检索工具的机读版本就是数据库。

数据库可以由一至多个文档构成。鉴于一个文档可以构成数据库,所以,人们常常把数据库又称作文档。严格地讲,文档是数据库的下位概念。按编排结构和功能的不同,数据库又可分为顺排文档、倒排文档等。在书目型数据库中,所谓顺排文档,是以描述每篇文献内外特征的标引内容集合,即以文献记录作为信息存贮单元,按文献记录入藏的存取号从小到大顺序排列而形成的目录式文档。也叫主文档。它相当于印刷型检索工具的正文部分。所谓倒排文档,就是以从顺排文档记录中抽取出来的每一个文献特征标识作为信息存贮单元,并按某种顺序排列,同时,在每一标识后面注明相应的文献记录存取号,这样便构成了倒排文档。不同类型的特征标识可以形成不同的倒排文档。例如,著者倒排文档、主题词倒排文档等。它们相当于印刷型检索工具的作者索引、主题索引等。总之,倒排文档是从顺排文档派生出来的一种文档。因此,人们又称之为辅助文档。倒排文档的作用是供计算机直接检索使用。对于检索结果的显示打印、存盘等,则是通过顺排文档进行的。

文档是大量机读记录的有序集合。记录是数据库中一个完整的信息单元,通常是由一系列字段(Field)构成。在书目数据库中,所谓字段即指描述文献内外特征的各项标识内容。如文献记录的存取号、题名、作者、文献出处、主题词、文摘、语种、文献类型等等。每一项内容构成一个字段。某些字段还可分成若干个子字段。这些字段依其在检索中的作用和地位,又可分为基本索引字段和辅助索引字段两大类。前者揭示了文献主题内容特征,后者则揭示了文献的外表特征。各种数据库的记录格式不尽一样,但就其实质而言,多数是大同小异。详见以下各节。

第四章 数据库类型及医学数据库

一、数据库类型

作为信息用户,必须了解数据库的各种类型,以便根据不同信息需求选择合适的数据库。对于数据库类型的划分,有多种标准。按数据库所含信息的内容划分,主要有以下几种。

1. 书目型数据库(Bibliographic Database)

书目型数据库,又称为参考型数据库(Reference Database),主要是指二次文献数据库。它包括各种机读版的文摘、题录、目录等形式的数据库。它给用户提供一些简单而基本的信息以及原始文献的线索,指引用户根据文献线索去找原始文献,获取充分的信息。目前,在信息检索中,书目型数据库使用最广泛,地位最重要。

2. 事实型数据库 (Fact Database)

事实型数据库中存贮的数据一般是用来描述人物、机构、事物等非文献信息源的情况、过程、现象、特性等方面的事实性信息。诸如名人录、机构指南、产品目录、科研成果目录、研究或开发项目目录以及大事记之类,均可编成事实型数据库。名人录类数据库,有综合性,有专业性的,收录各种人物传记的信息,包括个人特长、生平事迹、主要业绩等。如果是当代在世的名人录,还可能提供工作单位、通讯地址等信息。公司和机构数据库,收录公司、机构的业务活动信息,包括人员、组织、资产、服务项目等情况。有的最终还提供单位地址、电话号码等信息。在医学方面,这类数据库也不少。例如美国MEDLARS系统的医生咨询数据库(PDQ数据库),它为医生们提供了有关癌症治疗和临床实验最新研究进展的内容。包括肿瘤的类型和预后、各种治疗方案以及从事肿瘤治疗方案研究的医生和保健机构的有关信息。

对于这类数据库,用户只要通过要查询的人物、机构或事物的名称及有关事项进行检索,就可能查到所需信息及其进一步追踪的线索。事实型数据库中,有些对用户还起着一种指示、介绍、牵线、搭桥的作用。它指引用户根据地址或电话等寻找合适的信息源,获取更详实的信息。因此,人们把这类数据库往往又称作指南型数据库或指示型数据库(Referral Database)。

3. 数值型数据库 (Numeric Database)

数值型数据库主要为用户提供数值性数据类信息。包括各种统计数据、科学实验数据、科学测量数据等等。例如,医学上使用的化学制剂、药物的各种理化参数、人体生理上的各种数值,均可入数值型数据库。有的数据库,除了提供数值性数据外,还包含有文本数据,主要是有关数值的定义和数值项的说明、解释之类等必不可少的文字。有人把这类数据库另称之为“文本-数值型数据库”(Textual Numeric Database),在此,一并归入数值型数据库。

数值型数据库为用户提供一些能够直接使用的信息,无须再去追查原始文献,大大节省了时间。这是与书目型数据库明显不同之处。这些数值信息是由专门人员从各方面的文献中收集、提取而来,经过仔细加工、积累并存贮起来,形成所谓的“浓缩信息”。

数值型数据库已成为世界各国信息检索工作开发的热点,而且,规模越来越大。例如,美国国立医学图书馆编制的化学物质毒性数据库RTECS (Registry of Toxic Effects of Chemical Substances),包含有10万多种物质的急、慢性毒力试验数据。

4. 全文型数据库 (Fulltext Database)

全文数据库是一种存贮文献全文或节选其中主要部分而形成的原始文献数据库。有的有对应的印刷型文本,有的则是纯电子出版物。最早出现的是法律条文方面全文数据库。随后几乎各个领域都出现了全文数据库。据资料表明,目前,全文数据库约占整个数据库的1/4~1/3。医学方面的全文数据库也越来越多。目前在DIALOG中就有48种医药期刊全文数据库。由美国医学药物协会编制的药物信息全文数据库DIF (Drug Information Fulltext)就是其中之一。它在DIALOG系统中的文档号为229。

上述的数值型数据库、全文型数据库又被人们称之为源数据库(Source Database)。除此之外,还有图像型数据库、术语型数据库也属于源数据库。这些数据库均为用户提供直接使用的信息,不再需要追踪原始文献,这是源数据库与参考数据库的根本区别。

二、医学数据库