医学信息检索

实用教程

主 审·王娟萍
主 编·何 怡 刘 毅

天津科学技术出版社

医学信息检索实用教程

主　审　王娟萍
主　编　何　怡　刘　毅
副主编　常　红　王淑琴　程　鸿

天津科学技术出版社

图书在版编目(CIP)数据

医学信息检索实用教程 / 何怡，刘毅主编. —天津：天津科学技术出版社，2009. 8
ISBN 978-7-5308-5264-4

Ⅰ. 医… Ⅱ. ①何… ②刘… Ⅲ. 医药学—情报检索—教材
Ⅳ. G252.7

中国版本图书馆 CIP 数据核字(2009)第 139011 号

责任编辑：宋庆伟
责任印刷：王 莹

天津科学技术出版社出版
出版人：胡振泰
天津市西康路 35 号 邮编 300051
电话：(022)23332379(编辑室) 23332393(发行部)
网址：www.tjkjcbs.com.cn
新华书店经销
三河市腾飞印务有限公司印刷

开本 787×1092 1/16 印张 13 字数 180 000
2009 年 8 月第 1 版第 1 次印刷
定价：25.90 元

编写者名单

主　审　王娟萍

主　编　何　怡　刘　毅

副主编　常　红　王淑琴　程　鸿

编写者　（以汉语拼音为序）

常　红（天津医科大学）

何　怡（天津医科大学）

黎小沛（天津医科大学）

思金华（天津中医药大学）

唐万斌（天津医科大学）

王　蕾（天津中医药大学）

王娟萍（南开大学）

王淑琴（武警医学院）

席　薇（天津医科大学）

闫　蓓（武警医学院）

张惠荣（天津医科大学附属肿瘤医院）

郑　凯（天津中医药大学）

前言

文献检索课是一门以图书馆学、情报学为理论基础，以计算机和网络为手段，侧重培养学生信息意识以及获取和利用文献信息能力的科学方法课。在培养学生的创新意识、自学能力和独立研究问题能力等方面具有非常重要的作用。在当今信息社会，让医学生具备获得最新医学信息的能力已成为高等医学教育中不可忽视的重要一环。为了适应网络时代知识经济的发展以及专业人员对于医学信息利用的需要，培养能力型高素质的医学人才，我们编写了这本《医学信息检索实用教程》。

本书在原有《网络生物医学信息检索与利用》(第二版)的基础上作了一些调整与更新。为了适应部分数据库内容及功能的变化，重新编写了CBM、万方、超星数字图书馆、OVID、EBSCO、SCI、Elsevier SDOL等数据库，增加了CNKI、CMCC、CMCI、BIOSIS Previews、EMBase、检索策略的制定与检索报告的书写等内容。

新版教材共八章。第一章计算机检索与数据库，着重介绍计算机和数据库检索的基本知识。基于长年文检课教学经验，重点介绍两种常用的检索语言：分类语言—《中国图书馆分类法》以及主题语言—颇具权威性的医学主题词表—美国《医学主题词表》(MeSH)。结合本书所介绍的数据库，总结归纳出若干项常用的检索技术，便于使用者系统全面地了解。第二章至第四章分别为文摘数据库、全文数据库和引文数据库的介绍。所述数据库几乎涉及国内外现有的与生物医学有关的重要数据库。第五章为国内外医学专业搜索引擎的介绍。第六章介绍检索策略的制定与检索报告的书写方法。第七章为医学论文与综述的撰写介绍，概述了医学论文及综述的特点、类型，详细介绍了撰写医学论文和综述的基本步骤及方法。本章内容主要为研究生、高年级本科生撰写医学论文及综述提供参考。第八章为医学信息的查新咨询，并附详细的科技查新报告样例。本章内容为配合目前医学临床与科研日益增多的信息查新活动而特设的。

本书内容新颖，简明扼要，通俗易懂，注重实践技能的培养，实用性强。既可作为高等医学院校研究生、七年制、本科生的教材，也可作为医药卫生工作者、医学教学与科研人员以及医学信息工作者的参考书。

本书编写是在天津医科大学、南开大学、天津中医药大学、武警医学院和内蒙古医学院各位老师的通力合作下顺利完成的。特邀南开大学图书馆科技查新专家王娟萍研究馆员担当本书主审。王老师还亲自撰写了第二章第六节化学文摘(CA)光盘数据库、第八章医学信息的查新咨询两部分内容。

在本书编写过程中，编写者参考了大量的中外有关文献资料，对所引用的文献作者特在此表示衷心的感谢。由于编写者水平有限，编写时间紧迫，加之网络信息资源变化更新较快，书中错误疏漏在所难免，恳请使用者批评指正。

编　者

2009年6月

目录

第一章
计算机检索与数据库

一 网络生物医学信息资源的类型与特点

(一)网络生物医学信息资源的类型

网上生物医学信息资源极为丰富,数量巨大,形式多样,彼此间互相交叉,要对其进行准确的分类是比较困难的。一般情况下,可将网上生物医学信息资源分为以下几类。

1.电子出版物资源 主要包括电子期刊、电子图书和电子报纸。

(1)电子期刊 如CNKI的"中国期刊全文数据库"库,提供9056种数字化期刊的全文文献(截至2009年6月8日)。万方数字资源系统的"中国数字化期刊群",收录5500种各学科领域核心期刊。SpringerLink数据库提供2099种全文电子期刊。一些著名期刊如Science、BMJ等也在网上提供免费全文。

(2)电子图书 如《超星数字图书馆》提供4万余种生物医学图书,SpringerLink数据库提供图书、科技丛书(Book Series)、参考工具书33477余种(截至2009年6月8日)。电子图书一般可供用户网上在线阅读或根据需要整本或按章节下载。

(3)电子报纸 如CNKI的"中国重要报纸全文数据库"。国内如健康报(http://www.jkb.com.cn)、中国医学论坛报 (http://www.cmt.com.cn)。国外如 Science Daily(http://www.sciencedaily.com)等。多为各大报纸的网络版,一般提供免费阅览。

2.生物医学数据库资源 主要包括文献型数据库、数值或事实型数据库、多媒体数据库。

(1)文献型数据库 包括题录文摘数据库和全文型数据库。题录文摘数据库如《中国生物医学文献数据库》(CBM)、PubMed等;全文型数据库如《中文科技期刊数据库》(重庆维普)、EBSCO、OVID等。

(2)数值或事实型数据库 主要包括基因库、核酸序列(GenBank)、蛋白质结构库等分子生物学数据库以及毒理学、药物方面的事实型数据库。

(3)多媒体数据库 如各种医学图谱库、医学影像库、病理切片库等。

3.医学新闻资源 主要包括生物医学电子公告(BBS)和网络新闻组(newsgroup,group,forum)。BBS是网上使用较为广泛的交流方式。通过其论坛,可就某些问题展开讨论,是普及医学知识的理想途径。新闻组又称为讨论组,是网上获取信息、进行交流的重要工具。医学专业人员或对某个领域的主题感兴趣的患者,通过新闻组,互发信件或传递信息,就共同关心的问题进行广泛交流。

4.医药市场信息资源 如"医药信息指南"(万方),是专业化的生物医药网络信息资源导航系统,由从事医药信息检索工作多年的专家在对网络生物医学、药学信息资源进行深入研究、评价、筛选的基础上编辑而成,是医药行业工作人员的必备资源。

5.医学教育资源 主要包括医学继续教育与培训资源和病人教育资源。医学继续教育与培训资源国内主要有中华医学视听网(http://www.cma-cmc.com.cn)以及各医科大学网站中

的医学教育栏目。对于病人教育资源的获取，首选学术机构网上资源，如国外著名的医学学会和协会的网站。一些网络电子期刊(如 JAMA)的网站也提供丰富可靠的病人教育资源。

6.生物医学类公共软件资源　主要指实验数据分析、统计或基因同源性比较等免费公用软件，对于从事生物医学基础研究、流行病学研究和科研管理人员都是非常有价值的。

7.其他医学信息资源　网络医学信息几乎覆盖了医学科研、临床、教学等各个方面，其他医学信息资源主要包括：学位论文(CNKI《中国优秀硕士学位论文全文数据库》《中国博士学位论文全文数据库》)、专利文献(万方《专利技术数据库》)、标准文献(万方《中外标准文献数据库》)、医学会议信息(万方《中国医学学术会议论文文摘数据库》)等特种文献资源。此外，还包括医院(万方《中国医院名录数据库》)、医学院校(万方《中国医药院校名录数据库》)和医生信息(万方《中国医药名人数据库》)等。

(二)网络生物医学信息资源的特点

1.数量巨大、来源丰富　Internet 是一个开放的信息资源存储与传播的主要媒介之一，它集各领域的信息资源于一体，供网上用户自由共享。同时，任何机构、任何个人都可以将自己拥有且愿意让他人共享的信息在网上发布，这些导致了网上信息资源数量巨大，且增长迅速。生物医学信息资源不仅有专门的生命科学相关机构提供，还有大量的非生物医学机构特别是商业公司提供，其信息来源十分丰富。

2.内容庞杂、质量不一　网上信息资源包罗万象，几乎覆盖了人类知识的各个领域。既有社会科学、自然科学等学术信息资源，也有与大众日常生活息息相关的诸如体育、娱乐、旅游、消遣等信息资源。由于网络所具有的随意性和自由度，造成信息内容庞杂，质量参差不齐。作为网上资源重要组成部分的生物医学信息资源也不例外。

3.类型齐全、形式多样　网络生物信息资源可分为文本、声音、图形、图像、动画、电影、音乐等多种信息类型。类型齐全，形式多样，不拘一格。

4.更新快、时效性强　由于计算机网络的高速传输特性，使网络上的信息资源瞬息万变。故网上医学信息的更新周期短、内容新颖、时效性强。

二　计算机检索的基本知识

(一)计算机信息检索的概念及原理

1.计算机信息检索的概念　计算机信息检索包括两个过程：一是信息存储，即把大量分散无序的信息集中起来，经过加工，使之有序化、系统化，并按一定的格式存储起来建成数据库；二是信息检索，即按照信息用户的需求，利用计算机在已建成的数据库中进行信息查寻。信息存储和信息检索的两个过程是相辅相成、密不可分的，所以信息检索的全称又叫“信息存储和检索”。

信息存储即数据库的建立。存储文献信息前首先要对信息进行标引，形成文献的特征标识，并将特征标识输入计算机进行排序，形成多种索引，组织成具有检索价值的数据库系统。文献特征标识一般分为文献的外表特征和内容特征。所谓外表特征是指文献的题名、作者、出版物名称、卷期、页码、出版时间、语种、文献类型等项目。所谓内容特征是指文献所论述的学科属性、学科分类、主题内容等。标引文献前，首先要对文献内容进行主题分析，把握所论述的中心内容，形成主题概念，然后选用特定的检索语言(主题词表、分类表)来表达主题概念，最后将这种标识按其内容和出处进行编排，并输入检索系统。

信息检索即数据库的利用，是信息存储的逆过程。信息检索是根据信息需求，确定能代表信息需求的若干检索词，并将检索词转换成特征标识，依据检索技术构建检索式，利用计算机在数据库中查找文献线索，最后找到原始文献。信息存储是信息检索的基础，信息检索是信息存储的目的以及信息利用的手段。

2.计算机信息检索的原理　简单地说，就是指通过一定的方法和手段，使信息存储与检索两个过程所采用的特征标识达到一致，以便有效地获取和利用文献。

3.信息检索的目的和意义

(1)继承和借鉴前人成果，避免重复研究　整个科学发展表明，积累、继承和借鉴前人或他人的研究成果是科学发展的重要前提。因此，科研人员在开始着手研究一项课题前，必须通过信息检索来了解这个课题是如何提出来的，前人或他人在此方面已经做了哪些工作，是如何做的，有何成果和经验、教训，还存在什么问题等。只有这样，才能正确地制订研究方案，避免重复研究，提高研究起点，降低研究过程中获取信息和知识的成本。相反，如果继承和借鉴工作做得不好，就容易造成低水平的重复研究。

(2)节省科研工作时间，提高科研效率　医学文献信息的特点表现为：①增长迅速、数量庞大；②内容交叉渗透、分散重复；③语种不断增加；④更新周期缩短、传播速度加快；⑤日益向多元化(电子化、网络化、数字化)发展。这些在一定程度上增加了科研人员查找文献信息的难度。信息检索是医学科研的前期工作，是科研工作的重要组成部分，其检索方法和技能是科技人员应当掌握的基本功之一。科研人员如果掌握了信息检索的方法，能熟练地查找自己所需的文献信息，无疑将大大地缩短查找文献信息的时间，同时，提高科研工作效率，缩短科研周期，达到多出成果、早出成果的目的。

(二)检索语言

检索语言是一种在文献存储和检索过程中共同使用，用于描述检索系统中文献的内容特征及外表特征、表达用户检索提问的专门语言。其作用是描述文献特征，表达信息提问，并使两者达到相互沟通。检索语言可分为两种类型。

1.按检索语言所使用语词的受控情况划分

(1)规范化检索语言(或称受控语言)　是指对文献检索用语的概念加以人工控制和规范，把检索语言中各种同义词、多义词、近义词等进行规范化处理，使每个检索词只能表达一个概念。如：主题词 neoplasms(MeSH)。使用规范化的语言能够相对提高检索的效率，但在选词上对检索者和医学信息存储人员要求比较严格。

(2)非规范化检索语言(或称自然语言)　它对检索用语中的各种同义词、多义词、近义词等不加处理，所以又称自然语言。如表达“肿瘤”的英文词有 cancer、neoplasm、oncoma、tumor、tumour 等，均可以作为检索词进行检索。

2.按检索语言所描述的文献内容特征划分

(1)分类检索语言　用分类号作为语言来表达概念，将各种概念按学科性质进行分类和系统排列，如《中国图书馆分类法》(简称《中图法》)。中文数据库分类检索语言一般采用《中图法》，如《中文科技期刊数据库》(重庆维普)、CBM 等。

《中图法》(书目文献出版社出版，第四版，1999 年)是一部等级体系分类语言，按照从总到分、从一般到具体的编排原则，逐级地进行概念的划分和概括，产生许多不同级别、层层隶属的概念，构成一个秩序井然、层层展开的概念等级体系。《中图法》将知识门类分为五大部

类:①马克思主义、列宁主义、毛泽东思想、邓小平理论;②哲学;③社会科学;④自然科学;⑤综合性图书。在基本部类的基础上,组成22个基本大类,并用22个字母分别作为各大类的代码。各级类目的分类号由字母和数字组合而成,每一级类目又分为若干个下位类目。“医药卫生”使用“R”作为类目代码,其等级结构举例如图1-1。

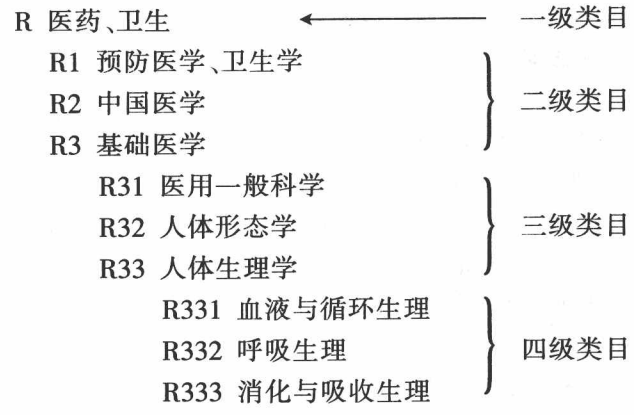

图1-1 《中图法》类目表等级关系图

(2)主题检索语言　它是用词语来表达各种概念。主题检索语言中目前使用较多的是主题词和关键词两种检索语言。

①关键词(keyword)是为适应计算机自动编制索引的需要而产生的。所谓关键词是指出现在文献的题名、文摘或全文中的,能表达文献实质内容的或者能被人们作为检索入口的关键性专业名词术语。关键词属于非规范化的检索词,词与词之间没有语法关系,适用于计算机检索。例如:一篇文献的题名为“高血压药物治疗的现状与进展”,根据该文题名、文摘及全文所论述的内容,选取“高血压、药物治疗、β受体阻滞剂、钙拮抗剂、血管紧张素转换酶抑制剂、抗高血压药物、利尿剂”等词作为关键词;而“现状”“进展”因只起到辅助描述作用,故不作为关键词(选自《中文科技期刊数据库》)。采用关键词检索,其优点是能够及时反映文献的最新信息,用法简便,一般用户不经系统培训也能容易掌握利用。但同一主题概念的文献标引相对分散,容易导致漏检和误检。如果使用不当,会影响文献的查全率和查准率。

②主题词(subject, subject headings)又称叙词(discriptor),系来源于系统词表并经过规范化处理的检索词。主题词既能表达文献的内容特征,词间又有严密的语义关系,使用主题词检索文献可保证较高的查全率和查准率,但对于检索者来说要求较高。

主题词的主要特点:一是经过严格规范化处理的词或词组,保证语词与概念的一一对应;二是可用于进行概念组配检索。检索时,检索者只需根据需要,临时从词表中选出相应的主题词,并按组配规则,任意扩大或缩小检索范围,体现了主题检索语言的灵活性。

主题词来源于主题词表。医学主题词表是提供检索者对自然医学词汇进行规范处理,并使用规范化的主题词检索文献的辅助工具。随着科学的发展及文献中用词的变化,主题词表一般会不断增删、修订,定期更新。

下面简要介绍颇具代表性的医学主题词表——MeSH。

3.《医学主题词表》(MeSH)　《医学主题词表》(Medical Subject Headings,简称MeSH)由

美国国立医学图书馆(National Library of Medicine,简称 NLM)编辑出版,它是一部具有权威性的、可扩充的动态性医学主题词表。有些中文生物医学数据库(如 CBM)也采用 MeSH 作为标引文献的主题词表。

(1)MeSH 的构造　MeSH 每年出版一册,主要包括主题词变更表、字顺表、树状结构表和副主题词表等四个部分。

①主题词变更表:当年主题词的增删更改情况(Deleted Headings、New Headings)。

②字顺表(Alphabetic List):全部主题词(包括少量款目词)按英文字母顺序排列,内嵌参照系统(用代参照、属分参照和相关参照)描述词间关系。字顺表中的医学词汇之间存在着诸如同义、交叉、隶属等关系。针对这些语义关系,MeSH 在字顺表中建立了一套完整的参照系统,以指引检索者正确地选择主题词进行检索。参照系统共有三组,分别表示三种不同的词间语义关系。

a.用代参照:揭示词间等同关系,其参照符号为 see(见)和 X(代)。例如 Kidney Circulation see Renal Circulation 及 Renal Circulation X Kidney Circulation。

b.属分参照:揭示词间上下等级关系,参照符号为 see under(属)和 XU(分)。例如 Hepatitis, Viral Non-A Non-B see under Hepatitis, Viral, Human 及 Hepatitis, Viral, Human XU Hepatitis, Viral Non-A Non-B。1991 年后,所有次要主题词均升级为主要主题词,该参照被取消。

c.相关参照:揭示部分主题词之间意义上的相关关系,参照符号为 see related(参见)和 XR(反参)。例如 Acquired Immunodeficiency Syndrome see related Lymphoma, AIDS-Related 及 Lymphoma, AIDS-Related XR Acquired Immunodeficiency Syndrome,表示两个词间既非同义关系,也无属分关系,仅在概念上有一定的相关性。

MeSH 的主题词可以是单词或词组。单词如 asthma、hypertension,词组如 lung diseases、myocardial infarction。词组又有正置和倒置两种形式。正置形式即按词组的自然语序组成的词,如: myocardial infarction、kidney failure。倒置形式如:

Hepatitis (肝炎)
Hepatitis A (甲型肝炎)
Hepatitis A Virus (甲型肝炎病毒)
Hepatitis, Alcoholic (肝炎,嗜酒型)
Hepatitis, Animal (肝炎,动物型)
Hepatitis Antibodies (肝炎抗体)
Hepatitis B (乙型肝炎)

倒置形式的主题词在 MeSH 中所占比例较大,其目的是使同属某一大概念的文献能按某一主题词字顺相对集中地排列在一起,便于族性词类检索。若检索者在使用所选主题词查找文献未果的情况下,应考虑主题词的倒置形式。

③树状结构表(Tree Structures):将主题词按照学科分类集中,反映主题词之间概念上的逻辑隶属关系。

字顺表反映词间的横向关系,树状结构表则显示词间的纵向隶属关系,两者以树状结构

号互相沟通。将两者配合使用，可帮助检索者进行专指性检索和扩展检索。

④副主题词表(Subheadings)：副主题词是对主题词的进一步限制，使主题词的专指性更强。MeSH 共有 80 余个副主题词。

主题词一般表达某个确切的概念，如某种疾病、药物、器官等，而副主题词一般是对某一类事物的某一方面的概述，如对某种疾病的诊断、治疗，某种药物的治疗应用、副作用等。副主题词的作用主要是限定主题词的范围，使主题词具有更强的专指性，缩小检索范围，加快检索速度。因为检索者检索时一般只需有关主题的某一个或某几个方面的文献信息，如果没有副主题词对其进行限定，检索出来的文献的准确性就不会很高。

副主题词不能单独使用，必须与主题词组配在一起使用。但并非每个副主题词都能与任何主题词组配使用，主题词与副主题词之间必须遵循一定的逻辑关系。副主题词表对每一个副主题词的组配范围作了说明。以下是各种副主题词“治疗”的组配范围限定，从中可以看出主题词与副主题词的组配规律和使用方法。

Diet Therapy(DH)饮食疗法。与疾病主题词组配，表明对疾病所作的伙食和营养安排。但不包括维生素或矿物质的补充，对后者可用副主题词“药物疗法(Drug Therapy)”。

Drug Therapy(DT)药物疗法。与疾病主题词组配，表明通过投给药品、化学物质和抗生素治疗疾病。但不包括免疫治疗和用生物制品治疗，对后者用副主题词“治疗(Therapy)”。对于饮食疗法和放射疗法，分别用专指的副主题词。

Radiotherapy(RT)放射疗法。与疾病主题词组配，表明电离和非电离放射的治疗应用，包括放射性核素疗法。

Surgery(SU)外科手术。与器官、部位、组织和疾病的主题词组配，表明以手术治疗疾病，包括用激光切除组织。但不包括移植术，对后者用副主题词“移植(Transplantation)”。

Therapy(TH)治疗。与疾病主题词组配，用于除药物疗法、饮食疗法、放射疗法和外科手术以外的治疗手段，包括综合治疗。

(2)选择主题词时应注意的问题　在检索文献时，主题词的选择应遵循以下几点。

①首选专指词。如一篇关于“影响 1 岁以下婴幼儿行为的因素”的文献，将被标引在专指词“Infant Behavior(婴儿行为)”，而主题词“Child Behavior(儿童行为)”下将无此文献。

a.注意同义词、近义词的转换，如肾循环：Kidney Circulation see Renal Circulation。

b.注意主题词的倒置形式，如检索“肾性高血压”的文献，用检索词“Renal Hypertension”在 MeSH 中是找不到的，它的主题词为“Hypertension, Renal”。

c.注意主题词的增删、变更情况，如查找“由专业人员来做的家庭护理”方面的文献，2001 年以前的文献需用主题词“Family Health”或“Nursing Care”；而 2001 年新增主题词“Family Nursing”，则不能再用前者查找。此处还需注意与主题词“Home Nursing”的区别，后者虽从 1966 年以来就是主题词，但其表示的是“非专业人员的家庭护理”，故不能选用。

②次选主副组配词。如检索有关“公共卫生标准”方面的文献，可直接选用“Public Health/standards(公共卫生/标准)”即可。

在实际检索过程中，一个课题往往包含多个概念，这时需要分析它们之间的内在关系，选

择最佳组配形式。如要查找“用生姜食盐乙醇防止常见食品的污染”方面的文献，应选择主副组配“Food Contamination/prevention & control(食品污染/预防和控制)”。在组配时应注意采用概念组配，而不是字面组配。

③再选上位词。当某些文献主题专指性较强或者是一些新的名词术语、新药、新物质、新发现的疾病、新的实验方法等，在 MeSH 中既无专指词，又无法组配检索时，可根据其学科属性选择一个与之最邻近的上位概念的主题词来进行检索。如检索非甲非乙型肝炎(Hepatitis, Viral Non-A Non-B)方面的文献，可以用其上位概念“人类病毒性肝炎(Hepatitis, Viral, Human)”进行检索。又如检索有关“污水处理”方面的文献，只能用其上位概念“废弃物处理，液体(Waste Disposal, Fluid)”，而不能用“废弃物处理(Waste Management)”。注意上位词的选择必须是与所表达的概念最邻近的词。另外在主副组配中若无法找到最专指的主题词，也要求选用最邻近的词。

④靠词检索。采用在含义上相似或相近的主题词来检索文献。有些词虽没有专指性的主题词，但却具有意义上接近的主题词概念。如“血管离断术抢救门脉高压引起的上消化道出血”的文献，其中“上消化道出血”这一主题概念用靠词标引法标引为“胃肠出血(Gastrointestinal Hemorrhage)”较合适。又如“红细胞存活(Erythrocyte Survival)”可标引为意义上接近的“红细胞衰老(Erythrocyte Aging)”。

三　数据库的类型与结构

计算机信息检索系统从物理构成上说，包括计算机硬件、软件、数据库、通讯线路和检索终端五个部分。一般而言，软件由计算机信息检索系统的开发商制作，通讯线路、硬件和检索终端只要满足计算机检索系统的要求均不需要用户多加考虑。对用户来说，需要了解的是数据库的类型和结构，以便根据不同的检索要求选择适合的数据库和检索途径。

1.数据库的概念　数据库(database)是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，是计算机技术与信息检索技术相结合的产物，是检索系统的信息源，也是用户检索的对象。

2.数据库的类型　对于数据库类型的划分有多种标准，依据数据库中存储的信息内容可以将其分为四种类型。

(1)书目数据库(bibliographic database)　书目数据库是文献检索中最为常见的一种数据库，它提供文献的外表特征和内容特征，如文献的题名、作者、文献出处(出版物名称、出版时间、卷期、页码)、关键词、摘要部分等信息。检索结果只提供文献的线索而非原始文献。如许多图书馆提供的基于网络的联机公共检索目录(Web-based Online Public Access Catalogue)、MEDLINE、《CA》(化学文摘)、《CSA Illumina》(剑桥科学文摘网络版数据库)、CBM等题录、文摘型数据库。

(2)全文数据库(full text database)　存储的是原始文献的全文，包括印刷版的电子版和纯电子出版物，如《中文科技期刊数据库》(维普)、《万方数据资源系统》《OVID》《EBSCO》等。全文数据库因其直接提供原始文献，免去了用户查询书目数据库后还需寻找原文的麻烦，是近年来发展迅速且前景广阔的一类数据库。

(3)数值和事实型数据库

①数值型数据库(numeric database)是以数值为主要内容的数据库，除存储各类数值如

科学技术数据、统计数据、实验数据、人口数据外，还存储运算公式、图谱、表格等。

②事实型数据库(fact database)直接提供可用的事实，“事实”既可以是有数字又有文字的统计资料，也可以是纯文字的知识资料或信息资料，还可以是一篇叙述性文献，如人物传记数据库、自然及社会资源统计数据库等。

数值和事实型数据库是建立在科学调查与研究的基础上，是科学研究成果的信息总汇。通过这些数据库，可以获得大量的信息，并在前人的基础上进行更为深入的研究。如美国国立医学图书馆编制的化学物质毒性数据库 RTECS，包含了 10 万余种化学物质的急慢性毒理实验数据。美国国立癌症研究院建立的美国医生数据咨询库 PDQ(Physician Data Query)，提供有关癌症治疗和临床实验最新研究进展的内容。通过美国国立分子生物学信息资源中心(NCBI)的 GenBank(GenBank 核酸序列数据库)，可以较完整地获得 DNA 的序列信息，同时，研究者也可以将测定的序列结果上报给该数据库加以认定、发表和交流。万方公司提供的《医药名人库》囊括我国著名的医学家、药物学家、生物学家及从事医药卫生管理和政策制定的科技负责人的全面信息。

3.数据库的结构　虽然数据库的类型较多，但其结构却基本相同。从用户的角度来看，数据库主要由文档、记录、字段三个层次构成。

(1)文档(file)　是数据库中一部分记录的有序集合。为了便于用户检索，一般将数据库所包含的数以万计的记录划分为若干个文档。文档又分为顺排文档(serial file)和倒排文档(inverted file)两种。

顺排文档是数据库的主体，又称主文档。顺排文档是将文献数据库中全部记录按一定顺序(记录顺序号从小到大)排列而成的文献记录集合。如果以顺排文档为单位检索，检索速度比较慢。倒排文档是将文献数据库中全部记录的全部文献特征标识按一定顺序排列而成的集合，即从记录中抽取有检索意义的数据，如主题词、关键词、著者姓名等文献信息特征作为检索标识，按一定顺序(字母顺序)排列而成的文档。故倒排文档在一个数据库中可以有若干个，如主题词索引、著者索引、刊名索引等。其目的是加快数据库的检索速度。在实际的检索过程中，计算机首先按用户键入检索词的字顺从指定的倒排文档中找到相匹配的索引词，然后再根据索引词后所附的记录顺序号到主文档(顺排文档)中调出所需记录内容。

(2)记录(record)　记录是由若干字段组成的构成文献数据库的信息单元。在全文数据库中，一条记录相当于一篇完整的文献。在书目数据库中，一条记录相当于一条题录。其他类型的数据库中，一条记录则代表一个信息单元。每条记录描述了一个原始信息(如文献、专著、专利说明书)的外表特征和内容特征。

(3)字段(field)　字段是比记录更小的单位，是构成记录的基本单元。书目数据库的记录含有题名、著者、出版物名称、出版时间、关键词、主题词、文摘等字段。

四　计算机检索途径与检索技术

(一)计算机检索途径

检索途径是指以记录的某一特征为检索切入点进行检索。计算机检索途径通常体现为字段检索，常用的计算机检索途径如下。

1.自由词检索　自由词又称文本词(text word)，是作者本人写文章时所使用的自然词语，包括标题词(title word)、关键词(keyword)、文摘词(abstract word)、全文词(full text

word)。自由词不受词表约束,同一概念用词取决于作者的偏爱。

2.主题词检索　用来表达文献的主题概念、经过规范化处理的词或词组。主题词检索是以主题词为检索标识，它是一种特性检索。外文数据库（如 MEDLINE,PubMed）一般采用 MeSH,中文数据库(如 CBM)对于西医方面的文献也采用 MeSH 标引文献。但有的数据库(如 CSA Illumina)拥有自己的主题词表。故检索时,切忌使用一个主题词同时进行多个数据库的检索。由于主题词是规范化的词,能在一定程度上避免漏检现象,故为最佳检索途径。在各学科及其分支交叉渗透日益严重的今天,越来越多的科技人员选择主题词途径来检索文献。

3.分类检索　根据文献内容在学科分类体系中的位置作为文献信息的检索途径,它的检索标识是所给定的分类号，是以学科概念上下左右的关系来反映事物的隶属和平行关系,能够满足检索者对文献族性检索的需要。中文数据库(如 CBM、中文科技期刊数据库)的文献一般按《中图法》进行分类。

4.作者检索　以文献署名作者或编者的姓名作为检索入口。当对某专业学科领域的某专家比较熟悉,想了解其最近研究的进展或新的成果,可通过作者途径来查找。查找方法比较简单,但要注意作者姓名的键入形式。检索时,作者的姓(last name，surname，family name)在前,名(first name，given name)在后,更多的情况是名只用首字母。欧美人姓名习惯是名在前,姓在后,所以检索时,必须进行姓与名的转换。一般数据库要求姓名之间用空格或逗号分隔。例如:"smith m""smith，m"均可(如 Elsevier SDOL)。在西文数据库中如果检索中国作者发表的文献,也是姓在前,名(拼音首字母)在后。作者检索有时会出现同名同姓但并非同人的情况,此时,可借助文献主题、期刊名称和作者单位等加以鉴别。

5.刊名检索　以期刊名称作为检索入口,检索特定期刊特定卷期次刊载的文献。有的数据库(如 SDOL、SpringerLink、万方数据资源系统等)提供刊名浏览,找到并点击刊名链接即可;有的则须键入期刊名称。外文期刊名的键入方法有刊名全称和刊名缩写两种情况。各个数据库对键入刊名的要求不尽相同。PubMed 要求键入期刊全称或 MEDLINE 的规范刊名缩写。SpringerLink 允许键入刊名的一部分，如键入 "cancer immu"，可检索到含有键入词的刊名 "Cancer Immunology，Immunotherapy"。对于刊名的缩写与全称转换把握不大的用户，可查 PubMed 中的 Journals Database 等辅助工具。

6.机构检索　以文献作者所在机构名称作为检索入口,检索该机构科研人员发表文献的情况。

7.引文检索　引文检索是以被引用文献即参考文献(references,cited paper)作为检索起点来查找引用文献即列有参考文献的文献(citing paper)的过程。一般以被引用文献的作者作为检索词,也有(如 EBSCO)用被引用文献所刊载的刊名或文献题名作为检索词进行引文检索的。由于被引用文献和引用文献在内容上或多或少有关联,所以,通过一个知名学者或一篇质量较高的文献进行引文检索,常常可以获得一系列主题相关、内容上有所继承发展的新文献。

8.限定检索(limit)　不少数据库(如 CBM、EBSCO、PubMed、CSA Illumina、SDOL)提供 "Limit"检索功能,用来对检出结果进行数量限制。常用的限定选项有:只检索有全文的文献(Full Text Only)、只检索包含摘要的文献 (Only Items with Abstracts)、提供参考文献(References Available)、期刊文献 (Journal Articles Only)、专家评审刊 (Scholarly(Peer Reviewed)Journals)、文献类型(Article Type，Document Type，Publication Types)、带有图像的文献(Articles with Images)、英文文献(English Only)、出版日期(Published Date)等选项。

9.默认检索(default)　默认检索又称缺省检索、隐含检索，是指在检索系统预先设定的多个字段中进行检索。各个数据库的默认检索字段不尽相同。例如：CBM 中的“缺省”检索是指在中文题名、摘要、作者、关键词、主题词和刊名 6 个主要字段中进行检索。

10.其他检索途径　数据库依其所收录的内容提供相应的检索途径。学位论文数据库(如 CNKI)提供学位授予单位、导师等检索途径；专利数据库(如万方)提供专利名称、发明人等检索途径；《CA》提供化学物质登记号(CAS registry Number)、分子式(Formula)等检索途径；《中文科技期刊数据库》(维普)在“期刊搜索”中提供国际标准连续出版物编号(ISSN)等检索途径。

(二)计算机检索技术

计算机信息检索过程实际上是将检索词与文献记录标引词进行对比匹配的过程。为了提高检索效率，计算机检索系统常采用一些运算方法，从概念相关性、位置相关性等方面对检索词进行技术处理。依据本教材所介绍主要数据库的检索技术并结合文献信息检索实践，介绍一些常用的信息检索技术和方法。

1.布尔逻辑检索(Boolean searching)　布尔逻辑检索是检索系统中应用最为广泛的检索技术，它采用布尔逻辑表达式来表明用户的检索要求。布尔逻辑表达式由检索词以及用于表达词与词之间逻辑关系的运算符构成。布尔逻辑运算符有三种：and、or、not。对逻辑运算符使用的技巧决定检索结果的满意程度。

(1)AND　称为逻辑“与”，表示“相交”关系，其含义是两个以上概念相交的部分即为所需部分。例如：检索式“A AND B”表示检索的文献必须同时包含 A 和 B 两个检索词。有的数据库用“*”表示逻辑“与”。其作用是缩小检索范围，提高查准率。

(2)OR　称为逻辑“或”，表示“并列”关系，其含义是两个以上概念的总和。例如：检索式“A OR B”表示检索的文献包含 A 或包含 B，或同时包含 A 和 B 两个检索词。有的数据库用“+”表示逻辑“或”。其作用是扩大检索范围，提高查全率。在用关键词检索时，需用“OR”将各同义词和近义词连接起来，避免漏检。

(3)NOT　称为逻辑“非”，表示“排斥”关系，其含义是检索出来的文献不包括“NOT”后面的检索词，其目的是将命中文献中不需要的文献排除出去，缩小检索范围。例如：检索“A NOT B”表示检索的文献只包含 A，排除不需要的检索词 B 或同时有 A 和 B 的检索词。有的数据库用“-”表示逻辑“非”。

在一个检索式中如果含有两个以上的逻辑运算符就要注意运算顺序：(　　)>NOT>AND>OR，即先运算括号内的逻辑关系，再依次运算“非”“与”“或”的关系。

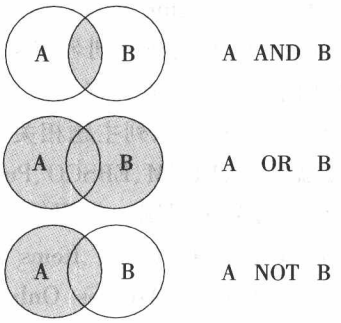

图 1-2　逻辑运算示意图