医学信息检索与利用

YIXUE XINXI JIANSUO YU LIYONG

范吉莲　张静昌　主编

第二军医大学出版社

医学信息检索与利用

主　编　范吉莲　张静昌

副主编　邱君瑞　彭　骏

编　者　（以姓氏拼音排序）

范吉莲　彭　骏

邱君瑞　徐　维

张静昌

第二军医大学出版社

内 容 提 要

本书旨在提供快速而高效地获取医学信息的技能与方法。根据当前信息环境的特点，以及医药卫生工作者的信息需求，本书融合了传统文献检索和现代信息检索的理论，以信息检索语言、计算机检索基础、互联网及其搜索引擎知识为基础，选择中外经典和重要的医学数据库进行系统论述，介绍了信息检索基础知识、图书馆资源利用、常用生物医学文摘数据库、全文数据库、引文检索、搜索引擎、常用医学网站、生物信息数据库、网上免费电子期刊、专利检索、会议信息检索、机构及个人信息搜索、循证医学信息检索、医学情报调研、综述文献写作等内容。

全书内容简明新颖，实用性强。既可作为医药院校研究生、本科生教学用书，又可作为医药卫生工作者案头参考书。

图书在版编目(CIP)数据

医学信息检索与利用/范吉莲，张静昌主编．上海：第二军医大学出版社，2006.8

ISBN 7-81060-593-3

Ⅰ.医… Ⅱ.①范… ②张… Ⅲ.医药学-情报检索-医学院校-教材… Ⅳ.G252.7

中国版本图书馆 CIP 数据核字(2006)第 066856 号

医学信息检索与利用

主　　编　范吉莲　张静昌

责任编辑　王　楠

第二军医大学出版社出版发行

上海市翔殷路 800 号　邮政编码:200433

发行科电话/传真:021-65493093

全国各地新华书店经销

上海第二教育学院印刷厂印刷

开本:787×1092　1/16　印张:21　字数:514.8千字

2006 年 8 月第 1 版　2006 年 8 月第 1 次印刷

印数　1-5000

ISBN 7-81060-593-3/G·049

定价：36 元

前　言

第二军医大学的《医学文献检索与利用》课程教学始于20世纪80年代，从最初面向研究生开设医学文献检索讲座，逐渐发展到面向本科生开设选修课、必修课。在2000年学校实行教改时，《医学信息检索与利用》课程实行分段教学，分为图书馆资源利用、医学信息检索和医学信息交流，并编写了各阶段的试用教材和检索实习手册。

近年来，医学文献数据库品种不断增加，数据库供应商之间并购与重组频繁，数据库操作系统更新加快，互联网上的医学信息资源也愈加丰富，亟需出版一本能够反映最新的医学信息检索与利用知识的教科书。正是在这样的背景下，《医学信息检索与利用》在原第二军医大学图书馆馆长耿亦兵主编的试用教材基础上，将原来的三本教材浓缩成一本，在做了充实和调整后正式出版。教材中新增了中外文全文数据库检索、网上医学专业信息搜索、循证医学信息检索等内容，删减的内容主要是书本式检索工具、网页制作等。

全书共分为11章。第一章"绪论"介绍了信息检索基础、检索语言、计算机检索原理、检索策略与步骤，是整个课程学习的基础知识。第二章"图书馆资源利用"内容简明易懂，可以自学。本校是安排在新生入学时进行教学，让大学生在第一时间了解图书馆资源，学会有效地利用图书馆。第三至第六章是中外数据库检索，是本课程学习的重点内容。第七章中的Google搜索引擎知名度高，检索功能强大，为学生必修内容。第八章的常用医学网站和医学专业信息搜索可选择讲解，留出一定时间让学生自学和上机实习。第九章"检索技能的综合应用"，是将前面所学习过的知识融会贯通，是学生在进行综合检索实习前的必修内容。第十章概述了循证医学及其证据检索。第十一章介绍了医学情报调研的基本方法、资料的搜集与整理、综述文献写作等内容。

第一、六、十章由邱君瑞编写，第二、三章由范吉莲编写（第三章第四节由徐维编写），第四、七、八、十一章由彭骏编写，第五、九章由张静昌编写。

本书编写历时半年，编写人员勤勉工作，孜孜不倦，付出了大量的辛勤劳动。虽经认真研究，反复推敲，但受编者学术水平所限，加之时间紧迫，其中的疏漏不当甚至错误之处恐怕在所难免，真诚希望专家学者批评指正。

本书在编写过程中，参阅了大量的国内外相关文献信息，在此向有关专家学者表示衷心感谢。本书的出版得到了第二军医大学出版社的大力支持，在此深表谢意。

编　者

2006年8月

目 录

第一章 绪 论

第一节 信息检索基础

自从人类进入文明社会以来，人们就一直在追求信息处理和信息传递能力的提高。从语言的出现到文字的发明，从印刷术的面世到无线电的应用以及互联网的出现和发展，每一次都是信息表达、信息存储以及信息传播手段上的重大变革。20世纪以来，随着科学技术的空前进步，信息、能源和材料已构成现代社会文明的三大支柱。我们随时随地都在自觉不自觉地接受、传递、存储和利用各种信息，毫无疑问，人类已经进入信息时代。

一、信息概述

（一）信息的内涵与外延

自20世纪50年代信息概念被正式提出以来，信息得到了广泛而深入的研究，但由于信息涉及的领域广、内容丰富，至今都没有形成统一的认识。已有的较具代表性的信息定义如下：①信息是客观世界中各种事物变化与特征的最新反映，是客观事物状态经过传递后的再现；②信息是实现事物间根据某种自然的规律和人为的约定建立联系的一种形式，是被表现出来的事物增添了的确定性或被取消了的不确定性；③信息是经过传递为接收者所理解，并对解决面临问题有用的、预先不知道的新报道和新知识；④信息是物质存在的一种形式，它是以物质的属性和运动状态为内容，并且总是借助于一定的物质载体传输和储存的。

目前较为普遍的信息定义为：信息是通过一定的物质载体形式反映出来的事物存在的状态、运动形式、运动规律及其相互作用的表征。

信息普遍存在于整个自然界，无处不在，无时不有，其外延是相当广泛的，宇宙万物所产生和彼此交流的内容都可成为信息，如病人体温的升高与下降是其体征信息；天空乌云密布是暴风雨来临的预示信息；敌军布防的情报是战斗决策信息；DNA的密码是生命繁衍所依靠的遗传信息；医学成果的论文是有关学术知识的文献信息。

本书将研究内容界定在文献信息的范畴内。

（二）信息的属性

虽然信息在不同的背景条件下有不同的理解，但信息一般都具有以下属性：

1. 客观性

信息不是虚无缥缈的事物，它的存在可以被人们感知、获取、传递和利用。信息是现实世界中各种事物运动与状态的反映，其存在是不以人的意志为转移的。客观性、真实性是信息最重要的本质特征。

2. 时效性

由于事物是在不断变化着的，那么表征事物存在方式和运动状态的信息也必然会随之改变。在现代社会中，信息的使用周期越来越短，信息的价值实现取决于对其把握和运用的及时

性。如果不能及时地利用最新信息，信息就会贬值，甚至毫无价值，这就是信息的时效性，即时间与效能的统一性。它既表明了信息的时间价值，也表明了信息的经济价值。

3. 载体性

信息必须依附于一定的载体（如声波、电磁波、纸张、化学材料、磁性材料等）才能流通和传递，否则，信息的价值就不能体现。信息可以存储在不同的载体上，但其内容并不因记录手段或物质载体的改变而发生变化。例如关于医疗机械的信息，不论是刊登在报刊上、发布在电视节目中，还是存储在光盘数据库中，其信息内容和价值是同样的。

4. 传递性

信息依附于一定的物质载体后，其传递和流通便成为可能。信息的传递性是指信息从信源出发，经过信息载体的传递，被信宿接收并进行处理和利用的特性。不同载体的信息可以通过计算机、人际交流、文献交流或大众传媒等手段传递给信息用户，这种跨越时空的传递特性是实现信息资源共享的基础，是将信息利用最大化的保证。

5. 可塑性

信息在流通和使用过程中，人们借助于先进的技术，可以对其进行综合、分析及加工处理。也就是把信息从一种形式变换成另一种形式，如可以将一本图书加工为题录或文摘等形式，从而方便用户的选择和利用。不过，在信息的加工过程中，信息量会减少或增加。用户可根据检索需要选择不同的信息形式。

6. 共享性

共享性是指同一信息同时或不同时被多个用户使用，而信息的提供者并不因此而失去信息内容和信息量。信息的共享性可以提高信息的利用率，人们可以利用他人的研究成果进一步创造，避免重复研究，可以节约资源。

（三）信息运动的基本方式

1. 定向运动

在人为控制下，信息按人类社会的组织结构以一定的方式，沿着固定的方向运动。比如在部门和组织内部都规定有各自的定向方式的信息传递系统，以实现信息的上情下达和下情上达。

2. 辐射运动

信息生成后便以辐射方式向周围传播和扩散，比如通过广播、电视、报刊杂志等。辐射方式的信息运动，效果是范围广、速度快，它是信息社会化传播的主要方式，具有较强的社会影响力。

3. 无序运动

信息在人类社会的作用下有时往往作无规则、无休止的运动。比如人际交谈、新闻报道等，信息内容越是独特，与人们的现实生活关系越是密切，其无序运动越是强烈。信息的无序运动往往会影响个人的声誉和社会的安定。信息工作人员往往也是从信息的无序运动中收集所需信息。

二、信息源

信息广泛存在于自然界、生物界和人类社会中。随着科学技术的进步，信息的表现形式呈多样化态势。了解信息的不同形式不仅有助于我们加深对信息内涵及其特征的认识，也能为

检索和利用信息打下坚实的基础 。

（一）信息的载体形式

信息可通过不同的手段记录、存储在不同的载体中，按其载体形式可分为印刷型、缩微型、声像型及电子型信息。

1. 印刷型信息

印刷型信息又称书本型信息。它是以纸张为载体、以印刷为记录手段而产生的一种传统的信息形式，如图书、期刊、报纸、印刷型的检索工具等。其优点是便于阅读和流通，符合人们的阅读习惯；缺点是存储密度低，收藏和管理需要较大的空间和人力。

2. 缩微型信息

为了弥补印刷型信息的不足，缩微型信息应运而生。它是一种以缩微胶片或平片为载体，利用缩微摄影技术为记录手段而产生的信息形式。随着激光和全息摄影技术的应用，又出现了超级缩微胶片和特级缩微胶片，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，保存期长，便于收藏和管理；缺点是必须借助缩微阅读机才能阅读。

3. 声像型信息

声像型信息又称视听资料。这是一种以磁性和感光材料为存储介质，借助特殊的机械装置直接把图像和声音记录下来的一种信息形式。主要载体有录音带、唱片、激光唱盘、录像带、电影胶片、幻灯片等。其优点是既能闻其声又能观其像，直观、亲切，表现力强。与印刷型信息相比，声像型信息更能提高人们理解信息的能力。

4. 电子型信息

电子型信息是指以数字代码方式将图、文、声、像等信息存储到磁、光、电介质上并通过计算机阅读的信息，如各种电子图书、电子期刊、联机数据库、网络数据库、网络新闻、光盘数据库等。该类信息在计算机与网络技术的支持下，通过编码和程序设计，将信息变为数字语言和机器语言并存储在磁带、光盘、磁盘等介质上，从而建立起相应的文献数据库。其特点是存储量大，出版周期短，传递迅速，存取速度快，可以融文本、图像、声音等多媒体信息于一体，易复制，共享性好。

随着计算机技术与通信技术的发展与融合，又产生了一种依托于新型载体的文献信息源，这就是多媒体型(Multi-media)信息。多媒体即多种信息媒体，它采用计算机、通信、数字、超文本(Hypertext)或超媒体(Hypermedia)技术，不仅实现了文字、图像、动画、声音等的多位一体及人机交互对话，而且使全球信息共享成为可能。多媒体型的文献信息源实际上是以上数种载体形式的混合型，是一种立体式的信息源。

（二）信息的级别

依内容性质和加工程度的不同，信息可分为以下 3 个级别：

1. 一次信息

一次信息又称原始信息。它是指著者以本人的研究成果为依据撰写并公开发表或出版的信息。主要包括专著、期刊论文、科技报告、会议论文、专利说明书、学位论文等。一次信息是检索的主要对象。信息检索的最终目的就是查找到最适用的一次信息。

2. 二次信息

一次信息的数量极为庞大，在内容上是分散的、无系统的，也就是“无序”的，不便于管理和利用。为了方便用户选择和利用，信息工作者对一次信息进行再加工，通过整理、提炼和浓缩，

并按其外部特征(如题名、著者等)或内容特征(如分类号、主题词等)将其"有序化",形成另一类新的信息形式,如印刷型的目录、索引、文摘、题录或电子型的书目数据库、文摘数据库及题录数据库等就属于二次信息。通过阅读二次信息可以快速地了解一次信息的大致内容,选择并查找所需的一次信息。信息检索主要讲述的就是二次信息的编排体系和使用方法。

3. 三次信息

利用二次信息,选择有关的一次信息加以分析、综合而编纂出的第三层次的信息形式为三次信息,如专题报告、综述,以及词典、手册、百科全书、年鉴等工具书。三次信息具有系统性、综合性、知识性和概括性的特点。因此,要在浩瀚的信息中查找所需的特定的一次信息,往往离不开二次和三次信息。

(三)信息的类型

根据出版形式的不同,信息可以划分为以下类型:

1. 图书

图书是论述或介绍某一领域知识的出版物,常见的有专著、教科书、科普读物、词典、手册、百科全书等。图书往往是著者在收集大量第一手资料基础上,经分析归纳后编写而成的。其特点是内容比较系统、全面、成熟、可靠,但出版周期较长,报道速度相对较慢。图书主要用于需对大范围的问题获得一般性的知识或对陌生的问题需要初步了解的场合。

2. 期刊

期刊一般是指名称固定、出版形式一致的定期或不定期连续出版物。期刊论文内容新颖,报道速度快,信息含量大,是传递信息、交流学术思想最基本的文献形式,可以最快地反映最新研究成果。期刊是科技人员获取信息的主要来源,受到高度重视。大多数检索工具(或数据库)也以期刊论文作为报道的主要对象。对某一问题需要深入了解时,较普遍的办法是查阅期刊论文。

3. 会议论文

会议论文是指在国际或国内重要的学术或专业性会议上发表的论文,其学术性强,往往代表着某一领域内的最新成就,反映了国内外科技发展水平和趋势,是获得最新情报的一个重要来源。

4. 科技报告

科技报告是指国家政府部门或科研单位关于某项研究成果的总结报告,或是研究过程中的阶段性进展报告。在内容方面,报告比期刊论文等更为专深、详尽与可靠,是一种不可多得的情报源。科技报告可分成技术报告、技术备忘录、札记、通报等几种类型。有些报告因涉及尖端技术或国防问题等,所以又分绝密、秘密、内部限制发行和公开发行几个等级。目前国际上较著名的科技报告是美国政府的四大报告,即 PB、AD、NASA 和 DOE 报告。

5. 专利文献

专利文献是实行专利制度的国家及国际性专利组织在审批过程中产生的官方文件及出版物的总称,它通常以专利说明书、公报、文摘、索引、权利说明书等形式出版。所谓专利说明书是指专利申请人向专利局递交的有关发明目的、构成和效果的技术文件。专利说明书的内容比较具体,有的还有附图,通过它可以了解该项专利的主要技术内容。由于只有符合新颖性、创造性和实用性的发明创造才能获得专利权,所以专利说明书对于技术人员,特别是医疗仪器的研究人员来说,是一种切合实际、启迪思维的重要情报源。

6. 标准文献

标准文献是指标准化工作的文件。作为一种规章性文献，它具有一定的法律约束力。标准文献是一种很重要的信息源，任何一个国家的标准文献都反映着该国的生产技术水平和经济实力，而国际现行标准则代表了当前世界水平。国际标准和工业先进国家的标准常是科研生产活动的重要依据和情报来源。国际上最重要的标准化组织是国际标准化组织(ISO)。

7. 学位论文

学位论文是指为申请硕士、博士等学位而提交的学术论文。学位论文的质量参差不齐，但都是就某一专题进行研究而作的总结，多数有一定的独创性。

8. 技术档案

技术档案是指在科研生产活动中形成的，有具体事物的技术文件、图纸、图表、照片和原始记录等。详细内容包括任务书、协议书、技术指标、审批文件、研究计划、方案大纲、技术措施、调查材料、设计资料、试验等。这些材料是科研工作中用以积累经验、吸取教训的重要文献。技术档案一般为内部使用，不公开出版发行，有些有密级限制，因此在参考文献和检索工具中极少引用。

9. 政府出版物

政府出版物是指各国政府部门及其设立的专门机构发表的文献。它的内容十分广泛，既有科学技术方面的，也有社会经济方面的。就文献的性质而言，政府出版物可分为行政性文件(如国会记录、政府法令、方针政策、规章制度以及调查统计资料等)和科学技术文献两部分。

除以上介绍的信息类型外，信息的出版形式还包括产品资料、报纸、计算机软件等。

三、信息检索

(一) 信息检索的意义和作用

1. 信息检索是信息素质教育的主要内容

21 世纪是经济信息化、社会信息化的时代。终身教育、开放教育、能力导向学习成为教育理念的重要内涵。为满足知识创新和终身学习的需求，发达国家和地区纷纷将信息素质或信息能力教育作为 21 世纪人才能力的重要内容。目前，美国从小学、中学到大学都已全面将信息素质纳入正式的课程设置中。信息素质是一个带根本性的、重要的教育议题，是未来信息社会衡量国民素质和生产力的重要指标。

信息素质有其自身的内容结构，包括信息意识、信息能力和信息道德。信息意识是指人对各种信息的自觉反应；信息能力包括信息技术应用能力，信息查询、获取能力，信息组织、加工和分析能力；信息道德是指整个信息活动中的道德，是调节信息生产者、信息加工者、信息传递者及信息使用者之间相互关系的行为规范的总和。

通过信息检索知识的系统学习，学生能意识自身潜在的信息需求，并将其转化为现实的信息需求，进而能充分、正确地表达出来。而且会拥有对信息的查询、获取、分析和应用能力，对信息进行去伪存真，去粗取精，提炼、吸取符合自身需要的信息。可见，信息检索是当代学生必须具备的能力，是学生信息素质教育的重要内容。

2. 信息检索是科学研究的重要环节

一个科技工作者创新成果的多少，一个科研项目科技水平的高低，都与其开发、占有和利用信息资源的能力大小息息相关。因为科学研究具有连续性和继承性，没有继承就没有创新。正如伟大的科学家牛顿所说："如果我比别人看得远些的话，那是因为我站在巨人的肩膀上。"

这句名言极其深刻地概括了科学研究的连续性和继承性的道理。

信息检索是科学研究的重要环节。科技工作者在科学研究中，从选题、立项、试验、撰写研究报告、研究成果鉴定到申报奖项，每一个环节都离不开信息检索。据统计，科研人员在整个研究过程中，查阅文献信息的时间要占全部科研时间的40%左右。只有大量搜集、整理、分析与利用信息，才能弄清楚古今中外进行过哪些研究、运用什么理论、采用何种方法、取得什么成果、达到何种水平，哪些研究领域还没有涉及，哪些研究项目具有可行性、重要性和发展前景。掌握了这些信息，首先可以了解国内外科技发展水平与动向，利用已有的科研成果，避免重复他人的劳动，把自己的研究工作建立在一个较高的起点上；其次，通过信息这一智慧的火种，可以使科研人员开阔视野，发展思路，启迪创造力，开拓更新的、更高层次的、更广阔的研究领域；再次，掌握信息检索技术与方法，可以大大提高信息检索效率，为科研工作赢得大量宝贵时间，缩短科研周期，加速科研进程，创造出更多的具有高附加值的技术成果。

（二）信息检索的含义

信息检索（Information Retrieval）是指将信息按一定方式组织和存储起来，并针对用户的需求找出所需信息的过程，所以又称为信息存储与检索。对于信息专业人员来说，信息检索是指对分散的、无序的、海量的信息，进行组织、整理、加工和存储，建立可供检索的检索工具的过程；对于信息用户来说，信息检索是指从检索工具中查找所需要信息的过程。信息检索的目的是为了解决特定的信息需求和满足信息用户的需要。

（三）信息检索的原理

广义的信息检索包括信息的存储和检索两个过程。信息的存储就是将搜集到的一次信息，经过著录其特征（如题名、著者、主题词、分类号等）而形成款目，并将这些款目组织起来成为二次信息的过程。信息的检索是针对已存储好的二次信息库进行的，是存储的逆过程。存储是为了检索，而为了快速而有效地检索，就必须存储。没有存储，检索就无从谈起。这是存储与检索相辅相成、相互依存的辩证关系。但存储与检索所依据的规则必须一致，也就是说，存储者与检索者必须遵守相同的规则。

信息存储与检索共同遵循的规则称之为信息检索语言（详见本章第二节）。只要存储者和检索者用同一种规则来标引要存入的信息特征和要查找的检索提问，使它们变成一致的标识形式，当检索提问标识和信息特征标识一致时，相关的信息就能被检索出来。

可见，信息检索正是以信息的存储与检索之间的相符性为基础的，如图1-1-1所示。如果两个过程不能相符，那么信息检索就失去了基础。同样，要是检索不到所需的信息，存储也就失去了意义。

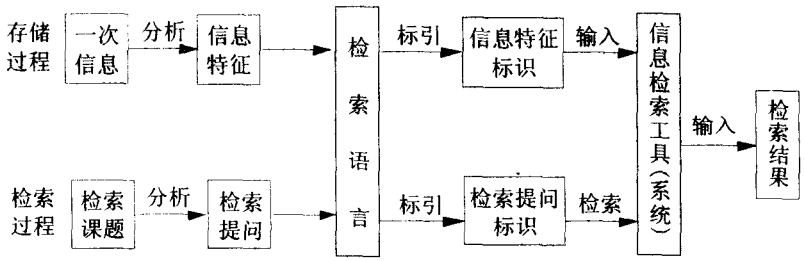

图1-1-1　信息检索原理图

（四）信息检索语言

从图 1-1-1 可以看出，用户要检索到适当的信息，关键是要使自己使用的提问标识与检索系统使用的检索标识一致。二者达到统一的关键在于使用相同的检索语言。因此，检索语言在信息检索中起着至关重要的作用。

所谓检索语言，是指在信息检索过程中用来描述信息特征和表达信息需求提问的一种专门的语言，它是掌握信息检索技能所必须具备的知识。

（五）信息检索的类型

根据检索内容的不同，信息检索可以分为：

1. 数据检索

数据检索是以特定的数值性数据为检索对象，包括各种统计数字、图表、化学结构式、计算公式等，如某种药物的理化常数、常用剂量、结构式等。

2. 事实检索

事实检索即通过对存储的文献中已有的基本事实进行检索，或对数据进行处理（逻辑推理）后得出新的事实的过程，如我国首例艾滋病是何年发现的？我国艾滋病的发病率是多少？

3. 文献检索

文献检索是以文献为检索内容的检索，包括书目检索和全文检索两种。书目检索是以文献线索为检索内容的信息检索；全文检索以文献所含的全部信息作为检索内容，同时要求检出文献的全文，是计算机信息检索的发展方向。

目前，文献仍是存储信息的主要形式。因此，人们对数据和事实的检索，在很多情况下都要借助于文献检索。数据和事实检索是要检索出包含在文献中的具体信息；文献检索则是要检索出包含所需要信息的文献。文献检索是最典型和最重要也是最常利用的信息检索。掌握了文献检索的方法就能以最快的速度，在最短的时间内，以最少的精力了解前人和别人取得的经验和成果。

4. 图像检索

图像检索即以图形、图像或图文信息为检索内容的信息检索。

5. 多媒体检索

多媒体检索是以文字、图像、声音等多媒体信息为检索内容的信息检索。

第二节 检索语言

一、检索语言及其作用

（一）检索语言的概念

检索语言是应信息的加工、存储和检索的共同需要而编制的专门语言，是表达一系列概括信息内容和检索课题内容的概念及其相互关系的一种概念标识系统。简言之，检索语言是用来描述信息源特征和进行检索的语言，可分为规范化语言（人工语言）和非规范化语言（自然语言）两类。

（二）检索语言的作用

检索语言在信息检索中起着极其重要的作用，它是沟通信息存储与信息检索两个过程的

桥梁。在信息存储过程中，用它来描述信息的内容和外部特征，从而形成检索标识；在信息检索过程中，用它来描述检索提问，从而形成提问标识；当提问标识与检索标识完全匹配或部分匹配时，结果即为命中文献（详见本章第一节的检索原理图）。

检索语言的主要作用如下：①标引信息内容及其外表特征，保证不同标引人员表征信息内容的一致性；②对内容相同及相关的信息加以集中或揭示其相关性；③使信息的存储集中化、系统化、组织化，便于检索者按照一定的排列次序进行有序化检索；④便于将标引用语和检索用语进行相符性比较，保证不同检索人员表述相同信息内容的一致性，以及检索人员与标引人员对相同信息内容表述的一致性；⑤保证检索者按不同需要检索信息时，都能获得最高查全率和查准率。

二、检索语言类型

（一）人工语言

对于高要求的检索来说，控制是绝对必要的。人工语言就是对概念及其标识系统实施严格规范的检索语言。人工语言按其结构原理可分为分类语言、主题语言、代码语言3种类型。

1. 分类语言

分类语言是将表示各种知识领域（学科及其研究问题）的类目按知识分类原理进行系统排列并以代表类目的数字、字母符号（分类号）作为信息主题标识的一类信息检索语言。分类语言的主要特点是按学科、专业集中信息，并从知识分类角度揭示各类信息在内容上的区别和联系，提供从知识分类检索信息的途径。它犹如一张知识地图，能够使检索者沿着大大小小知识领域的隶属并列关系，找到所要到达的目的地——记载着所需要的知识的文献信息。

分类语言分为两种类型：应用概念划分与概括的逻辑方法构成的体系等级分类语言以及应用概念的分析与综合原理构成的组配分类语言。

分类语言是信息工作中运用特别广泛的信息组织方法，最近几年在网络信息资源组织中的作用也逐渐得以体现，其中体系分类语言是最常用的。它的最大优点是具有按学科或专业集中地、系统地揭示信息内容的功能，可使检索者鸟瞰一个学科或专业信息的全貌，并可触类旁通。这对于系统地掌握和利用一个学科或专业范围的信息是很方便和有效的。由于人们一般都是在某个专业范围内从事科研、生产、教学、管理等活动的，比较习惯于从学科、专业出发去获取信息，所以它的这一优点特别重要。

目前，国内外最常用的分类语言包括《中国图书馆图书分类法》、《杜威十进分类法》、《国际十进分类法》、《美国国会图书馆图书分类法》等。

2. 主题语言

主题语言就是以自然语言的语词经过规范处理（采取人工控制措施）后直接作为信息内容主题标识。主题标识按字顺排列，并用参照系统和其他方法来间接地显示概念之间的关系。主题语言的特点是按事物集中信息，用参照系统等方法间接显示概念或事物之间的关系，提供事物名称的字顺检索途径。检索者只要确知他所需要检索对象事物的名称，就可从主题检索系统的字顺中直接查出该对象事物及其各方面的有关信息，并可通过参照系统等扩大检索范围。

主题语言包括标题法、单元词法、叙词法、关键词法、自由标引法等。一般来说，用主题语言组织与揭示信息具有直接和直观的特点，即主题语言用于组织信息不仅具有“直呼其名、依名查检”的直接性，而且其标识基本上是独立完整的事物概念。在网络环境中，主题语言检索

系统得到发展与完善，尤其是关键词检索，在网络中的应用相当广泛，有相当一部分网络信息资源浏览器与搜索引擎都以关键词为组织与揭示信息的重要途径与方法，如 Google、Baidu、Yahoo 等。

目前，国内外常用的主题语言有《美国国会图书馆标题表》、《医学主题词表》、《汉语主题词表》等。

3. 代码语言

代码语言是指对事物的某方面特征，用某种代码系统来表示和排列事物概念，从而提供检索的检索语言。例如，根据化合物的分子式这种代码语言，可以构成分子式索引系统，允许用户从分子式出发，检索相应的化合物及其相关的信息。

（二）自然语言

我们通常所说的自然语言指人们日常说话、写文章和思想交流所用的各种语言，而信息检索中的自然语言是指文献作者或文摘、提要的作者原来使用的语言。检索系统直接使用不经过控制的自然语言中的语词作标识，进行信息资源的标引和检索。

使用自然语言标引和检索的实践，可以追溯到我国唐代类书的编制和西方 13～14 世纪的圣经语词索引。作为一种标引和检索方法得到社会广泛使用，则是在计算机出现以后。20 世纪 50 年代后期，美国卢恩等人首先将计算机用于关键词索引的编制，其后，各种直接以自然语言为标识的检索系统也随之出现。这种检索系统以各种类型的电子文本为基础，一般不对词汇进行控制，或只进行少量控制，因此处理速度快、成本低。70 年代后自然语言迅速发展，与受控语言并驾齐驱。随着电子文本的使用日益广泛和网络的出现，目前，这种方法已逐步发展为主要的检索方式。

目前用于标引和检索的自然语言主要有以下几种形式：关键词法、文本检索、自由标引、自然语言入口词检索和自动标引。此外，还有按词频统计计算方式实施的抽词标引。自动赋词标引、自动分类等则是在自动抽词的基础上，依据自然语言语词与控制词、分类号对应表和转换规则等，将自然语言的语词转换成规范化的标引词和分类号。这两种标引方式已经超出了单纯自然语言的范围，是自然语言与人工语言的结合。这类标引在计算机检索环境下使用时，一般可以同时以自然语言语词或规范词的形式进行检索查找。自动标引目前仍处于探索阶段。为了提高自动标引质量，一些系统往往采用人机结合的方式，对计算机实施的抽词标引加以人工鉴别，调整补充。汉字由于语词之间不像西文那样留有空格，计算机无法自动识别，进行抽词标引前一般须先解决分词问题。

计算机检索系统中最常用的文本检索通常可以有两种实现方式：一是将所有的基本词记入倒排档，检索实际上不是在文本中进行，而是在索引中进行。这类系统一般使用停用词排除无检索意义的语词，将停用词表以外的词收入索引。二是不建立倒排档，直接对检索文本逐字顺序匹配。为了便于用户使用，满足用户在检索中可能出现的各种查全、查准的需求，文本检索系统发展了一系列检索的技术方法。常用的文本检索技术包括布尔逻辑检索、截词检索、精确检索、限定范围检索等（详见本章第三节）。

三、中国图书馆图书分类法

（一）简介

中国图书馆图书分类法，简称中图法，是由北京图书馆、中国科学技术情报所等单位于

1971年共同编制完成的，1974年出版，并经过多次修订与再版，1999年3月第四版出版。

中图法是体系等级分类语言，它采用五分法，即把知识学科分成5个基本部类："马克思主义、列宁主义、毛泽东思想、邓小平理论"、"哲学、宗教"、"社会科学"、"自然科学"和"综合性图书"。在各基本部类下再展开，共形成22个大类，用1个英文大写字母表示一级类目名（工业技术类除外），并以字母的顺序反映大类的顺序。根据各学科的内容需要，大类下再依次分二级、三级、四级、五级子类目，各子类目用阿拉伯数字表示。

中图法是国内最有影响力，使用最广泛的文献信息组织工具，国内大多数图书馆、情报机构使用中图法。国内也有一些大型书目数据库使用中图法作为组织数据的工具，并提供分类检索途径，如《中国生物医学文摘数据库》（CBMDisc）、清华同方的《中国学术期刊全文数据库》、维普的《中国科技期刊数据库》等。目前图书情报界的学者们正致力于中图法在网络信息资源组织中的应用研究。

（二）中图法的医药卫生大类

中图法用"R"代表"医药卫生"大类。该大类又分为17个基本类目（二级类目）：

R1	预防医学、卫生学	R74	神经病学与精神病学
R2	中国医学	R75	皮肤病学与性病学
R3	基础医学	R76	耳鼻咽喉科学
R4	临床医学	R77	眼科学
R5	内科学	R78	口腔科学
R6	外科学	R79	外国民族医学
R71	妇产科学	R8	特种医学
R72	儿科学	R9	药学
R73	肿瘤学		

每一个二级类目，又分为若干三级、四级、五级类目，下面是"R2 中国医学"的类目划分：

（三级类）		（四级类）		（五级类）	
R21	中医预防、卫生学	R241	中医诊断学	R246.1	内科
R22	中医基础理论	R242	中医治疗学	R246.2	外科针刺麻醉法
R24	中医临床学	R245	针灸学、针灸疗法	R246.3	妇产科
R25	中医内科	R246	针灸疗法与临床应用	R246.4	小儿科
R26	中医外科	R247	其他疗法	R246.5	肿瘤科
R271	中医妇产科	R248	中医护理学	R246.8	五官科
R272	中医儿科	R249	医案医话	R246.82	眼科
R273	中医肿瘤科			R246.83	口腔科
R274	中医骨伤科			R246.9	其他

四、医学主题词表

医学主题词（Medical Subject Headings，MeSH）是美国国立医学图书馆（NLM）编制的医学领域内权威的主题语言，在医学信息检索系统中得到了广泛的应用，如世界上最大的生物医学数据库Medline是由该词表索引和组织文献数据的；国内最大的中文生物医学文献数据库CBMDisc也用MeSH组织文献；美国NLM用MeSH组织馆藏；国内大多数医学图书馆和情